Elementary Statistics and Computer Application

- **Content Creator:** Dr. Raju Prasad Paswan, Assam Agricultural University, Jorhat

- **Course Reviewer:** Dr. S.S. Sidhu, Punjab Agricultural University, Ludhiana

Lesson Number	Lesson Title
Lesson 1	Introduction to Statistics
Lesson 2	Statistical Data (Basic Concepts, Classification and Tabulation)
Lesson 3	Frequency Distribution and Presentation of Data
Lesson 4	Measures of Central Tendency
Lesson 5	Measures of Dispersion
Lesson 6	Basic Concepts of Probability and Some Theoretical Distributions
Lesson 7	Sample Survey
Lesson 8	Tests of Significance
Lesson 9	Simple Correlation
Lesson 10	Regression Analysis
Lesson 11	Experimental Designs
Lesson 12	Factorial Experiments
Lesson 13	Introduction to Computers
Lesson 14	Classification of Computer
Lesson 15	Basic Concepts of Operating System
Lesson 16	Basic Concepts of DOS and Windows Operating

	System
Lesson 17	MS-Office
Lesson 18	Basic Concepts of Programming
Lesson 19	Introduction to Multimedia Applications
Lesson 20	Introduction to Visual Basic Programming

Elementary Statistics and Computer Application

LESSON 1

Introduction to Statistics

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 1	Introduction To Statistics
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. Meaning of statistics and its usage in real life.
2. Drawbacks and limitations.
3. Applications of statistics in agriculture.
4. Distrusts of Statistics

1.1 REASONS FOR LEARNING STATISTICS

- It is important to develop the ability to extract meaningful information from raw datato make better decisions. It is possible only through the careful analysis of data guided by **statistical thinking**.
- The reason for analysis of data is an understanding of variation and its causes in any phenomenon. Since, variation is present in all phenomena, therefore knowledge of it leads to better decisions about a phenomenon that produced the data.

  It is from this perspective that the learning of statistics enables the decision-maker to understand how to:

  - Present and describe information (data) so as to improve decisions;
  - Draw conclusions about the large population based upon information obtained from samples;
  - Seek out relationship between pair of variables to improve processes;
  - Obtain reliable forecasts of statistical variables of interest;
- Thus, a statistical study might be a simple exploration enabling us to gain insight into a virtually unknown situation or it might be a sophisticated analysis to produce numerical confirmation or a reflection of some widely held belief.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1.2 GROWTH AND DEVELOPMENT OF STATISTICS

- The views commonly held about **statistics** are numerous, but often incomplete. It has different meanings to different people depending largely on its use. For example,
    - for a cricket fan, statistics refers to numerical information or data relating to the runs scored by a cricketer;
    - for an environmentalist, statistics refers to information on the quantity of pollutants released into the atmosphere by all types of vehicles in different cities;
    - for the census department, statistics consists of information about the birth rate and the sex ratio in different states;
    - for a share broker, statistics is the information on changes in share prices over a period of time; **and so on**.
- An average person perceives statistics as a **column of figures, various types of graphs, tables and charts** showing the increase and/or decrease in per capita income, exports, imports, and crime rate and so on they come across in newspapers, magazines/journals, reports/bulletins, radio, and television etc.

1.3 ORIGIN OF STATISTICS

- Statistics was born as the **science of kings.** It had its origin in the needs of the administrators in the ancient days for collecting and maintaining quantitative information about their population and wealth with a view to make policies for citizens.
- 'Statistics' seems to have been derived from the Latin word 'Status' or Italian word 'Statista' or the German word 'Statistik' carrying the same meaning of a **political state**

1.4 DEFINITION OF STATISTICS

- The word "Statistics" is used in **two senses:** Plural and Singular.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- In its plural form, **it refers to the statistical data collected in a systematic manner with some definite aim or object in view**.
- In **singular form**, it means, **statistical methods** or the **subject itself**. It includes the methods and principles concerned with collection, analysis and interpretation of numerical data.
    - **STATISTICS IN PLURAL SENSE – HORACE SECRIST**

      "By statistics, we mean aggregate of facts, affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other."
    - **STATISTICS IN SINGULAR SENSE – CROXTON AND COWDEN**

      "Statistics may be defined as the science of collection, presentation, analysis and interpretation of numerical data."
- As a subject, there are two branches of statistics: **Mathematical Statistics and Applied Statistics.**
    - **Mathematical Statistics** is a branch of mathematics and is theoretical. It deals with the basic theory about how a particular statistical method is developed.
    - **Applied statistics**, on the other hand, uses statistical theory in formulating and solving problems in other subject areas such as economics, agriculture, sociology, medicine, business/industry, education, and psychology.
- The whole of statistical methods can also be classified into two heads: **Descriptive Statistics and Inferential Statistics**.
    - Statistical methods involving the collection, presentation, and characterization of a set of data in order to describe the various

features of that set of data is known as **descriptive statistics**. It includes diagrammatic/graphic methods (bar charts, ogives etc.) and numeric measures (mean, variance etc.)

- Statistical methods which facilitate estimating the characteristics of a population or making decisions concerning a population on the basis of sample results is **inferential statistics**.

1.5 CHARACTERISTICS OF STATISTICAL DATA

Statistical data are:

- Aggregate of facts, and must be enumerated or estimated.
- Affected (to a marked extent) by multiplicity of causes.
- Must be numerically expressed.
- Collected in a systematic manner and for a predetermined purpose.
- Capable of being related to each other.

1.6 APPLICATION OF STATISTICS IN AGRICULTURE

- Data and numerical information have played a very vital role in the growth and development of agriculture, especially in the developed countries. In an agrarian country, like India, the utility of agricultural statistics is even more important. The quantitative agricultural researches, in fact, are largely based on statistical data and its scope is very widening.
- Some of its mentionable utility are in the fields of:
  - **Land utilization and irrigation**, including the net area sown, gross cultivated area, current flow, cultivable waste, irrigated area in Kharif and Rabi seasons etc.
  - **Forestry,** including knowledge about forest cover, variety etc.
  - **Agricultural production** including arable, plantations, livestock and fisheries.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Agricultural prices and wages**.
- **Statistics relating to agricultural organization and farming structure**, e.g., persons employed in agriculture, their status, land held under various tenure, number of draught animals, implements, farm building, etc.
- **Statistics and economics of production and marketing**, e.g., cost of production, input-output ratio, marketing changes, marketing spread over, etc.
- **General statistics**, literacy among those employed in agriculture, health, sanitation.
- **Statistics relating to weather and climate**, rainfall and its distribution, temperature (and its range, soil and its pH value, etc.
- **Forecast** weather, crops and prices.

1.7 LIMITATIONS OF STATISTICS

Some of the common limitations of statistics are:

- It deals only with aggregates.
- Statistical methods can be applied only to quantitative data**.**
- Laws of statistics are true on an average.
- Statistical results are only approximately correct.
- Statistics does not reveal the entire story; (it is a helping hand).
- Statistics can be misused

1.8 DISTRUSTS OF STATISTICS

- Though statistics today finds it application in all branches of study, some people do not accept the role of statistics as they express various discontents regarding the subject. Some of the reasons for expressing such divergent views regarding the nature and functions of Statistics are as follows:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Figures are innocent, easily believable and more convincing. The facts supported by figures are psychologically more appealing.
- Figures put forward for arguments may be inaccurate or incomplete and might lead to wrong inferences.
- Figures though accurate, might be manipulated to conceal the truth and present a distorted picture of facts to the public to meet their own motives.
- Statistics neither proves anything nor disproves anything. It is only a tool, which if rightly applied may prove extremely useful, and if misused, might be disastrous.

- To conclude, it is not the subject of Statistics that is to be blamed but those people who twist the numerical data and misuse them either due to ignorance or deliberately for personal motives. The science of Statistics is the most useful servant but only of great value to those who understand it anduse it judiciously.

****

GLOSSARY:

- **Statistics in plural sense:** In its plural form, it refers to the statistical data collected in a systematic manner with some definite aim or object in view.
- **Statistics in singular sense**: In singular form, it means, statistical methods or the subject itself.
- **Mathematical statistics:** Theoretical branch of statistics that deals with the basic theory about how a particular statistical method is developed.
- **Applied statistics:** It a branch of statistics that applies statistical theory for solving problems in other subject areas
- **Descriptive statistics:** Statistical methods involving the collection, presentation, and characterization of a set of data in order to describe the various features of that set of data.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Inferential statistics:** Statistical methods which facilitate estimating the characteristics of a population or making decisions concerning a population on the basis of sample results.

REFERENCES

- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.
- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson2

Statistical Data

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 2	Statistical Data
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. Overview of different types of statistical data.
2. Different methods of data collection.
3. Classification of data and its advantages.
4. Tabulation of data and its advantages.

2.1 STATISTICAL DATA AND ITS TYPES.

- Statistical data are the basic material method to make an effective decision in a particular situation.
- Statistical data are numerically expressed aggregate of facts, collected in a systematic manner for a predetermined purpose and placed in relation to each other.
- Types of Data: (a) Categorical data and (b) Measurement data.
    - **Categorical data**: The objects being studied are grouped into categories based on some qualitative trait. The resulting data are merely labels or categories. Categorical data are of two types: Nominal and Ordinal.

    Example: Nominal: Information type: *Do you practice Yoga?* Measurement type: **Yes/ No**.Ordinal: Suppose a group of people were asked to take varieties of biscuit and classify each biscuit on a rating scale of 1 to 5, representing strongly dislike, dislike, neutral, like, strongly like.

    - **Measurement Data:** The objects being studied are 'measured' based on some quantitative trait. The resulting data are set of numbers. Measurement (or numerical) data are two types: Discrete and Continuous. Example: Discrete: Information type: *How many books do you have in your library?*

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Measurement type: **Number** (say 2500 nos.). Continuous: Information type: *What is your height?* Measurement type: **Centimetres or Inches** (say 162 cms)

Table: 2.1 Different measurement scales on statistical data:

Statistical Data	Scale of Measurement	Characteristics of Measurement	Example
Categorical	Nominal	No order, distance or unique origin	What is your gender?
Categorical	Ordinal	Order but no distance or unique origin	How do you feel today?
Measurement	Interval	Both order and distance but not unique origin	Beauty score, Intelligence score, etc.
Measurement	Ratio	Order, distance and unique origin	Height, Weight, Age, etc.

2.2 SOURCES OF DATA/ COLLECTION OF DATA.

- Collection of data means the methods that are to be employed for getting the required information from the units under investigation. It depends on the nature object and scope of enquiry.
- Based on the method of collection of data, statistical data can be either primary or secondary.
    - **Primary data:** Data which are collected for the first time by the investigator himself for a specific purpose are known as primary data. Such data are collected by means of a census or sample survey.
    - **Secondary data:** Secondary data are those which have been previously collected by some agency for their own purpose but are now used in a different connection.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

E.g. suppose we want to conduct an inquiry with the cost of living of the workers in a certain factory. If the facts selecting to this inquiry are collected by the investigators from the workers themselves such data would be termed as primary data. The term secondary data refers to the statistical material collection from someone else's record and not by original observations. Thus, if the above example if we collect the data from the records of the trade union or from some other source the data will be called secondary data.

- The **difference between primary data and secondary data** is largely a matter of degree. Data which are primary in the hands of one may be secondary for others.
    - Primary data are those data which are collected for the first time and thus original in character, whereas, secondary data are obtained from someone else's records.
    - Primary data are in the shape of raw data to which statistical methods are applied but the secondary data are like finished products since they have been processed statistically.
    - After statistical treatment the primary data loss their original shape and becomes secondary data. Primary data once published become secondary data.
- **Collection of Primary Data**: For the collection of primary data, the following methods can be adopted:
    - Direct Personal Investigation
      (+) gives more reliable results because of personal approach
      (-) costly and time consuming
      (-) bias of the investigator comes into play
    - Indirect oral Investigation
      (+) Economic, time saving & even the original units are not cooperating,
      information about them is possible.
      (-) Third party information are not reliable at times.
    - Investigation through local agencies

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

(+) very cheap, coverage is wide and local expertise yields results easily.
(-) reliability of data may be a matter of doubt.

- Mailed Questionnaire method
  (+) least expensive.
  (+) bias of the investigator is completely ruled out.
  (-) cannot be applied if the informants are illiterate.
  (-) success of the method depends upon the co-operation of the respondents
- Schedules sent through enumerators
  (+) enumerators go to the respondents with the schedule and record their
  replies. So, can be applied even when the informants are illiterate.
  (+) yields maximum possible results.
  (+) method of substitution can be applied if there is non-response.
  (-) expensive and time consuming.
  (-) accuracy, largely dependent upon the skill and efficiency of the investigator.

[Note: In some cases, a combination of two or more of the methods may be used**]**

- **Collection of Secondary Data:** Following are some of the common sources of secondary data:
  - Government Publications: Usually information is collected from publications of central, state or international institutions. E.g. Official Publications of the international organizations like the U.N.O, W.H.O, I.L.O, World Bank etc. Official publications of the Govt. Organizations like Central Statistical Organization (C.S.O), National Sample Survey Office (N.S.S.O), Directorate of Economics and Statistics, etc.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Semi-Govt. Publications: In this method, the information is collected from the statistical material published by the semi-govt. bodies like municipal corporation, District boards etc.
- Other sources: Necessary information can also be collected from the publications of trade associations or chambers of commerce, newspapers and periodicals, Research bureau/Universities/Institutes etc.

2.3 CLASSIFICATION OF DATA

- Classification of data is the process of arranging data in groups/classes on the basis of certain properties. Statistical data are classified after taking into account the nature, scope, and purpose of an investigation.
- Generally, data are classified on the basis of the following four bases:
  - **Geographical Classification**: In geographical classification, data are classified on the basis of geographical or location differences such as—cities, districts, or villages between various elements of the data set.

City	**Mumbai**	**Kolkata**	**Delhi**	**Chennai**
Population density (per sq km)	654	685	423	205

  - **Chronological Classification**: When data are classified on the basis of time, the classification is known as chronological classification. Such classifications are also called time series because data are usually listed in chronological order starting with the earliest period.

Year	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Population (crore)	31.9	36.9	45.9	54.7	75.6	85.9

  - **Qualitative Classification**: In qualitative classification, data are classified on the basis of descriptive characteristics or on the

basis of attributes like sex, literacy, region, caste, or education, which cannot be quantified.

- **Quantitative Classification**: In this classification, data are classified on the basis of some characteristics which can be measured such as height, weight, income, expenditure, production, or sales. Quantitative variables can be divided into the following two types:
  - Discrete data: A set of data is said to be discrete if the values or observations belonging to it are distinct and separate i.e. they can be counted and listed out. E.g. No. of students in your class, no. of stars in the sky etc.
  - Continuous data: A set of data is said to be continuous if the values/observations belonging to it may take on any value within a finite or infinite interval. They are usually described using intervals on the, real line. E.g. height of students, temperature of a place, weight, life time of a bulb etc.

- Some of the requisites of a good classification are:
  - The classes should be clearly defined and should not lead to any ambiguity.
  - The classes should be exhaustive.
  - The classes should be mutually exclusive and non-overlapping.
  - The classes should be of equal width.
  - The open-end classes should be avoided.
  - The number of classes should neither be too large nor too small. It should preferably lie between 5 and 15.

