www.tnschoolbooks.in

GOVERNMENT OF TAMIL NADU

STATISTICS

HIGHER SECONDARY FIRST YEAR

Untouchability is Inhuman and a Crime

A publication under Free Textbook Programme of Government of Tamil Nadu

Department Of School Education

11th_Statistics_FM.indd 1 13/12/2021 20:08:16

www.tnschoolbooks.in

Government of Tamil Nadu

First Edition - 2018

Revised Edition - 2019, 2021, 2022

Reprint - 2020, 2023, 2024

(Published under New Syllabus)

NOT FOR SALE

Content Creation

State Council of Educational Research and Training

© SCERT 2018

Printing & Publishing

Tamil NaduTextbook and Educational Services Corporation

www.textbooksonline.tn.nic.in

11th_Statistics_FM.indd 2 22/12/2023 15:35:10

www.tnschoolbooks.in

CONTENTS

STATISTICS

Chapter No	Title	Page No	Month
Chapter 1	Scope of Statistics and Types of Data	01	June
Chapter 2	Collection of Data and Sampling Methods	17	June
Chapter 3	Classification and Tabulation of Data	43	July
Chapter 4	Diagrammatic and Graphical Representation of Data	72	July
Chapter 5	Measures of Central Tendency	110	August
Chapter 6	Measures of Dispersion	157	August/ September
Chapter 7	Mathematical Methods	191	October
Chapter 8	Elementary Probability Theory	218	October
Chapter 9	Random Variable and Mathematical Expectation	253	November
Chapter 10	Probability Distributions	288	November/ December

KJSF6

E-book

UAZOC

Assessment

11th_Statistics_FM.indd 3 13/12/2021 20:08:16

KEY FEATURES OF THE BOOK

Profile of a Statistician	Presents a brief history and contribution of a statistician
Learning Objectives	Goals to transform the classroom processes a learner centric
DO YOU KNOW?	Amazing facts, Rhetorical questions to lead students to Statistical inquiry
Note	Additional inputs to content is provided
Activity	Directions are provided to students to conduct activities in order to explore, enrich the concept
Infographics	Visual representation of the lesson to enrich learning
ZNYPS	To motivate the students to further explore the content digitally and take them to virtual world
Success Story	Success Stories given as a source of inspiration
Points to Remember	Summary of each lesson is given at the end
ICT	To enhance digital skills among students
Evaluation	Assess students to pause, think and check their understanding
Glossary	Explanation of scientific terms

11th_Statistics_FM.indd 4 13/12/2021 20:08:16

Career in Statistics

After completion of Higher Secondary Course, the subject Statistics is an essential part of the curriculum of many undergraduate, postgraduate, professional courses and research level studies. At least one or more papers are included in the Syllabus of the following courses:

Under Graduate Courses	Post Graduate Courses	Competitive Eaminations
B.A.(Economics)	M.A.(Economics)	UPSC
B.Com	M.Com	TNPSC
B.B.A	M.B.A	Staff Selection Commission Examinations
B.C.A	M.C.A	I.A.S
B.Sc.(Maths)	M.Sc	I.F.S
B.Pharm	M.Pharm	and many more
B.Ed	M.Ed	
B.Stat	M.Stat	
B.E	M.E	
Diploma Courses	C.A	
	I.C.W.A	
	Actuarial science	

Specialized fields in Statistics : Colleges/universities, Indian Statistical Institute(ISI) offer a number of specialisations in statistics at undergraduate, postgraduate and research level. A candidate with bachelor's degree in statistics can also apply for Indian Statistical Services (ISS).

Job Titles	Job Areas
• Statisticians	• Census
• Business Analyst	• Ecology
• Mathematician	• Medicine
• Professor	• Election
• Risk Analyst	• Crime
• Data Analyst	• Economics
• Content Analyst	• Education
• Statistics Trainer	• Film
• Data Scientist	• Sports
• Consultant	• Tourism
• Biostatistician	
• Econometrician	

Skills Required for a statistician

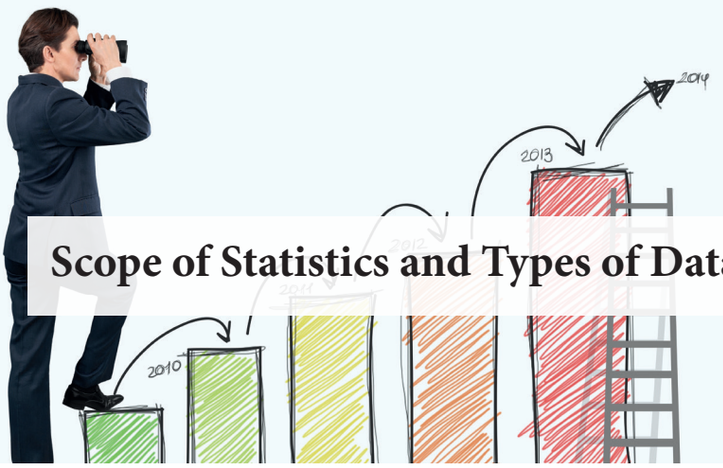
- Strong Foundation in Mathematical Statistics
- Logical Thinking & Ability to Comprehend Key Facts
- Ability to Interact with people from various fields to understand the problems
- Strong Background in Statistical Computing
- Ability to stay updated on recent literature & statistical software
- Versatility in solving problems

11th_Statistics_FM.indd 5 13/12/2021 20:08:16

www.tnschoolbooks.in

Chapter

1



Scope of Statistics and Types of Data



Prsanta Chandra Mahalanobis
(29 June, 1893-28 June, 1972)

P C Mahalanobis, known as father of Indian Statistics, received his early schooling at the Brahmo Boys School and graduation in Physics in 1912 through Presidency college, Kolkata. He left for England in1913 for higher studies at University of London. Mahalanobis was introduced to the journal Biometrika, that motivated him for the interest in statistics. A statistical laboratory was developed by him at Presidency College and later this formed the foundation for the famous Indian Statistical Institute at Kolkata. One of his important contribution was in designing and conducting large scale surveys. He introduced the concept of pilot surveys and advocated the usefulness of sampling methods.

DO YOU KNOW?

June 29th the birth anniversary of **P.C. Mahalanobis** is commemorated as the National Statistics Day of India.

'Statistics is the grammar of science' -*Karl Pearson*

Learning Objectives

- Highlights the origin and growth of statistics
- Introduces the meaning and definitions of statistics
- Presents the scope and functions of statistics
- Explains the applications of statistics in different fields
- Introduces the meaning of data
- Distinguishes the different types of data

UZMDV

Introduction

In this unit, we present the meaning of statistics, various definitions, origin and growth, functions, scope and applications to different fields such as Agriculture, Economics and many more. We also define 'data', various types of data and their measurement scales.

Ch1_11th Stats.indd 1 07/12/2021 11:42:14

1.1 Origin and Growth of Statistics

The origin of statistics can be traced back to the primitive man, who put notches on trees to keep an account of his belongings. During 5000 BCE, kings used to carry out census of populations and resources of the state. Kings of olden days made their crucial decisions on wars, based on statistics of infantry, cavalry and elephantary units of their own and that of their enemies. Later it enhanced its scope in their kingdoms' tax management and administrative domains. Thus, the word 'Statistics' has its root either to Latin word 'Status' or Italian word 'Statista' or German word 'Statistik' each of which means a 'political state'. The word 'Statistics' was primarily associated with the presentation of facts and figures pertaining to demographic, social and political situations prevailing in a state/government. Its evolution over time formed the basis for most of the science and art disciplines. Statistics is used in the developmental phases of both theoretical and applied areas, encompassing the field of Industry, Agriculture, Medicine, Sports and Business analytics.

DO YOU KNOW?

The word 'data' was first used in 1640's. In 1946, the word 'data' also meant for "transmittable and storable computer information". In 1954, a term called 'data processing' was introduced. The plural form of 'datum' is 'data'. It also means "given" or "to give" in Latin.

In olden days statistics was used for political- war purpose. Later, it was extended to taxation purposes. This is evident from Kautilya's Arthasastra (324 – 300 BCE). Akbar`s finance minister Raja Thodarmall collected information regarding agricultural land holdings. During the seventeenth century, statistics entered in vital statistics, which is the basis for the modern day Actuarial Science. Gauss introduced the theory of errors in physical sciences at the end of eighteenth century.



Statistics is concerned with scientific method for collecting, organizing, summarizing, presenting, analyzing and interpreting of data. The word statistics is normally referred either as numerical facts or methods.

Statistics is used in two different forms-singular and plural. In plural form it refers to the numerical figures obtained by measurement or counting in a systematic manner with a definite purpose such as number of accidents in a busy road of a city in a day, number of people died due to a chronic disease during a month in a state and so on. In its singular form, it refers to statistical theories and methods of collecting, presenting, analyzing and interpreting numerical figures.

Though the importance of statistics was strongly felt, its tremendous growth was in the twentieth century. During this period, lot of new theories, applications in various disciplines were introduced. With the contribution of renowned statisticians several

Ch1_11th Stats.indd 2 07/12/2021 11:42:14

theories and methods were introduced, naming a few are Probability Theory, Sampling Theory, Statistical Inference, Design of Experiments, Correlation and Regression Methods, Time Series and Forecasting Techniques.

In early 1900s, statistics and statisticians were not given much importance but over the years due to advancement of technology it had its wider scope and gained attention in all fields of science and management. We also tend to think statistician as a small profession but a steady growth in the last century is impressive. It is pertinent to note that the continued growth of statistics is closely associated with information technology. As a result several new inter- disciplines have emerged. They are Data Mining, Data Warehousing, Geographic Information System, Artificial Intelligence etc. Now-a-days, statistics can be applied in hardcore technological spheres such as Bioinformatics, Signal processing, Telecommunications, Engineering, Medicine, Crimes, Ecology, etc.

Today's business managers need to learn how analytics can help them make better decisions that can generate better business outcomes. They need to have an understanding of the statistical concepts that can help analyze and simplify the flood of data around them. They should be able to leverage analytical techniques like decision trees, regression analysis, clustering and association to improve business processes.

1.2 Definitions

Statistics has been defined by various statisticians.

- *'Statistics is the science of counting'* -***A. L .Bowley***
- *'Statistics is the science which deals with the collection, presentation, analysis and interpretation of numerical data'* - ***Croxton and Cowden***
- **Wallist** and **Roberts** defines statistics as *"Statistics is a body of methods for making decisions in the face of uncertainty"*
- **Ya-Lun-Chou** slightly modifies Wallist and Roberts definition and come with the following definition : *"Statistics is a method of decision making in the face of uncertainty on the basis of numerical data and calculated risk."*

It may be seen that most of the above definitions of statistics are restricted to numerical measurements of facts and figures of a state. But modern thinkers like Secrist defines statistics as

'By statistics we mean the aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated to reasonable standards of accuracy collected in a systematic manner for a predetermined purpose and placed in relation to each other'.

Ch1_11th Stats.indd 3 07/12/2021 11:42:14

Among them, the definition by Croxton and Cowden is considered as the most preferable one due to its comprehensiveness. It is clear from this definition that statistics brings out the following characteristics.

Characteristics of Statistics:

(1) Aggregate of facts collected in systematic manner for a specific purpose.

Statistics deals with the aggregate of facts and figures. A single number cannot be called as statistics. For example, the weight of a person with 65kg is not statistics but the weights of a class of 60 persons is statistics, since they can be studied together and meaningful comparisons are made in relation to the other. This reminds us of Joseph Stalin's well known quote, "One death is a tragedy; a million is a statistics." Further the purpose for which the data is collected is to be made clear, otherwise the whole exercise will be futile. The data so collected must be in a systematic way and should not be haphazard.

(2) Affected by large number of causes to marked extent.

Statistical data so collected should be affected by various factors at the same time. This will help the statistician to identify the factors that influence the statistics. For example, the sales of commodities in the market are affected by causes such as supply, demand, and import quality etc. Similarly, as mentioned earlier if a million deaths occur the policy makers will be immediately in action to find out the causes for these deaths to see that such events will not occur.

(3) Numerically expressed.

The statistical facts and figures are collected numerically for meaningful inference. For instance, the service provided by a telephone company may be classified as poor, average, good, very good and excellent. They are qualitative in nature and cannot be called statistics. They should be expressed numerically such as 0 to denote poor,1 for average, 2 for good, 4 to denote very good and 5 for excellent. Then this can be regarded as statistics and is suitable for analysis. The other types of quality characteristics such as honesty, beauty, intelligence, defective etc which cannot be measured numerically cannot be called statistics. They should be suitably expressed in the form of numbers so that they are called statistics.

(4) Enumerated or estimated with a reasonable degree of accuracy.

The numerical data are collected by counting, measuring or by estimating. For example, to find out the number of patients admitted in a hospital, data is collected by actual counting or to find out the obesity of patients, data are collected by actual measurements

Ch1_11th Stats.indd 4 07/12/2021 11:42:14

on height and weight. In a large scale study like crop estimation, data are collected by estimation and using the powerful sampling techniques, because the actual counting may or may not be possible. Even if it is possible, the measurements involve more time and cost. The estimated figures may not be accurate and precise. However certain degree of accuracy has to be maintained for a meaningful analysis.

(5) To be placed in relation to the other.

One of the main reasons for the collection of statistical data is for comparisons In order to make meaningful and valid comparisons, the data should be on the same characteristic as far as possible. For instance, we can compare the monthly savings of male employees to that of the female employees in a company. It is meaningless if we compare the heights of 20 year-old boys to the heights 20 year- old trees in a forest.

Having looked into various definitions given by different authors to the term statistics in different contexts it would be appropriate to define

"Statistics in the sense of data are numerical statements of facts capable of analysis and interpretation".

"Statistics in the sense of science is the study of principles and methods used in the collection, presentation, analysis and interpretation of numerical data in any sphere of enquiry".

1.3 Functions of Statistics

The functions of statistics can be elegantly expressed as 7 - C's as :

S.NO	Functions	What it does
1	Collection	The basic ingredient of statistics is data. It should be carefully and scientifically collected
2	Classification	The collected data is grouped based on similarities so that large and complex data are in understandable form.
3	Condensation	The data is summarized, precisely without losing information to do further statistical analysis.
4	Comparison	It helps to identify the best one and checking for the homogeneity of groups,
5	Correlation	It enables to find the relationship among the variables
6	Causation.	To evaluate the impact of independent variables on the dependent variables.
7	Chance	Statistics helps make correct decisions under uncertainty.

Ch1_11th Stats.indd 5 07/12/2021 11:42:14

1.4 Scope and Applications

In ancient times the scope of statistics was limited. When people hear the word 'Statistics' they think immediately of either sports related numbers or a subject they have studied at college and passed with minimum marks. While statistics can be thought in these terms there is a wide scope for statistics. Today, there is no human activity which does not use statistics. There are two major divisions of statistical methods called descriptive statistics and inferential statistics and each of the divisions are important and satisfies different objectives. The descriptive statistics is used to consolidate a large amount of information. For example, measures of central tendency, like mean are descriptive statistics. Descriptive statistics just describes the data in a condensed form for solving some limited problems. They do not involve beyond the data at hand.

Inferential statistics, on the other hand, are used when we want to draw meaningful conclusions based on sample data drawn from a large population. For example, one might want to test whether a recently developed drug is more efficient than the conventional drug. Hence, it is impossible to test the efficiency of the drug by administering to each patient affected by a particular disease, but we will test it only through a sample. A quality control engineer may be interested in the quality of the products manufactured by a company. He uses a powerful technique called acceptance sampling to protect the producer and consumer interests. An agricultural scientist wanted to test the efficacy of fertilizers should test by designed experiments. He may be interested in farm size, use of land and crop harvested etc. One advantage of working in statistics is that one can combine his interest with almost any field of science, technology or social sciences such as Economics Commerce, Engineering Medicine, and Criminology and so on.

The profession of statistician is exciting, challenging and rewarding. Statistician is the most prevalent title but professionals like Risk analyst, Data analyst, Business analyst have been engaged in work related to statistics. In view of the overwhelming demand for Statistics many universities in India and elsewhere have been offering courses in statistics at graduate and Master's level. We have mentioned earlier that statistics has applications to almost all fields. Here in this section we highlight its applications to select branches.

1.4.1 Statistics and actuarial science

Actuarial science is the discipline that extensively applies statistical methods among other subjects involved in insurance and financial institutions. The professionals who qualify in actuarial science course are called actuaries. Actuaries, in the earlier days used deterministic models to assess the premiums in insurance sector. Nowadays, with modern computers and sophisticated statistical methods, science has developed vastly. In India, from 2006 a statutory body has been looking after the profession of actuaries.

Ch1_11th Stats.indd 6 07/12/2021 11:42:14

1.4.2 Statistics and Commerce

Statistical methods are widely used in business and trade solutions such as financial analysis, market research and manpower planning. Every business establishment irrespective of the type has to adopt statistical techniques for its growth. They estimate the trend of prices, buying and selling, importing and exporting of goods using statistical methods and past data. Ya-Lun-Chou says "It is not an exaggeration to say that today nearly every decision in business is made with the aid of statistical data and statistical methods."

Success Story

In 2004, a hurricane named *Sandy* hit the United States, tens of thousands of households were affected by it with bad weather and power failure. A Multi National Company, the largest retailer across the globe conducted a vast data analysis through its comprehensive database system and came out with surprising results. Emergency equipments and frozen bakery products were badly needed by the households during such disasters. This data analysis helped the MNC to predict the next hurricane named France in 2012. So it dispatched emergency equipments, flashlights and bakery products like strawberry pop tarts to all its retail outlets near the hurricane hit places. Those products were sold extremely well in large number by that particular MNC, whereas other retailers could not do. Such Big data analysis and prediction henceforth has helped many Multi National Companies to reduce their time, cost and labour.



1.4.3 Statistics and Economics

Statistical methods are very much useful to understand economic concepts, such as mandatory policy and public finance. In the modern world, economics is taught as an exact service which makes extensive use of statistics. Some of the important statistical techniques used in economic analysis are: Times series, Index Numbers, Estimation theory and Tests of significance, stochastic models. According to Engeberg "No Economist would attempt to arrive at a conclusion concerning the production or distribution of wealth without an exhaustive study of statistical data." In our country many state governments have a division called Department of Economics and Statistics for the analysis of Economic data of the state.

Ch1_11th Stats.indd 7 07/12/2021 11:42:14

1.4.4 Statistics and Medicine

In medical field, statistical methods are extensively used. If we look at the medical journals one can understand to what extent the statistical techniques play a key role. Medical statistics deals with the applications of statistical methods like tests of significance and confidence intervals to medicine and health science including epidemiology, public health. Modern statistical methods helps the medical practitioners to understand how long a patient affected by a dreaded disease will survive and what are the factors that influence a patient to be alive or dead.



A Digital Revolution in Healthcare

- The Economist, Feb 03,2018

We know that the internet already enables patients to seek online consultations when and where it suits them. Now technologies such as the smartphones allow people to monitor their own health.

The possibilities multiply when you add to crucial missing ingredients- access to your own medical records and the ability, easily to share information with those you trust. That allows you to reduce inefficiencies in your own treatment and also to provide date to help train medical algorithms. You can enhance your own care and everyone else's too.

1.4.5 Statistics and Agriculture

Experimentation and inference based on these experiments are the key features of general scientific methodology. Agricultural scientists conduct experiments and make inferences to decide whether the particular variety of crop gives a better yield than others or a particular type of fertilizer etc,. There are several institutes where research is being done by making use of statistical methods like analysis of variance (ANOVA), factorial experiments etc., falls under the hut of Design of experiments. There is a separate institute (IASRI), New Delhi, carrying out research in agricultural statistics.

1.4.6 Statistics and Industry

Statistical methods play a vital role in any modern use of science and technology. Many statistical methods have been developed and applied in industries for various problems. For example, to maintain the quality of manufactured products the concept of statistical quality control is used. The quality in time domain study of mechanical, electrical or electronic items the concept of 'Reliability' has emerged. Total quality management and six-sigma theories make use of statistical concepts.

1.4.7 Statistics and Information Technology

Information Technology is the applications of computers and telecommunication equipments to store, retrieve, transmit and manipulate data. Now-a-days, several industries are involved in information technology and massive amounts of data are stored every day. These data are to be analyzed meaningfully so that the information contained in the data

Ch1_11th Stats.indd 8 07/12/2021 11:42:14

is used by the respective users. To address this issue, fields such as data mining, Machine learning have emerged. Data mining an interdisciplinary sub field of computer science is the computational process of discovering patterns in large data sets involving methods such as artificial intelligence and statistics. Persons trained in statistics with computing knowledge have been working as data analytics to analyze such huge data.

1.4.8 Statistics and Government

Statistics provides statistical information to government to evolve policies, to maintain law and order, to promote welfare schemes and to other schemes of the government. In other words, statistical information is vital in overall governance of the state. For instance statistics provide information to the government on population, agricultural production, industrial production, wealth, imports, exports, crimes, birth rates, unemployment, education, minerals and so on.

1.5 Big Data

Big Data is a term used for a collection of data sets that are large and complex, which is difficult to store and process using available database management tools or traditional data processing applications. Daily we upload millions of bytes of data. 90 % of the world's data has been created in last few years.

Applications of Big Data

We cannot talk about data without talking about the people, because those are the once who are getting benefited by Big Data applications. Almost all the industries today are leveraging Big Data applications in one or the other way.

Smarter Healthcare: Making use of the petabytes of patient's data, the organization can extract meaningful information and then build applications that can predict the patient's deteriorating condition in advance.



Retail: Retail has some of the tightest margins, and is one of the greatest beneficiaries of big data. The beauty of using big data in retail is to understand consumer behavior. Suggestion based on the browsing history of the consumer, they supply their product to increase their sales.

Manufacturing: Analyzing big data in the manufacturing industry can reduce component defects, improve product quality, increase efficiency, and save time and money.

Ch1_11th Stats.indd 9 07/12/2021 11:42:15

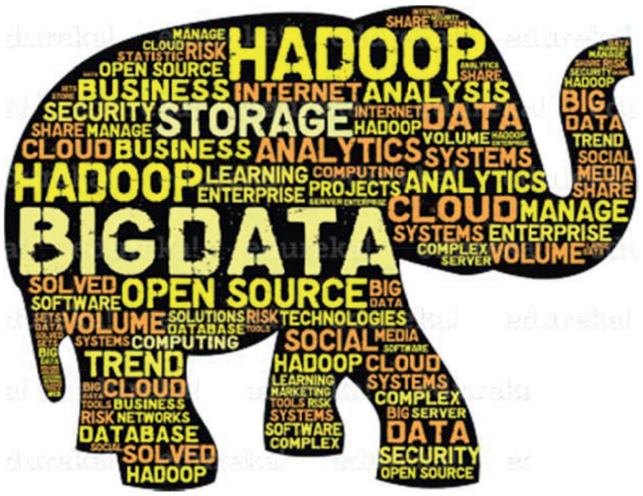
Traffic control: Traffic congestion is a major challenge for many cities globally. Effective use of data and sensors will be key to managing traffic better as cities become increasingly densely populated.



Search Quality: Every time we are extracting information from google, we are simultaneously generating data for it. Google stores this data and uses it to improve its search quality.

Sales promotion: Prominent sports persons or celebrities are selected as Brand Ambassadors for their products by the prominent industries through big data got from social media or from other organizations.

Challenges with Big Data



We have a few challenges which come along with Big Data those are data complexity, storage, discovery analytics and lack of talent. But we have several advance programming language that can handle the issue of Big data, like Hadoop, Mapreduce, Scale etc., and many of this languages like open source, Java-based programming framework that supports the storage and processing of extremely large data sets in a distributed computing environment.

1.6 Variable and Types of Data

Information, especially facts or numbers collected for decision making is called data. Data may be numerical or categorical. Data may also be generated through a variable.

Variable: A variable is an entity that varies from a place to place, a person to person, a trial to trial and so on. For instance the height is a variable; domicile is a variable since they vary from person to person.

A variable is said to be quantitative if it is measurable and can be expressed in specific units of measurement (numbers).

Ch1_11th Stats.indd 10 07/12/2021 11:42:15

A variable is said to be qualitative if it is not measurable and cannot be expressed in specific units of measurement (numbers). This variable is also called categorical variable.

The variable height is a quantitative variable since it is measurable and is expressed in a number while the variable domicile is qualitative since it is not measured and is expressed as rural or urban. It is noted that they are free from units of measurement.

Quantitative Data:

Quantitative data (variable) are measurements that are collected or recorded as a number. Apart from the usual data like height, weight etc.,

Qualitative Data:

Qualitative data are measurements that cannot be measured on a natural numerical scale. For example, the blood types are categorized as O, A, B along with the Rh factors. They can only be classified into one of the pre assigned or pre designated categories.

1.7 Measurement Scales

There are four types of data or measurements scales called nominal, ordinal, interval and ratio. These measurement scale is made by Stanley Stevens.

1.7.1 Nominal scales:

Nominal measurement is used to label a variable without any ordered value. For example, we can ask in a questionnaire 'What is your gender? The answer is male or female. Here gender is a nominal variable and we associate a value 1 for male and 2 for a female.'

They are numerical for name sake only. For example, the numbers 1,2,3,4 may be used to denote a person being single, married, widowed or divorced respectively. These numbers do not share any of the properties of numbers we deal with in day to life. We cannot say $4 > 1$ or $2 < 3$ or $1+3 = 4$ etc. The order of listings in the categories is irrelevant here. Any statistical analysis carried out with the ordering or with arithmetic operations is meaningless.

1.7.2 Ordinal scales:

These data share some properties of numbers of arithmetic but not all properties. For example, we can classify the cars as small, medium and big depending on the size.

In the ordinal scales, the order of the values is important but the differences between each one is unknown. Look at the example below.

Ch1_11th Stats.indd 11 07/12/2021 11:42:15

How did you feel yesterday after our trip toVedanthangal? The answers would be:

(1) Very unhappy (2) Unhappy (3) Okay (4) Happy (5) Very happy

In each case, we know that number 5 is better than number 4 or number 3, but we don't know how much better it is. For example, is the difference between "Okay" and "Unhappy" the same as the difference between "Very Happy" and "Happy?" In fact we cannot say anything.

NOTE

In the 'ordinal scales' it is the order that matters, and that is all we get from these. It is to be noted that the mean cannot be computed from the ordinal data, but either median or mode can be computed.

Similarly, a medical practitioner can say the condition of a patient in the hospital as good, fair, serious and critical and assign numbers 1 for good, 2 for fair, 3 for serious and 4 for critical. The level of seriousness can be from 1 to 4 leading to $1 < 2$ or $2 < 3$ or $3 < 4$. However, the value here just indicates the level of seriousness of the patient only.

1.7.3 Interval scales:

In an interval scale one can also carryout numerical differences but not the multiplication and division. In other words, an interval variable has the numerical distances between any two numbers. For example, suppose we are given the following temperature readings in Fahrenheit: $60^\circ, 65^\circ, 88^\circ, 105^\circ, 115^\circ$, and 120°. It can be written that $105^\circ > 88^\circ$ or $60^\circ < 65^\circ$ which means that 105° is warmer than 88° and that 60° is colder than 65°. It can also be written that $65^\circ - 60^\circ = 120^\circ - 115^\circ$ because equal temperature differences are equal conveying the same amount of heat needed to increase the temperature from an object from 60° to 65° and from 115° to 125°. However it does not mean that an object with temperature 120° is twice as hot as an object with temperature 60°, though 120° divided by 60° is 2. The reason is converting the temperature in to Celsius we have:

$$60^\circ \text{F} = \frac{5}{9}[60^\circ - 32^\circ]C = 15.570^\circ C \text{ and}$$

$$120^\circ \text{F} = \frac{5}{9}[120^\circ - 32^\circ]C = 48.870^\circ C$$

NOTE

It is impossible to compute ratios without a real origin as zero.

In the above equations, it is clear from the left hand side that 120° F is twice of 60° F while the right hand side says 48.87° C is more than three times of $15.57^\circ C$. The reason for the difficulty is that the Fahrenheit and Celsius scales have artificial origins namely zeros(freezing point of centigrade measure is 0^o C and the freezing point of Farenheit is 32^o F) and there is no such thing as 'no temperature.'

Ch1_11th Stats.indd 12 07/12/2021 11:42:17

1.7.4 Ratio scales:

Ratio scales are important when it comes to measurement scales because they tell us about the order, they tell us the exact value between units, and they also have an absolute zero–which allows for a wide range of both descriptive and inferential statistics to be applied. Good examples of ratio variables include height and weight. Ratio scales provide a wealth of possibilities when it comes to statistical analysis. These variables can be meaningfully added, subtracted, multiplied, divided). Central tendency can be measured by mean, median, or mode; Measures of dispersion, such as standard deviation and coefficient of variation can also be calculated from ratio scales.

In summary, nominal variables are used to "name," or label a series of values. Ordinal scales provide good information about the order of choices, such as in a customer satisfaction survey. Interval scales give us the order of values plus the ability to quantify the difference between each one. Finally, Ratio scales give us the ultimate–order, interval values, plus the ability to calculate ratios since a "true zero" can be defined. The distinction made here among nominal, ordinal, interval and ratio data are very much important as these concepts used in computers for solving statistical problems using statistical packages like SPSS, SAS, R, STATA etc.,

Points to Remember	
● Characteristics of statistics	Aggregate of facts collected in systematic manner for a specific purpose. Affected by large number of causes to marked extent. Numerically expressed. Enumerated or estimated with a reasonable degree of accuracy. To be placed in relation to the other.
● Functions of Statistics	Collection, Classification, Condensation, Comparison, Correlation, Causation, Chance
● Scope and Applications	Actuarial science, Commerce, Economics, Medicine, Agriculture, Industry, Information Technology Government
● Applications of Big Data	Smarter Healthcare, Retail, Traffic control Manufacturing, Search Quality, Sales promotion
● Types of Data	Quantitative Data, Qualitative Data
● Measurement Scales	Nominal scale, Ordinal scale, Interval scale, Ratio scale

Ch1_11th Stats.indd 13 07/12/2021 11:42:17

EXERCISE 1

5HBK2

I. Choose the best answer:

1. The number of days of absence per year that a worker has is an example of
   (a) Nominal scale (b) Ordinal scale (c) Interval scale (d) Ratio scale

2. The data that can be classified according to colour is
   (a) Nominal scale (b) Ordinal scale (c) Interval scale (d) Ratio scale

3. The rating of movies as good , average and bad is
   (a) Nominal scale (b) Ordinal scale (c) Interval scale (d) Ratio scale

4. The temperature of a patient during hospitalization is 100^0F is in
   (a) Nominal scale (b) Ordinal scale (c) Interval scale (d) Ratio scale

II. Fill in the blanks:

5. Statistics mainly deals with_____
6. Statistics is broadly classified as_____and _____
7. The measure used in statistics during primitive days is_____
8. Age of a student is _____variable
9. The founder of Indian Statistical Institute (ISI) is_____
10. Temperature in your city is in _____ scale of measurement.
11. Intelligence of a student is a _____ variable.
12. Colour of a car is in _____ scale of measurement.
13. The name of your representative in parliament is in _____ measurement.
14. Performance of a salesman is good. Good is in _____ scale of measurement.

III. Answer shortly:

15. Define statistics.
16. What is the meaning of data?
17. Explain the role of statistics in Actuarial science.

Ch1_11th Stats.indd 14 07/12/2021 11:42:17

IV. Answer briefly:

18. Discuss the definition of Statistics due to Croxdon and Cowden.

19. List the characteristics of statistics.

20. What do you understand by Qualitative variable and quantitative variables?

V. Answer in detail:

21. Write a brief note on the contributions by P C Mahalonobis.

22. Write a note on the origin and growth of statistics

23. Explain the functions of statistics

24. State the applications of statistics in Agriculture and industry

ANSWERS

I. 1. d 2. a 3. b 4. c

II. 5. numerical facts 6. singular and plural 7. counting 8. quantitative
9. P.C.Mahalanobis 10. interval 11. qualitative 12. nominal
13. nominal 14. ordinal

Ch1_11th Stats.indd 15 07/12/2021 11:42:17

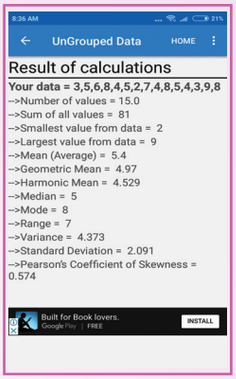


ICT CORNER

STATISTICAL ANALYSER

STATS IN YOUR PALM

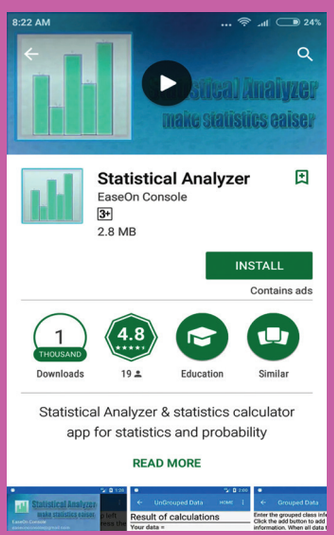
This activity is to calculate mean, median, range variance, standard deviation.



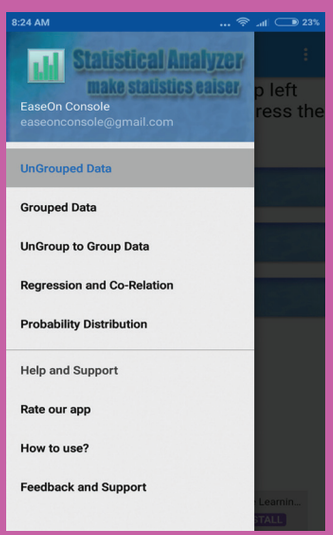
Steps:

- This is an android app activity. Open the browser and type the URL given (or) scan the QR code. (Or) search for "**Statistical Analyzer**" in google play store.
- (i) Install the app and open the app, (ii) click "**Menu**", (iii) In the menu page click "**Ungrouped data**" menu.
- Type the raw data followed by comma for each entry and click "**CALCULATE**" to get the output.

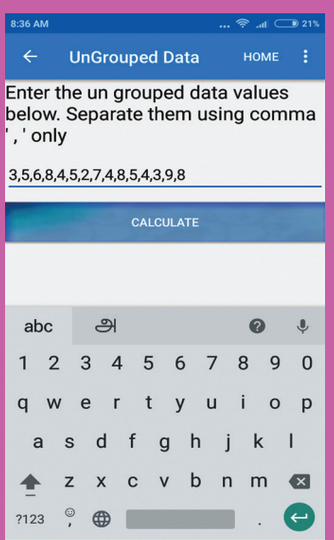
Step-1



Step-2



Step-3



Pictures are indicatives only*

URL:

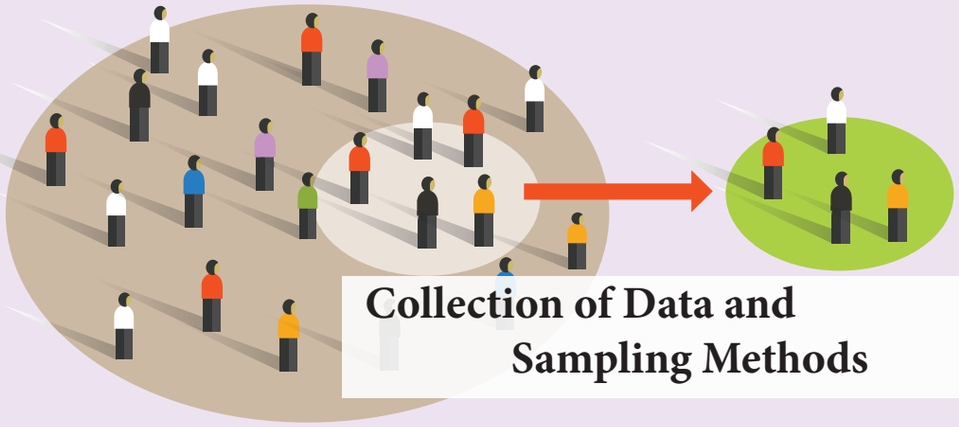
https://play.google.com/store/apps/details?id=com.easeonconsole.malik.statisticcalculator

Ch1_11th Stats.indd 16 07/12/2021 11:42:19

www.tnschoolbooks.in

Chapter

2



Collection of Data and Sampling Methods



Pandurang Vasudeo Sukhatme
(27 July, 1911 -
28 January 1997)

Pandurang Vasudeo Sukhatme was born on 27th July 1911 in the village Budh, district Satara, near Pune. During 1933-36, he studied at the University College, London and was awarded a Ph.D. in 1936 and a D.Sc. Degree in 1939.When he was in London, Prof. Sukhatme came under the influence of such eminent personalities in Statistics as R.A.Fisher, Jerzy Neyman and E.S.Person and did valuable research in Statistical Theory of Sampling. Sukhatme was appointed as Statistical Advisor to the Council as head of Statistical Unit in ICAR. On account of his dynamic leadership,statistical branch of ICAR eventually grew to become a full-fledged Indian Agricultural Statistics Research Institute (IASRI) exclusively devoted to research in agricultural statistics

Prof. Sukhatme was known in the field of nutrition for the Sukhatme-Margen hypothesis which implies the following: *At low levels of calorie intake, energy is used with greater metabolic efficiency and efficiency decreases as the intake increases over the homeostatic range.*

He received several awards including the Padma Bhushan by the President of India in 1973,the Hari Om Ashram Trust Award by the University Grants Commission in 1983 and P.C.Mohalanobis Birth Centenary Award of the Indian Science Congress Association in 1994.

'Statistics is the science of learning from experience' *- Bradley Efron*

Learning Objectives

- Emphasis the necessity of data collection
- Distinguishes between primary and secondary data
- Introduces methods of collecting primary data with their advantages and disadvantages

Ch2_11th Stats.indd 17 07/12/2021 11:58:32

- Designs a questionnaire for the collection of data.
- Describes Secondary data
- Explains the advantages of Sampling over Census method
- Describes Probability sampling methods and their appropriateness.
- Explains the uses of Non-Probability sampling
- Differentiates Sampling and Non-sampling errors

Introduction

Statistical data are the basic 'ingredients' of Statistics on which statistician work. A set of numbers representing records of observations is termed statistical data. The need to collect data arises in every sphere of human activity. However that 'Garbage in garbage out' applies in Statistics too. Hence adequate care must be taken in the collection of data. It is a poor practice to depend on whatever data available.

Data collection process:

There are five important questions to ask in the process of collecting data: What? How? Who? Where? When?

QUESTION	RELATED ACTIVITY
What data is to be collected?	Decide the relevant data of the study
How will the data be collected?	Choice of a data collection instrument
Who will collect the data?	Method of enquiry: Primary / Secondary
Where the data will be collected?	Decide the Population oftt the survey
When will the data be collected?	Fixing the time schedule

This unit addresses the above Questions in detail

2.1 Categories and Sources of Data

There are two categories of data namely primary data and secondary data.

Primary data are that information which is collected for the first time, from a Survey, or an observational study or through experimentation. For example

- A survey is conducted to identify the reasons from the parents for selection of a particular school for their children in a locality.
- Information collected from the observations made by the customers based on the service they received.

Ch2_11th Stats.indd 18 07/12/2021 11:58:32

- To test the efficacy of a drug, a randomized control trial is conducted using a particular drug and a placebo.

Let us see the detailed methods of collecting Primary data in the following Section

2.2 Methods of collecting primary data

In this section, we present different methods of collecting primary data. In this context, we define an Investigator or Interviewer as one who conducts the statistical enquiry and the person from whom the information is collected is called a Respondent.

The primary data comes in the following three formats.

(i) Survey data: The investigator or his agency meets the respondents and gets the required data.

(ii) Experimental data (field/laboratory): The investigator conducts an experiment, controlling the independent variables and obtains the corresponding values of the dependent variable.

(iii) Observational data: In the case of a psychological study or in a medical situation, the investigator simply observes and records the information about respondent. In other words the investigator behaves like a spectator.

The various methods used to collect primary data are:

(i) Direct Method

(ii) Indirect Method

(iii) Questionnaire Method

(iv) Local Correspondents Method

(v) Enumeration Method

2.2.1 Direct Method:

There are four methods under the direct method

(a) Personal Contact Method

As the name says, the investigator himself goes to the field, meets the respondents and gets the required information. In this method, the investigator personally interviews the respondent either directly or through phone or through any electronic media. This method is suitable when the scope of investigation is small and greater accuracy is needed.

Ch2_11th Stats.indd 19 07/12/2021 11:58:32

Merits:



- This method ensures accuracy because of personal interaction with the investigator.
- This method enables the interviewer to suitably adjust the situations with the respondent.

Limitations:

- When the field of enquiry is vast,this method is more expensive, time consuming and cumbersome.
- In this type of survey, there is chance for personal bias by the investigator in terms of asking 'leading questions'.

(b) Telephone Interviewing

In the present age of communication explosion, telephones and mobile phones are extensively used to collect data from the respondents. This saves the cost and time of collecting the data with a good amount of accuracy.

(c) Computer Assisted Telephone Interviewing (CATI)

With the widespread use of computers, telephone interviewing can be combined with immediate entry of the response into a data file by means of terminals, personal computers, or voice data entry. Computer – Assisted Telephone Interviewing (CATI) is used in market research organizations throughout the world.

(d) Computer Administered Telephone Survey

Another means of securing immediate response is the computer-administered telephone survey. Unlike CATI, there is no interviewer. A computer calls the phone number, conducts the interview, places data into a file for later tabulation, and terminates the contact. The questions are voice synthesized and the respondent's answer and computer timing trigger continuation or disconnect. The last three methods save time and cost, apart from minimizing the personal bias.

2.2.2 Indirect Method:

The indirect method is used in cases where it is delicate or difficult to get the information from the respondents due to unwillingness or indifference. The information about the respondent is collected by interviewing the third party who knows the respondent well.

Ch2_11th Stats.indd 20 07/12/2021 11:58:33

Instances for this type of data collection include information on addiction, marriage proposal, economic status, witnesses in court, criminal proceedings etc. The shortcoming of this method is genuineness and accuracy of the information, as it completely depends on the third party.

2.2.3 Questionnaire Method

A questionnaire contains a sequence of questions relevant to the study arranged in a logical order. Preparing a questionnaire is a very interesting and a challenging job and requires good experience and skill.

The general guidelines for a good questionnaire:

- The wording must be clear and relevant to the study
- Ability of the respondents to answer the questions to be considered
- Avoid jargons
- Ask only the necessary questions so that the questionnaire may not be lengthy.
- Arrange the questions in a logical order.
- Questions which hurt the feelings of the respondents should be avoided.
- Calculations are to be avoided.
- It must be accompanied by the covering letter stating the purpose of the survey and guaranteeing the confidentiality of the information provided.

Editing the preliminary questionnaire

Once a preliminary draft of the questionnaire has been designed, the researcher is obligated to critically evaluate and edit, if needed. This phase may seem redundant, given all the careful thoughts that went into each question. But recall the crucial role played by the questionnaire.

Pre Test

Once the rough draft of the questionnaire is ready, pretest is to be conducted. This practice of pretest often reveals certain short comings in the questions, which can be modified in the final form of the questionnaire. Sometimes, the questionnaire is circulated among the competent investigators to make suggestions for its improvement. Once this has been done and suggestions are incorporated, the final form of the questionnaire is ready for the collection of data.

Ch2_11th Stats.indd 21 07/12/2021 11:58:33

Advantages:

- In a short span of time, vast geographical area can be covered.
- It involves less labor.

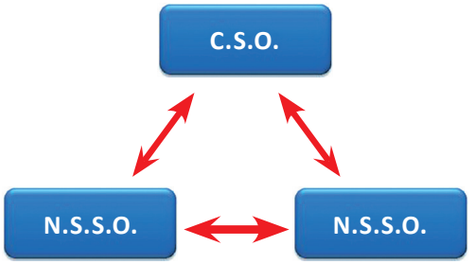
Limitations:

- This method can be used only for the literate population.
- Some of the mailed questionnaires may not be returned.
- Some of the filled questionnaires may not be complete.
- The success of this method depends on the nature of the questions and the involvement of the respondents.

Schedule:

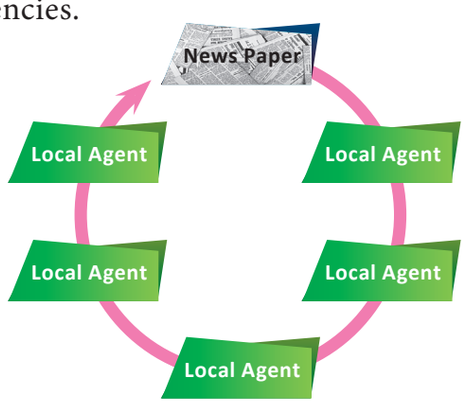
Schedule is a structure of a set of questions on a given topic which are asked by an investigator. Population census and some personal interview method are the examples of using schedules.

2.2.4 Local Correspondents Method



In this method, the investigator appoints local agents or correspondents in different places. They collect the information on behalf of the investigator in their locality and transmit the data to the investigator or headquarters. This method is adopted by newspapers and government agencies.

For instance, the Central Statistical Organization (CSO) of Government of India has local correspondents NSSO. Through them they get the required data. Newspaper publishers appoint agents to collect news for their dailies. These people collect data in their locality on behalf of the publisher and transmit them to the head office.



This method is economical and provides timely information on a continuous basis. It involves high degree of personal bias of the correspondents.

2.2.5 Enumeration method:

In this method, the trained enumerators or interviewers take the schedules themselves, contact the informants, get replies and fill them in their own hand writing.

Ch2_11th Stats.indd 22 07/12/2021 11:58:33

Thus, schedules are filled by the enumerator whereas questionnaires are filled by the respondents. The enumerators are paid honorarium. This method is suitable when the respondents include illiterates. The success of this method depends on the training imparted to the enumerators. The voters' list preparation, information on ration card for public distribution in India, etc., follow this method of data collection. National Sample Survey Office (NSSO) collects information using schedules depending on the theme.

2.3 Secondary data

Secondary data is collected and processed by some other agency but the investigator uses it for his study. They can be obtained from published sources such as government reports, documents, newspapers, books written by economists or from any other source., for example websites. Use of secondary data saves time and cost. Before using the secondary data scrutiny must be done to assess the suitability, reliability, adequacy, and accuracy of the data.

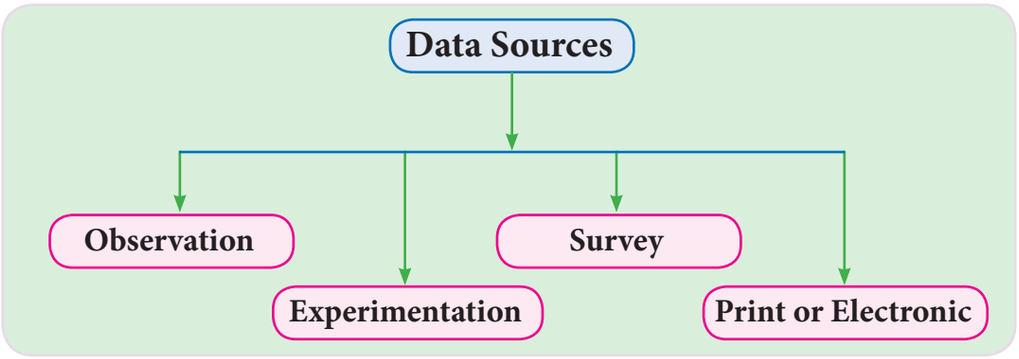
Sources of Secondary Data

The secondary data comes from two main sources, namely published or unpublished.

The published sources include:

- Government Publications - Reserve Bank of India (RBI) Bulletin, Statistical Abstracts of India by Central Statistical Organization (CSO), Statistical Abstracts of Tamil Nadu by the Department of Economics and Statistics, Government of Tamil Nadu.
- International Publications - Publications of World Health Organizations, World Bank, International Labor Organizations, United Nations Organizations
- Publications of Research institutes – Indian Council of Medical Research (ICMR), Indian Council of Agricultural Research (ICAR).
- Journals or Magazines or Newspapers - Economic Times, Business Line

The data which are not published are also available in files and office records Government and Private organizations. The different sources described above are schematically described below.



Ch2_11th Stats.indd 23 07/12/2021 11:58:33

The following table compares the two types of data.

Comparison between Primary and Secondary data

Primary	Secondary
It is collected for the first time	Compiled from already existing sources
It is collected directly by the investigator or by his team	Complied by persons other than the persons who collected the data
It costs more	It costs less
It requires more time	It requires considerably less time
Possibility of having personal bias	Personal bias is minimized

2.4 Population:

The concept of population and sample are to be understood clearly as they have significance role in the context of statistics.

The word population or statistical population is used for aggregate of all units or objects within the purview of enquiry. We may be interested in the level of education in a college of Tamil Nadu. Then all the students in the college will make up the population. If the study is aimed to know the economic background of Hr. Secondary students of a school, then all the students studying +1 and +2 classes of that school is the population. The population may contain living or non-living things. All the flowers in a garden or all the patients in a hospital are examples of populations in different studies.

Finite Population

A population is called finite if it is possible to count or label its individuals. It may also be called a Countable Population. The number of vehicles passing in a highway during an hour, the number of births per month in a locality and the number of telephone calls made during a specific period of time are examples of finite populations. The number of units in a finite population is denoted by N and is called the size of the population.

Infinite Population

Sometimes it is not possible to count or label the units contained in the population. Such a population is called infinite or uncountable. The number of germs in the body of a sick patient is uncountable.

Sampling from a finite population will be considered in the rest of the chapter.

Ch2_11th Stats.indd 24 07/12/2021 11:58:33

Sampling from an infinite population can be handled by considering the distribution of the population. A random sample from an infinite population is considered as a random sample from a probability distribution. This idea will be used when we study testing of significance in the second year.

2.5 Census Method

The census method is also called complete enumeration method. In this method, information is collected from each and every individual in the statistical population.

Census of India is one of the best example. It is carried out once in every ten years. An enquiry is carried out, covering each and every house in India. It focuses on demographic details. They are collected and published by the Register general of India.

Appropriateness of this method:

The complete enumeration method is preferable provided the population is small and not scattered. Otherwise, it will have the following disadvantages.

Disadvantages:

- It is more time consuming, expensive and requires more skilled and trained investigators.
- More errors creep in due to the volume of work.
- Complete enumeration cannot be used if the units in the population destructive in nature. For example, blood testing, testing whether the rice is cooked or not in a kitchen,
- Testing the life times of bulbs etc.,
- When area of the survey is very large and there is less knowledge about the population, this method is not practicable. For example the tiger population in India, number trees in a forest cannot be enumerated using census method.

2.6 Sampling method:

In view of all these difficulties one has to resort to sampling methods for collecting thedata.

Sample is small proportion of the population taken from the population to study the characteristics of the population. By observing the sample one can make inferences about the population from which it is taken.

Ch2_11th Stats.indd 25 07/12/2021 11:58:33

Sampling is a technique adopted to select a sample. The sample must represent or exactly duplicate the characteristics of the population under study. In suchcase that sample is called as a representative sample. The sampling method used for selecting a sample is important in determining how closely the sample resembles the population, in determining.



The Tamil proverb "ஒரு பானை சோற்றுக்கு ஒரு சோறு பதம்" describes precisely about sampling that "one grain suffices to test a whole pot of cooked rice".

Sampling unit is the basic unit to be sampled from the population which cannot be further subdivided for the purpose of sampling. Head of the house is the sampling unit for the household survey. In the study to know the average age of a class, student is the sampling unit.

Sampling frame to adopt a sampling procedure it is necessary to prepare a list such that there exists, one to one correspondence between sampling units and numbers. Such a list or map is called sampling frame. A list of villages in a district, Student list of +1 and +2 students in the above said example, A list of houses in a household survey etc.

Sample size is the number of units in the sample.

Merits and Limitations of Sampling

The prime objective of the sampling is to get the representative sample which will provides the desired information about the population with maximum accuracy at a given cost.

Merits

- Cost: Expenditure on conducting the survey is less compared to complete enumeration.
- Time: The consumption of time is relatively less in a sample study than potentially generated voluminous data.
- Accuracy: It is practically proved that the results based on representative samples more reliable than the complete enumeration.
- In the case of destructive type situations, sampling method is the only way.

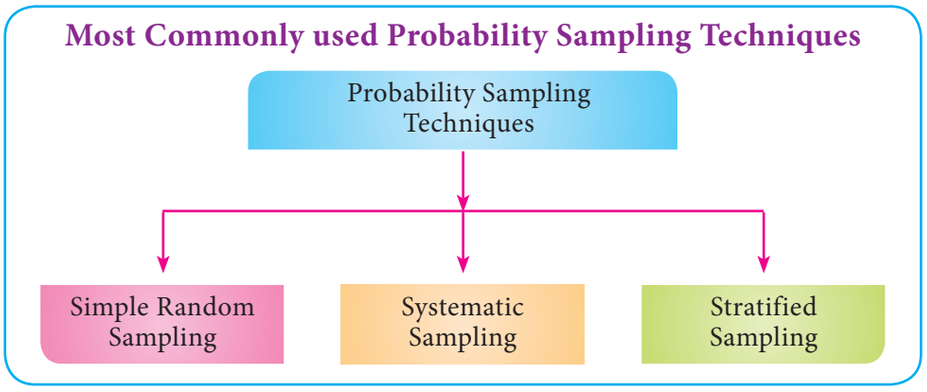
Limitations

- Accuracy depends on the honesty of the investigator
- There is possibility for sampling error.

Ch2_11th Stats.indd 26 07/12/2021 11:58:33

2.7 Probability sampling:

Probability sampling or random sampling involves a random selection process to include a unit of the population in the sample. In random sampling, all items have a positive probability of selection. Random sampling techniques remove the personal bias of the investigators. Some of the important methods of probability sampling are: Simple Random Sampling, Stratified Random Sampling and Systematic Sampling.



2.7.1 Simple random sampling

This method selects a sample in such a way that each possible sample to have an equal probability of being selected or each item in the entire population to have an equal probability of being included in the sample. The following are the methods of selecting a simple random sample.

(1) Lottery Method

Suppose that we have to select a random sample of size n from a finite population of size *N*. First assign numbers 1 to N to all the N units of the population. Then write numbers 1 to *N* on different identical slips or cards so that a card is not distinguishable from another. They are folded and mixed up in a drum or a box or a container. A blindfold selection is made. The selected card may be replaced or may not be replaced before the next draw. The Required numbers of slips are selected for the desired sample size under any one of these methods forms a simple random sample. The selection of items thus depends on chance.

Simple random sampling without replacement (SRSWOR):

If the selected cards are not replaced before the next draw, such a sampling is called without replacement.

Ch2_11th Stats.indd 27 07/12/2021 11:58:33

Simple random sampling with replacement (SRSWR):

If the selected cards are replaced before the next draw, such a sampling is called sampling with replacement

Remark:

If the population size is large, this method is cumbersome. The alternative method is using of table of random numbers.

(2) Table of Random numbers

The easiest way to select a sample randomly is to use random number tables. There are several tables of random numbers. Some of them are

(i) Kendall and Smith random number table.

(ii) Tippet's random number table.

(iii) Fishers and Yates random number table.

Method of using Random Number table.

In a given finite population, the method of selecting a random sample from a random number table is given below:

The table contains different numbers consisting of 0, 1, 2,… ,9 and possesses the essential characteristics which ensures random sampling. A part of the table is given as Appendix. Referring to this page, selection of a random sample is explained by taking a finite population of size 100 units and selecting a sample of 10 units. The steps to be followed are listed as under.

(i) A list of all 100 units in the finite population is prepared and each unit is assigned a serial number ranging from 00 to 99.

(ii) Any number in the table is chosen at random and it is the starting point for selecting the sampling units.

(iii) From the starting point we can make a move on to the next number either vertically, horizontally or diagonally.

(iv) Since our numbered population consists of two digits from 00 to 99 , we confine ourselves to reading only two digit numbers without omitting any number that comes forward.

(v) As we proceed, random numbers read from the table which are above 99 are ignored and those numbers which are less than are equal to 99 are recorded. This process is continued until we reach 10 such random numbers.

Ch2_11th Stats.indd 28 07/12/2021 11:58:33

(vi) If sampling without replacement scheme is followed, a random number once recorded appears again, the same is omitted and we move on to the next number.

(vii) The 10 such selected random numbers are compared with the labeled (numbered) population units and the corresponding units are selected to get a simple random sample of size 10.

Example 2.1

Select a random sample of size 15 from the random number table from a finite population of size 220.

Step 1 : Assign serial numbers 00 to 219 to the 220 units of the population.

Step 2 : Since the maximum digit is 3, select a three digit as starting point. Let the starting point in the random number table is 066.(first two digits of the entry in the 11^{th} row and 4^{th} column).

Step 3 : Continue in the column downwards to select random numbers of size 15.

Step 4 : The numbers chosen to have SRS are 066, 147, 119, 194, 093, 180, 092, 127, 211, 087, 002, 214, 176, 063 and 176. If we get any random number greater than 219, then the multiple of 219 is to be subtracted from the selected number such that the resultant value is less than 219.

Illustration:

Assume that the random is 892. But here, the maximum value assigned for the population is 219 . Hence we select '4'as the multiplier of 219. Then the random number correspond to 892 is 016. (See: $892 - 4 \times 219 = 892 - 876 = 16$)

Case(1) : Under SRSWR: Here the random sample is 066, 147, 119, 194, 093, 180, 092, 127, 211, 087, 002, 214, 176, 063 and 176.

Case(2): Under SRSWOR: Here the random sample is 066, 147, 119, 194, 093, 180, 092, 127, 211, 087, 002, 214, 176, 063and 157.

[**Note:** 176 is removed from the selection since it is repeated again]

Merits:

- It is more representative of the universe.
- It is free from personal bias and prejudices.

Ch2_11th Stats.indd 29 07/12/2021 11:58:33

- The method is simple to use.
- It is to assess sampling error in the method.

Demerits:

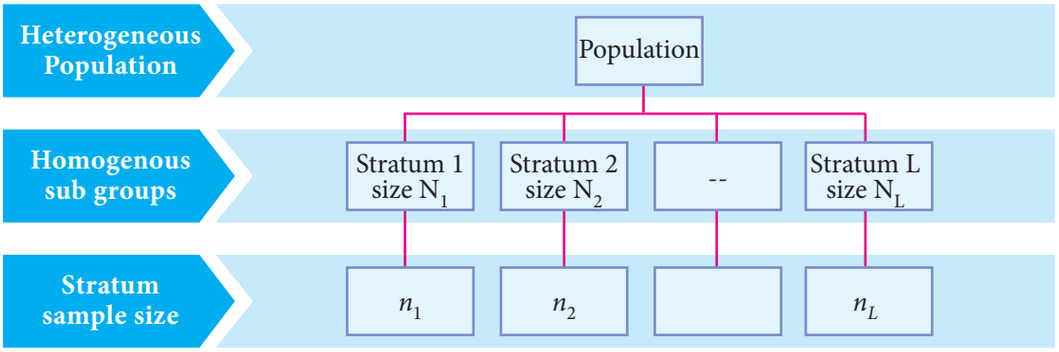
- If the units are widely dispersed, the sample becomes unrepresentative.
- The method is not applicable when the units are heterogeneous in nature.
- Units selected may be too widely dispersed; thus, adherence to the whole sample may be difficult. Remark: Simple random sampling is more suitable when the population is homogeneous. If the population is heterogeneous we must go in for stratified random sampling.

NOTE

Random number can also be generated through scientific calculator or through the computer.

2.7.2 Stratified Random Sampling

In Stratified random sampling, the heterogeneous population of size N units is sub-divided into L homogeneous non overlapping sub populations called Strata, the ith stratum having N_i units (i =1, 2, 3,…,L) such that $N_1+N_2+\ldots+N_L = N$.



A sample size being n_i from i^{th} stratum (i=1, 2,…,L) is independently taken by simple random sampling in such way that $n_1+n_2+\ldots+n_L = n$. A sample obtained using this procedure is called a stratified random sample.

To determine the sample size for each stratum, there are two methods namely proportionate allocation and optimum allocation. In proportionate allocation sample size is determined as proportionate to stratum size. If the stratum size large, that stratum will get more representation in the sample. If the stratum size small, that stratum will get less representation in the sample. The sample size for the ith stratum can be determined using the formula $n_i = (n/N)^* N_i$. The optimum allocation method uses variation in the stratum and cost to determine the stratum sample size ni .

Ch2_11th Stats.indd 30 07/12/2021 11:58:34

Example 2.2

To study about the introduction of NEET exam, the opinions are collected from 3 schools. The strength of the schools are 2000, 2500 and 4000. It is fixed that the sample size is 170. Calculate the sample size for each school?

Solution

Here $N = 2000 + 2500 + 4000 = 8500$ and $n = 170$ then $n_1 = n_2 = n_{13} = ?$

$N_1 = 2000, \quad N_2 = 2500, \quad N_3 = 4000$

$$n_1 = (n/N) \times N_1 = (170 / 8500) \times 2000 = 40$$

$$n_2 = (n/N) \times N_2 = (170 / 8500) \times 2500 = 50$$

$$n_3 = (n/N) \times N_3 = (170 / 8500) \times 4000 = 80$$

Therefore 40 students from school 1, 50 students from school 2 and 80 students from school 3, are to be selected using SRS to obtain the required stratified random sample

The main objective of stratification is to give a better cross-section of the population for a higher degree of relative precision. The criteria used for stratification are States, age and sex, academic ability, marital status etc,. In many practical situations when it is difficult to stratify with respect to the characteristic under study, administrative convenience may be considered as the basis for stratification

Merits

- It provides a chance to study of all the sub-populations separately.
- An optimum size of the sample can be determined with a given cost, precision and reliability.
- It is a more precise sample.
- Representation of sub groups in the population
- Biases reduced and greater precise.

Limitations

- There is a possibility of faulty stratification and hence the accuracy may be lost.
- Proportionate stratification requires accurate information on the proportion of population in each stratum.

2.7.3 Systematic Sampling

In systematic sampling, the population units are numbered from 1 to N in ascending order. A sampling interval, denoted by k, is determined as $k = \frac{N}{n}$, where n denotes the

Ch2_11th Stats.indd 31 07/12/2021 11:58:34

required sample size. Then $n-1$ such sampling intervals each consisting of k units will be formed. A number is selected at random from the first sampling interval. Let it be number i where $i \leq k$. This number is the random starting point for the whole selection of the sample. The unit corresponding to i is the first unit in the sample. The subsequent sampling units are the units in the following positions:

$$i,\ k+i,\ 2k+i,\ 3k+i,\ \ldots,\ nk$$

The layout for systematic sampling								
Sampling interval 1	1	2	3	4	...	i - Random start	...	k
Sampling interval 2	k+1	k+2	k+3	k+4	...	$k+i$	...	2k
Sampling interval 3	2k+1	2k+2	2k+3	2k+4	...	2$k+i$	...	3k
Sampling interval 4	3k+1	3k+2	3k+3	3k+4	...	3$k+i$	...	4k
	...	...	...	...	...	...	...	...
Sampling interval n	(n-1)k+1	(n-1)k+2	(n-1)k+3	(n-1)k+4	...	(n-1)$k+i$	...	nk

Thus, with selection of the first unit, the whole sample is selected automatically. As the first unit could have been any of the k units, the technique will generate k systematic samples with equal probability. If N is not an integral multiple of n, then sizes of a few possible systematic samples may vary by one unit.

NOTE

This sample is also called ***quasi random sample*** since the first unit only is selected at random and all the subsequent units are not selected randomly.

Merits:

- This method is simple and convenient.
- Less time consuming.
- It can be used in infinite population.

Limitation:

- Since it is a quasi random sampling, the sample may not be a representative sample.

Ch2_11th Stats.indd 32 07/12/2021 11:58:34

Example 2.3

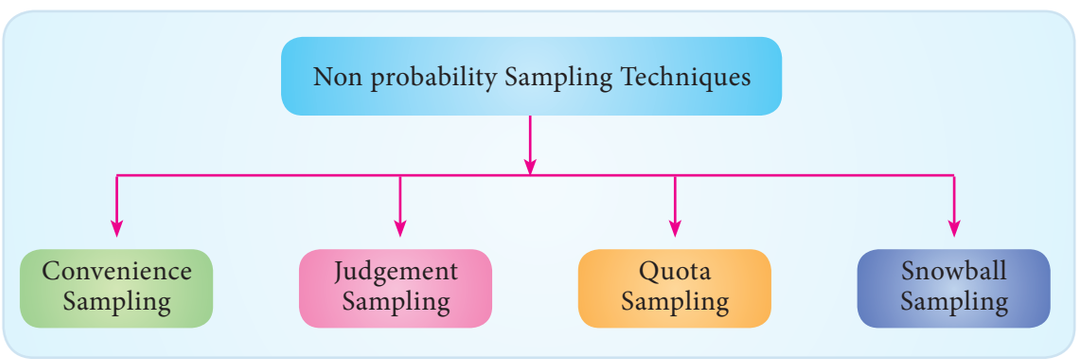
Suppose a systematic random sample of size $n = 10$ is needed from a population of size $N = 200$, the sampling interval $k = \frac{N}{n} = \frac{200}{10} = 20$. The first sampling interval consists of numbers 1 to 20. If the randomly selected number (random starter) is 7, the systematic sample will consist units corresponding to positions 7, 27,47,67,87,107,127,147,167,187.

Applications:

- Systematic sampling is preferably used when the information is to be collected from trees in a highway, houses in blocks, etc,.
- This method is often used in industry, where an item is selected for testing from a production line (say, every fifteenth item in the order of production) to ensure that equipments are working satisfactorily.
- This technique could also be used in a sample survey for interviewing people. A market researcher might select every 10th person who enters a particular store, after selecting a person at random as a random start.

2.8 Non-probability sampling:

Non probability sampling is the sampling procedure in which samples are selected based on the subjective judgment of the researcher, rather than random selection. This is used when the representativeness of the population is not the prime issue. Convenience or judgments of the investigators play an important role in selecting the samples. In general, there are four types of non probability sampling called convenience sampling, judgment sampling, quota sampling and snowball sampling.



2.8.1 Convenience Sampling:

The samples are drawn at the convenience of the investigator. The investigator pick up cases which are easily available units keeping the objectives in mind for the study.

Ch2_11th Stats.indd 33 07/12/2021 11:58:34

Merits:

- Useful for pilot study.
- Use the results that are easily available.
- Processes of picking people in the most convenient and faster way to immediately get their reactions to a certain hot and controversial topic.
- Minimum time needed and minimum cost incurs.

Limitations:

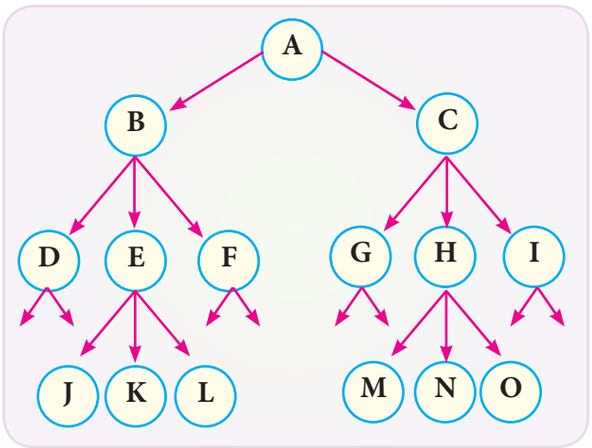
- High risk of selection bias.
- May provide misleading information.
- Not representative sample. Errors occur in the form of the members of the population who are infrequent or non users of that location and who are not related with the study.

NOTE

The use of convenience sampling technique is discouraged by many researchers due to inability to generalize research findings, the relevance of bias and high sampling error. Nevertheless convenience sampling may be the only option available in certain situations. For example, "it may be that a researcher intends to study the 'customer satisfaction of Jet Airways' he use the convenient sampling because he has been able to negotiate access through available contacts".

2.8.2 Snowball Sampling:

In this type, initial group of respondents are selected. Those respondents are requested to provide the names of additional respondents who belong to the target population of interest. It is a sampling method that involves the assistance of study subjects to identify other potential subjects in studies where subjects are hard to locate such as sex workers, drug abusers, etc. This type of sampling technique works like a chain referral. Therefore it is also called chain referral sampling.



Merits:

- Appropriate for small specialized population.
- Useful in studies involving respondents rare to find.

Ch2_11th Stats.indd 34 07/12/2021 11:58:34

Limitations:

- It takes more time
- Most likely not representative
- Members of the population, who are little known, disliked or whose opinions conflict with the respondents, have low probability of being included.

2.8.3 Judgement Sampling.

The investigator believes that in his opinion, some objects are the best representative of the population than others. It involves "hand picking" of sampling units. That is the interviewer uses his judgment in the selection of the sample that who should represent the population to serve the investigator's purpose. It is usually used when a limited number of individuals possess the trait of interest. This type of sampling is also known as purposive sampling. This is useful when selecting specific people, specific events, specific prices of data, etc.

For example, Selecting members for a competition like quiz, oratorical contest to represent a school.

Merits:

- Low expense.
- Minimum time needed.
- Easy

Limitations:

- Highly subjective.
- Generalization is not appropriate.
- Certain members of the population will have a smaller chance or no chance of selection compared to others.
- This method does not give representative part of the population, since favoritism is involved.

2.8.4 Quota Sampling

This is another non-probability sampling method. In this method, the population is divided into different groups and the interviewer assign quotas to each group. The selection of individuals from each group is based on the judgment of the interviewer. This type of sampling is called quota sampling. Specified sizes of number of certain types of peoples are included in the sample.

Ch2_11th Stats.indd 35 07/12/2021 11:58:34

Merits:

- The selection of the sample in this method is quick, easy and cheaper.
- May control sample characteristics.
- More chance of representative.

Limitations:

- Selection bias.
- The sample is not a true representative and statistical properties cannot be applied.

Example 2.4

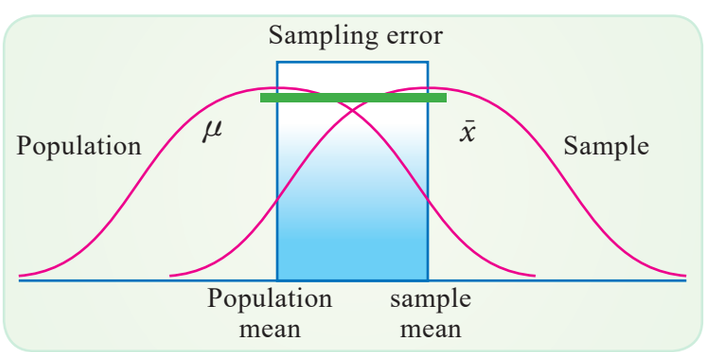
A selection committee wants to compose a cricket team (11 players) for a test match.

Groups	Pace bowlers	Spinners	All-rounders	Batsmen	Wicket keepers
	Players in this category	Players in this category	Players in this category	Players in this category	Players in this category
Quota	2	3	3	2	1
Select the players using judgment sampling to fulfill the requirement of the respective quota.					

In the composition of a cricket team, the selection committee forms groups compartmentalize as pace bowlers, spinners, all-rounders, batsmen and wicket keepers. The committee fixed quota for each group based on the pitch and the opponent teams' strength. Then, from each group they select the required number of players using judgement. Look the table.

2.9 Sampling and Non-sampling errors

2.9.1 Sampling error



The purpose of sample is to study the population characteristics. The sample size is not equal to population size except in the case complete enumeration. Therefore, the statistical measurements like mean of the sample and mean of the population differ.

Ch2_11th Stats.indd 36 07/12/2021 11:58:34

If $\bar{x}$ is the sample mean and μ is the population mean of the characteristic X then the sampling error is $\bar{x} - \mu$. The sampling error may be positive or negative or zero.

2.9.2 Non-Sampling Errors

The non-sampling errors arise due to various causes right from the beginning stage when the survey is planned and designed to the final stage where the data are processed and analyzed. Non sampling errors are more serious than the sampling errors because a sampling error can be minimized by taking a large sample. It is difficult to minimize non sampling errors, even if a large sample is taken. The main sources of non-sampling errors are now described.

(i) Non-response:

The errors due to non-response may occur due to omission or lapse on the part of the interviewer, or the refusal on the part of the respondents to questions or because of the non-availability of the individuals during the period of survey.

(ii) Errors in measurement:

The measuring device may be biased or inaccurate. The respondent may not know the correct answer and may give imprecise answers. Common examples are questions on age, income, and events that happened in the past. The interviewer may also fail to record the responses correctly. Errors in measurement include errors in coding, editing, and tabulation.

Coverage errors:

The coverage errors are classified as **'under coverage errors'** and **'over coverage errors'.** Under-coverage errors occur in the following situations:

- The selected unit in the sampling frame is not interviewed by the investigator.
- The selected unit is incorrectly classified as ineligible for surveys
- The unit is omitted or skipped by the interviewer.

Similarly, over-coverage occurs under the following situations:

- The sampling frame covers ineligible units.
- The frame may contain the same unit more than once.

The errors cannot be ignored since the cumulative effect of these errors affect the objectives of the survey.

Organising a sample survey

The above said things provide a comprehensive idea about collection of data. However, when one decides to collect data through sampling the following steps are to be followed.

Ch2_11th Stats.indd 37 07/12/2021 11:58:34

Stage I: Developing a sample plan

Definite sequence of steps the interviewer ought to go in order to draw and ultimately arrive at the final sample.

(i) Define the relevant population.

(ii) Obtain a population list, if possible: may only be some type of a sampling frame.

(iii) Fix the sample size

(iv) Choose the appropriate sampling.

(v) Draw the sample.

(vi) Assess the validity of the sample.

(vii) Resample if necessary.

Stage II: Pilot survey or Pilot Study

It is a guiding survey, usually on a small scale, carried out before the main survey. The information received by pilot survey is utilized in improving the efficiency of the large scale main survey. It helps in:

(i) Estimating the cost of the regular survey

(ii) Correcting the questionnaire of the survey

(iii) Training the field workers.

(iv) Removing the faults of the field organization.

(v) Deciding about the other details of the survey.

Stage III. Dealing with Non- respondents

Procedures will have to be devised to deal with those who do not give information.

Points to Remember

- Data are the ingredients on which statistics works.
- Data type may be primary data or secondary data.
- Source of getting data depends on the problem of study.
- Each method of collection of data has its own advantage and disadvantages. Hence an appropriate method should be used in data collection.
- Studying all individuals in a population is not viable or feasible or impractical, it is advantageous to study the characteristics of the population using sample characteristics.
- In sampling we should specifically specify the population be able to list the sampling frame and use an appropriate sampling method.
- Although several methods are of selecting samples are available, random sampling method is more preferable as it is a representative of the population and have some theoretical advantages.

Ch2_11th Stats.indd 38 07/12/2021 11:58:34

EXERCISE 2

I. Choose the best answer:

1. Which one of the method is not a primary data collection method

   (a) Questionnaire method
   (b) Date collected from published sources
   (c) Local correspondent method
   (d) Indirect investigation

2. Which one of the method is Probability sampling?

   (a) Quota sampling
   (b) Snowball sampling
   (c) systematic sampling
   (d) Convenience sampling

3. Which one of the method is quasi probability sampling?

   (a) Quota sampling
   (b) Snowball sampling
   (c) Systematic sampling
   (d) Convenience sampling

4. 'A schedule' in the context of data collection refers to:

   (a) Questionnaire used by the investigator
   (b) Program schedule of the data collection
   (c) Instrument used in enumeration method
   (d) Secondary data.

5. Which one is false in the questionnaire method?

   (a)Vast coverage in less time
   (b) This method can be adopted to any respondent
   (c) Response rate may be low
   (d) It offers greater anonymity.

6. Opinion poll in a study is conducted:

   (a) Before the process start
   (b) After the process
   (c) Middle of the process
   (d) At any point of time of the process

7. Exit survey is conducted:

   (a) Before the process start
   (b) After the process
   (c) Middle of the process
   (d) At any point of time of the process

Ch2_11th Stats.indd 39 07/12/2021 11:58:34

8. When the researcher uses the data of an agency, then the data is called:

(a) Quantitative data (b) Qualitative data

(c) Secondary data (d) Primary data

II . Fill in the blanks:

9. Data means _____

10. A questionnaire contains _____

11. Sample size means_____

12. Snow ball sampling is _____ sampling

13. _____ is also called quasi random sampling.

III . Answer shortly :

14. What do you understand by data?

15. Define population.

16. What is a complete enumeration?

17. What is a schedule?

18. What do you mean by destructive type?

19. Define sample.

20. Define sampling error.

21. What is the prime concern about a sample?

22. Define sampling frame.

23. State the principle of stratification.

24. What are 'under coverage errors'?

IV . Answer in brief:

25. Distinguish between primary data and secondary data.

26. List out the precautions to be taken while using the secondary data.

27. Distinguish between sample and sampling.

Ch2_11th Stats.indd 40 07/12/2021 11:58:34

28. Distinguish between random sample and simple random sampling .

29. Distinguish between sampling error and non sampling error.

30. Distinguish between questionnaire and schedule.

31. Explain the method of snowball sampling. Under what circumstances it is more suitable.

32. Why systematic sample is called quasi random sample?

33. State the sources of non responses.

34. Define pretest and state its advantage.

35. List the sources of non sampling error.

V. Answer in detail:

36. Describe various methods of collecting primary data and comment on their relative merits and demerits.

37. What are the guiding considerations in the construction of questionnaire?. Explain

38. Discuss the advantages of sampling method over census method of data collection.

39. Under what circumstances stratified random sampling procedure is considered appropriate? How would you select such a sample? Explain by means of an example.

40. What is non probability sampling? Explain each one with the help of examples.

41. Write an essay about non sampling errors.

Answers

I. 1. b. 2. c. 3. c. 4. c. 5. b. 6. a. 7. b. 8. c.

II.9. ingredients of statistics 10. sequence of questions

11. number of units in the sample

12. non probability 13. systematic sampling

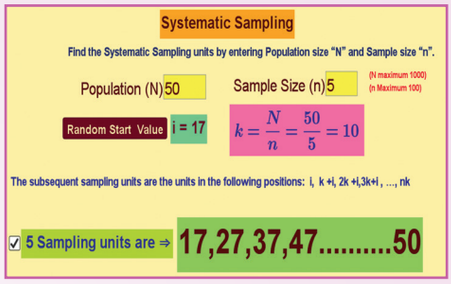
Ch2_11th Stats.indd 41 07/12/2021 11:58:34



ICT CORNER

SAMPLING METHODS -SYSTEMATIC SAMPLING

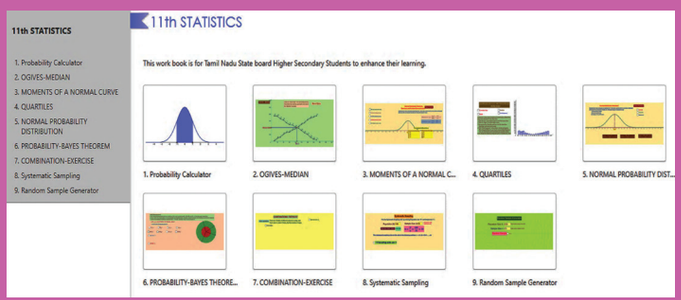
Expected final outcome



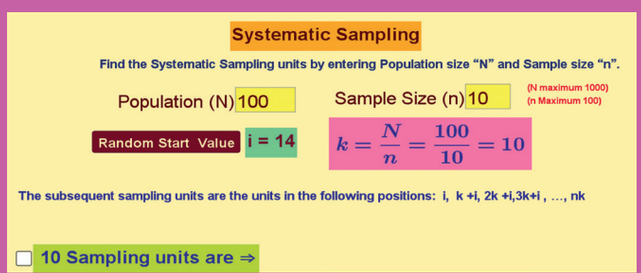
Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for statistics are given, open the worksheet named "**Systematic Sampling**"
- Type the population value and sample size values in the respective boxes and click box sampling units to get the systematic sampling units for your data.

Step-1



Step-2



Pictures are indicatives only*

URL:

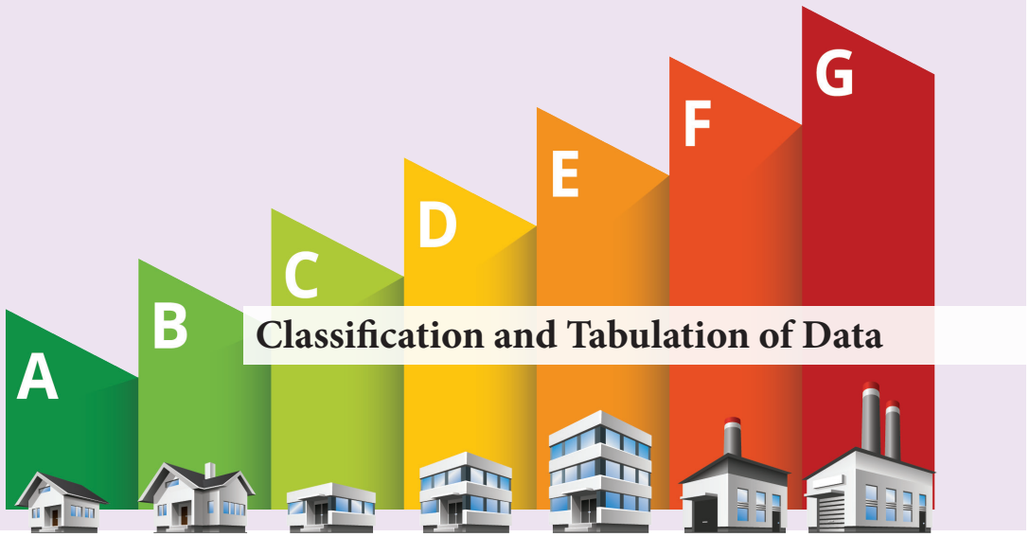
https://ggbm.at/uqVhSJWZ

Ch2_11th Stats.indd 42 07/12/2021 11:58:36

www.tnschoolbooks.in

Chapter

3



Classification and Tabulation of Data



Prof. John Wilder Tukey
(16 June, 1915- 26 June, 2000)

John Wilder Tukey, an American mathematician born on June 16,1915 in New Bedford, Massachusetts, USA, obtained M. Sc., in Chemistry (1937) from Brown University. Later he received a Ph.D. in Mathematics from Princeton University. In his career, Tukey worked on developing statistical methods for computers at AT&T Bell Laboratories and invented the term 'bit' as a contraction of 'binary digit'. He introduced the box plot technique in his book entitled 'Exploratory Data Analysis'. His varied interest in Statistics led him to develop the Tukey's Range test, Tukey lambda distribution, Tukey's test of additivity and Tukey's lemma. He received the IEEE Medal of Honor(1982) for his contribution to the Spectral Analysis of Random Processes and the FFT Algorithm. He passed away in New Brunswick, New Jersey on 26 July, 2000.

'Knowledge is power, and data is just data. No matter how much data you have on hand, if you don't have a way to make sense of it, you really have nothing at all' - unknown

R1KBY

Learning Objectives

- Emphasizes the importance and objectives of classification and tabulation
- Distinguishes various types of classification and types of tables
- Explains the meaning and formation of frequency distributions
- Explains stem and leaf plot method
- Illustrates the procedures with numerical examples

Ch3_11th Stats.indd 43 07/12/2021 12:00:32

Introduction

In the previous chapter, discussions were on various types of data, census and sampling and methods of collecting data with suitable examples. It may be recalled that by data we mean the collected set of a number of related observations. The data as such are complex and voluminous in nature in many instances. It is essential that the data so collected must be arranged for proper understanding and for further statistical analysis.

3.1 Classification of Data



Few 1000 years ago the ancient Tamil poet Tholkappiyar classified things on earth into two categories as living and non-living things. He also classified the living beings based on their six senses.

The data that are unorganized or have not been arranged in any way are called raw data. The ungrouped data are often voluminous, complex to handle and hardly useful to draw any vital decisions. Hence, it is essential to rearrange the elements of the raw data set in a specific pattern. Further, it is important that such data must be presented in a condensed form and must be classified according to homogeneity for the purpose of analysis and interpretation. An arrangement of raw data in an order of magnitude or in a sequence is called **array**. Specifically, an arrangement of observations in an ascending or a descending order of magnitude is said to be an **ordered array**.

Classification is the process of arranging the primary data in a definite pattern and presenting in a systematic form. *Horace Secrist* defined classification as the process of arranging the data into sequences and groups according to their common characteristics or separating them into different but related parts. It is treated as the process of classifying the elements of observations or things into different groups or classes or sequences according to the resemblances and similarities of their character. It is also defined as the process of dividing the data into different groups or classes which are as homogeneous as possible within the groups or classes, but heterogeneous between themselves.

Objectives of Classification

Classification of data has manifold objectives. The salient features among them are the following:

- It explains the features of the data.
- It facilitates comparison with similar data.
- It strikes a note of homogeneity in the heterogeneous elements of the collected information.
- It explains the similarities which may exist in the diversity of data points.
- It is required to condense the mass data in such a manner that the similarities and dissimilarities are understood.

Ch3_11th Stats.indd 44 07/12/2021 12:00:32

- It reduces the complexity of nature of data and renders the data to comprehend easily.
- It enables proper utilization of data for further statistical treatment.

3.2 Types of Classification

The raw data can be classified in various ways depending on the nature of data. The general types of classification are: (i) Classification by Time or Chronological Classification (ii) Classification by Space or Spatial Classification (iii) Classification by Attribute or Qualitative Classification and (iv) Classification by Size or Quantitative Classification. Each of these types is now described.

3.2.1 Classification by Time or Chronological Classification

The method of classifying data according to time component is known as classification by time or chronological classification. In this type of classification, the groups or classes are arranged either in the ascending order or in the descending order with reference to time such as years, quarters, months, weeks, days, etc. Illustrations for statistical data to be classified under this type are listed below:

- Number of new schools established in Tamil Nadu during 1995 – 2015
- Pass percentage of students in SSLC Board Examinations over a period of past 5 years
- Index of market prices in stock exchanges arranged day-wise
- Month-wise salary particulars of employees in an industry
- Particulars of outpatients in a Primary Health Centre presented day-wise.

Example 3.1

The classification of data relating to the price of 10 gms of gold in India during 2001 - 2012 is given in Table 3.1

Table 3.1

Price of 10 gms of Gold in India

Year	Price in ₹	Year	Price in ₹	Year	Price in ₹
2001	4300	2005	7000	2009	14500
2002	4990	2006	8400	2010	18500
2003	5600	2007	10800	2011	26400
2004	5850	2008	12500	2012	31799

Ch3_11th Stats.indd 45 07/12/2021 12:00:32

Example 3.2

The classification of data relating to the population of India from 1961 to 2011 is provided in Table 3.2:

Table 3.2
Population of India from 1961 to 2011

Year	1961	1971	1981	1991	2001	2011
Population (in crores)	43.92	54.82	68.33	84.64	102.87	121.02

3.2.2 Classification by Space (Spatial) or Geographical Classification

The method of classifying data with reference to geographical location such as countries, states, cities, districts, etc., is called classification by space or spatial classification. It is also termed as geographical classification. The following are some examples:

- Number of school students in rural and urban areas in a State
- Region-wise literacy rate in a state
- State-wise crop production in India
- Country-wise growth rate in South East Asia

Example 3.3

The classification of data relating to number of schools and types of schools in 7 major cities of Tamil Nadu as per the Annual Budget Report 2012 – 2013 is given in Table 3.3

Table 3.3
Number of Schools and Types of Schools

District	Primary School	Middle School	High School	Hr. Sec. School	Total
Chennai	697	203	206	448	1554
Coimbatore	1090	307	185	306	1888
Madurai	1314	332	172	254	2075
Trichy	1260	350	187	199	1996
Salem	1402	445	213	231	2291
Tirunelveli	1786	437	178	251	2652
Erode	986	357	146	176	1665

Example 3.4

Average yield of rice (Kg/hec) during 2014-15 as per the records of Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, Government of India, in five states in India is given in Table 3.4

Ch3_11th Stats.indd 46 07/12/2021 12:00:32

Table 3.4

Average Yield of Rice during 2014 - 15

State	Yield (Kg/hec)
Tamilnadu	3191
Karnataka	2827
Kerala	2818
Uttarpradesh	2082
West Bengal	2731

3.2.3 Classification by Attributes or Qualitative classification

The method of classifying statistical data on the basis of attribute is said to be classification by attributes or qualitative classification. Examples of attributes include nationality, religion, gender, marital status, literacy and so on.

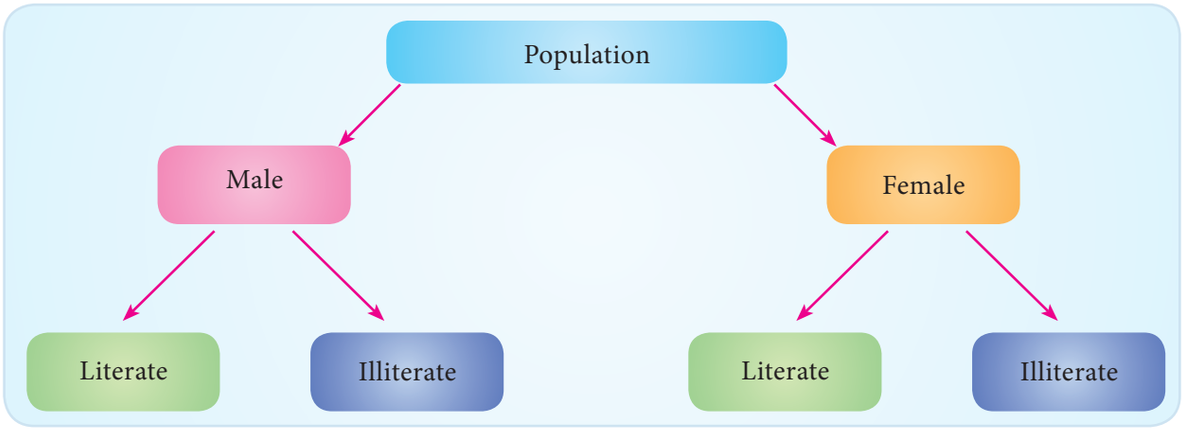


Fig: 3.1

Classification according to attributes is of two kinds: simple classification and manifold classification.

In simple classification the raw data are classified by a single attribute. All those units in which a particular characteristic is present are placed in one group and others are placed in another group. The classification of individuals according to literacy, gender, economic status would come under simple classification.

In manifold classification, two or more attributes are considered simultaneously. When more attributes are involved, the data would be classified into several classes and subclasses depending on the number of attributes. For example, population in a country can be classified in terms of gender as male and female. These two sub-classes may be further classified in terms of literacy as literate and illiterate.

Ch3_11th Stats.indd 47 07/12/2021 12:00:32

While classifying the data according to attributes, it is essential to ensure that the attributes involved have to be defined without ambiguity. For example, while classifying income groups, the investigator has to define carefully the different non-overlapping income groups.

Example 3.5

The classification of students studying in a school according to gender is given in Table 3. 5

Table 3.5

Gender-wise and class-wise information about students in a School

Class	Boys	Girls
VI	82	34
VII	74	43
VIII	92	27
IX	87	32
X	90	30
XI	75	25
XII	78	22

3.2.4 Classification by Size or Quantitative Classification

When the characteristics are measured on numerical scale, they may be classified on the basis of their magnitude. Such a classification is known as classification by size or quantitative classification. For example data relating to the characteristics such as height, weight, age, income, marks of students, production and consumption, etc., which are quantitative in nature, come under this category.



Colours of vegetables, Types of vegetables	Qualitative data (Non-numerical)
Weight of vegetables, Cost of vegetables	Quantitative data (Numerical)

Example 3.6

The classification of data relating to nutritive values of three items measured per 100 grams is provided in Table 3.6

Ch3_11th Stats.indd 48 07/12/2021 12:00:32

Table 3.6

Nutritive values of Sugar, Jaggery and Honey

Item	Energy K calories	Carbohydrate (in gm)	Calcium (in mg)	Iron (in mg)
Sugar	398	99.4	12	0.15
Jaggery	383	95.0	80	2.65
Honey	313	79.5	5	0.69

Source: National Institute of Nutrition, ICMR, Hyderabad.

In the classification of data by size, data may also be classified deriving number of classes based on the range of observations and assigning number of observations lying in each class. The following is another example for classification by size.

Example 3.7

The classification of 55 students according to their marks is given in Table 3.7

Table 3.7

Classification of students with respect to their marks

Marks	0 - 5	5 - 10	10 - 15	15 - 20	20 - 25	25 – 30	30 - 35
Number of Students	2	6	13	17	11	4	2

Rules for Classification of Data

There are certain rules to be followed for classifying the data which are given below.

- The classes must be exhaustive, i.e., it should be possible to include each of the data points in one or the other group or class.
- The classes must be mutually exclusive, i.e., there should not be any overlapping.
- It must be ensured that number of classes should be neither too large or nor too small. Generally, the number of classes may be fixed between 4 and 15.
- The magnitude or width of all the classes should be equal in the entire classification.
- The system of open end classes may be avoided.

Ch3_11th Stats.indd 49 07/12/2021 12:00:33

3.3 Tabulation

A logical step after classifying the statistical data is to present them in the form of tables. A table is a systematic organization of statistical data in rows and columns. The main objective of tabulation is to answer various queries concerning the investigation. Tables are very helpful while carrying out the analysis of collected data and subsequently for drawing inferences from them. It is considered as the final stage in the compilation of data and forms the basis for its further statistical treatment.

Advantages of Tabulation

- It is a logical step of presenting statistical data after classification.
- It enables the reader to understand the required information with ease as the information is contained in rows and columns with figures.
- It enables the investigator to present the data in a brief or condensed and compact form.
- Comparison is made simple by displaying data to be compared in a single table.
- It is easy to remember the data points or items if they are properly placed in the form of table, as it provides a kind of visual aid.
- It facilitates easy computation and helps easy detection of errors and omissions.
- It enables the reader to refer the data to be presented in a manner that suits for further statistical treatment and for making valid conclusions.

3.4 Types of Tables

Statistical tables can be classified under two general categories, namely, general tables and summary tables.

General tables contain a collection of detailed information including all that is relevant to the subject or theme. The main purpose of such tables is to present all the information available on a certain problem at one place for easy reference and they are usually placed in the appendices of reports.

Summary tables are designed to serve some specific purposes. They are smaller in size than general tables, emphasize on some aspect of data and are generally incorporated within the text. The summary tables are also called derivative tables because they are derived from the general tables. The information contained in the summary table aims at analysis and inference. Hence, they are also known as interpretative tables.

Ch3_11th Stats.indd 50 07/12/2021 12:00:33

The statistical tables may further be classified into two broad classes namely simple tables and complex tables. A simple table summarizes information on a single characteristic and is also called a univariate table.

Example 3.8

The marks secured by a batch of students in a class test are displayed in Table 3.8

Table 3.8
Marks of Students

Marks	0 - 10	10 - 20	20 - 30	30 – 40	40 - 50	50 - 60
Number of Students	10	12	17	20	15	6

This table is based on a single characteristic namely marks and from this table one may observe the number of students in each class of marks. The questions such as the number of students scored in the range 50 – 60, the maximum number of students in a specific range of marks and so on can be determined from this table.

A complex table summarizes the complicated information and presents them into two or more interrelated categories. For example, if there are two coordinate factors, the table is called a two-way table or bi-variate table; if the number of coordinate groups is three, it is a case of three-way tabulation, and if it is based on more than three coordinate groups, the table is known as higher order tabulation or a manifold tabulation.

Example 3.9

Table 3.9 is an illustration for a two-way table, in which there are two characteristics, namely, marks secured by the students in the test and the gender of the students. The table provides information relating to two interrelated characteristics, such as marks and gender of students. It is observed from the table that 26 students have scored marks in the range 40 – 50 and among them students, 16 are males and 10 are females.

Table 3.9
Marks of Students

Marks	**Number of Students**		**Total**
	Males	**Females**	
30 – 40	8	6	14
40 – 50	16	10	26
50 – 60	14	16	30
60 - 70	12	8	20
70 – 80	6	4	10
Total	56	44	100

Ch3_11th Stats.indd 51 07/12/2021 12:00:33

Example 3.10

Table 3.10 is an example for a three – way table with three factors, namely, marks, gender and location.

Table 3.10
Marks of Students

Marks	Males		Total	Females		Total	Total		Total
	Urban	Rural		Urban	Rural		Urban	Rural	
30 – 40	4	4	8	4	2	6	8	6	14
40 – 50	10	6	16	5	5	10	15	11	26
50 – 60	8	6	14	9	7	16	17	13	30
60 - 70	7	5	12	5	3	8	12	8	20
70 – 80	5	1	6	2	2	4	7	3	10
Total	34	22	56	25	19	44	59	44	100

From this table, one may get information relating to the distribution of students according marks, gender and geographical location from where they hail.

3.5 Components of a Table

Generally a table should be comprised of the following components:

- Table number and title
- Stub (the headings of rows)
- Caption (the headings of columns)
- Body of the table
- Foot notes
- Sources of data.

(i) **Table Number and Title:** Each table should be identified by a number given at the top. It should also have an appropriate short and self explanatory title indicating what exactly the table presents.

(ii) **Stub:** Stubs stand for brief and self explanatory headings of rows.

(iii) **Caption:** Caption stands for brief and self explanatory headings of columns. It may involve headings and sub-headings as well.

(iv) **Body of the Table:** The body of the table should provide the numerical information in different cells.

Ch3_11th Stats.indd 52 07/12/2021 12:00:33

(v) Foot Note: The explanatory notes should be given as foot notes and must be complete in order to understand them at a later stage.

(vi) Source of Data: It is always customary to provide source of data to enable the user to refer the original data. The source of data may be provided in a foot note at the bottom of the table.

A typical format of a table is given below:

Table Number

Title of the Table

Stub heading	Caption (Column headings)	Total
Stub (Row entries)	Body	
Total		

Foot note (if any)

Source of Data (if any)

General Precautions for Tabulation

The following points may be considered while constructing statistical tables:

- A table must be as precise as possible and easy to understand.
- It must be free from ambiguity so that main characteristics from the data can be easily brought out.
- Presenting a mass of data in a single table should be avoided. Displaying the data in a single table would increase the chances for occurrence of mistakes and would make the table unwieldy. Such data may be presented in more than one table such that each table should be complete and should serve the purpose.
- Figures presented in columns for comparison must be placed as near to each other as possible. Percentages, totals and averages must be kept close to each other. Totals to be compared may be given in bold type wherever necessary.
- Each table should have an appropriate short and self- explanatory title indicating what exactly the table presents.
- The main headings and subheadings must be properly placed.
- The source of the data must be indicated in the footnote.
- The explanatory notes should always be given as footnotes and must be complete in order to understand them at a later stage.

Ch3_11th Stats.indd 53 07/12/2021 12:00:33

- The column or row heads should indicate the units of measurements such as monetary units like Rupees, and other units such as meters, etc. wherever necessary.
- Column heading may be numbered for comparison purposes. Items may be arranged either in the order of their magnitude or in alphabetical, geographical, and chronological or in any other suitable arrangement for meaningful presentation.
- Figures as accurate as possible are to be entered in a table. If the figures are approximate, the same may be properly indicated.

3.6 Frequency Distribution

A tabular arrangement of raw data by a certain number of classes and the number of items (called frequency) belonging to each class is termed as a frequency distribution. The frequency distributions are of two types, namely, discrete frequency distribution and continuous frequency distribution.

3.6.1 Discrete Frequency Distribution

Raw data sometimes may contain a limited number of values and each of them appeared many numbers of times. Such data may be organized in a tabular form termed as a simple frequency distribution. Thus the tabular arrangement of the data values along with the frequencies is a simple frequency distribution. A simple frequency distribution is formed using a tool called 'tally chart'. A tally chart is constructed using the following method:

- Examine each data value.
- Record the occurrence of the value with the slash symbol (/), called tally bar or tally mark.
- If the tally marks are more than four, put a crossbar on the four tally bar and make this as block of 5 tally bars (𝍸)
- Find the frequency of the data value as the total number of tally bars i.e., tally marks corresponding to that value.

Example 3.11

The marks obtained by 25 students in a test are given as follows: 10, 20, 20, 30, 40, 25, 25, 30, 40, 20, 25, 25, 50, 15, 25, 30, 40, 50, 40, 50, 30, 25, 25, 15 and 40. The following discrete frequency distribution represents the given data:

Ch3_11th Stats.indd 54 07/12/2021 12:00:33

Table 3.11

Marks Scored by the Students

Marks	Tally Bars	Frequency
10	/	1
15	//	2
20	///	3
25	𝍸 //	7
30	////	4
40	𝍸	5
50	///	3
Total		25



Number of fruits	Discrete data
Weight of fruits, Cost of fruits	Continuous data

3.6.2 Continuous Frequency Distribution:

It is necessary to summarize and present large masses of data so that important facts from the data could be extracted for effective decisions. A large mass of data that is summarized in such a way that the data values are distributed into groups, or classes, or categories along with the frequencies is known as a continuous or grouped frequency distribution.

Example 3.12

Table 3.12 displays the number of orders for supply of machineries received by an industrial plant each week over a period of one year.

Table 3.12

Supply Orders for Machineries of an Industrial Plant

Number of Orders Received	Number of Weeks
0 – 4	2
5 – 9	8
10 – 14	11
15 – 19	14

Ch3_11th Stats.indd 55 07/12/2021 12:00:33

20 – 24	6
25 – 29	4
30 – 34	3
35 – 39	2
40 – 44	1
45 – 49	1

This table is a grouped frequency distribution in which the number of orders are given as classes and number of weeks as frequencies. Some terminologies related to a frequency distribution are given below.

Class: If the observations of a data set are divided into groups and the groups are bounded by limits, then each group is called a class.

Class limits: The end values of a class are called class limits. The smaller value of the class limits is called lower limit (L) and the larger value is called the upper limit(U).

Class interval: The difference between the upper limit and the lower limit is called class interval (I). That is, I = U – L.

Class boundaries: Class boundaries are the midpoints between the upper limit of a class and the lower limit of its succeeding class in the sequence. Therefore, each class has an upper and lower boundaries.

Width : Width of a particular class is the difference between the upper class boundary and lower class boundary.

Mid- point: Half of the difference between the upper class boundary and lower class boundary.

In Example 3.12, the interval 0 - 4 is a class interval with 0 as the lower limit nd 4 as the upper limit. The upper boundary of this class is obtained as midpoint of the upper limit of this class and lower limit of its succeeding class. Thus the upper boundary of the class 0 - 4 is 4.5. The lower class boundary of this is 0 - 0.5 which is - 0.5. The lower boundary of the class 5 - 9 is clearly 4.5. Similarly, the other boundaries of different classes can be found. The width of the classes is 5.

3.6.3 Inclusive and Exclusive Methods of Forming Frequency Distribution

Formation of frequency distribution is usually done by two different methods, namely inclusive method and exclusive method.

Ch3_11th Stats.indd 56 07/12/2021 12:00:33

Inclusive method

In this method, both the lower and upper class limits are included in the classes. Inclusive type of classification may be used for a grouped frequency distribution for discrete variable like members in a family, number of workers etc., It cannot be used in the case of continuous variable like height, weight etc., where integral as well as fractional values are permissible. Since both upper limit and lower limit of classes are included for frequency calculation, this method is called inclusive method.

Exclusive method

In this method, the values which are equal to upper limit of a class are not included in that class and instead they would be included in the next class. The upper limit is not at all taken into consideration or in other words it is always excluded from the consideration. Hence this method is called exclusive method .

Example 3.13

The marks scored by 50 students in an examination are given as follows:

23, 25, 36, 39, 37, 41, 42, 22, 26, 35, 34, 30, 29, 27, 47, 40, 31, 32, 43, 45, 34, 46, 23, 24, 27, 36, 41, 43, 39, 38, 28, 32, 42, 33, 46, 23, 34, 41, 40, 30, 45, 42, 39, 37, 38, 42, 44, 46, 29, 37.

It can be observed from this data set that the marks of 50 students vary from 22 to 47. If it is decided to divide this group into 6 smaller groups, we can have the boundary lines fixed as 25, 30, 35, 40, 45 and 50 marks. Then, we form the six groups with the boundaries as 21 - 25, 26 - 30, 31 - 35, 36 - 40, 41 – 45 and 46 - 50.

The continuous frequency distribution formed by inclusive and exclusive methods are displayed in Table 3.13(i) and Table 3.13(ii), respectively.

Table 3.13(i)

Marks secured by students (Inclusive Method)

Marks	x, Integer value	Tally Marks	No. of Students
21-25	$21 \le x \le 25$	𝍸 /	6
26-30	$26 \le x \le 30$	𝍸 ///	8
31-35	$31 \le x \le 35$	𝍸 ///	8
36-40	$36 \le x \le 40$	𝍸 𝍸 //	12
41-45	$41 \le x \le 45$	𝍸 𝍸 //	12
46-50	$46 \le x \le 50$	////	4
		Total	**50**

Ch3_11th Stats.indd 57 07/12/2021 12:00:33

Table 3.13(ii)

Marks secured by students (Exclusive Method)

Marks	x, Integer value	Tally Marks	No. of Students
20-25	$20 \leq x < 25$	𝍸	5
25-30	$25 \leq x < 30$	𝍸 //	7
30-35	$30 \leq x < 35$	𝍸 ////	9
35-40	$35 \leq x < 40$	𝍸 𝍸 /	11
40-45	$40 \leq x < 45$	𝍸 𝍸 //	12
45-50	$45 \leq x < 50$	𝍸 /	6
		Total	**50**

True class intervals

In the case of continuous variables, we take the classes in such a way that there is no gap between successive classes. The classes are defined in such a way that the upper limit of each class is equal to lower limit of the succeeding class. Such classes are known as true classes. The inclusive method of forming class intervals are also known as not-true classes. We can convert the not-true classes into true-classes by subtracting 0.5 from the lower limit of the class and adding 0.5 to the upper limit of each class like 19.5 - 25.5, 25.5 - 30.5, 30.5 – 35.5, 35.5 – 40.5, 40.5 - 45.5, 45.5 – 50.5.

Open End Classes

When a class limit is missing either at the lower end of the first class interval or at the upper end of the last classes or when the limits are not specified at both the ends, the frequency distribution is said to be the frequency distribution with open end classes.

Example 3.14

Salary received by 113 workers in a factory are classified into 6 classes. The classes and their frequencies are displayed in Table 3.14 Since the lower limit of the first class and the upper limit of the last class are not specified, they are open end classes.

Table 3.14

Open-Ended Frequency Table

Salary Range in Rs.	Number of workers
Below 10000	18
10000 - 20000	23
20000 - 30000	30

Ch3_11th Stats.indd 58 07/12/2021 12:00:33

30000 - 40000	20
40000 - 50000	12
50000 and above	10

3.6.4 Guidelines on Compilation of Continuous Frequency Distribution

The following guidelines may be followed for compiling the continuous frequency distribution.

- The values given in the data set must be contained within one (and only one) class and overlapping classes must not occur.
- The classes must be arranged in the order of their magnitude.
- Normally a frequency distribution may have 8 to 10 classes. It is not desirable to have less than 5 and more than 15 classes.
- Frequency distributions having equal class widths throughout are preferable. When this is not possible, classes with smaller or larger widths can be used. Open ended classes are acceptable but only in the first and the last classes of the distribution.
- It should be noted that in a frequency distribution, the first class should contain the lowest value and the last class should contain the highest value.
- The number of classes may be determined by using the Sturges formula $k = 1 + 3.322\log_{10}N$, where N is the total frequency and k is the number of classes.

3.7 Cumulative Frequency Distribution

Cumulative frequency corresponding to a class interval is defined as the total frequency of all values less than upper boundary of the class. A tabular arrangement of all cumulative frequencies together with the corresponding classes is called a cumulative frequency distribution or cumulative frequency table.

The main difference between a frequency distribution and a cumulative frequency distribution is that in the former case a particular class interval according to how many items lie within it is described, whereas in the latter case the number of items that have values either above or below a particular level is described.

There are two forms of cumulative frequency distributions, which are defined as follows:

Ch3_11th Stats.indd 59 07/12/2021 12:00:33

(i) *Less than Cumulative Frequency Distribution:* In this type of cumulative frequency distribution, the cumulative frequency for each class shows the number of elements in the data whose magnitudes are less than the upper limit of the respective class.

(ii) *More than Cumulative Frequency Distribution:* In this type of cumulative frequency distribution, the cumulative frequency for each class shows the number of elements in the data whose magnitudes are larger than the lower limit of the class.

Example 3.15

Construct less than and more than cumulative frequency distribution tables for the following frequency distribution of orders received by a business firm over a number of weeks during a year.

Number of order received	0 - 4	5 - 9	10 - 14	15 - 19	20 - 24
Number of weeks	2	8	11	14	6
Number of order received	25 - 29	30 - 34	35 - 39	40 - 44	45 - 49
Number of weeks	4	3	2	1	1

Solution:

For the data related to the number of orders received per week by a business firm during a period of one year, the less than and more than cumulative frequencies are computed and are given in Table 3.15

Table 3.15

Cumulative Frequency Distribution for the number of orders received by a Business firm

Given data		**Less than ogive**		**More than ogive**	
Number of Orders Received	**Number of Weeks**	**Upper limit**	**Less than Cumulative Frequencies**	**Lower limit**	**More Than Cumulative Frequencies**
0 – 4	2	4	2	0	52
5 – 9	8	9	10	5	50
10 – 14	11	14	21	10	42
15 – 19	14	19	35	15	31
20 – 24	6	24	41	20	17
25 – 29	4	29	45	25	11
30 – 34	3	34	48	30	7

Ch3_11th Stats.indd 60 07/12/2021 12:00:33

35 – 39	2	39	50	35	4
40 – 44	1	44	51	40	2
45 – 49	1	49	52	45	1

Relative-Cumulative Frequency Distributions

The relative cumulative frequency is defined as the ratio of the cumulative frequency to the total frequency. The relative cumulative frequency is usually expressed in terms of a percentage. The arrangement of relative cumulative frequencies against the respective class boundaries is termed as relative cumulative frequency distribution or percentage cumulative frequency distribution.

Example 3.16

For the data given in Example 3.15 find the relative cumulative frequencies.

Solution:

For the data given in Example 3.15 the less-than and more-than cumulative frequencies are obtained and given in Table 3.15 The relative cumulative frequency is computed for each class by dividing the respective class cumulative frequency by the total frequency and is expressed as a percentage. The cumulative frequencies and related cumulative frequencies are tabulated in Table 3.16

Table 3.16
Relative Cumulative Frequency Distribution for the number of orders received by a Business Firm

Number of Orders Received	Number of Weeks	Less than Cumulative Frequencies	More Than Cumulative Frequencies	Relative (less than) Cumulative Frequencies	Relative (more than) Cumulative Frequencies
0 – 4	2	2	52	3.85	100.00
5 – 9	8	10	50	19.23	96.15
10 – 14	11	21	42	40.38	80.77
15 – 19	14	35	31	67.31	59.62
20 – 24	6	41	17	78.85	32.69
25 – 29	4	45	11	86.54	21.15
30 – 34	3	48	7	92.31	13.46
35 – 39	2	50	4	96.15	7.69
40 – 44	1	51	2	98.08	3.85
45 – 49	1	52	1	100.00	1.92

Ch3_11th Stats.indd 61 07/12/2021 12:00:33

3.8 Bivariate Frequency Distributions

It is known that the frequency distribution of a single variable is called univariate distribution. When a data set consists of a large mass of observations, they may be summarized by using a two-way table. A two-way table is associated with two variables, say X and Y. For each variable, a number of classes can be defined keeping in view the same considerations as in the univariate case. When there are m classes for X and n classes for Y, there will be m × n cells in the two-way table. The classes of one variable may be arranged horizontally, and the classes of another variable may be arranged vertically in the two way table. By going through the pairs of values of X and Y, we can find the frequency for each cell. The whole set of cell frequencies will then define a bivariate frequency distribution. In other words, a bivariate frequency distribution is the frequency distribution of two variables.

Table 3.17 shows the frequency distribution of two variables, namely, age and marks obtained by 50 students in an intelligent test. Classes defined for marks are arranged horizontally (rows) and the classes defined for age are arranged vertically (columns). Each cell shows the frequency of the corresponding row and column values. For instance, there are 5 students whose age fall in the class 20 – 22 years and their marks lie in the group 30 – 40.

Table 3.17

Bivariate Frequency Distribution of Age and Marks

Marks	Age in Years				Total
	16 – 18	18 - 20	20 - 22	22 – 24	
10 – 20	2	1	1	-	4
20 – 30	3	2	3	1	9
30 – 40	3	3	5	6	17
40 – 50	2	2	3	4	11
50 – 60	-	1	2	2	5
60 – 70	-	1	2	1	4
Total	**10**	**10**	**16**	**14**	**50**

3.9 Stem and Leaf Plot (Stem and Leaf Diagram)

The *stem* and *leaf* plot is another method of organizing data and is a combination of sorting and graphing. It is an alternative to a tally chart or a grouped frequency distribution. It retains the original data without loss of information. This is also a type of bar chart, in which the numbers themselves would form the bars.

Ch3_11th Stats.indd 62 07/12/2021 12:00:33

Stem and *leaf* plot is a type of data representation for numbers, usually like a table with two columns. Generally, *stem* is the label for left digit (leading digit) and leaf is the label for the right digit (trailing digit) of a number.

For example, the leaf corresponding to the value 63 is 3. The digit to the left of the leaf is called the stem. Here the stem of 63 is 6. Similarly for the number 265, the leaf is 5 and the stem is 26.

The elements of data 252, 255, 260, 262, 276, 276, 276, 283, 289, 298 are expressed in Stem and leaf plot as follows:

Actual data	Stem (Leading digits)	Leaf (Trailing digits)
252, 255	25	2 5
260, 262	26	0 2
276, 276, 276	27	6 6 6
283, 289	28	3 9
298	29	8

NOTE

In Stem and Leaf plot, we display stem and leaf columns only.

From the Stem and Leaf plot, we find easily the smallest number is 252 and the largest number is 298.

Also, in the class 270 – 280 we find 3 items are included and that group has the highest frequency.

The procedure for plotting a Stem and Leaf diagram is illustrated through an example given below:

Example 3.17

Construct a Stem and Leaf plot for the given data.

1.13, 0.72, 0.91, 1.44, 1.03, 0.88, 0.99, 0.73, 0.91, 0.98, 1.21, 0.79, 1.14, 1.19, 1.08, 0.94, 1.06, 1.11, 1.01, 1.39

Solution:

Step 1: Arrange the data in the ascending order of magnitude:

0.72, 0.73, 0.79, 0.88, 0.91, 0.91, 0.94, 0.98, 0.99, 1.01,

1.03, 1.06, 1.08, 1.11, 1.13, 1.14, 1.19, 1.21, 1.39, 1.44

Step 2: Separate the data according to the first digit as shown

0.72, 0.73, 0.79

0.88

0.91, 0.91, 0.94, 0.98, 0.99

Ch3_11th Stats.indd 63 07/12/2021 12:00:33

1.01, 1.03, 1.06, 1.08
1.11, 1.13, 1.14, 1.19
1.21
1.39
1.44

Step 3: Now construct the stem and leaf plot for the above data.

Stem (Leading digits)	Leaves (Trailing digits)
0.7	2 3 9
0.8	8
0.9	1 1 4 8 9
1.0	1 3 6 8
1.1	1 3 4 9
1.2	1
1.3	9
1.4	4



NOTE

Stem and Leaf diagrams normally have one set of leaves for any stem value but either two or more sets of leaves are possible.

Using a Stem and Leaf plot, finding the Mean, Median, Mode and Range

We know how to create a stem and leaf plot. From this display, let us look at how we can use it to analyze data and draw conclusions. First, let us recall some statistical terms already we used in the earlier classes.

- The mean is the data value which gives the sum of all the data values, divided by the number of data values.
- The median is the data value in the middle when the data is ordered from the smallest to the largest.
- The mode is the data value that occurs most often. On a stem and leaf plot, the mode is the repeated leaf.
- The range is the difference between the highest and the least data value.

Example 3.18

Determine the mean, median, mode and the range on the stem and leaf plot given below:

Stem	Leaf
25	2 5
26	0 2
27	6 6 6
28	3 9
29	8

Ch3_11th Stats.indd 64 07/12/2021 12:00:33

Solution:

From the display, combine the stem with each of its leaves. The values are in the order from the smallest to the largest on the plot. Therefore, keep them in order and list the data values as follows:

252, 255, 260, 262, 276, 276, 276, 283, 289, 298

To determine the mean, add all the data values and then divide the sum by the number of data values.

$(252 + 255 + 260 + 262 + 276 + 276 + 276 + 283 + 289 + 298) \div 10$

$= 2727 \div 10 = 272.7$

Mean = 272.7.

The data is already arranged in ascending order. Therefore, identify the number in the middle position of the data. In this case, two data values share the middle position. To find the median, find the mean of these two middle data values.

The two middle numbers are 276 and 27 6.

The median is $(276 + 276) \div 2 = 276$.

The mode is the data value that occurs more frequently. Looking at the stem and leaf plot, we can see the data value 276 appears thrice.

Therefore the mode is 276.

Recall that the range is the difference of the greatest and least values. On the stem and leaf plot, the greatest value is the last value and the smallest value is the first value.

The range is 298 – 252 = 46.

Points to Remember

- Data array enables one to extract supplementary information from the data.
- It is essential that statistical data must be presented in a condensed form through classification.
- The process of dividing the data into different groups or classes which are as homogeneous as possible within the groups or classes, but heterogeneous between themselves is said to be classification

Ch3_11th Stats.indd 65 07/12/2021 12:00:33

- There are four methods of classification namely (1) classification by time or chronological classification (2) classification by space or spatial classification (3) classification by attribute or qualitative classification and (4) classification by size or quantitative classification.
- General tables contain a collection of detailed information including all that is relevant to the subject or theme in row-column format.
- The title of a statistical table must be framed in such a way that it describes the contents of the table appropriately
- Every statistical table is assigned with a table number which helps to identify the appropriate table for the intended purpose and to distinguish one table from the other, in the case of more than one table.
- Captions or column headings and stubs or row headings must be given in short and must be self-explanatory.
- The information contained in the summary table aim at comparison of data, and enable conclusion to be drawn.
- A simple frequency distribution , also called as frequency table, is a tabular arrangement of data values together with the number of occurrences, called frequency, of such values.
- A standard form into which the large mass of data is organized into classes or groups along with the frequencies is known as a grouped frequency distribution.
- The Stem and Leaf plot is another method of organizing data and is a combination of sorting and graphing. It retains the original data without loss of information.
- Stem and Leaf plot is a type of data representation for numbers, usually like a table with two columns. Generally stem is the label for left digit (leading digit) and leaf is the label for the right digit(trailing digit) of a number.

EXERCISE 3

QZ67U

I. Choose the best answer:

1. Raw data means

(a) primary data (b) secondary data

(c) well classified data (d) none of these.

Ch3_11th Stats.indd 66 07/12/2021 12:00:34

2. Classification is the process of arranging the data in

(a) different rows (b) different columns

(c) different rows and columns

(d) grouping of related facts in different classes.

3. In chronological classification, data are classified on the basis of

(a) attributes (b) time (c) classes (d) location.

4. The data classified on the basis of location is known as

(a) chronological (b) geographical

(c) qualitative (d) quantitative classification

5. Column heading of a table is known as

(a) stub (b) caption (c) note (d) title

6. An arrangement of data values together with the number of occurrences forms

(a) a table (b) a frequency distribution

(c) a frequency curve (d) a cumulative distribution

7. The class interval of the type 10-14, 15-19, 20-24, 25-29, 30-34 represents

(a) inclusive type (b) exclusive type

(c) open-end type (d) none

8. In a *stem and leaf* plot, stem is the label for ---------- digit

(a) leading (b) trailing

(c) middle (d) none.

II . Fill in the blanks:

9. An arrangement of raw data in an order of magnitude or in a sequence is called _____

10. The process of arranging the primary data in a definite pattern and presenting in a systematic form is known as _____

11. The method of classifying statistical data on the basis of descriptive characteristics is called _____

12. The method of classifying data with reference to location such as countries, states, cities, districts, etc., is called _____

Ch3_11th Stats.indd 67 07/12/2021 12:00:34

13. _____is the systematic organization of statistical data in rows and columns.

14. A simple table summarising information on a single characteristic is also called a _____

15. The term stub stands for _____

16. The numerical difference between the lower and upper boundaries of a class is called _____

17. If the lower limit and the upper limit of a class are 10 and 19 respectively, the mid point of the class is _____

18. Sturges formula for finding number of classes to construct a continuous frequency table is _____

19. _____ is defined as the ratio of the cumulative frequency to the total frequency.

20. The *stem and leaf* plot retains the _____ data without loss of information.

21. _____frequency distribution is the frequency distribution of two variables.

III. Answer shortly :

22. What is an array?

23. Define classification of data.

24. List out various types of classification.

25. Define a statistical table.

26. What are known as stubs and captions?

27. What is bi-variate table?

IV. Answer in brief:

28. State the objectives of classification.

29. Give an illustration for a simple table.

30. What are the advantages of tables?

31. Define one-way and two-way table.

32. What is discrete frequency distribution?

33. Explain open-end class interval with example.

Ch3_11th Stats.indd 68 07/12/2021 12:00:34

34. What is relative-cumulative frequency distribution?

35. Find the less than and more than cumulative frequencies for the following distribution

Classes	15-20	20-25	25-30	30-35	35-40	40-45	Total
Frequencies	5	8	17	24	16	10	80

V. Answer in detail :

36. Explain various types of classification.

37. What are the precautions to be considered for tabulation of data.

38. Explain the major types of statistical tables.

39. Distinguish between inclusive method and exclusive method of forming frequency distribution with suitable examples.

40. Construct frequency distribution table for the following data by (i) inclusive method (ii) exclusive method

67,34,36,48,49,31,61,34,43,45,38,32,27, 61, 29,47, 36, 50,46,30,46,32,30,33,45,49,48, 41, 53, 36, 37, 47, 47,30, 46, 57, 39, 45, 42, 37

41. Construct a bi-variate frequency distribution table for the following data of twenty students.

Marks in Economics: 15 12 17 20 23 14 20 18 15 21 10 16 22 18 16 15 17 19 15 20

Marks in Statistics : 20 21 22 21 23 20 22 21 24 23 22 24 22 23 20 23 20 22 24 23

42. The total scores in a series of basketball matches were 216, 223, 183,219, 228, 200, 217, 208, 195, 172, 210, 213, 208, 192, 197, 185, 213, 219. Construct a stem and leaf plot for the given data.

43. Determine the mean, median, mode and the range for the given stem and leaf plot given below:

Stem (Leading digits)	Leaves (Trailing digits)
0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7

Ch3_11th Stats.indd 69 07/12/2021 12:00:34



NOTE

The teacher and students can create their own problems similar to the above and enlarge the exercise problems.



Activity

1. Collect data about the mode of transport of your school students. Classify the data and tabulate it.
2. Prepare a table for the expenditure of various food items at home for a month.
3. In your class, measure the height (or weight) of each student. Plot the results on a stem and leaf plot
4. Collect the important and relevant tables from various sources and include these in your album.
5. Prepare a frequency table for the height (or weight) of your class students using Computer Spread Sheets.

Answers:

I.1. (a) 2. (b) 3. (b) 4. (b) 5. (b) 6. (b) 7. (a) 8. (a)

II. 9. array 10. classification 11. qualitative classification

12. geographical classification 13. Tabulation

14. uni-variate table 15. row headings 16. Width 17. 14.5

18. $k = 1 + 3.322\log_N 10$ 19. Relative cumulative frequency distribution

20. Original 21. Bi-variate.

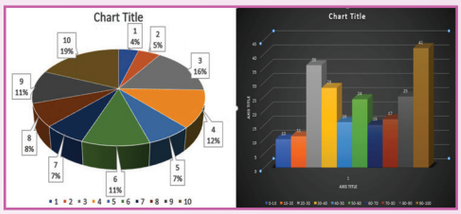
Ch3_11th Stats.indd 70 07/12/2021 12:00:34



ICT CORNER

DIAGRAMMATIC AND GRAPHICAL REPRESENTATION OF DATA

This activity is for drawing charts using excel for the given data is given in easy steps



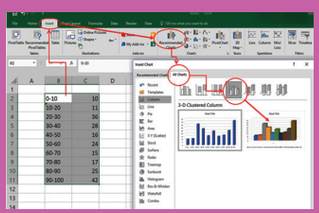
We have collection of data for the given interval. Let us draw the bar graph and pie chart for this data. Follow the steps given below:

Steps:

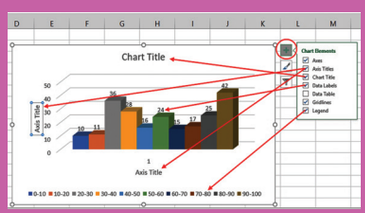
- Open an excel sheet and type the data a shown in the figure. Select all the data and click insert menu at the top so that charts menu will appear. In that select column or bar chart 3D cluster column.
- Now 3D-cluster column chart will appear for your data. You can edit this by selecting the menu chart elements which appear on the right side. Check the check boxes and edit chart title, axes titles, data labels and legend.
- Now you can beautify the chart by selecting the 2nd menu on the right-hand side called styles. (Explore)
- Now you can draw pie chart for the frequency. Select only the frequencies and select insert and in the charts select insert pie or doughnut 3D pie. You will get the pie chart. Now beautify it in the style menu on the right side.

You can copy, resize or save this chart and use it any document.

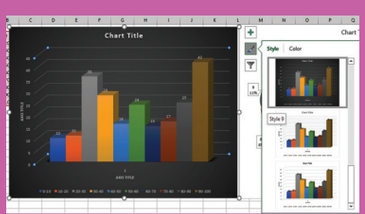
Step-1



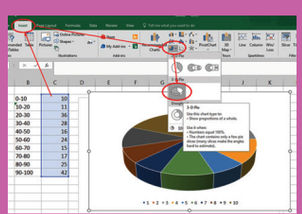
Step-2



Step-3



Step-4



Pictures are indicatives only*

URL:

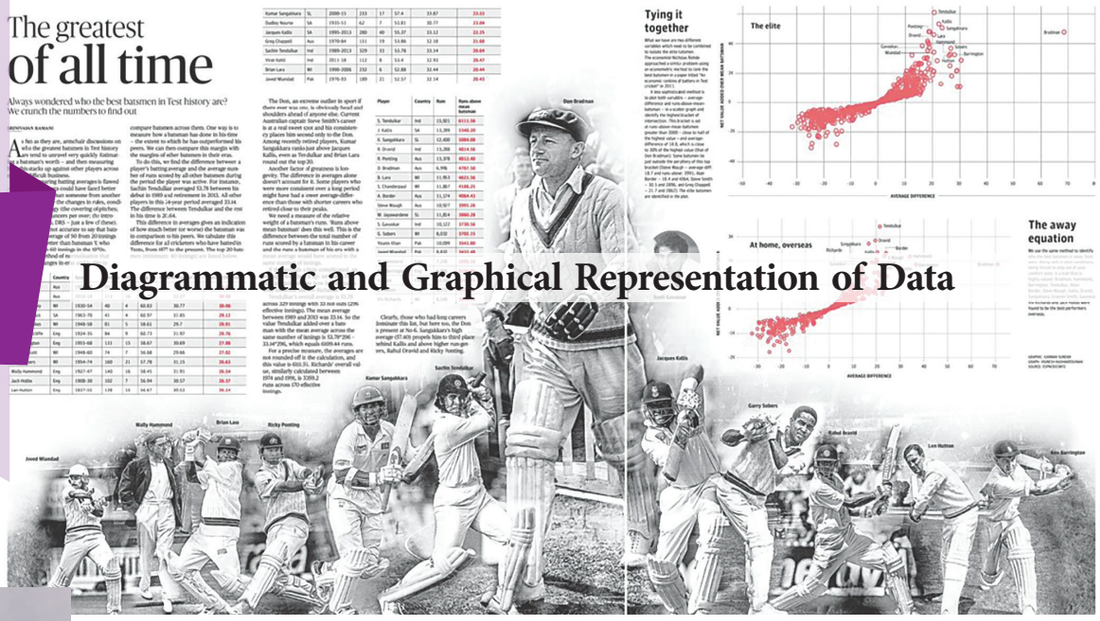
https://www.calculatorsoup.com/calculators/statistics/stemleaf.php

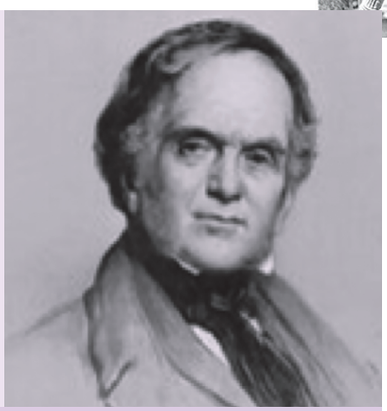
Ch3_11th Stats.indd 71 07/12/2021 12:00:36

Chapter

4

Diagrammatic and Graphical Representation of Data





William Palyfair
(22 Sep, 1759- 11 Feb, 1823)

William Playfair was a Scottish engineer and political economist, the founder of graphical methods of statistics.

He invented several types of diagrams: In 1786, the line, area and bar chart of economic data, and in 1801 the pie chart and circle graph, used to show part-whole relations.

'Good statistical inference never strays very far from data'
- Brian S Yandell

Learning Objectives

- Presents the data in diagrams
- Understands the various types of diagrams
- Compares the tabular data with diagrammatic representation of data
- Represents the data in a graph
- Enumerates the unknown value using graphs
- Distinguishes diagrammatic and graphical representation of data

Introduction

In the preceding unit, we discussed the techniques of classification and tabulation that help in condensing and presenting the data in a tabular form. These ways of presentation of data in numbers is dull and uninteresting to the common man. For that,

Ch4_11th Stats.indd 72 07/12/2021 12:02:04

the most convincing and appealing ways are highlighting the salient features of statistical data through visual/pictorial presentation using diagrams and graphs. It is an accepted fact that the pictorial representation is more appealing, attractive and has long last effect. Moreover, a layman who averse to numbers can understand the diagrams more easily. Hence, the newspapers, magazines, journals, advertisements etc., present their numerical facts through diagrams and graphs. It is imperative on the part of a student to understand and apply the pictorial representation in a real life situation.

4.1 Meaning and Significance of Diagrams and Graphs

Diagrams:

A diagram is a visual form for presenting statistical data for highlighting the basic facts and relationship which are inherent in the data. The diagrammatic presentation is more understandable and it is appreciated by everyone. It attracts the attention and it is a quicker way of grasping the results saving the time. It is very much required, particularly, in presenting qualitative data.

Graphs:

The quantitative data is usually represented by graphs.Though it is not quite attractive and understandable by a layman, the classification and tabulation techniques will reduce the complexity of presenting the data using graphs. Statisticians have understood the importance of graphical presentation to present the data in an interpretable way. The graphs are drawn manually on graph papers.

Significance of Diagrams and Graphs:

Diagrams and graphs are extremely useful due to the following reasons:

(i) They are attractive and impressive

(ii) They make data more simple and intelligible

(iii) They are amenable for comparison

(iv) They save time and labour and

(v) They have great memorizing effect.



Cricket is the game in which, graphs and charts are used predominantly to represent data than many other games.

4.2 Rules for Constructing Diagrams

While constructing diagrams for statistical data, the following guidelines are to be kept in mind:

Ch4_11th Stats.indd 73 07/12/2021 12:02:04

- A diagram should be neatly drawn in an attractive manner
- Every diagram must have a precise and suitable heading
- Appropriate scale has to be defined to present the diagram as per the size of the paper
- The scale should be mentioned in the diagram
- Mention the values of the independent variable along the X-axis and the values of the dependent variable along the Y-axis
- False base line(s) may be used in X-axis and Y-axis, if required
- Legends should be given for X-axis, Y-axis and each category of the independent variable to show the difference
- Foot notes can be given at the bottom of the diagram, if necessary.

Legend means a brief verbal description about the shades/colours applied in the chart/graph. The legend is also known as chart key. It is most often located on the right hand side of the graph/chart and can sometimes be surrounded by a border

4.3 Types of Diagrams

In practice, varieties of diagrams are used to present the data. They are explained below.

4.3.1 Simple Bar Diagram

Simple bar diagram can be drawn either on horizontal or vertical base. But, bars on vertical base are more common. Bars are erected along the axis with uniform width and space between the bars must be equal. While constructing a simple bar diagram, the scale is determined as proportional to the highest value of the variable. The bars can be coloured to make the diagram attractive. This diagram is mostly drawn for categorical variable. It is more useful to present the data related to the fields of Business and Economics.

Example 4.1

The production cost of the company in lakhs of rupees is given below.

(i) Construct a simple bar diagram.

(ii) Find in which year the production cost of the company is (a) maximum (b) minimum (c) less than 40 lakhs.

(iii) What is the average production cost of the company?

Ch4_11th Stats.indd 74 07/12/2021 12:02:04

(iv) What is the percentage increase from 2014 to 2015?

Year	Production Cost
2010	55
2011	40
2012	30
2013	25
2014	35
2015	70

Solution:

(i) We represent the above data by simple bar diagram in the following manner:

Step-1: Years are marked along the X-axis and labelled as 'Year'.

Step-2: Values of Production Cost are marked along the Y-axis and labelled as 'Production Cost (in lakhs of ₹).

Step-3: Vertical rectangular bars are erected on the years marked and whose height is proportional to the magnitude of the respective production cost.

Step-4: Vertical bars are filled with the same colours.

The simple bar diagram is presented in Fig.4.1.

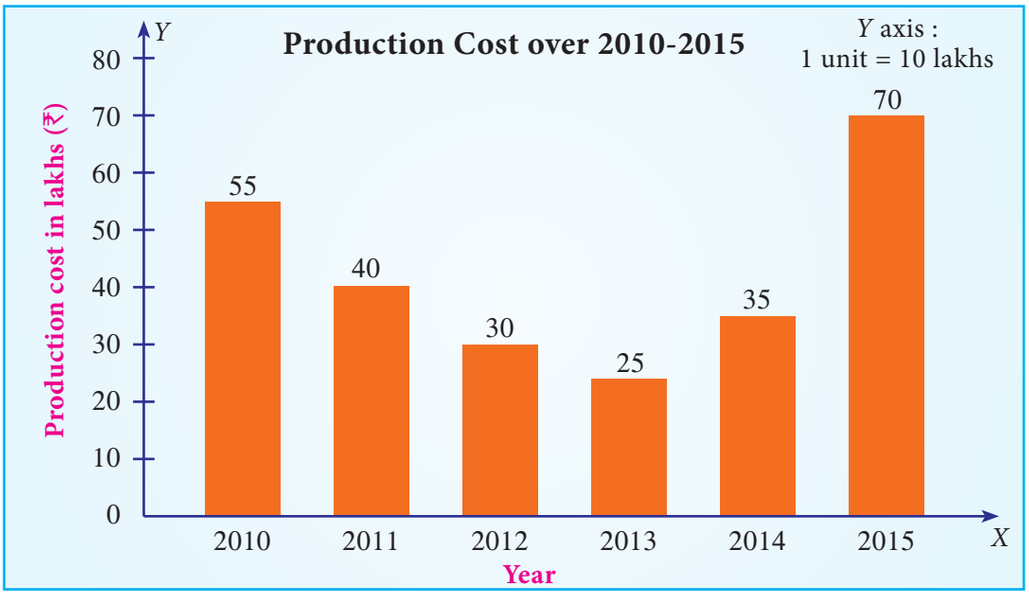


Fig 4.1

(ii) (a) The maximum production cost of the company was in the year 2015.

(b) The minimum production cost of the company was in the year 2013.

Ch4_11th Stats.indd 75 07/12/2021 12:02:04

(c) The production cost of the company during the period 2012- 2014 is less than 40 lakhs.

(iii) Average production Cost of the company

$$= \frac{55 + 40 + 30 + 25 + 35 + 70}{6}$$

$$= ₹ 42.5 \text{ Lakhs}$$

(iv) Percentage increase in the production cost of the company is

$$= \frac{70 - 35}{35} \times 100$$

$$= 100\%$$

4.3.2 Pareto Diagram:



Vilfredo Pareto (1848-1923), born in Paris in an Italian aristocratic family, studied Engineering and Mathematics at the University of Turin. During his studies at the University of Lousane in Switzerland, Pareto derived a complicated mathematical formula to prove the distribution of income and wealth in society is not random. Approximately 80% of total wealth in a society lies with only 20% of the families. The famous law about the 'Vital few and trivial many' is widely known as 'Pareto Principle' in Economics.

Pareto diagram is similar to simple bar diagram. But, in Pareto diagram, the bars are arranged in the descending order of the heights of the bars. In addition, there will be a line representing the cumulative frequencies (in %) of the different categories of the variable. The line is more useful to find the vital categories among trivial categories

Example 4.2

Administration of a school wished to initiate suitable preventive measures against breakage of equipment in its Chemistry laboratory. Information collected about breakage of equipment occurred during the year 2017 in the laboratory are given below:

Ch4_11th Stats.indd 76 07/12/2021 12:02:05

Equipment	No. of breakages
Burette	45
Conical flask	75
Test tube	150
Pipette	30

Draw Pareto Diagram for the above data. Which equipment requires more attention in order to reduce breakages?

Solution:

Since we have to find the vital few among the several, we draw Pareto diagram.

Step 1 : Arrange the equipment according to the descending order of the number of breakages.

Step 2 : Find the percentage of breakages for each equipment using the formula $= \frac{No.\ of\ Breakages}{Total\ No.\ of\ Breakages} \times 100$

Step 3 : Calculate cumulative percentage for each equipment.

Step 4 : Mark the equipment along the X-axis and the number of breakages along the Y-axis. Construct an attached simple bar diagram to this data. In an attached simple bar diagram, the vertical bars are erected adjacently.

Step 5 : Mark the cumulative no. of breakages for each equipment corresponding to the mid-point of the respective vertical bar.

Step 6 : Draw a free hand curve joining those plotted points.

Equipment	No. of Breakages (Frequency)	No. of Breakages in percentage	Cumulative No. of Breakages in percentage
Test tube	150	$\frac{150}{300} \times 100 = 50$	50
Conical flask	75	$\frac{75}{300} \times 100 = 25$	75
Burette	45	$\frac{45}{300} \times 100 = 15$	90
Pipette	30	$\frac{30}{300} \times 100 = 10$	100
Total	**300**	**100**	

Ch4_11th Stats.indd 77 07/12/2021 12:02:06

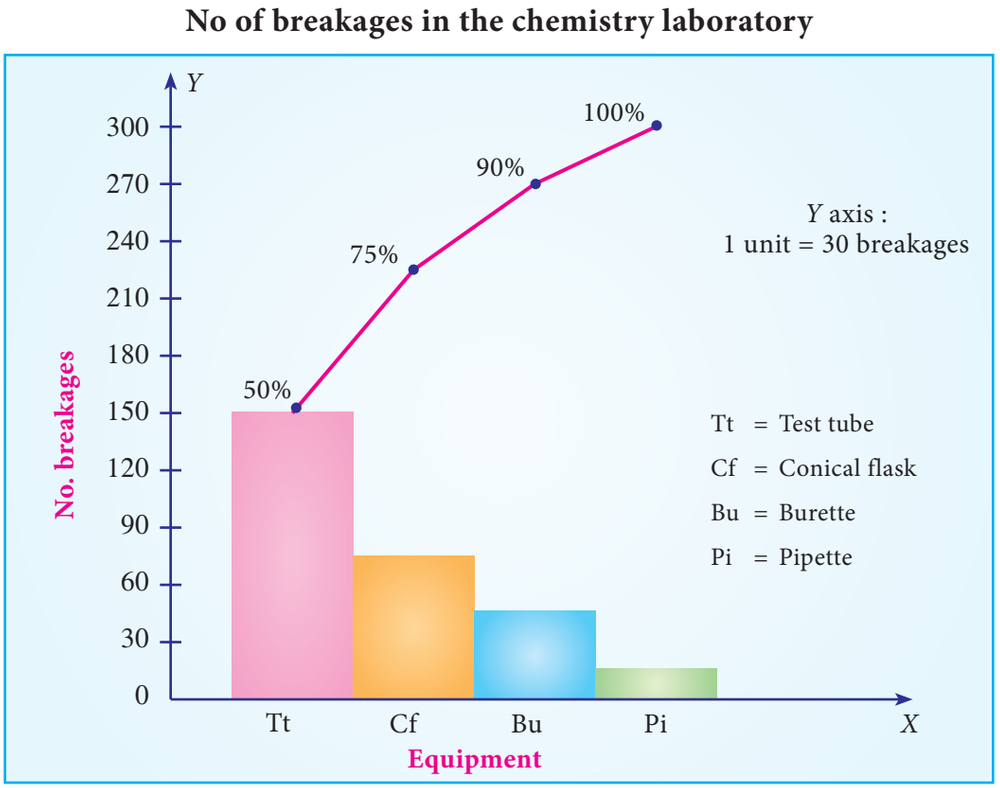


Fig 4.2: Pareto Diagram for No. of Breakages in the Chemistry Laboratory

From Fig 4.2, it can be found that 50% of breakages is due to Test tube, 25% due to Conical Flask. Therefore, the School Administration has to focus more attention on reducing the breakages of Test Tubes and Conical Flasks.

4.3.3 Multiple Bar Diagram

Multiple bar diagram is used for comparing two or more sets of statistical data. Bars with equal width are placed adjacently for each cluster of values of the variable. There should be equal space between clusters. In order to distinguish bars in each cluster, they may be either differently coloured or shaded. Legends should be provided.

Example 4.3

The table given below shows the profit obtained before and after tax payment(in lakhs of rupees) by a business man on selling cars from the year 2014 to 2017.

Year	Profit before tax	Profit after tax
2014	195	80
2015	200	87
2016	165	45
2017	140	32

(i) Construct a multiple bar diagram for the above data.

(ii) In which year, the company earned maximum profit before paying the tax?

Ch4_11th Stats.indd 78 07/12/2021 12:02:06

(iii) In which year, the company earned minimum profit after paying the tax?

(iv) Find the difference between the average profit earned by the company before paying the tax and after paying the tax.

Solution:

Since we are comparing the profit earned before and after paying the tax by the same Company, the multiple bar diagram is drawn. The diagram is drawn following the procedure presented below:

Step 1 : Years are marked along the X-axis and labeled as "Year".

Step 2 : Values of Profit before and after paying the tax are marked along the Y-axis and labeled as "Profit (in lakhs of ₹)".

Step 3 : Vertical rectangular bars are erected on the years marked, whose heights are proportional to the respective profit. The vertical bars corresponding to the profit earned before and after paying the tax in each year are placed adjacently.

Step 4 : The vertical bars drawn corresponding to the profit earned before paying the tax are filled with one type of colour. The vertical bars drawn corresponding to the profit earned after paying the tax are filled with another type of colour. The colouring procedure should be applied to all the years uniformly.

Step 5 : Legends are displayed to describe the different colours applied to the bars drawn for profit earned before and after paying the tax.

The multiple bar diagram is presented in Fig 4.3.

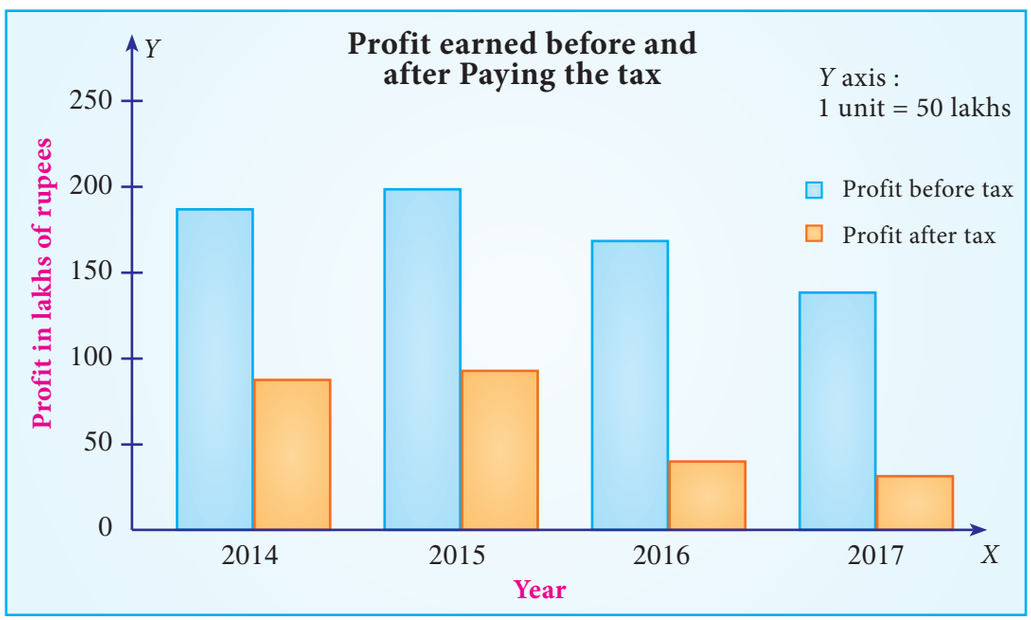


Fig 4.3 Multiple Bar Diagram for Profit by the Company earned before and after paying the Tax

Ch4_11th Stats.indd 79 07/12/2021 12:02:07

(i) The company earned the maximum profit before paying the tax in the year 2015.

(ii) The company earned the minimum profit after paying the tax in the year 2017.

(iii) The average profit earned before paying the tax $= \frac{700}{4} =$ ₹ 175 lakhs

The average profit earned after paying the tax $= \frac{244}{4} =$ ₹ 61 lakhs

Hence, difference between the average profit earned by the company before paying the tax and after paying the tax is

$= 175 - 61 =$ ₹ 114 lakhs.

4.3.4 Component Bar Diagram(Sub-divided Bar Diagram)

A component bar diagram is used for comparing two or more sets of statistical data, as like multiple bar diagram. But, unlike multiple bar diagram, the bars are stacked in component bar diagrams. In the construction of sub-divided bar diagram, bars are drawn with equal width such that the heights of the bars are proportional to the magnitude of the total frequency. The bars are positioned with equal space. Each bar is sub-divided into various parts in proportion to the values of the components. The subdivisions are distinguished by different colours or shades. If the number of clusters and the categories in the clusters are large, the multiple bar diagram is not attractive due to more number of bars. In such situation, component bar diagram is preferred.

Example 4.4

Total expenditure incurred on various heads of two schools in an year are given below. Draw a suitable bar diagram.

Expenditure Head	Amount (in lakhs)	
	School I	School II
Construction/Repairs	80	90
Computers	35	50
Laboratory	30	25
Watering plants	45	40
Library books	40	30
Total	230	235

Which school had spent more amount for

(a) construction/repairs (b) Watering plants?

Ch4_11th Stats.indd 80 07/12/2021 12:02:07

Solution :

Since we are comparing the amount spent by two schools in a year towards various expenditures with respect to their total expenditures, a component bar diagram is drawn.

Step 1 : Schools are marked along the *X*-axis and labeled as "School".

Step 2 : Expenditure Head are marked along the *Y*-axis and labeled as "Expenditure (₹ in lakhs)".

Step 3 : Vertical rectangular bars are erected for each school, whose heights are proportional to their respective total expenditure.

Step 4 : Each vertical bar is split into components in the order of the list of expenditure heads. Area of each rectangular box is proportional to the frequency of the respective expenditure head/component. Rectangular boxes for each school are coloured with different colours. Same colours are applied to the similar expenditure heads for each school.

Step 5 : Legends are displayed to describe the colours applied to the rectangular boxes drawn for various expenditure heads.

The component bar diagram is presented in Fig 4.4.

Expenditure Head	Amount (₹ in lakhs)			
	School I		School II	
	Amount Spent	Cumulative Amount Spent	Amount Spent	Cumulative Amount Spent
Construction/Repairs	80	80	90	90
Computers	35	115	50	140
Laboratory	30	145	25	165
Watering plants	45	190	40	205
Library books	40	230	30	235

Ch4_11th Stats.indd 81 07/12/2021 12:02:07

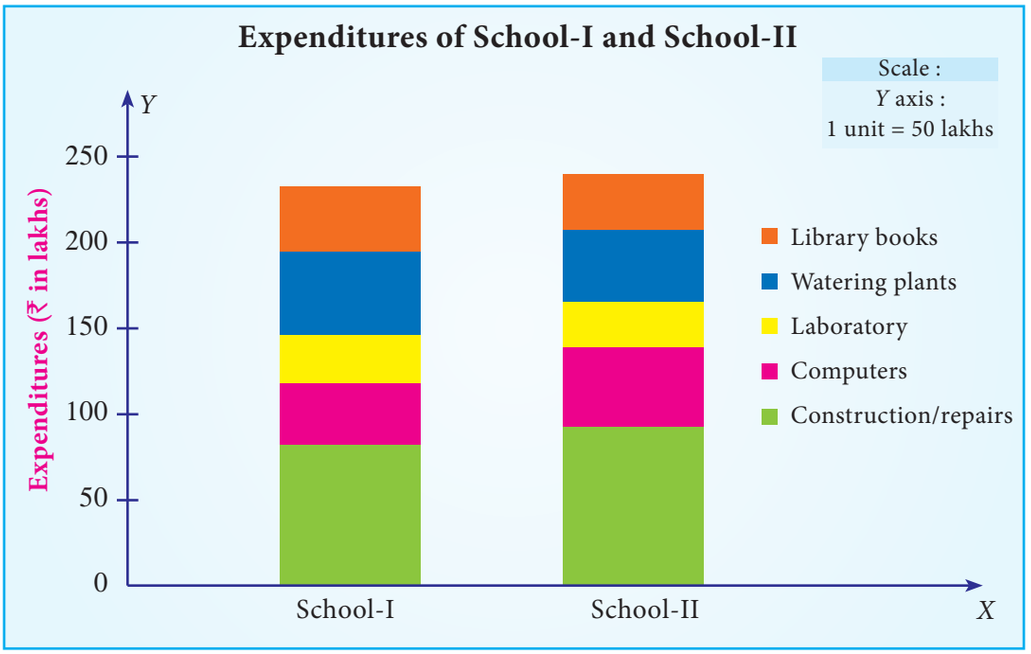


Fig 4.4 Component Bar diagram for expenditures of School I and School II

(i) School- II had spent more amount towards Construction/Repairs.

(ii) School- I had spent more amount towards Watering plants.

4.3.5 Percentage Bar Diagram

Percentage bar diagram is another form of component bar diagram. Here, the heights of the components do not represent the actual values, but percentages. The main difference between sub-divided bar diagram and percentage bar diagram is that, in the former, the height of the bars corresponds to the magnitude of the value. But, in the latter, it corresponds to the percentages. Thus, in the component bar diagram, heights of the bars are different, whereas in the percentage bar diagram, heights are equal corresponding to 100%. Hence, percentage bar diagram will be more appealing than sub-divided bar diagram. Also, comparison between components is much easier using percentage bar diagram.

Example 4.5

Draw the percentage sub-divided bar diagram to the data given in Example 4.4. Also find

(i) The percentage of amount spent for computers in School I

(ii) What are the expenditures in which School II spent more than School I.

Ch4_11th Stats.indd 82 07/12/2021 12:02:07

Solution:

Since we are comparing the amount spent by two schools in a year towards various expenditures with respect to their total expenditures in percentages, a percentage bar diagram is drawn.

Step 1 : Schools are marked along the *X*-axis and labeled as "School".

Step 2 : Amount spent in percentages are marked along the Y-axis and labeled as "Percentage of Expenditure (₹ in lakhs)".

Step 3 : Vertical rectangular bars are erected for each school, whose heights are taken to be hundred.

Step 4 : Each vertical bar is split into components in the order of the list of percentage expenditure heads. Area of each rectangular box is proportional to the percentage of frequency of the respective expenditure head/component. Rectangular boxes for each school are coloured with different colours. Same colours are applied to the similar expenditure heads for each school.

Step 5 : Legends are displayed to describe the colours applied to the rectangular boxes drawn for various expenditure heads.

The percentage bar diagram is presented in Fig 4.5.

Expenditure Head	Amount (₹ in lakhs)					
	School I			School II		
	Amount spent	Percentage of Amount spent	Cumulative Percentage	Amount spent	Percentage of Amount spent	Cumulative Percentage
Construction / Repairs	80	35	35	90	38	38
Computers	35	15	50	50	21	59
Laboratory	30	13	63	25	11	70
Watering plants	45	20	83	40	17	87
Library books	40	17	100	30	13	100
Total	**230**	**100**		**235**	**100**	

Ch4_11th Stats.indd 83 07/12/2021 12:02:07

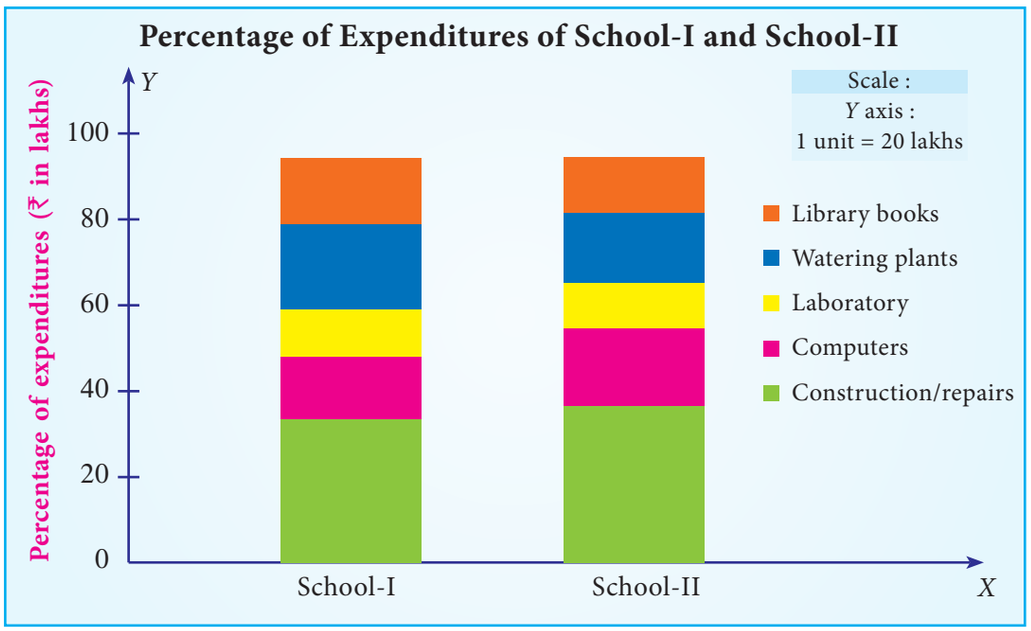


Fig 4.5 Percentage Bar diagram for expenditures of School I and School II

(i) 21% of the amount was spent for computers in School I

(ii) 38% of expenditure was spent for construction/Repairs by School II than School I.