2.4 TABULATION OF DATA

- Tabulation is another way of summarizing and presenting the given data in a systematic form in rows and columns. Such presentation facilitates comparisons by bringing related information close to each other and helps in further statistical analysis and interpretation.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Some of the important advantages of tabulation are:
    - It simplifies complicated data.
    - Facilitates comparison of data.
    - Provides a basis for analysis and interpretation of data.
    - Helps as a reference.
- Different Parts of a Table: The different parts of a table are listed below:
    - **Table number**: A table should be numbered for easy identification and reference in future. The table number may be given either in the center or side of the table but above the top of the title of the table.
    - **Title of the table**: Each table must have a brief, self-explanatory and complete title which is to be written either below the table number or after the table number in the same line. The title should convey the full description of the contents in the table.
    - **Caption and stubs**: The heading for columns and rows are called caption and stub, respectively. They must be clear and concise. Two or more columns or rows with similar headings may be grouped under a common heading to avoid repetition. Such arrangements are called sub-captions or sub-stubs.
    - **Body:** The body of the table should contain the numerical information. The numerical information is arranged according to the descriptions given for each column and row.
    - **Totals:** The totals and sub-totals of all the rows and the columns should be given in the table.
    - **Footnotes**: Anything written below the table is called a footnote. The purpose of footnote is to clarify some of the specific items given in the table.
    - **Source:** the source or sources of the data embodied in the table should be mentioned under the table. It is mentioned below the footnote.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- A sample table is drawn below for better understanding:

Table: 2.2 Number of candidates appearing for different examinations of a University

Table No. Caption Title Sub-captions

Faculty (Stubs)	Examination						Total
	H.S.		U.G.		P.G.		
	M	F	M	F	M	F	
Arts							
Science							
Commerce							
Total							

- N.B.: F-Female, M-Male
- Source: AAU Magazine, 2018

- **Tips for Constructing a Table**: While tabulating, the following may be borne in mind:
  - A rough draft of the table should be prepared first. Before presenting the final table the rough draft should be examined very carefully.
  - Figures that are to be compared are to be placed as near to each other as possible.
  - A suitable heading should be given to the table. It should be brief, comprehensive and self- explanatory.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Heading of columns and rows should be brief and clear.
- The columns and rows are to be arranged in a logical order.
- The table should be simple and self-explanatory.

****

GLOSSARY:

- **Statistical Data:** These are numerically expressed aggregate of facts, collected in a systematic manner for a predetermined purpose and placed in relation to each other. They are either categorical or of measurement.
- **Primary data:** Data which are collected for the first time by the investigator himself for a specific purpose are known as primary data.
- **Secondary data:** Secondary data are those which have been previously collected by some agency for their own purpose but are now used in a different connection.
- **Classification of data**: It is the process of arranging data in groups/classes on the basis of certain properties.
- **Tabulation of data:** It is a way of summarizing and presenting the given data in a systematic form in rows and columns, which facilitates comparison.

REFERENCES

- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.
- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

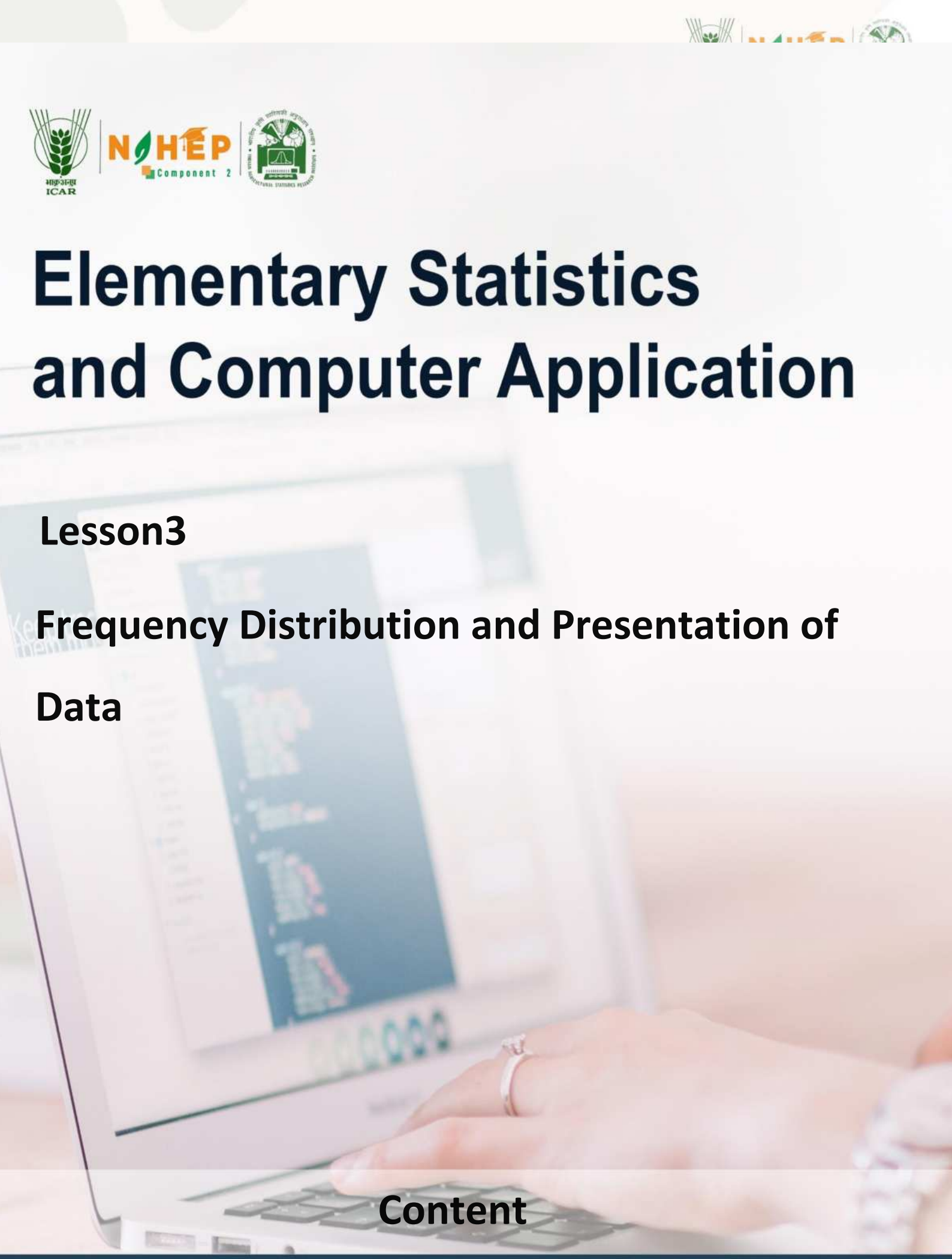
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson3

Frequency Distribution and Presentation of Data



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications

Course Name	Elementary Statistics And Computer Application
Lesson No. 3	Frequency Distributions And Presentation Of Data
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. Basic idea of a frequency distribution.
2. Different measures for diagrammatic representation of data
3. Different tools for graphical representation of data.

3.1 FREQUENCY AND FREQUENCY DISTRIBUTION

- The word 'frequency' is derived from 'how frequently' a variable occurs. The number of units associated with each value of the variable is called frequency of that value. Suppose the variable takes the value 15 and it occurs 3 times than 3 is called the frequency of the value 15.
- A systematic presentation of the values taken by variables together with corresponding frequencies is called a Frequency Distribution of the variable. If it is presented in tabular form it is called Frequency Table.
- Few examples of instances where frequency distributions would be useful are when (i) a marketing manager wants to know how many units of each product sells in a particular region during a given period, (ii) a financial analyst wants to keep track of the number of times the shares of manufacturing and service companies to be or gain order a period of time, etc.

3.2 GROUPED DATA AND UNGROUPED DATA

- Data is often described as ungrouped or grouped.
- Ungrouped data is data given as individual data points.
    - E.g.: Ungrouped data without a frequency distribution 1,3,6,4,5,6,3,4,6,3,6
    - E.g.: Ungrouped data with a frequency distribution (also known as Discrete frequency distribution) Here, class intervals are not present.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Number of children	No. of families
0	15
1	20
2	22
3	16
4	7

- Grouped data is data given in intervals.
    - E.g.: Grouped data (also known as continuous frequency distribution). These are formed with class intervals.

Marks	No. of Students
0 – 20	15
20 – 40	20
40 – 60	28
60 – 80	22
80– 100	15

3.3 COMMON TERMINOLOGIES:

- **Class Intervals**: The class intervals are the subsets into which the data is grouped.
- **Class Limits**: The lowest and the highest limit that can be included in the class is known as the class limits of that particular class.
- **Class Frequency**: The number of observations corresponding to a particular class is known as the frequency of that class.
- **Mid-point**: The mid-point or mid-value (M.V) of a class is determined as :

  (upper limit of the class + lower limit of the class)/2

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Inclusive Type and Exclusive Type of Class intervals**: The classes of the type 15 – 19, 20 – 24, 25 – 29, etc. in which both the upper and lower limits are included are called 'inclusive type'. For e.g. 20 – 24, includes all the values from 20 to 24, both inclusive.

  On the other hand, the classes of the type 15 – 20, 20 – 25, 25 – 30, etc. in which the lower limit is included and the upper limit is excluded are called 'exclusive type' of class intervals. For e.g. 20 – 25, includes all the values from 20 to 24. The value 25 will be included in the next class interval 25 – 30.

- **Class Boundaries**: The upper and lower class limits of the exclusive type classes are called class boundaries.

  Upper class boundary = upper class limit + d/2

  Lower class boundary = lower class limit + d/2

  where, *d = gap between the upper limit of any class and lower limit of the succeeding class.*

- **Width of a Class**: The width of a class interval is the difference between the class boundaries.

  Width of a class = upper class boundary – lower class boundary

- **Determination of number of class intervals:** To decide approximate number of classes in a frequency distribution Sturge's formula can be used. According to Sturge's rule, the number of classes can be determined by the formula

$k = 1 + 3.222 \log_{10} N$

where, k is the number of classes and $\log_{10} N$ is the logarithm of the total number of observations.

3.4 TYPES OF FREQUENCY/ FREQUENCY DISTRIBUTIONS

- **Relative Frequency Distribution:** The frequency of a class when expressed as a rate of the total frequency of the distribution is called relative frequency (R.F) distribution.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Percentage Frequency Distribution:** The percentage frequency (P.F) distribution is the ratio of class frequency to the total frequency expressed as percentage.
- **Frequency Density:** It is the ratio of the class frequency to the width of that class.
- **Cumulative Frequency:** If for each class, the frequencies given are the aggregate of the preceding frequencies they are known as cumulative frequencies. Cumulative frequencies (c.f.) can be presented either in 'less than' or 'more than' type.
    - In a less than (L.T) cumulative frequency distribution, the frequencies of each class interval are added successively from top to bottom.
    - In the more than (M.T) cumulative frequency distribution, the frequencies of each class interval are added successively from bottom to top.

Example: Below are given the marks of 50 students in a particular examination:

40	41	50	66	69	57	60	62	49	50
50	52	57	60	22	23	43	50	70	40
11	13	25	33	41	15	53	62	63	67
65	79	71	45	41	52	53	22	30	35
29	45	49	45	55	59	34	39	40	41

Prepare a frequency distribution table of the exclusive type by taking suitable class intervals.

Solution: Here, Largest observation, L =79

Smallest observation, S=11

No. of classes = $1 + 3.222 \log_{10} N$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

= 6.6 ~ 7 (since, N=50).

From the given information, we prepare the following table:

I	II	III	IV	V	VI	VII	VIII
Class interval	**Tally marks**	**Freq. (f)**	**M.V (x)**	**R.F**	**P.F**	**Cumulative Frequency**	
						L.T	M.T
10-20	///	3	15	3/50=	0.06x100%	3	50
20-30	~~////~~	5	25	0.06	=6%	8	47
30-40	~~////~~	5	35	0.10	10%	13	42
40-50	~~////// ///~~	13	45	0.10	10%	26	37
50-60	~~////// //~~	12	55	0.26	26%	38	24
60-70	~~//// ////~~	9	65	0.24	24%	47	12
70-80	///	3	75	0.18	18%	50	3
				0.06	6%		
TOTAL	--	50	--	1	100%	--	--

3.5 DIAGRAMMATIC AND GRAPHICAL PRESENTATION OF DATA

- According to Willford I. King, 'One of the chief aims of statistical science is to render the meaning the masses of figures clear and comprehensible at a glance'. This is often best accomplished by presenting the data in a pictorial form
- Instead of going through the mass data and to understand its nature, diagrammatic and graphical representations are more intelligible, attractive and appealing to the eye of the observer. They give a bird's eye-view of the data. They facilitate comparison of various aspects of data.

3.6 ADVANTAGES AND LIMITATIONS OF DIRAGRAMS AND GRAPHS

- **Advantages**
    - The graphs and diagrams give an attractive and elegant presentation. They give delight to the eye and add the spark of interest.
    - The graphic and diagrammatic representation of the data leaves good visual impact. They have greater memorizing value than figures.
    - The diagrams and graphs are not only attractive and impressive but they save time also. Because through the diagrams, it is possible to have an immediate grasp of significance.
    - With the help of diagrams and graphs, comparisons of groups and series of figures can be made easily.
    - Forecasting becomes easier with the help of the graphs.
    - The partition values and mode are also can be determined by graphs.
- **Limitations**
    - The graphs and diagrams provide an approximate picture of the data and not the accurate one.
    - They can be used only for comparative study.
    - They are capable of representing only homogeneous and comparable data.

3.7 DIAGRAMMATIC FORM OR DIAGRAMS

- A diagram makes visual comparison clear and it further facilitates understanding of the salient features of the data. Further it develops faster the idea about the relative magnitudes for the purpose of comparison.
- There are a variety of diagrams used to represent statistical data. Different types of diagrams, used to describe sets of data, are divided into the following categories:
    - **Dimensional diagrams**: One or two or three dimensional diagrams

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Pictograms or Ideographs**
- **Cartographs or Statistical maps**

3.7.1 ONE-DIMENSIONAL DIAGRAMS:

- They are in shape of vertical or horizontal lines or bars. The lengths of the lines or bars are in proportion to the different figures they represent.
- In one-dimensional diagrams the values of the variable (the characteristic to be measured) are scaled along the horizontal axis and the number of observations (or frequencies) along the vertical axis of the graph. The plotted points are then connected by straight lines to enhance the shape of the distribution.
- The height of such boxes (rectangles) measures the number of observations in each of the classes.
- These diagrams are called one-dimensional diagrams because only the length (height) of the bar (not the width) is taken into consideration.
- The following points must be kept in mind while constructing one-dimensional diagrams:
  - The width of all the bars drawn should be same.
  - The gap between one bar and another must be uniform.
  - There should be a common base to all the bars.
  - It is desirable to write the value of the variable represented by the bar at the top end so that the reader can understand the value without looking at the scale.
  - The frequency, relative frequency, or per cent frequency of each class interval is shown by drawing a rectangle whose base is the class interval on the horizontal axis and whose height is the corresponding frequency, relative frequency, or per cent frequency.
  - The value of variables (or class boundaries in case of grouped data) under study are scaled along the

horizontal axis, and the number of observations (frequencies, relative frequencies or percentage frequencies) are scaled along the vertical axis.

- The bar diagrams of common use in the one-dimensional diagrams are of the following main types:
    - Simple bar diagram
    - Multiple bar diagram
    - Sub-divided bar diagram
    - Percentage bar diagram

Simple bar diagram: In simple bar diagrams one bar represents only one figure and as such there will be as many bars as the number of figures. Such diagrams represent only one particular type of data. For example, the number of students in a university year after year can be represented by such bars.

Here the magnitudes of the items are represented by thick bars of uniform width with equal spacing between any two consecutive bars. The lengths of the bars are proportional to the magnitudes of the items. The bars can be drawn either vertical or horizontal depending upon the type of variable- numerical or categorical.

Bar diagrams can be used in case of discrete series, series of individual observation. They cannot be used in continuous series. Bars are not suitable to represent the long period time series also. Wherever there is continuity in data bar diagrams should not be used.

Example: Draw a simple bar diagram to represent the following statistics relating to the area under different crops:

Crops	**Million Acres**
Rice	80.3
Wheat	27.6

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Jowar	21.4
Other food crops	88.2
Oil seeds	17.6
Cotton	14.5
Other fibers	3.1
Fodder crops	10.2
Other non food crops	3.9

Solution: The Acreage (million acres) under different crops is shown by the following simple bar diagram.