4.3.6 Pie Diagram

The Pie diagram is a circular diagram. As the diagram looks like a pie, it is given this name. A circle which has 360° is divided into different sectors. Angles of the sectors, subtending at the center, are proportional to the magnitudes of the frequency of the components.

Procedure:

The following procedure can be followed to draw a Pie diagram for a given data:

(i) Calculate total frequency, say, N.

(ii) Compute angles for each component using the formula. $\frac{class\ frequemcy}{N} \times 360$

(iii) Draw a circle with radius of sufficient length as a horizontal line.

(iv) Draw the first sector in the anti-clockwise direction at an angle calculated for the first component.

(v) Draw the second sector adjacent to the first sector at an angle corresponding to the second component.

Ch4_11th Stats.indd 84 07/12/2021 12:02:08

(vi) This process may be continued for all the components.

(vii) Shade/colour each sector with different shades/colours.

(viii) Write legends to each component.

Example 4.6

Draw a pie diagram for the following data (in hundreds) of house hold expenditure of a family.

Items	Expenditure
Food	87
Clothing	24
Recreation	11
Education	13
Rent	25
Miscellaneous	20

Also find

(i) The central angle of the sector corresponding to the expenditure incurred on Education

(ii) By how much percentage the recreation cost is less than the Rent.

Solution :

The following procedure is followed to draw a Pie diagram for a given data:

(i) Calculate the total expenditure, say, N.

(ii) Compute angles for each component food, clothing, recreation, education, rent and miscellaneous using the formula $\frac{class\ frequency}{N} \times 360$

Item	Expenditure	Angle of the circle
Food	87	$\frac{87}{180} \times 360 = 174$
Clothing	24	$\frac{24}{180} \times 360 = 48$
Recreation	11	$\frac{11}{180} \times 360 = 22$
Education	13	$\frac{13}{180} \times 360 = 26$
Rent	25	$\frac{25}{180} \times 360 = 50$

Ch4_11th Stats.indd 85 07/12/2021 12:02:09

Miscellaneous	20	$\frac{20}{180} \times 360 = 40$
Total	N=180	360

(iii) Draw a circle with radius of sufficient length as a horizontal line.

(iv) Draw the first sector in the anti-clockwise direction at an angle calculated for the first component food.

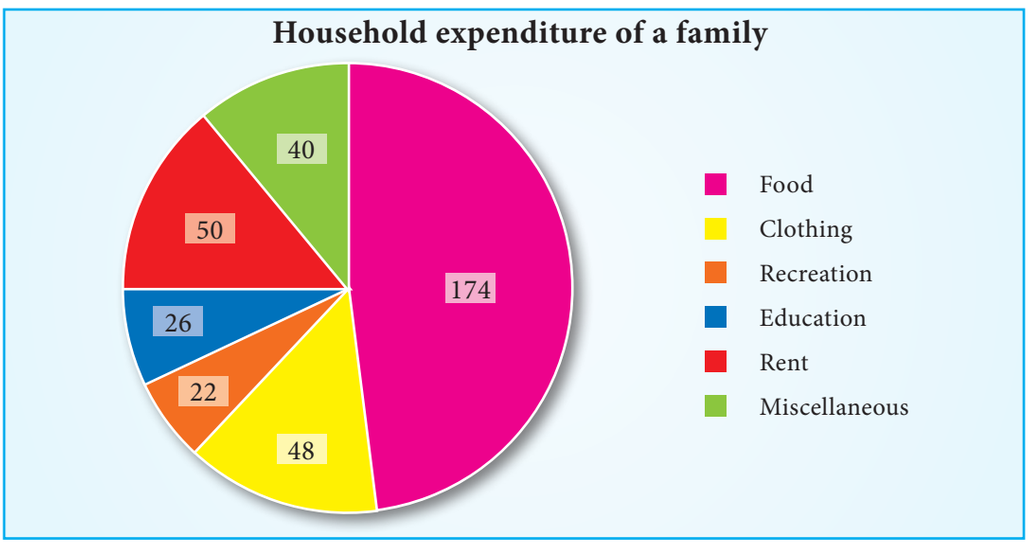
(v) Draw the second sector adjacent to the first sector at an angle corresponding to the second component clothing.

(vi) This process is continued for all the components namely recreation, education, rent and miscellaneous.

(vii) Shade/colour each sector with different shades/colours.

(viii) Write legends to each component.

The pie diagram is presented in Fig 4.6.



The central angle of the sector corresponding to the expenditure incurred on Education is 26°

Recreation cost is less than rent by 28°

4.3.7 Pictogram

Pictograms are diagrammatic representation of statistical data using pictures of resemblance. These are very useful in attracting attention. They are easily understood. For the purpose of propaganda, the pictorial presentations of facts are quite popular

Ch4_11th Stats.indd 86 07/12/2021 12:02:09

and they also find places in exhibitions. They are extensively used by the government organizations as well as by private institutions. If needed, scales can be fixed.

Despite its visual advantages, pictogram has limited application due to the usage of pictures resembling the data. It can express an approximate value than the given actual numerical value..

Example 4.7

The following table gives the sugarcane production in tonnes per acre for various years.

Year	2013	2014	2015	2016	2017
Sugar Cane (in tonnes per acre)	10	13	9	15	18

Represent the above data by pictogram.

Solution :

The above data is represented by pictogram in the following manner:

2013	
2014	
2015	
2016	
2017	

 = 1 tonne

Example 4.8

The Pictogram given below shows the number of persons who have traveled by train from Chennai to Rameshwaram on each day of a week

Ch4_11th Stats.indd 87 07/12/2021 12:02:10

Sunday	
Monday	
Tuesday	
Wednesday	
Thursday	
Friday	
Saturday	

= 100 persons

From the Pictogram find:

(i) Number of travelers travelled during the week

(ii) On which day there was a maximum rush in the train

(iii) The difference between the maximum and minimum number of travelers.

Solution :

(i) Here total number of is 48, and each represents 100 persons.

Hence number of travelers travelled during the week is $48 \times 100 = 4800$.

(ii) The maximum rush in the train is on Thursday.

(iii) Maximum number of persons travelled on Thursday = 10

Hence the number of persons travelled on Thursday is $10 \times 100 = 1000$

Minimum number of persons travelled on Wednesday = 4

Hence the number of persons travelled on Wednesday is $4 \times 100 = 400$

Therefore difference between maximum and minimum number of travelers is 1000-400=600 persons.

4.4 Types of Graphs

Graphical representation can be advantageous to bring out the statistical nature of the frequency distribution of quantitative variable, which may be discrete or continuous.

Ch4_11th Stats.indd 88 07/12/2021 12:02:10

The most commonly used graphs are

(i) Histogram

(ii) Frequency Polygon

(iii) Frequency Curve

(iv) Cumulative Frequency Curves (Ogives)

4.4.1 Histogram

A histogram is an attached bar chart or graph displaying the distribution of a frequency distribution in visual form. Take classes along the X-axis and the frequencies along the Y-axis. Corresponding to each class interval, a vertical bar is drawn whose height is proportional to the class frequency.

Limitations:

We cannot construct a histogram for distribution with open-ended classes. The histogram is also quite misleading, if the distribution has unequal intervals.

Example 4.9

Draw the histogram for the 50 students in a class whose heights (in cms) are given below.

Height	111 – 120	121 – 130	131 – 140	141 – 150	151 – 160	161 – 170
Number of students	4	11	15	9	8	3

Find the range, whose height of students are maximum.

Solution:

Since we are displaying the distribution of Height and Number of students in visual form, the histogram is drawn.

Step 1 : Heights are marked along the X-axis and labeled as "Height(in cms)".

Step 2 : Number of students are marked along the Y-axis and labeled as "No. of students".

Step 3 : Corresponding to each Heights, a vertical attached bar is drawn whose height is proportional to the number of students.

The Histogram is presented in Fig 4.7.

For drawing a histogram, the frequency distribution should be continuous. If it is not continuous, then make it continuous as follows.

Ch4_11th Stats.indd 89 07/12/2021 12:02:10

Height (in Cm)	No.of Students
110.5 - 120.5	4
120.5 - 130.5	11
130.5 - 140.5	15
140.5 - 150.5	9
150.5 - 160.5	8
160.5 - 170.5	3

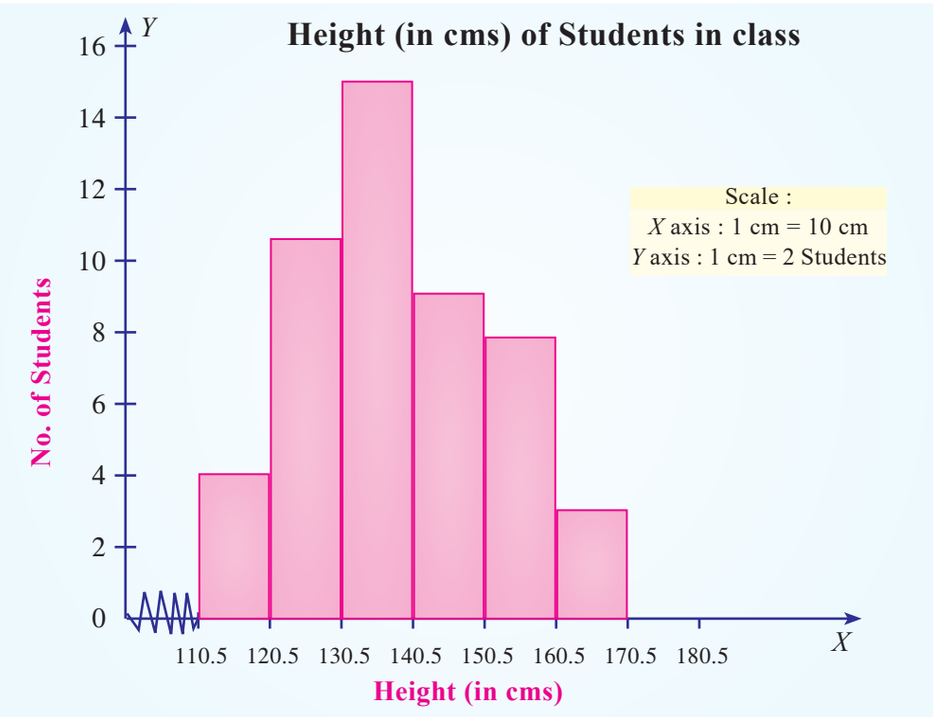


Fig 4.7 Histogram for heights of students in a class

The tallest bar shows that maximum number of students height are in the range 130.5 to 140.5 cm

Example 4.10

The following table shows the time taken (in minutes) by 100 students to travel to school on a particular day

Time	0-5	5-10	10-15	15-20	20-25
No. of Students	5	25	40	17	13

Draw the histogram. Also find:

(i) The number of students who travel to school within 15 minutes.

(ii) Number of students whose travelling time is more than 20 minutes.

Ch4_11th Stats.indd 90 07/12/2021 12:02:10

Solution:

Since we are displaying the distribution of time taken (in minutes) by 100 students to travel to school on a particular day in visual form, the histogram is drawn.

Step 1 : Time taken are marked along the X-axis and labeled as "Time (in minutes)".

Step 2 : Number of students are marked along the Y-axis and labeled as "No. of students".

Step 3 : Corresponding to each time taken, a vertical attached bar is drawn whose height is proportional to the number of students.

The Histogram is presented in Fig 4.8.

(i) 5+25+40=70 students travel to school within 15 minutes

(ii) 13 students travelling time is more than 20 minutes

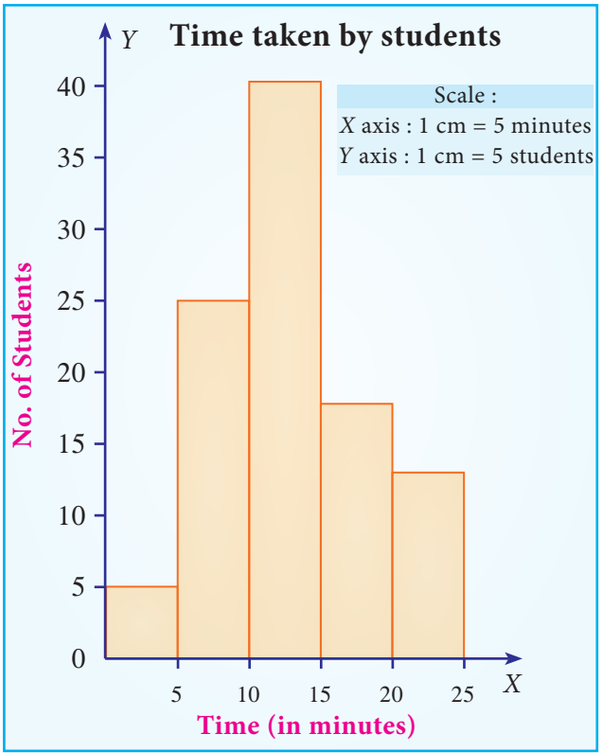


Fig 4.8 Histogram for time taken by students to travel to school

Example 4.11

Draw a histogram for the following 100 persons whose daily wages (in ₹) are given below.

Daily wages	0 – 50	50 – 100	100 – 200	200 – 250	250 – 450	450 – 500
Number of persons	5	10	16	7	48	14

Also find:

(i) Number of persons who gets daily wages less than ₹ 200?

(ii) Number of persons whose daily wages are more than ₹ 250?

Solution:

Since we are displaying the distribution of 100 persons whose daily wages in rupees in visual form, the histogram is drawn.

Step 1 : Daily wages are marked along the *X*-axis and labeled as "Daily Wages (in ₹)".

Ch4_11th Stats.indd 91 07/12/2021 12:02:10

Step 2 : Number of Persons are marked along the Y-axis and labeled as "No. of Persons".

Step 3 : Corresponding to each daily wages, a vertical attached bar is drawn whose height is proportional to the number of persons.

The Histogram is presented in Fig 4.9.

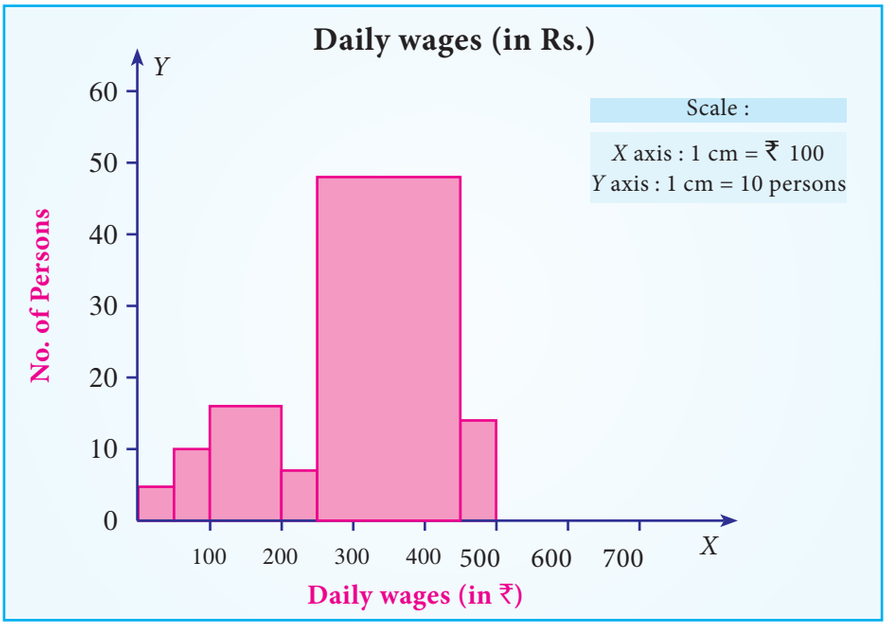


Fig 4.9 Histogram of daily wages (in ₹) for persons

(i) 5+10+16=31 persons get daily wages less than ₹ 200.

(ii) 48+14=62 persons get more than Rs 250.

4.4.2 Frequency Polygon

Frequency polygon is drawn after drawing histogram for a given frequency distribution. The area covered under the polygon is equal to the area of the histogram. Vertices of the polygon represent the class frequencies. Frequency polygon helps to determine the classes with higher frequencies. It displays the tendency of the data. The following procedure can be followed to draw frequency polygon:

(i) Mark the midpoints at the top of each vertical bar in the histogram representing the classes.

(ii) Connect the midpoints by line segments.

Example 4.12

A firm reported that its Net Worth in the years 2011-2016 are as follows:

Ch4_11th Stats.indd 92 07/12/2021 12:02:10

Year	2011 -2012	2012 – 2013	2013 – 2014	2014 – 2015	2015 - 2016
Net Worth (₹ in lakhs)	100	112	120	133	117

Draw the frequency polygon for the above data

Solution:

Since we are displaying the distribution of Net worth in the years 2011-2016, the Frequency polygon is drawn to determine the classes with higher frequencies. It displays the tendency of the data.

The following procedure can be followed to draw frequency polygon:

Step 1 : Year are marked along the *X*-axis and labeled as 'Year'.

Step 2 : Net worth are marked along the *Y*-axis and labeled as 'Net Worth (in lakhs of ₹)'.

Step 3 : Mark the midpoints at the top of each vertical bar in the histogram representing the year.

Step 4 : Connect the midpoints by line segments.

The Frequency polygon is presented in Fig 4.10.

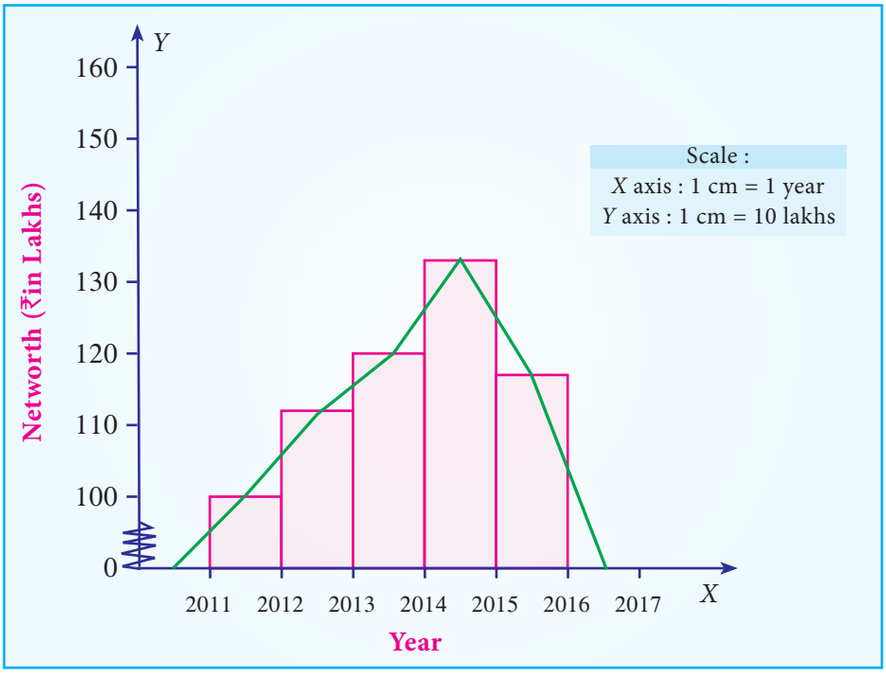


Fig 4.10 Frequency polygon for Net Worth in the years 2011-2016

4.4.3 Frequency Curve

Frequency curve is a smooth and free-hand curve drawn to represent a frequency distribution. Frequency curve is drawn by smoothing the vertices of the frequency

Ch4_11th Stats.indd 93 07/12/2021 12:02:10

polygon. Frequency curve provides better understanding about the properties of the data than frequency polygon and histogram.

Example 4.13

The ages of group of pensioners are given in the table below. Draw the Frequency curve to the following data.

Age	65 - 70	70 - 75	75 - 80	80 - 85	85 - 90
No.of pensioners	38	45	24	10	8

Solution:

Since we are displaying the distribution of Age and Number of Pensioners, the Frequency curve is drawn, to provide better understanding about the age and number of pensioners than frequency polygon.

The following procedure can be followed to draw frequency curve:

Step 1 : Age are marked along the *X*-axis and labeled as 'Age'.

Step 2 : Number of pensioners are marked along the *Y*-axis and labeled as 'No. of Pensioners'.

Step3 : Mark the midpoints at the top of each vertical bar in the histogram representing the age.

Step 4 : Connect the midpoints by line segments by smoothing the vertices of the frequency polygon

The Frequency curve is presented in Fig 4.11.

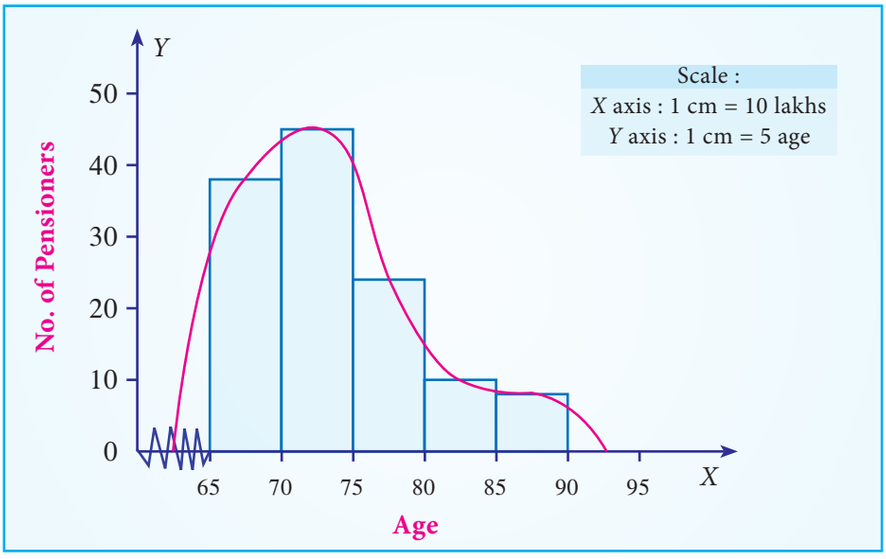


Fig 4.11 Frequency curve for Age and No. of pensioners

Ch4_11th Stats.indd 94 07/12/2021 12:02:10

4.4.4 Cumulative frequency curve (Ogive)

Cumulative frequency curve (Ogive) is drawn to represent the cumulative frequency distribution. There are two types of Ogives such as 'less thanOgive curve' and 'more thanOgive curve'. To draw these curves, we have to calculate the 'less than' cumulative frequencies and 'more than' cumulative frequencies. The following procedure can be followed to draw the ogive curves:

Less than Ogive: Less than cumulative frequency of each class is marked against the corresponding upper limit of the respective class. All the points are joined by a free-hand curve to draw **the less than ogive** curve.

More than Ogive: More than cumulative frequency of each class is marked against the corresponding lower limit of the respective class. All the points are joined by a free-hand curve to draw the **more than ogive** curve.

Both the curves can be drawn separately or in the same graph. If both the curves are drawn in the same graph, then the value of abscissa (x-coordinate) in the point of intersection is the median.

Median is a measure of central tendency, which divides the given data/distribution into two equal parts. It is discussed much in detail in Unit V

If the curves are drawn separately, median can be calculated as follows:

Draw a line perpendicular to Y-axis at $y=N/2$. Let it meet the Ogive at C. Then, draw a perpendicular line to X-axis from the point C. Let it meet the X-axis at M. The abscissa of M is the median of the data.

Example 4.14

Draw the less than Ogive curve for the following data:

Daily Wages (in Rs.)	70- 80	80- 90	90-100	100-110	110-120	120-130	130-140	140-150
No. of workers	12	18	35	42	50	45	20	8

Also, find

(i) The Median

(ii) The number of workers whose daily wages are less than ₹ 125.

Solution:

Since we are displaying the distribution of Daily Wages and No. of workers, the Ogive curve is drawn, to provide better understanding about the wages and No. of workers.

Ch4_11th Stats.indd 95 07/12/2021 12:02:11

The following procedure can be followed to draw Less than Ogive curve:

Step 1 : Daily wages are marked along the *X*-axis and labeled as "Wages(in ₹)".

Step 2 : No. of Workers are marked along the *Y*-axis and labeled as "No. of workers".

Step 3 : Find the less than cumulative frequency, by taking the upper class-limit of daily wages. The cumulative frequency corresponding to any upper class-limit of daily wages is the sum of all the frequencies less than the limit of daily wages.

Step 4 : The less than cumulative frequency of Number of workers are plotted as points against the daily wages (upper-limit). These points are joined to form less than ogive curve.

The Less than Ogive curve is presented in Fig 4.12.

Daily wages (less than)	No of workers
80	12
90	30
100	65
110	107
120	157
130	202
140	222
150	230

Daily Wages of Workers

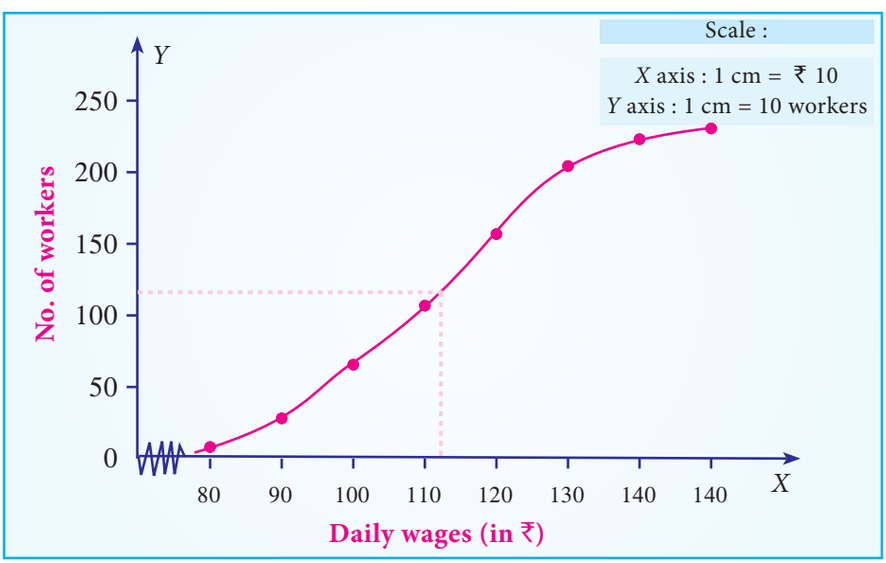


Fig 4.12 Less than Ogive curve for daily wages and number of workers

Ch4_11th Stats.indd 96 07/12/2021 12:02:11

(i) Median = ₹ 113

(ii) 183 workers get daily wages less than ₹ 125

Example 4.15

The following table shows the marks obtained by 120 students of class IX in a cycle test-I . Draw the more than Ogive curve for the following data :

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of students	2	6	8	20	30	22	18	8	4	2

Also, find

(i) The Median

(ii) The Number of students who get more than 75 marks.

Solution:

Since we are displaying the distribution marks and No. of students, the more than Ogive curve is drawn, to provide better understanding about the marks of the students and No. of students.

The following procedure can be followed to draw More than Ogive curve:

Step 1 : Marks of the students are marked along the *X*-axis and labeled as 'Marks'.

Step 2 : No. of students are marked along the *Y*-axis and labeled as 'No. of students'.

Step 3 : Find the more than cumulative frequency, by taking the lower class-limit of marks. The cumulative frequency corresponding to any lower class-limit of marks is the sum of all the frequencies above the limit of marks.

Step 4 : The more than cumulative frequency of number of students are plotted as points against the marks (lower-limit). These points are joined to form more than ogive curve.

The More than Ogive curve is presented in Fig 4.13.

Marks More than	**No of Students**
0	120
10	118

Ch4_11th Stats.indd 97 07/12/2021 12:02:11

20	112
30	104
40	84
50	54
60	32
70	14
80	6
90	2

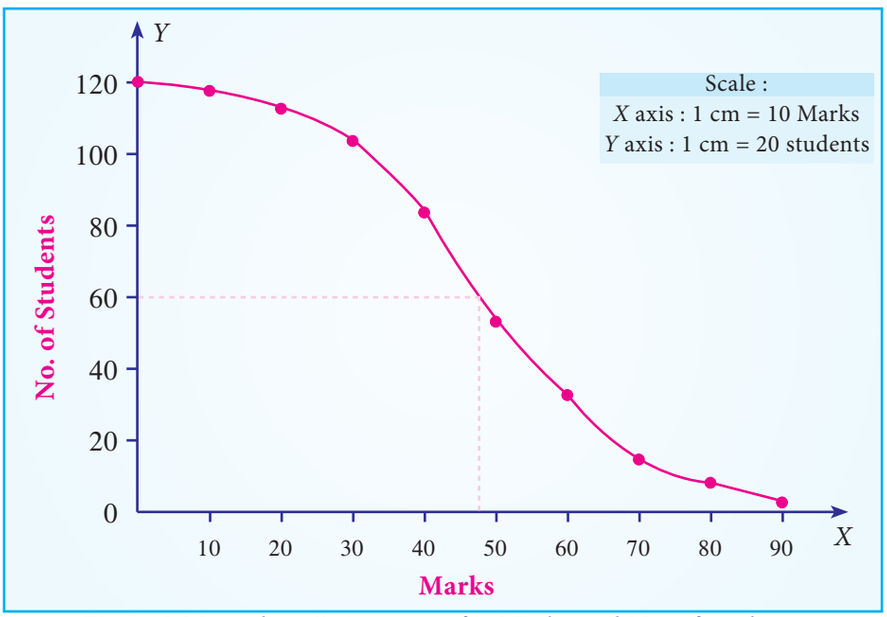


Fig 4.13 More than Ogive curve for Marks and No. of students

(i) Median = 47 students

(ii) 7 students get more than 75 marks.

Example 4.16

The yield of mangoes were recorded (in kg)are given below:

Graphically,

(i) find the number of trees which yield mangoes of less than 55 kg.

(ii) find the number of trees from which mangoes of more than 75 kg.

(iii) find the median.

Ch4_11th Stats.indd 98 07/12/2021 12:02:11

Draw the Less than and More than Ogive curves. Also, find the median using the Ogive curves

Yield (in kg)	No. of trees
40 – 50	10
50 – 60	15
60 – 70	17
70 – 80	14
80 – 90	12
90 – 100	2
Total	70

Solution:

Since we are displaying the distribution of Yield and No. of trees, the Ogive curve is drawn, to provide better understanding about the Yield and No. of trees

The following procedure can be followed to draw Ogive curve:

Step 1 : Yield of mangoes are marked along the *X*-axis and labeled as 'Yield (in Kg.)'.

Step 2 : No. of trees are marked along the *Y*-axis and labeled as 'No. of trees'.

Step 3 : Find the less than cumulative frequency, by taking the upper class-limit of Yield of mangoes. The cumulative frequency corresponding to any upper class-limit of Mangoes is the sum of all the frequencies less than the limit of mangoes.

Step 4 : Find the more than cumulative frequency, by taking the lower class-limit of Yield of mangoes. The cumulative frequency corresponding to any lower class-limit of Mangoes is the sum of all the frequencies above the limit of mangoes.

Step 5 : The less than cumulative frequency of Number of trees are plotted as points against the yield of mangoes (upper-limit). These points are joined to form less than ogive curve.

Step 6 : The more than cumulative frequency of Number of trees are plotted as points against the yield of mangoes (lower-limit). These points are joined to form more than O give curve.

Ch4_11th Stats.indd 99 07/12/2021 12:02:11

Less than Ogive		More than Ogive	
Yield less than	No. of trees	Yield greater than	No. of trees
50	10	40	70
60	25	50	60
70	42	60	45
80	56	70	28
90	68	80	14
100	70	90	2

The Ogive curve is presented in Fig 4.14.

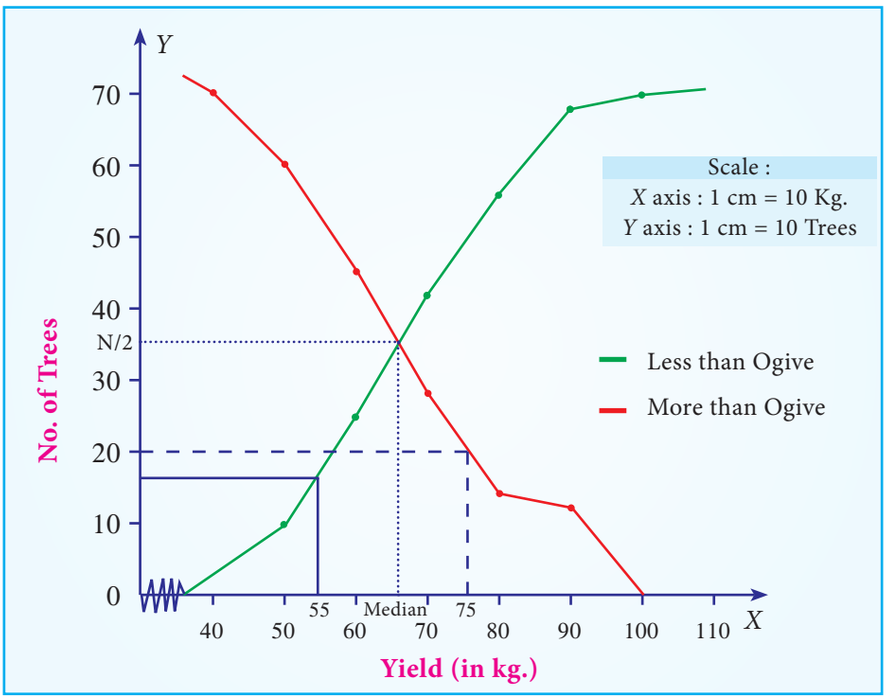


Fig 4.14 Ogive curve for Yield of mangoes and number of trees

(i) 16 trees yield less than 55 kg

(ii) 20 trees yield more than 75 kg

(iii) Median =66 kg

4.5 Comparison of Tables, Diagrams and Graphs

Data may be presented in the form of tables as well as using diagrams and graphs. Tables can be compared with graphs and diagrams on the basis of various characteristics as follows:

Ch4_11th Stats.indd 100 07/12/2021 12:02:11

(i) Table contains precise and accurate information, whereas graphs and diagrams give only an approximate idea.

(ii) More information can be presented in tables than in graphs and diagrams.

(iii) Tables require careful reading and are difficult to interpret, whereas diagrams and graphs are easily interpretable.

(iv) For common men, graphs and diagrams are attractive and more appealing than tables.

(v) Diagrams and graphs exhibit the inherent trends in the distribution easily on comparable mode than the tables.

(vi) Graphs and diagrams can be easily misinterpreted than tables.

Comparison between diagrams and graphs

(i) Diagrams can be drawn on plain papers, whereas graphs require graph papers.

(ii) Diagrams are appropriate and effective to present information about one or more variables. Normally, it is difficult to draw graphs for more than one variable in the same graph.

(iii) Graphs can be used for interpolation and/or extrapolation, but diagrams cannot be used for this purpose.

(iv) Median can be determined using graphs, but not using diagrams.

(v) Diagrams can be used for comparison of data/variables, whereas graphs can be used for determining the relationship between variables.

Points to Remember

- A diagram is a Visual aid for presenting statistical data.
- Simple bar diagram can be drawn either on horizontal or vertical base.It is used in Business and Economics.
- Pareto diagram is similar to simple bar diagram. Here the bars are arranged in the descending order of the heights of the bars. Also ,there will be a line representing the cumulative frequencies (in %) of the different categories of the variable.
- Multiple bar diagram is used for comparing two or more sets of statistical data.
- A component bar diagram is used for comparing two or more sets of statistical data, as like multiple bar diagram. But, unlike multiple bar diagram, the bars are stacked in component bar diagrams.

Ch4_11th Stats.indd 101 07/12/2021 12:02:11

- Percentage bar diagram is another form of component bar diagram. Here, the heights of the components do not represent the actual values, but percentages.
- The Pie diagram is a circular diagram. As the diagram looks like a pie, it is given this name. A circle which has 360° is divided into different sectors. Angles of the sectors, subtending at the center, are proportional to the magnitudes of the frequency of the components.
- Pictograms are diagrammatic representation of statistical data using pictures of resemblance. These are very useful in attracting attention.
- Histogram is an attached bar chart or graph displaying the distribution of a frequency distribution in visual form.
- Frequency curve is a smooth and free-hand curve drawn to represent a frequency distribution.
- Cumulative frequency curve (Ogive) is drawn to represent the cumulative frequency distribution. There are two types of Ogives such as '*less than*Ogive curve' and '*more than*Ogive curve'.
- If both the curves are drawn in the same graph, then the value of x-coordinate in the point of intersection is the median.

EXERCISE 4

AX7V4

I. Choose the best answer:

1. Which one of the following diagrams displays the class frequencies at the same height?

   (a) Simple bar diagram (b) percentage bar diagram
   (c)Sub-divided bar diagram (d) multiple bar diagram

2. Which one of the following diagrams use pictures to present the data?

   (a) Pictogram (b) Pareto Diagram (c) Pie diagram (d) Histogram

3. In which one of the following diagrams, data is transformed into angles?

   (a) Pictogram (b) Pareto Diagram (c) Pie diagram (d) Histogram

4. In which one of the following diagrams, bars are arranged in the descending order of their heights?

   (a) Pictogram (b) Pareto Diagram (c) Pie diagram (d) Histogram

5. The bars are _____ in multiple bar diagrams.

   (a) Sub-divided (b) placed adjacently
   (c) placed adjacently and sub-divided (d) sub-divided and are of equal height

Ch4_11th Stats.indd 102 07/12/2021 12:02:11

6. In which one of the following diagrams, the heights of the bars are proportional to the magnitude of the total frequency?

(a) Simple bar diagram (b) percentage bar diagram

(c) Sub-divided bar diagram (d) multiple bar diagram

II. Fill in the blanks:

7. _____ is useful for interpolation and extrapolation.

8. Frequency curve is a _____ curve drawn from the histogram.

9. _____ diagram uses pictures to present the data.

10. Circular diagram is known as _____

11. While drawing less than Ogive, cumulative frequencies are marked against the _____ of the respective classes.

12. In _____ curve, cumulative frequencies are marked against the lower limit of the respective classes.

13. Intersection of less than Ogive and more than Ogive gives _____

14. Frequency curve is drawn from frequency polygon by _____ the vertices.

15. Frequency polygon is drawn from ____

16. Area under the frequency polygon and the area of the histogram are ____

III. Answer shortly :

17. What is a diagram?

18. What is a graph?

19. List various types of diagrams.

20. List various types of graphs.

21. What is the use of simple bar diagram?

22. Distinguish the simple bar diagram and the Pareto diagram.

23. What are the different types of Ogives?

24. Write down the formula used for computing the angles of the components in Pie diagram?

25. How do you present the data using pictograms?

26. Distinguish multiple bar diagrams and component bar diagrams.

Ch4_11th Stats.indd 103 07/12/2021 12:02:11

IV. Answer in brief:

27. Write down the method of constructing Pie diagram.

28. Write down the significance of diagrams and graphs.

29. What are the features of percentage bar diagram?

30. How will you construct histogram for a given grouped data?

31. Mention any two features of tabulated data distinguishing it from diagrams and graphs.

32. Distinguish the two types of Ogives.

V. Answer in detail:

33. What are the general rules to be followed for constructing diagrams?

34. Profit (in ₹ '000) earned by a company during the period 2011-2016 are given below. Draw simple bar diagram to this data showing profit .

Year	Profit (₹'000)
2011	15000
2012	18000
2013	20000
2014	16000
2015	13000
2016	17000

Also,

(a) find the year in which the company earned maximum profit.

(b)find the year in which the company earned minimum profit.

(c) find the difference between the profit earned in the years 2012 and 2013.

35. Draw multiple bar diagram for the following data showing the monthly expenditure (in ₹ '000)of Shop A and Shop B.

Expenditure	Amount spent by	
	Shop A	Shop B
Rent	10	18
Investment	70	90
Salary	20	35
Electricity bill	10	15
Miscellaneous expenditure	5	7

Ch4_11th Stats.indd 104 07/12/2021 12:02:11

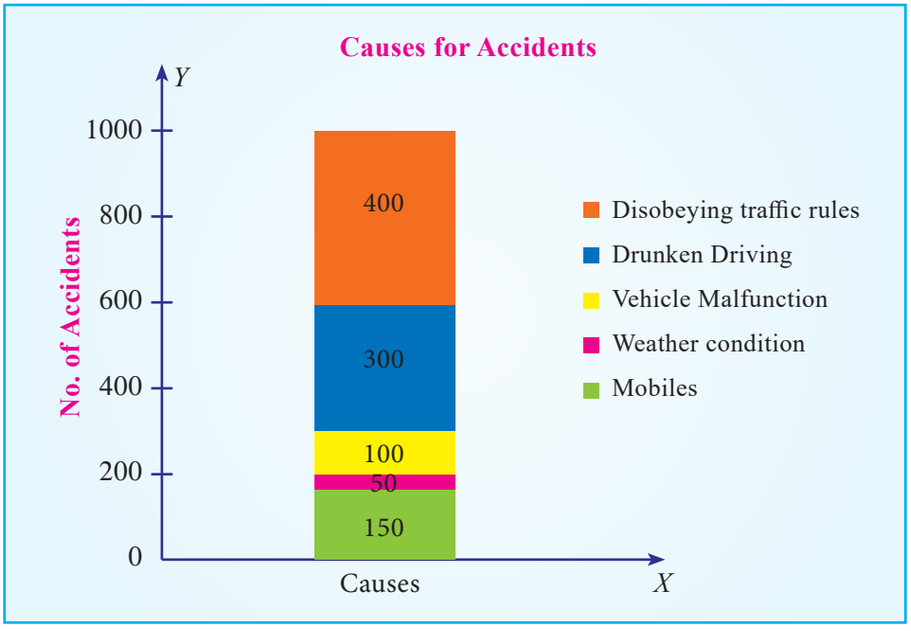
36. Draw component bar diagram to the data given in Exercise V (35).

37. Represent the data given in Exercise V(35) by percentage bar diagram.

38. The following table shows the number of students studying in three schools, preferred Walking/Cycling to go to their schools. Draw multiple bar diagram to the data.

Preference of Students	School A	School B	School C
Walking	400	550	150
Cycling	450	250	350

39. Component bar diagram showing the information about various causes and the number of road accidents occurred in a state during a period of one year are given below.



Using the diagram answer the following:

(i) ---------- caused more number of road accidents.

(ii) Minimum of road accidents is due to ----------

40. The number of students studying various undergraduate degree programmes in three Colleges are given in the following table.

Year of Study	College		
	A	B	C
First year	450	350	400

Ch4_11th Stats.indd 105 07/12/2021 12:02:11

Second year	250	250	350
Third year	225	200	300

Draw sub – divided bar diagram.

41. The following table shows the details about the expenditures (in percentages) of Indian Hotel industries under various components.

Component of Expenditure	Percentage of Expenditure
Administrative	30
Salary	20
Maintenance	14
Food and Bevarages	12
Electricity	16
Savings	8

Draw pie diagram to represent the data. Also, find

(a) the angle of the sector corresponding to Salary?

(b) the difference between the angles of the sectors corresponding to Electricity and Maintenance.

42. The pictogram given below shows the number of mails received by the Directorate of Schools over a period of 6 months.

June	✉✉✉✉✉
July	✉✉✉✉
August	✉✉✉
September	✉✉✉✉✉✉
October	✉✉
November	✉✉✉✉

✉ =100 mails.

From the above pictogram find

(a) The total no. of mails received from June to November.

(b) During which month i) maximum ii) minimum no. of mails were received.

(c) What is the percentage of decrease in the receipt of mails from September toOctober?

(d) The average number of mails received per month.

Ch4_11th Stats.indd 106 07/12/2021 12:02:14

43. Draw histogram for the following data.

Age	0-20	20-40	40-60	60-80	80-100
Number of persons	10	45	36	28	5

From the histogram say whether there is symmetry in the data or not.

44. A factory produces bolts. In the Quality Control test conducted on 500 bolts, the weights (grams) of the bolts were recorded. Draw frequency polygon and frequency curve to the data.

Weight(gram)	40-50	50-60	60-70	70-80	80-90
Number of bolts	30	90	130	210	40

45. Compare the uses of tables with diagrams and graphs for presenting statistical data.

46. Compare diagrams with graphs.

47. Details about Stipend (in Rs.) given to the apprentices in an organization are given below. Draw less than Ogive to the data.

Stipend in (Rs)	No. of apprentices
2000 – 3000	4
3000 – 4000	6
4000 – 5000	13
5000 – 6000	25
6000 – 7000	32
7000 – 8000	19
8000 – 9000	8
- 10000	3

(a) Estimate, from the curve, the number of apprentices whose stipend is less than Rs.5500.

(b) Find the median of the data.

48. Draw more than Ogive curve for the following data showing the marks secured by the students of Class XI in a school.

Marks	No. of students
0 – 10	2
10 – 20	4
20 – 30	9

Ch4_11th Stats.indd 107 07/12/2021 12:02:14

30 – 40	10
40 – 50	8
50 – 60	5
60 – 70	3
70 - 80	2

(a) Estimate the total number of students who secured marks more than 33.

(b) Find the median of the data.

49. The lifetime (in hours) of 100 bulbs observed in a Quality Control testis given below.

Lifetime (in hours)	600-650	650-700	700-750	750-800	800-850
No. of Bulbs	6	14	40	34	6

(a) Draw less than Ogive and more than Ogive curves?

(b) Find the median lifetime of bulbs graphically.

ANSWERS

I. 1. (b) 2. (a) 3. (c) 4. (b) 5. (b) 6. (c)

II. 7. Graphs 8. Smooth and free hand 9. Pictogram

10. Pie diagram 11. Upper limit 12. More than Ogive

13. Median 14. Smoothening 15. Histogram 16. equal

V 34. (a) 2013 (b) 2015 (c) 2000000

39. (1) Disobeying traffic rules (2) Weather conditions.

41. (a) 72° (b) 8° 42. (a) 2400

(b) (i) September (ii) October (c) 66.7% (d) 400

47. (a) 38 apprentices (b) 6300 48. (a) 26 students (b) 36

49. (ii) 737.5 hrs

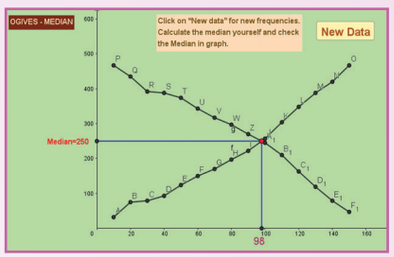
Ch4_11th Stats.indd 108 07/12/2021 12:02:15



ICT CORNER

OGIVES-MEDIAN

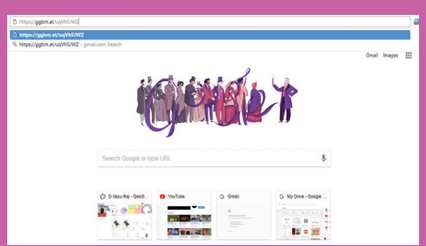
This activity is to find median of a statistical data using the Ogive curve. Also helps the students to learn how to draw Ogives.



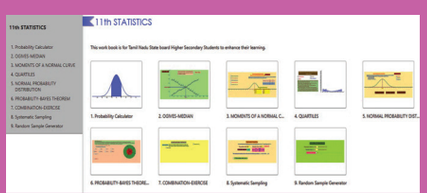
Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for statistics are given, Open the worksheet named "**Ogives-Median**"
- Ogives work sheet will open. On the right-hand side x value and frequencies are given in first two columns. 3rd and 4th columns are less than cumulative frequency and more than cumulative frequency respectively. Observe the data and compare with the curves on the left-hand side. You can press new data to get new problems as many times you want to work.
- For each problem you can see the median frequency and the median value on left side and bottom. If necessary, you can create your own problem by typing new frequencies in 2nd column.

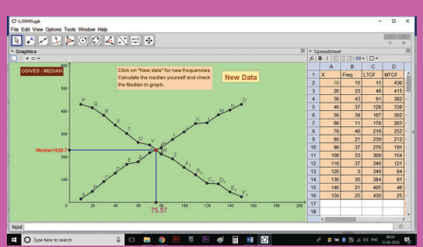
Step-1



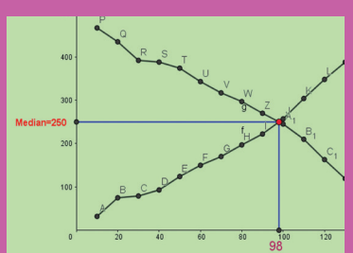
Step-2



Step-3



Step-4



Pictures are indicatives only*

URL:

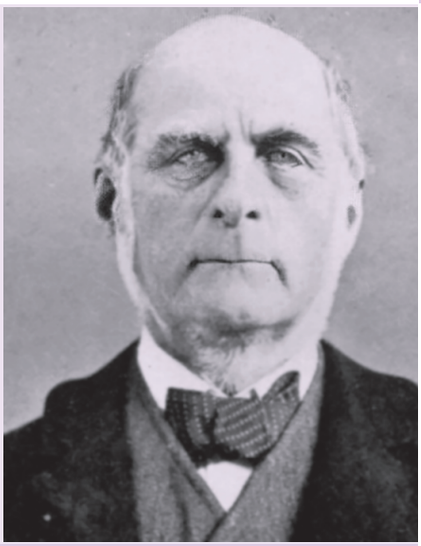
https://ggbm.at/uqVhSJWZ

MSYT8

Ch4_11th Stats.indd 109 07/12/2021 12:02:17

Chapter 5

Measures of Central Tendency



Sir Francis Galton

(16 Feb, 1822- 14 Jan, 1911)

Sir Francis Galton, born in England was a statistician, sociologist, psychologist anthropologist, eugenicist, tropical explorer, geographer, inventor, meteorologist, and psychometrician. Galton produced over 340 papers and several books. He also created the statistical concept of correlation and regression. He was the first to apply statistical methods to the study of human differences and inheritance of intelligence, and introduced the use of questionnaires and surveys for collecting data on human communities.

As an initiator of scientific meteorology, he devised the first weather map, proposed a theory of anticyclones, and was the first to establish a complete record of short-term climatic phenomena on a European scale. He also invented the Galton Whistle for testing differential hearing ability.

'The inherent inability of the human mind to grasp in it's entirely a large body of numerical data compels us to see relatively few constants that will adequately describe the data'.

- Prof. R. A. Fisher

Learning Objectives

KRNWT

- Knows the average as the representation of the entire group
- Calculates the mathematical averages and the positional averages
- Computes quartiles, Deciles, Percentiles and interprets
- Understands the relationships among the averages and stating their uses.

Introduction

Human mind is incapable of remembering the entire mass of unwieldy data.Having learnt the methods of collection and presentation of data, one has to condense the data to

Ch5_11th Stats.indd 110 07/12/2021 12:03:39

get representative numbers to study the characteristics of data. The characteristics of the data set is explored with some numerical measures namely measures of central tendency, measures of dispersion, measures ofskewness, and measures of kurtosis. This unit focuses on "Measure of central tendency". The measures of central tendency are also called "the averages".

In practical situations one need to have a single value to represent each variable in the whole set of data. Because, the values of the variable are not equal, however there is a general tendency of such observations to cluster around a particular level. In this situation it may be preferable to characterize each group of observations by a single value such that all other values clustered around it. That is why such measure is called the measure of central tendency of that group. A measure of central tendency is a representative value of the entire group of data. It describes the characteristic of the entire mass of data. It reduces the complexity of data and makes them amenable for the application of mathematical techniques involved in analysis and interpretation of data.

5.1 Definition of Measures of Central Tendency

Various statisticians have defined the word average differently. Some of the important definitions are given below:

"Average is an attempt to find one single figure to describe whole of figure"

– Clark and Sekkade

"Average is a value which is typical or representative of a set of data"

– Murray R. Speigal.

"The average is sometimes described as number which is typical of the whole group"

– Leabo.

It is clear from the above definitions that average is a typical value of the entire data and is a measures of central tendency.

5.2 Characteristics for a good statistical average

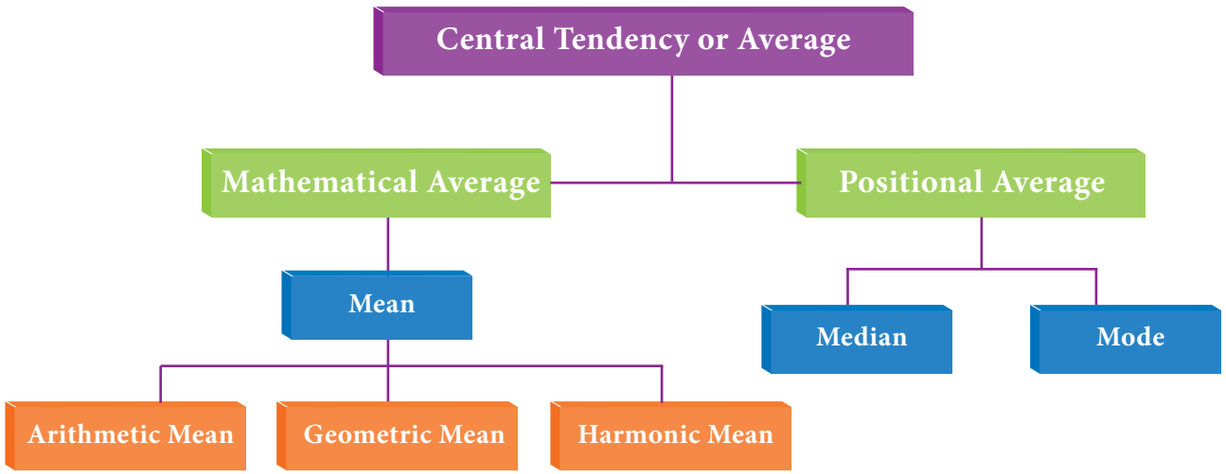
The following properties should be possessed by an ideal average.

- It should be well defined so that a unique answer can be obtained.
- It should be easy to understand, calculate and interpret.
- It should be based on all the observations of the data.

Ch5_11th Stats.indd 111 07/12/2021 12:03:39

- It should be amenable for further mathematical calculations.
- It should be least affected by the fluctuations of the sampling.
- It should not be unduly affected by the extreme values.

5.3 Various measures of central tendency



5.3.1 Arithmetic Mean

(a) To find A.M. for Raw data

For a raw data, the arithmetic mean of a series of numbers is sum of all observations divided by the number of observations in the series. Thus if $x_1, x_2, \ldots, x_n$ represent the values of n observations, then arithmetic mean (A.M.) for n observations is: (direct method)

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

There are two methods for computing the A.M :

(i) Direct method (ii) Short cut method.

NOTE

Normal health parameters such as blood pressure, pulse rate, blood cell count, BMI, blood sugar level etc., are calculated averages of people in a particular region and always vary among individuals.

Example 5.1

The following data represent the number of books issued in a school library on selected from 7 different days 7, 9, 12, 15, 5, 4, 11 find the mean number of books.

Solution:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Ch5_11th Stats.indd 112 07/12/2021 12:03:39

$$\bar{x} = \frac{7+9+12+15+5+4+11}{7}$$

$$= \frac{63}{7} = 9$$

Hence the mean of the number of books is 9

(ii) Short-cut Method to find A.M.

Under this method an assumed mean or an arbitrary value (denoted by A) is used as the basis of calculation of deviations (d_i) from individual values. That is if $d_i = x_i - A$ then

$$\bar{x} = A + \frac{\sum_{i=1}^{n} d_i}{n}$$

Example 5.2

A student's marks in 5 subjects are 75, 68, 80, 92, 56. Find the average of his marks.

Solution:

Let us take the assumed mean, A = 68

x_i	$d_i = x_i - 68$
75	7
68	0
80	12
56	-12
92	24
Total	**31**

$$\bar{x} = A + \frac{\sum_{i=1}^{n} d_i}{n}$$

$$= 68 + \frac{31}{5}$$

$$= 68+6.2 = 74.2$$

The arithmetic mean of average marks is 74.2

(b) To find A.M. for Discrete Grouped data

If $x_1, x_2, ..., x_n$ are discrete values with the corresponding frequencies $f_1, f_2, \ldots, f_n$. Then the mean for discrete grouped data is defined as (direct method)

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

Ch5_11th Stats.indd 113 07/12/2021 12:03:40

In the short cut method the formula is modified as

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \quad \text{where} \quad d_i = x_i - A$$

Example 5.3

A proof reads through 73 pages manuscript The number of mistakes found on each of the pages are summarized in the table below Determine the mean number of mistakes found per page

No of mistakes	1	2	3	4	5	6	7
No of pages	5	9	12	17	14	10	6

Solution:

(i) Direct Method

x_i	f_i	$f_i x_i$
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
Total	N=73	299

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

$$= \frac{299}{73}$$

$$= 4.09$$

The mean number of mistakes is 4.09

(ii) Short-cut Method

x_i	f_i	$d_i = x_i - A$	$f_i d_i$
1	5	-3	-15
2	9	-2	-18

Ch5_11th Stats.indd 114 07/12/2021 12:03:41

3	12	-1	-12
4	17	0	0
5	14	1	14
6	10	2	20
7	6	3	18
	$\Sigma f_i=73$		$\Sigma f_i d_i=7$

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N}$$

$$= 4 + \frac{7}{73}$$

$$= 4.09$$

The mean number of mistakes = 4.09

(c) Mean for Continuous Grouped data:

For the computation of A.M for the continuous grouped data, we can use direct method or short cut method.

(i) Direct Method:

The formula is

$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$, x_i is the midpoint of the class interval

(ii) Short cut method

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times C$$

$$d = \frac{x_i - A}{c}$$

where A - any arbitrary value

c - width of the class interval

x_i is the midpoint of the class interval.

Ch5_11th Stats.indd 115 07/12/2021 12:03:42

Example 5.4

The following the distribution of persons according to different income groups

Income (in ₹1000)	0 – 8	8 – 16	16 – 24	24 – 32	32 – 40	40 – 48
No of persons	8	7	16	24	15	7

Find the average income of the persons.

Solution :

Direct Method:

Class	f_i	x_i	$f_i x_i$
0-8	8	4	32
8 – 16	7	12	84
16-24	16	20	320
24-32	24	28	672
32-40	15	36	540
40-48	7	44	308
Total	N =77		1956

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

$$= \frac{1956}{77}$$

$$=25.40$$

Short cut method:

Class	f_i	x_i	$d_i=(x_i-A)/c$	$f_i d_i$
0 – 8	8	4	-3	-24
8 – 16	7	12	-2	-14
16 – 24	16	20	-1	-16
24 – 32	24	28 A	0	0
32 – 40	15	36	1	15
40 – 48	7	44	2	14
Total	**N= 77**			**-25**

Ch5_11th Stats.indd 116 07/12/2021 12:03:42

$$\overline{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \times C$$

$$= 28 + \frac{-25}{77} \times 8 \quad = 25.40$$

Merits

- It is easy to compute and has a unique value.
- It is based on all the observations.
- It is well defined.
- It is least affected by sampling fluctuations.
- It can be used for further statistical analysis.

Limitations

- The mean is unduly affected by the extreme items (outliers).
- It cannot be determined for the qualitative data such as beauty, honesty etc.
- It cannot be located by observations on the graphic method.

When to use?

Arithmetic mean is a best representative of the data if the data set is homogeneous. On the other hand if the data set is heterogeneous the result may be misleading and may not represent the data.

Weighted Arithmetic Mean

The arithmetic mean, as discussed earlier, gives equal importance (or weights) to each observation in the data set. However, there are situations in which values of individual observations in the data set are not of equal importance. Under these circumstances, we may attach, a weight, as an indicator of their importance to each observation value.

Definition

Let $x_1, x_2, ..., x_n$ be the set of n values having weights $w_1, w_2, ..., w_n$ respectively, then the weighted mean is,

$$\overline{x_w} = \frac{w_1 x_1 + w_2 x_2 + ... + w_n x_n}{w_1 + w_2 + ... + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Ch5_11th Stats.indd 117 07/12/2021 12:03:44

Uses of weighted arithmetic mean

Weighted arithmetic mean is used in:

- The construction of index numbers.
- Comparison of results of two or more groups where number of items in the groups differs.
- Computation of standardized death and birth rates.

Example 5.5

The weights assigned to different components in an examination or Component Weightage Marks scored

Component	Weightage	Marks scored
Theory	4	60
Practical	3	80
Assignment	1	90
Project	2	75
	10	

Calculate the weighted average score of the student who scored marks as given in the table

Solution:

Component	Marks scored (x_i)	Weightage (w_i)	$w_i x_i$
Theory	60	4	240
Practical	80	3	240
Assignment	90	1	90
Project	75	2	150
Total		10	720

Weighted average, $\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$

$= 720/10$

$= 72$

NOTE

If the weights are not assigned, then the mean is

$= (60 + 80 + 90 + 75)/4$

$= 76.25$

Ch5_11th Stats.indd 118 07/12/2021 12:03:44

Combined Mean:

Let $\overline{x}_1$ and $\overline{x}_2$ are the arithmetic mean of two groups (having the same unit of measurement of a variable), based on n1 and n2 observations respectively. Then the combined mean can be calculated using

$$\text{Combined Mean} = \bar{x}_{12} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

Remark : The above result can be extended to any number of groups.

Example 5.6

A class consists of 4 boys and 3 girls. The average marks obtained by the boys and girls are 20 and 30 respectively. Find the class average.

Solution:

$n_1 = 4,\ \overline{x}_1 = 20,\ n_2 = 3,\ \overline{x}_2 = 30$

$$\text{Combined Mean} = \bar{x}_{12} = \frac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$$

$$= \left[\frac{4 \times 20 + 3 \times 30}{4 + 3}\right]$$

$$= \left[\frac{80 + 90}{7} = \frac{170}{7}\right] = 24.3$$

5.3.2 Geometric Mean(GM)

(a) G.M. For Ungrouped data

The Geometric Mean (G.M.) of a set of n observations is the nth root of their product. If $x_1, x_2, \ldots, x_n$ are n observations then

$$\text{G. M.} = \sqrt[n]{x_1, x_2 \ldots x_n} = (x_1 . x_2 \ldots x_n)^{\frac{1}{n}}$$

Taking the nth root of a number is difficult. Thus, the computation is done as under

$$\log \text{G.M.} = \log (x_1 . x_2 \ldots x_n)$$

$$= (\log x_1 + \log x_2 + \ldots + \log x_n)$$

$$= \frac{\sum_{i=1}^{n} \log x_i}{n}$$

$$\text{G.M.} = \text{Antilog} \frac{\sum_{i=1}^{n} \log x_i}{n}$$

Ch5_11th Stats.indd 119 07/12/2021 12:03:46

Example 5.7

Calculate the geometric mean of the annual percentage growth rate of profits in business corporate from the year 2000 to 2005 is given below

50, 72, 54, 82, 93

Solution:

x_i	50	72	54	82	93	**Total**
log x_i	1.6990	1.8573	1.7324	1.9138	1.9685	**9.1710**

$$\text{G.M.} = \text{Antilog}\ \frac{\sum_{i=l}^{n} \log x_i}{n}$$

$$= \text{Antilog}\ \frac{9.1710}{5}$$

$$= \text{Antilog}\ 1.8342$$

$$\text{G. M.} = 68.26$$

Geometrical mean of annual percentage growth rate of profits is 68.26

Example 5.8

The population in a city increased at the rate of 15% and 25% for two successive years. In the next year it decreased at the rate of 5%. Find the average rate of growt

Solution:

Let us assume that the population is 100

Percentage rise in population	**Population at the end of year** x_i	**log**x_i
15	115	2.0607
25	125	2.0969
5	95	1.9777
		6.1353

$$\text{G.M} = \text{Antilog}\ \frac{\sum_{n}^{n} \log x_i}{n}$$

Ch5_11th Stats.indd 120 07/12/2021 12:03:47

$$= \text{Antilog } \frac{(6.1353)}{3}$$

$$= \text{Antilog } (2.0451)$$

$$= 110.9$$

(b) G.M. For Discrete grouped data

If $x_1, x_2, \ldots\ldots x_n$ are discrete values of the variate x with corresponding frequencies $f_1, f_2, \ldots f_n$. Then geometric mean is defined as

$$\text{G. M. } = \text{Antilog } \frac{\sum_{i=1}^{n} f_i \log x_i}{N} \text{ with usual notations}$$

Example 5.9

Find the G.M for the following data, which gives the defective screws obtained in a factory.

Diameter (cm)	5	15	25	35
Number of defective screws	5	8	3	4

Solution:

x_i	f_i	$\log x_i$	$f_i \log x_i$
5	5	0.6990	3.4950
15	8	1.1761	9.4088
25	3	1.3979	4.1937
35	4	1.5441	6.1764
	N=20		23.2739

$$\text{G.M} = \text{Antilog}$$

$$= \text{Antilog } \frac{\sum_{i=1}^{n} f_i \log x_i}{N}$$

$$= \text{Antilog } \frac{23.2739}{20}$$

$$= \text{Antilog } 1.1637$$

$$\text{G.M} = 14.58$$

Ch5_11th Stats.indd 121 07/12/2021 12:03:48

(c) G.M. for Continuous grouped data

Let x_i be the mid point of the class interval

$$\text{G. M.} = \text{Antilog}\left[\frac{\sum_{i=l}^{n} f_i \log x_i}{N}\right]$$

Example 5.10

The following is the distribution of marks obtained by 109 students in a subject in an institution. Find the Geometric mean.

Marks	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40
No. of Students	6	10	18	30	15	12	10	6	2

Solution:

Marks	**Mid point (x_i)**	f_i	**log x_i**	$f_i \log x_i$
4-8	6	6	0.7782	4.6692
8-12	10	10	1.0000	10.0000
12-16	14	18	1.1461	20.6298
16-20	18	30	1.2553	37.6590
20-24	22	15	1.3424	20.1360
24-28	26	12	1.4150	16.980
28-32	30	10	1.4771	14.7710
32-36	34	6	1.5315	9.1890
36-40	38	2	1.5798	3.1596
Total		**N =109**		**137.1936**

$$\text{G.M.} = \text{Antilog}\left[\frac{\sum_{i=l}^{n} f_i \log x_i}{N}\right]$$

$$= \text{Antilog}\left[\frac{137.1936}{109}\right] = \text{Antilog}\,[1.2587]$$

$$\text{G. M.} = 18.14$$

Geometric mean marks of 109 students in a subject is 18.14

Ch5_11th Stats.indd 122 07/12/2021 12:03:49

Merits of Geometric Mean:

- It is based on all the observations
- It is rigidly defined
- It is capable of further algebraic treatment
- It is less affected by the extreme values
- It is suitable for averaging ratios, percentages and rates.

Limitations of Geometric Mean:

- It is difficult to understand
- The geometric mean cannot be computed if any item in the series is negative or zero.
- The GM may not be the actual value of the series
- It brings out the property of the ratio of the change and not the absolute difference of change as the case in arithmetic mean.

5.3.3 Harmonic Mean (H.M.)

Harmonic Mean is defined as the reciprocal of the arithmetic mean of reciprocals of the observations.

(a) H.M. for Ungrouped data

Let $x_1, x_2, ..., x_n$ be the n observations then the harmonic mean is defined as

$$\text{H. M.} = \frac{n}{\sum_{i=l}^{n}\left(\frac{1}{x_i}\right)}$$

Example 5.11

A man travels from Jaipur to Agra by a car and takes 4 hours to cover the whole distance. In the first hour he travels at a speed of 50 km/hr, in the second hour his speed is 65 km/hr, in third hour his speed is 80 km/hr and in the fourth hour he travels at the speed of 55 km/hr. Find the average speed of the motorist.

Solution:

x	50	65	80	55	**Total**
$1/x$	0.0200	0.0154	0.0125	0.0182	**0.0661**

Ch5_11th Stats.indd 123 07/12/2021 12:03:50

$$\text{H. M.} = \frac{n}{\sum\left(\frac{1}{x_i}\right)}$$

$$= \frac{4}{0.0661} = 60.5 \text{ km/hr}$$

Average speed of the motorist is 60.5km/hr

(b) H.M. for Discrete Grouped data:

For a frequency distribution

$$\text{H. M.} = \frac{N}{\sum_{i=l}^{n} f_i\left(\frac{1}{x_i}\right)}$$

Example 5.12

The following data is obtained from the survey. Compute H.M

Speed of the car	130	135	140	145	150
No of cars	3	4	8	9	2

Solution:

x_i	f_i	$\frac{f_i}{x_i}$
130	3	0.0296
135	4	0.0091
140	8	0.0571
145	9	0.0621
150	2	0.0133
Total	N = 26	0.1852

$$\text{H. M.} = \frac{N}{\sum_{i=l}^{n} f_i\left(\frac{1}{x_i}\right)}$$

$$= \frac{26}{0.1852}$$

$$\text{H.M} = 140.39$$

Ch5_11th Stats.indd 124 07/12/2021 12:03:51

(c) H.M. for Continuous data:

The Harmonic mean H.M. = $\dfrac{N}{\sum_{i=l}^{n} f_i\left(\dfrac{1}{x_i}\right)}$

Where x_i is the mid-point of the class interval

Example 5.13

Find the harmonic mean of the following distribution of data

Dividend yield (percent)	2 – 6	6 – 10	10 – 14
No. of companies	10	12	18

Solution:

Class Intervals	**Mid-value (x_i)**	**No. of companies (f_i)**	**Reciprocal ($1/x_i$)**	f_i **($1/x_i$)**
2 – 6	4	10	¼	2.5
6 – 10	8	12	1/8	1.5
10 – 14	12	18	1/12	1.5
Total		N = 40		5.5

The harmonic mean is H.M. = $\dfrac{N}{\sum_{i=l}^{n} f_i\left(\dfrac{1}{x_i}\right)} = \dfrac{40}{5.5} = 7.27$

Merits of H.M:

- It is rigidly defined
- It is based on all the observations of the series
- It is suitable in case of series having wide dispersion
- It is suitable for further mathematical treatment
- It gives less weight to large items and more weight to small items

Limitations of H.M:

- It is difficult to calculate and is not understandable
- All the values must be available for computation
- It is not popular due to its complex calculation.
- It is usually a value which does not exist in series

Ch5_11th Stats.indd 125 07/12/2021 12:03:51

When to use?

Harmonic mean is used to calculate the average value when the values are expressed as value/unit. Since the speed is expressed as km/hour, harmonic mean is used for the calculation of average speed.

Relationship among the averages:

In any distribution when the original items are different the A.M., G.M. and H.M would also differ and will be in the following order:

$$\text{A.M.} \geq \text{G.M} \geq \text{H.M}$$

5.3.4 Median

Median is the value of the variable which divides the whole set of data into two equal parts. It is the value such that in a set of observations, 50% observations are above and 50% observations are below it. Hence the median is a positional average.

(a) Median for Ungrouped or Raw data:

In this case, the data is arranged in either ascending or descending order of magnitude.

(i) If the number of observations n is an odd number, then the median is represented by the numerical value of x, corresponds to the positioning point of $\frac{n+1}{2}$ in ordered observations. That is,

$$\text{Median} = \text{value of } \left(\frac{n+1}{2}\right)^{th} \text{ observation in the data array}$$

(ii) If the number of observations n is an even number, then the median is defined as the arithmetic mean of the middle values in the array That is,

$$\text{Median} = \frac{\text{value of } \left(\frac{n}{2}\right)^{th} \text{ observation} + \text{value of } \left(\frac{n}{2}+1\right)^{th} \text{ observation}}{2}$$

Example 5.14

The number of rooms in the seven five stars hotel in Chennai city is 71, 30, 61, 59, 31, 40 and 29. Find the median number of rooms

Solution:

Arrange the data in ascending order 29, 30, 31, 40, 59, 61, 71

$n = 7$ (odd)

Ch5_11th Stats.indd 126 07/12/2021 12:03:52

$$\text{Median} = \frac{7+1}{2} = 4^{\text{th}} \text{ positional value}$$

$$\text{Median} = 40 \text{ rooms}$$

Example 5.15

The export of agricultural product in million dollars from a country during eight quarters in 1974 and 1975 was recorded as 29.7, 16.6, 2.3, 14.1, 36.6, 18.7, 3.5, 21.3

Find the median of the given set of values

Solution:

We arrange the data in descending order

36.6, 29.7, 21.3, 18.7, 16.6, 14.1, 3.5, 2.3

$$n = 8 \text{ (even)}$$

$$\text{Median} = \frac{4^{\text{th}} \text{item} + 5^{\text{th}} \text{item}}{2}$$

$$= \frac{18.7 + 16.6}{2}$$

$$= 17.65 \text{ million dollars}$$

Cumulative Frequency

In a grouped distribution, values are associated with frequencies. The cumulative frequencies are calculated to know the total number of items above or below a certain limit.This is obtained by adding the frequencies successively up to the required level. This cumulative frequencies are useful to calculate median, quartiles, deciles and percentiles.

(b) Median for Discrete grouped data

We can find median using following steps

(i) Calculate the cumulative frequencies

(ii) Find $\frac{N+1}{2}$, Where $N = \sum f$ = total frequencies

(iii) Identify the cumulative frequency just greater than $\frac{N+1}{2}$

(iv) The value of x corresponding to that cumulative frequency $\frac{N+1}{2}$ is the median.

Ch5_11th Stats.indd 127 07/12/2021 12:03:53

Example 5.16

The following data are the weights of students in a class. Find the median weights of the students

Weight(kg)	10	20	30	40	50	60	70
Number of Students	4	7	12	15	13	5	4

Solution:

Weight (kg) x	Frequency f	Cumulative Frequency $c.f$
10	4	4
20	7	11
30	12	23
40	15	38
50	13	51
60	5	56
70	4	60
Total	N = 60	

Here, $N = \sum f = 60$

$$\frac{N+1}{2} = 30.5$$

The cumulative frequency greater than 30.5 is 38.The value of x corresponding to 38 is 40. The median weight of the students is 40 kgs

(c) Median for Continuous grouped data

In this case, the data is given in the form of a frequency table with class-interval etc., The following formula is used to calculate the median.

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where

l = Lower limit of the median class

N = Total Numbers of frequencies

f = Frequency of the median class

NOTE

If the class intervals are given in inclusive type, convert them into exclusive type and call it as true class interval and consider the lower limit in it.

Ch5_11th Stats.indd 128 07/12/2021 12:03:54

m = Cumulative frequency of the class preceding the median class

c = the class interval of the median class.

From the formula, it is clear that one has to find the median class first. Median class is, that class which correspond to the cumulative frequency just greater than $\frac{N}{2}$.

Example 5.17

The following data attained from a garden records of certain period Calculate the median weight of the apple

Weight in grams	410 – 420	420 – 430	430 – 440	440 – 450	450 – 460	460 – 470	470 – 480
Number of apples	14	20	42	54	45	18	7

Solution:

Weight in grams	**Number of apples**	**Cumulative Frequency**
410 – 420	14	14
420 – 430	20	34
430 – 440	42	76
440 – 450	54	130
450 – 460	45	175
460 – 470	18	193
470 – 480	7	200
Total	**N = 200**	

$$\frac{N}{2} = \frac{200}{2} = 100.$$

Median class is 440 – 450

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

$l = 440$, $\frac{N}{2} = 100$, $m = 76$, $f = 54$, $c = 10$

$$\text{Median} = 440 + \frac{100 - 76}{54} \times 10$$

$$= 440 + \frac{24}{54} \times 10 = 440 + 4.44 = 444.44$$

The median weight of the apple is 444.44 grams

Ch5_11th Stats.indd 129 07/12/2021 12:03:55

Example 5.18

The following table shows age distribution of persons in a particular region:

Age (years)	No. of persons (in thousands)
Below 10	2
Below20	5
Below30	9
Below 40	12
Below 50	14
Below 60	15
Below 70	15.5
Below 80	15.6

Find the median age.

Solution:

We are given upper limit and less than cumulative frequencies. First find the class-intervals and the frequencies. Since the values are increasing by 10, hence the width of the class interval is equal to 10.

Age groups	No. of persons (in thousands) f	cf
0 – 10	2	2
10 – 20	3	5
20 – 30	4	9
30 – 40	3	12
40 – 50	2	14
50 – 60	1	15
60 – 70	0.5	15.5
70 – 80	0.1	15.6
Total	**N = 15.6**	

$$\left(\frac{N}{2}\right) = \frac{15.6}{2} = 7.8$$

Median lies in the 20 – 30 age group

Ch5_11th Stats.indd 130 07/12/2021 12:03:55

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

$$= 20 + \frac{7.8 - 5}{4} \times 10$$

$$\text{Median} = 27 \text{ years}$$

Example 5.19

The following is the marks obtained by 140 students in a college. Find the median marks

Marks	Number of students
10-19	7
20-29	15
30-39	18
40-49	25
50-59	30
60-69	20
70-79	16
80-89	7
90-99	2

NOTE

In this problem the class intervals are given in inclusive type, convert them into exclusive type and call it as true class interval.