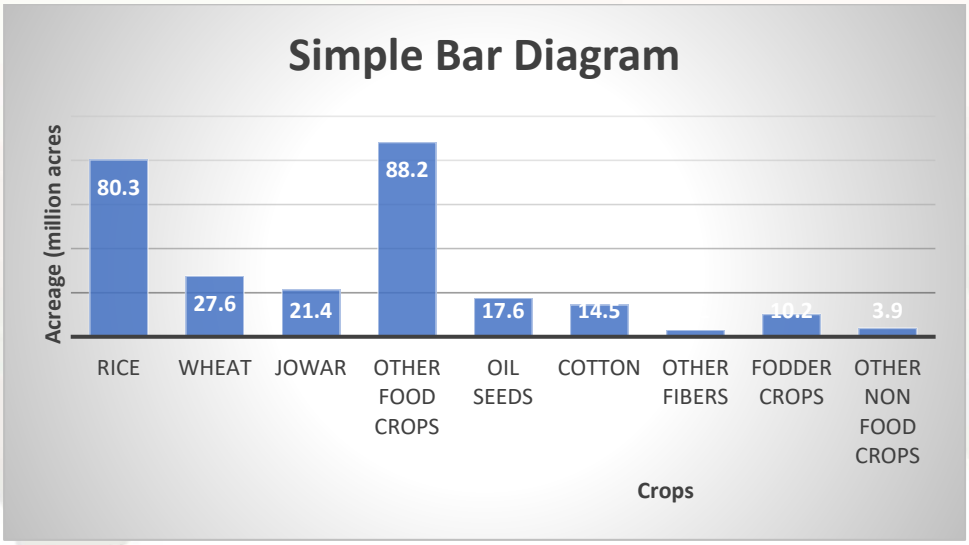


Fig : Simple Bar Diagram representing the Acreage (million acres) under different crops

Multiple bar diagram: Multiple bar diagrams are prepared on the basis of simple bar diagrams. These diagrams represent more than one sets of inter-related data at a time and so two or more bars (as the case may be) are constructed at a time side by side. Population data relating to occupational structure, civil conditions, age, etc. are usually represented by such diagrams.

The technique of drawing such a chart is same as that of a single bar chart with a difference that each set of data is represented in different shades or colors on the same scale.

Example: The following table provides the percentage of cultivators and percentage of cultivated area in different sizes of holdings in U.P. Depict the data in a bar diagram.

Size of holding (acres)	Percentage of cultivators	Percentage of area
Up to 1	37.8	6.0
1 – 2	18.0	8.1
2 – 3	11.6	8.7
3 – 4	8.1	8.4
4 – 5	5.7	7.6
5 – 6	4.2	6.8
6 – 7	3.0	5.9
7 – 8	2.3	5.1
8 – 9	1.8	4.4
9 – 10	1.4	3.9

Solution: The percentage of cultivators and the cultivated area in different sizes of holdings is shown by the following multiple bar diagram. Corresponding to each holding size, two adjacent bars are constructed where one of the two represents the percentage of cultivators and the other percentage of cultivated area.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

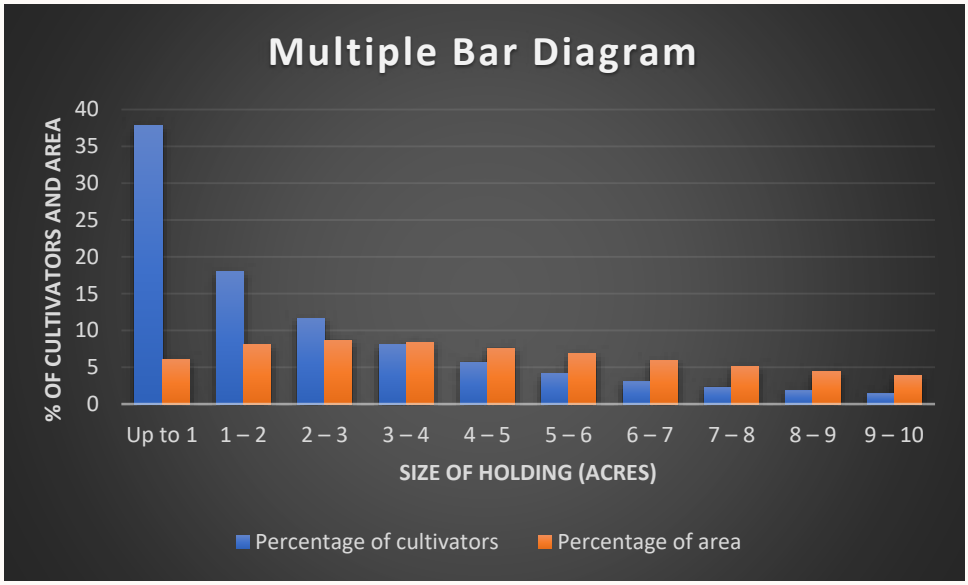


Fig.: Multiple Bar Diagram representing % of cultivators cultivated area in different sizes of holdings

Sub-divided bar diagram: Sub-divided bar diagrams are used to present such data which are to be shown in the parts or which are totals of various sub-divisions. Subdivided bar charts are suitable for expressing information in terms of ratios or percentages. For example, net per capita availability of food grains, results of a college faculty-wise in last few years, and so on. While constructing these charts the various components in each bar should be in the same order to avoid confusion. Different shades must be used to represent various ratio values but the shade of each component should remain the same in all the other bars.

Example: The following table shows the expenditure incurred by the central government and the governments of States of type A and B in the 1st five years plan under the six major heads. Represent the data by means of a suitable diagram.

Expenditure in the 1st five years plan (in crores of rupees)			
Subject	**Central**	**Type A States**	**Type B States**
Agril. & development	186.3	127.3	37.6

Irrigation & power	265.9	206.1	81.5
Transport & Communication	409.5	56.5	17.4
Industries	146.7	17.9	7.1
Social Services	191.4	192.3	28.9
Miscellaneous	40.7	10.0	7.0
Totals	1240.5	610.1	173.2

Solution: The data can be represented by means of a sub-divided bar diagram. Three bars are constructed and their heights are in proportion of the total expenditure of the three types of governments respectively. Further each bar is sub-divided into six parts and the height of each art is proportional to the expenditure incurred on it.

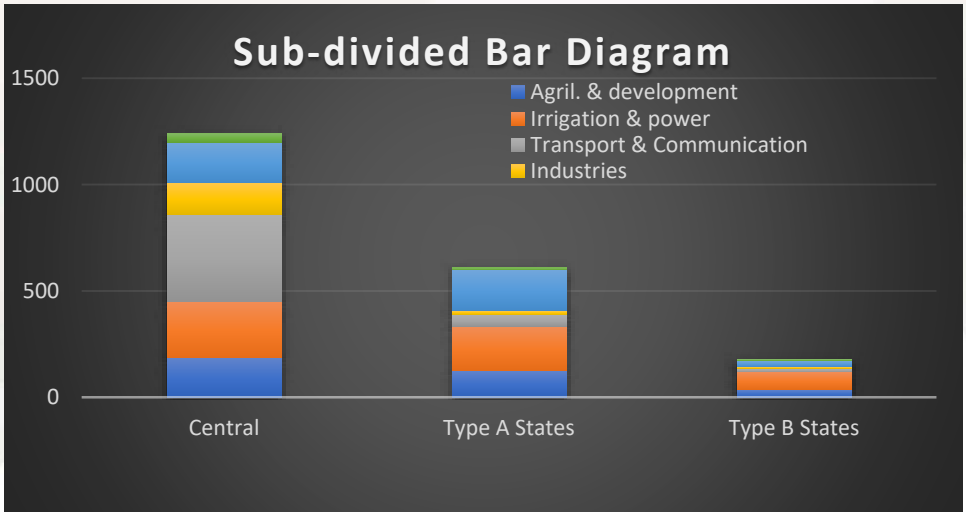


Fig : Sub-divided Bar Diagram representing the expenditure incurred by the central government and the governments of States of type A and B

Percentage bar diagram: When the relative proportions of components of a bar are more important than their absolute values, then each bar can be constructed with same size to represent 100%. The component

values are then expressed in terms of percentage of the total to obtain the necessary length for each of these in the full length of the bars.

Example : With the help of percentage bar diagram represent the following data.

Faculty	College A	College B
Arts	350	200
Science	500	250
Commerce	650	150
Law	300	120
Totals	1800	720

Solution: To draw the percentage bar diagram, first we have to change the absolute magnitudes into the percentages.

Faculty	College (A)		College (B)	
	Absolute magnitude	Relative magnitude in %	Absolute magnitude	Relative magnitude in %
Arts	350	$\frac{350}{1800}$x 100 = 19.44	200	$\frac{200}{720}$x 100 = 27.78
Science	500	$\frac{500}{1800}$x 100 = 27.76	250	$\frac{250}{720}$x 100 = 34.72
Commerce	650	$\frac{650}{1800}$x 100 = 36.13	150	$\frac{150}{720}$x 100 = 20.83
Law	300	$\frac{300}{1800}$x 100 = 16.66	120	$\frac{120}{720}$x 100 = 16.67
Total	1800	100	720	100

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

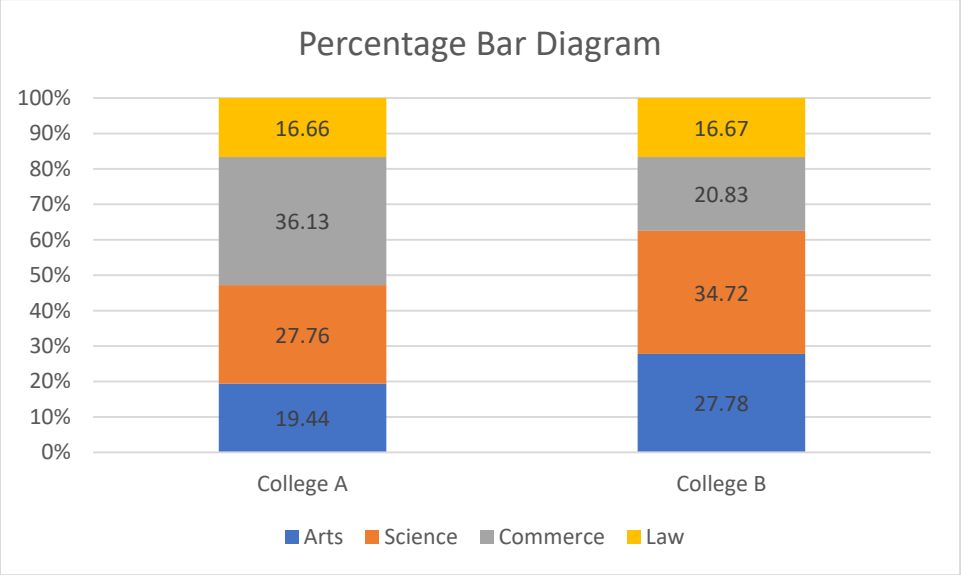


Fig : Percentage Bar Diagram representing faculty wise number of students in

College A and College B

3.7.2TWO-DIMENSIONAL DIAGRAMS (AREA DIAGRAMS)

- In one-dimensional diagrams or charts, only the length of the bar is taken into consideration. But, in two-dimensional diagrams, both its height and width are taken into account for presenting the data. These diagrams, also known as surface diagrams or area diagrams. They are in the shape of rectangles, squares or circles. The areas of squares, rectangles or circles are in proportion to the size of items which they represent.

  Pie Diagrams or Circle: These diagrams are used to show the total number of observations of different types in the data set on a percentage basis rather than on an absolute basis through a circle. Circles with area proportional to magnitudes are drawn to represent the total magnitude. Then the circle representing the total are divided into different segments according to the magnitude of the components. Pie diagrams are also called circular or angular diagrams.

If T is the total magnitude and R is the magnitude of a component, then the degree of any component part is given as $\left(\frac{R}{T}360\right)°$

Example : The following table gives the total outlay on rural development proposed in the first Five Years Plan and its break down into major items. Give a suitable diagrammatic representation of the data.

Item	**Amount (in crores of rupees)**
Agri & community development	360.43
Irrigation	167.97
Irrigation & Power (multipurpose projects)	265.90
Power	127.54
Transport & Communications	497.10
Industry	173.04
Social Services	339.81
Rehabilitation	85.00
Miscellaneous	51.99
Total	**2068.78**

Solution: The suitable diagram for representing the above data is sub-divided Pie diagram. The total area of the circle will represent the total amount i.e. 2068.78 crores of rupees and its components will represent the various items of expenditure under the plan. Now, let us prepare the following table:

Item	**Amount (in crores of rupees)**	**Angle**
Agri & community development	360.43	$\frac{360.43}{2068.78} * 360^0 = 62.7$
Irrigation	167.97	$\frac{167.97}{2068.78} * 360^0 = 29.2$

Irrigation & Power (multipurpose projects)	265.90	$\frac{265.90}{2068.78} * 360^0 = 46.3$
Power	127.54	$\frac{127.54}{2068.78} * 360^0 = 22.2$
Transport & Communications	497.10	$\frac{497.10}{2068.78} * 360^0 = 86.5$
Industry	173.04	$\frac{173.04}{2068.78} * 360^0 = 30.1$
Social Services	339.81	$\frac{339.81}{2068.78} * 360^0 = 59.1$
Rehabilitation	85.00	$\frac{85.00}{2068.78} * 360^0 = 14.8$
Miscellaneous	51.99	$\frac{51.99}{2068.78} * 360^0 = 9.0$
Total	2068.78	360

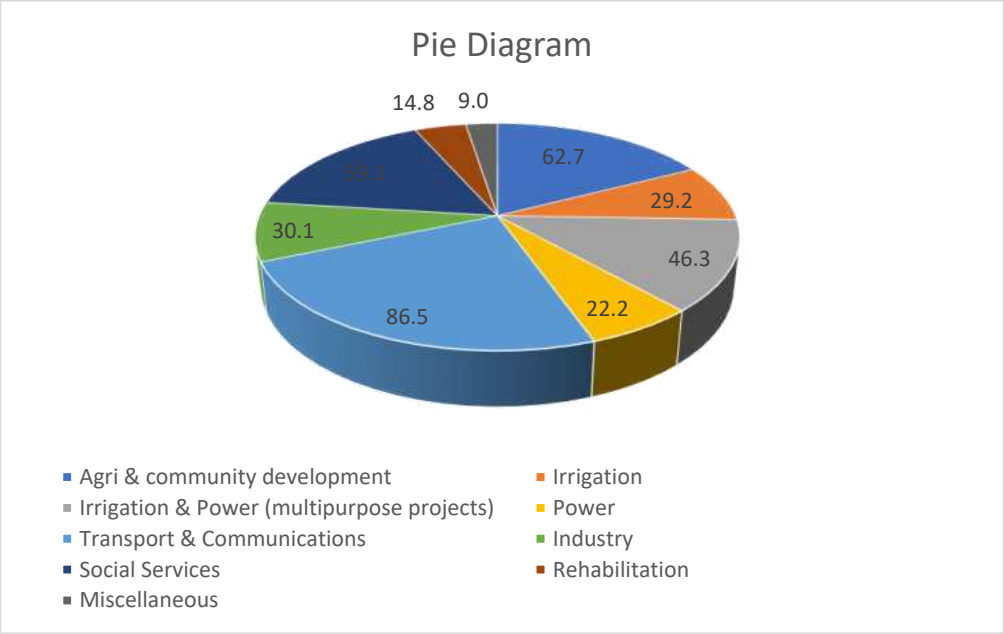


Fig: Pie Diagram representing the break down into major items of the total outlay on rural development

3.7.3 THREE-DIMENSIONAL DIAGRAMS

Cylinders, spheres, cubes, and so on are known as three-dimensional diagrams because three dimensions—length, breadth, and depth, are

taken into consideration for representing figures. Here the volumes of cubes, blocks or cylinders are in proportion to given values.

3.8 GRAPHICAL REPRESENTATIONS OF DATA

- Graphical representation makes the unwieldy data look more fascinating and brings the prominent and salient features of the data to light at a glance. Data of frequency distributions and time series are generally represented by graphs.
- With the help of graphs, the relationships among the variables can be studied. Generally, the horizontal scale is used for the value of the variable or magnitude of the class while the vertical scale is used for representing the frequencies of the values of the variable.
- In order to make frequency distributions easily understandable three types of graphs are usually drawn. These are:
  - Histogram
  - Frequency polygon and Frequency curve
  - Cumulative frequency curve or ogive

Histogram: The graph by which the frequencies of various class intervals of a frequency distribution with the help of adjacent vertical rectangles are shown is called a histogram. First of all, the actual class-intervals are to be marked on the x-axis choosing a suitable scale. Then taking these as bases, rectangles are to be drawn continuously on these bases. The heights of these rectangles vary in two cases, *i.e.* equal and unequal class intervals.

(a) Equal class-intervals: Here the heights of the rectangles are proportional to the corresponding frequency of the class so that the areas of the adjacent vertical rectangles are equal to the frequencies of the corresponding classes.

(b) Unequal class-intervals: If the classes are not of equal width, then the heights of the corresponding rectangles are so adjusted that the area of the rectangle is equal to the frequency of the corresponding class.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Thus, frequency of adjusted class = $\frac{freuency\ of\ the\ class}{width\ of\ that\ class}$

so that area of the rectangle = width of class x frequency of adjusted class = frequency of the class.