Solution:

Class boundaries	*f*	*Cf*
9.5 -19.5	7	7
19.5-29.5	15	22
29.5- 39.5	18	40
39.5-49.5	25	65
49.5-59.5	30	95
59.5-69.5	20	115
69.5-79.5	16	131
79.5-89.5	7	138
89.5-99.5	2	140
Total	**N =140**	

Ch5_11th Stats.indd 131 07/12/2021 12:03:55

$$\text{Median} = l + \left(\frac{\frac{N}{2} - m}{f} \right) \times c$$

$$\frac{N}{2} = \frac{140}{2} = 70$$

Here $l = 49.5,\ f = 30,\ m = 65,\ c = 10$

$$\text{Median} = 49.5 + \left(\frac{70 - 65}{30} \right) \times 10$$

$$= 49.5 + 1.67$$

$$= 51.17$$

Graphical method for Location of median

Median can be located with the help of the cumulative frequency curve or 'ogive'. The procedure for locating median in a grouped data is as follows:

Step 1 : The class intervals, are represented on the horizontal axis (x-axis)

Step 2 : The cumulative frequency corresponding to different classes is calculated. These cumulative frequencies are plotted on the vertical axis (y-axis) against the upper limit of the respective class interval

Step 3 : The curve obtained by joining the points by means of freehand is called the *'less than ogive'*.

Step 4 : A horizontal straight line is drawn from the value $\frac{N}{2}$ or $\frac{N+1}{2}$ on the y-axis parallel to x- axis to meet the ogive. (depending on N is odd or even)

Step 5 : From the point of intersection, draw a line, perpendicular to the horizontal axis which meet the x axis at m say.

Step 6 : The value m at x axis gives the value of the median.

Remarks:

(i) Similarly 'more than' ogives, can be drawn by plotting more than cumulative frequencies against lower limit of the class. A horizontal straight line is drawn from the value $\frac{N}{2}$ or $\frac{N+1}{2}$ on the y-axis parallel to x-axis to meet the ogive. A line is drawn perpendicular to x-axis meets the point at m, say, the X coordinate of m gives the value of the median.

(depending on N is odd or even)

Ch5_11th Stats.indd 132 07/12/2021 12:03:56

(ii) When the two ogive curves are drawn on the same graph, a line is drawn perpendicular to x-axis from the point of intersection, meets the point at m, say. The x coordinate m gives the value of the median.

Example 5.20

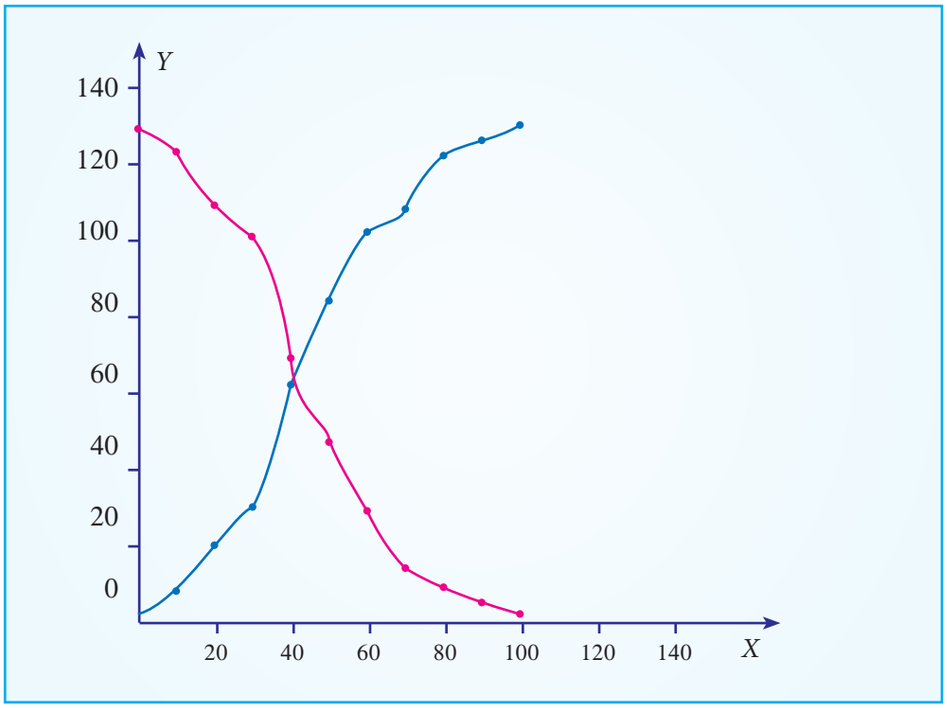
Draw ogive curves for the following frequency distribution and determine the median.

Age groups	No. of people
0 – 10	6
10 – 20	12
20 – 30	10
30 – 40	32
40 – 50	22
50 – 60	18
60 – 70	15
70 – 80	5
80 – 90	4
90 – 100	3

Solution:

Class boundary	Cumulative Frequency	
	Less than	More than
0	0	127
10	6	121
20	18	109
30	28	99
40	60	67
50	82	45
60	100	27
70	115	12
80	120	7
90	124	3
100	127	0

Ch5_11th Stats.indd 133 07/12/2021 12:03:56



The median value from the graph is 42

Merits

- It is easy to compute. It can be calculated by mere inspection and by the graphical method
- It is not affected by extreme values.
- It can be easily located even if the class intervals in the series are unequal

Limitations

- It is not amenable to further algebraic treatment
- It is a positional average and is based on the middle item
- It does not take into account the actual values of the items in the series

5.3.5 Mode

According to Croxton and Cowden, 'The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated'.



In a busy road, where we take a survey on the vehicle - traffic on the road at a place at a

Ch5_11th Stats.indd 134 07/12/2021 12:03:56

particular period of time, we observe the number of two wheelers is more than cars, buses and other vehicles. Because of the higher frequency, we say that the modal value of this survey is 'two wheelers'

Mode is defined as the value which occurs most frequently in a data set. The mode obtained may be two or more in frequency distribution.

Computation of mode:

(a) For Ungrouped or Raw Data:

The mode is defined as the value which occurs frequently in a data set

Example 5.21

The following are the marks scored by 20 students in the class. Find the mode
90, 70, 50, 30, 40, 86, 65, 73, 68, 90, 90, 10, 73, 25, 35, 88, 67, 80, 74, 46

Solution:

Since the marks 90 occurs the maximum number of times, three times compared with the other numbers, mode is 90.

Example 5.22

A doctor who checked 9 patients' sugar level is given below. Find the mode value of the sugar levels. 80, 112, 110, 115, 124, 130, 100, 90, 150, 180

Solution:

Since each values occurs only once, there is no mode.

Example 5.23

Compute mode value for the following observations.

2, 7, 10, 12, 10, 19, 2, 11, 3, 12

Solution:

Here, the observations 10 and 12 occurs twice in the data set, the modes are 10 and 12.

For discrete frequency distribution, mode is the value of the variable corresponding to the maximum frequency.



NOTE

It is clear that mode may not exist or mode may not be unique.

Ch5_11th Stats.indd 135 07/12/2021 12:03:56

Example 5.24

Calculate the mode from the following data

Days of Confinement	6	7	8	9	10
Number of patients	4	6	7	5	3

Solution:

Here, 7 is the maximum frequency, hence the value of x corresponding to 7 is 8. Therefore 8 is the mode.

(b) Mode for Continuous data:

The mode or modal value of the distribution is that value of the variate for which the frequency is maximum. It is the value around which the items or observations tend to be most heavily concentrated. The mode is computed by the formula.

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

Modal class is the class which has maximum frequency.

f_1 = frequency of the modal class

f_0 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class

c = width of the class limits

Remarks

(i) If $(2f_1\text{-}f_0\text{-}f_2)$ comes out to be zero, then mode is obtained by the following formula taking absolute differences $M_0 = l + \left(\frac{(f_1 - f_0)}{|f_1 - f_0| + |f_1 - f_2|} \times C \right)$

(ii) If mode lies in the first class interval, then f_0 is taken as zero.

(iii) The computation of mode poses problem when the modal value lies in the open-ended class.

Example 5.25

The following data relates to the daily income of families in an urban area. Find the modal income of the families.

Ch5_11th Stats.indd 136 07/12/2021 12:03:57

Income (₹)	0-100	100-200	200-300	300-400	400-500	500-600	600-700
No.of persons	5	7	12	18	16	10	5

Solution:

Income (₹)	No.of persons (f)	
0-100	5	
100-200	7	
200-300	12	f_0
300-400	18	f_1
400-500	16	f_2
500-600	10	
600-700	5	

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$$

The highest frequency is 18, the modal class is 300-400

Here, $l = 300$, $f_0 = 12$, $f_1 = 18$, $f_2 = 16$,

$$\text{Mode} = 300 + \frac{18 - 12}{2 \times 18 - 12 - 16} \times 100$$

$$= 300 + \frac{6}{36 - 28} \times 100$$

$$= 300 + \frac{6}{8} \times 100$$

$$= 300 + \frac{600}{8} = 300 + 75 = 375$$

The modal income of the families is 375.

Determination of Modal class:

For a frequency distribution modal class corresponds to the class with maximum frequency. But in any one of the following cases that is not easily possible.

(i) If the maximum frequency is repeated.

(ii) If the maximum frequency occurs in the beginning or at the end of the distribution

(iii) If there are irregularities in the distribution, the modal class is determined by the method of grouping.

Ch5_11th Stats.indd 137 07/12/2021 12:03:58

Steps for preparing Analysis table:

We prepare a grouping table with 6 columns

(i) In column I, we write down the given frequencies.

(ii) Column II is obtained by combining the frequencies two by two.

(iii) Leave the Ist frequency and combine the remaining frequencies two by two and write in column III

(iv) Column IV is obtained by combining the frequencies three by three.

(v) Leave the Ist frequency and combine the remaining frequencies three by three and write in column V

(vi) Leave the Ist and 2nd frequencies and combine the remaining frequencies three by three and write in column VI

NOTE

In discrete or continuous series, an error of judgement is possible by finding y inspection. In these cases, where the difference between the maximum frequency and the frequency preceeding or succeeding is very small and the items are heavily concentrated on either side, under such circumstance the value of mode is determined by preparing grouping table and through analysis table.

Mark the highest frequency in each column. Then form an analysis table to find the modal class. After finding the modal class use the formula to calculate the modal value.

Example 5.26

Calculate mode for the following frequency distribution:

Size	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	9	12	15	16	17	15	10	13

Ch5_11th Stats.indd 138 07/12/2021 12:03:58

Solution:

class	f	2	3	4	5	6
0-5	9	21		36		
5-10	12		27			
10-15	15	31			43	
15-20	16					48
20-25	17	32	33	48		
25-30	15					
30-35	10	23	25		42	38
35-40	13					

Analysis Table:

Columns	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
1					1			
2					1	1		
3				1	1			
4				1	1	1		
5		1	1	1				
6			1	1	1			
Total		1	2	4	5	2		

The maximum occurred corresponding to 20-25, and hence it is the modal class.

$$\text{Mode} = l + \frac{f_1 f_0}{2f_1 - f_0 - f_2} \times C$$

Here, $l = 20$, $f_0 = 16$, $f_1 = 17$, $f_2 = 15$

Ch5_11th Stats.indd 139 07/12/2021 12:03:59

$$= 20 + \frac{17-16}{2\times 17-16-15}\times C$$

$$= 20 + \frac{1}{34-31}\times 5$$

$$= 20 + \frac{5}{3} = 20 + 1.67 = 21.67$$

Mode = 21.67

(d) Graphical Location of Mode

The following are the steps to locate mode by graph

(i) Draw a histogram of the given distribution.

(ii) Join the rectangle corner of the highest rectangle (modal class rectangle) by a straight line to the top right corner of the preceding rectangle. Similarly the top left corner of the highest rectangle is joined to the top left corner of the rectangle on the right.

(iii) From the point of intersection of these two diagonal lines, draw a perpendicular line to the x –axis which meets at M.

(iv) The value of x coordinate of M is the mode.

Example 5.27

Locate the modal value graphically for the following frequency distribution

Class Interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	5	8	12	7	5	3

Solution:

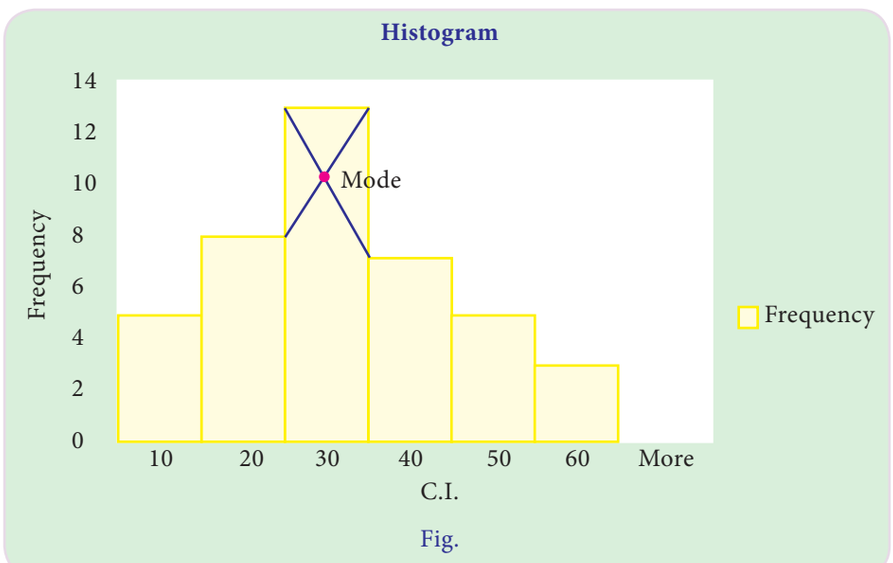


Fig.

Ch5_11th Stats.indd 140 07/12/2021 12:04:00

Merits of Mode:

- It is comparatively easy to understand.
- It can be found graphically.
- It is easy to locate in some cases by inspection.
- It is not affected by extreme values.
- It is the simplest descriptive measure of average.

Demerits of Mode:

- It is not suitable for further mathematical treatment.
- It is an unstable measure as it is affected more by sampling fluctuations.
- Mode for the series with unequal class intervals cannot be calculated.
- In a bimodal distribution, there are two modal classes and it is difficult to determine the values of the mode.

5.4 Empirical Relationship among mean, median and mode

A frequency distribution in which the values of arithmetic mean, median and mode coincide is known of symmetrical distribution, when the values of mean, median and mode are not equal the distribution is known as asymmetrical or skewed. In moderately skewed asymmetrical distributions a very important relationship exists among arithmetic mean, median and mode.

Karl Pearson has expressed this relationship as follows

Mode = 3 Median – 2 Arithmetic Mean

Example 5.28

In a moderately asymmetrical frequency distribution, the values of median and arithmetic mean are 72 and 78 respectively; estimate the value of the mode.

Solution:

The value of the mode is estimated by applying the following formula:

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean} = 3\,(72) - 2\,(78)$$

$$= 216 - 156 = 60$$

$$\text{Mode} = 60$$

Ch5_11th Stats.indd 141 07/12/2021 12:04:00

Example 5.29

In a moderately asymmetrical frequency distribution, the values of mean and mode are 52.3 and 60.3 respectively, Find the median value.

Solution:

The value of the median is estimated by applying the formula:

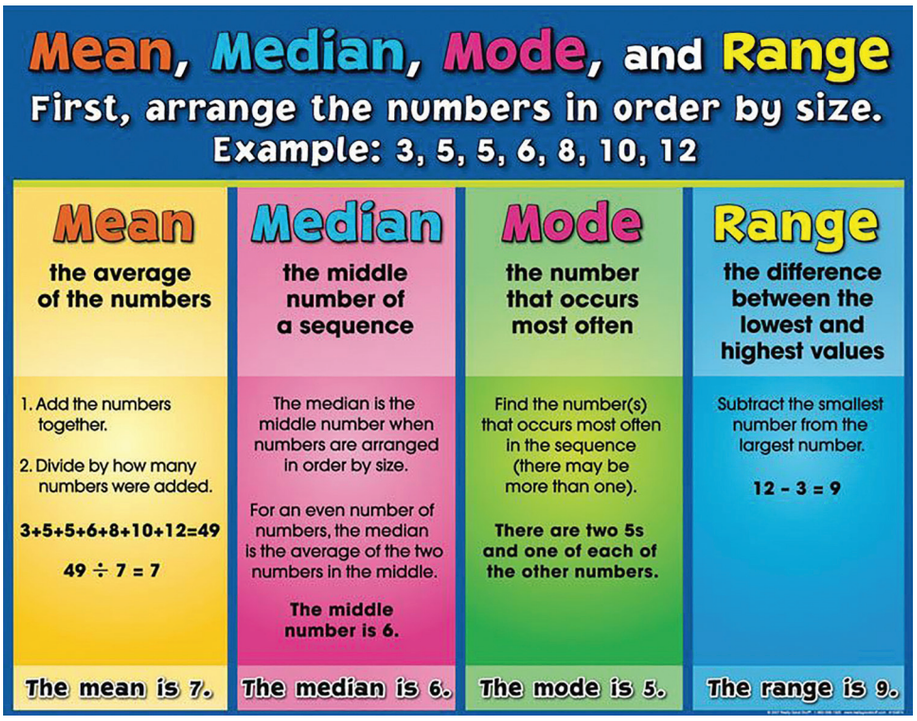
$$\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$$

$$60.3 = 3\,\text{Median} - 2 \times 52.3$$

$$3\,\text{Median} = 60.3 + 2 \times 52.3$$

$$60.3 + 104.6 = 164.9$$

$$\text{Median} = \frac{164.9}{3} = 54.966 = 54.97$$



5.5 Partition Measures

5.5.1 Quartiles

There are three quartiles denoted by Q_1, Q_2 and Q_3 divides the frequency distribution in to four equal parts

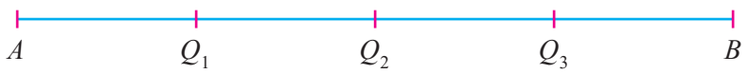


Fig.

Ch5_11th Stats.indd 142 07/12/2021 12:04:00

That is 25 percent of data will lie below Q_1, 50 percent of data below Q_2 and 75 percent below Q_3. Here Q_2 is called the Median. Quartiles are obtained in almost the same way as median

Quartiles for Raw or Ungrouped data:

If the data set consist of n items and arranged in ascending order then

$Q_1 = \left(\frac{n+1}{4}\right)^{th}$ item, $Q_2 = \left(\frac{n+1}{2}\right)^{th}$ item and $Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$ item

Example 5.30

Compute Q_1 and Q_3 for the data relating to the marks of 8 students in an examination given below 25, 48, 32, 52, 21, 64, 29, 57

Solution:

$n = 8$

Arrange the values in ascending order

21, 25, 29, 32, 48, 52, 57, 64 we have

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= \left(\frac{8+1}{4}\right)^{th} \text{ item}$$

$$= 2.25^{\text{th}} \text{ item}$$

$$= 2^{\text{nd}} \text{ item} + \left(\frac{1}{4}\right)\ (3^{\text{rd}} \text{ item} - 2^{\text{nd}} \text{ item})$$

$$= 25 + 0.25\ (29 - 25)$$

$$= 25 + 1.0$$

$$Q_1 = 26$$

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= 3 \times (2.25)^{th} \text{ item}$$

$$= 6.75^{\text{th}} \text{ item}$$

$$= 6^{\text{th}} \text{ item} + 0.75\ (7^{\text{th}} \text{ item} - 6^{\text{th}} \text{ item})$$

Ch5_11th Stats.indd 143 07/12/2021 12:04:02

$= 52 + (0.75)\ (57 - 52)$

$= 52 + 3.75$

$Q_3 = 55.75$

Quartiles for Discrete Series (grouped data)

Step 1 : Find cumulative frequencies

Step 2 : Find $\left(\frac{N+1}{4}\right)$

Step 3 : See in the cumulative frequencies, the value just greater than $\left(\frac{N+1}{4}\right)$, the corresponding value of x is Q_1

Step 4 : Find $3\left(\frac{N+1}{4}\right)$

Step 5 : See in the cumulative frequencies, the value just greater than $3\left(\frac{N+1}{4}\right)$ then the corresponding value of x is $Q3$.

Example 5.31

Compute Q_1 and Q_3 for the data relating to age in years of 543 members in a village

Age in years	20	30	40	50	60	70	80
No. of members	3	61	132	153	140	51	3

Solution:

x	f	cf
20	3	3
30	61	64
40	132	196
50	153	349
60	140	489
70	51	540
80	3	543

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ item

$= \left(\frac{543+1}{4}\right)^{th}$ item

$= 136^{th}$ item

Ch5_11th Stats.indd 144 07/12/2021 12:04:04

Q_1 = 40 years

$Q_3 = 3\left(\frac{N+1}{4}\right)^{th}$ item

$= 3\left(\frac{543+1}{4}\right)^{th}$ item

$= 3 \times 136^{th}$ item

$= 408^{\text{th}}$ item

Q_3 = 60 years

Quartiles for Continuous series (grouped data)

Step 1 : Find Cumulative frequencies

Step 2 : Find $\left(\frac{N}{4}\right)$

Step 3: Q_1 class is the class interval corresponding to the value of the cumulative frequency just greater than $\left(\frac{N}{4}\right)$

Step 4 : Q3 class is the class interval corresponding to the value of the cumulative frequency just greater than 3 $\left(\frac{N}{4}\right)$

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1 \quad \text{and} \quad Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times C_3$$

where N=∑f = total of all frequency values

l_1 = lower limit of the first quartile class

f_1 = frequency of the first quartile class

c_1 = width of the first quartile class

m_1 = c.f. preceding the first quartile class

l_3 = lower limit of the 3rd quartile class

f_3 = frequency of the 3rd quartile class

m_3 = c.f. preceding the 3rd quartile class

c_3 = width of the third quartile class

Ch5_11th Stats.indd 145 07/12/2021 12:04:06

Example 5.32

Calculate the quartiles Q_1 and Q_3 for wages of the labours given below

Wages (Rs.)	30-32	32-34	34-36	36-38	38-40	40-42	42-44
Labourers	12	18	16	14	12	8	6

Solution:

x	*f*	*cf*
30 – 32	12	12
32 – 34	18	30
34 – 36	16	46
36 – 38	14	60
38 – 40	12	72
40 – 42	8	80
42 – 44	6	86
	86	

$$\frac{N}{4} = \frac{86}{4} = 21.5$$

lies in the group 32 – 34

$$Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1$$

$$= 32 + \frac{21.5 - 12}{18} \times 2$$

$$= 32 + \frac{19}{18} = 32 + 1.06$$

$$= \text{Rs. } 33.06$$

$$\frac{3N}{4} = \frac{3 \times 86}{4} = 64.5$$

$\therefore Q_3$ lies in the group 38 – 40

$$Q_3 = l_3 + \frac{3\left(\frac{N}{4}\right) - m_3}{f_3} \times C_3$$

$$= 38 + \frac{64.5 - 60}{12} \times 2$$

$$= 38 + \frac{4.5}{12} \times 2$$

$$= 38 + 0.75 = \text{Rs. } 38.75$$

Ch5_11th Stats.indd 146 07/12/2021 12:04:08

5.5.2 Deciles

Deciles are similar to quartiles. Quartiles divides ungrouped data into four quarters and Deciles divide data into 10 equal parts .

Example 5.33

Find the D_6 for the following data

11, 25, 20, 15, 24, 28, 19, 21

Solution:

Arrange in an ascending order

11,15,19,20,21,24,25,28

$$D_6 = \left(\frac{6(n+1)}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{6(8+1)}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{6(9)}{10}\right)^{th} \text{ item}$$

$$= [5.4]^{th} \text{ item}$$

$$= 5^{th} \text{ item} + (0.4)\ (6^{th} \text{ item} - 5^{th} \text{ item})$$

$$D_6 = 21+(0.4)(24-21)$$

$$= 21+(0.4)(3)$$

$$= 21+1.2$$

$$= 22.2$$

Example 5.34

Calculate D_5 for the frequency distribution of monthly income of workers in a factory

Income (in thousands)	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 28	28 – 32
No of persons	10	12	8	7	5	8	4	6

Ch5_11th Stats.indd 147 07/12/2021 12:04:09

Solution:

Class	f	cf
0-4	10	10
4-8	12	22
8-12	8	30
12-16	7	37
16-20	5	42
20-24	8	50
24-28	4	54
28-32	6	60
	N=60	

$$D_5 = \left(\frac{5N}{10}\right)^{th} \text{ item}$$

$$= \left(\frac{5(60)}{10}\right)^{th} \text{ item}$$

$$= 30^{th} \text{ item}$$

This item in the interval 8–12

$l = 8,\ m = 22,\ f = 8,\ c = 4,\ N = 60$

$$D_5 = l + \left(\frac{\frac{5N}{10} - m}{f}\right) \times c$$

$$= 8 + \left(\frac{30 - 22}{8}\right) \times 4$$

$$= 8 + \frac{8}{8} \times 4$$

$$D_5 = 12$$

5.5.3 Percentiles

The percentile values divide the frequency distribution into 100 parts each containing 1 percent of the cases. It is clear from the definition of quartiles, deciles and percentiles

Ch5_11th Stats.indd 148 07/12/2021 12:04:10

Relationship

$P_{25} = Q_1$

P_{50} = Median = Q_2

P_{75} = 3rd quartile = Q_3

Example 5.35

The following is the monthly income (in 1000) of 8 persons working in a factory. Find P_{30} income value

10,14, 36, 25, 15, 21, 29, 17

Solutions:

Arrange the data in an ascending order.

$n = 8$

10,14,15,17,21,25,29,36

$$P_{30} = \left(\frac{30(n+1)}{100}\right)^{th} \text{ item}$$

$$= \left(\frac{30 \times 9}{100}\right)^{th} \text{ item}$$

$= 2.7^{th}$ item

$= 2^{nd}$ item + 0.7 (3^{rd} items – 2^{nd} items)

= 14+0.7(15-14)

= 14+0.7

P_{30} = 14.7

Example 5.36

Calculate P_{61} for the following data relating to the height of the plants in a garden

Heights (in cm)	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
No of plants	18	20	36	40	26	16

Ch5_11th Stats.indd 149 07/12/2021 12:04:10

Solution:

Class	*f*	*cf*
0 – 5	18	18
5 – 10	20	38
10 – 15	36	74
15 – 20	40	114
20 – 25	26	140
25 – 30	16	156
	N=156	

$$\frac{61N}{100} = \left(\frac{61 \times 156}{100}\right)$$

$$= 95.16 \text{ item}$$

This item is the interval 15-20. Thus

$l = 15, m = 74, f = 40, c = 5$

$$P_{61} = l + \left(\frac{\frac{61N}{100} - m}{f}\right) \times c$$

$$= 15 + \left(\frac{95.16 - 74}{40}\right) \times 5$$

$$= 15 + \frac{21.16}{40} \times 5$$

$$= 17.645$$

$P_{61} = 17.645$

Points to Remember

- A central tendency is a single figure that represents the whole mass of data
- Arithmetic mean or mean is the number which is obtained by adding the values of all the items of a series and dividing the total by the number of items.
- When all items of a series are given equal importance than it is called simple arithmetical mean and when different items of a series are given different weights according with their relative importance is known as weighted arithmetic mean.

Ch5_11th Stats.indd 150 07/12/2021 12:04:11

- Median is the middle value of the series when arranged in ascending order
- When a series is divided into more than two parts, the dividing values are called partition values.
- Mode is the value which occurs most frequently in the series,that is modal value has the highest frequency in the series.

EXERCISE 5

B4Z8P

I. Choose the best answer:

1. Which of the following is a measure of central value?

   (a) Median (b) Deciles

   (c) Quartiles (d) Percentiles

2. Geometric Mean is better than other means when

   (a) the data are positive as well as negative

   (b) the data are in ratios or percentages

   (c) the data are binary

   (d) the data are on interval scale

3. When all the observations are same, then the relation between A.M., G.M. and H.M. is:

   (a) A.M. = G.M. = H.M. (b) A.M. < H.M. < G.M.

   (c) A.M. < G.M. < H.M. (d) A.M. > G.M. > H.M.

4. The median of the variate values 11, 7, 6, 9, 12, 15, 19 is

   (a)9 (b)12 (c)15 (d)11

5. The middle values of an ordered series is called

   (a) 50^{th} percentile (b) 2^{nd} quartile

   (c) 5^{th} decile (d) all the above

6. Mode is that value in a frequency distribution which possesses

   (a) minimum frequency (b) maximum frequency

   (c) frequency one (d) none of the above

Ch5_11th Stats.indd 151 07/12/2021 12:04:11

7. For decile, the total number of partition values are:

(a) 5 (b) 8 (c) 9 (d) 10

8. The mean of the squares of first eleven natural numbers is:

(a) 46 (b) 23 (c) 48 (d) 42

9. Histogram is useful to determine graphically the value of

(a) mean (b) median (c) mode (d) all the above

10. What percentage of values lies between 5^{th} and 25^{th} percentile?

(a)15% (b) 30% (c) 75% (d) none of the above

II. Fill in the blanks

11. In an open end distribution _____ cannot be determined.

12. The sum of the deviations from mean is _____

13. The distribution having two modes is called _____

14. Second quartile and _____ deciles are equal.

15. Median is a more suited average for grouped data with _____ classes.

III. Very Short Answer Questions:

16. What is meant by measure of central tendency?

17. What are the desirable characteristics of a good measure of central tendency?

18. What are the merits and demerits of the arithmetic mean?

19. Express weighted arithmetic mean in brief.

20. Define Median. Discuss its advantages and disadvantages.

IV. Short Answer Questions :

21. The monthly income of ten families of a certain locality is given in rupees as below.

Family	A	B	C	D	E	F	G	H	I	J
Income (₹)	85	70	10	75	500	8	42	250	40	36

Calculate the arithmetic mean.

Ch5_11th Stats.indd 152 07/12/2021 12:04:11

22. The mean of 100 items are found to be 30. If at the time of calculation two items are wrongly taken as 32 and 12 instead of 23 and 11. Find the correct mean.

23. A cyclist covers his first three kms at an average speed of 8 kmph. Another two kms at 3 kmph and the last two kms at 2 kmph. Find the average speed for the entire journey.

24. The mean marks of 100 students were found to 40. Later it was discovered that a score of 53 was misread as 83. Find the corrected mean corresponding to the corrected score.

25. In a moderately asymmetrical distribution the values of mode and mean are 32.1 and 35.4 respectively. Find the median value.

26. Calculate D_9 from the following frequency distribution

x	58	59	60	61	62	63	64	65	66
f	2	3	6	15	10	5	4	3	2

27. Calculated P_{40} The following is the distribution of weights of patients in an hospital

Weight (in kg)	40	50	60	70	80	90	100
No of patients	15	26	12	10	8	9	5

V. Calculate the following:

28. Find the mean and median:

Wages (₹)	60 – 70	50 – 60	40 – 50	30 – 40	20 – 30
No. of labourers	5	10	20	5	3

29. The following data relates to the marks obtain by students in a school find the median

Marks	>10	>20	>30	>40	>50	>60	>70	>80	>90
No. of Students	70	62	50	38	30	24	17	9	4

30. The number of telephone calls received in 245 successive one minute intervals at an exchange are shown in the following frequency distribution:

Ch5_11th Stats.indd 153 07/12/2021 12:04:12

No. of calls	0	1	2	3	4	5	6	7
Frequency	14	21	25	43	51	40	39	12

Evaluate the mean, median and mode.

31. Calculate the geometric and the harmonic mean of the following series of monthly expenditure of a batch of students.

125, 130, 75, 10, 45, 0.5, 0.40, 500, 150, 5

32. Find out the mode of the following series:

Wages (₹)	Below 25	25 – 50	50 – 75	75 – 100	100 – 125	Above 125
No. of persons	10	30	40	25	20	15

33. Find the median, lower quartile, 7^{th} decile and 85^{th} percentile of the frequency distribution given below:

Marks in Statistics	Below 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	Above 70
No. of Students	8	12	20	32	30	28	12	4

34. The following data gives the distribution of heights of a group of 60 college students:

Height (in cms)	Number of students
145.0 – 149. 9	2
150.0 – 154.9	5
155.0 – 159.9	9
160.0 – 164.9	15
165.0 – 169.9	16
170.0 – 174.9	7
175.0 – 179.9	5
180.0 – 184.9	1

Draw the histogram for this distribution and find the modal height. Check this result by using the algebraic formula.

Ch5_11th Stats.indd 154 07/12/2021 12:04:12

35. Find the median and the quartiles for the following datas:

Size	Frequency	Size	Frequency
4	40	12	50
5	48	13	52
6	52	14	41
7	56	15	57
8	60	16	63
9	63	17	52
10	57	18	48
11	55	19	43

36. In a class of 50 students, 10 have failed and their average of marks is 2.5. The total marks secured by the entire class were 281. Find the average marks of those who have passed.

Answers

I.1. (a) 2. (b) 3. (a) 4. (d) 5. (d) 6. (b) 7. (c) 8. (a) 9. (c) 10. (d)

II. 11. Mean 12. Zero 13. Bimodal 14. 5th

15. Open end IV. 21. Mean = 111.60 22. Mean = 29.90

23. average speed = 3.4 mph 24. corrected mean = 39.7

25. Median = 34.3 26. $D_9 = 65$ 27. $P_{40} = 50$

V. 28. Mean = 47.09, Median = 46.75 29. Median = 43.75

30. Mean = 3.76, Median = 4, Mode = 4

31. G.M. = 57.73 and H.M. = 2.058 32. Mode = Rs. 60

33. Median = 40.33, First quartile = 28.25, 7th decile = 50.07, 85th percentile = 57.9

35. Median = 11, First quartile = 8, Third quartile = 15

36. Average marks of those who passed = 6.4

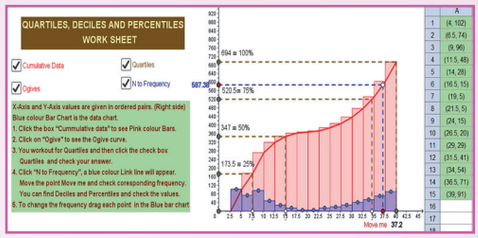
Ch5_11th Stats.indd 155 07/12/2021 12:04:12



ICT CORNER

QUARTILES, DECILES, PERCENTILES

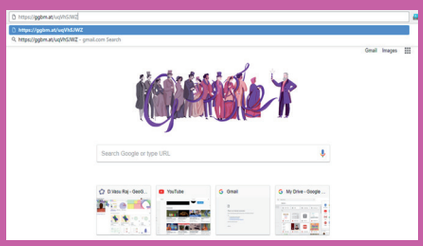
This activity helps to understand about quartiles, deciles and percentiles diagrammatically



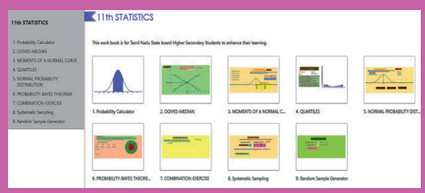
Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for Statistics are given, Open the worksheet named "**Quartiles**"
- Quartiles page will open. In that blue colour histogram is seen. If you click on "**Cumulative data**" pink colour chart for "**Cumulative frequency**" will appear. If you click on "**Ogive**" red colour ogive curve will appear. Now you calculate the quartiles, deciles and percentiles using the formula given in the text book.
- Now If press "**Quartiles**" check box to see the quartiles and check your answer. Now press "**N to frequency**" check box and move "**Blue dot**" on the x-axis to check your deciles and percentiles. If necessary You can create your own problem by typing new frequencies

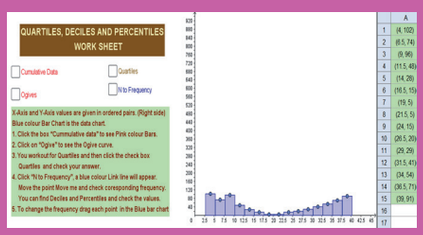
Step-1



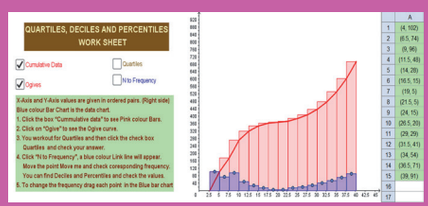
Step-2



Step-3



Step-4



Pictures are indicatives only*

URL:

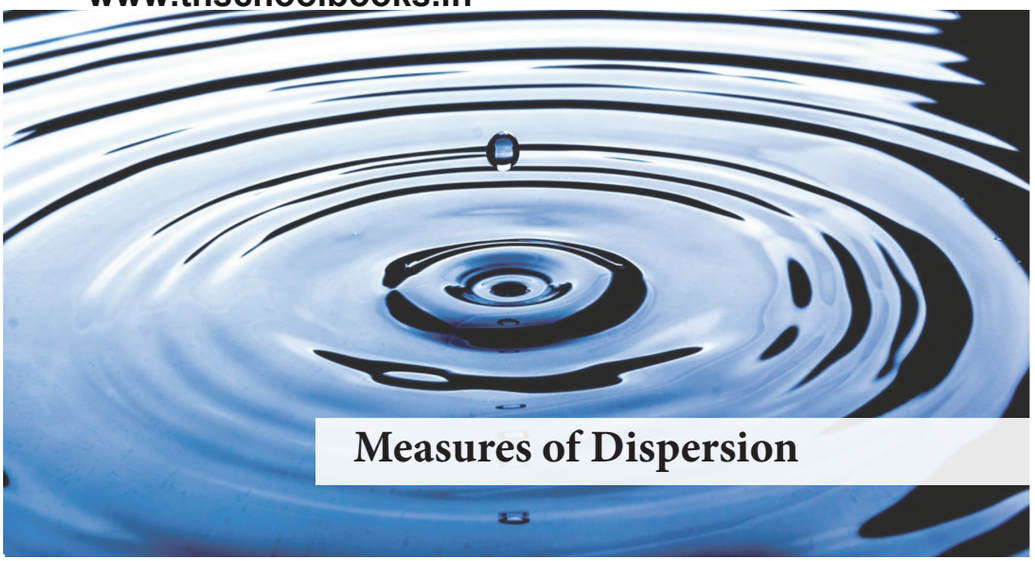
https://ggbm.at/uqVhSJWZ

JV29L

Ch5_11th Stats.indd 156 07/12/2021 12:04:13

Chapter

6



Measures of Dispersion



Karl Pearson
(27 March 1857 – 27 April 1936)

Karl Pearson was an English mathematician and biostatistician. He has been credited with establishing the discipline of mathematical statistics. He founded the world's first university Statistics department at University College London in 1911, and contributed significantly to the field of biometrics, meteorology, theories of social Darwinism and eugenics. In fact, Pearson devoted much time during 1893 to 1904 in developing statistical techniques for biometry. These techniques, which are widely used today for statistical analysis.

'The never was in the world two opinions alike, no more than two hairs or two grains; the most universal quality is diversity'.
- Michel de Montaigne

Learning Objectives

- Provides the importance of the concept of variability (dispersion)
- Measures the spread or dispersion and Identifiers the causes of dispersion
- Describes the spread - range and standard deviations
- Describes the role of Skewness and Kurtosis
- Explains about moments
- Illustrates the procedure to draw Box plot.

DI5PY

Introduction

The measures of central tendency describes the central part of values in the data set appears to concentrate around a central value called average. But these measures do

Ch6_11th Stats.indd 157 07/12/2021 12:05:58

not reveal how these values are dispersed (spread or scattered) on each side of the central value. Therefore while describing data set it is equally important to know how for the item in the data are close around or scattered away from the measures of central tendency.

Example 6.1

Look at the runs scored by the two cricket players in a test match:

Players	I Innings	II Innings	Mean
Player 1	0	100	50
Player 2	40	60	50

Comparing the averages of the two players we may come to the conclusion that they were playing alike. But player 1 scored 0 runs in I innings and 100 in II innings. Player 2 scored nearly equal runs in both the innings. Therefore it is necessary for us to understand data by measuring dispersion.

6.1 Characteristics of a good Measure of Dispersion

An ideal measure of dispersion is to satisfy the following characteristics.

(i) It should be well defined without any ambiguity.

(ii) It should be based on all observations in the data set..

(iii) It should be easy to understand and compute.

(iv) It should be capable of further mathematical treatment.

(v) It should not be affected by fluctuations of sampling.

(vi) It should not be affected by extreme observations.

6.2 Types of measures of dispersion

Range, Quartile deviation, Mean deviations, Standard deviation and their Relative measures

The measures of dispersion are classified in two categories, namely

(i) Absolute measures

(ii) Relative measures.

6.3 Absolute Measures

It involves the units of measurements of the observations. For example, (i) the dispersion of salary of employees is expressed in rupees, and (ii) the variation of time

Ch6_11th Stats.indd 158 07/12/2021 12:05:58

required for workers is expressed in hours. Such measures are not suitable for comparing the variability of the two data sets which are expressed in different units of measurements.

6.3.1 Range

Raw Data:

Range is defined as difference between the largest and smallest observations in the data set. Range(R) = Largest value in the data set (L) –Smallest value in the data set(S)

$$R = L - S$$

Grouped Data:

For grouped frequency distribution of values in the data set, the range is the difference between the upper class limit of the last class interval and the lower class limit of first class interval.

Coefficient of Range

The relative measure of range is called the coefficient of range

Co efficient of Range = (L- S) / (L+ S)

Example 6.2

The following data relates to the heights of 10 students (in cms) in a school. Calculate the range and coefficient of range.

158, 164, 168, 170, 142, 160, 154, 174, 159, 146

Solution:

L=174 S=142

Range = L – S = 174 – 142 = 32

Coefficient of range = (L - S)/(L + S)

= (174-142) / (174+142) = 32 / 316 = 0.101

Example 6.3

Calculate the range and the co-efficient of range for the marks obtained by 100 students in a school.

Ch6_11th Stats.indd 159 07/12/2021 12:05:58

Marks	60-63	63-66	66-69	69-72	72-75
No. of students	5	18	42	27	8

Solution:

L = Upper limit of highest class = 75

S = lower limit of lowest class = 60

Range = L-S = 75-60 = 15

Co - efficient of Range = (L- S) / (L+ S)

= 15/ (75 + 60) = 15 / 135=0.111

Merits:

- Range is the simplest measure of dispersion.
- It is well defined, and easy to compute.
- It is widely used in quality control, weather forecasting, stock market variations etc.

Limitations:

- The calculations of range is based on only two values – largest value and smallest value.
- It is largely influenced by two extreme values.
- It cannot be computed in the case of open-ended frequency distributions.
- It is not suitable for further mathematical treatment.

6.3.2 Inter Quartile Range and Quartile Deviation

The quartiles Q_1, Q_2 and Q_3 have been introduced and studied in Chapter 5.

Inter quartile range is defined as: Inter quartile Range (IQR) = $Q_3 - Q_1$

Quartile Deviation is defined as, half of the distance between Q_1 and Q_3.

Quartile Deviation Q.D = $\frac{Q_3 - Q_1}{2}$

It is also called as semi-inter quartile range.

Ch6_11th Stats.indd 160 07/12/2021 12:05:58

Coefficient of Quartile Deviation:

The relative measure corresponding to QD is coefficient of QD and is defined as:

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Merits:

- It is not affected by the extreme (highest and lowest) values in the data set.
- It is an appropriate measure of variation for a data set summarized in open-ended class intervals.
- It is a positional measure of variation; therefore it is useful in the cases of erratic or highly skewed distributions.

Limitations:

- The QD is based on the middle 50 per cent observed values only and is not based on all the observations in the data set, therefore it cannot be considered as a good measure of variation.
- It is not suitable for mathematical treatment.
- It is affected by sampling fluctuations.
- The QD is a positional measure and has no relationship with any average in the data set.

6.3.3 Mean Deviation

The Mean Deviation (MD) is defined as the arithmetic mean of the absolute deviations of the individual values from a measure of central tendency of the data set. It is also known as the average deviation.

The measure of central tendency is either mean or median. If the measure of central tendency is mean (or median), then we get the mean deviation about the mean (or median).

$$\text{MD (about mean)} = \frac{\sum |D|}{n} \qquad D = (x - \bar{x})$$

$$\text{MD (about median)} = \frac{\sum |{}^{D}m|}{n} = D_m = x - \text{Median}$$

The coefficient of mean deviation (CMD) is the relative measure of dispersion corresponding to mean deviation and it is given by

$$\text{Coefficient of Mean Deviation (CMD)} = \frac{MD\,(mean\ or\ median)}{mean\ or\ median}$$

Ch6_11th Stats.indd 161 07/12/2021 12:06:00

Example 6.4

The following are the weights of 10 children admitted in a hospital on a particular day.

Find the mean deviation about mean, median and their coefficients of mean deviation.

7, 4, 10, 9, 15, 12, 7, 9, 9, 18

Solution:

$n = 10$; Mean: $\bar{x} = \frac{\sum x}{n} = \frac{100}{10} = 10$

Median: The arranged data is: 4, 7, 7, 9, 9, 9, 10, 12, 15, 18

Median $= \frac{9+9}{2} = \frac{18}{2} = 9$:

Marks (x)	$\lvert D \rvert = \lvert x - \bar{x} \rvert$	$\lvert D_m \rvert = \lvert x$- Median$\rvert$
7	3	2
4	6	5
10	0	1
9	1	0
15	5	6
12	2	3
7	3	2
9	1	0
9	1	0
18	8	9
Total =100	**30**	**28**

Mean deviation from mean $= \frac{\sum |D|}{n} = \frac{30}{10} = 3$

Co-efficient mean deviation about mean $= \frac{\text{Mean deviation about mean}}{\bar{x}} = \frac{3}{10} = 0.3$

Mean deviation about median $= \frac{\sum |Dm|}{n}$

$= \frac{28}{10} = 2.8$

Co- efficient mean deviation about median $= \frac{\text{Mean deviation about median}}{\text{median}}$

$= \frac{2.8}{9} = 0.311$

Ch6_11th Stats.indd 162 07/12/2021 12:06:03

6.3.4 Standard Deviation

Consider the following data sets.

10, 7, 6, 5, 4, 3, 2
10, 10, 10, 9, 9, 9, 2 ,2
10, 4, 4, 3, 2, 2, 2

It is obvious that the range for the three sets of data is 8. But a careful look at these sets clearly shows the numbers are different and there is a necessity for a new measure to address the real variations among the numbers in the three data sets. This variation is measured by standard deviation. The idea of standard deviation was given by Karl Pearson in 1893.

Definition:

'Standard deviation is the positive square root of average of the deviations of all the observation taken from the mean.' It is denoted by a greek letter σ.

(a) Ungrouped data

$x_1, x_2, x_3 \ldots x_n$ are the ungrouped data then standard deviation is calculated by

1. **Actual mean method:** Standard deviation $\sigma = \sqrt{\frac{\sum d^2}{n}}$, $d = x - \bar{x}$

2. **Assumed mean method:** Standard deviation $\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}$, $d = x - A$

(b) Grouped Data (Discrete)

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \text{ , } d = x - A$$

Where, f = frequency of each class interval

N = total number of observation (or elements) in the population

x = mid – value of each class interval

where A is an assumed A.M.

(c) Grouped Data (continuous)

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \times C, \quad d = \frac{x - A}{c}$$

Where, f = frequency of each class interval

N = total number of observation (or elements) in the population

Ch6_11th Stats.indd 163 07/12/2021 12:06:04

c = width of class interval

x = mid-value of each class interval

where A is an assumed A.M.

Variance : Sum of the squares of the deviation from mean is known as Variance.

The square root of the variance is known as standard deviation.

NOTE

The simplified form of standard deviation formula may also be used

1. $\sigma = \frac{1}{n}\sqrt{n\sum d^2 - \left(\sum d\right)^2}$ (for raw data)
2. $\sigma = \frac{1}{N}\sqrt{N\sum fd^2 - \left(\sum fd\right)^2} \times C$ (for grouped data) where $d = (x\text{-A})/c$

Example 6.5

The following data gives the number of books taken in a school library in 7 days find the standard deviation of the books taken

7, 9, 12, 15, 5, 4, 11

Solution:

Actual mean method

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

$$\frac{7+9+.....+11}{7} = \frac{63}{7} = 9$$

x	$d = x - \bar{x}$	d^2
7	-2	4
9	0	0
12	3	9
15	6	36
5	-4	16
4	-5	25
11	2	4
		94

$$\sigma = \sqrt{\frac{\sum d^2}{n}}$$

$$= \sqrt{\frac{94}{7}}$$

$$= \sqrt{13.43}$$

$$= 3.66$$

NOTE

We can use two methods to find standard deviation

1. Direct method
2. Shortcut method

Ch6_11th Stats.indd 164 07/12/2021 12:06:05

Merits:

- The value of standard deviation is based on every observation in a set of data.
- It is less affected by fluctuations of sampling.
- It is the only measure of variation capable of algebraic treatment.

Limitations:

- Compared to other measures of dispersion, calculations of standard deviation are difficult.
- While calculating standard deviation, more weight is given to extreme values and less to those near mean.
- It cannot be calculated in open intervals.
- If two or more data set were given in different units, variation among those data set cannot be compared.

Example 6.6

Raw Data:

Weights of children admitted in a hospital is given below calculate the standard deviation of weights of children.

13, 15, 12, 19, 10.5, 11.3, 13, 15, 12, 9

Solution:

$$\text{A.M.,} \quad \bar{x} = \frac{\sum x}{n}$$

$$= \frac{13 + 15 + \ldots\ldots + 49}{10}$$

$$= \frac{129.8}{10}$$

$$= 12.98$$

Deviation from actual mean

x	$d = x\text{-}12.98$	d^2
13	0.02	0.0004
15	2.02	4.0804
12	–0.98	0.9604
19	6.02	36.2404
10.5	2.48	6.1504

Ch6_11th Stats.indd 165 07/12/2021 12:06:06

11.3	−1.68	2.8224
13	0.02	0.0004
15	2.02	4.0804
12	−0.98	0.9604
9	−3.98	15.8404
***n* =10**		**71.136**

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum d^2}{n}}$$

$$= \sqrt{\frac{71.136}{10}}$$

$$= 2.67$$

NOTE

If the mean value is not an integer, the calculation is difficult. In such a case we use the alternative formula for the calculation.

Example 6.7

Find the standard deviation of the first '*n*' natural numbers.

Solution:

The first n natural numbers are 1, 2, 3,…, n. The sum and the sum of squares of these n numbers are

$$\sum x_i = 1+2+3+.....+n = \frac{n(n+1)}{2}$$

$$\sum x^2{}_i = 1^2+2^2+3^2+.....n = \frac{n(n+1)(2n+1)}{6}$$

$$\text{Mean } \bar{x} = \frac{1}{n}\sum x_i = \frac{n(n+1)}{2n} = \frac{(n+1)}{2}$$

$$\frac{\sum x_i^2}{n} = \frac{(n+1)(2n+1)}{6}$$

$$\text{Standard deviation, } \sigma = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

$$= \sqrt{\frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}}$$

$$= \sqrt{\frac{2(n+1)(2n+1) - 3(n+1)^2}{12}}$$

$$= \sqrt{\frac{(n+1)[2(2n+1) - 3(n+1)]}{12}}$$

$$= \sqrt{\frac{(n+1)(n-1)}{12}} = \sqrt{\frac{(n^2-1)}{12}}$$

$$\sigma = \sqrt{\frac{(n^2-1)}{12}}$$

Ch6_11th Stats.indd 166 07/12/2021 12:06:09

Example 6.8

The wholesale price of a commodity for seven consecutive days in a month is as follows:

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

Calculate the variance and standard deviation.

Solution:

The computations for variance and standard deviation is cumbersome when x values are large. So, another method is used, which will reduce the calculation time. Here we take the deviations from an assumed mean or arbitrary value A such that $d = x - A$

In this question, if we take deviation from an assumed A.M. =255. The calculations then for standard deviation will be as shown in below Table;

Observations (x)	$d = x - A$	d^2
240	–15	225
260	5	25
270	15	225
245	–10	100
255 A	0	0
286	31	961
264	9	81
	35	**1617**

$$\text{Variance} = \sigma^2 = \frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2$$

$$= \frac{1617}{7} - \left(\frac{35}{7}\right)^2$$

$$= 231 - 5^2$$

$$= 231 - 25$$

$$= 206$$

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

$$\sigma = \sqrt{206} = 14.35$$

Ch6_11th Stats.indd 167 07/12/2021 12:06:10

Example 6.9

The mean and standard deviation from 18 observations are 14 and 12 respectively. If an additional observation 8 is to be included, find the corrected mean and standard deviation.

Solution:

The sum of the 18 observations is = $n \times \bar{x} = 18 \times 14 = 252$.

The sum of the squares of these 18 observations

$$\sigma^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$

$$12^2 = \frac{\sum x^2}{18} - 14^2$$

$$144+196 = \frac{\sum x^2}{18}$$

$$\frac{\sum x^2}{18} = 340$$

$$\sum x^2 = 340 \times 18 = 6120$$

When the additional observation 8 is included, then n=19 ,

$\Sigma x = 252+8 = 260$

Therefore, Corrected Mean = 260/19 = 13.68

Corrected $\sum x^2 = \sum x^2 + 8^2$

$= 6120+64$

$= 6184$

Corrected Variance $\sigma^2 = \frac{6184}{19} - 13.68^2$

$= 325.47 - 187.14$

$= 138.33;$

Corrected Standard deviation $\sigma = \sqrt{138.33}$

$\sigma = 11.76$

Example 6.10

A study of 100 engineering companies gives the following information

Ch6_11th Stats.indd 168 07/12/2021 12:06:11

Profit (₹ in Crore)	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Number of Companies	8	12	20	30	20	10

Calculate the standard deviation of the profit earned.

Solution:

A = 35 C = 10

Profit (Rs. in Crore)	Mid-value (x)	$d = \frac{x-A}{C}$	f	fd	fd^2
0 – 10	5	-3	8	-24	72
10 – 20	15	-2	12	-24	48
20 – 30	25	-1	20	-20	20
30 – 40	35	0	30	0	0
40 – 50	45	1	20	20	20
50 – 60	55	2	10	20	40
Total			**100**	**-28**	**200**

$$\text{Standard deviation } \sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

$$= \sqrt{\frac{200}{100} - \left(\frac{-28}{100}\right)^2} \times 10$$

$$= \sqrt{2 - (0.078)} \times 10$$

$$= 13.863$$

Activity

Find the standard deviation for this problem using the other two formulae.

6.4 Combined Mean and Combined Standard Deviation

Combined arithmetic mean can be computed if we know the mean and number of items in each group of the data.

$\overline{x_1}, \overline{x_2}, \sigma_1, \sigma_2$ are mean and standard deviation of two data sets having n_1 and n_2 as number of elements respectively.

$$\text{combined mean } \overline{x_{12}} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2} \quad \text{(if two data sets)}$$

$$\overline{x_{123}} = \frac{n_1\overline{x_1} + n_2\overline{x_2} + n_3\overline{x_3}}{n_1 + n_2 + n_3} \quad \text{(if three data sets)}$$

Ch6_11th Stats.indd 169 07/12/2021 12:06:14

Combined standard deviation

$$\sigma_{12} = \sqrt{\frac{n_1\left(\sigma_1^{\,2} + d_1^{\,2}\right) + n_2\left(\sigma_2^{\,2} + d_2^{\,2}\right)}{n_1 + n_2}}$$

$$d_1 = \overline{x_{12}} - \overline{x_1}$$

$$d_2 = \overline{x_{12}} - \overline{x_2}$$

Example 6.11

From the analysis of monthly wages paid to employees in two service organizations X and Y, the following results were obtained

	Organization *X*	**Organization *Y***
Number of wage-earners	550	650
Average monthly wages	5000	4500
Variance of the distribution of wages	900	1600

(i) Which organization pays a larger amount as monthly wages?

(ii) Find the combined standard deviation?

Solution:

(i) For finding out which organization X or Y pays larger amount of monthly wages, we have to compare the total wages:

Total wage bill paid monthly by X and Y is

$X: n_1 \times \bar{x}_1 = 550 \times 5000 = ₹\ 27,50,000$

$Y: n_2 \times \bar{x}_2 = 650 \times 4500 = ₹\ 29,25,000$

Organization Y pays a larger amount as monthly wages as compared to organization X.

(ii) For calculating the combined variance, we will first calculate the combined mean

$$\overline{x_{12}} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2}$$

$$= \frac{2750000 + 29250000}{550 + 650}$$

$$= \text{Rs. } 4729.166$$

Ch6_11th Stats.indd 170 07/12/2021 12:06:15

Combined standard deviation

$$d_1 = \overline{x_{12}} - \overline{x_1} = 4729.166 — 5000 = -270.834$$

$$\sigma_{12} = \sqrt{\frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2}}$$

$$= \sqrt{\frac{550(900 + 73,351.05) + 650(1600 + 52,517.05)}{550 + 650}}$$

$$= \sqrt{\frac{4,08,38,080.55 + 3,51,76,082.50}{1200}} = \sqrt{633445} = 251.68$$

6.5 Relative Measures

It is a pure number independent of the units of measurements. This measure is useful especially when the data sets are measured in different units of measurement.

For example, suppose a nutritionist would like to compare the obesity of school children in India and England. He collects data from some of the schools in these two countries. The weight is normally measured in kilograms in India and in pounds in England. It will be meaningless, if we compare the obesity of students using absolute measures. So it is sensible to compare them in relative measures.

6.5.1 Coefficient of Variation

The standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated. The standard deviation of heights of students cannot be compared with the standard deviation of weights of students, as both are expressed in different units, ie., heights in centimeter and weights in kilograms. Therefore the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure is known as the coefficient of variation.

The coefficient of variation is obtained by dividing the standard deviation by the mean and multiplying it by 100. Symbolically,

$$\text{Coefficient of Variation } (C.V) = \frac{\sigma}{\bar{x}} \times 100$$

If we want to compare the variability of two or more series, we can use $C.V$. The series or groups of data for which the $C.V$ is greater indicate that the group is more variable, less stable, less uniform, less consistent or less homogeneous. If the $C.V$ is less, it indicates that the group is less variable, more stable, more uniform, more consistent or more homogeneous.

Ch6_11th Stats.indd 171 07/12/2021 12:06:16

Merits:

- The $C.V$ is independent of the unit in which the measurement has been taken, but standard deviation depends on units of measurement. Hence one should use the coefficient of variation instead of the standard deviation.

Limitations:

- If the value of mean approaches 0, the coefficient of variation approaches infinity. So the minute changes in the mean will make major changes.

Example 6.12

If the coefficient of variation is 50 per cent and a standard deviation is 4, find the mean.

Solution:

$$\text{Coefficient of Variation} = \frac{\sigma}{\bar{x}} \times 100$$

$$50 = \frac{4}{\bar{x}} \times 100$$

$$\bar{x} = \frac{4}{50} \times 100 = 8$$

Example 6.13

The scores of two batsmen, A and B, in ten innings during a certain season, are as under:

A: Mean score = 50; Standard deviation = 5

B: Mean score = 75; Standard deviation = 25

Find which of the batsmen is more consistent in scoring.

Solution:

$$\text{Coefficient of Variation } (C.V) = \frac{\sigma}{\bar{x}} \times 100$$

$$C.V \text{ for batsman } A = \frac{5}{50} \times 100 = 10\%$$

$$C.V \text{ for batsman } B = \frac{25}{75} \times 100 = 33.33\%$$

The batsman with the smaller $C.V$ is more consistent.

Since for Cricketer A, the $C.V$ is smaller, he is more consistent than B.

Ch6_11th Stats.indd 172 07/12/2021 12:06:18

Example 6.14

The weekly sales of two products A and B were recorded as given below

Product *A*	59	75	27	63	27	28	56
Product *B*	150	200	125	310	330	250	225

Find out which of the two shows greater fluctuations in sales.

Solution:

For comparing the fluctuations in sales of two products, we will prefer to calculate coefficient of variation for both the products.

Product A: Let $A = 56$ be the assumed mean of sales for product A.

Sales(x)	**Frequency(f)**	**A= 56** $d = x - A$	fd	fd^2
27	2	-29	-58	1682
28	1	-28	-28	784
56 A	1	0	0	0
59	1	3	3	9
63	1	7	7	49
75	1	19	19	361
Total	**7**		**-57**	**2885**

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

$$= 56 - \frac{57}{7} = 47.86$$

$$\text{Variance } \sigma^2 = \frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2$$

$$= \frac{2885}{7} - \left(\frac{-57}{7}\right)^2$$

$$= 412.14 - 66.30 = 345.84$$

$$\text{Standard deviation } \sigma = \sqrt{345.84} = 18.59$$

$$\text{Coefficient of variation } (C.V) = \frac{\sigma}{\bar{x}} \times 100$$

$$= \frac{18.59}{47.86} \times 100$$

$$= 38.84\ \%$$

Ch6_11th Stats.indd 173 07/12/2021 12:06:20

Product *B*

Sales(x)	Frequency(f)	A= 225 $d = x - A$	fd^2	fd^2
125	1	-100	-100	10,000
150	1	-75	-75	5625
200	1	-25	-25	625
225	1	0	0	0
250	1	25	25	625
310	1	85	85	7225
330	1	105	105	11,025
Total	**7**		**15**	**35,125**

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

$$= 225 + \frac{15}{7}$$

$$= 225 + 2.14 = 227.14$$

$$\text{Variance } \sigma^2 = \frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2$$

$$= \frac{35125}{7} - \left(\frac{15}{7}\right)^2$$

$$= 5017.85 - 4.59$$

$$= 5013.26$$

$$\text{Standard deviation} = \sqrt{5013.26} = 70.80$$

$$\text{Coefficient of variation } (C.V)\ B = \frac{70.80}{227.14} \times 100$$

$$= 31.17\%$$

Since the coefficient of variation for product A is more than that of product B,

Therefore the fluctuation in sales of product A is higher than product B.

6.6 Moments

6.6.1 Raw moments:

Raw moments can be defined as the arithmetic mean of various powers of deviations taken from origin. The r^{th} Raw moment is denoted by μ_r', r = 1,2,3... Then the first raw moments are given by

Ch6_11th Stats.indd 174 07/12/2021 12:06:21

Raw moments	Raw data ($d=x-A$)	Discrete data ($d=x-A$)	Continuous data ($d=(x-A)/c$)
μ_1'	$\frac{\sum d}{n}$	$\frac{\sum fd}{N}$	$\frac{\sum fd}{N}\times c$
μ_2'	$\frac{\sum d^2}{n}$	$\frac{\sum fd^2}{N}$	$\frac{\sum fd^2}{N}\times c^2$
μ_3'	$\frac{\sum d^3}{n}$	$\frac{\sum fd^3}{N}$	$\frac{\sum fd^3}{N}\times c^3$
μ_4'	$\frac{\sum d^4}{n}$	$\frac{\sum fd^4}{N}$	$\frac{\sum fd^4}{N}\times c^4$

6.6.2 Central Moments:

Central moments can be defined as the arithmetic mean of various powers of deviation taken from the mean of the distribution. The r^{th} central moment is denoted by μ_r, $r = 1, 2, 3...$

Central moments	Raw data	Discrete data	Continuous data $d'=\frac{(x-\bar{x})}{c}$
μ_1	$\frac{\sum(x-\bar{x})}{n}=0$	$\frac{\sum f(x-\bar{x})}{N}=0$	$\frac{\sum fd'}{N}\times c$
μ_2	$\frac{\sum f(x-\bar{x})^2}{N}=0$	$\frac{\sum f(x-\bar{x})^2}{N}=\sigma^2$	$\frac{\sum fd'^2}{N}\times c^2$
μ_3	$\frac{\sum(x-\bar{x})^3}{n}$	$\frac{\sum f(x-\bar{x})^3}{N}$	$\frac{\sum fd'^3}{N}\times c^3$
μ_4	$\frac{\sum(x-\bar{x})^4}{n}$	$\frac{\sum f(x-\bar{x})^4}{N}$	$\frac{\sum fd'^4}{N}\times c^4$

In general, given n observation $x_1, x_2, \ldots, x_n$ the r^{th} order raw moments ($r = 0,1,2,...$) are defined as follows:

$$\mu'_r = \frac{1}{N}\sum f(x-A)^r \text{ (about } A)$$

$$\mu'_r = \frac{\sum fx^r}{N} \text{ (about origin)}$$

$$\mu_r = \frac{1}{N}\sum f(x-\bar{x})^r \text{ (about mean)}$$

NOTE

Raw moments are denoted by μ'_r and central moments are denoted by μ_r

Ch6_11th Stats.indd 175 07/12/2021 12:06:27

6.6.3 Relation between raw moments and central moments

$$\mu_1 = 0$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu_1')^4$$

Example 6.15

The first two moments of the distribution about the value 5 of the variable, are 2 and 20.find the mean and the variance.

Solution:

$\mu'_1 = 2, \mu'_2 = 20$ and $A = 5$

$$\overline{x} = \mu'_1 + A$$

$$\overline{x} = 2+5=7$$

$$\sigma^2 = \mu'_2 - (\mu'_1)^2$$

$$\sigma^2 = 20 - 2^2 = 16$$

Mean = 7 and Variance = 16

6.7 SKEWNESS AND KURTOSIS

There are two other comparable characteristics called skewness and kurtosis that help us to understand a distribution.

6.7.1 Skewness

Skewness means 'lack of symmetry'. We study skewness to have an idea about the shape of the curve drawn from the given data. When the data set is not a symmetrical distribution, it is called a skewed distribution and such a distribution could either be positively skewed or negatively skewed.



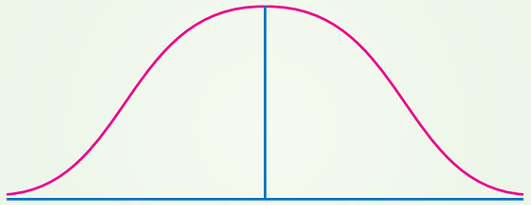
The concept of skewness will be clear from the following three diagrams showing a symmetrical distribution, a positively skewed distribution and negatively skewed distribution.

We can see the symmetricity from the following diagram.

Ch6_11th Stats.indd 176 07/12/2021 12:06:29

(a) Symmetrical Distribution

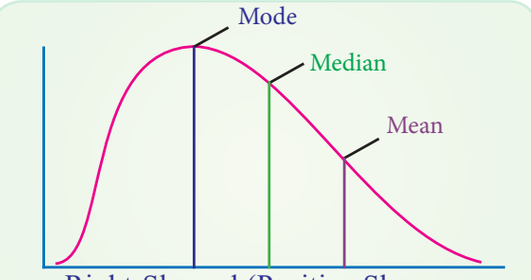
It is clear from the diagram below that in a symmetrical distribution the values of mean, median and mode coincide. The spread of the frequencies is the same on both sides of the centre point of the curve.



Mean = Median = Mode
Fig.6.1

(b) Positively Skewed Distribution

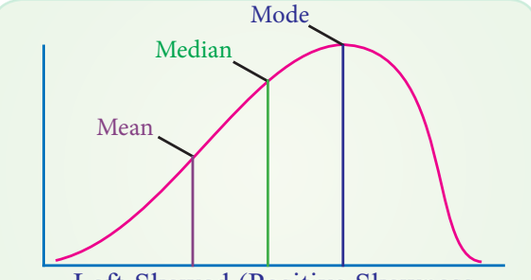
In the positively skewed distribution the value of the mean is maximum and that of mode is least – the median lies in between the two. In the positively skewed distribution the frequencies are spread out over a greater range of values on the high-value end of the curve (the right-hand side) than they are on the low – value end. For a positively skewed distribution, Mean>Median> Mode



Right-Skewed (Positive Skewness
Fig.6.2

(c) Negatively skewed distribution

In a negatively skewed distribution the value of mode is maximum and that of mean least-the median lies in between the two. In the negatively skewed distribution the position is reversed, i.e., the excess tail is on the left-hand side.



Left-Skewed (Positive Skewness
Fig.6.3

It should be noted that in moderately symmetrical distribution the interval between the mean and the median is approximately one-third of the interval between the mean and the mode. It is this relationship which provides a means of measuring the degree of skewness.

d. Some important Measures of Skewness

(i) Karl-Pearson coefficient of skewness

(ii) Bowley's coefficient of skewness

(iii) Coefficient of skewness based on moments

Ch6_11th Stats.indd 177 07/12/2021 12:06:29

(i) Karl – Person coefficient of skewness

According to Karl-Pearson the absolute measure of skewness = Mean – Mode.

Karl- Pearson coefficient of skewness = $\frac{Mean - Mode}{S.D}$

Example 6.16

From the known data, mean = 7.35, mode = 8 and Variance = 1.69 then find the Karl- Pearson coefficient of skewness.

Solution:

Karl- Pearson coefficient of skewness = $\frac{Mean - Mode}{S.D}$

Variance = 1.69

Standard deviation = $\sqrt{1.69} = 1.3$

Karl- Pearson coefficient of skewness = $\frac{7.35 - 8}{1.3}$

$= \frac{-0.65}{1.3} = -0.5$

(ii) Bowley's coefficient of skewness

In Karl Pearson method of measuring skewness the whole of the series is needed. Prof. Bowley has suggested a formula based on position of quartiles. In symmetric distribution quartiles will be equidistance from the median. $Q_2 - Q_1 = Q_3 - Q_2$, but in skewed distributions it may not happen. Hence

Bowley's coefficient of skewness (SK) = $\frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

Example 6.17

If $Q_1 = 40$, $Q_2 = 50$, $Q_3 = 60$, find Bowley's coefficient of skewness

Solution:

Bowley's coefficient of skewness

$= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

Bowley's coefficient of skewness

$= \frac{40 + 60 - 2 \times 50}{60 - 40} = \frac{0}{20} = 0$

$\therefore$ Given distribution is symmetric.

NOTE

If the difference between the mean and median or mean and mode is greater, the data is said to be more dispersed.

Ch6_11th Stats.indd 178 07/12/2021 12:06:30

(iii) Measure of skewness based on Moments

The Measure of skewness based on moments is denoted by β_1 and is given by

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Example 6.18

Find β_1 for the following data $\mu_1 = 0$, $\mu_2 = 8.76$, $\mu_3 = -2.91$

Solution:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$\beta_1 = \frac{(-2.91)^2}{(8.76)^3} = \frac{8.47}{672.24}$$

$$= 0.0126$$

6.7.2 Kurtosis

Kurtosis in Greek means 'bulginess'. In statistics kurtosis refers to the degree of flatness or peakedness in the region about the mode of a frequency curve. The degree of kurtosis of distribution is measured relative to the peakedness of normal curve. In other words, measures of kurtosis tell us the extent of which a distribution is more peaked or flat-topped than the normal curve.

The following diagram illustrates the shape of three different curves mentioned below:

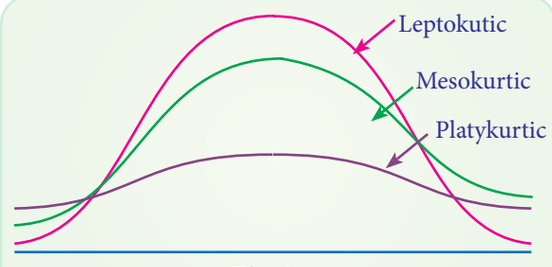


Fig.6.4

If a curve is more peaked than the normal curve, it is called 'leptokurtic'. In such a case items are more closely bunched around the mode. On the other hand if a curve is more flat-topped than the normal curve, it is called 'platykurtic'. The bell shaped normal curve itself is known as 'mesokurtic'. We can find how much the frequency curve is flatter than the normal curve using measure of kurtosis.

Measures of Kurtosis

The most important measure of kurtosis is the value of the coefficient. It is defined as: coefficient of kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$

Ch6_11th Stats.indd 179 07/12/2021 12:06:31

NOTE

The greater the value of β_2 , the more peaked the distribution.

(i) The normal curve and other curves with $\beta_2 = 3$ are called mesokurtic.

(ii) When the value of β_2 is greater than 3, the curve is more peaked than the normal curve, i.e., leptokurtic.

(iii) When the value of β_2 is less than 3 the curve is less peaked than the normal curve, i.e., platykurtic.

Example 6.19

Find the value of β_2 for the following data $\mu_1 = 0\ \mu_2 = 4\ \mu_3 = 0\ \mu_4 = 37.6$

Solution:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\beta_2 = \frac{37.6}{4^2} = \frac{37.6}{16} = 2.35 < 3$$

$\beta_2 < 3$, The curve is platykurtic.

6.8 Box plot

A box plot can be used to graphically represent the data set. These plots involve five specific values:

(i) The lowest value of the data set (i.e., minimum), (ii) Q_1 (iii) The median (iv) Q_3, (v) The highest value of the data set (i.e maximum)

These values are called a five- number summary of the data set.

A box plot is a graph of a data set obtained by drawing a horizontal line from the minimum data value to Q_1 and a horizontal line from Q_3 to the maximum data value, and drawing a box by vertical lines passing through Q_1 and Q_3, with a vertical line inside the box passing through the median or Q_2.

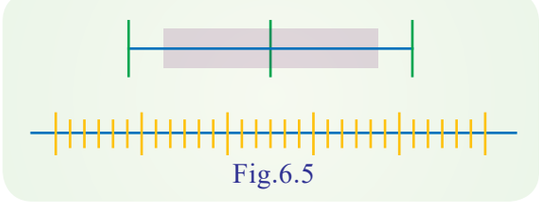


Fig.6.5

Ch6_11th Stats.indd 180 07/12/2021 12:06:32

6.8.1 Description of boxplot

(i) If the median is near the center of the box, the distribution is approximately symmetric

(ii) If the median falls to the left of the center of the box, the distribution is positively skewed.

(iii) If the median falls to the right of the center of the box, the distribution is negatively skewed.

(iv) If the lines are about the same length, the distribution is approximately symmetric

(v) If the right line is larger than the left line. the distribution is positively skewed.

(vi) If the left line is larger than the right line. the distribution is negatively skewed.

Remark:

(i) The line drawn from minimum value of the dataset to Q_1 and Q_3 to the maximum value of the data set is called whisker.

(ii) Box plot is also called Box – Whisker plot.

(iii) A box and whisker plot illustrate the spread of the distribution and also gives an idea of the shape of the distribution

Example 6.20

The following data gives the Number of students studying in XI standard in 10 different schools 89,47,164,296,30,215,138,78,48, 39 construct a boxplot for the data.

Solution:

Step 1: Arrange the data in order

30,39,47,48,78,89,138,164,215,296

Step 2: Find the median

$$\text{Median} = \frac{78+89}{2} = 83.5$$

Step 3: Find Q_1

30,39,47,48,78

$Q_1 = 47$

Ch6_11th Stats.indd 181 07/12/2021 12:06:32

Step 4: Find Q_3

89,138,164,215,296

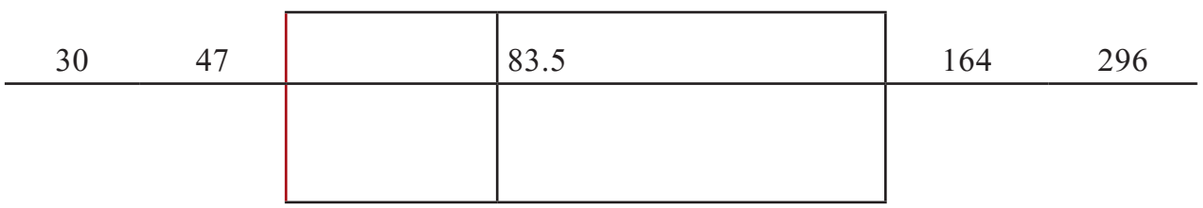
$Q_3 = 164$

Step 5: Find the minimum and maximum values.

Step 6: Locate the lowest value, Q_1, median, Q_3 and the highest value on the scale.

Step 7: Draw a box through Q_1 and Q_3

Box plot



Example 6.21

Construct a box –whisker plot for the following data

96, 151, 167, 185, 200, 220, 246, 269, 238, 252, 297, 105, 123, 178, 202

Solution:

Step 1: Arrange the data in code

96,105,123,151,167,178,185,200,202,220,238,246,252,269,297.

Step 2: Find the Median

8^{th} term Median = 200

Step 3: Find Q_1 (middle of previous terms of 200)

96,105,123,151,167,178,185

$Q_1 = 151$

Step 4 : Find Q_3 (middle of successive terms of 200)

202,220,238,246,252,269,297

$Q_3 = 246$

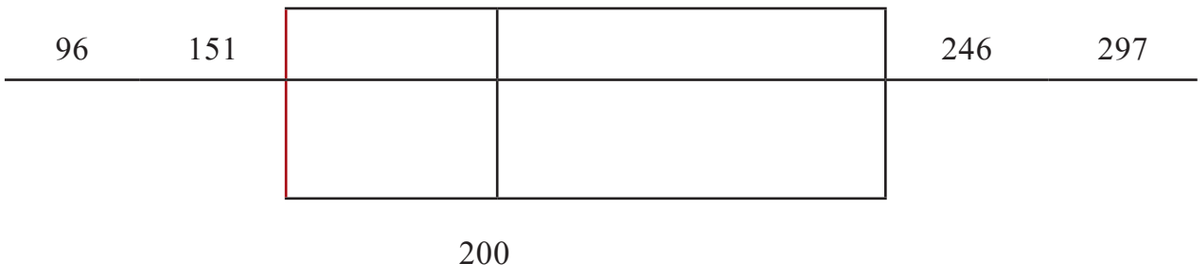
Step 5: Minimum value = 96, Maximum Value = 297

Ch6_11th Stats.indd 182 07/12/2021 12:06:32

Step 6: Draw a scale for the data on the x axis

Step 7: Locate the five numbers in the scale and draw a box around

Box plot



Points to Remember

- The range is the difference between the largest and smallest observations.
- The inter quartile range (IQR) is the difference between the upper and lower quartiles.
- The variance is the average of the squares of the values of $x-\overline{x}$
- The standard deviation (s.d) is the square root of the variance and has the same units as x
- If a population is approximately **symmetric** in a sample the mean and the median will have similar values. Typically their values will also be close to that of the mode of the population (if there is one!)
- A population that is not symmetric is said to be **skewed**. A distribution with a long 'tail' of high values is said to be **positively skewed**, in which case the mean is usually greater than the mode or the median. If it has a long tail of low values it is said to be **negatively skewed**, then the mean is likely to be the lowest of the three location measures of the distribution
- Box plots (Box-whisker diagrams): indicate the least and greatest values together with the quartiles and the median.

Ch6_11th Stats.indd 183 07/12/2021 12:06:32

EXERCISE : 6

PBS29

I. Choose the best answer:

1. When a distribution is symmetrical and has one mode, the highest point on the curve is called the

   (a) Mode (b) Median (c) Mean (d) All of these.

2. When referring to a curve tails to the left end, you would call it.

   (a) Symmetrical (b) Negatively skewed
   (c) Positively skewed (d) All of these

3. Disadvantages of using the range as a measure of dispersion include all of the following except

   (a) It is heavily influenced by extreme values

   (b) It can change drastically from one sample to the next

   (c) It is difficult to calculate

   (d) It is determined by only two points in the data set.

4. Which of the following is true?

   (a) The variance can be calculated for grouped or ungrouped data.

   (b) The standard deviation can be calculated for grouped or ungrouped data.

   (c) The standard deviation can be calculated for grouped or ungrouped data but the variance can be calculated only for ungrouped data.

   (d) (a) and (b), but not (c).

5. The squareroot of the variance of a distribution is the

   (a) Standard deviation (b) Mean
   (c) Range (d) Absolute deviation

6. The standard deviation of a set of 50 observations in 8. If each observation is multiplied by 2, then the new value of standard deviation will be:

   (a) 4 (b) 2 (c)16 (d) 8

7. In a more dispersed (spread out) set of data:

   (a) Difference between the mean and the median is greater

   (b) Value of the mode is greater

Ch6_11th Stats.indd 184 07/12/2021 12:06:32

(c) Standard deviation is greater

(d) Inter-quartile range is smaller

8. Which of the following is a relative measure of dispersion?

(a) standard deviation (b) variance
(c) coefficient of variation (d) all of the above

9. If quartile deviation is 8, then value of the standard deviation will be:

(a) 12 (b) 16 (c) 24 (d) none of the above

10. If the difference between the mean and the mode is 35 and the standard deviation is 10 then the coefficient of skewness is

(a) 2.5 (b) 1.5 (c) 3.5 (d) 6.5

II. Fill in the Blanks:

11. The difference between the values of the first and third quartiles is the _____

12. The measure of the average squared difference between the mean and each item in the population is the _____.The positive square root of this value is the _____

13. The expression of the standard deviation as a percentage of the mean is the _____

14. The number of observations lies above or below the median is called the _____

15. If $\beta_2 = 3$ the curve is_____

III Answer shortly:

16. What is dispersion? What are various measures of dispersion?

17. What is meant by relative measure of dispersion? Describe its uses.

18. Define mean deviation. How does it differ from standard deviation?

19. What is standard deviation? Explain its important properties? What is variance?

20. What are the measures of skewness?

21. Write the measures used in box plot.

Ch6_11th Stats.indd 185 07/12/2021 12:06:32

IV Answer in brief:

22. Explain dispersion and write their uses?

23. What are the requisites of a good measure of variation?

24. Explain how measures of central tendency and measures of variations are complementary to each other in the context of analysis of data.

25. Distinguish between absolute and relative measures of variation. Give a broad classification of the measures of variation.

26. Explain and illustrate how the measures of variation afford a supplement to averages in frequency distribution.

27. What you understand by 'coefficient of variation?' Discuss its importance in business problems.

28. When is the variance equal to the standard deviation? Under what circumstances can variance be less than the standard deviation?

29. A retailer uses two different formulas for predicting monthly sales. The first formula has an average of 700 records, and a standard deviation of 35 records. The second formula has an average of 300 records, and a standard deviation of 16 records. Which formula is relatively less accurate?

30. In a small business firm, two typists are employed-typist A and typist B. Typist A types out, on an average, 30 pages per day with a standard deviation of 6. Typist B, on an average, types out 45 pages with a standard deviation of 10. Which typist shows greater consistency in his output?

V Calculate the following:

31. Calculate mean deviation about mean for the following frequency distribution:

Age in Years	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45
No.of persons	7	10	16	32	24	18	10	5	1

32. The mean of two sample sizes 50 and 100 respectively are 54.1 and 50.3 and the standard deviations are 8 and 7. Find the mean and standard deviation of the sample size of 150 obtained by combining the two samples.

Ch6_11th Stats.indd 186 07/12/2021 12:06:32

33. Following data represents the life of two models of refrigerators A and B.

Life (No. of years)		0-2	2-4	4-6	6-8	8-10	10-12
Refrigerator	Model A	5	16	13	7	5	4
	Model B	2	7	12	19	9	1

Find the mean life of each model. Which model has greater uniformity? Also obtain mode for both models.

34. Calculate the quartile deviation and coefficient of quartile deviation from the following data:

Size	06	09	12	15	18
Frequency	7	12	19	10	2

35. Calculate the appropriate measure of dispersion from the following data:

Wages in ₹	Below 35	35-37	38-40	41-43	Above 43
Earners	14	60	95	24	7

36. Two brands of tyres are tested with the following results:

Life (in 1000 miles)	20-25	25-30	30-35	35-40	40-45	45-50
Brand A	8	15	12	18	13	9
Brand B	6	20	32	30	12	0

Which is more consistent ?

37. Find the S.D. for the number of days patients admitted in a hospital.

Days of confinement	5	6	7	8	9
No. of patients	18	14	9	3	1

38. Calculate the quartile deviation and its coefficient from the following data:

Class Interval	10 -15	15 - 20	20 - 25	25 - 30	30 - 35
Frequency	8	12	14	10	6

Ch6_11th Stats.indd 187 07/12/2021 12:06:32

39. Calculate the Mean Deviation from mean and its Coefficient from the following data, relating to Height (to the nearest cm) of 100 children:

Height(cms)	60	61	62	63	64	65	66	67	68
No. of Children	2	0	15	29	25	12	10	4	3

40. Find the standard deviation for the distribution given below:

x	1	2	3	4	5	6	7
Frequency	10	20	30	35	14	10	2

41. Calculate the coefficient of range separately for the two sets of data:

Set 1	10	20	9	15	10	13	28
Set 2	35	42	50	32	49	39	33

42. Blood serum cholesterol levels of 10 persons are 240, 260, 290, 245, 255, 288, 272, 263, 277 and 250. Calculate the standard deviation with the help of assumed mean.

43. Two groups of people played a game which reveals the quick operation of a particular key in a computer and the fraction of reaction times nearest to the tenth of a second is given below .

	Minimum	I Quartile	Median	III Quartile	Maximum
Group I	0.6	0.8	1.0	1.5	1.9
Group II	0.4	0.7	1.0	1.3	1.6

Draw two Box-Whisker plots and compare reaction times of the two groups.

44. Draw Box – Whisker plot for the following

(i) 3, 5, 10, 11, 12, 16, 17, 17, 19, 20, 22

(ii) –7, –5, –4, –4, –3, –3, –2, –1, 0, 1, 4, 6, 8, 9

Answers

I.1. d 2. b 3. c 4. d 5. a 6. c 7. a 8. c 9. d 10. c

II. 11. IQR 12. variance, standard deviation 13.coefficient of variation

14. 50% 15. normal (or) symmetric V.29. 2nd Formula 30. typist B

31. 7.13 32. 51.57, 7.67 33. (i) 5.12, 6.16 (ii) model B

34. 3, 0.25 35. QD=1.8, CQD=0.047 36. brand B 37. 1.03 38. 4.94,0.2265

39. 1.24, 0.0194 40. 1.4 41. 0.51,0.22 42. 16.5

Ch6_11th Stats.indd 188 07/12/2021 12:06:32

Self-Practice Problems

1. The following samples shows the weekly number of road accidents in a city during a two-year period:

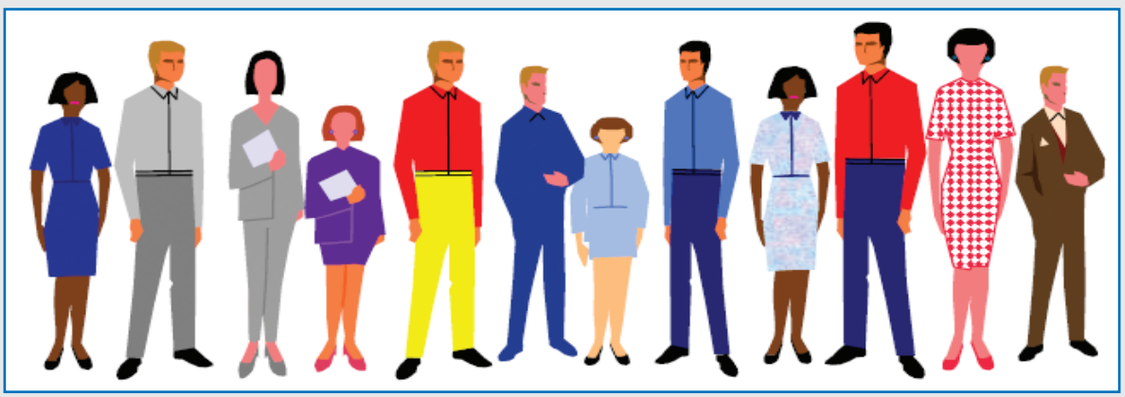
Number of Accidents	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	Total
Frequency	5	12	32	27	11	9	4	3	1	104

2. The cholera cases reported in different hospitals of a city in a rainy season are given below:

   Calculate the quartile deviation for the given distribution and comment upon the meaning of your result.

Activity

Age Group (years)	< 1	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	> 40
Frequency	15	113	122	91	110	119	132	65	46	15



Based on the figure answer the following question:

1. Whether the mean height can be taken as a representative of this group? If not what would be appropriate measure?

2. What would be the shape of the distribution of the height? In other words whether it is Symmetry?, if not what is the nature of skewness?

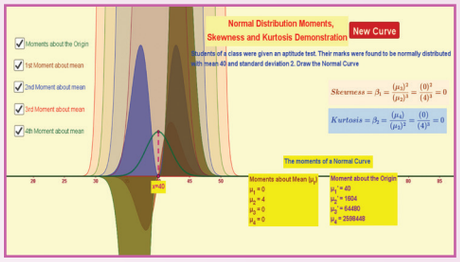
Ch6_11th Stats.indd 189 07/12/2021 12:06:33



ICT CORNER

MOMENTS OF NORMAL CURVE

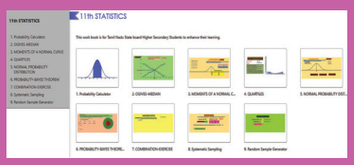
This activity helps to understand what is moments in probability using normal probability distribution



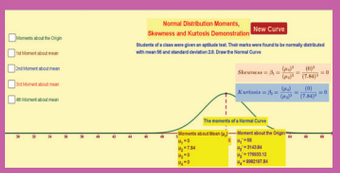
Steps:

- Open the browser and type the URL given (or) scan the QR code. GeoGebra work book called "11th Standard Statistics" will appear. In this several work sheets for statistics are given, open the worksheet named "Moments of Normal Curve"
- Normal curve will appear. You can change the normal curve by pressing on "New Curve". Observe the changes.
- You have to observe that for normal curve skewness and kurtosis is always zero. First you observe the calculation on the screen by changing the curve as in step-2. The reason you can find in next steps.
- First click on "Moments about the origin" check box. You can see the graph for moments about the origin in different colours. You get various values for moment about the origin. Observe the data at right bottom.
- If you click on "2nd moment about Mean" check box you see the curve is entirely on the top and it has some value.
- If you click on 1st, 3rd and 4th moment about mean you get top and bottom equal curves which are positive and negative. That is why these values are zero. That is why skewness and kurtosis formula lead to zero. (Observe)

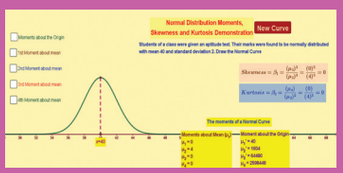
Step-1



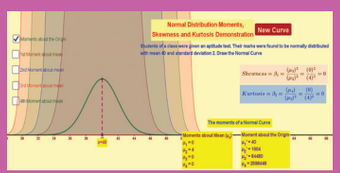
Step-2



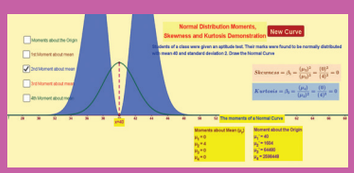
Step-3



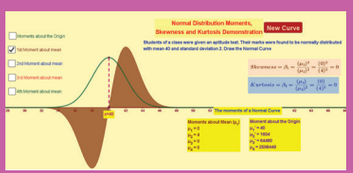
Step-4



Step-5



Step-6



Pictures are indicatives only*

URL:

https://ggbm.at/uqVhSJWZ

18H3Z

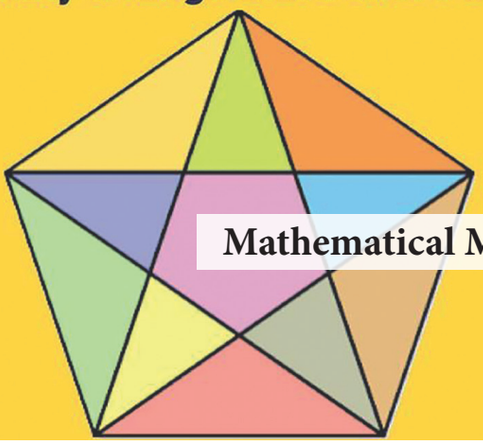
Ch6_11th Stats.indd 190 07/12/2021 12:06:35

www.tnschoolbooks.in

Chapter

7

How many triangles are there below?



Mathematical Methods



Srinivasa Ramanujan

(22 December, 1887 - 26 April, 1920)

Srinivasa Ramanujan, was an Indian Mathematician born on December 22, 1887, at Erode inTamilnadu, whose contributions to the **theory of numbers** include pioneering discoveries of the properties of the partition function. In 1911 Ramanujan published the first of his papers in the ***Journal of the Indian Mathematical Society.*** His genius slowly gained recognition, and in 1913 he began a correspondence with the British mathematician **Godfrey H. Hardy** that led to a special scholarship from the University of Madras and a grant from Trinity College, Cambridge.

Overcoming his religious objections, Ramanujan traveled to England in 1914, where Hardy tutored him and collaborated with him in some research.In England Ramanujan made further advances, especially in the **partition of numbers.** His papers were published in English and European journals, and in 1918 he was elected to **the Royal Society of London**.

In 1917 Ramanujan had contracted tuberculosis, but his condition improved sufficiently for him to return to India in 1919. He died the following year at Kumbakonam in Tamilnadu on April 26,1920, generally unknown to the world at large but recognized by mathematicians as a phenomenal genius. Ramanujan left behind three notebooks and a sheaf of pages (also called the **"lost notebook"**) containing many unpublished results that mathematicians continued to verify long after his death..

"Numbers are my friends".

- Srinivasa Ramanujan

ZS6L5

Learning Objectives

- Understands permutations and combinations
- Knows about the Binomial, Exponential and Logarithmic expansions
- Understands the concept of differentiation and applies in solving problems
- Understands the concept of integration and applies in solving problems

Ch7_11th Stats.indd 191 07/12/2021 12:08:32

Introduction

Before we proceed to further statistical concepts and calculations, we need to have a very good knowledge on some more new mathematical concepts, rules and formulae to understand in a better way the theory and problems in statistics. Hence we introduce some algebraic methods and elementary calculus in this chapter.

7.1 Fundamental Principles of counting

We use two fundamental principles of counting in solving the problems. They are addition rule on counting and multiplication rule on counting.

Fundamental Principle of addition on counting:

If an operation can be performed in m ways and if another operation can be performed in n ways and only one operation can be done at a time, then either of the two operations can be done at a time can be performed in $m + n$ ways.

Example 7.1

In a box there are 5 red balls and 6 green balls. A person wants to select either a red ball or a green ball. In how many ways can the person make this selection?.

Solution:

Selection of a red ball from 5 balls in 5 ways.

Selection of a green ball from 6 balls in 6 ways.

By the fundamental Principle of addition, selection of a red ball or a green ball in (5+6) =11 ways.

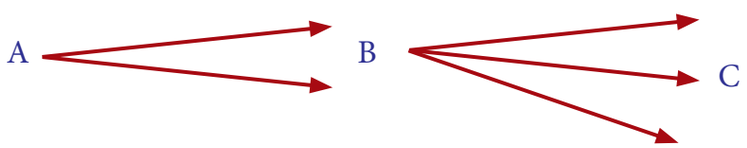
Fundamental Principle of multiplication on counting:

If an operation can be performed in m ways and if another operation in n ways independent of the first, then the number of ways of performing both the operations simultaneously in m × n ways.

Example 7.2

A person has to travel from a place A to C through B. From A to B there are two routes and from B to C three routes. In how many ways can he travel from A to C?.

Solution:



Ch7_11th Stats.indd 192 07/12/2021 12:08:32

The person can travel from A to B in 2 ways and the person can travel from B to C in 3 ways.

By the Fundamental Principle of multiplication, the person can travel from A to C simultaneously in $2 \times 3 = 6$ ways.

[Note: Observe the answer part and the diagram carefully.]

Example 7.3

A company allots a code on each different product they sell. The code is made up of one English letter and two digit numbers. How many different codes are possible?

Solution:

There are 26 English Letters (A to Z) and other two digit numbers (0 to 9) are given.

Letter	Number	Number
26 ways	10 ways	10 ways

The letter place can be filled in 26 ways with the 26 alphabets *A* to Z.

The ten's place can be filled in 10 ways with the digits 0 to 9.

The unit's place also can be filled in 10 ways with the digits 0 to 9.

So the number of product codes can be formed in $26 \times 10 \times 10$ ways $= 2600$ ways.

Example 7.4

How many four digit numbers can be formed by using the digits 2, 5, 7, 8, 9, if the repetition of the digits is not allowed?.

Solution:

Thousand	Hundred	Ten	One
5 ways	4 ways	3 ways	2 ways

The thousand's place can be filled with the 5 digits in 5 ways.

Since the repetition is not allowed, the hundred's place can be filled with the remaining 4 ways.

Similarly, for the ten's place can be filled with the remaining 3 digits in 3 ways and the unit's place can be filled with the remaining 2 digits in 2 ways.

Ch7_11th Stats.indd 193 07/12/2021 12:08:32

Therefore the number of numbers formed in $5 \times 4 \times 3 \times 2 = 120$ ways.

$\therefore$ 120 four digit numbers can be formed.

Factorial:

The consecutive product of first n natural numbers is known as factorial n and is denoted as $n!$ or $\lfloor\underline{n}$

That is $n! = n \times (n-1) \times \ldots \times 3 \times 2 \times 1$

$3! = 3 \times 2 \times 1$

$4! = 4 \times 3 \times 2 \times 1$

$5! = 5 \times 4 \times 3 \times 2 \times 1$

$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1$

Also $6! = 6 \times (5 \times 4 \times 3 \times 2 \times 1) = 6 \times (5!)$

This can be algebraically expressed as $n! = n\,(n-1)!$

Note that $1! = 1$ and $0! = 1$.

7.2 Permutations

Permutation means arrangement of things in different ways. Let us take 3 things A, B, C for an arrangement. Out of these three things two at a time, we can arrange them in the following manner.

AB	AC	BC
BA	CA	CB

Here we find 6 arrangements. In these arrangements, order of arrangement is considered. Note that the arrangement AB and the arrangement BA are different.

The number of arrangements of the above is given as the number of permutation of 3 things taken 2 at a time which gives the value 6.

This is written symbolically $3P_2 = 6$

Thus the number of arrangements that can be formed out of n things taken r at a time is known as the number of permutation of n things taken r at a time and is denoted as nP_r or $P(n, r)$

We write nP_r as $nP_r = n\,(n-1)\,(n-2)\,\ldots\,[n-(r-1)]$

Ch7_11th Stats.indd 194 07/12/2021 12:08:33

The same nP_r can be written in factorial notation as follows:

$$nP_r = \frac{n!}{(n-r)!}$$

For example to find $10\,P_3$, we write in factorial notation as

$$10\,P_3 = \frac{10!}{(10-3)!}$$

$$10\,P_3 = \frac{10!}{7!} = \frac{10 \times 9 \times 8 \times (7!)}{(7!)}$$

$$= 10 \times 9 \times 8$$

$$= 720$$

Also we get the value for $10P_3$ as follows:

$$10\,P_3 = 10 \times (10\text{-}1) \times (10\text{-}2)$$

$$= 10 \times 9 \times 8$$

$$= 720$$

[To find $10\,P_3$, Start with 10, write the product of 3 consecutive natural numbers in the descending order]

Example 7.5

In how many ways can five students stand for a photograph in a row?

Solution:

The number of ways in which 5 students can stand in a row is same as the number of arrangements of 5 different things taken all at a time.

NOTE

(1) Number of all permutations of n things taken all at a time is n!

(2) $n\,P_0 = 1$, $n\,P_1 = n$ and $n\,P_n$ =n!

This can be done in $5P_5$ ways and $5P_5 = 5! = 120$ ways.

Permutation of objects not all distinct:

The number of permutations of n objects, where p_1 objects are of one kind, p_2 are of second kind … p_k are of k^{th} kind and the rest, if any, are of different kind is given by

$$\frac{n_!}{p_1!p_2!....p_k!}$$

Example 7.6

Find the number of permutations of the letters in the word 'STATISTICS'

Ch7_11th Stats.indd 195 07/12/2021 12:08:34

Solution:

Here there are 10 objects (letters) of which there are 3S, 3 T, 2 I and 1A and 1C

Therefore the required number of arrangements is

$$= \frac{10!}{3!2!2!1!1!}$$

$$= 50400$$

Example 7.7

If $10\,P_r = 720$ find the value of r.

$$10\,P_r = 720$$

$$10\,P_r = 10 \times 9 \times 8$$

$$10\,P_r = 10\,P_3$$

$$r = 3$$

7.3 Combinations

Combination is a selection of objects without considering the order of arrangements. For example out of three things A, B, C we have to select two things at a time. This can be selected in three different ways as follows.

AB AC BC

Here the selection of object AB and BA are one and the same. The order of arrangement is not considered in combination. Hence the number of combinations from 3 different things taken 2 at a time is 3.

This is written symbolically $3C_2 = 3$.

Now we use the formula to find combination.

The number of combination of n different things, taken r at a time is given by

$$nC_r = \frac{np_r}{r!} \quad \text{or} \quad nC_r = \frac{n!}{(n-r)!r!}$$

NOTE

(i) nC_r is also denoted by $C(n, r)$ or $\binom{n}{r}$

(ii) $n\,C_0 = 1$, $nC_1 = n$, $nC_n = 1$

Example 7.8

Find $10C_3$ and $8C_4$

Solution:

$$10C_3 = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120$$

Ch7_11th Stats.indd 196 07/12/2021 12:08:35

$$8C_4 = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 70$$

[To find $10\,C_3$: In the numerator, first write the product of 3 natural numbers starting from 10 in the descending order and in the denominator write the factorial 3 and then simplify].

Compare $10C_8$ and $10C_2$

$$10\,C_8 = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3}{4 \times 3 \times 2 \times 1} = \frac{10 \times 9}{2 \times 1} = 45$$

$$10\,C_2 = \frac{10 \times 9}{2 \times 1} = 45$$

From the above we find $10\,C_8 = 10\,C_2$

Therefore this is also written as $10\,C_8 = 10\,C_{(10-8)} = 10\,C_2$

This is very useful, when the difference between n and r is very high in $n\,C_r$.

This property of combination is given as a result, $n\,C_r = n\,C_{n-r}$

To find $200\,C_{198}$, we can use the above formula as follows:

$$200\,C_{198} = 200\,C_{200-198}$$
$$= 200\,C_2$$
$$= \frac{200 \times 199}{2 \times 1}$$
$$= 19900.$$

Example 7.9

Out of 13 players, 11 Players are to be selected for a cricket team. In how many ways can this be done?

Solution:

Out of 13 Players, 11 Players are selected in $13C_{11}$ ways

i.e., $13C_{11} = 13C_2 = \frac{13 \times 12}{2 \times 1} = 78$

Example 7.10

In how many ways a committee of 5 members can be selected from 6 men and 5 women, consisting of 2 men and 3 women?

Solution:

For a committee, 2 men and 3 women members are to be selected. From 6 men, 2 men are selected in $6C_2$ ways. From 5 women, 3 women are selected in $5C_3$ ways.

Ch7_11th Stats.indd 197 07/12/2021 12:08:36

Hence a committee of 5 members (2 men and 3 women) is selected in

$6C_2 \times 5C_3$ ways

i.e., $\frac{6 \times 5}{2 \times 1} \times \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 150$ ways.

Example 7.11

How many triangles can be formed by joining the vertices of a pentagon of five sides.

Solution:

There are 5 vertices in a pentagon. One triangle is formed by selecting a group of 3 vertices from given 5 vertices. This can be done in $5C_3$ ways.

i.e., Number of triangles $= 5C_3 = \frac{5 \times 4 \times 3}{3 \times 2 \times 1} = 10$

Example 7.12

A question paper contains section A with 5 questions and section B with 7 questions. A student is required to attempt 8 questions in all, selecting at least 3 from each section. In how many ways can a student select the questions?

Solution:

Selection of 8 questions from 12 questions and at least 3 from each section is given below

Section A 5Questions	Section B 7Questions	Combinations	NO. of ways
3	5	$5C_3 \times 7C_5$	210
4	4	$5C_4 \times 7C_4$	175
5	3	$5C_5 \times 7C_3$	35
		Total	420

Therefore total number of selection is 420

Example 7.13

If $6P_r = 360$ and $6C_r = 15$ find r.

Solution:

From the formula,

$$nC_r = \frac{{}^nP_r}{r!}$$

$$6C_r = \frac{6p_r}{r!}$$

Ch7_11th Stats.indd 198 07/12/2021 12:08:37

Here, $15 = \frac{360}{r!}$

i.e., $r! = \frac{360}{15} = 24$

$r! = 4 \times 3 \times 2 \times 1$

$r! = 4!$

$\therefore \quad r = 4$

Example 7.14

If $nC_8 = nC_7$, find nC_{15}

Solution:

$nC_8 = nC_7$

$nC_{n-8} = nC_7$

$n-8 = 7$

$n = 15$

Now, $nC_{15} = 15C_{15} = 1$.

7.4 Introduction to Binomial, Exponential and Logarithmic series

7.4.1 Binomial series

A binomial is an algebraic expression of two terms. Now let us see the following binomial expansion and the number pattern we get adjacent to it.

Binomials	Expansions	Number pattern
$(x+a)^0$	1	1
$(x+a)^1$	$x+a$	1 1
$(x+a)^2$	$x^2+2xa+a^2$	1 2 1
$(x+a)^3$	$x^3+3x^2a+3xa^2+a^3$	1 3 3 1
$(x+a)^4$	$x^4+4x^3a+6x^2a^2+4xa^3+a^4$	1 4 6 4 1
$(x+a)^5$	$x^5+5x^4a+10x^3a^2+10x^2a^3+5xa^4+a^5$	1 5 10 10 5 1

From the above, we observe that the binomial coefficients form a number pattern which is in a triangular form. This pattern is known as Pascal's triangle. [In pascal's triangle, the binomial coefficients appear as each entry is the sum of the two above it.]

Ch7_11th Stats.indd 199 07/12/2021 12:08:37

Binomial theorem for a positive integral index:

For any natural number n

$$(x+a)^n = nc_0 x^n a^0 + nc_1 x^{n-1} a^1 + nc_2 x^{n-2} a^2 + \ldots + nc_r x^{n-r} a^r + \ldots + nc_n a^n$$

In the above binomial expansion, we observe,

(i) The $(r+1^{)\text{th}}$ term is denoted by $T_{r+1} = nc_r x^{n-r} a^r$

(ii) The degree of 'x' in each term decreases while that of 'a' increases such that the sum of the power in each term equal to n

(iii) nc_0, nc_1, nc_2, … nc_r, … nc_n are binomial coefficients they are also written as $c_0, c_1, c_2 \ldots, c_n$

(iv) From the relation $nc_r = nc_{n-r}$ we see that the coefficients of term equidistant from the beginning and the end are equal.

Example 7.15

Expand $(2x+y)^5$ using binomial theorem.

Solution:

$$(x+a)^n = nc_0 x^n a^0 + nc_1 x^n - 1_a + nc_2 x^{n-2} a^2 + \ldots + nc_r x^{n-r} a^r + \ldots + nc_n a^n$$

Here, $n = 5$, $X = 2x$, $a = y$

$$\begin{aligned}(2x+y)^5 &= 5C_0\ (2x)^5 (y)^0 + 5C_1\ (2x)^4 (y)^1 + 5C_2\ (2x)^3 (y)^2 + \\ &\quad 5C_3\ (2x)^2 (y)^3 + 5C_4\ (2x)^1 (y)^4 + 5C_5\ (2x)^0 (y)^5\end{aligned}$$

$$= (1)2^5x^5 + (5)\,2^4x^4\,y + (10)\,2^3x^3\,y^2 + (10)\,2^2x^2\,y^3 + (5)\,2^1x^1\,y^4 + (1)\,y^5$$

$$= 32x^5 + 80x^4y + 80x^3y^2 + 40x^2y^3 + 10xy^4 + y^5$$

$$(2x+y)^5 = 32x^5 + 80x^4y + 80x^3y^2 + 40x^2y^3 + 10xy^4 + y^5$$

Example 7.16

Find the middle terms of expansion $(3x+y)^5$

Solution:

In the expansion of $(3x+y)^5$ we have totally 6 terms. From this the middle terms are

T_3 and T_4

Ch7_11th Stats.indd 200 07/12/2021 12:08:37

To find T_3 put r = 2 in T_{r+1}

Here $n = 5,\ x = 3x,\ a = y$

$$T_{r+1} = nC_r\, x^{n-r}\, a^r$$

$$T_{2+1} = 5C_2\, (3x)^{5-2}\, (y)^2$$

$$= 5C_2\, 3^3\ x^3\, y^2$$

$$= 270x^3y^2$$

Similarly we can find by putting $r = 3$ in T_{r+1} to get T_4, then $T_4 = 90x^2y^3$

Binomial theorem for a rational index:

For any rational number other than positive integer

$$(1+x)^n = 1 + \frac{n}{1!}x + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \ldots, \quad \text{provided } |x| < 1.$$

NOTE

(i) If $n \in N$, $(1+x)^n$ is desired for all value of x and if n is a rational number other than the natural number then $(1+x)^n$ is desired only when $|x| < 1$.

(ii) If $n \in N$, then the expansion of $(1+x)^n$ contains only $(n+1)$ terms. If n is a rational number, then the expansion of $(1+x)^n$ contain infinitely many terms.

Some important expansions:

$$(1-x)^n = 1 - \frac{n}{1!}x + \frac{n(n-1)}{2!}x^2 - \frac{n(n-1)(n-2)}{3!}x^3 + \ldots$$

$$(1+x)^{-n} = 1 - \frac{n}{1!}x + \frac{n(n+1)}{2!}x^2 - \frac{n(n+1)(n+2)}{3!}x^3 + \ldots$$

$$(1-x)^{-n} = 1 + \frac{n}{1!}x + \frac{n(n+1)}{2!}x^2 + \frac{n(n+1)(n+2)}{3!}x^3 + \ldots$$

Special cases of infinite series:

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \ldots$$

$$(1-x)^{-1} = 1 + x + x^2 + x^3 + \ldots$$

$$(1+x)^{-2} = 1 - 2x + 3x^2 - 4x^3 + \ldots$$

$$(1-x)^{-2} = 1 + 2x + 3x^2 + 4x^3 + \ldots$$

Ch7_11th Stats.indd 201 07/12/2021 12:08:40

Example 7.17

Find the approximate value of $\sqrt[3]{1002}$ (correct to 3 decimal places) using Binomial series.

Solution:

$$\begin{aligned}
\sqrt[3]{1002} &= \sqrt[3]{1000+2} \\
&= \sqrt[3]{1000\left(1+\frac{2}{1000}\right)} \\
&= [(10^3)^{\frac{1}{3}}(1+0.002)^{\frac{1}{3}}] \\
&= 10[1+0.002]^{\frac{1}{3}} \\
&= 10[1+\frac{1}{3}[0.002]+\ldots] \\
&= 10[1+0.00066+\ldots] \\
&= 10[1+0.00066] \\
&= [10+0.0066] \\
&= 10.007
\end{aligned}$$

7.4.2 Exponential series

For all real values of x, $e^x = 1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots$ is known as exponential series, where e is an irrational number and the value of e to six decimal places is

$e = 2.718282$ … and

$e = 1+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\ldots\infty$, $e^{-1}=1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\ldots$

The other exponential expansions are

$e^{-x}= 1-\frac{x}{1!}+\frac{x^2}{2!}-\frac{x^3}{3!}+\ldots$

$\frac{e^x+e^{-x}}{2} = 1+\frac{x^2}{2!}+\frac{x^4}{4!}+\ldots$ and $\frac{e^x-e^{-x}}{2}\ x+\frac{x_3}{3!}+\frac{x^5}{5!}+\ldots$

$\frac{e+e^{-1}}{2} = 1+\frac{1!}{2!}+\frac{1}{4!}+\ldots$ and $\frac{e-e^{-1}}{2}1+\frac{1}{3!}+\frac{1}{5!}+\ldots$

Example 7.18

Show that $\frac{\frac{1}{1!}+\frac{1}{3}+\frac{1}{5!}+\ldots}{\frac{1}{2!}+\frac{1}{4!}+\frac{1}{6!}+\ldots} = \frac{e+1}{e-1}$

Ch7_11th Stats.indd 202 07/12/2021 12:08:43

Solution:

$$\frac{\frac{1}{1!}+\frac{1}{3}+\frac{1}{5!}+\ldots}{\frac{1}{2!}+\frac{1}{4!}+\frac{1}{6!}+\ldots} = \frac{\frac{e-e^{-1}}{2}}{\left[\frac{e+e^{-1}}{2}-1\right]}$$

$$= \left[\frac{e-e^{-1}}{2}\right] \div \left[\frac{e+e^{-1}-2}{2}\right]$$

$$= \frac{e-e^{-1}}{e+e^{-1}-2}$$

$$= \left[e-\frac{1}{e}\right] \div \left[e+\frac{1}{e}-2\right]$$

$$= \left[\frac{e^2-1}{e}\right] \div \left[\frac{e^2-2e+1}{e}\right]$$

$$= \frac{(e^2-1)}{(e^2-2e+1)}$$

$$= \frac{(e^2-1)}{(e-1)^2}$$

$$= \frac{(e-1)(e+1)}{(e-1)^2}$$

$$= \frac{e+1}{e-1}$$

7.4.3 Logarithmic series:

If $|x| < 1$, the series, $x - \frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\ldots$ converges to $\log(1+x)$

Some important deductions from the above series are

(i) $\log(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\ldots$

(ii) $\log(1-x) = -x-\frac{x^2}{2}-\frac{x^3}{3}-\frac{x^4}{4}+\ldots$

(iii) $\log(1+x) - \log(1-x) = 2\left(x+\frac{x^3}{3}+\frac{x^5}{5}+\ldots\right)$

or $\frac{1}{2}\log\left(\frac{1+x}{1-x}\right) = x+\frac{x^3}{3}+\frac{x^5}{5}+\ldots$

7.5 Introduction to Elementary calculus

Before going to understand the problems on continuous random variables, we need to know some fundamental knowledge about differentiation and integration, which are part of calculus in higher mathematics.

Hence, we introduce some simple concepts, techniques and formulae to calculate problems in statistics, which involve calculus.

Ch7_11th Stats.indd 203 07/12/2021 12:08:47

7.5.1 Differentiation

We Studied about functions and functional values in earlier classes. Functional value is an exact value. For some function $f(x)$, when $x = a$, we obtain the *functional value* as $f(a) = k$.

Another type of approximation gives the very nearest value to the functional value is known as *limiting value*. So the limiting value is an approximate value. This limiting value approaches the nearest to the exact value k.

Suppose the exact value is 4, the limiting value may be 4.00000001 or 3.999999994. Here we observe that the functional value and the limiting value are more or less the same and there is no significant difference between them.

Hence in many occasions we use the limiting values for some critical problems.

The limiting value of $f(x)$ when x approaches a number 2 is denoted by $\lim_{x \to 2} f(x) = f(2) = l$ (some existing value)

The special type of any existing limit, $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$ is called the *derivative* of the function f with respect to x and is denoted by $f'(x)$. If y is a function of x, and has a derivative, then the differential coefficient of y with respect to x is denoted by $\frac{dy}{dx}$ .This process of finding the limiting value is known as *differentiation*.

Some rules on differentiation:

(i) Derivative of a constant function is zero.
i.e., $f'(c) = 0$, where c is some constant.

(ii) If u is a function of x and k is some constant and dash denotes the differentiation, $[k\,u]' = k[u]'$

(iii) $(u \pm v)' = u' \pm v'$

(iv) $(u\,v)' = u'v + u\,v'$ (product rule)

(v) $\left[\frac{u}{v}\right]' = \frac{u'v - uv'}{v^2}$ (quotient rule)

Important formulae:

(i) $(x^n)' = n\,x^{n-1}$

(ii) $(e^x)' = e^x$

(iii) $(\log x)' = \frac{1}{x}$

Ch7_11th Stats.indd 204 07/12/2021 12:08:48

Example 7.19

Evaluate the following limits:

(i) $\lim_{x \to 5} \frac{x^3 - 25}{x - 3}$ (ii) $\lim_{x \to 2} \frac{x^2 - 4}{x - 2}$

Solution:

(i) $\lim_{x \to 5} \frac{x^3 - 25}{x - 3} = \frac{5^3 - 25}{5 - 3} = \frac{100}{2} = 50$

(ii) $\lim_{x \to 2} \frac{x^2 - 4}{x - 2} = \frac{0}{0}$, an indeterminate form. Therefore first factorise and simplify and then apply the same limit to get the limiting value.

$$\lim_{x \to 2} \frac{x^2 - 4}{x - 2} = \lim_{x \to 2} \frac{(x + 2)(x - 2)}{x - 2}$$

$$= \lim_{x \to 2} (x + 2)$$

$$= 2+2 = 4$$

$$\therefore \quad \lim_{x \to 2} \frac{x^2 - 4}{x - 2} = 4$$

Example 7.20

Find the derivative of the following with respect to x:

(i) x^{15}-10 (ii) x^3+3x^2-6 (iii) x^4e^x (iv) $\frac{x^2 - 1}{x + 3}$

Solution:

(i) Let $y = x^{15}$-10

$$\frac{dy}{dx} = 15x^{15-1} - 0$$

$$= 15x^{14}$$

(ii) Let $y = x^3+3x^2$-6

$$\frac{dy}{dx} = 3x^2+3(2x)-0$$

$$= 3x^2+6x$$

(iii) Let $y = x^4e^x$

This is of the type

$$[uv]' = u'v + uv'$$

Ch7_11th Stats.indd 205 07/12/2021 12:08:51

$$= [x^4]' (e^x) + (x^4) [e^x]'$$

$$= 4x^3e^x + x^4e^x$$

$$= (4x^3 + x^4)e^x$$

(iv) Let $y = \frac{x^2-1}{x+3}$

This is of the type

$$\left[\frac{u}{v}\right]' = \frac{u'v - uv'}{v^2}$$

$$\frac{dy}{dx} = \left[\frac{x^2-1}{x+3}\right]'$$

$$= \frac{(x^2-1)'(x+3) - (x^2-1)(x+3)'}{(x+3)^2}$$

$$= \frac{(2x)(x+3) - (x^2-1)\times 1}{(x+3)^2}$$

$$= \frac{2x^2+6x-x^2+1}{(x+3)^2}$$

$$= \frac{x^2+6x+1}{(x+3)^2}$$

Repeated differentiation:

If the derivative of a function is again differentiated with respect to the same variable, we say that the differentiation is the second order differentiation and is denoted as $\frac{d}{dx}(\frac{dy}{dx}) = \frac{d^2y}{dx^2}$ *or* D^2y *or* Y_2

For example if $y = (x^3+4x^2+7)$, then $\frac{dy}{dx} = (3x^2+8x)$

Again differentiating with respect to x, we get

$$\frac{d^2y}{dx} = \frac{d}{dx}(\frac{dy}{dx}) = \frac{d}{dx}(3x^2+8x)$$

$$= 6x+8$$

NOTE

Repeated or successive differentiation may be extended to any higher order derivation.

7.5.2 Integration

Integration is the reverse process of differentiation. It is also called anti-derivative.

Suppose the derivative of x^5 is $5x^4$. Then the integration of $5x^4$ with respect to x is x^5. we use this in symbol as follows:

Ch7_11th Stats.indd 206 07/12/2021 12:08:53

$$\frac{d}{dx}(x^5) = 5x^4 \Rightarrow \int 5x^4\,dx = x^5$$

Similarly,

$$\frac{d}{dx}(x^7) = 7x^6 \Rightarrow \int 7x^6\,dx = x^7$$

$$\frac{d}{dx}(e^x) = e^x \Rightarrow \int e^x\,dx = e^x$$

$$\frac{d}{dx}(\log x) = \frac{1}{x} \Rightarrow \int \frac{1}{x}\,dx = \log x$$

and so on.

NOTE

While differentiating the constant term we get zero. But in the reverse process, that is on integration, unless you know the value of the constant we cannot include. That is why we include an arbitrary constant c to each integral value.

Therefore for the above examples, we usually write

$$\int 7x^6\,dx = x^7 + c$$

$$\int e^x\,dx = e^x + c$$

These integrals are also called improper integrals or indefinite integrals

Rules and some formulae on integration:

(i) $\int k\,dx = kx$

(ii) $\int x^n\,dx = \frac{x^{n+1}}{n+1}$

(iii) $\int e^x\,dx = e^x$

(iv) $\int \frac{1}{x}\,dx = \log x$

(v) $\int (u \pm v)\,dx = \int u dx \pm \int v dx$

Example 7.21

Integrate the following with respect to x.

(i) x^7 (ii) $\frac{1}{x^6}$ (iii) $\sqrt{x}$ (iv) $x^5 - 4x^2 + 3x + 2$

Ch7_11th Stats.indd 207 07/12/2021 12:08:56

Solutions:

(i) $\int x^7\,dx = \frac{x^{7+1}}{7+1} = \frac{x^8}{8} + c$

(ii) $\int \frac{1}{x^6}\,dx$

$= \int x^{-6}\,dx$

$= \frac{x^{-6+1}}{-6+1} = \frac{x^{-5}}{-5}$

$= -\frac{1}{5x^5} + c$

(iii) $\int \sqrt{x}\,dx$

$= \int x^{1/2}\,dx$

$= \frac{x^{\frac{1}{2}+1}}{\frac{1}{2}+1}$

$= \frac{x^{\frac{3}{2}}}{\frac{3}{2}}$

$= \frac{2}{3}x^{\frac{3}{2}} + c$

(iv) $\int (x^5 - 4x^2 + 3x + 2)\,dx$

$= \left(\frac{x^6}{6}\right) - 4\left(\frac{x^3}{3}\right) + 3\left(\frac{x^2}{2}\right) + 2x + c$

The above discussed integrals are known as improper integrals or indefinite integrals . For the proper or definite integrals we have the limiting points at both sides. These are called the lower limit and the upper limit of the integral.

This integral $\int f(x)\,dx$ is an indefinite integral. Integrating the same function with in the given limits a and b is known as the definite integral. We write this in symbol as

$\int_a^b f(x)\,dx$ =k(a constant value)

In a definite integral where a is known as the lower limit and b is known as the upper limit of the definite integral. To find the value of definite integral, we do as follows:

Suppose $\int f(x)\,dx = F(x)$

Then $\int_a^b f(x)\,dx = F(b) - F(a)$

Ch7_11th Stats.indd 208 07/12/2021 12:09:01

Example 7.22

Evaluate the following definite integrals:

(i) $\int_1^3 x^3\,dx$ (ii) $\int_{-1}^{+1} 5x^4\,dx$ (iii) $\int_1^2 \frac{5}{x}\,dx$

Solutions:

(i) $\int_1^3 x^3\,dx = \left[\frac{x^4}{4}\right]_1^3$

$= \frac{1}{4}[3^4 - 1^4]$

$= \frac{1}{4}[81 - 1]$

$= \frac{1}{4}[80]$

$= 20$

(ii) $\int_{-1}^{+1} 5x^4\,dx = 5\int_{-1}^{1} x^4\,dx$

$= 5\left[\frac{x^5}{5}\right]_{-1}^{1}$

$= [x^5]_{-1}^{1}$

$= [1^5 - (-1)^5]$

$= 1-(-1)$

$= 1+1$

$= 2$

(iii) $\int_1^2 \frac{5}{x}\,dx$

$= 5\int_1^2 \frac{1}{x}\,dx$

$= 5\,[\log x]_1^2$

$= 5[log2 - log1]$

$= 5log2$

7.5.3 Double integrals

A double integral is an integral of two variable function $f(x,y)$ over a region R If $R=[a, b] \times [c, d]$ then the double integral can be done by iterated Integration(integrate first with respect to y and then with respect to x)

Ch7_11th Stats.indd 209 07/12/2021 12:09:04

The notation used for double integral is $\int_a^b \int_c^d f(x,y)\,dy\,dx$

Here the function $f(x,y)$ is integrated with respect to y first and treat $f(x)$ constant and then integrate with respect to x and apply limits of x and simplify

Example 7.23

Evaluate: $\int_0^1 \int_1^2 x^2 y\,dx\,dy$, for $0 \le x \le 1,\ 1 \le y \le 2$

Solutions:

Let us first integrate with respect to y and then with respect to x. Hence the double integral is written as

$$\int_0^1 x^2 \left[\int_1^2 y\,dy\right] dx$$

Now integrate the inner integral only and simplify.

$$= \int_0^1 x^2 \left[\frac{y^2}{2}\right]_1^2 dx$$

$$= \int_0^1 \frac{x^2}{2}[4-1]\,dx$$

NOTE

We can also change the order of integration

Again integrate with respect to x

$$= \frac{3}{2}\int_0^1 x^2\,dx$$

$$= \frac{3}{2}\left[\frac{x^3}{3}\right]_0^1 = \frac{1}{2}[1^3-0]$$

$$= \frac{1}{2}$$

Example 7.24

Evaluate: $\int_0^2 \int_0^1 [2y^2x^2+3]\,dy\,dx$

Solution:

$$= \int_0^2 \int_0^1 [2y^2x^2+3]\,dy\,dx = \int_0^2 \left(\int_0^1 \left(2y^2x^2+3\right)dy\right)dx$$

$$= \int_0^2 \left[\frac{2x^2}{3}y^3+3y\right]_0^1 dx = \int_0^2 \left[2x^2\left(\frac{1}{3}\right)+3(1)-0\right]dx$$

$$= \left[\frac{2}{3}\left(\frac{x^3}{3}\right)+3x\right]_0^2 = \left[\frac{16}{9}+6\right]$$

$$= \frac{70}{9}$$

Ch7_11th Stats.indd 210 07/12/2021 12:09:07

Example 7.25

Evaluate $\int_0^1 \int_0^1 16abxydx\,dy$

Solution:

$$\int_0^1 \int_0^1 16abxydx\,dy$$

$$= 16ab \int_0^1 \left[\int_0^1 xydx \right] dy$$

$$= 16ab \int_0^1 \left[\frac{x^2}{2} y \right]_0^1 dy$$

$$= 16ab \int_0^1 \left[\frac{y}{2} \right] dy$$

$$= 8ab \left[\frac{y^2}{2} \right]_0^1 = 4ab[1-0]$$

$$= 4ab$$

Points to Remember

- Fundamental Principle of multiplication on counting: If an operation can be performed in m ways and if another operation in n ways independent of the first, then the number of ways of performing both the operations simultaneously in $m \times n$ ways.
- Factorial: The consecutive product of first n natural numbers is known as factorial n and is denoted as $n!$ or $\lfloor \underline{n}$
- Permutations : Thus the number of arrangements that can be formed out of n things taken r at a time is known as the number of permutation of n things taken r at a time and is denoted as nP_r; $nP_r = n(n-1)(n-2) - - - - [n-(r-1)] = \frac{n!}{(n-r)!}$
- Combinations : The number of combination of n different things, taken r at a time is given by $nC_r = \frac{np_r}{r!} = \frac{n!}{(n-r)!r!}$, $nC_r = nC_{n-r}$
- Binomial series : For any natural number n

  $(x+a)^n = nc_0 x^n a^0 + nc_1 x^n - 1_a + nc_2 x^{n-2} a^2 + \ldots + nc_r x^{n-r} a^r + \ldots + nc_n a^n$

  For any rational number other than positive integer

  $(1+x)^n = 1 + \frac{n}{1!}x + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \ldots$, provided $|x| < 1$.
- Exponential series : For all real values of x, $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$ is known as exponential series, Here, e= 2.718282

Ch7_11th Stats.indd 211 07/12/2021 12:09:09

- Differentiation : The special type of any existing limit, $\lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$ is called the derivative of the function f with respect to x and is denoted by $f'(x)$. If y is a function of x, and has a derivative, then the differential coefficient of y with respect to x is denoted by $\frac{dy}{dx}$.

- Some rules on differentiation:

(i) Derivative of a constant function is zero. $f'(c) = 0$, where c is some constant.

(ii) If u is a function of x and k is some constant and *dash* denotes the differentiation, $[k\,u]' = k\,[\,u\,]'$

(iii) $(u \pm v)' = u' \pm v'$

(iv) $(u\,v)' = u'v + u\,v'$

(v) $\left[\frac{u}{v}\right]' = \frac{u'v - uv'}{v^2}$

- Important formulae:

(i) $(x^n)' = n\,x^{n-1}$

(ii) $(e^x)' = e^x$

(iii) $(\log x)' = \frac{1}{x}$

- Integration

Integration is the reverse process of differentiation. It is also called anti-derivative.

Rules and some formulae on integration:

(i) $\int k\,dx = kx$

(ii) $\int x^n dx = \frac{x^{n+1}}{n+1}$

(iii) $\int e^x dx = e^x$

(iv) $\int \frac{1}{x} dx = \log x$

(v) $\int (u \pm v)\,dx = \int u dx \pm \int v dx$

- Definite Integral : $\int_a^b f(x)\,dx = F(b) - F(a)$

- Double integrals : The double integral defined on the function $f(x,y)$ is denoted as $\int_a^b \int_c^d f(x,y)\,dy\,dx$. The function $f(x,y)$ is integrated with respect to y first by treating $f(x)$ as constant and then integrate with respect to x and apply limits of x and simplify we get the value of integral.

Ch7_11th Stats.indd 212 07/12/2021 12:09:10

EXERCISE 7

F3MYU

I. Choose the best answer:

1. The factorial n is also written as

(a) $n(n\text{-}1)!$ (b) $n(n+1)!$ (c) $(n\text{-}1)!$ (d) $(n+1)!$

2. The value of $10P_2$ is

(a) 10 (b) 45 (c) 90 (d) 20

3. The number of different four letter words can be formed with the words 'DATE' is

(a) 4 (b) 8 (c) 24 (d)48

4. The value of $50C_{50}$ is equal to

(a) 50 (b) 25 (c)1 (d)0

5. The value of $20C_{18}$ is

(a) 190 (b) 180 (c) 360 (d) 95

6. $\lim_{x \to 0} \frac{x^2-1}{x-1}$ is equal to

(a) –1 (b) 0 (c) 1 (d) 2

7. Derivative of $\log x$ is equal to

(a) 1 (b) $\frac{1}{x}$ (c) e^x (d) $\log x$

8. Derivative of x^9 is

(a) x^8 (b) $9x^8$ (c) $8x^9$ (d) $8x^8$

9. The integral value of x^{11} is

(a) $\frac{1}{12}x^{12}$ (b) x^{12} (c) $11x^{10}$ (d) $10x^{11}$

10. $\int e^x dx$ is equal to

(a) e^{x^2} (b)e^x (c) e^{-x} (d) xe^x

11. $\int_0^1 x^3 \, dx$ is

(a) $\frac{1}{4}$ (b) $\frac{1}{2}$ (c) 1 (d) 3

Ch7_11th Stats.indd 213 07/12/2021 12:09:11

II. Fill in the blanks:

12. The number of all permutations of n distinct things taken all at a time is _____

13. Number of ways of 4 students can stand in a queue is _____

14. From a group of 10 students 2 students are selected in _____ ways

15. The value of nC_n is equal to _____

16. The value of $21C_3$ is equal to _____

17. $\lim_{x \to 5} \frac{2x+3}{x+5}$ is equal to _____

18. The derivative of e^x is equal to _____

19. The integral of $\frac{1}{x}$ is _____

20. The value of $\int 6x^5\,dx$ is _____

21. The value of $\int_0^1 x^{10}\,dx$ is _____

III . Very Short Answer Questions :

22. A boy has 6 pants and 10 shirts. In how many ways can he wear them?

23. Evaluate (i) $4P_4$ (ii) $10P_4$ (iii) $100P_2$

24. How many different four letter code words can be formed with the letter 'SWIPE' no letter is not repeated.

25. In a competition, in how many ways can first and second place be awarded to 10 people?

26. Evaluate $10C_4$, $22C_3$, $100C_{98}$

27. Evaluate (i) $\lim_{x \to 2} \frac{x^2-4x+7}{x+8}$ (ii) $\lim_{x \to 5} \frac{x^2-25}{x-5}$

28. Find the derivative of x^2e^x

IV. Short Answer Questions :

29. How many different words can be formed with letters of the word 'PROBABILITY'?

30. A bag contains 7 blue balls and 5 red balls. Determine the number of ways in which 3 blue and 2 red balls can be selected.

Ch7_11th Stats.indd 214 07/12/2021 12:09:12

31. How many triangles can be formed by joining the vertices of a hexagon?

32. In how many ways can 3 vowels and 2 consonants be chosen from {E,Q,U,A,T,I,O,N}?

33. Different the following with respect to x:
(i) $e^x(x^2-5)$ (ii) $\frac{x^2-1}{x+7}$

34. Find $\int (3x^3-2x^2+6x-7)\,dx$

35. Find $\int (\frac{2}{x}+\frac{2}{x^2}-e^x+3)\,dx$

36. Find $\int_0^2 \frac{3}{4}(2-x)\,dx$

37. Find $\int_{-1}^{1} 2x\,dx$

38. Evaluate $\int_0^2 \int_0^1 [2x+3y]\,dydx$

V. Calculate the following :

39. A class contains 12 boys and 10 girls. From the class, 10 students are to be chosen for a competition under the condition that at least 4 boys and at least 4 girls must be represented. In how many ways can the selection are made

40. What is the number of ways of choosing 4cards from a pack of 52 playing cards? In how many of these (i) 3 cards are of the same suit (ii) 4 cards are belong to different suits (iii) two are red cards and two are black cards

41. There are 3 questions in the first section, 3 questions in the second section and 2 questions in the third section in a question paper of an exam. The student has to answer any 5 questions, choosing atleast one from each section. In how many ways can the student answer the exam?

42. Differentiate the following with respect to x:
(i) $(x^3-4x^2+2x)(e^x+5)$ (ii) $\frac{x^2-7x}{x^2+8}$

43. Integrate the following with respect to x: (i) $\frac{x^2-5x+6}{x}$ (ii) $2\sqrt{x}+\frac{3}{x^4}$

44. Evaluate: (i) $\int_0^2 \frac{3}{4}x(2-x)\,dx$ (ii) $\int_1^2 \frac{12}{x}\,dx$

Ch7_11th Stats.indd 215 07/12/2021 12:09:15

45. Evaluate $\int_0^1 \int_0^1 x^2\, y dx dy$

46. Evaluate $\int_1^2 \int_2^4 6x\, y^2\, dx dy$

47. Evaluate $\int_0^2 \int_0^2 4mnxy dx dy$

48. $\int_1^2 \int_1^3 (4x-2y)\, dx dy$

ANSWERS:

I. 1. (a) , 2. (c) , 3. (c) , 4. (c) , 5. (a) , 6. (c) , 7. (b) , 8. (b) , 9. (a) , 10. (b) , 11. (a)

II. 12. n! 13. 24 , 14. 45, 15.1 , 16. 1330 , 17. $\frac{13}{10}$, 18. e^x , 19. $\log x$, 20. x^6+c, 21. $\frac{1}{11}$

III. 22. 60, 23. (i) 24 , (ii) 5040 , (iii) 9900 , 24.120 , 25. 90, 26. (i) 210 , (ii) 1540 , (iii) 4950 , 27. (i) $\frac{3}{10}$, (ii) 10 , 28. $e^x\ (x^2+2x)$

IV. 29. $\frac{11}{2!2!}$, 30. 350 , 31. 20 , 32. 30 , 33. (i) $e^x\ (x^2+2x-5)$, (ii) $\frac{x^2+14x+1}{(x+7)^2}$, 34. $\frac{3}{4}x^4-\frac{2}{3}\ x^3+3x^2-7x+c$ 35. $2\log x - \frac{2}{x} - e^x + 3x + c$ 36. $\frac{3}{2}$, 37. 0, 38. 7

V. 39. 497574 , 40. (i) $4(13C_3.39C_1)$, (ii) $(13C_1)^4$, (iii) $26C_2.26C_2$, 41. 48, 42. (i) $e^x\ (x^3-x^2-6x+2)+5(3x^2-8x+2)$, (ii) $\frac{7x^2-16x-56}{(x^2+8)^2}$ 43. (i) $\frac{x^2}{2}-5x+6\log x+c$, (ii) $\frac{4}{3}x^{\frac{3}{2}}-\frac{3}{x}+c$, 44. (i) 1, (ii) 12 log2 45. $\frac{1}{6}$, 46. 84 , 47. 16 mn , 48. 24

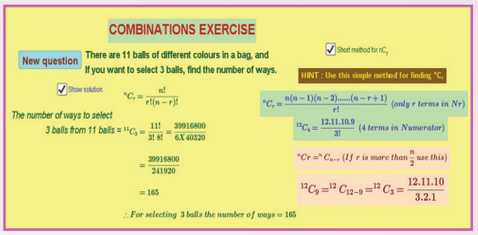
Ch7_11th Stats.indd 216 07/12/2021 12:09:20



ICT CORNER

SOME MATHEMATICAL METHODS-COMBINATION

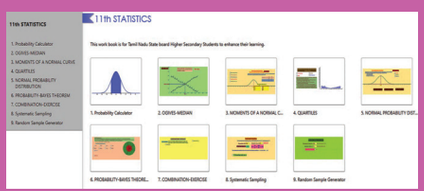
This activity helps to practice combinations. Which is an important fundamental for probability



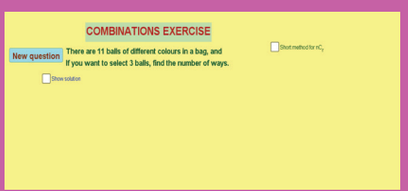
Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for statistics are given, open the worksheet named "**Problems in Combination**"
- A problem is given in the work sheet. Solve it. To check your answer, click on the box "**Show Solution**". You can change the question by pressing on "**NEW QUESTION**".
- Before going to the new question click on "**Short method for ^nC_r**" and follow the steps given carefully and practice it. Now go to new question and practice till you are thorough as it is important for doing probability.

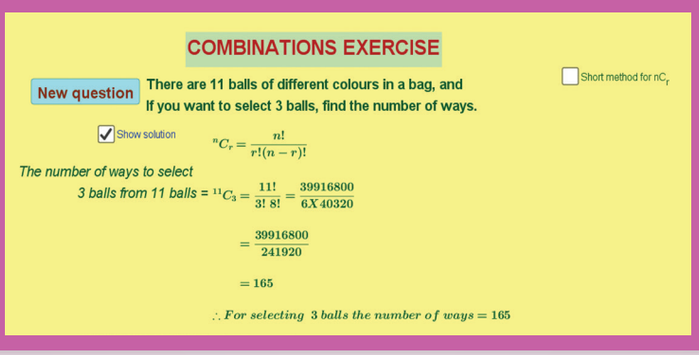
Step-1



Step-2



Step-3



Pictures are indicatives only*

URL:

https://ggbm.at/uqVhSJWZ

Ch7_11th Stats.indd 217 07/12/2021 12:09:21

Chapter

8

Elementary Probability Theory



Blaise Pascal

(19 June, 1623 - 19 August, 1662)

Blaise Pascal was a French mathematician, who laid the foundation for the modern theory of probabilities. He was a child prodigy.

In 1642, while still a teenager, he started some pioneering work on calculating machines. After three years of effort, he built 20 finished machines (called Pascal's calculators) over the following 10 years, establishing him as one of the first two inventors of the mechanical calculator.

Pascal was an important mathematician, he wrote a significant treatise on the subject geometry and projective geometry at the age of 16, and later corresponded with Pierre de Fermat on probability theory, strongly influencing the development of modern economics and social science. Pascal's results caused many disputes before being accepted.

'The scientific imagination always restrains itself within the limits of probability'.

- Thomas Huxley

X14H3

Learning Objectives

- Knows about random experiments, trials, outcomes, events, sample space
- Understands the theorems on probability and applies in problems.
- Understands Independent events and multiplication theorem on probability.
- Explains conditional probability.
- Knows the application of Bayes' Theorem.

Ch8_11th Stats.indd 218 07/12/2021 12:11:04

Introduction

Every scientific experiment conducted to investigate the patterns in natural phenomenon may result with "events" which may or may not happen. Most of the events in real life have uncertainty in their happening. For example,

Disintegration of a given atom of radium in a given time interval may or may not disintegrate;

A plant may or may not be infected by species during rainy season; The event of increase in the gold price under an economic condition in a country; A drug administered to a cancer patient for curing a disease in a period of time.

In all these cases, there is an amount of uncertainty prevails. Even for a student, asking a particular question in the examination from a particular portion of a subject is uncertain. Yet, the student is compelled to take a decision during the preparation of examination, whether to go for an in-depth study or leaving the question in choice. In a nutshell, one has to take a wise decision under the conditions of uncertainty. In such a situation, knowledge about the chance or probability for occurrence of an event of interest is vital and calculation of probability for happening of an event is imperative.

It is very much essential to determine a quantitative value to the chance or probability for the occurrence of random events in many real life situations. Before defining probability, students gain knowledge on the following essential terms.

8.1 Random experiment, Sample space, Sample point

Experiment: In Statistics, by the word experiment it means 'an attempt to produce a result'. It need not be a laboratory experiment.

Random Experiment: If an experiment is such that

(i) all the possible outcomes of the experiment are predictable, in advance

(ii) outcome of any trial of the experiment is not known, in advance, and

(iii) it can be repeated any number of times under identical conditions, is called a random experiment.

Sample space: The set of all possible outcomes of a random experiment is called the sample space of the experiment and is usually denoted by S (or Ω). If S contains only finite number of elements, it is termed as finite sample space. If S contains countable number of elements, S may be called as countable sample space or discrete sample space. Otherwise, S is called an uncountable sample space.

Ch8_11th Stats.indd 219 07/12/2021 12:11:04

Sample Point: The outcome of a random experiment is called a sample point, which is an element in S.

Example 8.1

Consider the random experiment of tossing a coin once "Head" and "Tail" are the two possible outcomes. The sample space is S = {H, T}. It is a finite sample space which is presented in fig. 8.1.



Fig. 8.1. Tossing of a Coin Once

Example 8.2

Suppose that a study is conducted on all families with one or two children. The possible outcomes, in the order of births, are: boy only, girl only, boy and girl, girl and boy, both are boys and both are girls. Then, the sample space is S = {b,g,bg,gb,bb,gg}. It is also a finite sample space. Here, 'b' represents the child is a boy and 'g' represents the child is a girl.

Example 8.3

Consider the experiment of tossing a coin until head appears. Then, the sample space of this experiment is S = { (H), (T,H), (T,T,H), (T,T,T,H), … }. This is a countable sample space. If head appears in the first trial itself, then the element of S is (H); if head appears in the second attempt then the element of S is (T,H); if head appears in the third attempt then the element of S is (T,T,H) and so on.

Example 8.4

In the experiment of observing the lifetime of any animate or inanimate things, the sample space is

$$S = \{x: x \geq 0\},$$

where x denotes the lifetime. It is an example for uncountable sample space.

Ch8_11th Stats.indd 220 07/12/2021 12:11:04

Event: A subset of the sample space is called an event. In this chapter, events are denoted by upper case English alphabets and the elements of the subsets by lower case English alphabets.

In Example 8.2, the event that the eldest child in the families is a girl is represented as

$$A = \{g, gb, gg\}$$

The event that the families have one boy is represented as

$$B = \{b, bg, gb\}.$$

In Example 8.4, A refers to the event that the refrigerator works to a maximum of 5000 hours. Then $A = \{ x: 0 < x \leq 5000 \}$ is the subset of

8.1.1 Mutually exclusive events

Two or more events are said to be mutually exclusive, when the occurrence of any one event excludes the occurrence of other event. Mutually exclusive events cannot occur simultaneously.

In particular, events A and B are said to be mutually exclusive if they are disjoint, that is, $A \cap B = \varnothing$



Consider the case of rolling a die. Let A = {1, 2, 3} and B = {4, 5, 6} be two events. Then we find $A \cap B = \varnothing$. Hence A and B are said to be mutually exclusive events.

8.2 Definitions of Probability

Probability is a measure of uncertainty. There are three different approaches to define the probability.

8.2.1 Mathematical Probability (Classical / a priori Approach)

If the sample space S, of an experiment is finite with all its elements being equally likely, then the probability for the occurrence of any event, A, of the experiment is defined as

$$\mathrm{P(A)} = \frac{No.of\ elements\ favourable\ to\ A}{No.\ of\ elements\ in\ S}$$

$$\mathrm{P(A)} = \frac{n(A)}{n(S)} \ .$$

The above definition of probability was used until the introduction of the axiomatic method. Hence, it is also known as classical definition of probability. Since this definition

Ch8_11th Stats.indd 221 07/12/2021 12:11:06

enables to calculate the probability even without conducting the experiment but using the prior knowledge about the experiment, it is also called as a priori probability.

Example 8.5

What is the chance of getting a king in a draw from a pack of 52 cards?

Solution:

In a pack there are 52 cards [n(s) = 52] which is shown in fig. 8.2

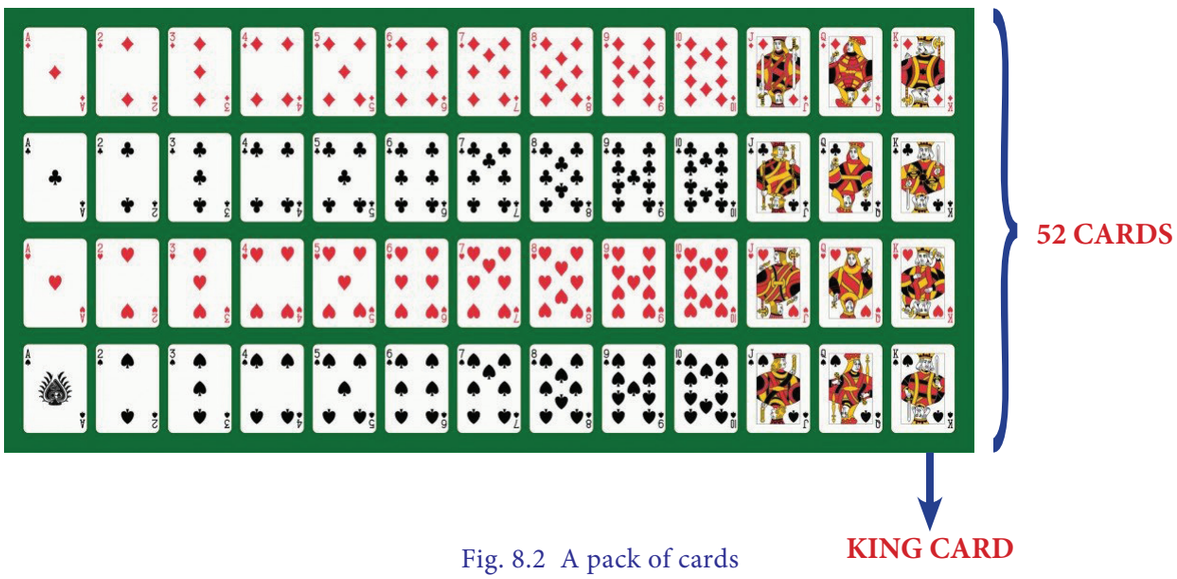


Fig. 8.2 A pack of cards

Let A be the event of choosing a card which is a king

In which, number of king cards $n(A) = 4$

Therefore probability of drawing a card which is king is = $P(A) = \frac{n(A)}{n(S)} = \frac{4}{52}$

Example 8.6

A bag contains 7 red, 12 blue and 4 green balls. What is the probability that 3 balls drawn are all blue?

Solution:

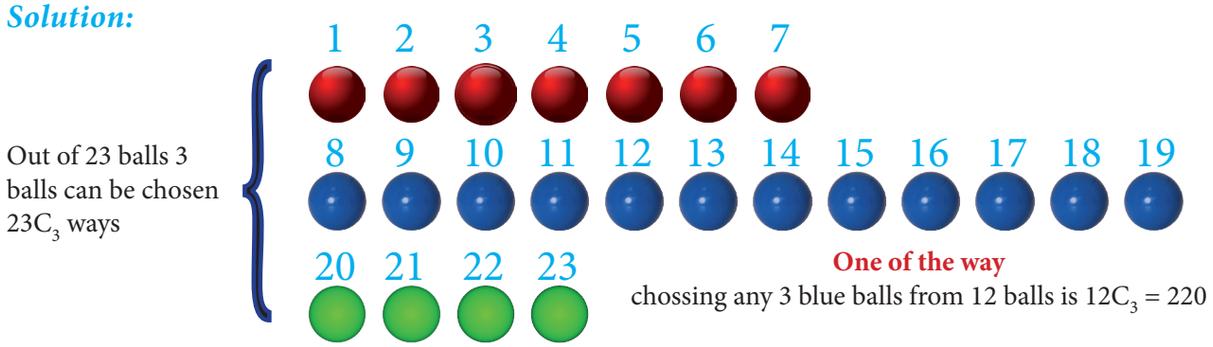


Fig. 8.3 Selection of 3 blue balls out of 23 balls

Ch8_11th Stats.indd 222 07/12/2021 12:11:08

From the fig. 8.3 we find that:

Total number of balls = 7+12+14=23 balls

Out of 23 balls 3 balls can be selected in = $n(s)= 23C_3$ ways

Let A be the event of choosing 3 balls which is blue

Number of possible ways of drawing 3 out of 12 blue balls is = $n(A)=12C_3$ ways

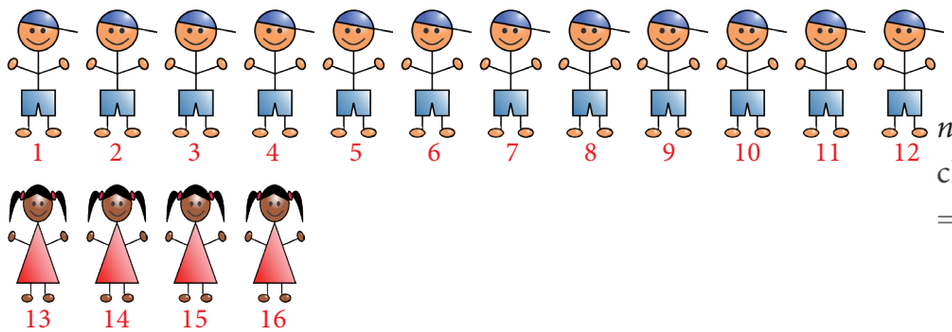
Therefore, $P(A) = \frac{n(A)}{n(S)} = \frac{12C_3}{23C_3} = \frac{220}{1771}$

$= 0.1242$

Example 8.7

A class has 12 boys and 4 girls. Suppose 3 students are selected at random from the class. Find the probability that all are boys.

Solution:



$n(s) = 16$

choosing 3 students

$= 16\ C_3$

chossing any 3 boys out of 12 boys is $12C_3 = 220$ ways

fig. 8.4 Selection of 3 Boys

From the fig 8.4, we find that:

Total number of students = 12+4=16

Three students can be selected out of 16 students in $16C_3$ ways

i.e. $n(s) = 16C_3 = \frac{16 \times 15 \times 14}{1 \times 2 \times 3} = 560$

Three boys can be selected from 12 boys in $12C_3$ ways

i.e. $n(A) = 12C3 = \frac{12 \times 11 \times 10}{1 \times 2 \times 3} = 220$

The required probability $P(A) = \frac{n(A)}{n(S)} = \frac{220}{560} = \frac{11}{28}$

$= 0.392$

Ch8_11th Stats.indd 223 07/12/2021 12:11:09

8.2.2 Statistical Probability (Relative Frequency/a posteriori Approach)

If the random experiment is repeated n times under identical conditions and the event A occurred in n(A) times, then the probability for the occurrence of the event A can be defined (Von Mises) as

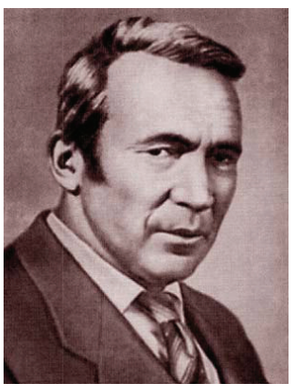
$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}.$$

Since computation of probability under this approach is based on the empirical evidences for the occurrence of the event, it is also knows as relative frequency or a posteriori probability.

NOTE

It should be noted that repeating some experiments is not always possible. In such cases, relative frequency approach cannot be applied. Classical approach to compute probability requires that the sample space should be a finite set. It is seldom possible. Hence, we present in the next section axiomatic approach to probability which overcomes the limitations of both mathematical probability and statistical probability.

8.3 Axioms of Probability



(A.N. Kolmogorov)

Andrey Nikolaevich Kolmogorov (1903–1987) was a 20th-century Soviet mathematician who made significant contributions to the mathematics of probability theory, topology, intuitionistic logic, turbulence, classical ... Wikipedia

A.N. Kolmogorov proposed the axiomatic approach to probability in 1933. An axiom is a simple, indisputable statement, which is proposed without proof. New results can be found using axioms, which later become as theorems.

8.3.1 Axiomatic approach to probability

Let S be the sample space of a random experiment. If a number $P(A)$ assigned to each event $A \in S$ satisfies the following axioms, then $P(A)$ is called the probability of A.

Axiom-1 : $P(A) \geq 0$

Axiom-2 : $P(S) = 1$

Axiom-3 : If $\{A_1, A_2, \ldots\}$ is a sequence of mutually exclusive events i.e., $A_i \cap A_j = \phi$

Ch8_11th Stats.indd 224 07/12/2021 12:11:10

when $i \neq j$, then

$$P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Axiom-3 also holds for a set of finite number of mutually exclusive events. If $A_1, A_2, \ldots,$ An are mutually exclusive events in S and n is a finite positive integer, then

$$P(A_1 \cup A_2 \ldots A_n) = P(A_1) + P(A_2) + \ldots + P(A_n).$$

It may be noted that the previous two approaches to probability satisfy all the above three axioms.

8.3.2 Basic Theorems of Probability

Theorem 8.1: The probability of impossible event is 0 i.e., $P(\phi) = 0$.

Proof: Let $A_1 = S$ and $A_2 = \phi$. Then, A_1 and A_2 are mutually exclusive.

$S = A_1 \cup A_2 = S \cup \phi$

Thus, by Axiom -3,

$$P(S) = P(S) + P(\phi)$$

Since by Axiom-2, $P(S) = 1$,

$$1 = 1 + P(\phi)$$

Hence, $P(\phi) = 0$.

Theorem 8.2: If S is the sample space and A is any event of the experiment, then $P(\bar{A}) = 1 - P(A)$.