- Remarks:
  - In case of open-end classes, the histograms cannot be constructed as the width of the corresponding class is not known.
  - With the help of histogram, mode can be located graphically. In this case, the upper left corner of highest rectangle is joined to the right adjacent rectangle's left corner and right upper corner of highest rectangle to left adjacent rectangle's right corner. From the intersecting point of these lines a perpendicular line is drawn to the X-axis. The X-reading at that point gives the mode of the distribution.

Example : (Equal width) The following is the frequency distribution of the yield of sugar cane in tones per acre.

Class:	35 – 40	40 – 45	45 – 50	50 – 55	55 – 60	60 – 65	65 – 70
Frequency:	7	8	12	26	32	15	9

Draw a histogram representing the above distribution.

Solution: In this example, the class intervals are equal and so the heights of the rectangles will be proportional to the frequencies.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

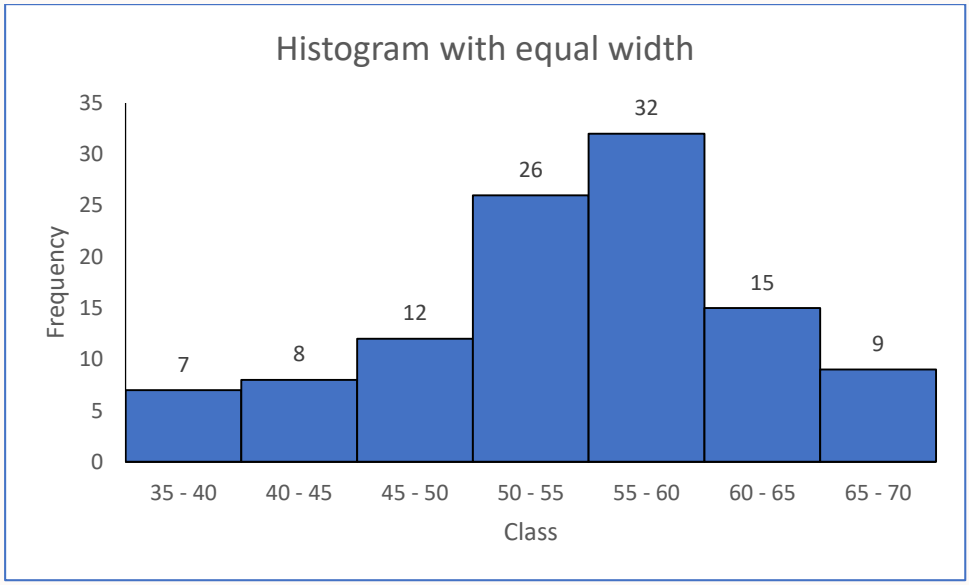


Fig: Histogram representing the yield of sugarcane in tones/acre

Remarks: Determination of the value of mode with the help of Histogram.

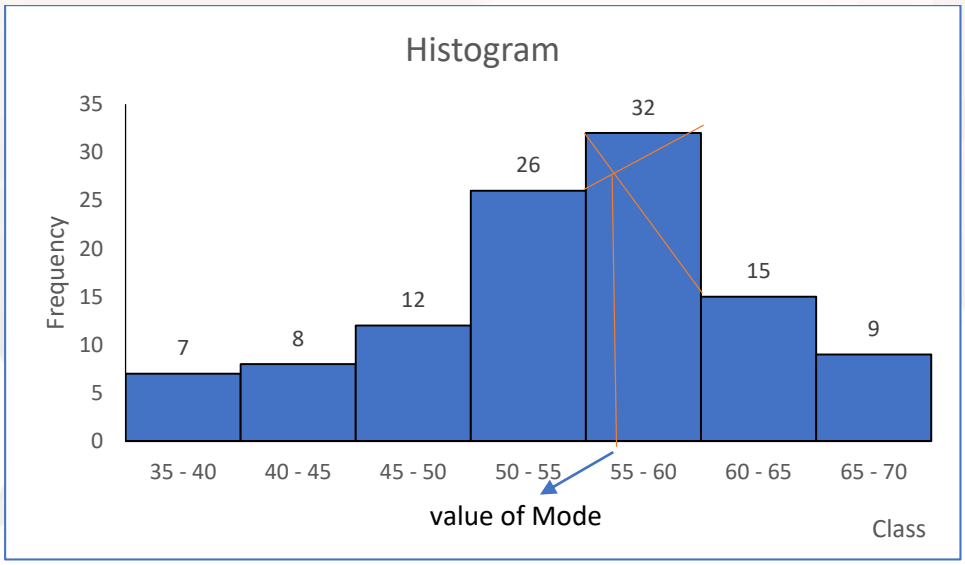


Fig: Determination of Mode with the help of Histogram

Example: (Unequal width) Represent the following data by means of a histogram

Weight (in kg.)	48 – 50	50 – 54	54 – 56	56 – 58	58 – 59	59 – 60	60 – 63
No. of boys	12	60	20	16	5	2	3

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Solution: Here the class intervals are of unequal width, so the heights of the rectangles will be proportional to the ratio of the frequencies to their respective class intervals. Thus, the heights of the rectangles are given in the following table against the respective classes:

Weight (in kg.)	No. of boys	Height of rectangle = frequency / class interval
48 – 50	12	12/2 = 6
50 – 54	60	60/4 = 15
54 – 56	20	20/2 = 10
56 – 58	16	16/2 = 8
58 – 59	5	5/1 = 5
59 – 60	2	2/1 = 2
60 - 63	3	3/3 = 1

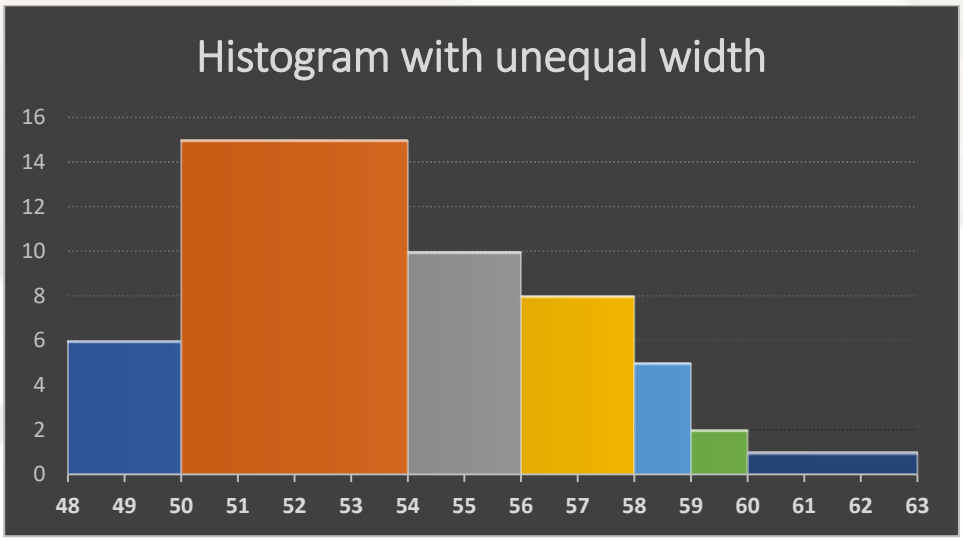


Fig: Histogram representing the weights of the boys

Frequency Polygon: It is a graphic representation of a frequency distribution whether ungrouped or discrete, grouped or continuous. For an ungrouped distribution, the frequency polygon is obtained by plotting the values of the variable along the horizontal axis and the frequencies along the vertical axis and then joining the plotted points by means of straight lines. The figure so obtained is known as the

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

frequency polygon for discrete series. For a grouped frequency distribution, the frequency polygon can be drawn in either of the following two ways.

- With the help of histogram: When the histogram has been drawn, the next step is to mark the mid-points of the tops of the set of adjacent rectangles. Join these marked points by means of straight lines. The figure so obtained is called frequency polygon.
- Without the help of histogram: plot the points with x-coordinates as the mid-values of corresponding classes and with y-coordinates as the frequencies of the mid points. These plotted points are joined by straight lines.

- **Remarks:**
  - Histogram is a two-dimensional diagram while frequency polygon is a line graph.
  - Two or more histograms cannot be drawn on the same graph paper while two or more frequency polygons can be drawn on the same graph paper.
  - In case the class intervals of a frequency distribution are not all equal then frequency polygon is usually not drawn.
  - The frequency polygons are formed as a closed figure so that the area under the curve should be equal to that of the histogram.

Example: (ungrouped data) An advertising company kept an account of response letters received each day over a period of 50 days. The observations were recorded as given in the following table:

Weekly wages (Rs.)	**No. of workers**
22	2
27	4
32	12
37	18
42	22

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

47	15
52	7

Solution: Here plot the frequencies (No. of workers) against the weekly wages respectively and then join these points by means of a straight line. Note that we will also plot the previous and next weekly wages (as analogous to the data set as 17 and 57 with frequencies 0) to start and end the polygon so that it touches the X-axis. These plot points are used only to give a closed shape to the polygon.

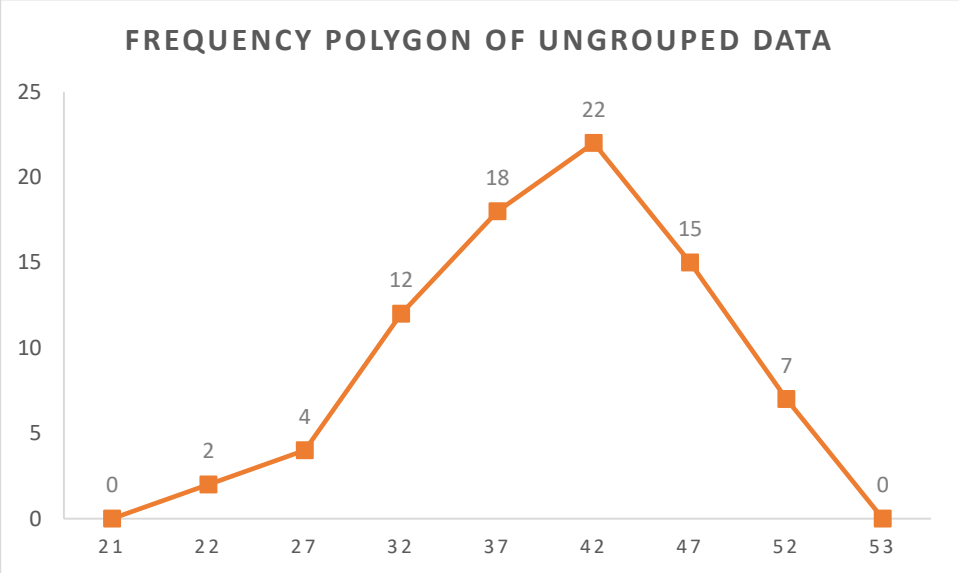


Fig: Frequency Polygon representing the no. of workers

Example: (grouped data with the help of histogram) Draw frequency polygon with the help of histogram for the following frequency distribution:

Weekly Wages (Rs.)	**No. of Workers**
30 – 31	2
32 – 33	9
34 – 35	25
36 – 37	30
38 – 39	49

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

40 – 41	62
42 – 43	39
44 – 45	20
46 – 47	11
48 – 49	3

Solution: First a histogram is drawn by taking the class boundaries at X-axis and the respective frequencies at Y-axis. Then by taking the mid-points of the tops of the set of adjacent rectangles, a straight line is joined through the marked points. To give a closed shape to the polygon, two class intervals are analogously taken at the start and end of the data (28-29 and 50- 51 with respective frequencies as 0).

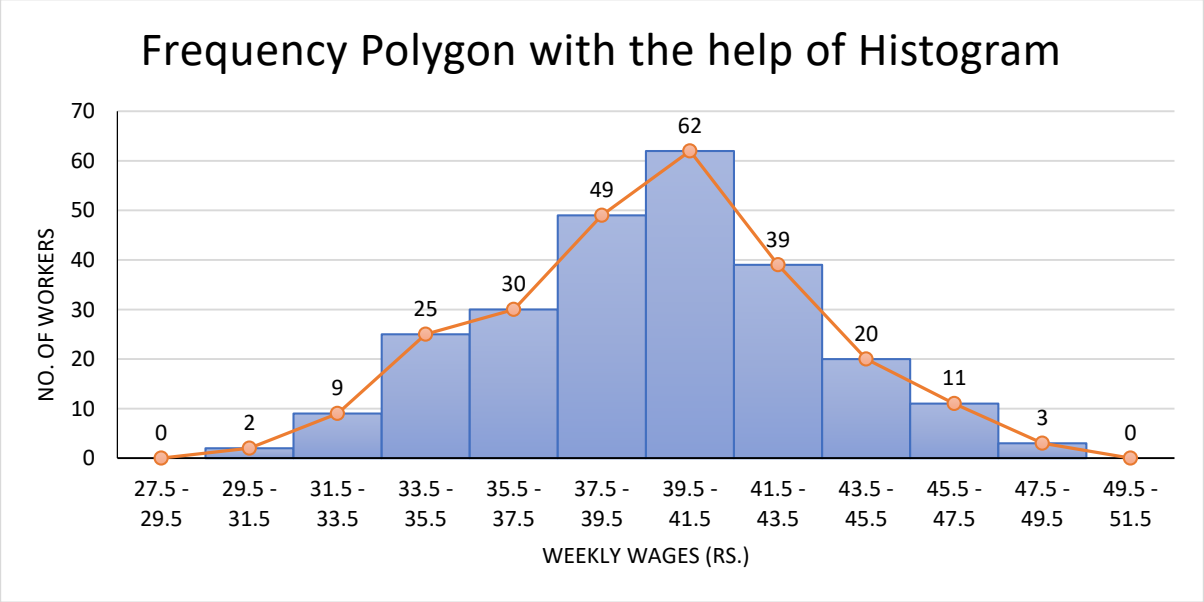


Fig: Frequency Polygon representing the no. of workers

Example: (grouped data without the help of histogram) The following table gives the distribution of height in inches for 100 students. Represent the data with the help of a frequency polygon without the help of histogram.

Interval	Frequency
57 – 60	3
60 – 63	12
63 – 66	31

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

66 – 69	37
69 – 72	16
72 – 75	1

Solution: Here the frequencies are plotted respectively against the mid points of the class intervals and then these points are joined by means of a straight line. The previous and next class intervals are plotted (analogously with the data set as taken as 54-57 and 75-78 with frequencies 0) to start and end the polygon.

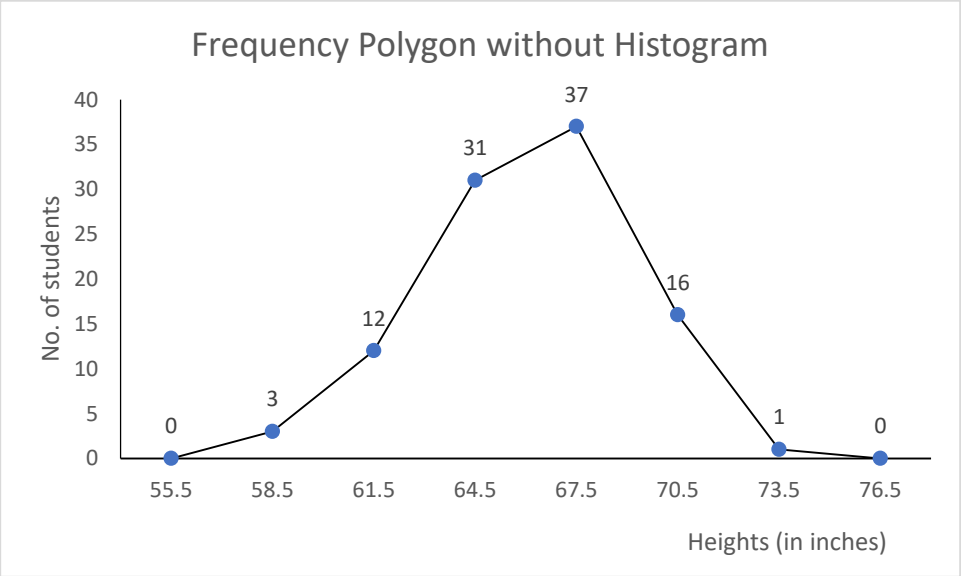


Fig: Frequency polygon representing the heights (in inches) of students

Frequency Curve: A frequency curve is obtained by smoothening the straight lines formed by joining the plotted points with their *x*-coordinates as the mid points of the classes and *y*-coordinates the corresponding frequencies, *i.e.* by smoothening the frequency polygon. The idea of smoothening is to remove the excess or deficit area present in the data and thus avoid sudden turns. In other words, the frequency curve is the limiting form of the frequency polygon when the number of vertices of the polygon is too large and the class intervals are too small.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Example: Draw a frequency curve for the following frequency distribution.