Proof:

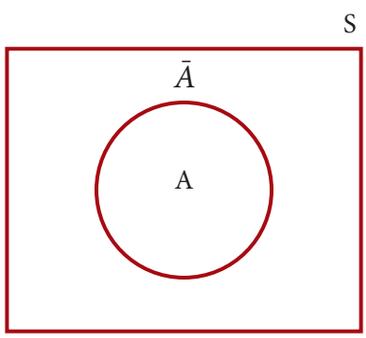


fig. 8.5 Venn diagram

Since A and $\bar{A}$ are mutually exclusive,

$$A \cup \bar{A} = S. \text{ (By Axiom -3)}$$

Ch8_11th Stats.indd 225 07/12/2021 12:11:11

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) \text{ (see figure 8.4)}$$

$$P(S) = P(A) + P(\bar{A})$$

$$1 = P(A) + P(\bar{A}) \quad \text{(by Axiom-2)}$$

It implies that $P(\bar{A}) = 1 - P(A)$.

NOTE

Since $\bar{A}$ is an event, by Axiom -1, $P(\bar{A}) \geq 0$.
i.e., $1 - P(A) \geq 0$ $\quad P(A) \leq 1$.

Thus, the probability for the occurrence of any event is always a real number between 0 and 1 i.e., $0 \leq P(A) \leq 1$.

Theorem 8.3: If A and B are two events in an experiment such that $A \subset B$, then $P(B\text{-}A) = P(B) - P(A)$.

Proof:

It is given that $A \subset B$.

The event B can be expressed as

$B = A \cup (B\text{-}A)$ (see Figure 8.6)

Since $A \cap (B\text{-}A) = \varphi$,

$$P(B) = P(A \cup (B\text{-}A))$$

Hence, by Axiom-3,

$$\Rightarrow \quad P(B) = P(A) + P(B\text{-}A)$$

Therefore, $P(B\text{-}A) = P(B) - P(A)$.

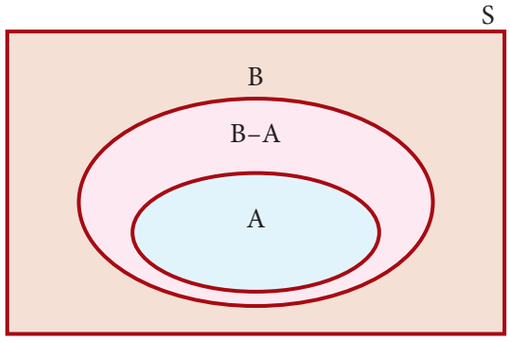


fig. 8.6 Venn diagram

Corollary: If $A \subset B$, then $P(A) \leq P(B)$.

Proof:

Since, by Axiom-1, $P(B-A) \geq 0$, it follows that

$P(B) - P(A) \geq 0$

$P(B) \geq P(A)$

$$\Rightarrow \quad P(A) \leq P(B).$$

Ch8_11th Stats.indd 226 07/12/2021 12:11:11

Example 8.8

In the experiment of tossing an unbiased coin (or synonymously balanced or fair coin), the sample space is $S = \{H, T\}$. What is the probability of getting head or tail?

Solution :

If the events A_1 and A_2 are defined as $A_1 = \{H\}$ and $A_2 = \{T\}$, then $S = A_1 \cup A_2$. Here, the events A_1 and A_2 are mutually exclusive, because they cannot occur together. Hence, by using Axiom-3, it can be written as

$$P(S) = P(A_1)+P(A_2)$$

Since the number of elementary events in S in favour of the occurrence of A_1 and A_2 is one each, they have equal chances to occur. Hence, $P(A_1) = P(A_2)$. Substituting this, it follows that

$$1 = P(A_1)+P(A_2) = 2P(A_1)$$

It gives that $P(A_1) = \frac{1}{2}$ and therefore $P(A_2) = \frac{1}{2}$.

$\Rightarrow P$(Getting a Head) $= \frac{1}{2} = P$(Getting a Tail).

Thus, a coin is called unbiased, if the probability of getting head is equal to that of getting tail.

Aliter: (Applying Classical approach)

Since $n(A_1) = 1 = n(A_2)$ and $n(S) = 2$,

$$P(A_1) = \frac{n(A_1)}{n(S)} = \frac{1}{2}.$$

Similarly, $P(A_2) = \frac{1}{2}$.

Remark:

If a biased coin is tossed and the outcome of head is thrice as likely as tail, then $P(A_1) = 3P(A_2)$.

Substituting this in $P(S) = P(A_1)+P(A_2)$, it follows that

$$1 = P(A_1)+P(A_2) = 4P(A_2)$$

It gives that $P(A_2) = \frac{1}{4}$ and hence $P(A_1) = \frac{3}{4}$.

It should be noted that the probabilities for getting head and tail differ for a biased coin.

Ch8_11th Stats.indd 227 07/12/2021 12:11:12

Example 8.9

Ammu has five toys which are identical and one of them is underweight. Her sister, Harini, chooses one of these toys at random. Find the probability for Harini to choose an underweight toy?

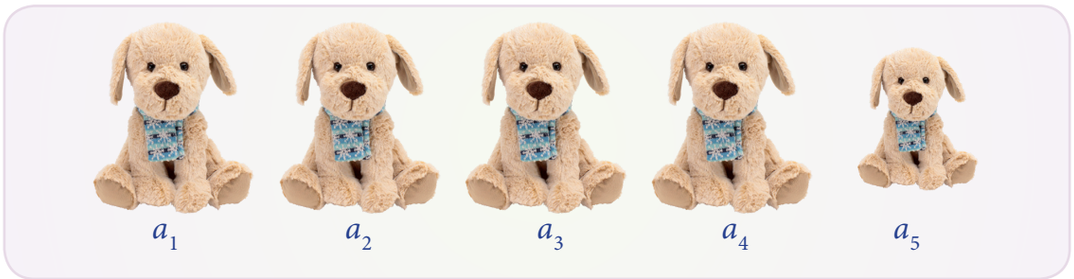


fig. 8.7 Identical toys with on underweight toy

Solution :

It is seen from fig. 8.7, the sample space is S = $\{a_1, a_2, a_3, a_4, a_5\}$.Define the events A_1, A_2, A_3, A_4 and A_5 as

A : Harini chooses the underweight toy

$$P(A) = \frac{n(A)}{n(S)} = \frac{1}{5}$$

Therefore, the probability for Harini to choose an underweight toy is 1/5.

Example 8.10

A box contains 3 red and 4 blue socks. Find the probability of choosing two socks of same colour.



= 7 socks (3 Red + 4 Blue)

i.e., $n = 7$

A_1 : choosing 2 socks from 3 socks

$n(A_1) = 3C_2$ ways = 3 ways

A_2 : choosing 2 socks from 4 socks

$n(A_2) = 4C_2$ ways = 6 ways

fig.8.8 choosing 2 socks of same colour

Solution :

From fig. 8.8, total number of socks = 3 + 4 = 7

If two socks are drawn at random, then

No. of ways of selecting 2 socks = $7C_2 = 21$

Ch8_11th Stats.indd 228 07/12/2021 12:11:14

A_1 = Selection of black socks,

$$n(A_1) = 3C_2 = 3$$

$$P(A_1) = \frac{n(A_1)}{n(S)} = \frac{3}{21}$$

A_2 = Selection of blue socks,

$$n(A_2) = 4C_2 = 6$$

$$P(A_2) = \frac{n(A2)}{n(S)} = \frac{6}{21}$$

then $A_1 \cup A_2$ represents the event of selecting 2 socks of same colour. Since the occurrence of one event excludes the occurrence of the other, these two events are mutually exclusive. Then, by Axiom-3,

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Therefore, $P(A_1 \cup A_2) = \frac{3}{21} + \frac{6}{21} = \frac{9}{21} = \frac{3}{7}$

Thus, the probability of selecting two socks of same colour is 3/7.

Example 8.11

Angel selects three cards at random from a pack of 52 cards. Find the probability of drawing:

(i) 3 spade cards.

(ii) one spade and two knave cards

(iii) one spade, one knave and one heart cards.

Solution:

Total no. of ways of drawing 3 cards = n(S) = 52 C_3 = 22100

(a) Let A_1 = drawing 3 spade cards.

Since there are 13 Spades cards in a pack of cards,

No. of ways of drawing 3 spade cards = n(A_1) = 13 C_3 = 286

Therefore, $P(A_1) = \frac{n(A_1)}{n(S)} = \frac{286}{22100}$

(b) Let A_2 = drawing one spade and two knave cards

No. of ways of drawing one spade card = $13C_1$ = 13

No. of ways of drawing two knave cards = $13C_2$ = 78

Ch8_11th Stats.indd 229 07/12/2021 12:11:15

Since drawing a spade and 2 knaves should occur together,

No. of ways drawing one spade and two knave cards = $n(A_2)$ = 13×78 =1014

$$\text{Therefore, } P(A_2) = \frac{n(A_2)}{n(S)} = \frac{13 \times 78}{22100}$$

$$\text{Hence, } P(A_2) = \frac{1014}{22100} = \frac{507}{11050}$$

(c) Let A_3 = drawing one spade, one knave and one heart cards

No. of ways of drawing one spade, one knave and one heart cards is

$$n(A_3) = 13C_1 \times 13C_1 \times 13C_1 = 13 \times 13 \times 13$$

$$\text{Therefore, } P(A_3) = \frac{n(A_3)}{n(S)} = \frac{13 \times 13 \times 13}{22100}$$

$$\text{Hence, } P(A_3) = \frac{2197}{22100}.$$

8.4 Addition Theorem of Probability

Theorem 8.4 : (Addition Theorem of Probability for Two Events)

If A and B are any two events in a random experiment, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof:

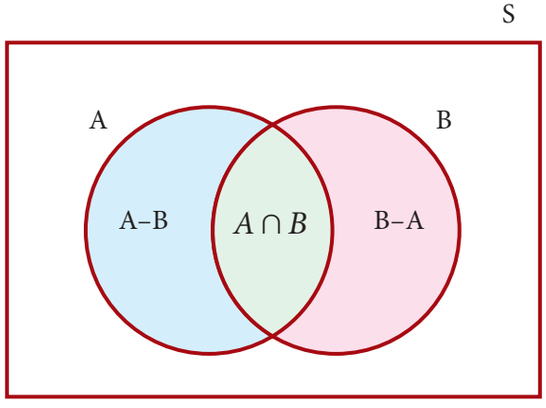


fig. 8.9 Venn diagram

For any two events A and B, the shaded region in fig. 8.9 represents the event A ∪ B.

$$A \cup B = A \cup (B - (A \cap B))$$

The events A and B-(A ∩B) are mutually exclusive.

Using Axiom 3,

$$P(A \cup B) = P(A \cup [B\text{-}(A \cap B)])$$

$$= P(A) + P[B\text{-}(A \cap B)] \qquad (8.1)$$

Ch8_11th Stats.indd 230 07/12/2021 12:11:17

Since $(A \cap B) \subset B$,,

$$B = (A \cap B) \cup (B - (A \cap B))$$

The events on the right hand side are disjoints. Hence by axiom 3

$$P(B) = P(A \cap B) + P(B - (A \cap B))$$

i.e. $$P[B - (A \cap B)] = P(B) - P(A \cap B) \qquad (8.2)$$

Substituting (8.2) in (8.1), it follows that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Corollary: If A, B and C are any three events, then

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Example 8.12

In the Annual sports meet, among the 260 students in XI standard in the school, 90 participated in Kabadi, 120 participated in Hockey, and 50 participated in Kabadi and Hockey. A Student is selected at random. Find the probability that the student participated in (i) Either Kabadi or Hockey, (ii) Neither of the two tournaments, (iii) Hockey only, (iv) Kabadi only, (v) Exactly one of the tournaments.

Solution:

$n(s)=260$

Let A : the event that the student participated in Kabadi

B : the event that the student participated in Hockey.

$n(A) = 90$; $n(B) = 120$; $n(A \cap B) = 50$

$$P(A) = \frac{n(A)}{n(S)} = \frac{90}{260}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{120}{260}$$

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{50}{260}$$

(i) The probability that the student participated in either Kabadi or Hockey is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\frac{90}{260} + \frac{120}{260} - \frac{50}{260} = \frac{160}{260} = \frac{8}{13}$$

Ch8_11th Stats.indd 231 07/12/2021 12:11:20

(ii) The probability that the student participated in neither of the two tournaments in

$$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) \text{ (By De Morgan's law } \overline{A \cup B} = \overline{A} \cap \overline{B})$$

$$= 1 - P(A \cup B)$$

$$= 1 - \frac{8}{13} = \frac{5}{13}$$

(iii) The probability that the student participated in Hockey only is

$$P(\overline{A} \cap B) = P(B) - P(A \cap B)$$

$$= \frac{120}{260} - \frac{50}{260} = \frac{70}{260} = \frac{7}{26}$$

(iv) The probability that the student participated in Kabadi only

$$P(A \cap \overline{B}) = P(A) - P(A \cap B)$$

$$= \frac{90}{260} - \frac{50}{260} = \frac{40}{260} = \frac{2}{13}$$

(v) The probability that the student participated in exactly one of the tournaments is

$P[(A \cap \overline{B}) \cup (\overline{A} \cap B)] = P(A \cap \overline{B}) + P(\overline{A} \cap B)$ [$\because A \cap \overline{B}, \overline{A} \cap B$ are mutually exclusive events]

$$= \frac{70}{260} + \frac{40}{260} = \frac{110}{260} = \frac{11}{26}$$

8.5 Conditional Probability

Consider the following situations:

(i) two events occur successively or one after the other (e.g) A occurs after B has occurred and

(ii) both event A and event B occur together.

Example 8.13

There are 4000 people living in a village including 1500 female. Among the people in the village, the age of 1000 people is above 25 years which includes 400 female. Suppose a person is chosen and you are told that the chosen person is a female. What is the probability that her age is above 25 years?

Ch8_11th Stats.indd 232 07/12/2021 12:11:22

Solution:

Here, the event of interest is selecting a female with age above 25 years. In connection with the occurrence of this event, the following two events must happen.

A: a person selected is female

B: a person chosen is above 25 years.

Situation1:

We are interested in the event B, given that A has occurred. This event can be denoted by $B|A$. It can be read as "B given A". It means that first the event A occurs then under that condition, B occurs. Here, we want to find the probability for the occurrence of $B|A$ i.e., $P(B|A)$. This probability is called conditional probability. In reverse, the probability for selecting a female given that a person has been selected with age above 25 years is denoted by $P(A|B)$.

Situation 2:

Suppose that it is interested to select a person who is both female and with age above 25 years. This event can be denoted by $A \cap B$.

Calculation of probabilities in these situations warrant us to have another theorem namely Multiplication theorem. It is derived based on the definition of conditional probability.

$$P(A) = P(\text{Selecting a female}) = \frac{1500}{4000}$$

$$P(A \cap B) = P(\text{Selecting a female with age above 25 years}) = \frac{400}{1500}$$

Hence, $$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{400}{1500} \times \frac{4000}{1500} = \frac{160}{225} = \frac{32}{45}.$$

8.5.1 Definition of Conditional of Probability

If $P(B) > 0$, the conditional probability of A given B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad .$$

If P(B) = 0, then $P(A \cap B) = 0$. Hence, the above formula is meaningless when $P(B) = 0$. Therefore, the conditional probability P(A|B) can be calculated only when P(B)>0.

The need for the computation of conditional probability is described in the following illustration.

Ch8_11th Stats.indd 233 07/12/2021 12:11:24

Illustration

A family is selected at random from the set of all families in a town with one twin pair. The sample space is

$$S = \{(\text{boy, boy}), (\text{boy, girl}), (\text{girl, boy}), (\text{girl, girl})\}.$$

Define the events

A: the randomly selected family has two boys, and

B: the randomly selected family has a boy.

Let us assume that all the families with one twin pair are equally likely. Since

$$A = \{(\text{boy, boy})\},$$

$$B = \{(\text{boy, boy}), (\text{boy, girl}), (\text{girl, boy})\},$$

$$A \cap B = A = \{(\text{boy, boy})\}.$$

Applying the classical definition of probability, it can be calculated that

$$P(B) = \frac{3}{4} \text{ and } P(A \cap B) = \frac{1}{4}.$$

Suppose that the randomly selected family has a boy. Then, the probability that the other child in the pair is a girl can be calculated using conditional probability as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/4}{3/4} = 1/3.$$

Example 8.14

A number is selected randomly from 11 through 19. Consider the events

$A = \{ 11,14, 16, 18, 19 \}$ $B = \{ 12, 14, 18, 19 \}$ $C = \{ 13, 15, 18, 19 \}$.

Find (i) P(A/B) (ii) P(A/C) (iii) P(B/C) (iv) P (B/A)

Solution:

given A $= \{ 11,14, 16, 18, 19 \}$ $B = \{ 12, 14, 18, 19 \}$ $C = \{ 13, 15, 18, 19 \}$.

$$A \cap B = \{ 14, 18, 19 \} \quad A \cap C = \{ 18, 19 \} = B \cap C$$

$$P(A) = \frac{5}{9} \qquad P(B) = \frac{4}{9} = P(C)$$

$$P(A \cap B) = \frac{3}{9} \qquad P(A \cap C) = \frac{2}{9} = P(B \cap C)$$

Ch8_11th Stats.indd 234 07/12/2021 12:11:25

Therefore, the probability for the occurrence of A given that B has occurred is

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{3}{9}}{\frac{4}{9}}$$

$$P(A/B) = \frac{3}{4}.$$

The probability for the occurrence of A given that C has occurred is

$$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{2}{9}}{\frac{4}{9}}$$

$$P(A/C) = 1/2.$$

Similarly, the conditional probability of B given C is

$$P(B/C) = \frac{P(B \cap C)}{P(C)} = \frac{\frac{2}{9}}{\frac{4}{9}}$$

$$P(B/C) = \frac{1}{2}$$

and the conditional probability of B given A is

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{3}{9}}{\frac{5}{9}}$$

$$P(B/A) = \frac{3}{5}.$$

Example 8.15

A pair of dice is rolled and the faces are noted. Let

A: sum of the faces is odd, B: sum of the faces exceeds 8, and

C: the faces are different then find (i) P (A/C) (ii) $P(B/C)$

Solution:

The outcomes favourable to the occurrence of these events are

A = { (1,2), (1,4), (1,6), (2,1), (2,3), (2,5), (3,2), (3,4), (3,6)
(4,1), (4,3), (4,5), (5,2), (5,4), (5,6), (6,1), (6,3), (6,5) }

B = { (3,6), (4,5), (4,6), (5,4), (5,5), (5,6), (6,3), (6,4), (6,5), (6,6) }

C = { (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,3), (2,4), (2,5), (2,6),
(3,1), (3,2), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,5), (4,6),
(5,1), (5,2), (5,3), (5,4), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5) }

Since A and B are proper subsets of C, $A \cap C = A$ and $B \cap C = B$.

Ch8_11th Stats.indd 235 07/12/2021 12:11:27

Here, $P(A) = \frac{18}{36} = \frac{1}{2}$

$P(B) = \frac{10}{36} = \frac{5}{9}$

$P(C) = \frac{30}{36} = \frac{5}{6}$.

Since, $A \cap C = A$. Hence,

$P(A \cap C) = P(A) = \frac{1}{2}$

$P(B \cap C) = P(B) = \frac{5}{9}$.

Hence, the probability for the sum of the faces is an odd number given that the faces are different is

$$\mathrm{P(A/C)} = \frac{P(A \cap C)}{P(C)} = \frac{\frac{1}{2}}{\frac{5}{6}} = \frac{3}{5}$$

Similarly, the probability for the sum of the faces exceeds 8 given that the faces are different is

$$\mathrm{P(B/C)} = \frac{P(B \cap C)}{P(C)} = \frac{\frac{2}{9}}{\frac{5}{6}} = \frac{4}{15}$$

8.5.2 Axioms

The conditional probabilities also satisfy the same axioms introduced in Section 8.3.

If S is the sample space of a random experiment and B is an event in the experiment, then

(i) $\mathrm{P(A/B)} \geq 0$ for any event A of S.

(ii) $\mathrm{P(S/B)} = 1$

(iii) If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i \mid B)$$

In continuation of conditional probability, another property of events, viz., independence can be studied. It is discussed in the next section. Also, multiplication theorem, a consequence of conditional probability, will be studied later.

8.6 Independent Events

For any two events A and B of a random experiment, if $P(A/B) = P(A)$, then knowledge of the event B does not change the probability for the occurrence of the event A. Such events are called independent events.

Ch8_11th Stats.indd 236 07/12/2021 12:11:29

If $P(A/B) = P(A)$, then

$$\frac{P(A \cap B)}{P(B)} = P(A).$$

$$\Rightarrow \quad P(A \cap B) = P(A) \times P(B).$$

Similarly, the relation $P(B/A) = P(B)$ also indicates the independence of the events A and B.

Definition :

Two events A and B are said to be independent of one another, if

$$P(A \cap B) = P(A) \times P(B).$$

NOTE

The greater the value of β_2, the more peaked the distribution.

(i) If A and B are not independent events, they are called dependent events.

(ii) The above definition of independence may be extended to finite number of events. If the events $A_1, A_2, \ldots, A_n$ satisfy

$$P(A_1 \cap A_2 \cap \ldots\ldots A_n) = P(A_1) \times P(A_2) \times \ldots \times P(A_n),$$

then $A_1, A_2, \ldots, A_n$ are called independent events.

Example 8.16

In tossing a fair coin twice, let the events A and B be defined as A: getting head on the first toss, B: getting head on the second toss. Prove that A and B are independent events.

Solution:

The sample space of this experiment is

$$S = \{HH, HT, TH, TT\}.$$

The unconditional probabilities of A and B are $P(A) = \frac{1}{2} = P(B)$.

The event of getting heads in both the tosses is represented by $A \cap B$. The outcome of the experiment in favour of the occurrence of this event is HH. Hence, $P(A \cap B) = \frac{1}{4}$.

$\therefore$ $P(A \cap B) = P(A) \times P(B)$ holds. Thus, the events A and B are independent events.

Ch8_11th Stats.indd 237 07/12/2021 12:11:30

Example 8.17

In the experiment of rolling a pair of dice, the events A, B and C are defined as A : getting 2 on the first die, B : getting 2 on the second die, and C : sum of the faces of dice is an even number. Prove that the events are pair wise independent but not mutually independent?

Solution:

$$S = \{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6),$$
$$(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),$$
$$(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),$$
$$(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),$$
$$(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),$$
$$(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \}$$
$$n(S) = 36$$

The outcomes which are favourable to the occurrence of these events can be listed below:

$$A = \{ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \}$$
$$n(A) = 6$$
$$P(A) = \frac{6}{36} = \frac{1}{6}$$
$$B = \{ (1,2), (2,2), (3,2), (4,2), (5,2), (6,2) \}$$
$$n(B) = 6$$
$$P(B) = \frac{6}{36} = \frac{1}{6}$$
$$C = \{ (1,1), (1,3), (1,5), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5),$$
$$(4,2), (4,4), (4,6), (5,1), (5,3), (5,5), (6,2), (6,4), (6,6) \}$$
$$n(C) = 18$$
$$P(C) = \frac{18}{36} = \frac{1}{2} .$$
$$A \cap B = \{ (2,2) \}$$
$$n(A \cap B) = 1$$
$$P(A \cap B) = \frac{1}{36}$$

Ch8_11th Stats.indd 238 07/12/2021 12:11:31

$$A \cap C = \{ (2,2), (2,4), (2,6) \}$$

$$n(A \cap C) = 3$$

$$P(A \cap C) = \frac{3}{36}$$

$$B \cap C = \{ (2,2), (4,2), (6,2) \}$$

$$n(B \cap C) = 3$$

$$P(B \cap C) = \frac{3}{36}$$

$$A \cap B \cap C = \{ (2,2) \}$$

$$n(A \cap B \cap C) = 1$$

$$P(A \cap B \cap C) = \frac{1}{36}$$

NOTE

If A and B are mutually exclusive events, $P(A \cap B) = 0$

The following relations may be obtained from these probabilities

$$P(A \cap B) = P(A) \times P(B)$$

$$\frac{1}{36} = \frac{1}{6} \times \frac{1}{6}$$

$$P(A \cap C) = P(A) \times P(C)$$

$$\frac{3}{36} = \frac{1}{6} \times \frac{1}{2}$$

$$P(B \cap C) = P(B) \times P(C)$$

$$\frac{3}{36} = \frac{1}{6} \times \frac{1}{2}$$

$$P(A \cap B \cap C) \neq P(A) \times P(B) \times P(C).$$

$$\frac{1}{36} \neq \frac{1}{6} \times \frac{1}{6} \times \frac{1}{2}$$

NOTE

If A and B are independent events, $P(A \cap B) = P(A) \times P(B)$

The above relations show that when the events A, B and C are considered in pairs, they are independent. But, when all the three events are considered together, they are not independent.

8.7 Multiplication Theorem on Probability

Theorem 8.5: (Multiplication Theorem of Probability)

If A and B are any two events of an experiment, then

$$P(A \cap B) = \begin{cases} P(A)\,P(B \mid A), & \text{if } P(A) > 0 \\ P(B)\,P(A \mid B), & \text{if } P(B) > 0 \end{cases}$$

Ch8_11th Stats.indd 239 07/12/2021 12:11:33

Proof:

If $P(B) > 0$, multiplying both sides of $P(A/B) = \frac{P(A \cap B)}{P(B)}$ by $P(B)$, we get

$P(A \cap B) = P(B)\, P(A/B)$

If $P(A) > 0$, interchanging A and B in $P(A/B) = \frac{P(A \cap B)}{P(B)}$, it also follows that

$P(A \cap B) = P(A)P(B/A)$.

Hence, the theorem is proved.

NOTE

If A and B are any two events of an experiment, then $P(A \cap B)$ can be calculated using the conditional probability of A given B or of B given A. Multiplication theorem of probability can be extended to compute $P(A \cap B \cap C)$ for the events A, B and C as follows:

$$P(A \cap B \cap C) = P(A)P(B/A)P(C/A \cap B).$$

Example 8.18

A box contains 7 red and 3 white marbles. Three marbles are drawn from the box one after the other without replacement. Find the probability of drawing three marbles in the alternate colours with the first marble being red.

Solution:

The event of interest is drawing the marbles in alternate colours with the first is red. This event can occur only when the marbles are drawn in the order (Red , White , Red)

If A and C represent the events of drawing red marbles respectively in the first and the third draws and B is the event of drawing white marble in the second draw, then the required event is $A \cap B \cap C$. The probability for the occurrence of $A \cap B \cap C$ can be calculated applying

$$P(A \cap B \cap C) = P(A)P(B/A)P(C/A \cap B)$$

Since there are 7 red and 3 white marbles in the box for the first draw,

$$P(A) = \frac{7}{10}$$

Now, there will be 6 red and 3 white marbles in the box for the second draw if the event A has occurred. Hence,

$$P(B/A) = \frac{3}{9}$$

Ch8_11th Stats.indd 240 07/12/2021 12:11:34

Similarly, there will be 6 red and 2 white marbles in the box for the third draw if the events A and B have occurred. Hence,

$P(C/A \cap B) = \frac{6}{8}$.

$\therefore \quad P(A \cap B \cap C) = \frac{7}{10} \times \frac{3}{9} \times \frac{6}{8} = \frac{7}{40}$ is the required probability of drawing three marbles in the alternate colours with the first marble being red.

Example 8.19

Three cards are drawn successively from a well-shuffled pack of 52 playing cards. Find the probability all three cards drawn successively is ace without replacing the card after each draw.

Solution:

Let A: all the three cards drawn are aces

At the first draw, there will be 4 aces among 52 cards. Having drawn an ace in the first draw, there will be 3 aces among 51 cards. Similarly, there will be 2 aces among 51 cards for the third draw.

Then, as discussed in Example 8.20, by Theorem 8.5

$$P(A) = \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50}$$

$$P(A) = \frac{1}{5525} .$$

Example 8.20

There are 13 boys and 6 girls in a class. Four students are selected randomly one after another from that class. Find the probability that: (i) all are girls, (ii) first two are boys and next are girls

Solution:

(i) Suppose that

B: all the randomly selected students are girls

There will be 6 girls among 19 students, in total, while selecting the first student; there will be 5 girls among 18 students, in total, while selecting the second student; 4 girls among 17 students, in total, while selecting the third student; and 3 girls among the remaining 16 students, in total, while selecting the fourth student.

Ch8_11th Stats.indd 241 07/12/2021 12:11:35

Then, by applying the Theorem-8.5 for simultaneous occurrence of these four events, it follows that

$$P(B) = \frac{6}{19} \times \frac{5}{18} \times \frac{4}{17} \times \frac{3}{16}$$

$$P(B) = \frac{5}{1292}.$$

(ii) Suppose that

C: In the randomly selected students the first two are boys and the next are girls

There will be 13 boys among the 19 students, in total, while selecting the first student; there will be 12 boys among 18 students, in total, while selecting the second student; 6 girls among 17 students, in total, while selecting the third student; and 5 girls among the remaining 16 students, in total, while selecting the fourth student.

Then, by applying the Theorem 8.5 for simultaneous occurrence of these four events, it follows that

$$P(C) = \frac{13}{19} \times \frac{12}{18} \times \frac{6}{17} \times \frac{5}{16} = \frac{65}{1292}$$

8.8 Bayes' Theorem and its Applications

In some cases, probability for the occurrence of an event of interest A may be difficult to compute from the given information. But, it may be possible to calculate its conditional probabilities $P(A/B)$ and $P(A/\bar{B})$ for some other event B of the same experiment. Then, $P(A)$ can be calculated applying the law of total probability. This theorem is a prelude for Bayes' theorem.

Theorem 8.6 (Law of Total Probability)

If $B_1, B_2, \ldots, B_n$ are mutually exclusive events such that $\bigcup_{j=1}^{n} B_j = S$ and $P(B_j) > 0$ for $j = 1, 2, \ldots, n$, Then for any event A

$$P(A) = P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \ldots + P(A/B_n)P(B_n).$$

In real life situations, decision making is an ongoing process. Situations may arise where we are interested in an event on an ongoing basis. Every time some new information may be available and based on this the probability of the event should be revised. This revision of probability with additional information is formalized in probability theory in the theorem known as Bayes' Theorem.

Ch8_11th Stats.indd 242 07/12/2021 12:11:37

Theorem 8.7 (Bayes' Theorem)

Let $B_1, \ldots, Bn$ be n mutually exclusive events such that where S is the sample space of the random experiment. If $P(B_j) > 0$ for $j = 1, 2, \ldots, n$, then for any event A of the same experiment with $P(A) > 0$,

$$P(B_j/A) = \frac{P(A/B_j)P(B_j)}{P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \ldots + P(A/B_n)P(B_n)}, \; j = 1, 2, \ldots, n.$$

[This Theorem is due to Rev. Thomas Bayes (1701-1761), an English philosopher and a priest. This work was published posthumously by his friend Richard Price during 1763 in the name of Bayes.]

Proof:

For each event B_j, $i = 1, 2, \ldots, n$, by the definition of conditional probability

$$P(B_j/A) = \frac{P(B_j \cap A)}{P(A)}$$

$$= \frac{P(A/B_j)P(B_j)}{P(A)} \quad \text{for } j = 1, 2, \ldots, n$$

Then, by the generalization of Theorem 8.5,

$$P(B_j/A) = \frac{P(A/B_j)P(B_j)}{P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \ldots + P(A/B_n)P(B_n)}, \; j = 1, 2, \ldots, n.$$

Example 8.21

Mr. Arivazhagan, Mr. Ilavarasan and Mr. Anbarasan attended an interview conducted for appointing a Physical Teacher in a school. Mr. Arivazhagan has 45% chance for selection, Mr. Ilavarasan has 28% chance and Mr.Anbarasan has 27% chance. Also, the chance for implementing monthly Mass Drill (MD) programme in the school is 42% if Mr.Arivazhagan is appointed; 40% if Mr.Ilavarasan is appointed; and 48% if Mr.Anbarasan is appointed.

Find the probability that the mass drill is implemented by if:

(i) Mr.Arivazhagan is appointed as the Physical Education Teacher.

(ii) Mr.Ilavarasan is appointed as the Physical Education Teacher.

(iii) Mr.Anbarasan is appointed as the Physical Education Teacher.

Ch8_11th Stats.indd 243 07/12/2021 12:11:38

Solution:

Let MD denote the event that the monthly Mass Drill programme is implemented in the school. Also, let

A: Mr.Arivazhagan is appointed

B: Mr.Ilavarasan is appointed

C: Mr.Anbarasan is appointed.S

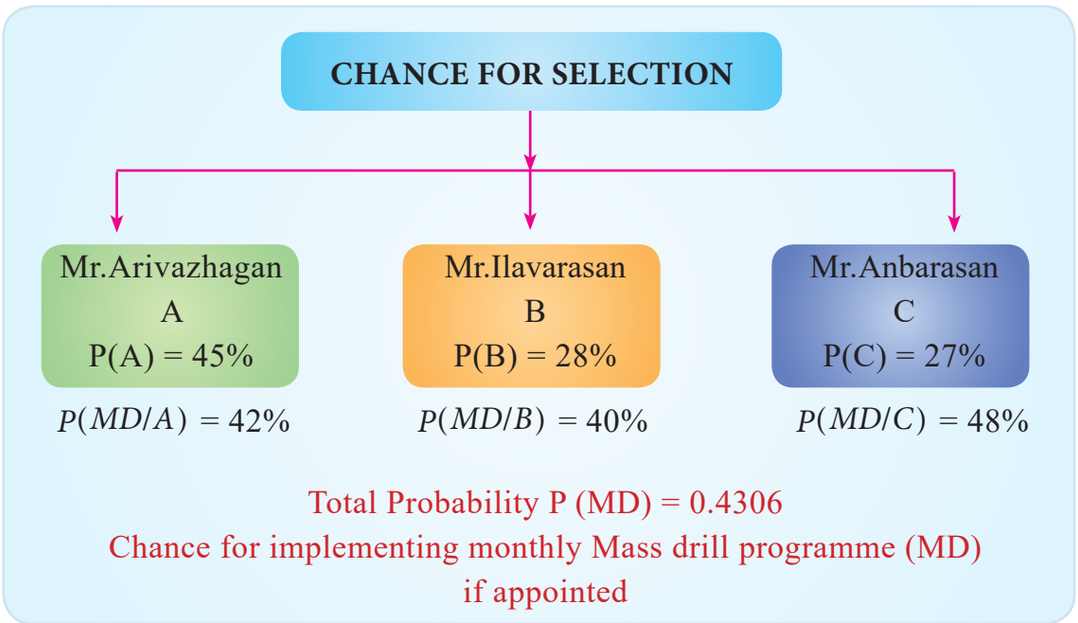


Fig. 8.10 Appointment of Physical Teacher

From fig. 8.10 the information given about these events are

$P(A) = 0.45$ $P(B) = 0.28$ $P(C) = 0.27$

$P(MD/A) = 0.42$ $P(MD/B) = 0.40$ $P(MD/C) = 0.48$

With these, the probability for implementing monthly Mass Drill programme in the school can be computed using total probability as

$$P(MD) = P(MD/A)P(A) + P(MD/B)P(B) + P(MD/C)P(C)$$

$$= (0.42 \times 0.45) + (0.40 \times 0.28) + (0.48 \times 0.27)$$

$$= 0.189 + 0.112 + 0.1296$$

$$\therefore \quad P(MD) = 0.4306.$$

If it is known that the monthly Mass Drill programme is implemented in the school,

Ch8_11th Stats.indd 244 07/12/2021 12:11:38

(i) the probability for Mr.Arivazhagan is appointed as the Physical Teacher can be calculated applying Theorem8.7 as

$$P(A/MD) = \frac{P(MD/A)P(A)}{P(MD/A)P(A)+P(MD/B)P(B)+P(MD/C)P(C)}$$

$$= \frac{0.42 \times 0.45}{(0.42 \times 0.45)+(0.40 \times 0.28)+(0.48 \times 0.27)}$$

$$= \frac{0.189}{0.4306}$$

$$\therefore \quad P(A/MD) = 0.4389.$$

(ii) the probability for Mr.Ilavarasan is appointed as the Physical Teacher can be calculated applying Theorem8.7 as

$$P(B/MD) = \frac{P(MD/B)P(B)}{P(MD/A)P(A)+P(MD/B)P(B)+P(MD/C)P(C)}$$

$$= \frac{0.40 \times 0.28}{(0.42 \times 0.45)+(0.40 \times 0.28)+(0.48 \times 0.27)}$$

$$\therefore \quad P(B/MD) = \frac{0.112}{0.4306}$$

$$= 0.2601.$$

(iii) the probability for Mr.Anbarasan is appointed as the Physical Teacher can be calculated applying Theorem8.7 as

$$P(C|MD) = \frac{P(MD/C)P(C)}{P(MD/A)P(A)+P(MD/B)P(B)+P(MD/C)P(C)}$$

$$= \frac{0.4 \times 0.27}{(0.42 \times 0.45)+(0.40 \times 0.28)+(0.28 \times 0.27)}$$

$$= \frac{0.1296}{0.4306}$$

$$P(C|MD) = 0.3010.$$

Example 8.22

Given three identical boxes I, II and III each containing two coins. In box I, both coins are gold coin, in box II, both are silver coins and in the box III, there is one gold and one silver coin. A person chooses a box at random and takes out a coin. If the coin is of gold, what is the probability that the other coin in the box is also of gold.

Solution:

In fig 8.11 given below if yellow colour denotes gold coin and grey colour denotes silver coin then:

Ch8_11th Stats.indd 245 07/12/2021 12:11:41

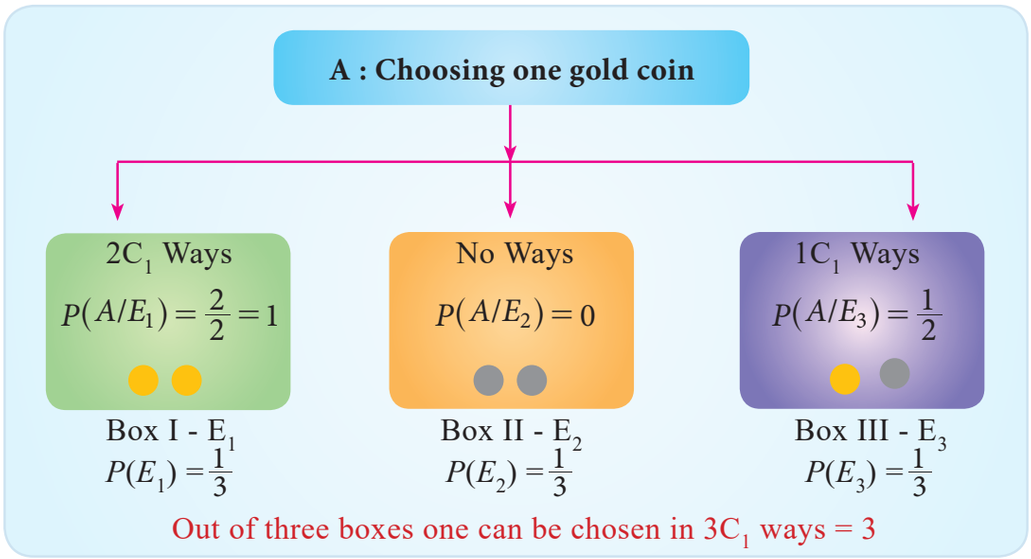


Fig. 8.11

If E_1, E_2 and E_3 be the events that the boxes I, II, III are chosen respectively.

Then $P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$

let A be the event that the gold coin is drawn

$P(A/E_1)$ = P (a gold coin from box 1) $= \frac{2}{2} = 1$

$P(A/E_2)$ = P (a gold coin from box 1I) $= \frac{0}{2} = 0$

$P(A/E_3)$ = P (a gold coin from box III) $= \frac{1}{2}$

Now, the probability that the other coin in the box is also gold is same as probability of choosing the box I and drawing a gold coin = $P(A/E_1)$

By Bayes' Theorem,

$$P(E_1/A) = \frac{P(E_1)P(A/E_1)}{\sum_{i=1}^{3} P(E_i)P(A/E_i)}$$

$$= \frac{\frac{1}{3} \times 1}{\frac{1}{3} \times 1 + \frac{1}{3} \times 0 + \frac{1}{3} \times \frac{1}{2}}$$

$$= \frac{2}{3}$$

Ch8_11th Stats.indd 246 07/12/2021 12:11:43

Points to Remember

- Probability of an event:

  (i) Classical Method: $P(A) = \frac{No.\ of\ outcomes\ in\ favour\ of\ A}{No.\ if\ elements\ in\ samplespace}$

  (ii) Relative Frequency Method: $P(A) = \underset{n \to \infty}{Lim} \frac{n(A)}{n}$

- Axioms

  (i) $P(A) \geq 0$

  (ii) $P(S) = 1$

  (iii) If $\{A_1, A_2, \ldots\}$ is a sequence of mutually exclusive events, then

  $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

- Probability of complementary event: $P(\bar{A}) = 1 - P(A)$

- Addition theorem of probability: $P(A \cup B) = P(A) + P(B)$, if A and B are mutually exclusive events

  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- If $A \subset B$, then $P(A) \leq P(B)$ and $P(B\text{-}A) = P(B) - P(A)$

- Conditional probability $P(A/B) = \frac{P(A \cap B)}{P(B)}, if\ P(B) > 0$

- A and B are independent events if $P(A \cap B) = P(A) \times P(B)$

- Multiplication theorem of probability $P(A \cap B) = \begin{cases} P(A/B) \times P(B), if\ P(B) > 0 \\ P(B/A) \times P(A), if\ P(A) > 0 \end{cases}$

- Bayes' theorem $P(B_j/A) = \frac{P(A/B_j) \times P(B_j)}{\sum_{i=1}^{n} P(A/B_j) \times P(B_j)}$, $j = 1, 2, \ldots n$

Ch8_11th Stats.indd 247 07/12/2021 12:11:45

EXERCISE 8

I Choose the best answer:

1. Which one of the following is not related to random experiment?

   (a) outcomes are predictable in advance

   (b) outcomes is unknown, in advance

   (c) experiment repeatable finite number of times

   (d) experiment is repeatable any number of times.



2. Mathematical probability may also be termed as

   (a) statistical probability (b) classical probability

   (c) Empirical probability (d) None of the above

3. In rolling of a die until 4 appears, the sample space is

   (a) a null set (b) a countable finite set

   (c) a countable infinite set (d) an uncountable set

4. A patient who has undergone a difficult surgery will survive a minimum of 12 years. Then, his survival time, x (in years), can be represented by

   (a) $\{x : 12 \leq x < \infty\}$ (b) $\{x : 12 \leq x < 24\}$

   (c) $\{x : 0 \leq x < 12\}$ (d) any interval along the real line.

5. If A and B are mutually exclusive events them $P(A \cup B)$ is equal to

   (a) $P(A) + P(B)$ (b) $P(A) - P(B)$

   (c) $P(A) + P(B) - P(A \cap B)$ (d) $P(A)P(B)$

6. If $A=\{1, 2\}$; $B=\{3, 4, 5\}$;$C=\{5, 6\}$ are events in S. Then the number sample point in S is

   (a)1 (b)4 (c) 3 (d) 6

7. Probability of not getting 3 when a die is thrown is

   (a) $\frac{1}{3}$ (b) $\frac{5}{6}$ (c) $\frac{1}{6}$ (d) $\frac{1}{4}$

8. If A_1 and A_2 are two events with $P(A_1)=\frac{4}{9}$ and $P(A_2)=\frac{3}{9}$ and $P(A_1 \cap A_2)=\frac{2}{9}$ then the probability that none of these two events occur is

   (a) $\frac{7}{9}$ (b) $\frac{4}{9}$ (c) $\frac{30}{81}$ (d) $\frac{2}{9}$

Ch8_11th Stats.indd 248 07/12/2021 12:11:47

9. If A and B are independent events with $P(A) = P(B)$; and $P(A \cap B) = a$; then $P(B)$ is

(a) $2a$ (b) $\sqrt{a}$ (c) $\frac{a}{2}$ (d) a^2

10. If A and B are two events with P(A/B)= 0.3, P(B/A) = 0.2. thenP(A) is

(a) $\frac{3}{10}$ (b) $\frac{7}{10}$ (c) $\frac{6}{7}$ (d) $\frac{1}{7}$

II. Fill in the blanks:

11. The probability of the entire sample space is _____

12. On throwing the single die, then the event of getting odd number or even number are _____ event.

13. Probability of getting a Monday in a week is _____

14. If A_1 and A_2 are independent events, then $P(A_1 \cup A_2) = p(A_1) +$ _____

15. If $A \subset B$, then $P(A)$___$P(B)$.

16. If probability for the occurrence of an event is 1, then the event is known as _____ event.

17. If $P(A) = 0$, then A is called _____ event.

18. Axiomatic approach to probability was proposed by _____

19. If A and B are two events with $P(A \cup B) = \frac{10}{15}$, then $P(\bar{A} \cap \bar{B})$ = _____

20. The conditional probability P(A/B) can be calculated using $P(B)$,if _____

III Very Short Answer Questions:

21. If E_1 and E_2 are two mutually exclusive events and Given that $P(E_2) = 0.5$ and $P(E_1 \cup E_2) = 0.7$ then find $P(E_1)$.

22. Find the probability that a leap year selected at random will contain 53 Fridays?

23. Two coins are tossed simultaneously. What is the probability of getting exactly two heads?

24. If $P(A \cap B) = 0.3$, $P(B) = 0.7$ find the value of $P(A/B)$

25. A box contains 5 red and 4 white marbles. Two marbles are drawn successively from the box without replacement and it is noted that the second one is white. What is the probability that the first is also white?

Ch8_11th Stats.indd 249 07/12/2021 12:11:48

26. Define:
(i) Event, (ii) Random Experiment, (iii) Sample space, (iv) Mutually exclusive events, (v) Exhaustive events (vi) Independent events, (vii) Dependent events.

27. Define mathematical probability.

28. Define statistical probability.

29. Define conditional probability.

30. State the Multiplication theorem on probability for any two events.

31. State the Multiplication theorem of probability for independent events.

IV. Short Answer Questions:

32. What is the chance that a non-leap year selected at random will contain 53 Sundays or 53 Mondays?

33. When two dice are thrown, find the probability of getting doublets. (same number on both dice).

34. A box containing 5 green balls land 3 red colour balls. Find the probability of selecting 3 green colour balls without replacement.

35. There are 5 items defective in a sample of 30 items. Find the probability that an item chosen at random from the sample is (i) defective (ii) non – defective

36. Given that $P(A) = 0.35$, $P(B) = 0.73$ and $P(A \cap B) = 0.14$, find $P(\overline{A} \cup \overline{B})$

37. State the axioms of probability.

38. State the theorem on total probability.

39. State Bayes' theorem.

40. A card is drawn at random from a well shuffled pack of 52 cards. What is the probability that it is (i) an ace (ii) a diamond card?

V Calculate the following:

41. State and prove addition theorem on probability.

42. An urn contains 5 red and 7 green balls. Another urn contains 6 red and 9 green balls. If a ball is drawn from any one of the two urns, find the probability that the ball drawn is green.

Ch8_11th Stats.indd 250 07/12/2021 12:11:48

43. In a railway reservation office, two clerks are engaged in checking reservation forms. On an average, the first clerk (A_1) checks 55% of the forms, while the second clerk (A_2) checks the remaining. A_1 has an error rate of 0.03 and A_2 has an error rate of 0.02. A reservation form is selected at random from the total number of forms checked during the day and is discovered to have some errors. Find the probability that the form is checked by A_1 and A_2 respectively.

44. In a university, 30% of the students are doing a course in statistics use the book authored by A_1, 45% use the book authored by A_2 and 25% use the book authored by A_3. The proportion of the students who learnt about each of these books through their teacher are $P(A_1)$ =0.50, $P(A_2)$=0.30, and $P(A_3)$=0.20. One of the student selected at random revealed that he learned the book he is using through their teachers. Find the probability that the book used it was authored by A_1, A_2, and A_3 respectively.

45. A bolt manufacturing company has four machines *A*, *B*, *C* and *D* producing 20%, 15%, 25% and 40% of the total output respectively. 5%, 4%, 3% and 2% of their output (in the same order) are defective bolts. A bolt is chosen at random from the factory and is found defective what is the probability of getting a defective bolt.

46. A city is partitioned into districts *A, B, C* having 20 percent, 40 percent and 40 percent of the registered voters respectively. The registered voters listed as Democrats are 50 percent in A, 25 percent in B and 75 percent in C. A registered voter is chosen randomly in the city. Find the probability that the voter is a listed democrat.

ANSWERS

I. 1. (*a*) 2. (*b*) 3. (c) 4. (a) 5. (a) 6. (d) 7. (b) 8. (b) 9. (b) 10. (c)

II. 11. one 12. mutually exclusive 13. $\frac{1}{7}$ 14. $P(A_1) + P(A_2)$

15. less than or equal to 16. sure event 17. impossible event

18. A.N. Kolmogorov 19. $\frac{1}{3}$ 20. greater than

III. 21. 0.2 22. $\frac{2}{7}$ 23. $\frac{1}{2}$ 24. $\frac{3}{7}$ 25. $\frac{4}{9}$

IV. 32. $\frac{2}{7}$ 33. $\frac{1}{6}$ 34. $\frac{5}{28}$ 35. (i) $\frac{1}{6}$ (ii) $\frac{5}{6}$ 36. 0.86 40. (i) $\frac{1}{3}$ (ii) $\frac{1}{4}$

V. 42. $\frac{71}{120}$ 43. 0.647, 0.353 44. 0.45, 0.40, 0.15 45. $\frac{63}{2000}$ 46. 50%

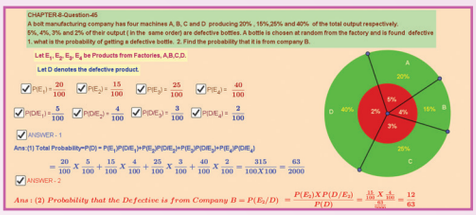
Ch8_11th Stats.indd 251 07/12/2021 12:11:51



ICT CORNER

ELEMENTARY PROBABILITY THEORY-BAYES THEOREM PROBLEM

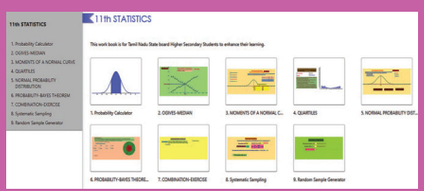
This activity is for conditional probability based Bayes theorem.
This helps to understand the problem.



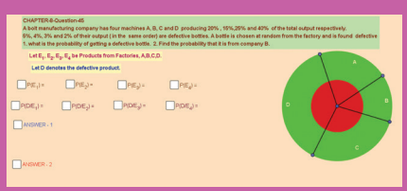
Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for Statistics are given, open the worksheet named "**Probability-Bayes theorem**"
- A text book exercise problem is explained step by step. Read the question carefully and compare with the diagram on right-hand side to assign the names as given and observe the check box names.
- Now click on the check boxes for separate probabilities and conditional probabilities. Now try to understand what conditional probability is, by observing the diagram.
- Now recall the formula for total probability **P(D)** and Bayes formula **$P(E_2/D)$** and work out.
- Now click on "**Answer-1**" and "**Answer-2**" and check your answer.

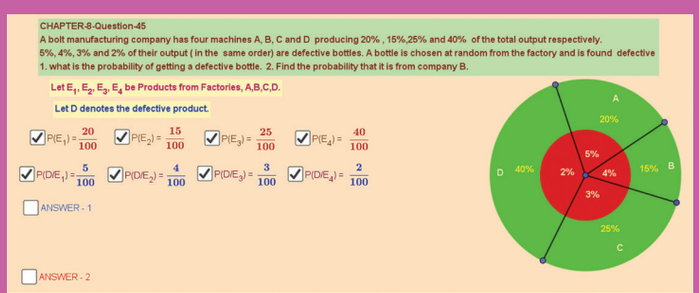
Step-1



Step-2



Step-3



Pictures are indicatives only*

URL:

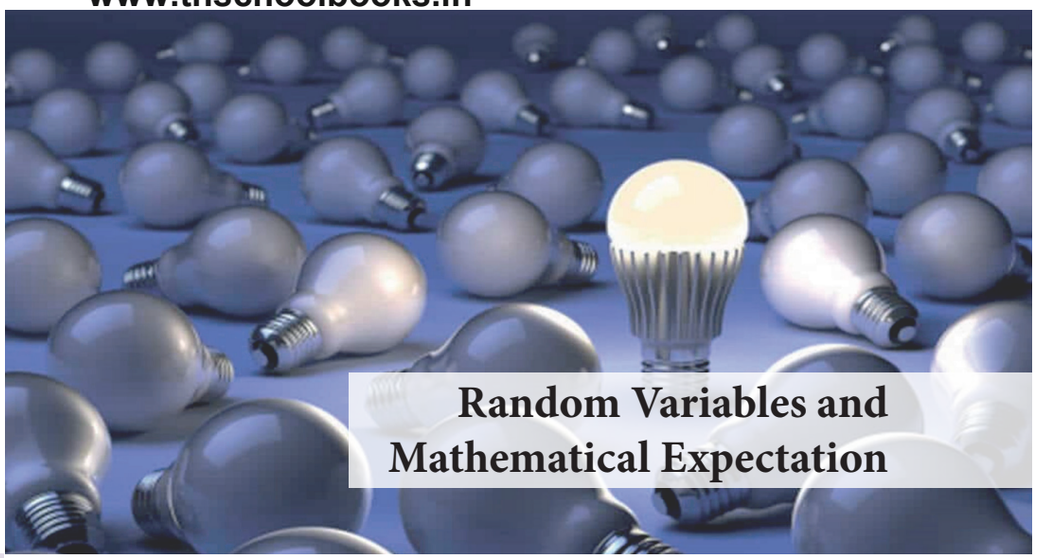
https://ggbm.at/uqVhSJWZ

Ch8_11th Stats.indd 252 07/12/2021 12:11:53

www.tnschoolbooks.in

Chapter

9



Random Variables and Mathematical Expectation



Siméon-Denis Poisson
(21 June, 1781 – 25 April, 1840)

Siméon-Denis Poisson, a French mathematician known for his work on definite integrals, electromagnetic theory, and probability.

Poisson's research was mainly on Probability. Poisson distribution was one of the important invention of Poisson. Poisson contributed to the law of large numbers and it was useful for the approximation of binomial distribution. Poisson distribution is now fundamental in the analysis of problems concerning radioactivity, traffic, and the random occurrence of events in time or space.

'Everything existing in the universe is the fruit of Chance'.

- Democritus

Learning Objectives

- Knows the relevance of Random Variables.
- Understands the types of Random Variables.
- Differentiates the various Types of probability functions.
- Enumerates the functions and properties of Cumulative Distribution.
- Understands the Concept of expectation.
- Applies the Theorems on Expectations.
- Identifies Moment generating functions and Characteristic Functions

Ch9_11th Stats.indd 253 07/12/2021 12:13:24

Introduction

We have studied the elementary probability in the previous chapter. As an application of probability, there are two more concepts namely random variables and probability distributions. Before seeing the definition of probability distribution, random variable needs to be explained. It has been a general notion that if an experiment is repeated under identical conditions, values of the variable so obtained would be similar. However, there are situations where these observations vary even though the experiment is repeated under identical conditions. As the result, the outcomes of the variable are unpredictable and the experiments become random.

We have already learnt about random experiments and formation of sample spaces. In a random experiment, we are more interested in, x number associated with the outcomes in the sample space rather than the individual outcomes. These numbers vary with different outcomes of the experiment. Hence it is a variable. That is, this value is associated with the outcome of the random experiment. To deal with such situation we need a special type of variable called random variable.

9.1 Definition of random variable

Definition

Let S be the sample space of a random experiment. A rule that assigns a single real number to each outcome (sample point) of the random experiment is called random variable.

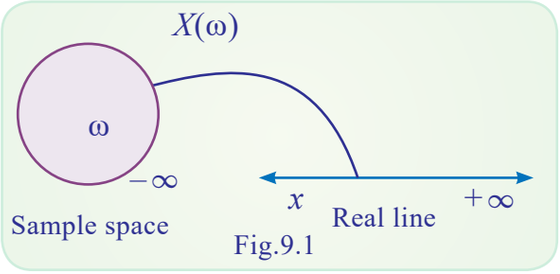


Fig.9.1

In other words, a random variable is a real valued function defined on a sample space S that is with each outcome ω of a random experiment there corresponds a unique real value x known as a value of the random variable X. That is $X(\omega) = x$.

Generally random variables are denoted by upper case alphabets like X, Y, Z … and their values or realizations are denoted by the corresponding lower case letters. For example, if X is a random variable, the realizations are x_1, x_2 …

For example,

(i) Consider the random experiment of rolling a die.

The sample space of the experiment is $S=\{1, 2, 3, 4, 5, 6\}$

Let X denotes the face of the die appears on top.

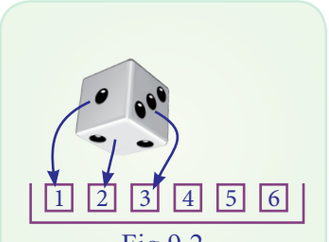


Fig.9.2

Ch9_11th Stats.indd 254 07/12/2021 12:13:24

The assigning rule is

$X(1) = 1$, $X(2) = 2$, $X(3) = 3$, $X(4)=4$, $X(5)=5$ and $X(6)=6$

Hence the values taken by the random variable X are 1,2,3,4,5,6. These values are also called the realization of the random variable X.

(ii) Random experiment : Two coins are tossed simultaneously.

Sample space : $S=\{HH, HT, TH, TT\}$

Assigning rule : Let X be a random variable defined as the number of heads comes up.

Sample Point ω	HH	HT	TH	TT
X (ω)	2	1	1	0

Here, the random variable X takes the values 0, 1, 2.

(iii) Experiment : Two dice are rolled simultaneously.

Sample space : $\{(1, 1),(1, 2),(1, 3),\ldots(6, 6)\}$

Assigning rule : Let X denote the sum of the numbers on the faces of dice

then $X_{ij} = i + j$, Here, i denotes face number on the first die and j denotes the face number on the second die.

Then X is a random variable which takes the values 2, 3, 4 . …. 12.

That is the range of X is {2, 3, 4…… 12}

9.2 Discrete and Continuous random variables

Random variables are generally classified into two types, based on the values they take such as Discrete random variable and Continuous random variable.

9.2.1 Discrete random variable

A random variable is said to be discrete if it takes only a finite or countable infinite number of values.

Ch9_11th Stats.indd 255 07/12/2021 12:13:25

For example,

(i) Consider the experiment of tossing a coin

If X (Head) =1, X (Tail) = 0

Then X takes the values either 0 or 1

This is a discrete random variable.

NOTE

Example 9.1, 9.2, and 9.3 are the random variables taking finite number of values, therefore they are discrete.

(ii) Consider the experiment of tossing a coin till head appears.

Let random variable X denote the number of trials needed to get a head. The values taken by it will be 1, 2, 3, ..

It is discrete random variable taking countable infinite values.

9.2.2 Continuous random variable:

A random variable X is said to be continuous, if it takes values in an interval or union of disjoint intervals. (A rigorous definition is beyond the scope of the book).

For example,

(i) If X is defined as the height of students in a school ranging between 120 cms and 180 cms, Then the random variable X is $\{x/120\ cms < x < 180\ \text{cms}\}$ is a continuous random variable.

(ii) Let the maximum life of electric bulbs is 1500 hrs. Life time of the electric bulb is the continuous random variables and it is written as $X = \{x/0 \leq x \leq 1500\}$

9.3 Probability mass function and probability density function

A probability function is associated with each value of the random variable. This function is used to compute probabilities for events associated with the random variables. The probability function defined for a discrete random variable is called probability mass function. The probability function associated with continuous random variable is called probability density function.

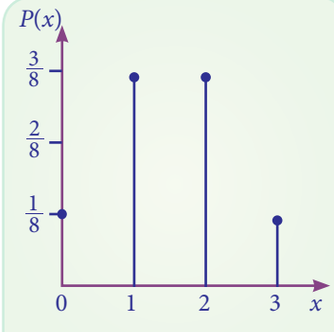


Fig.9.3

Ch9_11th Stats.indd 256 07/12/2021 12:13:25

9.3.1 Probability Mass Function.

If, X is a discrete random variable taking values $x_1, x_2 \ldots. x_n$ with respective probabilities $p(x_1), p(x_2) \ldots.. p(x_n)$ such that

(i) $p(x_i) \geq 0$, $\forall$ i (non- negative) and (ii) $\sum_{i=1}^{n} p(x_i) = 1$ then $p(x)$ is known as the probability mass function (p.m.f) of the discrete random variable X.

The pair $\{x_i, p(x_i); i = 1, 2, 3, \ldots\}$ is known as prabability distribution of X.

Example 9.1

A coin is tossed two times. If X is the number of heads, find the probability mass function of X.

Solution:

Since the coin is tossed two times, the sample space is S={HH, HT, TH, TT}

If X denotes the numbers of heads, the possible values of X are 0,1,2 with the following

$P(X = 0) = P$ (getting no head) $= \frac{1}{4}$

$P(X = 1) = P$ (getting one head) $= \frac{2}{4} = \frac{1}{2}$

$P(X = 2) = P$ (getting two heads) $= \frac{1}{4}$

NOTE

Probabilities are non negative and the total is 1.

The probability distribution of X is

X	0	1	2
$p(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Example 9.2

In example 9.3 the probability mass function of X is given in the following table

X	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The above table may be called as the probability distribution function of X.

9.3.2 Probability Density Function

When the random variable is continuous in the co-domain it is spread over it. A function $f(x)$ is defined on real line and satisfying the following conditions :

Ch9_11th Stats.indd 257 07/12/2021 12:13:28

(i) $f(x) \geq 0$, $\forall x$ (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$ is called the probability density function (p.d.f) of X.

Remark:

(i) Every integrable function satisfying the above two conditions is a probability density function of some random variable X

(ii) The probability that the value of X lies between two points 'a' and 'b' is $P(a < X < b) = \int_a^b f(x)dx$

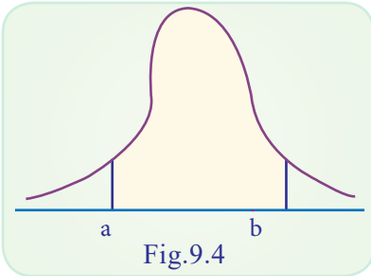


Fig.9.4

(iii) If X is discrete random variable then for any real x, $P(X = x)$ need not be zero.However in the case of continuous random variable $P(X = x) = 0$ holds always. $P(X - a) = \int_a^a f(x)dx = 0$

(iv) If X is a continuous random variable then for any $a < b$ $P(a < X < b) = p(a \leq X < b) = p(a < X \leq b) = p(a \leq X \leq b)$

Example 9.3

A continuous random variable X has probability density function given by $f(x) = \begin{cases} Ax^3, & 0 < x < 1 \\ 0, & otherwise \end{cases}$. Find A.

Solution:

Since $f(x)$ is a p.d.f

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_0^1 Ax^3 dx = 1$$

$$\Rightarrow \quad \frac{A}{4} = 1$$

$$\Rightarrow \quad A = 4$$

NOTE

Total probability of continuous random variable is 1 within the certain interval.

Example 9.4

Verify whether the following function is a probability density function

$$f(x) = \begin{cases} \frac{2x}{9}, & 0 < x < 3 \\ 0, & elsewhere \end{cases}$$

Solution :

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^3 \frac{2x}{9} dx$$

Ch9_11th Stats.indd 258 07/12/2021 12:13:30

$$\Rightarrow \qquad \frac{2}{9}\left[\frac{9}{2}\right] = 1$$

It is to be noted that (i) $f(x) \geq 0, \forall x$ (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

Hence, $f(x)$ is a p.d.f.

9.4 Distribution function and its properties

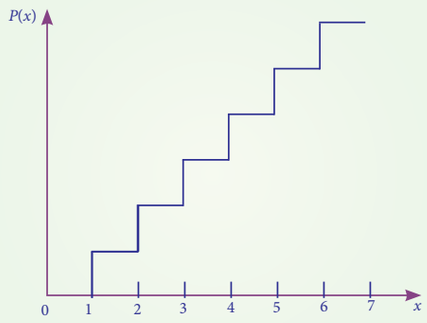


Fig.9.5

We get the probability of a given event at a particular point. If we want to have the probability upto the point we get the probability $P(X \leq x)$. This type of probability is known as probability mass function. We can also find how the probability is distributed within certain limits. [$P(X < x)$ or $P(X > x)$ or $P(a < x < b)$].

9.4.1 Distribution Function for discrete random variable

Definition: Let X be a random variable ,the cumulative distribution function (c.d.f) of a random variable X is defined as $F(x) = P(X \leq x), \forall x$. It is called simply as distribution function.

9.4.2 Properties:

(i) $0 \leq F(x) \leq 1, \ \forall x$, (non –negative)

(ii) $F(x) \leq F(y), \ \forall x < y$, (non –decreasing)

(iii) $\lim_{h \to 0} F(x+h) = F(x), \ \forall x$, ($F(x)$ is right continuous)

(iv) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$

(v) $F(\infty) = \lim_{x \to \infty} F(x) = 1$

NOTE

In the case of discrete random variable X

(i) $F(x) = \sum_{r=-\infty}^{x} P(x = r)$

(ii) $P(X = x) = F(x) - F(x-1)$

9.4.3 Distribution Function for continuous random variable

(i) $F(x) = \int_{-\infty}^{x} f(x)dx$

(ii) $f(x) = F'(x)$

(iii) $P(a < X < b) = P(a \leq X < b)$

$= P(a < X \leq b) = P(a \leq X \leq b)$

$= \int_{a}^{b} f(x)dx$

That is pdf can be obtained from distribution function by differentiating the distribution function and the distribution function can be obtained from pdf by integrating the pdf over the given range

Ch9_11th Stats.indd 259 07/12/2021 12:13:33

Properties

(i) $F(x)$ is a non decreasing function of x

(ii) $0<F(x)<1 \qquad -\infty < x < \infty$

(iii) $F(-\infty) = 0,$

(iv) $F(\infty) = 1$

(v) For any real constant a and b such that $a < b,\ p(\ a < X \leq b) = F(b) - F(a)$

(vi) $f(x) = \frac{d}{dx}(F(x)) \qquad$ i.e., $f(x) = F'(x)$

Example 9.5

A random variable X has the following probability mass function

X	0	1	2	3	4	5	6
$P(X = x)$	a	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$

(i) Find the value of 'a'

(ii) Find the c.d.f $F(x)$ of X

(iii) Evaluate : (a) $P(X \geq 4)$ (b) $P(X < 5)$ (c) $P(3 \leq X \leq 6)$

(iv) $P(X = 5)$ using $F(x)$

Solution:

(i) Since $P(X = x)$ is probability mass function $\sum P(X = x) = 1$

i.e., $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = 1$

$$a + 3a + 5a + 7a + 9a + 11a + 13a = 1$$

$$49a = 1 \Rightarrow a = \frac{1}{49}$$

(ii) Hence the c.d.f is

X	0	1	2	3	4	5	6
$P(x)$	$\frac{1}{49}$	$\frac{3}{49}$	$\frac{5}{49}$	$\frac{7}{49}$	$\frac{9}{49}$	$\frac{11}{49}$	$\frac{13}{49}$
$F(x)$	$\frac{1}{49}$	$\frac{4}{49}$	$\frac{9}{49}$	$\frac{16}{49}$	$\frac{25}{49}$	$\frac{36}{49}$	$\frac{49}{49} = 1$

Ch9_11th Stats.indd 260 07/12/2021 12:13:36

(iii) (a) $P(X \geq 4)$

$$= P(X = 4) + P(X = 5) + P(X = 6)$$

$$= 9a + 11a + 13a$$

$$= 33a$$

$$= 33 \times \frac{1}{49} = \frac{33}{49}$$

(b) $P(X<5)$ $= 1 - P(X \geq 5) = 1 - [P(X = 5) + P(X = 6)]$

$$= 1 - [11a + 13a]$$

$$= 1 - 24a$$

$$= 1 - \frac{24}{29}$$

$$= \frac{25}{29}$$

(c) $P(3 \leq \times \leq 6)$ $= P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6)$

$$= 7a + 9a + 11a + 13a$$

$$= 40$$

$$a = \frac{40}{49}$$

(iv) $P(X = 5)$ $= F(5) - F(5 - 1)$

$$= \frac{36}{49} - \frac{25}{49}$$

$$= \frac{11}{49}.$$

Example 9.6

Let X be a random variable with p.d.f

$$f(x) = \begin{cases} \frac{x}{2}; & 0 < x < 2 \\ 0 \ ; & otherwise \end{cases}$$

(i) Find the c.d.f of X , (ii) Compute $P\left(\frac{1}{2} < X \leq 1\right)$, (iii) Compute $P(X=1.5)$

Solution:

(i)The c.d.f of X : $F(x) = P(X \leq x)$

$$= \int_{-\infty}^{x} f(x)dx$$

Ch9_11th Stats.indd 261 07/12/2021 12:13:38

$$= \int_0^x \frac{x}{2}dx = \frac{x^2}{4}$$

$$\text{Hence, } F(x) = \begin{cases} 0 & if\ x < 0 \\ \frac{x^2}{4} & if\ 0 < x < 2 \\ 1 & if\ x \geq 2 \end{cases}$$

(ii) $P\left(\frac{1}{2} < X \leq 1\right) = F(1) - F\left(\frac{1}{2}\right)$

$$= \frac{1}{4} - \frac{1}{16} = \frac{3}{16}$$

This probability can be computed using p.d.f

$$\int_{\frac{1}{2}}^{1} \frac{x}{2}dx = \frac{1}{2}\left(\frac{x^2}{2}\right)_{\frac{1}{2}}^{1}$$

$$= \frac{3}{16}$$

(iii) $P\ (X = 1.5) = 0$

NOTE

For a continuous random variable probability at a particular point is 0

9.5 Joint and marginal probability mass functions

In real life situations we may be observed two or more random variables on the individuals simultaneously. For instance, blood pressure and cholesterol for each individual are measured simultaneously. In such cases we require the concept of bi-variate random variable represented by (X, Y), where X and Y are univariate random variables.

9.5.1 Definition (Joint p.m.f)

Let (X, Y) be a discrete bivariate random variable. Then $p\ (x, y)$ is called the joint probability mass function of (X, Y) if the following conditions are satisfied.

$p\ (x, y) \geq 0 \quad \forall\ x, y$

$\sum_{x,y} p(x,y) = 1$

NOTE

$p\ (x, y) = p\ (X = x, Y = y)$

Definition (Marginal Probability Mass Function)

Given a joint probability mass function $p\ (x, y)$, then $p\ (x) = \sum_y p(x,y)$ is called marginal probability mass function of X. Similarly $p\ (y) = \sum_x p(x,y)$ is called the marginal probability mass function of Y.

For example,

There are 10 tickets in a bag which are numbered 1, 2, 3, ...10. Two tickets are drawn at random one after the other with replacement.

Ch9_11th Stats.indd 262 07/12/2021 12:13:40

Here, random variable X denotes the number on the first ticket and random variable Y denotes the number on the second ticket.

9.5.2 Joint and marginal probability density functions

As we defined in section 9.5.1 the joint probability mass function, we define the joint probability density function.

Definition:

Let (X, Y) be a bivariate continuous random variables. The function $f(x, y)$ is called a bivariate probability density if the following conditions are satisfied.

(i) $f(x, y) \geq 0 \qquad \forall\ x, y$

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dx\,dy = 1$

The marginal probability density function of X is given by

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

and the marginal probability density function of Y is given by

$$h(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Example 9.7

Prove that the bivariate function given by $f(x, y) = \begin{cases} \frac{1}{8}(x+y) & , \quad 0 < x, y \leq 2 \\ 0 & , \quad otherwise \end{cases}$

Proof:

If f is a probability density function $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dx\,dy = 1$

$$= \frac{1}{8}\int_0^2\int_0^2 (x+y)dx\,dy$$

$$= \frac{1}{8}\left(\int_0^2\int_0^2 xdx\,dy + \int_0^2\int_0^2 ydx\,dy\right)$$

$$= \frac{1}{8}\left\{\int_0^2\left[\int_0^2 xdx\right]dy + \int_0^2\left[\int_0^2 ydy\right]dx\right\}$$

$$= \frac{1}{8}\left[\int_0^2 2dy + \int_0^2 2dx\right]$$

Ch9_11th Stats.indd 263 07/12/2021 12:13:42

$$= \frac{1}{8}\left[(2y)_0^2 + (2x)_0^2\right]$$

$$= \frac{1}{8}[(4-0)+(4-0)]$$

$$= \frac{1}{8} \times 8 = 1$$

Therefore, $f(x, y)$ is a probability density function.

Example 9.8

Joint p.d.f. of X, Y is $f(x, y) = \frac{3}{2}\begin{cases} x^2 y & , \ 0 \le x \le 1, 0 \le y \le 2 \\ 0 & , \ elsewhere \end{cases}$ then find the marginal density function of X and Y.

Solution:

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

$$f(x) = \frac{3}{2}\int_0^2 x^2 y dy$$

$$= \frac{3}{2}x^2\left[\frac{y^2}{2}\right]_0^2 = \frac{3}{2}x^2\left[\frac{4}{2} - \frac{0}{2}\right]$$

$$= \frac{3}{2}x^2 \times 2 = 3x^2$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

$$f(y) = \frac{3}{2}\int_0^1 x^2 y dy$$

$$= \frac{3}{2}y\left(\frac{x^3}{3}\right)_0^1 = \frac{3}{2}y\left(\frac{1}{3}\right) = \frac{y}{2}$$

Marginal density function of X

$$f(x) = \begin{cases} 3x^2 & , \ 0 \le x \le 1, \\ 0 & , \ elsewhere \end{cases}$$

Marginal density function of Y

$$f(y) = \begin{cases} \frac{y}{2} & , \ 0 \le y \le 2, \\ 0 & , \ elsewhere \end{cases}$$

Example 9.9

Joint p.d.f. of X, Y is $f(x, y) = \begin{cases} \frac{1}{10}(4x - 2y) & , \ 1 \le x \le 3, 1 \le y \le 2 \\ 0 & , \ elsewhere \end{cases}$. Find the marginal density function of X and Y

Ch9_11th Stats.indd 264 07/12/2021 12:13:48

Solution:

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$$

$$f(x) = \frac{1}{10}\int_{1}^{2} (4x - 2y)\,dy$$

$$= \frac{1}{10}\left[4xy - \frac{2y^2}{2}\right]_1^2$$

$$= \frac{1}{10}[4x(2-1) - (4-1)] = \frac{1}{10}(4x - 3)$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$$

$$f(y) = \frac{1}{10}\int_{1}^{3} (4x - 2y)\,dx$$

$$= \frac{1}{10}\left(\frac{4x^2}{2} - 2xy\right)_1^3$$

$$= \frac{1}{10}\left(2x^2 - 2xy\right)_1^3$$

$$= \frac{1}{10}[2(9-1) - 2y(3-1)]$$

$$= \frac{1}{10}(2 \times 8 - 6y + 2y)$$

$$= \frac{1}{10}(16 - 4y) = \frac{1}{5}(8 - 2y)$$

Marginal density function of X

$$f(x) = \begin{cases} \frac{1}{10}(4x-3) & , 1 \le x \le 3 \\ 0 & , \textit{elsewhere} \end{cases}$$

Marginal density function of Y

$$f(x, y) = \begin{cases} \frac{1}{5}(8-2y) & , 1 \le y \le 2 \\ 0 & , \textit{elsewhere} \end{cases}$$

9.6 Mathematical expectation

Probability distribution gives us an idea about the likely value of a random variable and the probability of the various events related to random variable. Even though it is necessary for us to explain probabilities using central tendencies, dispersion, symmetry and kurtosis. These are called descriptive measures and summary measures. Like frequency

Ch9_11th Stats.indd 265 07/12/2021 12:13:52

distribution we have to see the properties of probability distribution. This section focuses on how to calculate these summary measures. These measures can be calculated using

(i) Mathematical Expectation and variance.

(ii) Moment Generating Function .

(iii) Characteristic Function.

9.6.1 Expectation of Discrete random variable

Let X be a discrete random variable which takes the values $x_1, x_2 \ldots x_n$ with respective probabilities $p_1, p_2 \ldots p_n$ then mathematical expectation of X denoted by $E(X)$ is defined by

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$$

$$= \sum_{i=1}^{n} x_i p_i \qquad \text{where } \sum p_i = 1$$

Sometimes $E(X)$ is known as the mean of the random vairable X.

Result:

If $g(X)$ is a function of the random variable X, then $E\ g(X) = \sum g(x) p(X = x)$

Properties:

(i) $E(c) = c$ where c is a constant

Proof :

$$E(X) = \sum x_i p_i$$

$$E(c) = \sum c p_i = c \sum p_i = c \times 1 = c$$

(ii) $E(cX) = cE(X)$, where c is a constant

Proof:

$$E(X) = \sum x_i p_i$$

$$E(X) = \sum c x_i p_i$$

$$= c \sum x_i p_i$$

$$= c\ E(X).$$

(iii) $E(aX+b) = aE(X)+b$

Ch9_11th Stats.indd 266 07/12/2021 12:13:54

Variance of Discrete random variable

Definition: In a probability distribution Variance is the average of sum of squares of deviations from the mean. The variance of the random variable X can be defined as.

$$\text{Var}(X) = E(X - E(X))^2$$

$$= E(X^2) - (E(X))^2$$

Example 9.10

When a die is thrown X denotes the number turns up. Find $E(X)$, $E(X^2)$ and $\text{Var}(X)$.