Pulse rate (in beats)	No. of patients
50 – 55	8
55 – 60	12
60 – 65	18
65 – 70	15
70 – 75	8
75 – 80	6

Solution: A frequency curve can be drawn with or without the help of histogram. Here for the given data, a frequency curve is drawn without the help of histogram by plotting the mid points of the class intervals (pulse rate) along the X-axis and the frequencies (no. of patients) along Y-axis. Then these plotted points are joined by free-hand smooth curve.

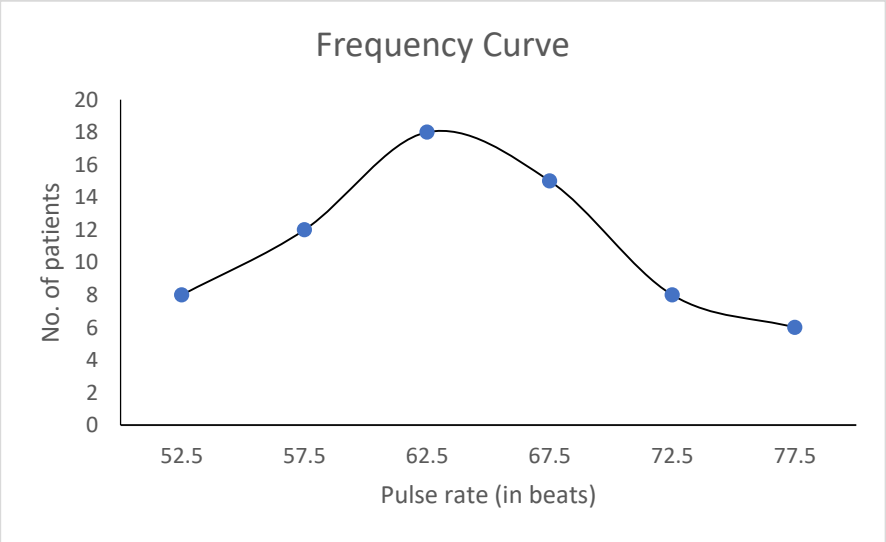


Fig: Frequency Curve representing Pulse rate (in beats) of a number of patients

Cumulative frequency curve or Ogive: The graphical representation of a cumulative frequency continuous distribution is known as the

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

cumulative frequency curve or ogive. Depending upon the pattern of drawing the graph there are two types of ogives, namely, 'less than ogive' and 'more than ogive'. To construct a 'less than ogive', the upper-class limits (or the upper-class boundaries) of the class intervals are taken as *x*-coordinates and the less than cumulative frequencies as the *y*-coordinates and then the plotted points are joined by a free-hand smooth curve. The frequency curve so obtained is the ogive by 'less than' method. However, if the points with lower limits (or the lower-class boundaries) are plotted as the x-coordinates and more than cumulative frequencies as the y-coordinates then the 'more than ogive' is obtained by joining the plotted points by a free-hand smooth curve.

- **Remarks:**
  - Two ogives, whether less than or more than type, can be readily compared by drawing them on the same graph paper. The presence of unequal class intervals poses no problem in their comparison, as it does in the case of comparison of two frequency polygons.
  - Ogive is helpful in determining the partitioning values (quartiles, defiles, etc.). In order to get the median of a distribution, a perpendicular line is to be drawn on the X-axis from the point of intersection of the less than olive and the more than give. The X-reading will give the value of the median.
  - The 'less than olive' is an increasing curve sloping upwards from left to right. Whereas, the 'more than give' is a decreasing curve sloping downwards from left to right.
  - The gives can determine the number or proportion of observations below or above a given value of the variable. Also, a number of points lying between certain class intervals can be determined.

Example: For the following frequency data construct a (i) less than ogive (ii) more than ogive

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Annual Profits ('000 Rs.)	No. of firms
25 – 34	5
35 – 44	19
45 – 54	13
55 – 64	11
65 – 74	2

Solution: To draw the 'less than' and 'more than' ogives we construct the following cumulative frequency table:

Annual Profits ('000 Rs)	Class Boundaries	No. of firms	Cumulative Frequency	
			Less than	More than
25 – 34	24.5 – 34.5	5	5	50
35 – 44	34.5 – 44.5	19	5+19 = 24	50 – 5 = 45
45 – 54	44.5 – 54.5	13	24+13 = 37	45 – 19 = 26
54 – 64	54.5 – 64.5	11	37+11 = 48	26 – 13 = 13
65 – 74	64.5 – 74.5	2	48+2 = 50	13 – 11 = 2
Total		50		

In order to draw the 'less than' ogive, upper class boundaries of the class intervals are marked on the X-axis. The points are plotted by taking the corresponding less than cumulative frequencies along the Y-axis and then the points are joined by a smooth curve with free hand.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

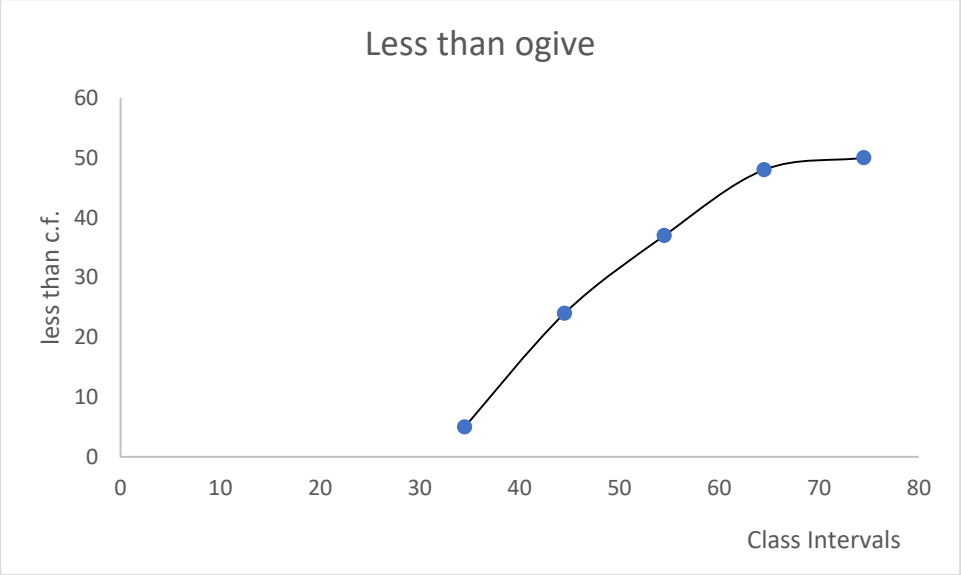


Fig: Less than ogive representing the annual profits of the number of firms

In a similar way, to draw the 'more than' ogive, the lower-class boundaries of the class intervals are marked on the X-axis and the corresponding more than cumulative frequencies are plotted along the Y-axis. Then the points are joined by a smooth curve with free hand.

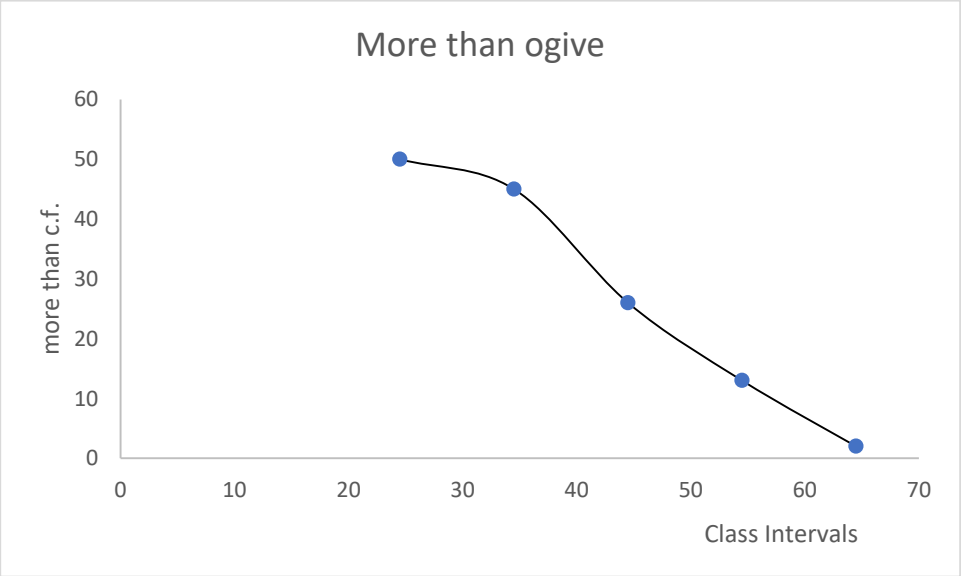


Fig: More than ogive representing the annual profits of the number of firms

The 'less than' ogive and 'more than' ogive both can be constructed in the same graph as follows:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

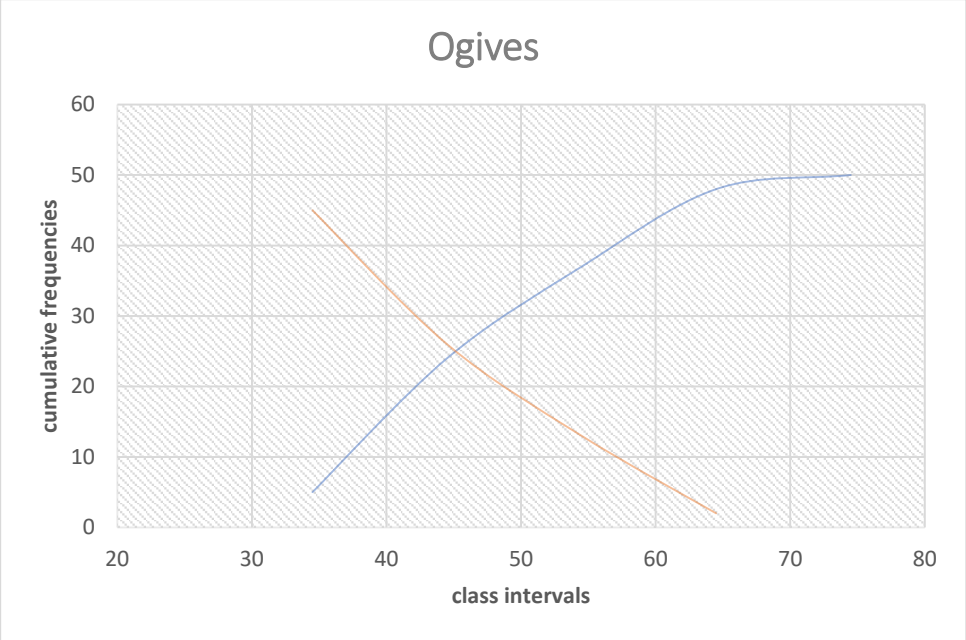


Fig: Less than and More than ogive representing the annual profits of the number of firms

Remarks: The value of median can be obtained from the less than and more than ogives as shown below:

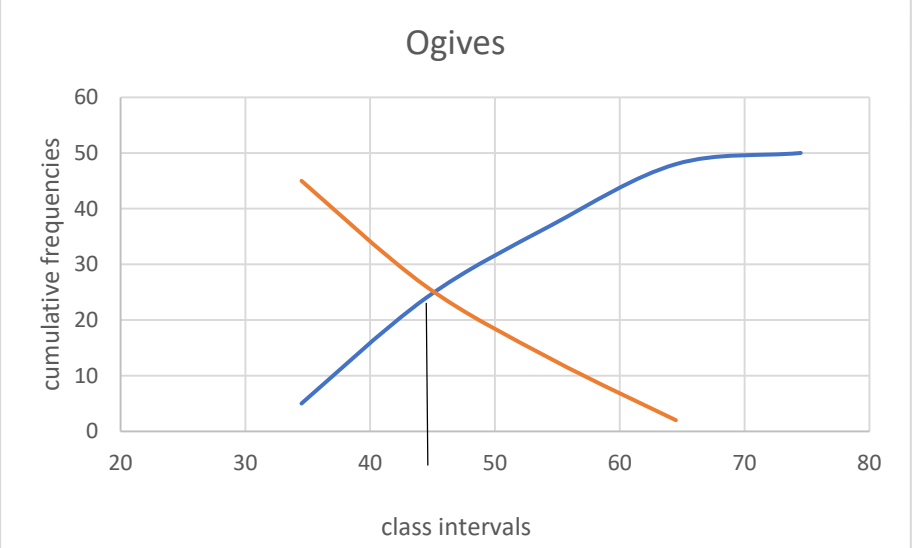


Fig: Determination of Median with the help of Ogives

****

GLOSSARY:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Frequency:**The number of units associated with each value of the variable is called frequency of that value.
- **Frequency distribution**: A systematic presentation of the values taken by variable together with corresponding frequencies is called a Frequency Distribution of the variable.
- **Cumulative Frequency:** If for each class, the frequencies given are the aggregate of the preceding frequencies they are known as cumulative frequencies.

REFERENCES

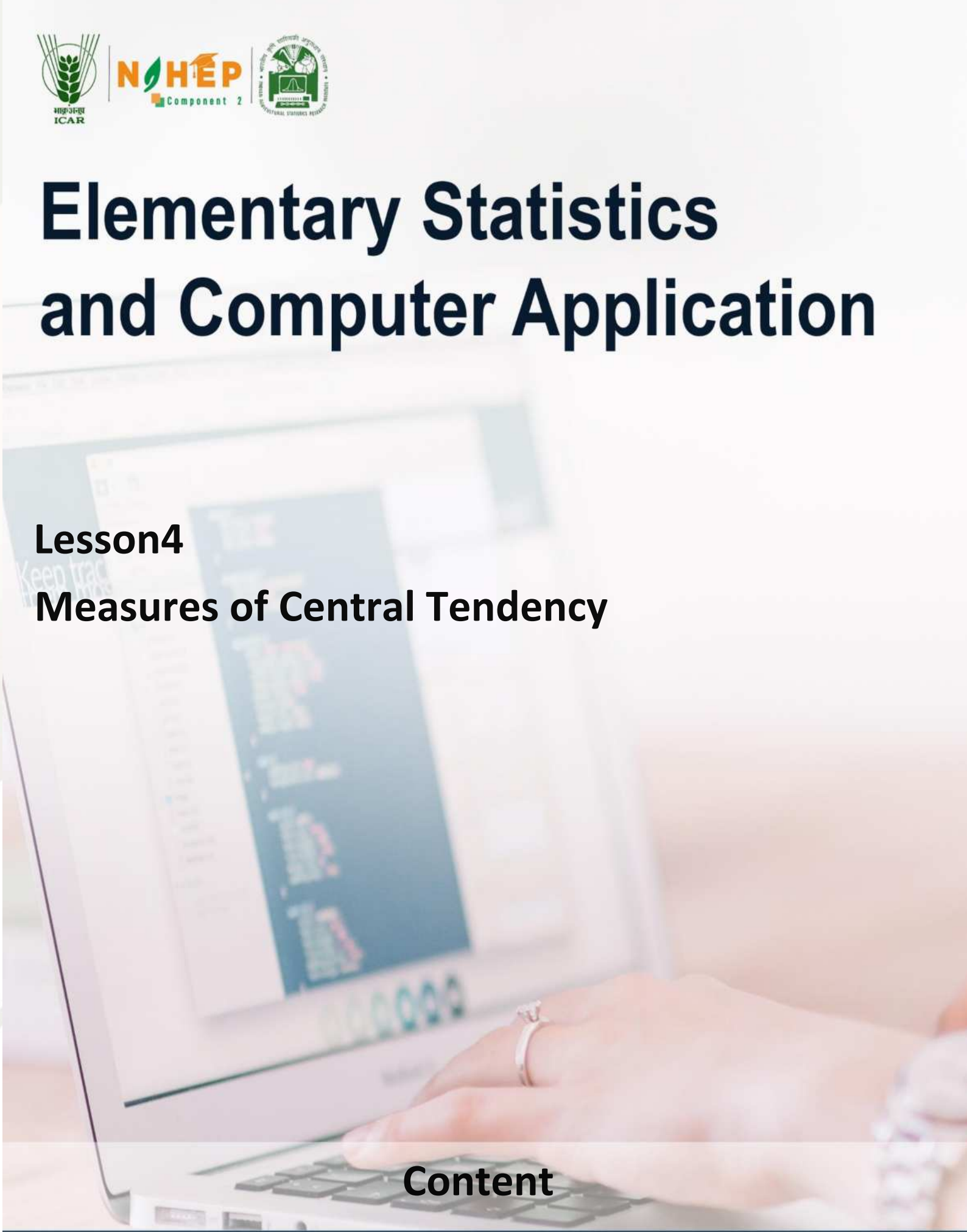
- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.
- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson4

Measures of Central Tendency



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 4	Measures of Central Tendency
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. Need for measures of central tendency.
2. Requisites for an ideal measure of central tendency.
3. Overview on different measures of central tendency: A.M., G.M, H.M, Median and Mode.
4. Discussion on different types of locational values.

4.1 NEED FOR MEASURES OF CENTRAL TENDENCY

- The term 'central tendency' was coined because observations (numerical values) in most data sets show a distinct tendency to group or cluster around a value of an observation located somewhere in the middle of all observations. It is necessary to identify or calculate this typical central value (also called average) to describe or project the characteristic of the entire data set. This descriptive value is the measure of the central tendency or location and methods of computing this central value are called measures of central tendency.

a. REQUISITES FOR AN IDEAL MEASURE OF CENTRALTENDENCY

- According to Professor Yule, the following are the characteristics to be satisfied by an ideal measure of central tendency:
    - It should be rigidly defined.
    - It should be readily comprehensible and easy to calculate.
    - It should be based on all the observations.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- It should be suitable for further mathematical treatment.
- It should be affected as little as possible by fluctuations of sampling.
- It should not be affected much by extreme values.

b. DIFFERENT MEASURES OF CENTRAL TENDENCY

- The various measures of central tendency or averages commonly used can be broadly classified in the following categories:
  - **Mean or Mathematical Averages**
    - Arithmetic Mean commonly called the Average
    - Geometric Mean
    - Harmonic Mean
  - **Median**
  - **Mode**

i. ARITHMETIC MEAN

Arithmetic mean of a set of observations is their sum divided by the number of observations, e.g., the arithmetic mean $\bar{x}$ of n observations $x_1, x_2, \ldots, x_n$ is given by:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In case of frequency distribution $(x_i, f_i), i=1,2,\ldots,n$, where f_i is the frequency of the variable x_i:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i} = \frac{1}{N}\sum_{i=1}^{n} f_ix_i \ , \quad \text{where} \sum_{i=1}^{n} f_i = N$$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

In case of grouped or continuous frequency distribution, x is taken as the mid-value of the corresponding class.

Example 1: In a survey of 5 textile companies, the profit (in Rs. lakh) earned during a year was 15, 20, 10, 35, and 32. Find the arithmetic mean of the profit earned.

Solution:

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{15 + 20 + 10 + 35 + 32}{5} = 22.4$$

Thus, the arithmetic mean of the profit earned by these companies during a year was Rs.22.4 lakh.

Example 2: The number of new orders received by a company over the last 25 working days were recorded as follows: 3, 0, 1, 4, 4, 4, 2, 5, 3, 6, 4, 5, 1, 4, 2, 3, 0, 2, 0, 5, 4, 2, 3, 3, 1. Calculate the arithmetic mean for the number of orders received over all similar working days

Solution: The arithmetic mean is

$$\bar{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = \frac{3 + 0 + 1 + \cdots + 3 + 1}{5} = 2.84$$

= 3 orders (approx.)

Alternatively, prepare the frequency table

No. of orders (x_i)	Frequency (f_i)	$f_i * x_i$
0	3	0
1	3	3

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

2	4	8
3	5	15
4	6	24
5	3	15
6	1	6
	$\sum_{i=1}^{n} f_i = 25$	71

Arithmetic Mean, $\bar{x} = \frac{1}{25}\sum_{i=1}^{7} f_i x_i = \frac{71}{25} = 2.84$ = 3 orders (approx.)

Example 3: Find the arithmetic mean for the following frequency distribution:

x:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

Solution:

x	f	f * x
1	5	5
2	9	18
3	12	36
4	17	68
5	14	70
6	10	60
7	6	42
TOTAL	73	299

$$\bar{x} = \frac{1}{73}\sum_{i=1}^{7} f_i x_i = \frac{299}{73} = 4.09$$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Example 4: Calculate the arithmetic mean of the marks from the following table:

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of students	12	18	27	20	17	6

Solution:

Marks	No. of students (f)	Mid-points (x)	f * x
0 – 10	12	5	60
10 – 20	18	15	270
20 – 30	27	25	675
30 – 40	20	35	700
40 – 50	17	45	765
50 – 60	6	55	330
Total	100		2800

$$\bar{x} = \frac{1}{100}\sum_{i=1}^{6} f_i x_i = \frac{2800}{100} = 28$$

Short-cut Method or Assumed Mean Method:

In case of ungrouped frequency distribution:

$$\bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i$$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

where, $d_i = x_i - A$, A is an arbitrary point.

In case of grouped or continuous frequency distribution:

$$\bar{x} = A + \frac{h}{N}\sum_{i=1}^{n} f_i d_i$$

where, $d_i = \frac{x_i - A}{h}$, A is an arbitrary point and h is the common width of class intervals.

Example 5: The human resource manager at a city hospital began a study of the overtime hours of the registered nurses. Fifteen nurses were selected at random, and following overtime hours during a month were recorded:

13 13 12 15 7 15 5 12 6 7 12 10 9

13 12 5 9 6 10 5 6 9 6 9 12

Calculate the arithmetic mean of overtime hours during the month.

Solution:

Overtime Hours (x_i)	No. of Nurses (f_i)	$d_i = x_i - A = x_i - 10$	$f_i d_i$
5	3	- 5	- 15
6	4	- 4	- 16
7	2	- 3	- 6
9	4	- 1	- 4
10 = A	2	0	0
12	5	2	10
13	3	3	9

15	2	5	10
	25		- 12

Here, A = 10 is taken as assumed mean. The required arithmetic mean of overtime hours is given as $\bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i = 10 + \frac{-12}{25} = 9.52$ hours

Example 6: A company is planning to improve plant safety. For this, accident data for the last 50 weeks was compiled. These data are grouped into the frequency distribution as shown below. Calculate the A.M. of the number of accidents per week.

No. of accidents	0 – 4	5 – 9	10 – 14	15 – 19	20 – 24
No. of weeks	5	22	13	8	2

Solution:

No. of Accidents	Mid-value (x_i)	$d_i = (x_i - A)/h = (x_i - 12)/5$	No. of Weeks (f_i)	$f_i d_i$
0 – 4	2	- 2	5	- 10
5 – 9	7	- 1	22	- 22
10 – 14	12 = A	0	13	0
15 – 19	17	1	8	8

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

20 – 24	22	2	2	4
			50	- 20

The required arithmetic mean

$$\bar{x} = A + \frac{h}{N}\sum_{i=1}^{n} f_i d_i = 12 + \frac{5}{50}(-20)$$

= 10 accidents per week.

Advantages:

- The calculation of arithmetic mean is simple and it is unique, that is, every data set has one and only one mean.
- The calculation of arithmetic mean is based on all values given in the data set.
- The arithmetic mean is reliable single value that reflects all values in the data set.
- The arithmetic mean is least affected by fluctuations in the sample size. In other words, its value, determined from various samples drawn from a population, varies by the least possible amount.
- It can be readily put to algebraic treatment.

Disadvantages:

- Arithmetic mean cannot be located graphically.
- The value of A.M. cannot be calculated accurately for unequal and open-ended class intervals either at the beginning or end of the given frequency distribution.
- The mean cannot be calculated for qualitative characteristics such as intelligence, honesty, beauty, or loyalty.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Arithmetic mean cannot be obtained if a single observation is missing or lost.

Properties of A.M.

- The sum of the deviations of the values from their A.M. is always zero, i.e., if $x_1, x_2, \ldots, x_n$ be the values then $\sum(x_i - \bar{x}) = 0$ or $\sum f_i(x_i - \bar{x}) = 0$.
- The sum of the squares of the deviations of a set of values is minimum when taken about mean, i.e.

  $Z = \sum_{i=1}^{n} f_i(x_i - A)^2$ is minimum at the point $A = \bar{x}$.
- If $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ be the A.M. of k-component series of sizes n_i (i=1,2,…,k) respectively, then the A.M. of the combined series is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_1 + n_2 + \cdots + n_k} = \frac{\sum n_i\bar{x}_i}{\sum n_i}$$

Uses:

AM is the most important, widely used and best measures of central tendency. In the determination of average income, average price, average cost of production, average sales, etc. i.e., those phenomena which are capable of direct quantitative measurement, AM is the most appropriate measure.

ii. GEOMETRIC MEAN

In many business and economics problems, we deal with quantities (variables) that change over aperiod of time. In such cases the aim is to know an average percentage change rather than simpleaverage value to represent the average growth or declining rate in the variable value over

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

a periodof time. The measure of central tendency used in such cases is called geometric mean(G.M.). The specific application of G.M. is to show multiplicative effects over time in compoundinterest and inflation calculations.

Geometric mean of a set of n observations is thenth root of their product. Thus, the geometric mean 'G' of n observations x_i, i=1,2,…,n is

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

Taking logarithm on both sides,

$$\log G = \frac{1}{n}(\log x_1 + \log x_2 + . \dots + \log x_n) = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

$$G = \text{Antilog}\left[\frac{1}{n}\sum_{i=1}^{n} \log x_i\right]$$

In case of frequency distribution $x_i|f_i$, i=1,2,…,n geometric mean, G is given by

$$G = \left[{x_1}^{f_1} . {x_2}^{f_2} \dots \dots {x_n}^{f_n}\right]^{\frac{1}{N}}, \text{ where } N = \sum_{i=1}^{n} f_i$$

Taking logarithms on both sides,

$$\log G = \frac{1}{N}(f_1 \log x_1 + f_2 \log x_2 + . \dots + f_n \log x_n) = \frac{1}{N}\sum_{i=1}^{n} f_i \log x_i$$

$$G = \text{Antilog}\left[\frac{1}{N}\sum_{i=1}^{n} f_i \log x_i\right], \qquad N = \sum_{i=1}^{n} f_i$$

Thus, the logarithm of G is the arithmetic mean of the logarithms of the given values.

In case of grouped or continuous frequency distributions, x is taken to be the value corresponding to the mid-point of the class-intervals.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Example 1: The annual growth (in %) of output of a company in the last five years is given as:

5.0 7.5 2.5 5.0 10.0

Find the average growth rate.

Solution: The average growth rate is given as:

$G = (5 \times 7.5 \times 2.5 \times 5 \times 10)^{1/5} = 5.9\ \%$

Alternatively, $G = \text{Antilog}\ \frac{1}{5}(\log 5 + \log 7.5 + \log 2.5 + \log 5 + \log 10)$

$= \text{Antilog}\ 0.734 = 5.9\%$

Example 2: Calculate the GM of the following distribution.

Class:	0 – 10	10 -20	20 – 30	30 – 40	40 – 50
Frequency:	10	17	25	20	8

Solution: Let us prepare the following table

Class	f	x	log x	f.log x
0 – 10	10	5	0.6990	6.9900
10 – 20	17	15	1.1761	19.9937
20 – 30	25	25	1.3979	34.9475
30 – 40	20	35	1.5441	30.8820
40 – 50	8	45	1.6532	13.2256

Total	N = 80			106.0388

$$G = \text{Antilog}\left[\frac{106.0388}{80}\right]$$

= Antilog (1.3255) = 21.16

Advantages:

- It is based upon all the-observations.
- The value of G.M. is not much affected byextreme observations and is computed by takingall the observations into account.
- It is useful for averaging ratio and percentage as well as in determining rate of increase anddecrease.
- In the calculation of G.M. more weight is given to smaller values and less weight to highervalues.
- It is suitable for algebraic manipulations.

Disadvantages:

- The calculation of G.M. as compared to A.M., is more difficult and complex.
- The value of G.M. cannot be calculated when any of the observations in the data set is eithernegative or zero.

Uses:

- To find the rate of population growth and the rate of interest.
- In the construction of index numbers.

iii. HARMONIC MEAN

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The harmonic mean (H.M.) of a set of observations is defined as the reciprocal of the arithmetic meanof the reciprocal of the given values. Thus, harmonic mean, H, of n observations x_i, i = 1,2,…,n is

$$H = \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right)} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}$$

In case of frequency distribution, $x_i|f_i$, i=1,2,…,n harmonic mean, H is given by

$$H = \frac{1}{\frac{1}{N}\left(\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}\right)} = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}\frac{f_i}{x_i}}, \qquad N = \sum_{i=1}^{n} f_i$$

Example 1: Calculate the harmonic mean of 4, 8 and 12.

Solution: The required HM is given as

$$H = \frac{1}{\frac{1}{3}\left(\frac{1}{4}+\frac{1}{8}+\frac{1}{12}\right)} = \frac{3}{\frac{6+3+2}{24}} = \frac{72}{11} = 6.55$$

Example 2: Calculate the HM for the following data

Class:	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency:	4	5	10	8	3

Solution: Let us prepare the following table

Class	f	x	f/x
10 – 20	4	15	0.27
20 – 30	5	25	0.20
30 – 40	10	35	0.29
40 – 50	8	45	0.18
50 – 60	3	55	0.05
Total	N = 30		0.99

The required HM is

$$H = \frac{30}{0.99} = 30.3$$

Example 3: Suppose a person travels 10 miles at 4 mph and again 12 miles at 5 mph. What is his average speed?

Solution: Here the rates are miles (x) per hour (f). The average speed is given as

$$H = \frac{f_1 + f_2}{\frac{f_1}{x_1} + \frac{f_2}{x_2}} = \frac{10 + 12}{\frac{10}{4} + \frac{12}{5}} = \frac{22}{4.9} = 4.49$$

Thus, the average speed is 4.49 mph.

Advantages:

- The H.M. is rigidly defined, based upon all the observations.
- While calculating H.M., more weightage is given to smaller values in a data set.
- H.M. is not affected by fluctuations of sampling.
- The formula for H.M. is suitable for further mathematical treatment.

Disadvantages:

- The H.M. of any data set cannot be calculated if it has negative and/or zero elements.
- The calculation of H.M. involves complicated calculations.

Uses:

The harmonic mean is particularly useful for computation of average rates and ratios. Such rates andratios are generally used to express relations between two different types of measuring units that canbe

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

expressed reciprocally. For example, distance (in km), and time (in hours).

Relationships among A.M., G.M. and H.M.

- A.M. ≥ G.M. ≥ H.M.
- A.M. x H.M. = $(G.M.)^2$

iv. MEDIAN

Median may be defined as the middle value in the data set when its elements are arranged in a sequential order, that is, in either ascending or descending order of magnitude. It is called a middle value in an ordered sequence of data in the sense that half of the observations are smaller and half are larger than this value. The median is thus a measure of the location or centrality of the observations.

The median can be calculated for both ungrouped and grouped data sets.

Discrete Series

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 (Med) is represented by the numerical value corresponding to the positioning point of (n + 1)/2 ordered observation.

That is, Med = Size or value of $\left(\frac{n+1}{2}\right)$th observation in the data array

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

(ii) If the number of observations (n) is an even number, then the median is defined as the arithmeticmean of the numerical values of $\left(\frac{n}{2}\right)$th and$\left(\frac{n}{2}+1\right)$th observations in the data array. That is,

$$\text{Med} = \frac{\frac{n}{2}\text{th obs.}+\left(\frac{n}{2}+1\right)\text{th obs.}}{2}$$

Example 1: Number of insects per plant is given as follows:17, 27, 30, 26, 24, 18, 19, 28, 23, 25, and 20. Findout the median value of number of insects perplant.

Solution: Let us arrange the data in ascendingorder of their values as follows: 17, 18, 19, 20,23, 24, 25, 26, 27, 28, and 30. Here, we have 11 observations, so the middle most observation is the $\left(\frac{11+1}{2}\right)$= 6th observation and thevalue of the sixth observation is 24. Hence, the median value of number of insects per plant is 24.

Example 2: Calculate the median of the following data that relates to the number of patientsexamined per hour in the outpatient ward (OPD) in a hospital: 10, 12, 15, 20, 13, 24, 17, 18.

Solution: Arranging the data in ascendingorder, we have

10 12 13 15 17 18 20 24

$$\text{Median} = \text{A.M. of } \frac{\frac{n}{2}\text{th obs.}+\left(\frac{n}{2}+1\right)\text{th obs.}}{2} = \frac{15+17}{2} = 16.$$

Thus, median number of patients examined per hour in OPD in a hospital are 16.

Continuous Series

To find the median value for continuous series, first identify the class interval which contains the median value or $\left(\frac{n}{2}\right)$ th observation of the

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

data set. To identify such class interval, find the cumulative frequency of each class until the class for which the cumulative frequency is equal to or greater than the value of $\left(\frac{n}{2}\right)$th observation. The following formula is used to determine the median of grouped data:

$$\text{Med} = l + \frac{\frac{n}{2} - f_c}{f} \times h$$

where l = lower class limit (or boundary) of the median class interval.

f_c= cumulative frequency of the class prior to the median class interval.

f = frequency of the median class.

h = width of the median class interval.

n = total number of observations in the distribution.

Example 3: A survey was conducted to determine the wages (in Rs.) of 43 labourers. The result of such a survey is as follows:

Wages (in Rs.)	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of labours	3	5	20	10	5

Solution:

Wages	No. of Workers (f)	cumulative frequency (c.f.)
20 – 30	3	3
30 – 40	5	8
40 – 50	20	28

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

50 – 60	10	38
60 – 70	5	43
	N=43	

Here N/2 = 43/2 = 21.5. The cumulative frequency just greater than or equal to 21.5 is 28 and the corresponding class is 40 – 50. Thus, median class is 40 – 50. Hence the value of median is

$$\text{Med} = 40 + \frac{21.5 - 8}{20} \times 10$$

$$= 40 + 6.75 = 46.75$$

Thus, the median wage is Rs. 46.75.

Example 4: For the following table calculate the value of median.

Class Interval	Frequency
Less than 150	150
151 – 200	580
201 – 250	900
251 – 300	500
301 – 350	600
351 and above	270

Solution:

Class Interval	Frequency	c.f.
Less than 150	150	150
151 – 200	580	730
201 – 250	900	1630
251 – 300	500	2130
301 – 350	600	2730

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

351 and above	270	3000
	N=3000	

Here N/2 = 3000/2 = 1500. Therefore, the median class is 201 – 250 and the value of the median is

$$\text{Med} = 201 + \frac{1500 - 730}{900} \times 50$$

$$= 201 + 42.77 = 243.77$$

Hence the value of the median is 243.77.

Advantages:

- Median is unique, i.e., like mean, there is only one median for a set of data.
- The value of median is easy to understand and may be calculated from any type of data.
- The extreme values in the data set does not affect the calculation of the median.
- The median is considered the best statistical technique for studying the qualitative attribute of an observation in the data set.
- The median value may be calculated for an open-ended distribution of data set.

Disadvantages:

- The median is not capable of algebraic treatment. For example, the median of two or more sets of data cannot be determined.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The value of median is affected more by sampling variations as compared with mean.
- Median is not based on all the observations.
- Since median is an average of position, therefore arranging the data in ascending or descending order of magnitude is time consuming in case of a large number of observations.

Uses:

The median is helpful in understanding the characteristic of a data set when

- observations are qualitative in nature
- extreme values are present in the data set

v. MODE

The mode is that value of an observation which occurs most frequently in the data set, that is, the point with the highest frequency. The concept of mode is of great use to large scale manufacturers of consumable items such as ready-made garments, shoe-makers, and so on. In all such cases it is important to know the size that fits most persons rather than 'mean' size.

Discrete Series: Mode of discrete series is the value which occurs most frequently.

Example 1: Consider the data set 1, 2, 3, 4, 4, 5.

The mode for this data is 4 because 4 occurs most frequently (twice) in the data.

Example 2: Look at the following discrete series:

Variable:	10	20	30	40	50
Frequency:	2	8	20	10	5

Here, as you can see the maximum frequency is 20, the value of mode is 30.

Continuous Series: In case of grouped data the following formula is used for calculating mode:

$$\text{Mo} = \text{l} + \frac{\text{f}_\text{m} - \text{f}_{\text{m}-1}}{2\text{f}_\text{m} - \text{f}_{\text{m}-1} - \text{f}_{\text{m}+1}} \times \text{h}$$

where l = lower limit (boundary) of the modal class

f_m= frequency of the model class

f_{m-1} = frequency of the class preceding the modal class

f_{m+1} = frequency of the class following the modal class

h = width of the modal class interval

Modal class is that class which has the highest frequency.

Example 3: Find the mode for the following distribution:

Wages (in Rs.)	Frequency
0 – 10	5
10 – 20	8
20 – 30	7
30 – 40	12
40 – 50	28
50 – 60	20
60 – 70	10
70 – 80	10

The value of mode is

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

$$\text{Mo} = 40 + \frac{28 - 12}{2 \times 28 - 12 - 20} \times 10$$

$$= 40 + 6.666 = 46.67$$

Thus, the value of the mode is Rs. 46.67.

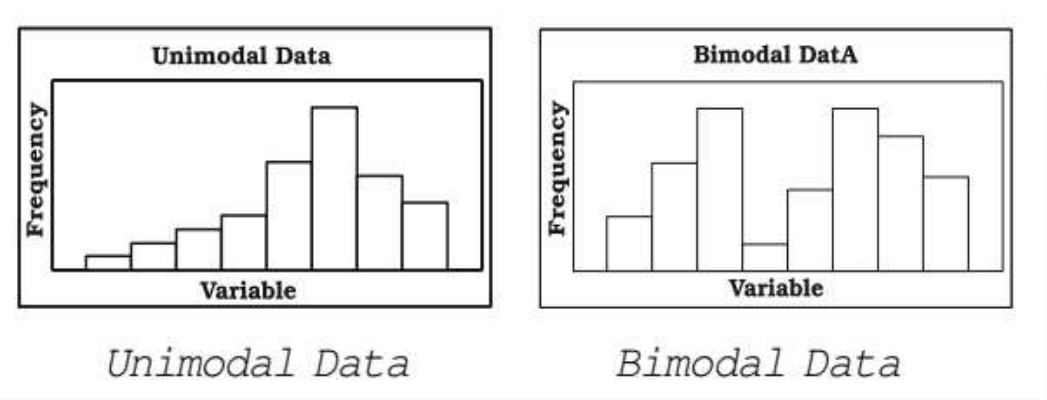
Advantages:

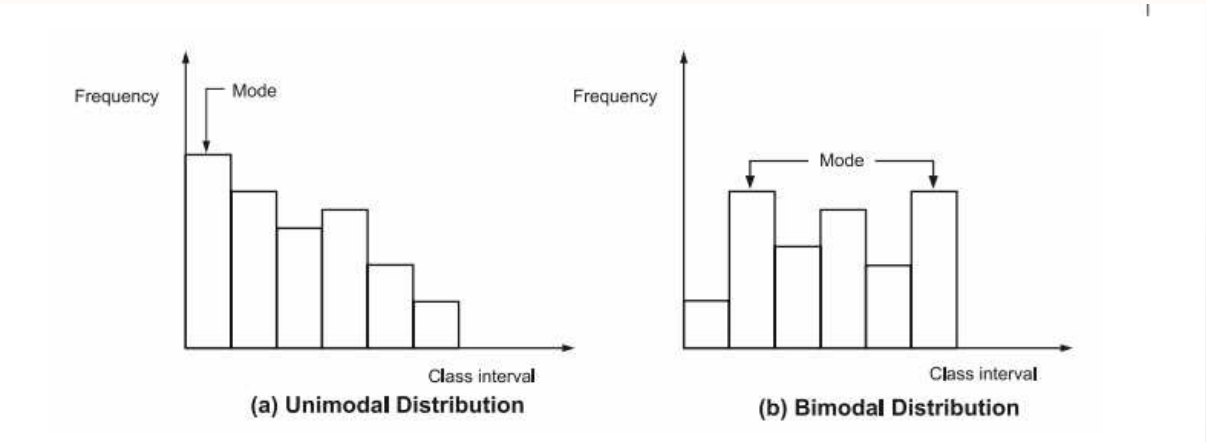
- Mode value is easy to understand and to calculate. Mode class can also be located by inspection.
- The mode is not affected by the extreme values in the distribution.
- The mode value can also be calculated for open-ended frequency distributions.
- The mode can be used to describe quantitative as well as qualitative data. For example, its value is used for comparing consumer preferences for various types of products, say cigarettes, soaps, toothpastes, or other products.

Disadvantages:

- Mode is not a rigidly defined measure as there are several methods for calculating its value.
- It is difficult to locate modal class in the case of multi-modal frequency distributions.
- Mode is not suitable for algebraic manipulations.
- When data sets contain more than one modes, such values are difficult to interpret and compare.
- Mode is a poor measure when most frequently occurring values of an observation do not appear close to the centre of the data or not even a unique value.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

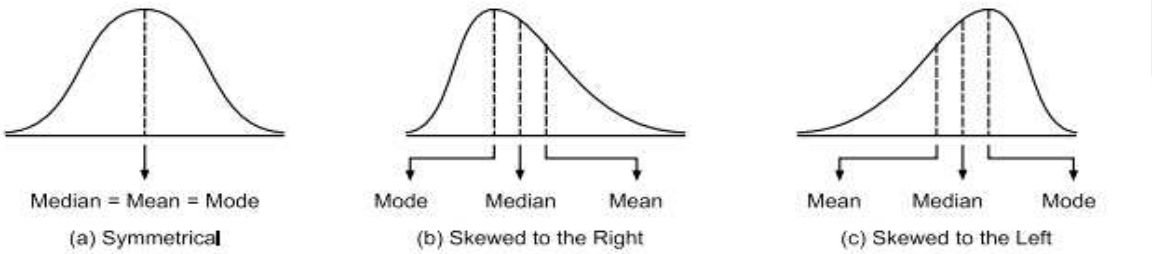




Uses: Mode is mostly used in business and commerce. Meteorological forecasts are based on mode.

c. RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE

In a unimodal and symmetrical distribution, the values of mean, median, and mode are equal as indicated in the following Figure. In other words, when all these three values are not equal to each other, thedistribution is not symmetrical.



Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

For asymmetrical distribution, Karl Pearson has suggested a relationship between these three measures of central tendency as:

$$\text{Mean} - \text{Mode} = 3\,(\text{Mean} - \text{Median})$$

Or, $\text{Mode} = 3\,\text{Median} - 2\,\text{Mean}$

d. PARTITION VALUES: QUARTILES, DECILES AND PERCENTILES

The measures of central tendency which are used for dividing the data into several equal parts are called partition values. A data set can be divided into four, ten, and hundred parts of equal size and the corresponding partition values are called quartiles, deciles, and percentiles. All thesevalues can be determined in the same way as median. The only difference is in their location.

Quartiles: The values of observations in a data set, when arranged in an ordered sequence, can bedivided into four equal parts, or quarters, using three quartiles namely Q_1, Q_2, and Q_3. The first quartile Q_1 divides a distribution in such a way that 25% (=n/4) of observations have a value less than Q_1 and 75% (= 3n/4) have a value more than Q_1. The second quartile Q_2 has the same number of observations above and below it. It is therefore same as median value. The quartile Q_3 divides the data set in such a way that 75% of the observations have a value less than Q_3 and 25% have a value more than Q_3.

The generalized formula for calculating quartiles in case of grouped data is:

$$Q_i = l + \frac{i\frac{N}{4} - f_c}{f} \times h; \quad i = 1,2,3$$

where l = lower limit of the i^{th} quartile class interval

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

f_c = cumulative frequency prior to the i^{th} quartile class interval

f = frequency of the i^{th} quartile class interval

h = width of the class interval

Deciles: The values of observations in a data set when arranged in an ordered sequence can be divided into ten equal parts, using nine deciles, D_i (i = 1, 2, . . ., 9). The generalized formula for calculating deciles in case of grouped data is:

$$D_i = l + \frac{i\frac{N}{10} - f_c}{f} \times h; \quad i = 1,2, \ldots ,9$$

where, the symbols have their usual meaning and interpretation.

Percentiles: The values of observations in a data when arranged in an ordered sequence can be divided into hundred equal parts using ninety-nine percentiles, P_i (i = 1, 2, . . ., 99). In general, the i^{th} percentile is a number that has i% of the data values at or below it and (100 – i)% of the data valuesat or above it. The lower quartile (Q_1), median and upper quartile (Q_3) are also the 25th percentile, 50th percentile and 75th percentile, respectively. The generalized formula for calculating percentiles in case ofgrouped data is:

$$P_i = l + \frac{i\frac{N}{100} - f_c}{f} \times h; \quad i = 1,2, \ldots ,99$$

where the symbols have their usual meaning and interpretation.

Example 1: The following distribution gives the pattern of overtime work per week done by100employees of a company. Calculate median, first quartile, seventh decile and sixtieth percentile.

Overtime hours: 10 – 15 15 – 20 20 – 25 25 – 30 30 – 35 35 – 40

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

No. of employees: 11 20 35 20 8 6

Solution: The calculation of median (Q_2), first quartile (Q_1), seventh decile (D_7) and sixtieth percentile (P_{60}) are as follows:

Overtime hours	No. of employees	c.f. (less than type)
10 – 15	11	11
15 – 20	20	31
20 – 25	35	66
25 – 30	20	86
30 – 35	8	94
35 – 40	6	100

Since the number of observations in this data set are 100, the median value is (n/2)th = (100 ÷ 2)th= 50th observation. This observation lies in the class interval 20–25. Thus, median is:

$$\text{Med} = l + \frac{(n/2) - cf}{f} \times h$$

$$= 20 + \frac{50 - 31}{35} \times 5 = 20 + 2.714 = 22.714 \text{ hours}$$

Q_1 = value of $(n/4)$ th observation = value of (100/4)th = 25th observation

Thus $Q_1 = l + \frac{(n/4) - cf}{f} \times h = 15 + \frac{25 - 11}{20} \times 5 = 15 + 3.5 = 18.5$ hours

D_7 = value of $(7n/10)$th observation = value of $(7 \times 100)/10$ = 70th observation

Thus $D_7 = l + \frac{(7n/10) - cf}{f} \times h = 25 + \frac{70 - 66}{20} \times 5 = 25 + 1 = 26$ hours

P_{60} = Value of $(60n/100)$ th observation = $60 \times (100/100)$ = 60th observation

Thus $P_{60} = l + \frac{(60 \times n/100) - cf}{f} \times h = 20 + \frac{60 - 31}{35} \times 5 = 24.14$ hours

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

4.6 Graphical Method for Calculating Partition Values:

The graphical method of determining various partition values can be summarized into followingsteps:

(i) Draw an ogive (cumulative frequency curve) by 'less than' method.

(ii) Take the values of observations or class intervals along the horizontal scale (i.e. x-axis) andcumulative frequency along vertical scale (i.e., y-axis).

(iii) Determine the median value, that is, value of (n/2)th observation, where n is the total number of observations in the data set.

(iv) Locate this value on the y-axis and from this point draw a line parallel to the x-axis meeting the ogive at a point, say P. Draw a perpendicular on x-axis from P and it meets the x-axis at a point, say M. Then abscissa of 'M' gives the value of median.

(v) To locate the values of Q_1 (or Q_3), mark the points along y-axis corresponding to N/4 (or 3N/4) and proceed exactly similarly.

(vi) Other partition values, viz., deciles and percentiles, can be similarly located from 'ogive'.

Example 2: Draw the cumulative frequency curve and locate the values of quartiles for the following distribution showing the number of marks of 59 students in Statistics.