Solution:

Let X denoted that number turns up in a die.

X takes the values 1, 2, 3, 4, 5, 6 with probabilities $\frac{1}{6}$ for each.

Therefore the probability distribution is

X	1	2	3	4	5	6
P(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = \Sigma x_i p_i$$

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_6 p_6$$

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \ldots + 6 \times \frac{1}{6}$$

$$= \frac{1}{6}(1+2+3+4+5+6)$$

$$= \frac{7}{2}$$

$$E(X^2) = \Sigma x_i^{\,2} p_i$$

$$E(X^2) = x_1^{\,2} p_1 + x_2^{\,2} p_2 + \ldots + x_6^{\,2} p_6$$

$$E(X^2) = \left(1^{2\times}\frac{1}{6}\right) + \left(2^{2\times}\frac{1}{6}\right) + \ldots \left(6^{2\times}\frac{1}{6}\right)$$

$$= \frac{1}{6}(1+4+9+16+25+36)$$

$$= \frac{1}{6}(91) = \frac{91}{6}$$

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2$$

Ch9_11th Stats.indd 267 07/12/2021 12:13:55

$$= \frac{91}{6} - \frac{49}{4}$$

$$\text{Var}(X) = \frac{35}{12}$$

Example 9.11

The mean and standard deviation of a random variable X are 5 and 4 respectively Find $E(X^2)$

Solution:

Given $E(X) = 5$ and $\sigma = 4$

$\therefore$ $\text{Var}(X) = 16$

But, $\text{Var}(X) = E(X^2) - [E(X)]^2$

$16 = E(X^2) - (5)^2$

$E(X^2) = 25 + 16 = 41$

Example 9.12

A player tosses two coins, if two head appears he wins ₹ 4, if one head appears he wins ₹ 2, but if two tails appears he loses ₹ 3. Find the expected sum of money he wins?

Solution:

Let X be the random variable denoted the amount he wins.

The possible values of X are 4, 2 and –3

$P(X = 4) =$ P(getting two heads) $= \frac{1}{4}$

$P(X = 2) =$ P(getting one head) $= \frac{1}{2}$

$P(X = 3) =$ P(getting No heads) $= \frac{1}{4}$

Probability distribution is

X	4	2	-3
$P(\mathbf{X}=x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

$$E(X) = \Sigma x_i p_i$$

$$E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3$$

$$E(X) = 4 \times \frac{1}{4} + 2 \times \frac{1}{2} - 3 \times \frac{1}{4} = 1.25$$

Ch9_11th Stats.indd 268 07/12/2021 12:13:56

Example 9.13

Let X be a discrete random variable with the following probability distribution

X	-3	6	9
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Find mean and variance

Solution:

$$\text{Mean} = E(X) = \Sigma x_i p_i$$

$$= x_1 p_1 + x_2 p_2 + x_3 p_3$$

$$= -3\times \frac{1}{6} + 6 \times \frac{1}{2} + 9 \times \frac{1}{3}$$

$$= \frac{11}{2}$$

$$E(X^2) = \Sigma x_i^{\,2} p_i$$

$$= x_1^{\,2} p_1 + x_2^{\,2} p_2 + x_3^{\,2} p_3$$

$$= \left(-3^{2\times}\frac{1}{6}\right) + \left(6^{2\times}\frac{1}{2}\right) + \ldots \left(9^{2\times}\frac{1}{3}\right)$$

$$= \frac{3}{2} + 18 + 27$$

$$= \frac{93}{2}$$

$$\text{Van}(X) = E(X^2) - (E(X))^2$$

$$= \frac{93}{2} - \left(\frac{11}{2}\right)^2$$

$$= \frac{93}{2} - \frac{121}{4}$$

$$= \frac{186 - 121}{4}$$

$$= \frac{65}{4}$$

$$\text{Mean} = \frac{11}{2},$$

$$\text{Var}(X) = \frac{65}{4}$$

9.6.2 Expectation of a continuous random variable

Let X be a continuous random variable with probability density function $f(x)$ then the mathematical expectation of X is defined as

Ch9_11th Stats.indd 269 07/12/2021 12:13:57

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$ provided the integral exists

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

NOTE

If $g(X)$ is a function of a random variable and $E[g(X)]$ exists then

Results:

(1) $E(c) = c$ where c is constant

Proof :

By definition, $E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$

$$E(c) = \int_{-\infty}^{\infty} cf(x)\,dx = c\int_{-\infty}^{\infty} f(x)\,dx$$

$$= c \times 1 = c \qquad \left(\because \int_{-\infty}^{\infty} f(x)\,dx = 1\right)$$

(2) $E(aX) = a\,E(X)$

Proof :

$$E(aX) = \int_{-\infty}^{\infty} axf(x)\,dx = a\int_{-\infty}^{\infty} xf(x)\,dx$$

$$= a\,E(X)$$

Example 9.14

The p.d.f. of a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{x}{2}, & 0 < x < 2 \\ 0, & elsewhere \end{cases}$$ find its mean and variance

Solution:

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$

$$E(X) = \int_{0}^{2} x\left(\frac{x}{2}\right)dx$$

$$= \frac{1}{2}\int_{0}^{2} x^2\,dx$$

$$= \frac{1}{2}\left[x^3/3\right]_0^2$$

Ch9_11th Stats.indd 270 07/12/2021 12:14:00

$$= \frac{1}{2}\left[\frac{8}{3} - 0\right]$$

$$= \frac{4}{3}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^2 \left(\frac{x}{2}\right) dx$$

$$= \frac{1}{2}\left[x^4/4\right]_0^2$$

$$= \frac{1}{8}\,[16\text{-}0]$$

$$= 2$$

$$\text{Variance } (X) = E(X^2) - [E(X)]^2$$

$$= 2 - \left(\frac{4}{3}\right)^2$$

$$= 2 - \frac{16}{9} = \frac{2}{9}$$

NOTE

The following results are true in both discrete and continuous cases.

(i) $E\,(1/X) \neq 1/E(X)$

(ii) $E\,[\log (X)] \neq \log E\,(X)$

(iii) $(E\,(X^2) \neq [E(X)]^2$

So far we have studied how to find mean and variance in the case of single random variables taken at a time but in some cases we need to calculate the expectation for a linear combination of random variables like $aX + bY$ or the product of the random variables $cX \times dY$ or involving more number of random variables. So here we see theorems useful in such situations.

9.6.3 Independent random variables

Random variables X and Y are said to be independent if the joint probability density function of X and Y can be written as the product of marginal densities of X and Y.

That is $f\,(x,y) = g(x)\,h(y)$

Here $g(x)$ marginal p.d.f. of X

$h(y)$ marginal p.d.f. of Y

9.7 Addition and Multiplication Theorem on Expectations

9.7.1 Addition Theorem on Expectations

1. Statement for Discrete random variable

If X and Y are two discrete random variables then

$$E\,(X+Y) = E(X) + E(Y)$$

Ch9_11th Stats.indd 271 07/12/2021 12:14:02

Proof

Let the random variable X assumes the values $x_1, x_2 \ldots x_n$ with corresponding probabilities $p_1, p_2 \ldots p_n$, and the random variable y assume the values $y_1, y_2 \ldots \ldots y_m$ with corresponding probabilities $p_1, p_2 \ldots p_m$

By definition,

$$E(X) = \sum x_i p_i \qquad \text{where } \sum p_i = 1$$

$$E(Y) = \sum y\, jpj \qquad \sum p_j = 1$$

Now,
$$E(X+Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij}(x_i + y_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} x_i + \sum_{j=1}^{n} p_{ij}(x_i + y_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} x_i + \sum_{i=1}^{n} \sum_{i=1}^{m} p_{ij} y_j$$

$$= \sum_{i=1}^{n} x_i p_i + \sum_{j=1}^{m} y_j p_j$$

$$= E(X) + E(Y)$$

2. Statement for Continuous random variable

Let X and Y are two continuous random variables with probability density functions $f(x)$ and $f(y)$ respectively. Then $E(X + Y) = E(X) + E(Y)$

Proof:

We know that

$$E(\mathrm{X}) = \int_{-\infty}^{\infty} x f(x)\, dx \qquad \text{and}$$

$$E(\mathrm{Y}) = \int_{-\infty}^{\infty} y f(y)\, dy$$

$$E(X+Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f(x, y)\, dx dy$$

$$= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x, y)\, dy \right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f(x, y)\, dx \right) dy$$

$$= \int_{-\infty}^{\infty} x f(x)\, dx + \int_{-\infty}^{\infty} y f(y)\, dy$$

$$= E(\mathrm{X}) + E(Y)$$

Ch9_11th Stats.indd 272 07/12/2021 12:14:04

$$E(aX+b) = a\,E(X) \pm b$$

$$E(aX \pm b) = \int_{-\infty}^{\infty} (ax \pm b) f(x)\,dx$$

$$= a \int_{-\infty}^{\infty} x f(x)\,dx \pm b \int_{-\infty}^{\infty} f(x)\,dx$$

$$= a\,E(X) \pm b \times 1 \qquad \text{as } \int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$= a\,E((X) \pm b)$$

Remarks:

1. **Statement:** $E(aX+b) = aE(X)+b$ where a and b are constants.

Proof: $E(aX+b) = E(aX)+E(b)$ by property 3

$= aE(X)+b$ by property 2

Similarly $E(aX-b) = aE(X)-b$

2. **Statement:** $E\left(\frac{ax+b}{c}\right) = \frac{[aE(x)+b]}{c}$

Proof: $E\frac{(aX+b)}{c} = \frac{E(aX+b)}{c}$

$= \frac{[aE(x)+b]}{c}$

3. **Statement:** $E(X - \overline{X}) = 0$

Proof: $E(X - \overline{X}) = E(X) - E(\overline{X})$ since $E(X) = \overline{X}$

$= \overline{X} - \overline{X} = 0$ $\overline{X}$ is a constant

Example 9.15

Find the expectation of the sum of the number obtained on throwing two dice.

Solution:

Let X&Y denote the number obtained on the I and II die respectively. Then each of them is a random variable which takes the value 1,2,3,4,5 and 6 with equal probability $\frac{1}{6}$

Ch9_11th Stats.indd 273 07/12/2021 12:14:06

$$E(X) = \sum x_i p_i$$

$$= 1\times\frac{1}{6}+2\times\frac{1}{6}+.....+6\times\frac{1}{6}$$

$$= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2}$$

Similarly,

$$E(Y) = \frac{7}{2}$$

Thus the expectation of the numbers obtained on two dices.

$X+Y$ takes the values 2, 3…12 with their probability given by

x	2	3	4	5	6	7	8	9	10	11	12
$P(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$E(X+Y) = 2\times\frac{1}{36}+3\times\frac{2}{36}+...+12\times\frac{1}{36}$$

$$= \frac{2+6+12+20+30+42+40+36+30+22+12}{36} = 7$$

$$E(X)+E(Y) = \frac{7}{2}+\frac{7}{2} = 7$$

$$\therefore \quad E(X+Y) = E(X) + E(Y)$$

Example 9.16

Let X and Y are two random variables with p.d.f given by

$$f(x,y) = \begin{cases} 4xy, & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

and their marginal density functions as

$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

and $g(y) = \begin{cases} 2y, & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$

prove that $E(X+Y) = E(X)+E(Y)$

Solution:

$$E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\,dxdy$$

$$= \int_0^1\int_0^1 (x+y)4xy\,dxdy$$

Ch9_11th Stats.indd 274 07/12/2021 12:14:10

$$= 4\left[\int_0^1\int_0^1 x^2 y\,dx\,dy + \int_0^1\int_0^1 xy^2\,dx\,dy\right]$$

$$= 4\left[\int_0^1\left(\int_0^1 x^2\,dx\right)y\,dy + \int_0^1\left(\int_0^1 y^2\,dy\right)x\,dx\right]$$

$$= 4[\int_0^1 \frac{1}{3}y\,dy + \int_0^1 \frac{1}{3}x\,dx]$$

$$= \frac{4}{3}\left[\int_0^1 y\,dy + \int_0^1 x\,dx\right]$$

$$= \frac{4}{3}[\frac{1}{2}+\frac{1}{2}] = \frac{4}{3} \qquad \ldots (1)$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$

$$= \int_0^1 x \times 2x\,dx$$

$$= 2\int_0^1 x^2\,dx$$

$$E(X) = 2\left[\frac{1}{3}\right] = \frac{2}{3}$$

$$\mathrm{E(Y)} = \int_{-\infty}^{\infty} yf(y)\,dy$$

$$= \int_0^1 y \times 2y\,dy$$

$$= 2\int_0^1 y^2\,dy$$

$$E(Y) = 2\left[\frac{1}{3}\right] = \frac{2}{3}$$

$$\mathrm{E(X)}+E(Y) = \frac{2}{3}+\frac{2}{3} = \frac{4}{3} \qquad \ldots (2)$$

From 1&2

$$E(X+Y) = E(X)+E(Y)$$

9.7.2 Multiplication Theorem on Expectation

Discrete random variable:

Statement: If X and Y are two independent variables then $E(XY) = E(X)\,E(Y)$

Ch9_11th Stats.indd 275 07/12/2021 12:14:14

Proof:

$$E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_{ij}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p_i p_j$$

$$= \left(\sum_{i=1}^{n} x_i p_i\right)\left(\sum_{j=1}^{m} y_i p_i\right)$$

$$= E(X)\, E(Y)$$

If X and Y are independent

$$P_{ij} = p_i p_j$$

Continuous random variable:

Statement: If X and Y are independent random variables Then $E(XY) = E(X)\, E(Y)$

Proof:

Now $$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x, y)\, dx dy$$

$$= = \int_{-\infty}^{\infty} x f(x)\, dx \int_{-\infty}^{\infty} y f(y)\, dy$$

$$= E(X)\, E(Y)$$

Example 9.17

Two coins are tossed one by one. First throw is considered as X and second throw is considered as Y following joint probability distribution is given by,

X / Y	**1**	**0**	**Total**
1	0.25	0.25	0.5
0	0.25	0.25	0.5
Total	0.5	0.5	1

[getting Head is taken as 1 and Tail is taken as 0]

Verify $E(XY) = E(X)\, E(Y)$

Solution:

A random variable XY can take the values 0 and 1

Ch9_11th Stats.indd 276 07/12/2021 12:14:15

$$E(XY) = \Sigma\Sigma\, xy\, p(x,y)$$

$$= 1 \times 0.25 + 0 \times 0.25 + 0 \times 0.25 + 0 \times 0.25$$

$$= 0.25$$

$$E(X) = \Sigma\, x\, p_i$$

$$= 1 \times 0.5 + 0 \times 0.5$$

$$= 0.5$$

$$E(Y) = \Sigma\, y\, p_j$$

$$= 1 \times 0.5 + 0 \times 0.5 = 0.5$$

$$E(X) \times E(Y) = 0.5 \times 0.5 = 0.25$$

$$E(XY) = E(X)\, E(Y)$$

[It is applicable only when X and Y are independent]

Example 9.18

The independent random variables X and Y have the p.d.f given by

$$f(x) = \begin{cases} 4ax, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$

$$f(y) = \begin{cases} 4by, & 0 \le y \le 1 \\ 0, & otherwise \end{cases}$$

Prove that $E(XY) = E(X)\, E(Y)$

Solution:

X and Y are independent

$$f(x,y) = f(x) \times f(Y)$$

$$f(x,y) = 4ax \times 4by$$

$$f(x,y) = \begin{cases} 16abxy, & 0 \le x,y \le 1 \\ 0, & otherwise \end{cases}$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\, dxdy$$

$$= \int_0^1\int_0^1 xy \times 16abxydxdy$$

Ch9_11th Stats.indd 277 07/12/2021 12:14:16

$$= 16ab \int_0^1 \left[\int_0^1 x^2 \, dx \right] y^2 \, dy$$

$$= 16ab \int_0^1 \frac{1}{3} y^2 \, dy$$

$$= 16ab \times \frac{1}{3} \times \frac{1}{3} = \frac{16ab}{9} \quad \rightarrow 1$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) \, dx$$

$$= \int_0^1 x \times 4ax \, dx = 4a \int_0^1 x^2 \, dx$$

$$= 4a\left(\frac{1}{3}\right) = \frac{4a}{3}$$

$$E(Y) = \int_{-\infty}^{\infty} y f(y) \, dy$$

$$= \int_0^1 y \times 4by \, dy$$

$$= 4b \int_0^1 y^2 \, dy$$

$$= 4b\left(\frac{1}{3}\right) = \frac{4b}{3}$$

$$E(X)\, E(Y) = \frac{4a}{3} \times \frac{4b}{3}$$

$$= \frac{16ab}{9} \quad \rightarrow 2$$

From 1 and 2,

$$E(XY) = E(X)\, E(Y)$$

9.8 Moments

Another approach helpful to find the summary measures for probability distribution is based on the 'moments'.We will discuss two types of moments.

(i) Moments about the origin. (Origin may be zero or any other constant say A). It is also called as raw moments.

(ii) Moments about the mean is called as central moments.

Ch9_11th Stats.indd 278 07/12/2021 12:14:20

9.8.1.Moments about the origin

Definition:

If X is a random variable and for each positive integer r (r = 1, 2, ...) the r^{th} raw moment about the origin '0' is $\mu'_r = E(X^r) = \sum x_i^{\,r}\ p_i$

First order raw Moment

Put $r = 1$, in the definition

$$\mu'_1 = E(X) = \sum x_i\, p_i = \overline{X}$$

This is called the mean of the random variable X. Hence the first order raw moment is mean.

Second order raw moment

Put $r = 2$ then

$$\mu'_2 = E(X^2) = \sum x_i^2 p_i$$

This is called second order raw moment about the origin.

9.8.2. Moments about the mean (Central moments)

Definition:

For each positive integer r, (r = 1, 2, ...), the rth order central moment of the random variable X is $\mu_r = E\,(X - \overline{X})^r$

First order central moment:

put r = 1 in the definition, then

$$\mu_1 = E\ (X - \overline{X}) = 0 \ \text{(always)}$$

Remark:

The algebraic sum of the deviations about the arithmetic mean is always 0

Second order central moment;

Put $r = 2$ in definition

$$\mu_2 = E\,(X - \overline{X})^2$$

$$= E\,(X^2 - 2\ X \times \overline{X} + \overline{X}^2\ )$$

Ch9_11th Stats.indd 279 07/12/2021 12:14:21

$$= E(X^2) - 2E(X)\,E(\overline{X}) + E(\overline{X})^2$$

$$= E(X^2) - 2\overline{X}\,\overline{X} + (\overline{X})^2$$

$$= E(X^2) - (\overline{X})^2$$

$$= E(X^2) - [E(X)]^2 \qquad \text{where } \overline{X} \text{ is constant and } E(X) = \overline{X}$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

This is called the 2nd central moment about the mean and is known as the variance of the random variable X.

i.e., $\quad$ Variance $=$ Var $(X) = \mu_2 = \mu'_2 - (\mu'_1)^2$

Standard Deviation (S.D) $= \sigma = \sqrt{variance}$

Some results based on variance:

(i) Var $(c) = 0$ i.e. Variance of a constant is zero

(ii) If c is constant then Var $(cX) = c^2$ Var(X)

(iii) If X is a random variable and c is a constant then Var $(X \pm c)$ = Var (X)

(iv) a and b are constants then Var $(aX \pm b) = a^2$ Var (X)

(v) a and b are constants then Var $(a \pm bX) = b^2$ Var (X).

(vi) If X and Y are independent random variables then Var $(X + Y)$
= Var (X) + Var(Y)

9.8.3. Moment generating function(M.G.F.)

Definition:

A moment generating function (m.g.f) of a random variable X about the origin is denoted by $M_x(t)$ and is given by

$M_x(t) = E(e^{tx})$, $|t| < 1$

(i) $M_x(t) = \sum (e^{tx})\, p(x)$ for a discrete distribution

(ii) $M_x(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$ for a continuous distribution

For a random variable X to find the moment about origin we use moment generating function.

Ch9_11th Stats.indd 280 07/12/2021 12:14:22

$E(X^r) = M_x^r(t)$ at t=0 $M_x^r(t)$

r^{th} differential of $M_x(t)$ is

$\mu_2 = \mu_2' - (\mu_1')^2$

To find the r^{th} moment of X about the origin.

We know that

$$M_x(t) = E(e^{tX})$$

$$= E\left[1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \ldots + \frac{(tX)^r}{r!} + \ldots\right]$$

$$= 1 + \frac{E(tX)}{1!} + \frac{E(tX)^2}{2!} + \frac{E(tX)^3}{3!} + \ldots + \frac{E(tX)^r}{r!} + \ldots$$

$$= 1 + t\frac{E(X)}{1!} + \frac{t^2E(X^2)}{2!} + \ldots + \frac{t^rE(X^r)}{r!} + \ldots$$

$$= 1 + t\mu'_1 + t^2\frac{\mu'_2}{2!} + \ldots + t^r\frac{\mu'_r}{r!} + \ldots$$

$$M_x(t) = \sum_{r=0}^{\infty} t^r \frac{\mu'_r}{r!}$$

From the series on the right hand side, μ_r' is the coefficient of $\frac{t^r}{r!}$ in $M_x(t)$.

Since $M_x(t)$ generates moments of the distribution and hence it is known as moment generating function.

Using the function, we can find mean and variance by using the first two raw moments.

$\mu_2 = \mu'_2 - (\mu'_1)^2$

9.8.4. Characteristic function

For some distribution, the M.G.F does not exist. In such cases we can use the characteristic function and it is more servicable function than the M.G.F.

Definition:

The characteristic function of a random variable X, denoted by $\phi_x(t)$, where $\phi_x(t) = E(e^{i\,t\,X})$ then

$$\phi_x(t) = E(e^{i\,tX}) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx \text{, for continuous random variable}$$

$$\phi_x(t) = \sum_e^{itx} p(x) \text{ , for discrete random variable}$$

where $i = \sqrt{-1}$

Ch9_11th Stats.indd 281 07/12/2021 12:14:26

Points to Remember

- Random variable is a real valued function defined on the sample space. It is always associated with the outcomes of the random experiment.
- Random variables are classified based on the nature of the range of values taken by it.
- Probability function is used to find the probabilities at the selected point or at the selected interval, in the range of the random variable.
- Cumulative distribution functions are used to find the probabilities in the subset of the range of the random variable.
- Summary measures of probability distribution (Mean and Variance) are obtained through the three approaches: (i) expectation (ii) m.g.f and (iii) characteristic function.

EXERCISE 9

ITK4X

I . Choose the best answer:

1. $f(x) = \begin{cases} kx^2 & 0 \le x \le 3 \\ 0 & elsewhere \end{cases}$ is a pdf then the value of K is

(a) $\frac{1}{3}$ (b) $\frac{1}{6}$ (c) $\frac{1}{9}$ (d) $\frac{1}{12}$

2. A random variable X has the following probability distribution

X	0	1	2	3	4	5
$P(X = x)$	$\frac{1}{4}$	$2a$	$3a$	$4a$	$5a$	$\frac{1}{4}$

Then $P(1 \le X \le 4)$ is

(a) $\frac{10}{21}$ (b) $\frac{2}{7}$ (c) $\frac{1}{2}$ (d) $\frac{1}{14}$

3. X is a discrete random variable which takes the value 0, 1, 2 and $P(X = 0) = \frac{144}{169}$; $P(X=1) = \frac{1}{169}$ and $P(X=2) = K$, then the value of K is

(a) $\frac{145}{169}$ (b) $\frac{24}{169}$ (c) $\frac{2}{169}$ (d) $\frac{143}{169}$

4. Given $E(X+C) = 8$ and $E(X-C) = 12$ then C is equal to

(a) -2 (b) 4 (c)-4 (d) 2

Ch9_11th Stats.indd 282 07/12/2021 12:14:28

5. The variance of random variable X is 4 its mean is 2 then $E(X^2)$ is

(a) 8 (b)6 (c) 4 (d) 2

6. Var $(4X+3)$ is

(a) 7 Var (X) (b) 16 Var (X) (c) 256 Var X (d) 0

7. The second order moment about mean is

(a) Mean (b)standard deviation (c)Variance (d) median

8. If variance = 20, second order moment about the origin = 276, then the mean of the random variable X is

(a) 16 (b)5 (c)4 (d) 2

II. Fill in the Blanks:

9. A variable whose values are real numbers and determined by the outcome of the random experiment is called _____

10. $F(X)$ is a distribution function then $F(-\infty)$ _____

11. Let X be a random variable and for any real constant a and b and $a \leq b$ then

$P(a \leq X \leq b) =$ ____

12. In the case of continuous random variable the probability at a point is always _____

13. If $F(X)$ is a cumulative distribution function of a continuous random variable X then $F'(X) =$ _____

14. $f(x)$ is a p.d.f. of continuous random variable then $\int_{-\infty}^{\infty} xf(x)\,dx$ ___

15. The Mathematical Expectation of a random variable X is also called as _____

III. Answer shortly:

16. Distinguish between discrete random variable and continuous random variable

17. Distinguish between probability mass function and probability density function.

18. Define mathematical Expectation of a discrete random variable.

Ch9_11th Stats.indd 283 07/12/2021 12:14:29

19. Define moment generating function.

20. State the Characteristic function for a continuous random variable.

21. If a random variable has the following probability distribution

X	o	1	2	3	4
$P(X=x)$	$3a$	$4a$	$6a$	$7a$	$8a$

find the value of 'a'.

22. Verify whether the following is a p.d.f. $f(x) = 5x^4$, $0<x<1$?

23. A random variable X has $E(X)=\frac{1}{2}$, $E(X^2)=\frac{1}{2}$ find its standard deviation

IV. Answer in brief:

24. Write down the properties of distribution function.

25. Find the probability distribution of 6's in throwing 3 dice once.

26. A box contains 6 red and 4 white balls. Three balls are drown at random obtain the probability distribution of the number of white balls drawn.

27. A random variable X has the density function) $f(x) = \begin{cases} \frac{1}{4} & -2 < x < 2 \\ 0 & elsewhere \end{cases}$

Then find (i) $P(-1 < X < 2)$ and (2) $P(X>1)$.

28. A game is played with single fair die. A player wins ₹ 20 if a 2 turns up ₹ 40 if a 4 turns up. Looses ₹ 30 if a 6 turns up. While he neither win nor looses If any other face turns up. Find his expectation of gain.

IV. Calculate the following:

29. In a continuous distribution the p.d.f. of X is

$f(x) = \begin{cases} \frac{3x(2-x)}{4} & 0 < x < 2 \\ 0 & elsewhere \end{cases}$ Find the mean and variance

30. If the *pdf* of a continuous random variable x is given by

$f(x) = \begin{cases} \frac{k(1-x^2)}{4} & 0 < x < 1 \\ 0 & elsewhere \end{cases}$

Find K and $E(X)$:

31. Two unbiased dice are thrown together at random. Find the expected value of a total number of points shown up.

Ch9_11th Stats.indd 284 07/12/2021 12:14:30

32. A probability distribution of a random variable X is given by

X	–2	3	1
P (x)	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Find $E(2X+5)$

33. If three coins are tossed find the mean and variance of the number of heads.

34. A discrete random variable has the following distribution function

X	0	1	2	3	4	5	6	7	8
P(x)	a	$3a$	$5a$	$7a$	$9a$	$11a$	$13a$	$15a$	$17a$

Find (i) a (ii) $P(X<3)$ (iii) $P(X\geq5)$ (iv) $P(3<X<7)$

35. Two cards are drawn with replacement from a well shuffled pack of 52 cards. Find the mean and variance of the number of Aces.

36. In an entrance examination student has answered all the 120 questions. Each question has 4 options and only one option is correct. A student gets one mark for correct answer and looses $\frac{1}{2}$ marks for wrong answer. What is the expectation of a mark scored by a student if he chooses the answer to each question at random?

37. A box contains 8 items of which 2 are defective. A man selects 3 items at random. Find the expected number of defective items he has drawn.

38. A box contains 4 white and 3 red balls. Find the probability distribution of number of red balls in 3 draws one by one from the box.

(i) With replacement (ii) Without replacement

39. The joint probability function of X and Y is given below:

X \ Y	1	3	9
2	0.1	0.1	0.05
4	0.2	0	0.1
6	0.1	0.15	K

Find (i) K (ii) $E(X+Y)$

Ch9_11th Stats.indd 285 07/12/2021 12:14:30

40. Find $E(3X)$ and $E(4Y)$ for the following joint probability distribution function of X and Y

X \ Y	1	2	3
−5	0	0.1	0.1
0	0.1	0.2	0.2
5	0.2	0.1	0

Answers:

I. 1. *c* 2. *c* 3. *b* 4. *a* 5. *a* 6. *b* 7. *c* 8. *a*

II. 9. random variable 10. 0 11. $F(b) - F(a)$ 12. 0 13. $f(x)$

14. $E(X)$ 15. arithmetic mean

III. 21. $\frac{1}{28}$ 22. 1 23. $\frac{1}{2}$

IV. 25.

X	0	1	2	3
$P(x)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

26.

X	0	1	2	3
$P(x)$	$\frac{5}{30}$	$\frac{15}{30}$	$\frac{9}{30}$	$\frac{1}{30}$

27. $\frac{3}{4}, \frac{1}{4}$ 28. 5

V. 29. 1, $\frac{1}{5}$ 30. 6, $\frac{3}{8}$ 31. 7 32. 7

33. $\frac{3}{2}, \frac{3}{4}$ 34. $\frac{1}{81}, \frac{1}{9}, \frac{56}{81}, \frac{11}{27}$

35. $\frac{2}{13}, \frac{24}{169}$ 36. −15 37. $\frac{3}{4}$

38. (i)

X	0	1	2	3
$P(x)$	$\frac{64}{343}$	$\frac{144}{343}$	$\frac{108}{343}$	$\frac{27}{343}$

(ii)

X	0	1	2	3
$P(x)$	$\frac{5}{30}$	$\frac{15}{30}$	$\frac{9}{30}$	$\frac{1}{30}$

39. 0.2, 8.7 40. −1.5, 2.0

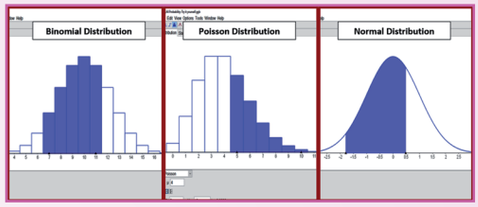
Ch9_11th Stats.indd 286 07/12/2021 12:14:35



ICT CORNER

PROBABILITY DISTRIBUTIONS

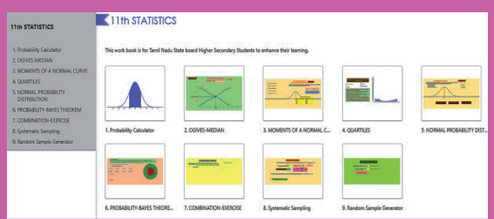
This activity probability distribution helps to understand binomial distribution, poisson distribution and normal distribution.



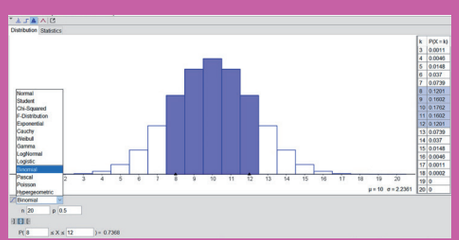
Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for Statistics are given, open the worksheet named "**Probability Calculator**"
- On left hand side bottom, in the drop up menu select binomial probability distribution. On the right-hand side, the data chart is given. If you select the particular data or a set of data, you can see the highlighted bar(s) and the respective probability at the left bottom.
- Similarly, you can select poisson probability distribution (or) normal probability distribution and find the probabilities.
- Also, you can select three menus as shown at the bottom to find left hand range (or) central range (or)right hand side range of probabilities.

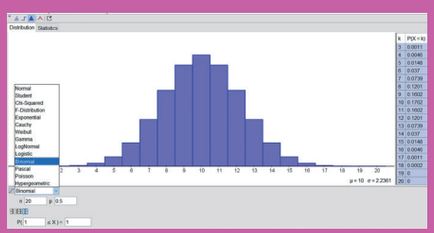
Step-1



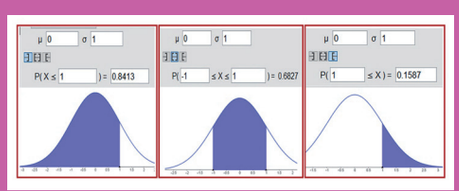
Step-2



Step-3



Step-4



Pictures are indicatives only*

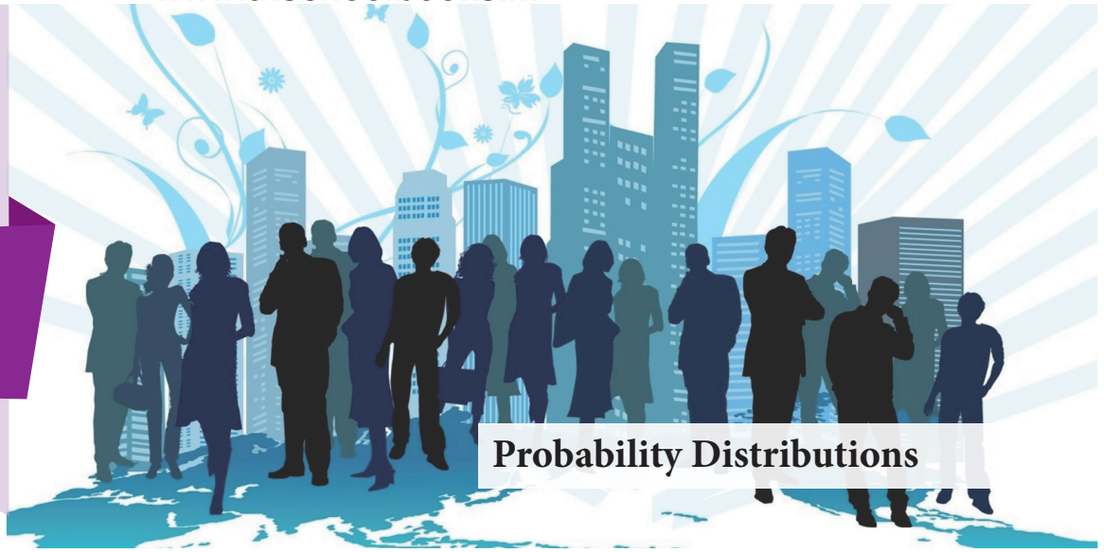
URL:

https://ggbm.at/uqVhSJWZ

Ch9_11th Stats.indd 287 07/12/2021 12:14:36

Chapter

10



Probability Distributions



Jacob Bernoulli
(6 Jan'1655 – 16 Aug'1705)

Jacob Bernoulli was born in Basel, Switzerland, is a mathematician. He contradicted to the desires of his parents, studied mathematics and astronomy and made tremendous achievement in later years. He traveled throughout Europe from 1676 to 1682, learning about the latest discoveries in mathematics and the sciences.

Jacob Bernoulli's first important contributions were a pamphlet on the parallels of logic and algebra published in 1685, after he worked on probability and geometry in 1687 . Jacob Bernoulli's paper of 1690 is important for the history of calculus, and in that integration was used with its complete meaning. In 1683 Bernoulli discovered the constant e by studying a question about compound interest which required him to find the value of the following expression $(1 + 1/n)^{1/n}$ (which is in fact e).

'Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write'. ***– H.G. Wells***

Learning Objectives

- Understands probability distributions
- Differentiates discrete and continuous distribution
- Estimates value using binomial, poisson distribution
- Calculates the values using uniform, normal distribution
- Applies the method of fitting various distributions.

Ch10_11th Stats.indd 288 07/12/2021 12:20:00

Introduction

In the earlier chapters we have learnt to construct probability distributions of certain Random variables. The probability of the various values of the random variables are obtained in accordance with the events and the nature of the experiment

In this chapter we are going to see some distributions called theoretical distributions. In these distributions, probabilities of the events are to be obtained using (formula) derived under certain conditions or assumptions. Of the many distributions available, the more common are Bernoulli, Binomial, Poisson, Normal, and Uniform distributions.

In practical situations one has to thoroughly understand the random environment and to describe it. It is followed by suggesting one of the above probability functions suitable to the situation and to obtain the requirement. The characteristics of the probability distributions such as Central Tendency, Dispersions, and Skewness are also to be studied.

10.1 Discrete distributions

10.1.1 Bernoulli's Distribution

It is discovered by a Swiss Mathematician James Bernoulli (1654-1705) for a trial which has only two outcomes viz. a success with probability p and a failure with probability $q = 1 - p$.

Definition

A random variable X is said to follow a Bernoulli distribution if its probability mass function is given by

$$P(X = x) = \begin{cases} p^x q^{1-x} & x = 0, 1 \\ 0 & otherwise \end{cases}$$

Characteristics of Bernoulli Distribution

(i) Number of trials is one

(ii) $q = 1 - p$

(iii) Constants of the distributions

(iv) (i) mean = p (ii) variance = pq (iii) standard deviation = $\sqrt{pq}$

Ch10_11th Stats.indd 289 07/12/2021 12:20:00

10.1.2 BINOMIAL DISTRIBUTION

Introduction

Binomial distribution was discovered by James Bernoulli (1654 – 1705) in the year 1700 and was first published in 1713 eight years after his death. The distribution of the Sum of n independent Bernoulli variables is known as a Binomial distribution.

That is the sum of outcome of n independent experiments of Bernoulli trials, in each of which the probability of success is constant p and the probability of failure is q=1– p is called Binomial experiment.

Definition

A random variable X denoting the number of successes in an outcome of a Binomial experiment having n trials and p as the probability of success in each trial is called Binomial random variable. Its probability mass function is given by

$$P(X = x) = \begin{cases} nC_x p^x q^{n-x} & \text{for } x = 0,1,2,...n \\ 0 & \text{otherwise} \end{cases} \qquad \text{where } q=1-p$$

NOTE

(i) If X is a Binomial variate, it is denoted by $X \sim B(n,p)$

(ii) n,p are the parameters of the distribution.

(iii) The probability values are the successive terms of the expansion of $(q+p)^n$

Conditions for Binomial Distribution

We get the Binomial Distribution under the following experimental conditions:

(i) The number of trials `n' is finite

(ii) The trials are independent of each other

(iii) The probability of success `p' is same for each trial

(iv) Each trial must result in a success or a failure.

NOTE

The problems relating to tossing of coins or throwing of dice or drawing cards with replacement from a pack of cards lead to Binomial distribution.

Characteristics of Binomial Distribution

(i) Binomial distribution is a discrete distribution i.e., X can take values 0, 1, 2, ... n where `n' is finite

Ch10_11th Stats.indd 290 07/12/2021 12:20:01

(ii) Constants of the distributions are:

Mean = np; Variance = npq; Standard deviation = $\sqrt{npq}$

Skewness = $\frac{q-p}{\sqrt{npq}}$; Kurtosis = $\frac{1-6pq}{npq}$

(iii) It may have one or two modes.

(iv) If $X \sim B(n_1, p)$ *and* $Y \sim B(n_2, p)$ and that X and Y are independent then $X + Y \sim B(n_1 + n_2, p)$

(v) If `$n$' independent trials are repeated $N$ times the expected frequency of `x' successes is $N \times {}^{n}C_{x}\ p^{x}\ q^{n-x}$

(vi) If $p = 0.5$, the distribution is symmetric.

Example 10.1

Comment on the following 'The mean of binomial distribution is 5 and its variance is 9'.

Solution:

Given mean $np = 5$ and variance $npq = 9$

$\therefore \frac{Value\ of\ variance}{Value\ of\ mean} = \frac{npq}{np} = \frac{9}{5} \therefore q = \frac{9}{5} > 1$ is not possible

as $0 \leq q \leq 1$ and hence the given statement is wrong.

Example 10.2

Eight coins are tossed simultaneously. Find the probability of getting atleast six heads.

Solution:

Here n=8 $p = P$ (head) $= \frac{1}{2}$ $q = 1 - \frac{1}{2} = \frac{1}{2}$

Trials satisfy conditions of Binomial distribution

Hence $P(X = x) = {}^{n}C_{x}\ p^{x}\ q^{n-x}$ $x = 0, 1, 2, \ldots n$

$= {}^{8}C_{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{8-x}$ $x = 0,1,2,\ldots.8$

$= {}^{8}c_{x}\left(\frac{1}{2}\right)^{x+8-x}$

$= {}^{8}C_{x}(\frac{1}{2})^{8}$

Ch10_11th Stats.indd 291 07/12/2021 12:20:04

$$\therefore P(X = x) = \frac{8C_x}{256}$$

P (getting atleast six heads)

$$= P(x \geq 6)$$

$$= P(x = 6) + P(x = 7) + P(x = 8)$$

$$= \frac{8C_6}{256} + \frac{8C_7}{256} + \frac{8C_8}{256}$$

$$= \frac{28}{256} + \frac{8}{256} + \frac{1}{256}$$

$$= \frac{37}{256}$$

Example 10.3

Ten coins are tossed simultaneously. Find the probability of getting (i) atleast seven heads (ii) exactly seven heads (iii) atmost seven heads.

Solution:

X denote the number of heads appear

$$P(X = x) = nC_x\, p^x\, q^{n-x},\ x = 0,1,2,\ldots n$$

Given: $p = P(\text{head}) = \frac{1}{2}$ $q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$ and $n = 10$

$$\therefore X \sim B\left(10, \frac{1}{2}\right)$$

$$= 10C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x}$$

$$P(X = x) = \frac{10C_x}{1024}$$

(i) P (atleast seven heads)

$$P(X \geq 7) = P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10)$$

$$= \frac{10C_7}{1024} + \frac{10C_8}{1024} + \frac{10C_9}{1024} \frac{10C_{10}}{1024}$$

$$= \frac{120}{1024} + \frac{45}{1024} + \frac{10}{1024} + \frac{1}{1024}$$

$$= \frac{176}{1024}$$

(ii) P (exactly 7 heads)

$$P(x=7) = \frac{10C_7}{1024} = \frac{120}{1024}$$

Ch10_11th Stats.indd 292 07/12/2021 12:20:06

(iii) P (atmost 7 heads)

$$= P(x \leq 7) = 1 - P(x > 7)$$

$$= 1 - \{P(x = 8) + P(x = 9) + P(x = 10)\}$$

$$= 1 - 1 - \left\{\frac{10C_8}{1024} + \frac{10C_9}{1024} + \frac{10C_{10}}{1024}\right\}$$

$$= 1 - \frac{56}{1024}$$

$$= \frac{968}{1024}$$

Example 10.4

With usual notation find p for Binomial random variable X if $n = 6$ and $9\,P(x = 4) = P(x = 2)$

Solution:

$$P(X = x) = {}^{n}C_x\, p^x\, q^{n-x},\ x = 0,1,2,\ldots n$$

$$X \sim B(6, p) \Rightarrow P(X = x) = {}^{6}C_x p^x\, q^{(6-x)}$$

Also $9 \times P(X = 4) = P(X = 2)$

$\Rightarrow$ $9 \times {}^{6}C_4 p^4 q^2 = {}^{6}C_2\, p^2\, q^4$

$\Rightarrow$ $9\,p^2 = q^2$

$\Rightarrow$ $3p = q$ as $p, q > 0$

$3p = 1 - p$

$4p = 1$

$\Rightarrow$ $p = \frac{1}{4} = 0.25$

Example 10.5

A Binomial distribution has parameters n=5 and p=1/4. Find the Skewness and Kurtosis.

Solution:

Here we are given n=5 and $p = \frac{1}{4}$

Skewness $= \frac{q - p}{\sqrt{npq}}$

Ch10_11th Stats.indd 293 07/12/2021 12:20:07

$$= \frac{\frac{3}{4} - \frac{1}{4}}{\sqrt{5 \times \frac{1}{4} \times \frac{3}{4}}}$$

$$= \frac{\frac{2}{4}}{\sqrt{\frac{15}{16}}}$$

$$= \frac{2}{\sqrt{15}}$$

Finding: The distribution is positively skewed.

Kurtosis

$$\text{Kurtosis} = \frac{1 - 6pq}{npq}$$

$$= \frac{1 - 6 \times \frac{1}{4} \times \frac{3}{4}}{\frac{15}{16}}$$

$$= \frac{\frac{-2}{16}}{\frac{15}{16}}$$

$$= \frac{-2}{15}$$

$$= -0.1333$$

Finding: The distribution is Platykurtic.

Example 10.6

In a Binomial distribution with 7 trials, $P(X=3)=P(X=4)$ Check whether it is a symmetrical distribution?

Solution:

A Binomial distribution is said to be symmetrical if $p= q= \frac{1}{2}$

Given: $P(X=3) = P(X=4)$

$X \sim B(n, p)$

$P(X = x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$, $x= 0,1,2,\ldots n$

${}^{n}C_{3}\, p^{3} q^{n-3} = {}^{n}C_{4}\, p^{4}\, q^{n-4}$

${}^{7}C_{3}\, p^{3} q^{4} = {}^{7}C_{4}\, p^{4} q^{3}$ note that ${}^{7}C_{3} = {}^{7}C_{4}$

On simplifying, we have $q = p$

Ch10_11th Stats.indd 294 07/12/2021 12:20:09

$$1-p = p$$

$$1 = 2p$$

$$p = \frac{1}{2}$$

$$q = \frac{1}{2}$$

Hence the given Binomial distribution is symmetrical.

Example 10.7

From a pack of 52 cards 4 cards are drawn one after another with replacement. Find the mean and variance of the distribution of the number of kings.

Solution:

Success X=event of getting king in a draw

p=probability of getting king in a single trial

$$p = \frac{4}{52}$$

$$= \frac{1}{13}$$

This is constant for each trial.

Hence, it is a binomial distribution with $n=4$, $p = \frac{1}{13}$ and $q = \frac{12}{13}$

$$\text{Mean} = np = 4\times\frac{1}{13} = \frac{4}{13}$$

$$\text{Variance} = npq = \frac{4}{13}\times\frac{12}{13} = \frac{48}{169}$$

Example 10.8

In a street of 200 families, 40 families purchase the Hindu newspaper. Among the families a sample of 10 families is selected, find the probability that

(i) Only one family purchase the news paper

(ii) No family purchasing

(iii) Not more than one family purchase it

Solution:

$X \sim B(n, p)$

$P(X = x) = nC_x\, p^x\, q^{n-x}$, $x = 0,1,2,...n$

Ch10_11th Stats.indd 295 07/12/2021 12:20:10

Let X denote the number of families purchasing Hindu Paper

p = Probability of their family purchasing the Hindu

$$p = \frac{40}{200} = \frac{1}{5}$$

$$q = \frac{4}{5}$$

$$n = 10$$

(i) Only one family purchase the Hindu

$$P(X=1) = nC_1\, p^1\, q^{n-1}$$

$$= 10C_1 \times \frac{1}{5} \times \left(\frac{4}{5}\right)^9$$

$$= 2 \times \left(\frac{4}{5}\right)^9$$

(ii) No family purchasing the Hindu

$$P(X=0) = 10C_0 (\frac{1}{5})^0 (\frac{4}{5})^{10}$$

$$= (\frac{4}{10})^{10}$$

(iii) Not more than one family purchasing The Hindu means that $X \leq 1$

$$P(X \leq 1) = P[x=0] + P[x=1]$$

$$= 10C_0 (\frac{1}{5})^0 (\frac{4}{5})^{10} + 10C_1 (\frac{1}{5})^1 (\frac{4}{9})^9$$

$$= (\frac{4}{5})^{10} + 10 \times (\frac{1}{5})^1 (\frac{4}{5})^9$$

$$= (\frac{4}{5})^9 \left[(\frac{4}{5}) + 2\right]$$

$$= (\frac{4}{5})^9 (\frac{14}{5})$$

Example 10.9

In a tourist spot, 80% of tourists are repeated visitors. Find the distribution of the numbers of repeated visitors among 4 selected peoples visiting the place. Also find its mode or the maximum visits by a visitor.

Solution:

Let the random variable X denote the number of repeated visitors.

$X \sim B\ (n, p)$

Ch10_11th Stats.indd 296 07/12/2021 12:20:11

$P(X = x) = nC_x\, p^x\, q^{n-x}$, x= 0,1,2,...n

It is a Binomial Distribution with n=4

$p = \frac{80}{100} = \frac{4}{5}$ $\quad q = 1\text{-}p = 1 - \frac{4}{5} = \frac{1}{5}$

$$P(x = 0) = 4C_0 (\tfrac{4}{5})^0 (\tfrac{1}{5})^4 = (\tfrac{1}{625})$$

$$P(x=1) = 4C_1 (\tfrac{4}{5})(\tfrac{1}{5})^3 = \tfrac{16}{625}$$

$$P(x=2) = 4C_2 (\tfrac{4}{5})^2(\tfrac{1}{5})^2 = (\tfrac{96}{625})$$

$$P(x = 3) = 4C_3 (\tfrac{4}{5})^3(\tfrac{1}{5})^1 = (\tfrac{256}{625})$$

$$P(x = 4) = 4C_4 \left(\tfrac{4}{5}\right)^4 \left(\tfrac{1}{5}\right)^0 = \left(\tfrac{256}{625}\right)$$

The probability distribution is given below.

$X=x$	0	1	2	3	4
$P(X=x)$	$\frac{1}{625}$	$\frac{16}{625}$	$\frac{96}{625}$	$\frac{256}{625}$	$\frac{256}{625}$

Model values are 3 and 4 (Maximum number of occurrences)

Example 10.10

In a college, 60% of the students are boys. A sample of 4 students of the college, is taken, find the minimum number of boys should it have so that probability up to that number is $\geq \frac{1}{2}$.

Solution:

It is given that 60% of the students of the college are boys and the selection probability for a boy is 60% or 0.6 As we are taking four samples, the number of trials n = 4. The selection process is independent.

$X \sim B(n, p)$

$P(X = x) = nC_x\, p^x\, q^{n-x}$, x= 0,1,2,...n

Let X be the number of boys so that $P(X \leq x) \geq \frac{1}{2}$

If $x = 0$ $\quad P(X \leq 0) = 4C_0 (\tfrac{3}{5})^0 (\tfrac{2}{5})^4 = \tfrac{16}{625} < \tfrac{1}{2}$

Ch10_11th Stats.indd 297 07/12/2021 12:20:15

$x=1$ $\quad P(X\leq 1) = P(x=0)+P(x=1)$

$$= 4C_0\left(\frac{3}{5}\right)^0\left(\frac{2}{5}\right)^4 + 4C_1\left(\frac{3}{5}\right)^1\left(\frac{2}{5}\right)^3$$

$$= \frac{16}{625} + \frac{96}{625} = \frac{112}{625} < \frac{1}{2}$$

$x=2$ $\quad P(X\leq 2) = P(x=0)+P(x=1)+P(x=2)$

$$= \frac{112}{625} + P(x=2) = \frac{112}{625} + 4C_2\left(\frac{3}{5}\right)^2\left(\frac{2}{5}\right)^2$$

$$= \frac{112}{625} + \frac{216}{625} = \frac{328}{625} > \frac{1}{2}$$

Therefore the sample should contain a minimum of 2 boys.

10.1.3 POISSION DISTRIBUTION

Introduction

In a Binomial distribution with parameter n and p if the exact value of n is not definitely known and if p is very small then it is not possible to the find the binomial probabilities .Even if n is known and it is very large, calculations are tedious. In such situations a distribution called Poisson distribution is very much useful.

In 1837 French mathematician Simeon Dennis Poisson derived the distribution as a limiting case of Binomial distribution. It is called after his name as Poisson distribution.

Conditions:

(i) The number of trails 'n' is indefinitely large i.e., $n \to \infty$

(ii) The probability of a success 'p' for each trial is very small i.e., $p \to 0$

(iii) $np = \lambda$ is finite

(iv) Events are Independent

Definition

A random variable X is said to follow a Poisson distribution if it assumes only non-negative integral values and its probability mass function is given by

$$P(X=x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}; & x = 0,1,2,..... \\ 0 & otherwise \end{cases}$$

Ch10_11th Stats.indd 298 07/12/2021 12:20:18

NOTE

(i) λ is called the parameter of the Poisson distribution.

(ii) $e = 1 + \frac{1}{1!} + \frac{1}{2!} + ... = 2.71828\ldots$ is an irrational number.[n]

Characteristics of Poisson Distribution

(i) Poisson distribution is a discrete distribution i.e., X can take values 0, 1, 2,…

(ii) p is small, q is large and n is indefinitely large i.e., $p \longrightarrow 0$ $q \longrightarrow 1$ and $n \longrightarrow \infty$ and np is finite

(iii) Values of constants : (a) Mean = λ = variance (b) Standard deviation = $\sqrt{\lambda}$ (c) Skewness = $\frac{1}{\sqrt{\lambda}}$ (iv) Kurtosis = $\frac{1}{\lambda}$

(iv) It may have one or two modes

(v) If X and Y are two independent Poisson variates, $X+Y$ is also a Poisson variate.

(vi) If X and Y are two independent Poisson variates, $X-Y$ need not be a Poisson variate.

(vii) Poisson distribution is positively skewed.

(viii) It is leptokurtic.

NOTE

As p is very small the probability of happening of the event is rare and this distribution is applicable where study of interest is on rare events

Some examples:

(i) The event of a student getting first mark in all subjects and at all the examinations

(ii) The event of finding a defective item from the production of a reputed company

(iii) The number of blinds born in a particular year

(iv) The number of mistakes committed in a typed page

(v) The number of traffic accidents per day at a busy junction.

(vi) The number of death claims received per day by an insurance company.

Ch10_11th Stats.indd 299 07/12/2021 12:20:18

Example 10.11

If 2% of electric bulbs manufactured by a certain company are defective find the probability that in a sample of 200 bulbs (i) less than 2 bulbs are defective (ii) more than 3 bulbs are defective. [e-4 = 0.0183]

Solution:

Let X denote the number of defective bulbs

$X \sim P(\lambda)$

$$\therefore \quad P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots \infty$$

Given $p = P$ (a defective bulb) = 2% = $\frac{2}{100}$ = 0.02

$$n = 200$$

$$\therefore \quad \lambda = np = 200 \times 0.02 = 4$$

$$\therefore \quad P(X = x) = \frac{e^{-4}4^x}{x!}, \quad x = 0, 1, 2, \ldots \infty$$

(i) P (less than 2 bulbs are defective)

$$= P(X < 2)$$

$$= P(x = 0) + P(x = 1)$$

$$= \frac{e^{-4}.4^0}{0!} + \frac{e^{-4},4^1}{1!}$$

$$= e^{-4}(1 + 4)$$

$$= 0.0183 \text{ x } 5$$

$$= 0.0915$$

(ii) P (more than 3 defectives)

$$= P(X > 3)$$

$$= 1 - P(X \leq 3)$$

$$= 1 - \{P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)\}$$

$$= 1 - \left\{\frac{e^{-4}.4^0}{0!} + \frac{e^{-4}.4^1}{1!} + \frac{e^{-4}.4^2}{2!} + \frac{e^{-4}.4^3}{3!}\right\}$$

$$= 1 - e^{-4}\{1 + 4 + 8 + 10.667\}$$

Ch10_11th Stats.indd 300 07/12/2021 12:20:19

$= 1 - 0.0183 \times 23.667$

$= 0.567$

Example 10.12

In a Poisson distribution $3P(X= 2) = P(X = 4)$. Find its parameter $'\lambda$

Solution:

The *pmf* of Poisson distribution is $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x= 0, 1, 2, ...\infty$,

Given $3\ P(X = 2)=P(X = 4)$

$$3.\frac{e^{-\lambda}\lambda^2}{2!} = \frac{e^{-\lambda}.\lambda^4}{4!}$$

$$\lambda^2 = \frac{3 \times 4!}{2!} = 36$$

$\therefore \lambda = 6$ as $\lambda > 0$

Example 10.13

Find the skewness and kurtosis of a Poisson variate with parameter 4.

Solution:

$$\lambda = 4$$

$$\text{Skewness} = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{4}} = \frac{1}{2}$$

$$\text{Kurtosis} = \frac{1}{\lambda} = \frac{1}{4}$$

Example 10.14

If there are 400 errors in a book of 1000 pages, find the probability that a randomly chosen page from the book has exactly 3 errors.

Solution:

Let X denote the number of errors in pages

$X \sim P(\lambda)$

$\therefore P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\qquad x= 0, 1, 2, ...\infty$

The average number of errors per page $= \frac{400}{1000}$

i.e., $\lambda = \frac{400}{1000} = 0.4$

Ch10_11th Stats.indd 301 07/12/2021 12:20:21

$$P(X=3) = \frac{e^{-\lambda}\lambda^3}{3!} = \frac{e^{-0.4}(0.4)^3}{3\times 2\times 1}$$

$$= \frac{0.6703\times 0.064}{6}$$

$$= 0.00715$$

Example 10.15

If X is a Poisson variate with $P(X=0) = 0.2725$, find $P(X=1)$

Solution:

$X \sim P(\lambda)$

$$\therefore P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots\infty$$

$P(X=0) = 0.2725$

$\frac{e^{-\lambda}\lambda^0}{0!} = 0.2725$

$e^{-\lambda} = 0.2725$

$\lambda = 1.3$ (from the table of values of e^{-m})

$$P(x=1) = \frac{e^{-\lambda}\lambda^1}{1!} = \frac{e^{-1.3}\times 1.3^1}{1!}$$

$$= 0.2725\times 1.3$$

$$= 0.3543$$

Example 10.16

The probability of safety pin manufactured by a firm to be defective is 0.04. (i) Find the probability that a box containing 100 such pins has one defective pin . (ii) Among 200 such boxes, how many boxes will have no defective pin

Solution:

Let X denote the number boxes with defective pins

$X \sim P(\lambda)$

$$\therefore P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots\infty$$

$p = 0.04$

Ch10_11th Stats.indd 302 07/12/2021 12:20:22

$$n = 100$$

$$\lambda = np = 4$$

(i) $P(X=1) = \frac{e^{-\lambda}\lambda}{1!} = e^{-4}(4) = 0.0183 \times 4$

$$= 0.0732$$

(ii) $P(X=0) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-4} = 0.0183$

Number of boxes having no defective pin $= 200 \times 0.0183$

$= 3.660$

$= 4$

10.2 Continuous distributions:

10.2.1 Rectangular or Uniform Distribution

Definition

A random variable X is said to have a continuous Uniform distribution over the interval (a, b) if its probability density function is

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases}$$

Characteristics of Uniform Distribution

(i) a and b are the parameters of the Uniform distribution and we write $X \sim U(a, b)$

(ii) The distribution is also known as Rectangular distribution, as the curve

(iii) $y = f(x)$ describes a rectangle over the x-axis and between ordinates at x = a and $x = b$.

(iv) f $X \sim U(-a, a)$ then its p.d.f. is $f(x) = \begin{cases} \frac{1}{2a} & -a < x < a \\ 0 & otherwise \end{cases}$

Constants of uniform distribution X ~ U (a, b)

(i) Mean $\mu = \frac{a+b}{2}$

(ii) Variance $\sigma^2 = \frac{(b-a)^2}{12}$

(iii) Median $= \frac{a+b}{2}$

(iv) Skewness = 0

(v) Kurtosis = $-\frac{6}{5}$

(vi) $Q_1 = \frac{3a+b}{4}$ (vii) $Q_3 = \frac{a+3b}{4}$

Ch10_11th Stats.indd 303 07/12/2021 12:20:25

Example 10.17

If $X \sim U(200, 250)$ find its p.d.f and $P(X > 230)$

Solution :

(i) For $X \sim U(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases}$$

Taking $a = 200$ and $b = 250$ the required p.d.f. is

$$f(x) = \begin{cases} \frac{1}{250-200} & 200 < x < 250 \\ 0 & otherwise \end{cases}$$

$$= \begin{cases} \frac{1}{50} & 200 < x < 250 \\ 0 & otherwise \end{cases}$$

(ii) $$P(x > 230) = \int_{230}^{\infty} f(x)dx = \int_{230}^{250} \frac{1}{50} dx$$

$$= \frac{1}{50}[x]_{230}^{250} = \frac{250-230}{50} = \frac{20}{50} = 0.4$$

Example 10.18

If X is a Uniform variate with the parameter 50 and100, find the mean, median and standard deviation.

Solution:

$X \sim U(50,100)$

Here a=50 b=100

$$\text{mean} = \text{median} = \frac{a+b}{2} = \frac{150}{2} = 75$$

$$\text{S.D} = \sqrt{\frac{(b-a)^2}{12}} = \frac{b-a}{\sqrt{12}} = \frac{100-50}{\sqrt{12}}$$

$$= \frac{50}{\sqrt{12}} = \frac{50}{3.464}$$

$$= 14.434$$

Example 10.19

If X is a random variable having a uniform distribution $U(a,b)$ such that $P(20<X<40)=0.2$ and mean = 150, find a and b.

Ch10_11th Stats.indd 304 07/12/2021 12:20:27

Solution:

For $X \sim U(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases}$$

$P(20<X<40)=0.2$

$$\int_{20}^{40} \frac{1}{b-a} dx = 0.2$$

$$\frac{1}{b-a}(x)_{20}^{40} = 0.2$$

$$\frac{1}{b-a}(20) = 0.2$$

$$b - a = 100 \quad \ldots (1)$$

$$\text{mean} = 150$$

$$\frac{a+b}{2} = 150$$

$$a+b = 300 \quad \ldots (2)$$

(1)+(2) implies $2b = 400$ and $b = 200$.

Substituting b in (2) we have $a + 200 = 300$ and that $a = 100$.

$a=100$, $b=200$

$X \sim U(100,200)$

Example 10.20

If X is a Uniform variable U(a,b) with first and third quartiles 100 and 200, find the p.d.f of X.

Solution :

$$Q_1 = 100$$

$$\frac{3a+b}{4} = 100$$

$$3a+b = 400 \quad \ldots (1)$$

$$Q_3 = 200$$

$$\frac{a+3b}{4} = 200$$

$$a+3b = 800 \quad \ldots (2)$$

Ch10_11th Stats.indd 305 07/12/2021 12:20:29

Solving (1) and (2) we get

$$a = 50 \qquad b = 250$$

$$f(x) = \begin{cases} \frac{1}{b-a} = \frac{1}{250-50} = \frac{1}{200}; 50 < x < 250 \\ 0 \qquad otherwise \end{cases}$$

Example 10.21

Electric trains on a certain line run every 15 minutes between mid- night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least ten minutes?

Let the random variable X denote the waiting time (in minutes).

The given assumption indicates that X is distributed Uniformly on (0,15).

$$P(X > 10) = \int_{10}^{15} f(x)dx = \frac{1}{15}\int_{10}^{15} 1dx$$

$$= \frac{1}{15}(x)_{10}^{15} = \frac{1}{15}(15-10) = \frac{1}{3}$$

10.2.2 NORMAL DISTRIBUTION

Introduction

The Normal distribution was first discovered by the English Mathematician De-Moivre in 1733, and he obtained this distribution as a limiting case of Binomial distribution and applied it to problems arising in the game of chance. In 1774 Laplace used it to estimate historical errors and in 1809 it was used by Gauss as the distribution of errors in Astronomy. Thus throughout the 18th and 19th centuries efforts were made for a common law for all continuous distributions which was then known as the Normal distribution.

Definition

A random variable X is said to have a Normal distribution with parameters μ and σ^2 if its probability density function is given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2}$ where

$-\infty < x < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$

It is denoted by $X \sim N(\mu, \sigma^2)$

Here μ is called as mean and σ^2 is variance of the distribution.

Ch10_11th Stats.indd 306 07/12/2021 12:20:30

NOTE

(i) $B(X: n, p)$ when $n \to \infty$ and p, q are not small will become a Normal distribution.

(i) $P(X, \lambda)$ when $\lambda \longrightarrow \infty$ will become a Normal distribution.

(ii) When $X \sim N(\mu, \sigma^2)$ then $Z = \frac{x-\mu}{\sigma}$ then $Z \sim N(0, 1)$

i.e., $f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \quad -\infty \le z \le \infty$

Z is known as a Standard Normal variate with mean 0 and variance 1.

To find probabilities of X we convert X into Z and then make use of standard normal table.

Properties of Normal Distribution

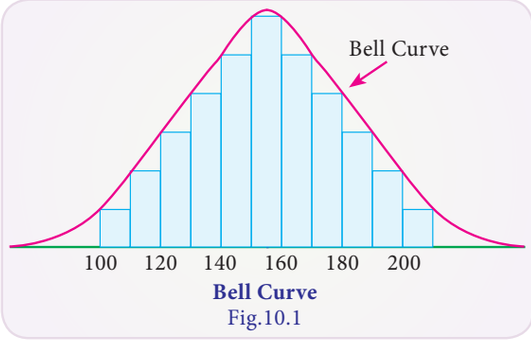
(i) The curve is bell shaped and is symmetric about $X = \mu$

(ii) Mean = Median = Mode = μ

(iii) Unimodal at $X = \mu$ and Model height $= \frac{1}{\sigma\sqrt{2\pi}}$

(iv) Skewness $\beta_1 = 0$ and Kurtosis $\beta_2 = 3$

(v) The points of inflections are at $x = \mu \pm \sigma$



Bell Curve
Fig.10.1

(vi) The X – axis is the asymptote to the curve

(vii) The mean deviation about mean is 0.8σ

(viii) Quartile deviation is 0.6745σ

(ix) If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ and X and Y are independent, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$


Fig.10.2

(x) Area properties:

(a) Total area $P(-\infty < X < \infty) = 1$

(b) Area about $X = \mu$: $P(-\infty < X < \infty) = P(\mu < X < \infty) = 0.5$

Ch10_11th Stats.indd 307 07/12/2021 12:20:32

(c) $P(\mu-\sigma < X < \mu+\sigma) = P(-1 < Z < 1)$

$= 2\,P(0 < Z < 1)$

$= 2\,(0.3413)$

$= 0.6826$

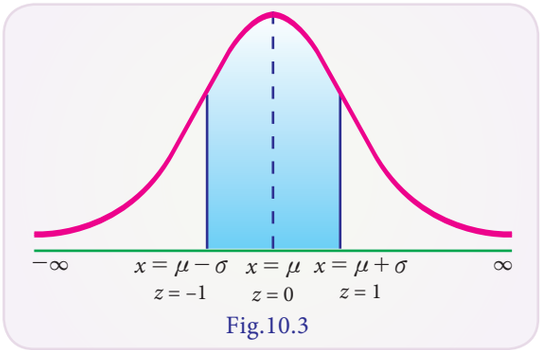


Fig.10.3

(d) $P(\mu-2\sigma < X < \mu+2\sigma)$

$= P(-2 < Z < 2)$

$= 2\,P(0 < Z < 2)$

$= 2\,(0.4772)$

$= 0.9544$

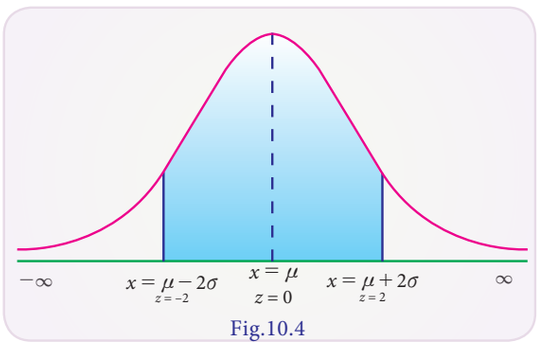


Fig.10.4

(e) $P(\mu-3\sigma < X < \mu+3\sigma)$

$= P(-3 < Z < 3)$

$= 2\,P(0 < Z < 3)$

$= 2\,(0.49865)$

$= 0.9973$

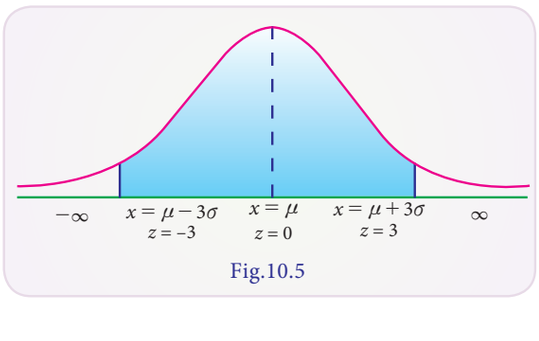


Fig.10.5

(f) $P(|X-\mu| > 3\sigma = P(|Z| > 3)$

$= 1 - P(|Z| < 3)$

$= 1 - P(-3 < Z < 3)$

$= 1 - 0.9973$

$= 0.0027$

Example 10.22

Find the area between $z = 0$ and $z = 1.56$

Solution:

$P(0 < Z < 1.56) = 0.4406$

(from the Normal Probability table)

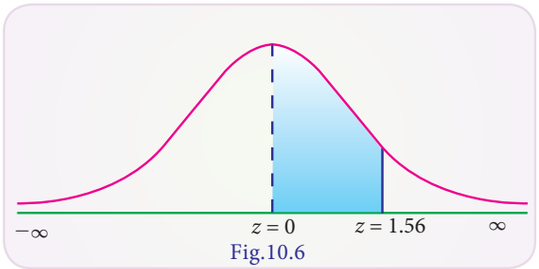


Fig.10.6

Ch10_11th Stats.indd 308 07/12/2021 12:20:34

Example 10.23

Find the area of the Standard Normal variate from – 1.96 to 0

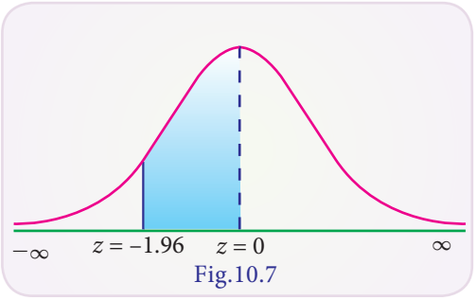


Fig.10.7

Solution:

$P(-1.96 < Z < 0) = P(0 < Z < 1.96)$ (by symmetry)

$= 0.4750$ (from the Normal Probability table)

Example 10.24

Find the area to the right of $Z = 0.25$

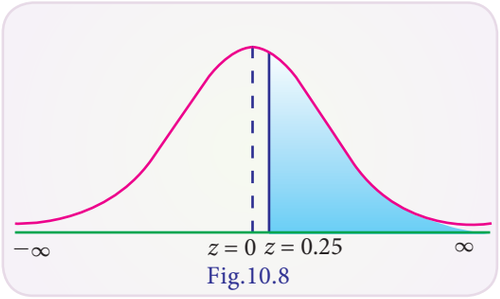


Fig.10.8

Solution:

$P(Z > 0.25) = P(0 < Z < \infty) - P(0 < Z < 0.25)$

$= 0.5000 - 0.0987$ (from table)

$= 0.4013$

Example 10.25

Find the area to the left of $Z = 1.5$

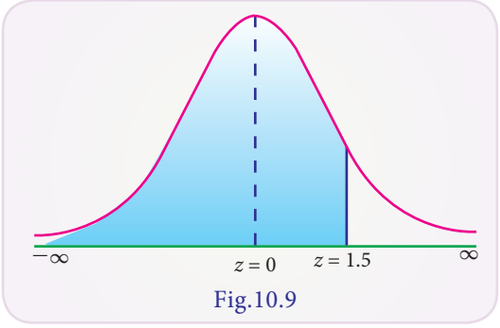


Fig.10.9

Solution:

$P(Z < 1.5) = P(-\infty < Z < 0) + P(0 < Z < 1.5)$

$= 0.5000 + 0.4332$ (from table)

$= 0.9332$

Example 10.26

Find the area between Z = -1 and $Z = 1.75$

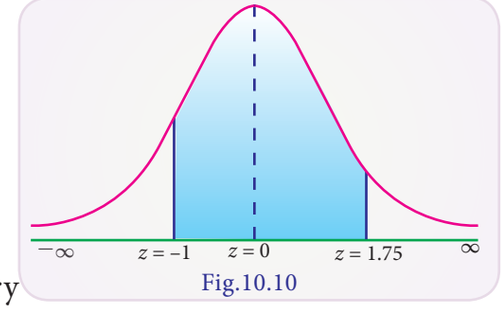


Fig.10.10

Solution:

$P(-1 < Z < 1.75)$

$= P(-1 < Z < 0) + P(0 < Z < 1.75)$

$= P(0 < Z < 1) + P(0 < Z < 1.75)$ by symmetry

$= 0.3413 + 0.4599 = 0.8012$

Ch10_11th Stats.indd 309 07/12/2021 12:20:34

Example 10.27

Find the maximum value of the p.d.f of the Normal distribution with mean 40 and standard derivation 10. Also find its points of inflection.

Solution:

$\mu = 40 \qquad \sigma=10$

Maximum value $= f(x) = \frac{1}{\sigma\sqrt{2\pi}} = \frac{1}{10\sqrt{2\pi}}$

Points of inflection $= \mu \pm \sigma$

$= 40 \pm 10$

$= 30$ and 50

Example 10.28

A Normal variable X has the mean 50 and the S.D 5. Find its mean deviation about mean and the quartile deviation.

Solution:

Given Mean $=\mu=50$, standard deviation $\sigma = 5$

Mean deviation above mean $= 0.8\ \sigma$

$= 0.8 \times 5 = 4$

Quartile deviation $= 0.6745\sigma$

$= 0.6745 \times 5$

$= 3.3725$

Example 10.29

Find the quartiles of the Normal distribution having mean 60 and S.D 10.

Solution:

$\mu=60$, $\sigma=10$

Let x_1 be the value such that the area from x_1 to μ is 25%

$\therefore \qquad P(x_1 < X < \mu) = 25\% = 0.25$

Ch10_11th Stats.indd 310 07/12/2021 12:20:35

$P(z_1 < Z < 0) = 0.25$ where $z_1 = \frac{x_1 - \mu}{\sigma} = \frac{x_1 - 60}{10}$

From Normal Probability table

$P(-0.675 < Z < 0) = 0.25015$

$\therefore \quad z_1 = -0.675$

$\therefore \quad \frac{x_1 - \mu}{\sigma} = -0.675$

$\frac{x_1 - 60}{10} = -0.675$

$X_1 - 60 = -6.75$

$X_1 = 53.25$ that is $Q_1 = 53.25$

Let x_2 be the value, so that

$\therefore \quad P(\mu < X < x_2) = 0.25$

$P(0 < Z < z_2) = 0.25$, where $z_2 = \frac{x_2 - \mu}{\sigma} = \frac{x_2 - 60}{10}$

From the table

$P(0 < Z < 0.675) = 0.25015$

$\therefore \quad Z_2 = 0.675$

$\frac{x_2 - 60}{10} = 0.675$

$X_2 = 66.75$ that is $Q_2 = 66.75$

Example 10.30

The height of the rose plants in a garden is Normally distributed with a mean 100cms. Given that 10% of the plants have height greater than104cm. Find (i) The S.D of the distribution (ii) The number of plants have height greater than 105cms if there were 200 plants in the garden.

Solution:

(i) The S.D of the distribution

Let X be the height of the rose plants that is Normally distributed.

Given: $P(100 < x < 104) = 40\ \%$

$P(0 < Z < 4/\sigma) = 0.40$ ... (1)

Ch10_11th Stats.indd 311 07/12/2021 12:20:36

But from table of area under Standard Normal curve

$P(0 < Z < 1.28) = 0.3997$ ... (2)

$\therefore$ From (1) and (2)

$$\frac{4}{\sigma} = 1.28$$

$$\sigma = 3.125$$

(ii) The number of plants with height greater than 105 cms if there were 200 plants in the garden.

$$X = 105$$

$$Z = \frac{x-\mu}{\sigma} = \frac{105-100}{3.125}$$

$$= \frac{5}{3.125}$$

$$= 1.6$$

$P(1.6 < Z < \infty) = 0.5\text{-}0.4452 = 0.0548$

Number of plants $= 200 \times 0.0548 = 10.9600 \approx 11$

$\therefore$ 11 plants have height greater than 105cms.

Example 10.31

Students of a class were given an aptitude test. Their marks were found to be normally distributed with mean 60 and standard deviation 5. What percentage of students scored (i) more than 60 marks (ii) less than 56 marks (iii) between 45 and 65 marks.