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70

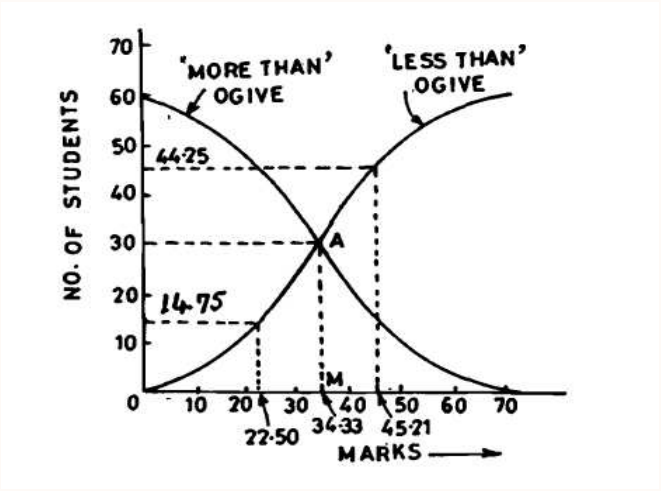
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

No. of students	4	8	11	15	12	6	3

Solution: To determine the quartiles graphically, let us prepare the following table

Marks	No. of students	Less than c.f.
0 – 10	4	4
10 – 20	8	12
20 – 30	11	23
30 – 40	15	38
40 – 50	12	50
50 – 60	6	56
60 – 70	3	59

Taking the marks along x-axis and c.f. along y-axis, plot thecumulative frequencies, viz., 4,12,23,...,59 against the upper limits of thecorresponding classes, viz., 10,20,...,70 respectively. The smooth curve obtainedon joining these points is called ogive or more particularly 'less than' ogive.



To locate graphically the value of median, mark a point corresponding toN/2 along y-axis. At this point draw a line parallel to x-axis meeting the ogive atthe point. 'A' (say). From 'A' draw a line perpendicular to x-axis meeting it in 'M'(say). Then abscissa of 'M' gives the value of median.

To locate the values of Q_1 (or Q_3), mark the points along y-axis correspondingtoN/4 (or3N/4) and proceedexactly similarly.

In the above example, we get from ogive, **Median= 34·33, Q_1-= 22·50 and Q_3= 45·21**.

****

GLOSSARY:

- **Arithmetic Mean:** Arithmetic mean of a set of observations is their sum divided by the number of observations
- **Geometric Mean:** Geometric mean of a set of n observations is the n^{th} root of their product.
- **Harmonic Mean:**The harmonic mean (H.M.) of a set of observations is defined as the reciprocal of the arithmetic mean of the reciprocal of the given values.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Median:** Median may be defined as the middle value in the data set when its elements are arranged in a sequential order
- **Mode:** The mode is that value of an observation which has the highest frequency.
- **Quartiles**: The values of observations in a data set, when arranged in an ordered sequence, can be divided into four equal parts.
- **Deciles:**The values of observations in a data set, when arranged in an ordered sequence, can be divided into ten equal parts.
- **Percentiles**: The values of observations in a data set, when arranged in an ordered sequence, can be divided into hundred equal parts

REFERENCES

- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.
- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

DESIGNED AND DEVELOPED UNDER THE AEGIS OF

NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Details

Course Name	Elementary Statistics And Computer Application
Lesson5	Measures Of Dispersion

Created by

Name	Role	University
Dr. Raju Prasad Paswan	Content Creator	Assam Agricultural University,Jorhat
Dr.S.S.Sidhu	Course Reviewer	Punjab Agricultural University, Ludhiana

Objectives:

- Need for studying measures of dispersion.
- Requisites for an ideal measure of dispersion.
- Overview on different absolute and relative measures of dispersion.

NEED FOR MEASURES OF DISPERSION

- Measures of central tendency helps to locate the centre of the distribution but they do not reveal how the items are spread out on either side of the central value.
- The dispersion of values is indicated by the extent to which these values tend to spread over an interval rather than cluster closely around an average.
- Scatteredness of data about an average is termed as dispersion or variation.

Significance of Measuring Dispersion:

Following are some of the purposes for which measures of variation are needed.

- Test the reliability of an average
- Control the variability
- Compare two or more sets of data with respect to their variability
- Facilitate the use of other statistical techniques

REQUISITES FOR AN IDEAL MEASURE OF DISPERSION

The essential requisites for a good measure of variation are:

- It should be rigidly defined.
- It should be based on all the values in the data set.
- It should be calculated easily, quickly, and accurately.
- It should not be unduly affected by the fluctuations of sampling and also by extreme observations.
- It should be amenable to further mathematical or algebraic manipulations.

CLASSIFICATION OF MEASURES OF DISPERSION

- The various measures of dispersion (variation) can be classified into two categories:

  1. Absolute measures are described by a number or value to represent the amount of variation or differences among values in a data set. Such a number or value is expressed in the same unit of measurement as the set of values in the data such as rupees, inches, feet, kilograms, or tonnes.

  2. **Relative measures** are described as the ratio of a measure of absolute variation to an average and is termed as coefficient of variation. The word 'coefficient' means a number that is independent of any unit of measurement.

Range And Its Coefficient

- The range is the simplest measure of dispersion and is based on the location of the largest and the smallest values in the data. Thus, the range is defined to be the difference between the largest and lowest observed values in a data set.
- Range (R) = Highest value of an observation – Lowest value of an observation
- **R = H – L**
- The relative measure of range, called the coefficient of range is obtained by applying the following formula:

- **Coefficient of range** $\frac{H - L}{H + L}$

Advantages of Range

- Range is rigidly defined, very simple to calculate and easy to understand.
- It is independent of the measure of central tendency.
- It is quite useful in cases where the purpose is only to find out the extent of extreme variation, such as industrial quality control, weather forecasting, fluctuations in share price, and so on.

Disadvantages of Range

- Range is not capable of further mathematical treatment.
- It is largely influenced by two extreme values and completely independent of the other values. For example, range of two data sets {1, 2, 3, 7, 12} and {1, 1, 1, 12, 12} is 11, but the two data sets differ in terms of overall dispersion of values.
- Range is much affected by fluctuation of sampling.
- It cannot be used in case of open-end distributions.

STANDARD DEVIATION, VARIANCE AND COEFFICIENT OF VARIATION

- The arithmetic mean of the squares of the deviations of the values of a variable from its arithmetic mean is called the variance of that variable denoted as σ^2 and the positive square root of variance is called the standard deviation (S.D.) and is denoted as σ.
- If a variable x takes values $x_1, x_2, \ldots, x_n$ and if be the arithmetic mean of these values then

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- and

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Advantages of Standard Deviation

- It is rigidly defined and the value of standard deviation is based on all observations in a set of data.
- It is the only measure of variation capable of algebraic treatment.
- It is less affected by fluctuations of sampling as compared to other measures of variation.

Disadvantages of Standard Deviation

- It is difficult to calculate as compared to other measures of dispersion.
- In comparison to mean deviation it is more affected by extreme values.
- It cannot be computed in case of open-end distribution.

Mathematical Properties of S.D:

- S.D. is independent of the change of origin.
- S.D. is dependent on the change of scale.
- S.D. of combined distribution: let a distribution having n_1 observations has mean and S.D. σ_1, and let another distribution having n_2 observations has mean and S.D. σ_2. Then the S.D. σ of the distribution obtained by combining the two distributions having altogether (n_1+n_2) observations is given by:

$$\sigma = \sqrt{\frac{n_1\sigma_1^{\,2} + n_2\sigma_2^{\,2} + n_1 d_1^{\,2} + n_2 d_2^{\,2}}{n_1 + n_2}}$$

$$d_1 = \bar{x}_1 - \bar{x} \quad d_2 = \bar{x}_2 - \bar{x} \qquad \bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

(*i.e.*, is the mean of the combined distribution)

Coefficient of Variation (C.V.)

- Standard deviation is an absolute measure of variation and expresses variation in the same unit of measurement as the arithmetic mean or the original data. A relative measure called the coefficient of variation (C.V.), developed by Karl Pearson is very useful measure for

1. comparing two or more data sets expressed in different units of measurement.
2. comparing data sets that are in same unit of measurement but the mean values of data sets in a comparable field are widely dissimilar.

- C.V. measures the standard deviation relative to the mean in percentages. In other words, CV indicates how large the standard deviation is in relation to the mean and is computed as follows:

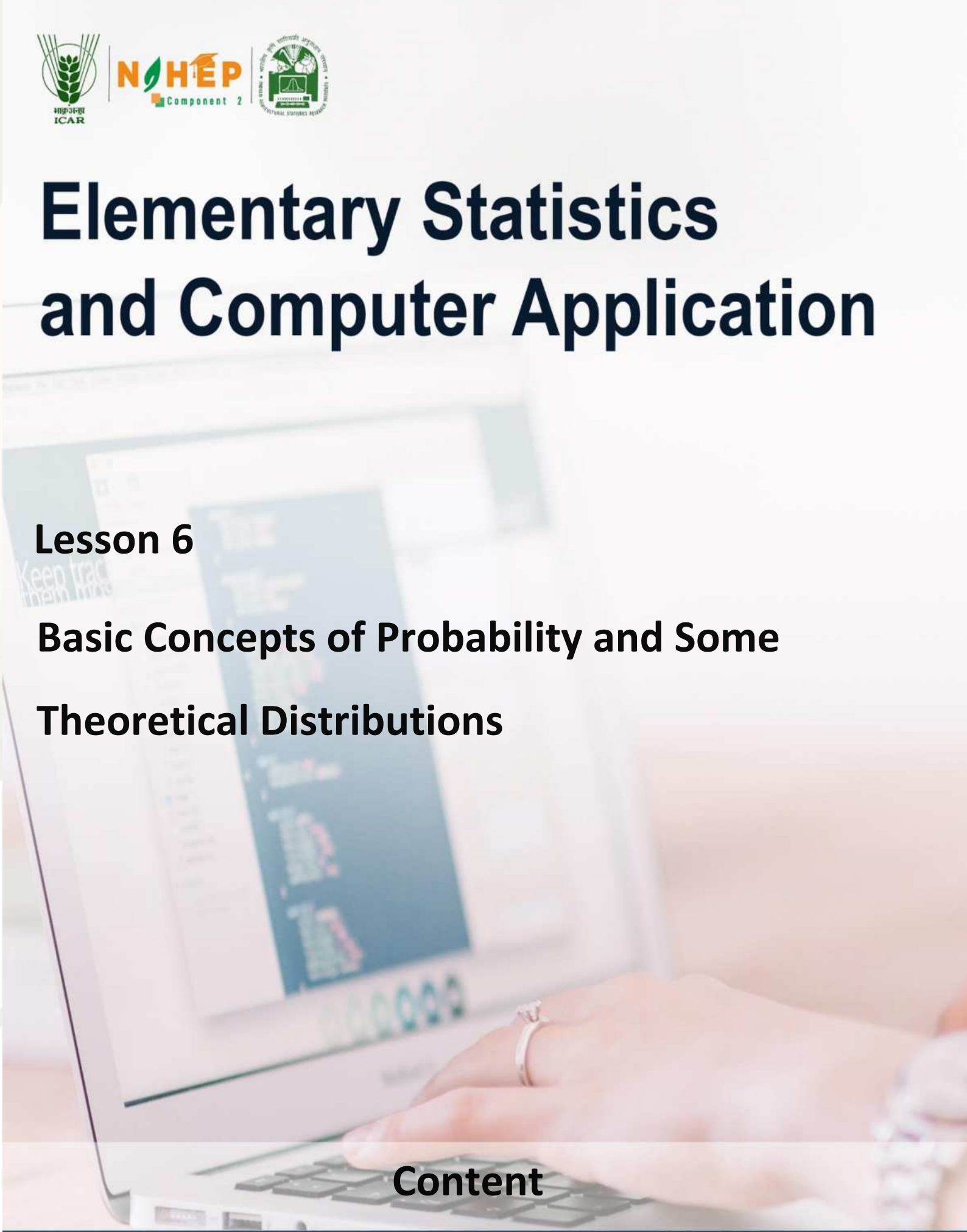
$$C.V. = \frac{\sigma}{\bar{x}} \times 100$$

THANK YOU

Elementary Statistics and Computer Application

Lesson 6

Basic Concepts of Probability and Some Theoretical Distributions



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 6	Basic Concepts Of Probability And Some Theoretical Distributions
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. Basic concept of probability.
2. Additive and multiplicative laws of probability.
3. Discrete distributions-Binomial and Poisson distribution.
4. Continuous distributions-Normal distribution.

6.1 INTRODUCTION

- Probability may be defined as the science that deals with uncertainties and helps us to take decisions about actions even in the midst of uncertainties.
- Probability theory had its origin in the games of chance.The theory of probability was historically used to simplify the complexities of betting and games. In course of time, the probability theory was incorporated in business process and decision making apparatus by business firms, government, professional and non-professional organizations.

6.2 BASIC TERMINOLOGY

- **Random Experiment:** It is an experiment, which if conducted repeatedly under homogeneous conditions does not give the same result. One may know the set of all possible outcomes that the experiment will result but the exact result cannot be predicted with certainty. For example, when a coin is tossed we get either head or tail, i.e., there are two possibilities out of which any one of the

outcomes may occur (i.e., either Head orTail).

- **Trial and Event:** Performing of a random experiment is called as a trial and the result of the random experiment thus performed is called as the event. For example, tossing of a coin is a trial and getting either head or tail is an event.

  Event is called Simple or Elementary, if it corresponds to a single possible outcome of the trial, otherwise it is compound or composite event. An event whether simple and composite is generally denoted by capital Latin letters like, A, B, E etc. Thus, in throwing a single die, the event of getting 5 is a simple event but the event of getting an odd number is a composite event as it consists of three elementary points 1, 3 and5.

- **Exhaustive Events or Cases**: The collection of all possible outcomes of a random experiment is called exhaustive cases. In the toss of a single coin we get either a head or a tail. Hence, there are two exhaustive cases. If two coins are tossed exhaustive cases will be 2^2 = 4. If two dice are thrown then the exhaustive cases will be 6^2 =36.

- **Favourable Cases**: The number of outcomes of a random experiment which result in the happening of a desired event are called as favourable cases. For example: In throwing a die the cases favourable to event of getting an even number is 3, and they are 2, 4 and 6. For example, in drawing a card from a pack of cards the cases favourable to getting a spade is 13, and to getting an ace is 4 and a king of

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

diamond is 1.

- **Mutually Exclusive Events or Cases**: Events are said to be mutually exclusive if the occurrence of one prevents the occurrence of all others. It means that if one of them occurs, it is certain that other events will not occur. Mutually exclusive events are also called as disjoint events. For example, in tossing of a coin head and tail are mutually exclusive events i.e. if head occurs tail will not occur and vice versa.

- **Equally Likely Cases**: The outcomes are said to be equally likely if there is no reason to prefer one rather than the other. For example, in tossing of a coin, all outcomes (H, T) and in rolling a dice all outcomes (1, 2, 3, 4, 5, 6) are equally likely if the coin or the dice are unbiased.

- **Independent Events**: Two events are said to be independent of each other if the occurrence or non– occurrence of one of them does not affect the occurrence or non–occurrence of the other. For example, in the toss of a coin repeatedly getting head in the first throw is independent of getting head in the second or in any of the subsequent throws. Again, if a coin is tossed and a die is thrown together, then the result of the toss is independent of the face value of the die.

- **Sample Point and Sample Space:** A sample point is an elementary event. The set of all possible outcomes i.e. the collection of all

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

possible sample points of a random experiment is called a sample space. For example in throwing of a die there are six sample points. All these sample points taken together forms the sample space. The sample space is generally denoted by 'S' or by '□' and a sample point is represented by 's'. In this case we have, S = {1,2,3,4,5}.

- **Complementary Event:** Two events A and B, are said to be complementary to each other, if the event B represents the non-occurrence of the event A. The event complementary to the event A is generally denoted by A^c, A□ or A. The complementary event A□, of the event A consists of all the sample points of the sample space that are not included in the event A. Thus we have,

$$P(A) + P(A□) = 1$$

For example, if A is the event of getting a multiple of 3 when a die is rolled. Then, we have A = { 3, 6}and A□ = {1, 2, 4, 5 }.

- **Union of Two Events:** The union of two events, generally denoted by A□B represents those sample points which belongs either to A or to B or to both. The union may be extended to any number of events.

- **Intersection of Two Events:** The intersection between the two events, generally denoted by A□B represents those sample points which belongs to both A and B. The intersection between n events consists of those sample points that is common to all the n events.

- **Conditional Probability:** Sometimes it so happens that the

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

probability of an event A (say) depends on the occurrence or non-occurrence of another event B. For example, the probability of a person carrying an umbrella is high if the day is cloudy or rainy and less otherwise. Thus, if A is the event that the person carries an umbrella and B be the event that the day is cloudy, then the event A is dependent on B. The symbol A|B is the event that the man carries an