Solution:

Given mean $\mu = 60$ and standard deviation $\sigma = 5$

Standard normal variate $Z = \frac{x-\mu}{\sigma} = \frac{x-60}{5}$

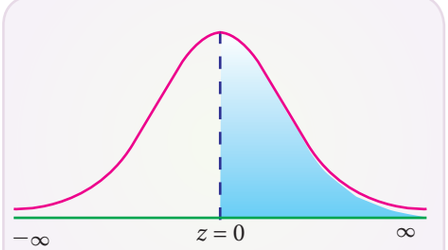


Fig.10.11

(i) P (more than 60) $= P(x > 60)$

$$= P\left(z > \frac{60-60}{5}\right)$$

$$= P(Z > 0)$$

$$= P(0 < Z < \infty)$$

$$= 0.5000$$

Student scored more than 60 marks = 0.5000 x 100

$$= 50\%$$

Ch10_11th Stats.indd 312 07/12/2021 12:20:37

(ii) P (less than 56 marks) $= P(X < 56)$

$$= P\left(Z < \frac{56-60}{5}\right)$$

$$= P(Z < -0.8)$$

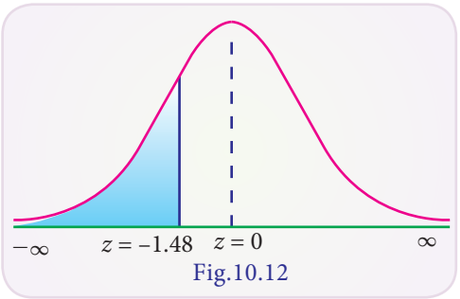


Fig.10.12

$$= P(-\infty < Z < 0) - P(-0.8 < Z < 0)$$

$$= 0.5000 - P(0 < Z < 0.8)$$

$$= 0.5000 - 0.2881$$

$$= 0.2119$$

$\therefore$ Number of students scored less than 50 marks $= 0.2119 \times 100$

$= 21.19\%$

(iii) P (between 45 and 65 marks) $= P(45 < x < 65)$

$$= P\left(\frac{45-60}{5} < z < \frac{65-60}{5}\right)$$

$$= P(-3 < z < 1)$$

$$= P(0 < Z < 3) + P(0 < Z < 1)$$

$$= 0.4986 + 0.3413$$

$$= 0.8399$$

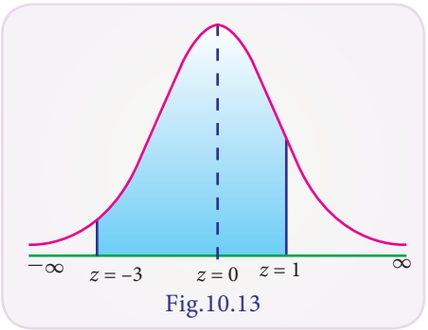


Fig.10.13

Number of students scored between 45 and 60 marks = 0.8399 x 100= 83.99%

Example 10.32

X has Normal distribution with mean 2 and standard deviation 3. Find the value of the variable x such that the probability of the interval from the mean to that value is 0.4115.

Solution:

Given $\mu = 2,\ \sigma = 3, z = \frac{x-\mu}{\sigma} = \frac{x-2}{3}$

Let $Z_1 = \frac{x_1 - 2}{3}$

We have $P(\mu < x < x_1) = 0.4115$

$P(0 < Z < z_1) = 0.4115$

But $P(0 < Z < 1.35) = 0.4115$ (from the Normal Probability table)

$\therefore \quad Z_1 = 1.35 \therefore \frac{x_1 - 2}{3} = 1.35$ (or) $x_1 = (1.35) \times 3 + 2 = 6.05$

Ch10_11th Stats.indd 313 07/12/2021 12:20:39

Example 10.33

In a Normal distribution 7% items are under 35 and 89% are under 63. Find its mean and standard deviation.

Solution:

$$Z = \frac{x - \mu}{\sigma}$$

We have $P(x < 35) = 7\% = 0.07$

If $z_1 =$ then $P(z_1 < Z < 0) = 0.50 - 0.07 = 0.43$

$\therefore z_1 = -1.48$ (from Normal Probability table)

$$\frac{35 - \mu}{\sigma} = -1.48$$

$$35 - \mu = -1.48\sigma \quad \ldots (1)$$

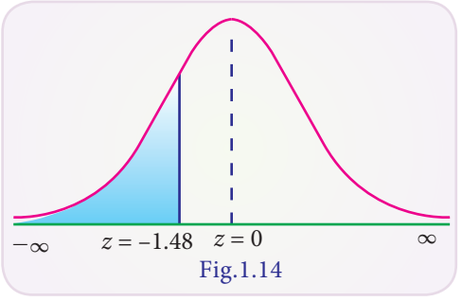


Fig.1.14

Also $P(X < 63) = 89\% = 0.89$

If $Z_2 =$ then $P(0 < Z < Z_2) = 0.89 - 0.50 = 0.39$

$Z_2 = 1.23$ (from Normal Probability table)

$$\frac{63 - \mu}{\sigma} = 1.23 \text{ (or) } 63 - \mu = 1.23\sigma \quad \ldots (2)$$

$$(2) - (1) \Rightarrow 28 = 2.71\sigma \text{ or } \sigma = \frac{28}{2.71} = 10.33$$

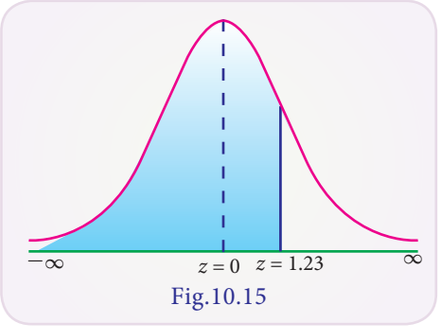


Fig.10.15

From (1) wehave

$$\mu = 35 + 1.48\sigma = 35 + 1.48\ (10.33)$$

$$\mu = 50.27$$

10.3 Fitting of Binomial, Poisson and Normal distributions

Introduction

Fitting of probability distribution to a series of observed data helps to predict the probability or to forecast the frequency of occurrence of the required variable in a certain desired interval.

There are many probability distributions of which some can be fitted more closely to the observed frequency of the data than others, depending on the characteristics of the variables. Therefore one needs to select a distribution that suits the data well.

Ch10_11th Stats.indd 314 07/12/2021 12:20:39

10.3.1 Fitting of Binomial Distribution

When a Binomial distribution is to be fitted to an observed data the following procedure is adopted:-

(i) Find mean $\overline{x} = \frac{\sum fx}{\sum f} = np$

(ii) Find $p = \frac{\overline{x}}{n}$

(iii) Find $q = 1 - p$

(iv) Write the probability mass function : $P(x) = nC_x p^x q^{n-x}$ $x = 0, 1, 2, \ldots n$

(v) Put $x = 0$; find $P(0) = nC_0 p^0 q^{n-0} = q^n$

(vi) Find the expected frequency of $X = 0$ i.e., $F(0) = N \times P(0)$, where $N = \sum f_i$

(vii) The other expected frequencies are obtained by using the recurrence formula

$F(x+1) = \frac{n-x}{x+1} \times \frac{p}{q} \times F(x)$

Example 10.34

A set of three similar coins are tossed 100 times with the following results

Number of heads	0	1	2	3
Frequency	36	40	22	2

Fit a binomial distribution and estimate the expected frequencies.

Solution :

x	f	fx
0	36	0
1	40	40
2	22	44
3	2	44
Total	100	90

(i) Mean $\overline{x} = \frac{\sum fx}{\sum f} = \frac{90}{100} = 0.9$

(ii) $p = \frac{\overline{x}}{n} = \frac{0.9}{3} = 0.3$

(iii) $q = 1 - p = 1 - 0.3 = 0.7$

(iv) $P(x) = nC_x p^x q^{n-x} = 3C_x \; 0.3_x \; 0.7^{3-x}$

Ch10_11th Stats.indd 315 07/12/2021 12:20:41

(v) $P(0) = 3C_0\ 0.3^0\ (0.7)^{3-0} = 0.7^3 = 0.343$

(vi) $F(0) = N \times P(0) = 100 \times 0.343 = 34.3$

(vii) $F(x+1) = \frac{n-x}{x+1} \times \frac{p}{q} \times F(x)$

$\therefore \quad F(1) = F(0+1) = \frac{3-0}{0+1} \times \frac{0.3}{0.7} \times 34.3 = 44.247$

$F(2) = F(1+1) = \frac{3-1}{1+1} \times \frac{0.3}{0.7} \times 44.247 = 19.03$

$F(3) = F(2+1) = \frac{3-2}{2+1} \times \frac{0.3}{0.7} \times 19.03 = 2.727$

Solution:

(i) The fitted binomial distribution is

$P(X = x) = 3C_x\, 0.3^x\, 0.7^{(3-x)}, \quad x = 0,1,2,3$

(ii) The expected frequencies are :

x	0	1	2	3	Total
Observed frequencies (O_i)	36	40	22	2	100
Expected Frequencies (E_i)	34	44	19	3	100

10.3.2 Fitting of Poisson Distribution

When a Poisson distribution is to be fitted to an observed data the following procedure is adopted:

(i) Find the mean: $\bar{x} = \frac{\Sigma fx}{\Sigma f}$

(ii) Poisson parameter = $\lambda = \bar{x}$

(iii) Probability mass function is: $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2\ldots.$

(iv) Put X = 0 and find $P(0) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda}$

(v) Expected frequency for $x = 0$ is $F(0) = N \times P(0)$, $\Sigma f_i = N$

(vi) Other expected frequencies can be found using

$F(x+1) = \frac{\lambda}{x+1} \times F(x)$ for $x = 0, 1, 2\ldots.$

Example 10.35

The following mistakes per page were observed in a book

Number of Mistakes (per page)	0	1	2	3	4
Number of pages	211	90	19	5	0

Fit a Poisson distribution and estimate the expected frequencies.

Ch10_11th Stats.indd 316 07/12/2021 12:20:43

Solution:

x	f	fx
0	211	0
1	90	90
2	19	38
3	5	15
4	0	0
Total	325	143

(i) Mean $\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{143}{325} = 0.44$

(ii) $\lambda = \bar{x} = \quad = 0.44$

(iii) $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-0.44}(0.44)^x}{x!}$

(iv) $P(0) \quad = \frac{e^{0.44} \times 0.44^0}{0!} = e^{-0.44} \quad = 0.6440$ (from the Poisson table)

(v) $F(0) \;=\; N \times P(0) \;= 325 \times 0.6440 = 209.43$

(vi) $F(x+1) = \frac{\lambda}{x+1} F(x)$

$F(1) \;=\; F(0+1) \;= \frac{0.44}{0+1} \times 209.43 = 92.15$

$F(2) \;=\; F(1+1) \;= \frac{0.44}{1+1} \times 92.15 = 20.27$

$F(3) \;=\; F(2+1) \;= \frac{0.44}{2+1} \times 20.27 = 2.973$

$F(4) \;=\; F(3+1) \;= \frac{0.44}{3+1} \times 2.97 = 0.33$

Result:

(1). Fitted Poisson distribution is $P(X = x) = \frac{e^{-0.44}0.44^x}{x!}$, $x = 0, 1, 2\ldots.$

(2). Expected frequencies are given below :

x	0	1	2	3	4	Total
Observed frequencies (O_i)	211	90	19	5	0	325
Expected Frequencies (E_i)	210	92	20	3	0	325

10.3.3 Fitting of Normal Distribution

In fitting a Normal distribution to the observed data, given in class intervals, we follow the following procedure:-

(i) Calculate μ and σ of the distribution

Ch10_11th Stats.indd 317 07/12/2021 12:20:45

(ii) Find x_i the lower class boundary

(iii) Find $z_i = \frac{x_i - \mu}{\sigma}$

(iv) Find Z_i (Zi) = $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz$

(v) Find $\triangle \phi$ (Z_i) = ϕ (Z_{i+1}) - ϕ (Z_i)

(vi) Find expected frequency $E_i = = N \triangle \phi(Z_i)$

Example 10.36

Find expected frequencies for the following data, if its calculated mean and standard deviation are 79.945 and 5.545.

Class	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
Frequency	3	21	150	335	326	135	26	4

Solution:

Given $\mu = 79.945, \sigma = 5.545$, and $N = 1000$

Hence the equation of Normal curve fitted to the data is

$$f(x) = \frac{1}{5.545\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-79.945}{5.545}\right)^2}$$

Theoretical Normal frequencies can be obtained as follows:

Class	Lower Class boundry (X_i)	$z_i = \frac{X_i - \mu}{\sigma}$	ϕ (Z_i) = $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz$	$= \phi$ (Z_{i+1}) $- \phi$ (Z_i)	Expected frequencies $N \triangle \phi(Z_i)$
Below 60	- ∞	- ∞	0	0.0001	0
60 - 65	60	-3.59693	0.001	0.0035	$3.5 \approx 4$
65 - 70	65	-2.69522	0.0036	0.0331	$33.1 \approx 33$
70 - 75	70	-1.79351	0.0367	0.15	150
75 - 80	75	-0.89179	0.1867	0.3173	$317.3 \approx 317$
80 - 85	80	0.009919	0.504	0.3146	$314.6 \approx 315$
85 - 90	85	0.911632	0.8186	0.1463	$146.3 \approx 146$
90 - 95	90	1.813345	0.9649	0.0318	$31.8 \approx 32$
95 – 100	95	2.715059	0.9967	0.0032	$3.2 \approx 3$
100 and Over	100	3.616772	0.9999		

Ch10_11th Stats.indd 318 07/12/2021 12:20:47

Points to Remember

- Binomial distribution is a discrete distribution.Its p.m.f is given by

$$P(X=x)=\begin{cases} nC_x p^x q^{n-x} & for\ x=0,1,2,...n \\ 0 & otherwise \end{cases}$$ where q=1-p

n,p are the parameters of the Binomial distribution.Then

(i)Mean = np; (ii)Variance = npq; (iii) Standard deviation = $\sqrt{npq}$

(iv)Skewness = $\frac{q-p}{\sqrt{npq}}$; (v)Kurtosis = $\frac{1-6pq}{npq}$

- Poisson distribution is a discrete distribution.Its p.m.f is given by

$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}$, x= 0, 1, 2, ...∞,

λ is the parameter of the Poisson distribution.Then

(i) Mean = λ = variance (ii) Standard deviation = $\sqrt{\lambda}$

(iii) Skewness = $\frac{1}{\sqrt{\lambda}}$ (iv) Kurtosis = $\frac{1}{\lambda}$

- Rectangular or Uniform distribution is a continuous distribution.The p.d.f isgiven by

$$f(x)=\begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & otherwise \end{cases}$$ a,b are the parameters of the rectangular distribution.Then

(i) Mean $\mu=\frac{a+b}{2}$ (ii) Variance $\sigma^2=\frac{(b-a)^2}{12}$

(iii) Median = $\frac{a+b}{2}$ (iv) Skewness = 0 (v)Kurtosis = $-\frac{6}{5}$

- A continuous random variable X with mean μ, and variance σ^2 as the parameters follows the Normal distribution. Its p.d.f is given by

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2}$$ where $-\infty<x<\infty$, $-\infty<\mu<\infty$ and $\sigma>0$

- In Normal distribution;

(i) Mean = Median = Mode = μ (ii) Modal height = $\frac{1}{\sigma\sqrt{2\pi}}$

(iii) Skewness β_1 = 0 (iv) Kurtosis β_2 = 3

- Area Properties of Normal Distribution

(i) P (- ∞<X<∞) = 1

(ii) P(μ - σ< X <μ + σ) = 0.6826

(iii) P (μ - 2σ< X < μ + 2σ) = 0.9544

(iv) P (μ - 3σ< X <μ + 3σ) = 0.9973

Ch10_11th Stats.indd 319 07/12/2021 12:20:48

EXERCISE – 10

I. Choose the best answer:

1. The mean and variance of a Bernoulli distribution with usual notations are
   (a) $p, \sqrt{pq}$ (b) p, pq (c) np, npq (d) λ

2. In getting a `6' in a throw of a die the Bernoulli distribution
   (a) $\left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{p-x}$ (b) $\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{1-x}$ (c) $\left(\frac{5}{6}\right)^x \left(\frac{1}{6}\right)^{1-x}$ (d) none

3. The mean and variance of $B(n, p)$ are
   (a) npq, np (b) $np, \sqrt{npq}$ (c) np, npq (d) $\sqrt{npq}, np$

4. In B (n, p) the probability `n' successes is
   (a) $nC_x p^x q^{n-x}$ (b) 1 (c) p^n (d) q^n

5. If $X \sim B(n_1, p)$ is independent of $Y \sim B(n_2, p)$ then $X + Y$ is distributed as
   (a) B $(n_1 + n_2, 2p)$ (b) B (n_1, p) (c) B $(n_1 + n_2, p)$ (d) B(n_2, p)

6. If the expectation of a Poisson variable is `1' then $P(x < 1)$ is
   (a) e^{-1} (b) $1-2e^{-1}$ (c) $1-\frac{5}{2e^{-1}}$ (d) none

7. Poisson distribution is a limiting case of Binomial distribution when
   (a) $n \to \infty$; $p \to 0$ and $np = \sqrt{m}$ (b) $n \to 0$; $p \to \infty$ and $p = 1/m$
   (c) $n \to \infty$; $p \to \infty$ and $np = m$ (d) $n \to \infty$; $p \to 0$ and $np = m$

8. Poisson distribution corresponds to
   (a) Rare events (b) Certain event
   (c) Impossible event (d) Almost sure event

9. If $\left(\frac{2}{3}+\frac{1}{3}\right)^9$ represents a Binomial distribution its standard deviation is
   (a) 2 (b) $\sqrt{2}$ (c) 6 (d) 3

10. In a Poisson distribution
    (a) Mean = Variance (b) Mean < Variance
    (c) Mean > Variance (d) Mean $\neq$ Variance

11. If X ~ U (-5, 5) then its p.d.f. f(x) in $-5 < x < 5$ is
    (a) $\frac{1}{5}$ (b) $\frac{1}{2 \times 5}$ (c) $\frac{1}{5^2}$ (d) $\frac{1}{2}$

Ch10_11th Stats.indd 320 07/12/2021 12:20:50

12. The Normal distribution is a limiting form of Binomial distribution if

(a) $n \longrightarrow \infty$ and $p \longrightarrow 0$ (b) $n \longrightarrow 0$ and $p \longrightarrow q$

(c) $n \longrightarrow \infty$ and $p \longrightarrow n$ (d) $n \longrightarrow \infty$ neither p nor q is small

13. Skewness and Kurtosis of $N(\mu, \sigma^2)$ are

(a) 0, 1 (b) 0, 3 (c) 0, 2 (d) 0, 0

14. In $N(\mu, \sigma^2)$ the value of $P(|X - \mu| < 2\sigma)$ is

(a) 0.9544 (b) 0.6826 (c) 0.9973 (d) 0.0027

15. If $X \sim N(6, 1.2)$ and $P(0 < Z < 1) = 0.3413$ then $P(4.8 < X < 7.2)$ is

(a) 0.3413 (b) 0.6587 (c) 0.6826 (d) 0.3174

II. Fill in the blanks:

16. Number of trials in a Bernoulli distribution is ----------

17. If the mean and variance of a binomial distribution are 8 and 4 respectively then $P(X = 1)$ is ----------

18. The trials in a Binomial distribution are ----------

19. If $X \sim U(a, b)$ then its variance is ----------

20. ---------- represents a Standard Normal distribution?

III. Answer shortly:

21. State the characteristics of a Bernoulli distribution.

22. State the conditions of a Binomial distribution.

23. Comment on: "the mean and the standard deviation of a Binomial distribution are 7 and 4".

24. For the Binomial distribution $(0.68 + 0.32)^{10}$ find the probability of two successes.

25. If on an average 8 ships out of 10 arrive safely at a port find the mean and standard deviation of the number of ships arriving safely out of 1600 ships.

26. Give any two examples of a Poisson distribution.

27. The variance of a Poisson distribution is 0.5. Find P(X=3) [e-0.5 = 0.6065]

28. State the characteristics of a Poisson distribution.

Ch10_11th Stats.indd 321 07/12/2021 12:20:50

29. If X is a Poisson variate with $P(X=1) = P(X = 2)$, find its mean.

30. Why U (a, b) is called a rectangular distribution?

31. If $X \sim$ U (-10, 10) find its mean and variance.

32. State the conditions for B (n, p) to become N (μ, σ^2)

33. State the limiting conditions for a Poisson distribution to become a Normal distribution.

34. Evaluate the value of P ($|X - \mu| < 3\sigma$), when $X \sim N(\mu, \sigma^2)$

35. Give a note about Standard Normal distribution.

36. State the properties concerned with the area of a Normal distribution.

IV. Answer in brief:

37. 10% of the screws manufactured by an automatic machine are found to be defective. 20 screws are selected at random find the probability that (i) exactly 2 are defective (ii) at most 3 are defective (iii)at least 2 are defective.

38. 20 wrist watches in a box of 100 are defective. If 10 watches are selected at random find the probability that (i) 10 are defective (ii) at most 3 are defective.

39. A manufacturer of television sets knows that an average of 5% of the product is defective. He sells television sets in consignment of 100 and guarantees that not more than 4 sets will be defective. What is the probability that a television set will fail to meet the guaranteed quality? [e-5 = 0.0067]

40. A factory employing a huge number of workers find that over a period of time, average absentee rate is three workers per shift. Calculate the probability that in a given shift (i) exactly 2 workers (ii) more than 4 workers will be absent?

41. The remuneration paid to 100 lecturers coaching for professional entrance examination are normally distributed with mean Rs. 700/- and standard deviation Rs. 50/-. Estimate the number of lectures whose remuneration will be (i) between Rs. 700/- and Rs. 720/- (ii) more than Rs. 750/-.

V. Calculate the following:

42. The skewness and kurtosis of a binomial distribution are 1/6 and – 11/36 respectively. Find the Binomial distribution.

Ch10_11th Stats.indd 322 07/12/2021 12:20:51

43. 4 dice are thrown 200 times. Fit a Binomial distribution using the following data and estimate the expected frequencies for getting the numbers 4, 5 or 6.

Getting 4,5,6	0	1	2	3	4	Total
Frequency	62	85	40	11	2	200

44. 100 car radios are inspected as they come off from the production line and number of defects per set is recorded as below:

No. of Defects	0	1	2	3	4
No. of sets	79	18	2	1	0

Fit a Poisson distribution and find the expected frequencies.

45. Assuming that one in 80 births are twins calculate the probability of 2 or more sets of twins on a day when 30 births occur. Compare the results obtained by using (i) the Binomial (ii) Poisson distributions.

46. Find the mean and standard deviation of the normal distribution in which 8% items are under 72 and 99% are over 53.

47. A normal distribution has mean 150 and S.D 10. Find the limits between which the middle 60% of the values lie.

48. A set of 4 coins are tossed 64 times. The Number of occurrence of heads tabulated as follows

Number of Heads	0	1	2	3	4
Number of times	3	15	23	17	6

Fit a Binomial distribution for the foresaid data and find the expected frequencies.

49. The Average number of fish caught per hour in a lake is 0.66. Find the probability to get (i) No fish (ii) 2 Fishes (iii) atmost 3 fishes (iv)atleast 1 fish in a period of 7 hours.

50. The Education department conducts a coaching programme. The scores got by the students in the examinations after the coaching programme were normally distributed and the Z scores for some of the students are given below

Rohit 1.1	Pavithra -2.00	David 0.0
Rahim 1.70	Priya -1.60	Fazeena -0.8

(1)Which of these students scored above the mean?

(2)Which of these students scored below the mean?

(3)If the mean score is 150 and S.D is 20. What is the score of each student?

Ch10_11th Stats.indd 323 07/12/2021 12:20:51

Answers:

I. 1. (b) 2. (b) 3. (c) 4. (c) 5. (c) 6. (a) 7. (d) 8. (a) 9. (b)

10. (a) 11. (b) 12. (d) 13. (b) 14. (a) 15. (c)

II. 16. 1 17. $\frac{1}{2^{12}}$ 18. Independent 19. $\frac{(b-a)^2}{12}$ 20. N (0, 1)

III. 24. $10C_2 0.32^2\ 0.68^8$ 25. Mean = 1280 SD = 16 27. 0.0126

29. Mean $\lambda = 2$ 31. Mean = 0; Variance = $\frac{100}{3}$ 34. 0.9973

IV. 37. (i) $0.9^{18} \times 1.90$ (ii) $0.9^{20} \times 7.1317$ (iii) $1\text{-}0.9^{20} \times 3.222$

38. (i) 0.2^{10} (ii) $(0.8)^{10}(8.19)$ 39. 0.5595 40. (i) 0.2240 (ii) 0.1847

41. (i) 16 (ii) 16

V. 43. (i) $p(x) = 4C_x\ (0.2575)^x\ (0.7425)^{4-x}$ (ii) 62, 84, 45,10, 1

44. (i) $p(x) = e^{-0.25}\ \frac{0.25^x}{x!}$ (ii) 78, 20, 2, 0, 0

45. (i) 0.05395 (ii) 0.05498 46. Mean = 101.12 Standard deviation = 20.65

47. (141.589, 158.411)

48. $p = 0.53125$

Number of Heads	0	1	2	3	4
Expected frequency	3	14	24	18	5

49. (i) 0.5168513, (ii) 0.1125702, (iii) 0.99530888, (iv) 0.990147

50. (i) Rohit, Rahim (ii) Pavithra, Priya, Fazeena

(iii)

Student	Rohit	Pavithra	David	Rahim	Priya	Fazeena
Mark	172	110	150	184	118	134

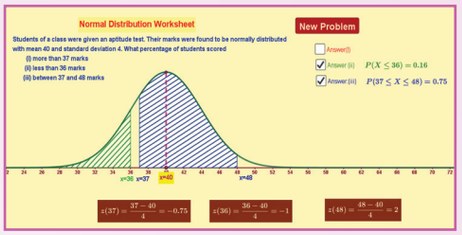
Ch10_11th Stats.indd 324 07/12/2021 12:20:52



ICT CORNER

NORMAL PROBABILITY DISTRIBUTION-EXERCISE

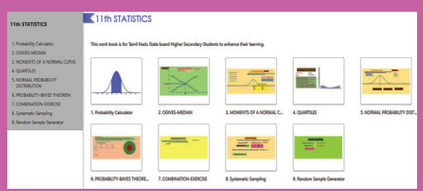
This activity normal probability distribution, is to learn a text book-based problem through interactive graphical representation.



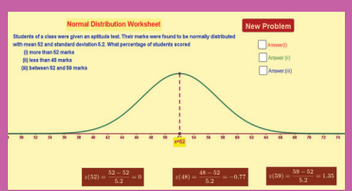
Steps:

- Open the browser and type the URL given (or) scan the QR code.
- GeoGebra work book called "**11th Standard Statistics**" will appear. In this several work sheets for statistics are given, open the worksheet named "**Normal Probability Distribution**"
- Exercise on book based normal probability distribution will appear. You can change the question as many times you like. For each problem the respective normal curve will appear with its mean. There are three check boxes to show answers and the respective shading on the curve.
- You work out each problem as learned in the class and click on the respective boxes to check your answer. After completing click on "**New Problem**" to get new problem.

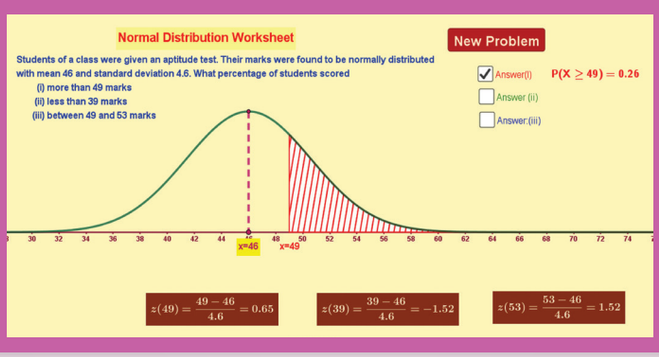
Step-1



Step-2



Step-3



Pictures are indicatives only*

URL:

https://ggbm.at/uqVhSJWZ

Ch10_11th Stats.indd 325 07/12/2021 12:20:54

11th standard – Statistics Practical

Introduction

Statistical tools are important for us in daily life. They are used in the analysis of data pertaining to various activities such as production, consumption, distribution, banking and insurance, trade, transport, etc. Practical work also gives students many opportunities to use their minds to adopt suitable statistical tools and methods on various types of analysis for the given sample data.

Objectives

- It facilitates comparison with similar data.
- Tabulation of data
- Compares the tabular data with diagrammatic representation of data
- Represents the data in a graph
- Presents the data in suitable diagrams
- Distinguishes diagrammatic and graphical representation of data
- Calculates the mathematical averages and the positional averages
- Computes quartiles, Deciles, Percentiles and interprets
- Measures the spread or dispersion
- Understands the theorems on probability and applies in problems
- Measures the Skewness
- Fittings Binomial and poisson distribution

Instructions To Students

Students must attend all the practical classes. They must also remember that there is a great degree of co-ordination between theory problems and practical problems.

- The following are some of the items that they must bring to the Practical Classes.
  - Practical observation note book
  - Practical record
  - Pencil sharpener
  - Eraser
  - A measuring scale
  - Graph sheets
  - Compass and protractor
  - Calculator
- Come prepared with theory part of the practical subject.
- They should submit the practical records periodically for correction and evaluation.
- They must maintain strict discipline and silence in the statistical laboratory.
- They should write the date and experiment number in their observation note books.

Practicals_11th_Stats.indd 326 2/3/2019 11:11:53 AM

The syllabus for 11th standard practicals are the following problems should be taken from the textbook examples or Exercises or relevant problems in real life situation. The question paper consists of two sections. Each section contains five questions. The students should answer four questions choosing two from each section.

Section A

1. Formation of Frequency Table – Bi-Variate and stem and leaf
2. Diagrammatic Representation of data – Pie Diagram and pareto diagram
3. Graphical Representation of data –Frequency polygon, frequency curve and Ogives
4. Measures of Central Tendency – Mean/ median/mode,GM and HM
5. Measures of Dispersion –CV, QD, Coefficient of Skewness and Box Whisker Plot

Section B

1. Simple problem - Probability and conditional Probability
2. Probability Mass Function and Probability Density Function - Computation of Mean and Variance for Single Random Variables
3. Fitting Binomial Distribution
4. Fitting Poisson distribution

The Outline of the each of the problems is as follows.

1. Aim or purpose:
2. Selection of the suitable statistical tool
3. The following procedure is to be followed for the Section A and Section B.

- Formula
- Substitution of data in the formula
- Calculation
- Result

Include graphs / diagrams wherever needed

Practicals_11th_Stats.indd 327 2/3/2019 11:11:53 AM

GLOSSARY	
Actuarial Science	நிபுணத்துவ அளவீட்டு அறிவியல்
Arithmetic Mean (AM)	கூட்டுச்சராசரி
Big data	பெருந்தரவுகள்
Binomial Distribution	ஈருறுப்புப் பரவல்
Binomial series	ஈருறுப்புத் தொடர்
Boxplot	கட்ட விளக்கப்படம்
Calculus	நுண்கணிதம்
Caption (Column heading)	அட்டவணையின் நிரல்தலைப்பு
Central Moments	மைய விலக்கப் பெருக்குத் தொகைகள்
Characteristic function	சிறப்பியல்புச் சார்பு
Chronological classification	காலம் சார் வகைப்படுத்துதல்
Coefficient of variation (CV)	மாறுபாட்டுக் கெழு
Combination	சேர்மானங்கள்
Component bar diagram	கூறுபட்டை விளக்கப்படம்
Conditional probability	நிபந்தனை நிகழ்தகவு
Continuous distribution	தொடர் நிகழ்வெண் பரவல்
Convenience sampling	ஏதுவான மாதிரி கணிப்பு முறை
Cumulative frequency curve (Ogive)	குவிவு நிகழ்வெண் வளைவரை (ஓஜைவ்)
Cumulative frequency distribution	குவிவு நிகழ்வெண் பரவல்
Deciles	பதின்மானங்கள்
Definite integral	வரையறுத்த தொகை
Differentiation	வகையிடல்
Discrete distribution	தனித்த நிகழ்வெண் பரவல்
Distribution function	பரவல் சார்பு
Enumerators	கணக்கெடுப்பாளர்
Factorial	வரிசை காரணிப்பெருக்கல், இயலெண் தொடர்பெருக்கல்
Frequency distribution	நிகழ்வெண் பரவல்
Frequency polygon	நிகழ்வெண் பன்முக வரைபடம்
Geographical classification	இடம் சார் வகைப்படுத்துதல்
Geometric Mean (GM)	பெருக்கு சராசரி
Hadoop	பெரும் தரவுகளைக் கையாளும் ஒரு மென்பொருள்
Harmonic Mean (HM)	இசைவுச் சராசரி
Histogram	பரவல் செவ்வகப் படம்
Independent events	சார்பற்ற நிகழ்ச்சிகள்
Integration	தொகையிடல்
Joint probability density function	இணைந்த நிகழ்தகவு அடர்த்திச் சார்பு
Joint probability mass function	இணைந்த நிகழ்தகவு திண்மைச் சார்பு
Judgement sampling	நோக்கமுடைய மாதிரி கணிப்பு முறை
Kurtosis	தட்டை அளவை
Marginal probability density function	இறுதிநிலை நிகழ்தகவு அடர்த்திச் சார்பு
Marginal probability mass function	இறுதிநிலை நிகழ்தகவு திண்மைச் சார்பு
Mathematical Expectation	கணித எதிர்பார்த்தல்
Mathematical probability	கணித நிகழ்தகவு (முந்தைய அணுகுமுறை)
Mean deviation	சராசரி விலக்கம்
Measurement scales	அளவீட்டு அளவைகள்
Measures of central tendency	மையப்போக்கு அளவைகள்
Measures of dispersion	சிதறல் அளவைகள்
Median	இடைநிலை அளவு

Glosary_11th_Stats.indd 328 07/12/2021 12:24:32

Mode	முகடு
Moment Generating Function(MGF)	விலக்கப் பெருக்கத்தொகை உருவாக்கும் சார்பு
Multiple bar diagram	பலகட்ட பட்டை விளக்கப்படம்
Mutually exclusive events	ஒன்றையொன்று விலக்கும் நிகழ்ச்சிகள்
Normal Distribution	இயல்நிலை பரவல்
Parameter	பண்பளவை
Pareto diagram	பெரிட்டோ வரைபடம்
Percentage bar diagram	விழுக்காடு பட்டை விளக்கப்படம்
Percentiles	நூற்றுமானங்கள்
Permutation	வரிசை மாற்றங்கள்
Pictogram	உருவ விளக்கப்படம்
Pie diagram	வட்ட விளக்கப்படம்
Poisson Distribution	பாய்சான் பரவல்
Population	முழுமைத் தொகுதி
Primary data	முதல்நிலை தரவுகள்
Probability density function (pdf)	நகழ்தகவு அடர்த்திச் சார்பு
Probability mass function (pmf)	நிகழ்தகவு திண்மைச் சார்பு
Probability sampling	நிகழ்தகவு மாதிரி கணிப்பு முறை
Qualitative classification	பண்பு சார் வகைப்படுத்துதல்
Quantitative classification	அளவின் மூலம் வகைப்படுத்துதல்
Quartiles	கால்மானங்கள்
Questionnaire	வினாவிடை பட்டியல்
Quota sampling	ஒதுக்கீட்டு மாதிரி கணிப்பு முறை
Random experiment	வாய்ப்பு சோதனை
Random variable	வாய்ப்பு மாறிகள்
Raw Moments	விலக்கப் பெருக்குத் தொகைகள்
Rectangular or Uniform Distribution	செவ்வக பரவல், சீரான பரவல்
Sample	மாதிரி, கூறு
Sampling	மாதிரி கணிப்பு
Sampling Error	மாதிரி கணிப்புப் பிழை
Secondary data	இரண்டாம் நிலை தரவுகள்
Simple bar diagram	எளிய பட்டை விளக்கப்படம்
Simple random sampling	எளிய வாய்ப்பு மாதிரி கணிப்பு முறை
Skewness	கோட்ட அளவை
Snowball sampling	பனிபந்து மாதிரி கணிப்பு முறை
Standard Deviation (SD)	திட்ட விலக்கம்
Standard normal variate Z	Z என்பது திட்ட இயல்நிலை மாறி
Statistical probability	புள்ளியியல் நிகழ்தகவு (பிந்தைய அணுகுமுறை)
Stem and Leaf Plot	தண்டு இலை வரைபடம்
Stratified random sampling	படுகை முறை மாதிரி கணிப்பு முறை
Stub (Row heading)	அட்டவணையின் நிரைதலைப்பு
Systematic random sampling	முறை சார்ந்த மாதிரி கணிப்பு முறை
Tabulation of data	தரவுகளை அட்டவணையிடுதல்
Variance	விலக்க வர்க்க சராசரி, மாறுபாடுட்டு அளவை

Glosary_11th_Stats.indd 329 07/12/2021 12:24:32

www.tnschoolbooks.in

LOGARITHM TABLE

											Mean Difference								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1.0	0.0000	0.0043	0.0086	0.0128	0.0170	0.0212	0.0253	0.0294	0.0334	0.0374	4	8	12	17	21	25	29	33	37
1.1	0.0414	0.0453	0.0492	0.0531	0.0569	0.0607	0.0645	0.0682	0.0719	0.0755	4	8	11	15	19	23	26	30	34
1.2	0.0792	0.0828	0.0864	0.0899	0.0934	0.0969	0.1004	0.1038	0.1072	0.1106	3	7	10	14	17	21	24	28	31
1.3	0.1139	0.1173	0.1206	0.1239	0.1271	0.1303	0.1335	0.1367	0.1399	0.1430	3	6	10	13	16	19	23	26	29
1.4	0.1461	0.1492	0.1523	0.1553	0.1584	0.1614	0.1644	0.1673	0.1703	0.1732	3	6	9	12	15	18	21	24	27
1.5	0.1761	0.1790	0.1818	0.1847	0.1875	0.1903	0.1931	0.1959	0.1987	0.2014	3	6	8	11	14	17	20	22	25
1.6	0.2041	0.2068	0.2095	0.2122	0.2148	0.2175	0.2201	0.2227	0.2253	0.2279	3	5	8	11	13	16	18	21	24
1.7	0.2304	0.2330	0.2355	0.2380	0.2405	0.2430	0.2455	0.2480	0.2504	0.2529	2	5	7	10	12	15	17	20	22
1.8	0.2553	0.2577	0.2601	0.2625	0.2648	0.2672	0.2695	0.2718	0.2742	0.2765	2	5	7	9	12	14	16	19	21
1.9	0.2788	0.2810	0.2833	0.2856	0.2878	0.2900	0.2923	0.2945	0.2967	0.2989	2	4	7	9	11	13	16	18	20
2.0	0.3010	0.3032	0.3054	0.3075	0.3096	0.3118	0.3139	0.3160	0.3181	0.3201	2	4	6	8	11	13	15	17	19
2.1	0.3222	0.3243	0.3263	0.3284	0.3304	0.3324	0.3345	0.3365	0.3385	0.3404	2	4	6	8	10	12	14	16	18
2.2	0.3424	0.3444	0.3464	0.3483	0.3502	0.3522	0.3541	0.3560	0.3579	0.3598	2	4	6	8	10	12	14	15	17
2.3	0.3617	0.3636	0.3655	0.3674	0.3692	0.3711	0.3729	0.3747	0.3766	0.3784	2	4	6	7	9	11	13	15	17
2.4	0.3802	0.3820	0.3838	0.3856	0.3874	0.3892	0.3909	0.3927	0.3945	0.3962	2	4	5	7	9	11	12	14	16
2.5	0.3979	0.3997	0.4014	0.4031	0.4048	0.4065	0.4082	0.4099	0.4116	0.4133	2	3	5	7	9	10	12	14	15
2.6	0.4150	0.4166	0.4183	0.4200	0.4216	0.4232	0.4249	0.4265	0.4281	0.4298	2	3	5	7	8	10	11	13	15
2.7	0.4314	0.4330	0.4346	0.4362	0.4378	0.4393	0.4409	0.4425	0.4440	0.4456	2	3	5	6	8	9	11	13	14
2.8	0.4472	0.4487	0.4502	0.4518	0.4533	0.4548	0.4564	0.4579	0.4594	0.4609	2	3	5	6	8	9	11	12	14
2.9	0.4624	0.4639	0.4654	0.4669	0.4683	0.4698	0.4713	0.4728	0.4742	0.4757	1	3	4	6	7	9	10	12	13
3.0	0.4771	0.4786	0.4800	0.4814	0.4829	0.4843	0.4857	0.4871	0.4886	0.4900	1	3	4	6	7	9	10	11	13
3.1	0.4914	0.4928	0.4942	0.4955	0.4969	0.4983	0.4997	0.5011	0.5024	0.5038	1	3	4	6	7	8	10	11	12
3.2	0.5051	0.5065	0.5079	0.5092	0.5105	0.5119	0.5132	0.5145	0.5159	0.5172	1	3	4	5	7	8	9	11	12
3.3	0.5185	0.5198	0.5211	0.5224	0.5237	0.5250	0.5263	0.5276	0.5289	0.5302	1	3	4	5	6	8	9	10	12
3.4	0.5315	0.5328	0.5340	0.5353	0.5366	0.5378	0.5391	0.5403	0.5416	0.5428	1	3	4	5	6	8	9	10	11
3.5	0.5441	0.5453	0.5465	0.5478	0.5490	0.5502	0.5514	0.5527	0.5539	0.5551	1	2	4	5	6	7	9	10	11
3.6	0.5563	0.5575	0.5587	0.5599	0.5611	0.5623	0.5635	0.5647	0.5658	0.5670	1	2	4	5	6	7	8	10	11
3.7	0.5682	0.5694	0.5705	0.5717	0.5729	0.5740	0.5752	0.5763	0.5775	0.5786	1	2	3	5	6	7	8	9	10
3.8	0.5798	0.5809	0.5821	0.5832	0.5843	0.5855	0.5866	0.5877	0.5888	0.5899	1	2	3	5	6	7	8	9	10
3.9	0.5911	0.5922	0.5933	0.5944	0.5955	0.5966	0.5977	0.5988	0.5999	0.6010	1	2	3	4	5	7	8	9	10
4.0	0.6021	0.6031	0.6042	0.6053	0.6064	0.6075	0.6085	0.6096	0.6107	0.6117	1	2	3	4	5	6	8	9	10
4.1	0.6128	0.6138	0.6149	0.6160	0.6170	0.6180	0.6191	0.6201	0.6212	0.6222	1	2	3	4	5	6	7	8	9
4.2	0.6232	0.6243	0.6253	0.6263	0.6274	0.6284	0.6294	0.6304	0.6314	0.6325	1	2	3	4	5	6	7	8	9
4.3	0.6335	0.6345	0.6355	0.6365	0.6375	0.6385	0.6395	0.6405	0.6415	0.6425	1	2	3	4	5	6	7	8	9
4.4	0.6435	0.6444	0.6454	0.6464	0.6474	0.6484	0.6493	0.6503	0.6513	0.6522	1	2	3	4	5	6	7	8	9
4.5	0.6532	0.6542	0.6551	0.6561	0.6571	0.6580	0.6590	0.6599	0.6609	0.6618	1	2	3	4	5	6	7	8	9
4.6	0.6628	0.6637	0.6646	0.6656	0.6665	0.6675	0.6684	0.6693	0.6702	0.6712	1	2	3	4	5	6	7	7	8
4.7	0.6721	0.6730	0.6739	0.6749	0.6758	0.6767	0.6776	0.6785	0.6794	0.6803	1	2	3	4	5	5	6	7	8
4.8	0.6812	0.6821	0.6830	0.6839	0.6848	0.6857	0.6866	0.6875	0.6884	0.6893	1	2	3	4	4	5	6	7	8
4.9	0.6902	0.6911	0.6920	0.6928	0.6937	0.6946	0.6955	0.6964	0.6972	0.6981	1	2	3	4	4	5	6	7	8
5.0	0.6990	0.6998	0.7007	0.7016	0.7024	0.7033	0.7042	0.7050	0.7059	0.7067	1	2	3	3	4	5	6	7	8
5.1	0.7076	0.7084	0.7093	0.7101	0.7110	0.7118	0.7126	0.7135	0.7143	0.7152	1	2	3	3	4	5	6	7	8
5.2	0.7160	0.7168	0.7177	0.7185	0.7193	0.7202	0.7210	0.7218	0.7226	0.7235	1	2	2	3	4	5	6	7	7
5.3	0.7243	0.7251	0.7259	0.7267	0.7275	0.7284	0.7292	0.7300	0.7308	0.7316	1	2	2	3	4	5	6	6	7
5.4	0.7324	0.7332	0.7340	0.7348	0.7356	0.7364	0.7372	0.7380	0.7388	0.7396	1	2	2	3	4	5	6	6	7

LogTable_11th_Stats.indd 330 2/3/2019 11:26:36 AM

www.tnschoolbooks.in

LOGARITHM TABLE

											Mean Difference								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
5.5	0.7404	0.7412	0.7419	0.7427	0.7435	0.7443	0.7451	0.7459	0.7466	0.7474	1	2	2	3	4	5	5	6	7
5.6	0.7482	0.7490	0.7497	0.7505	0.7513	0.7520	0.7528	0.7536	0.7543	0.7551	1	2	2	3	4	5	5	6	7
5.7	0.7559	0.7566	0.7574	0.7582	0.7589	0.7597	0.7604	0.7612	0.7619	0.7627	1	2	2	3	4	5	5	6	7
5.8	0.7634	0.7642	0.7649	0.7657	0.7664	0.7672	0.7679	0.7686	0.7694	0.7701	1	1	2	3	4	4	5	6	7
5.9	0.7709	0.7716	0.7723	0.7731	0.7738	0.7745	0.7752	0.7760	0.7767	0.7774	1	1	2	3	4	4	5	6	7
6.0	0.7782	0.7789	0.7796	0.7803	0.7810	0.7818	0.7825	0.7832	0.7839	0.7846	1	1	2	3	4	4	5	6	6
6.1	0.7853	0.7860	0.7868	0.7875	0.7882	0.7889	0.7896	0.7903	0.7910	0.7917	1	1	2	3	4	4	5	6	6
6.2	0.7924	0.7931	0.7938	0.7945	0.7952	0.7959	0.7966	0.7973	0.7980	0.7987	1	1	2	3	3	4	5	6	6
6.3	0.7993	0.8000	0.8007	0.8014	0.8021	0.8028	0.8035	0.8041	0.8048	0.8055	1	1	2	3	3	4	5	5	6
6.4	0.8062	0.8069	0.8075	0.8082	0.8089	0.8096	0.8102	0.8109	0.8116	0.8122	1	1	2	3	3	4	5	5	6
6.5	0.8129	0.8136	0.8142	0.8149	0.8156	0.8162	0.8169	0.8176	0.8182	0.8189	1	1	2	3	3	4	5	5	6
6.6	0.8195	0.8202	0.8209	0.8215	0.8222	0.8228	0.8235	0.8241	0.8248	0.8254	1	1	2	3	3	4	5	5	6
6.7	0.8261	0.8267	0.8274	0.8280	0.8287	0.8293	0.8299	0.8306	0.8312	0.8319	1	1	2	3	3	4	5	5	6
6.8	0.8325	0.8331	0.8338	0.8344	0.8351	0.8357	0.8363	0.8370	0.8376	0.8382	1	1	2	3	3	4	4	5	6
6.9	0.8388	0.8395	0.8401	0.8407	0.8414	0.8420	0.8426	0.8432	0.8439	0.8445	1	1	2	2	3	4	4	5	6
7.0	0.8451	0.8457	0.8463	0.8470	0.8476	0.8482	0.8488	0.8494	0.8500	0.8506	1	1	2	2	3	4	4	5	6
7.1	0.8513	0.8519	0.8525	0.8531	0.8537	0.8543	0.8549	0.8555	0.8561	0.8567	1	1	2	2	3	4	4	5	5
7.2	0.8573	0.8579	0.8585	0.8591	0.8597	0.8603	0.8609	0.8615	0.8621	0.8627	1	1	2	2	3	4	4	5	5
7.3	0.8633	0.8639	0.8645	0.8651	0.8657	0.8663	0.8669	0.8675	0.8681	0.8686	1	1	2	2	3	4	4	5	5
7.4	0.8692	0.8698	0.8704	0.8710	0.8716	0.8722	0.8727	0.8733	0.8739	0.8745	1	1	2	2	3	4	4	5	5
7.5	0.8751	0.8756	0.8762	0.8768	0.8774	0.8779	0.8785	0.8791	0.8797	0.8802	1	1	2	2	3	3	4	5	5
7.6	0.8808	0.8814	0.8820	0.8825	0.8831	0.8837	0.8842	0.8848	0.8854	0.8859	1	1	2	2	3	3	4	5	5
7.7	0.8865	0.8871	0.8876	0.8882	0.8887	0.8893	0.8899	0.8904	0.8910	0.8915	1	1	2	2	3	3	4	4	5
7.8	0.8921	0.8927	0.8932	0.8938	0.8943	0.8949	0.8954	0.8960	0.8965	0.8971	1	1	2	2	3	3	4	4	5
7.9	0.8976	0.8982	0.8987	0.8993	0.8998	0.9004	0.9009	0.9015	0.9020	0.9025	1	1	2	2	3	3	4	4	5
8.0	0.9031	0.9036	0.9042	0.9047	0.9053	0.9058	0.9063	0.9069	0.9074	0.9079	1	1	2	2	3	3	4	4	5
8.1	0.9085	0.9090	0.9096	0.9101	0.9106	0.9112	0.9117	0.9122	0.9128	0.9133	1	1	2	2	3	3	4	4	5
8.2	0.9138	0.9143	0.9149	0.9154	0.9159	0.9165	0.9170	0.9175	0.9180	0.9186	1	1	2	2	3	3	4	4	5
8.3	0.9191	0.9196	0.9201	0.9206	0.9212	0.9217	0.9222	0.9227	0.9232	0.9238	1	1	2	2	3	3	4	4	5
8.4	0.9243	0.9248	0.9253	0.9258	0.9263	0.9269	0.9274	0.9279	0.9284	0.9289	1	1	2	2	3	3	4	4	5
8.5	0.9294	0.9299	0.9304	0.9309	0.9315	0.9320	0.9325	0.9330	0.9335	0.9340	1	1	2	2	3	3	4	4	5
8.6	0.9345	0.9350	0.9355	0.9360	0.9365	0.9370	0.9375	0.9380	0.9385	0.9390	1	1	2	2	3	3	4	4	5
8.7	0.9395	0.9400	0.9405	0.9410	0.9415	0.9420	0.9425	0.9430	0.9435	0.9440	0	1	1	2	2	3	3	4	4
8.8	0.9445	0.9450	0.9455	0.9460	0.9465	0.9469	0.9474	0.9479	0.9484	0.9489	0	1	1	2	2	3	3	4	4
8.9	0.9494	0.9499	0.9504	0.9509	0.9513	0.9518	0.9523	0.9528	0.9533	0.9538	0	1	1	2	2	3	3	4	4
9.0	0.9542	0.9547	0.9552	0.9557	0.9562	0.9566	0.9571	0.9576	0.9581	0.9586	0	1	1	2	2	3	3	4	4
9.1	0.9590	0.9595	0.9600	0.9605	0.9609	0.9614	0.9619	0.9624	0.9628	0.9633	0	1	1	2	2	3	3	4	4
9.2	0.9638	0.9643	0.9647	0.9652	0.9657	0.9661	0.9666	0.9671	0.9675	0.9680	0	1	1	2	2	3	3	4	4
9.3	0.9685	0.9689	0.9694	0.9699	0.9703	0.9708	0.9713	0.9717	0.9722	0.9727	0	1	1	2	2	3	3	4	4
9.4	0.9731	0.9736	0.9741	0.9745	0.9750	0.9754	0.9759	0.9763	0.9768	0.9773	0	1	1	2	2	3	3	4	4
9.5	0.9777	0.9782	0.9786	0.9791	0.9795	0.9800	0.9805	0.9809	0.9814	0.9818	0	1	1	2	2	3	3	4	4
9.6	0.9823	0.9827	0.9832	0.9836	0.9841	0.9845	0.9850	0.9854	0.9859	0.9863	0	1	1	2	2	3	3	4	4
9.7	0.9868	0.9872	0.9877	0.9881	0.9886	0.9890	0.9894	0.9899	0.9903	0.9908	0	1	1	2	2	3	3	4	4
9.8	0.9912	0.9917	0.9921	0.9926	0.9930	0.9934	0.9939	0.9943	0.9948	0.9952	0	1	1	2	2	3	3	4	4
9.9	0.9956	0.9961	0.9965	0.9969	0.9974	0.9978	0.9983	0.9987	0.9991	0.9996	0	1	1	2	2	3	3	3	4

LogTable_11th_Stats.indd 331 2/3/2019 11:26:36 AM

www.tnschoolbooks.in

ANTI LOGARITHM TABLE

											Mean Difference								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
0.00	1.000	1.002	1.005	1.007	1.009	1.012	1.014	1.016	1.019	1.021	0	0	1	1	1	1	2	2	2
0.01	1.023	1.026	1.028	1.030	1.033	1.035	1.038	1.040	1.042	1.045	0	0	1	1	1	1	2	2	2
0.02	1.047	1.050	1.052	1.054	1.057	1.059	1.062	1.064	1.067	1.069	0	0	1	1	1	1	2	2	2
0.03	1.072	1.074	1.076	1.079	1.081	1.084	1.086	1.089	1.091	1.094	0	0	1	1	1	1	2	2	2
0.04	1.096	1.099	1.102	1.104	1.107	1.109	1.112	1.114	1.117	1.119	0	1	1	1	1	2	2	2	2
0.05	1.122	1.125	1.127	1.130	1.132	1.135	1.138	1.140	1.143	1.146	0	1	1	1	1	2	2	2	2
0.06	1.148	1.151	1.153	1.156	1.159	1.161	1.164	1.167	1.169	1.172	0	1	1	1	1	2	2	2	2
0.07	1.175	1.178	1.180	1.183	1.186	1.189	1.191	1.194	1.197	1.199	0	1	1	1	1	2	2	2	2
0.08	1.202	1.205	1.208	1.211	1.213	1.216	1.219	1.222	1.225	1.227	0	1	1	1	1	2	2	2	3
0.09	1.230	1.233	1.236	1.239	1.242	1.245	1.247	1.250	1.253	1.256	0	1	1	1	1	2	2	2	3
0.10	1.259	1.262	1.265	1.268	1.271	1.274	1.276	1.279	1.282	1.285	0	1	1	1	1	2	2	2	3
0.11	1.288	1.291	1.294	1.297	1.300	1.303	1.306	1.309	1.312	1.315	0	1	1	1	2	2	2	2	3
0.12	1.318	1.321	1.324	1.327	1.330	1.334	1.337	1.340	1.343	1.346	0	1	1	1	2	2	2	2	3
0.13	1.349	1.352	1.355	1.358	1.361	1.365	1.368	1.371	1.374	1.377	0	1	1	1	2	2	2	3	3
0.14	1.380	1.384	1.387	1.390	1.393	1.396	1.400	1.403	1.406	1.409	0	1	1	1	2	2	2	3	3
0.15	1.413	1.416	1.419	1.422	1.426	1.429	1.432	1.435	1.439	1.442	0	1	1	1	2	2	2	3	3
0.16	1.445	1.449	1.452	1.455	1.459	1.462	1.466	1.469	1.472	1.476	0	1	1	1	2	2	2	3	3
0.17	1.479	1.483	1.486	1.489	1.493	1.496	1.500	1.503	1.507	1.510	0	1	1	1	2	2	2	3	3
0.18	1.514	1.517	1.521	1.524	1.528	1.531	1.535	1.538	1.542	1.545	0	1	1	1	2	2	2	3	3
0.19	1.549	1.552	1.556	1.560	1.563	1.567	1.570	1.574	1.578	1.581	0	1	1	1	2	2	3	3	3
0.20	1.585	1.589	1.592	1.596	1.600	1.603	1.607	1.611	1.614	1.618	0	1	1	1	2	2	3	3	3
0.21	1.622	1.626	1.629	1.633	1.637	1.641	1.644	1.648	1.652	1.656	0	1	1	2	2	2	3	3	3
0.22	1.660	1.663	1.667	1.671	1.675	1.679	1.683	1.687	1.690	1.694	0	1	1	2	2	2	3	3	3
0.23	1.698	1.702	1.706	1.710	1.714	1.718	1.722	1.726	1.730	1.734	0	1	1	2	2	2	3	3	4
0.24	1.738	1.742	1.746	1.750	1.754	1.758	1.762	1.766	1.770	1.774	0	1	1	2	2	2	3	3	4
0.25	1.778	1.782	1.786	1.791	1.795	1.799	1.803	1.807	1.811	1.816	0	1	1	2	2	2	3	3	4
0.26	1.820	1.824	1.828	1.832	1.837	1.841	1.845	1.849	1.854	1.858	0	1	1	2	2	3	3	3	4
0.27	1.862	1.866	1.871	1.875	1.879	1.884	1.888	1.892	1.897	1.901	0	1	1	2	2	3	3	3	4
0.28	1.905	1.910	1.914	1.919	1.923	1.928	1.932	1.936	1.941	1.945	0	1	1	2	2	3	3	4	4
0.29	1.950	1.954	1.959	1.963	1.968	1.972	1.977	1.982	1.986	1.991	0	1	1	2	2	3	3	4	4
0.30	1.995	2.000	2.004	2.009	2.014	2.018	2.023	2.028	2.032	2.037	0	1	1	2	2	3	3	4	4
0.31	2.042	2.046	2.051	2.056	2.061	2.065	2.070	2.075	2.080	2.084	0	1	1	2	2	3	3	4	4
0.32	2.089	2.094	2.099	2.104	2.109	2.113	2.118	2.123	2.128	2.133	0	1	1	2	2	3	3	4	4
0.33	2.138	2.143	2.148	2.153	2.158	2.163	2.168	2.173	2.178	2.183	0	1	1	2	2	3	3	4	4
0.34	2.188	2.193	2.198	2.203	2.208	2.213	2.218	2.223	2.228	2.234	1	1	2	2	3	3	4	4	5
0.35	2.239	2.244	2.249	2.254	2.259	2.265	2.270	2.275	2.280	2.286	1	1	2	2	3	3	4	4	5
0.36	2.291	2.296	2.301	2.307	2.312	2.317	2.323	2.328	2.333	2.339	1	1	2	2	3	3	4	4	5
0.37	2.344	2.350	2.355	2.360	2.366	2.371	2.377	2.382	2.388	2.393	1	1	2	2	3	3	4	4	5
0.38	2.399	2.404	2.410	2.415	2.421	2.427	2.432	2.438	2.443	2.449	1	1	2	2	3	3	4	4	5
0.39	2.455	2.460	2.466	2.472	2.477	2.483	2.489	2.495	2.500	2.506	1	1	2	2	3	3	4	5	5
0.40	2.512	2.518	2.523	2.529	2.535	2.541	2.547	2.553	2.559	2.564	1	1	2	2	3	4	4	5	5
0.41	2.570	2.576	2.582	2.588	2.594	2.600	2.606	2.612	2.618	2.624	1	1	2	2	3	4	4	5	5
0.42	2.630	2.636	2.642	2.649	2.655	2.661	2.667	2.673	2.679	2.685	1	1	2	2	3	4	4	5	6
0.43	2.692	2.698	2.704	2.710	2.716	2.723	2.729	2.735	2.742	2.748	1	1	2	3	3	4	4	5	6
0.44	2.754	2.761	2.767	2.773	2.780	2.786	2.793	2.799	2.805	2.812	1	1	2	3	3	4	4	5	6
0.45	2.818	2.825	2.831	2.838	2.844	2.851	2.858	2.864	2.871	2.877	1	1	2	3	3	4	5	5	6
0.46	2.884	2.891	2.897	2.904	2.911	2.917	2.924	2.931	2.938	2.944	1	1	2	3	3	4	5	5	6
0.47	2.951	2.958	2.965	2.972	2.979	2.985	2.992	2.999	3.006	3.013	1	1	2	3	3	4	5	5	6
0.48	3.020	3.027	3.034	3.041	3.048	3.055	3.062	3.069	3.076	3.083	1	1	2	3	4	4	5	6	6
0.49	3.090	3.097	3.105	3.112	3.119	3.126	3.133	3.141	3.148	3.155	1	1	2	3	4	4	5	6	6

LogTable_11th_Stats.indd 332 2/3/2019 11:26:37 AM

www.tnschoolbooks.in

ANTI LOGARITHM TABLE

											Mean Difference								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
0.50	3.162	3.170	3.177	3.184	3.192	3.199	3.206	3.214	3.221	3.228	1	1	2	3	4	4	5	6	7
0.51	3.236	3.243	3.251	3.258	3.266	3.273	3.281	3.289	3.296	3.304	1	2	2	3	4	5	5	6	7
0.52	3.311	3.319	3.327	3.334	3.342	3.350	3.357	3.365	3.373	3.381	1	2	2	3	4	5	5	6	7
0.53	3.388	3.396	3.404	3.412	3.420	3.428	3.436	3.443	3.451	3.459	1	2	2	3	4	5	6	6	7
0.54	3.467	3.475	3.483	3.491	3.499	3.508	3.516	3.524	3.532	3.540	1	2	2	3	4	5	6	6	7
0.55	3.548	3.556	3.565	3.573	3.581	3.589	3.597	3.606	3.614	3.622	1	2	2	3	4	5	6	7	7
0.56	3.631	3.639	3.648	3.656	3.664	3.673	3.681	3.690	3.698	3.707	1	2	3	3	4	5	6	7	8
0.57	3.715	3.724	3.733	3.741	3.750	3.758	3.767	3.776	3.784	3.793	1	2	3	3	4	5	6	7	8
0.58	3.802	3.811	3.819	3.828	3.837	3.846	3.855	3.864	3.873	3.882	1	2	3	4	4	5	6	7	8
0.59	3.890	3.899	3.908	3.917	3.926	3.936	3.945	3.954	3.963	3.972	1	2	3	4	5	5	6	7	8
0.60	3.981	3.990	3.999	4.009	4.018	4.027	4.036	4.046	4.055	4.064	1	2	3	4	5	6	6	7	8
0.61	4.074	4.083	4.093	4.102	4.111	4.121	4.130	4.140	4.150	4.159	1	2	3	4	5	6	7	8	9
0.62	4.169	4.178	4.188	4.198	4.207	4.217	4.227	4.236	4.246	4.256	1	2	3	4	5	6	7	8	9
0.63	4.266	4.276	4.285	4.295	4.305	4.315	4.325	4.335	4.345	4.355	1	2	3	4	5	6	7	8	9
0.64	4.365	4.375	4.385	4.395	4.406	4.416	4.426	4.436	4.446	4.457	1	2	3	4	5	6	7	8	9
0.65	4.467	4.477	4.487	4.498	4.508	4.519	4.529	4.539	4.550	4.560	1	2	3	4	5	6	7	8	9
0.66	4.571	4.581	4.592	4.603	4.613	4.624	4.634	4.645	4.656	4.667	1	2	3	4	5	6	7	9	10
0.67	4.677	4.688	4.699	4.710	4.721	4.732	4.742	4.753	4.764	4.775	1	2	3	4	5	7	8	9	10
0.68	4.786	4.797	4.808	4.819	4.831	4.842	4.853	4.864	4.875	4.887	1	2	3	4	6	7	8	9	10
0.69	4.898	4.909	4.920	4.932	4.943	4.955	4.966	4.977	4.989	5.000	1	2	3	5	6	7	8	9	10
0.70	5.012	5.023	5.035	5.047	5.058	5.070	5.082	5.093	5.105	5.117	1	2	4	5	6	7	8	9	11
0.71	5.129	5.140	5.152	5.164	5.176	5.188	5.200	5.212	5.224	5.236	1	2	4	5	6	7	8	10	11
0.72	5.248	5.260	5.272	5.284	5.297	5.309	5.321	5.333	5.346	5.358	1	2	4	5	6	7	9	10	11
0.73	5.370	5.383	5.395	5.408	5.420	5.433	5.445	5.458	5.470	5.483	1	3	4	5	6	8	9	10	11
0.74	5.495	5.508	5.521	5.534	5.546	5.559	5.572	5.585	5.598	5.610	1	3	4	5	6	8	9	10	12
0.75	5.623	5.636	5.649	5.662	5.675	5.689	5.702	5.715	5.728	5.741	1	3	4	5	7	8	9	10	12
0.76	5.754	5.768	5.781	5.794	5.808	5.821	5.834	5.848	5.861	5.875	1	3	4	5	7	8	9	11	12
0.77	5.888	5.902	5.916	5.929	5.943	5.957	5.970	5.984	5.998	6.012	1	3	4	5	7	8	10	11	12
0.78	6.026	6.039	6.053	6.067	6.081	6.095	6.109	6.124	6.138	6.152	1	3	4	6	7	8	10	11	13
0.79	6.166	6.180	6.194	6.209	6.223	6.237	6.252	6.266	6.281	6.295	1	3	4	6	7	9	10	11	13
0.80	6.310	6.324	6.339	6.353	6.368	6.383	6.397	6.412	6.427	6.442	1	3	4	6	7	9	10	12	13
0.81	6.457	6.471	6.486	6.501	6.516	6.531	6.546	6.561	6.577	6.592	2	3	5	6	8	9	11	12	14
0.82	6.607	6.622	6.637	6.653	6.668	6.683	6.699	6.714	6.730	6.745	2	3	5	6	8	9	11	12	14
0.83	6.761	6.776	6.792	6.808	6.823	6.839	6.855	6.871	6.887	6.902	2	3	5	6	8	9	11	13	14
0.84	6.918	6.934	6.950	6.966	6.982	6.998	7.015	7.031	7.047	7.063	2	3	5	6	8	10	11	13	15
0.85	7.079	7.096	7.112	7.129	7.145	7.161	7.178	7.194	7.211	7.228	2	3	5	7	8	10	12	13	15
0.86	7.244	7.261	7.278	7.295	7.311	7.328	7.345	7.362	7.379	7.396	2	3	5	7	8	10	12	13	15
0.87	7.413	7.430	7.447	7.464	7.482	7.499	7.516	7.534	7.551	7.568	2	3	5	7	9	10	12	14	16
0.88	7.586	7.603	7.621	7.638	7.656	7.674	7.691	7.709	7.727	7.745	2	4	5	7	9	11	12	14	16
0.89	7.762	7.780	7.798	7.816	7.834	7.852	7.870	7.889	7.907	7.925	2	4	5	7	9	11	13	14	16
0.90	7.943	7.962	7.980	7.998	8.017	8.035	8.054	8.072	8.091	8.110	2	4	6	7	9	11	13	15	17
0.91	8.128	8.147	8.166	8.185	8.204	8.222	8.241	8.260	8.279	8.299	2	4	6	8	9	11	13	15	17
0.92	8.318	8.337	8.356	8.375	8.395	8.414	8.433	8.453	8.472	8.492	2	4	6	8	10	12	14	15	17
0.93	8.511	8.531	8.551	8.570	8.590	8.610	8.630	8.650	8.670	8.690	2	4	6	8	10	12	14	16	18
0.94	8.710	8.730	8.750	8.770	8.790	8.810	8.831	8.851	8.872	8.892	2	4	6	8	10	12	14	16	18
0.95	8.913	8.933	8.954	8.974	8.995	9.016	9.036	9.057	9.078	9.099	2	4	6	8	10	12	15	17	19
0.96	9.120	9.141	9.162	9.183	9.204	9.226	9.247	9.268	9.290	9.311	2	4	6	8	11	13	15	17	19
0.97	9.333	9.354	9.376	9.397	9.419	9.441	9.462	9.484	9.506	9.528	2	4	7	9	11	13	15	17	20
0.98	9.550	9.572	9.594	9.616	9.638	9.661	9.683	9.705	9.727	9.750	2	4	7	9	11	13	16	18	20
0.99	9.772	9.795	9.817	9.840	9.863	9.886	9.908	9.931	9.954	9.977	2	5	7	9	11	14	16	18	20

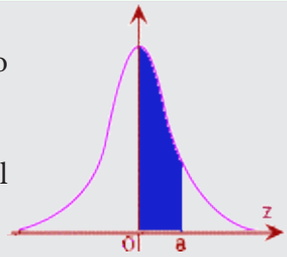
LogTable_11th_Stats.indd 333 2/3/2019 11:26:38 AM

www.tnschoolbooks.in

STANDARD NORMAL DISTRIBUTION TABLE

The Z-score values are represented by the column value + row value, up to two decimal places

The table is based on the upper right 1/2 of the Normal Distribution; total area shown is 0.5



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	Z
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	0.0
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	0.1
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	0.2
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	0.3
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	0.4
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	0.5
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	0.6
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	0.7
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	0.8
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	0.9
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	1.0
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	1.1
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	1.2
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	1.3
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	1.4
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441	1.5
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545	1.6
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633	1.7
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706	1.8
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767	1.9
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817	2.0
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857	2.1
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890	2.2
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916	2.3
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936	2.4
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952	2.5
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964	2.6
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974	2.7
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981	2.8
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986	2.9
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990	3.0

LogTable_11th_Stats.indd 334 2/3/2019 11:26:38 AM

LogTable_11th_Stats.indd 335

EXPONENTIAL FUNCTION TABLE (Values of e^{-m})

m	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.9900	0.9802	0.9704	0.9608	0.9512	0.9418	0.9324	0.9231	0.9139
0.1	0.8958	0.8869	0.8781	0.8694	0.8607	0.8521	0.8437	0.8353	0.8270
0.2	0.8106	0.8025	0.7945	0.7866	0.7788	0.7711	0.7634	0.7558	0.7483
0.3	0.7334	0.7261	0.7189	0.7118	0.7047	0.6977	0.6907	0.6839	0.6771
0.4	0.6637	0.6570	0.6505	0.6440	0.6376	0.6313	0.6250	0.6188	0.6126
0.5	0.6005	0.5945	0.5886	0.5827	0.5769	0.5712	0.5655	0.5599	0.5543
0.6	0.5434	0.5379	0.5326	0.5273	0.5220	0.5169	0.5117	0.5066	0.5016
0.7	0.4916	0.4868	0.4819	0.4771	0.4724	0.4677	0.4630	0.4584	0.4538
0.8	0.4449	0.4404	0.4360	0.4317	0.4274	0.4232	0.4190	0.4148	0.4107
0.9	0.4025	0.3985	0.3946	0.3906	0.3867	0.3829	0.3791	0.3753	0.3716

www.tnschoolbooks.in

2/3/2019 11:26:38 AM

Random Number Table

13962	70992	65172	28053	02190	83634	66012	70305	66761	88344
43905	46941	72300	11641	43548	30455	07686	31840	03261	89139
00504	48658	38051	59408	16508	82979	92002	63606	41078	86326
61274	57238	47267	35303	29066	02140	60867	39847	50968	96719
43753	21159	16239	50595	62509	61207	86816	29902	23395	72640
83503	51662	21636	68192	84294	38754	84755	34053	94582	29215
36807	71420	35804	44862	23577	79551	42003	58684	09271	68396
19110	55680	18792	41487	16614	83053	00812	16749	45347	88199
82615	86984	93290	87971	60022	35415	20852	02909	99476	45568
05621	26584	36493	63013	68181	57702	49510	75304	38724	15712
06936	37293	55875	71213	83025	46063	74665	12178	10741	58362
84981	60458	16194	92403	80951	80068	47076	23310	74899	87929
66354	88441	96191	04794	14714	64749	43097	83976	83281	72038
49602	94109	36460	62353	00721	66980	82554	90270	12312	56299
78430	72391	96973	70437	97803	78683	04670	70667	58912	21883
33331	51803	15934	75807	46561	80188	78984	29317	27971	16440
62843	84445	56652	91797	45284	25842	96246	73504	21631	81223
19528	15445	77764	33446	41204	70067	33354	70680	66664	75486
16737	01887	50934	43306	75190	86997	56561	79018	34273	25196
99389	06685	45945	62000	76228	60645	87750	46329	46544	95665
36160	38196	77705	28891	12106	56281	86222	66116	39626	06080
05505	45420	44016	79662	92069	27628	50002	32540	19848	27319
85962	19758	92795	00458	71289	05884	37963	23322	73243	98185
28763	04900	54460	22083	89279	43492	00066	40857	86568	49336
42222	40446	82240	79159	44168	38213	46839	26598	29983	67645

www.tnschoolbooks.in

LogTable_11th_Stats.indd 336 2/3/2019 11:26:38 AM

LogTable_11th_Stats.indd 337

Random Number Table

43626	40039	51492	36488	70280	24218	14596	04744	89336	35630
97761	43444	95895	24102	07006	71923	04800	32062	41425	66862
49275	44270	52512	03951	21651	53867	73531	70073	45542	22831
15797	75134	39856	73527	78417	36208	59510	76913	22499	68467
04497	24853	43879	07613	26400	17180	18880	66083	02196	10638
95468	87411	30647	88711	01765	57688	60665	57636	36070	37285
01420	74218	71047	14401	74537	14820	45248	78007	65911	38583
74633	40171	97092	79137	30698	97915	36305	42613	87251	75608
46662	99688	59576	04887	02310	35508	69481	30300	94047	57096
10853	10393	03013	90372	89639	65800	88532	71789	59964	50681
68583	01032	67938	29733	71176	35699	10551	15091	52947	20134
75818	78982	24258	93051	02081	83890	66944	99856	87950	13952
16395	16837	00538	57133	89398	78205	72122	99655	25294	20941
53892	15105	40963	69267	85534	00533	27130	90420	72584	84576
66009	26869	91829	65078	89616	49016	14200	97469	88307	92282
45292	93427	92326	70206	15847	14302	60043	30530	57149	08642
34033	45008	41621	79437	98745	84455	66769	94729	17975	50963
13364	09937	00535	88122	47278	90758	23542	35273	67912	97670
03343	62593	93332	09921	25306	57483	98115	33460	55304	43572
46145	24476	62507	19530	41257	97919	02290	40357	38408	50031
37703	51658	17420	30593	39637	64220	45486	03698	80220	12139
12622	98083	17689	59677	56603	93316	79858	52548	67367	72416
56043	00251	70085	28067	78135	53000	18138	40564	77086	49557
43401	35924	28308	55140	07515	53854	23023	70268	80435	24269
18053	53460	32125	81357	26935	67234	78460	47833	20496	35645

www.tnschoolbooks.in

2/3/2019 11:26:38 AM

www.tnschoolbooks.in

Statistics – Class XI
List of Authors and Reviewers

Domain Experts

Dr. G. Gopal
Professor & Head (Retd.), Dept. of Statistics
University of Madras, Chennai

Dr. G. Stephen Vincent
Associate Professor & Head (Retd.), Dept. of Statistics
St.Joseph's College, Trichy

Dr. R. Ravanan
Principal, Presidency College, Chennai.

Dr. K. Senthamarai Kannan
Professor, Dept. of Statistics, Manonmaniam
Sundaranar University, Tirunelveli.

Dr. A. Loganathan
Professor, Dept. of Statistics, Manonmaniam
Sundaranar University, Tirunelveli.

Dr. R.Vijayaraghavan
Professor, Dept. of Statistics,
Bharathiyar University, Coimbatore.

Dr. R. Kannan
Professor, Dept. of Statistics,
Annamalai University, Chidambaram.

Dr. N. Viswanathan
Associate Professor, Dept. of Statistics,
Presidency College, Chennai.

Dr. R.K. Radha
Assistant Professor, Dept. of Statistics,
Presidency College, Chennai.

Art and Design Team

Layout
Joy Graphics, Chennai

Illustrations
Gokula Krishnan

In-House
QC - Gopu Rasuvel,
Manohar Radhakrishnan,
Jerald wilson.

Wrapper- Kathir Arumugam

Co-ordination
Ramesh Munisamy

Typist
E.S.R. Rani Subbulakshmi
DIET-Vanaramutti, Thoothukudi Dt.
Manikandan, Chennai

Reviewers

Dr. M.R. Srinivasan
Professor & Head, Dept. of Statistics,
University of Madras, Chennai.

Dr. P. Dhanavanthan
Professor and Dean, Dept. of Statistics,
Pondicherry University, Pondicherry.

Tmt M.Susila
Addl. Director, Department of Economics and Statistics,
Chennai.

Content Writers

G. Gnanasundaram
HM (Retd.), SSV HSS, Parktown, Chennai.

P. Rengarajan
PG Asst., (Retd.), Thiyagarajar HSS, Madurai.

S. John Kennadi
PG Asst., St. Xavier's HSS, Purathakudi, Trichy.

AL.Nagammai
PG Asst., Sevasangam GHSS, Trichy.

M. Rama Lakshmi
PG Asst., Suguni Bai Sanathana Dharma GHSS, Chennai.

Maala Bhaskaran
PG Asst., GGHSS, Nandhivaram, Kanchipuram.

M. Boobalan
PG Asst., Zamindar HSS, Thuraiyur, Trichy.

R. Avoodaiappan
PG Asst., GGHSS, Ashok Pillar, Chennai.

G.K. Ganesan
PG Asst., KK Naidu HSS, Coimbatore.

Naga Madeshwaran
PG Asst., Presidency GHSS, Egmore, Chennai.

Sobana Rani
PG Asst., Ganesh Bai Galada Jain GHSS, Chennai.

K. Chitra
PG Asst., Tarapore and Loganathan GHSS, Chennai.

Academic Coordinator

M. Ramesh
B.T. Asst.,
Govt. Boys Hr. Sec. School, Alangayam, Vellore Dt.

ICT Coordinator

D. Vasuraj
BT Asst., PUMS, Kosapur, Puzhal Block,
Thiruvallur Dt.

**This book has been printed on 80 G.S.M.
Elegant Maplitho paper.**

Printed by offset at: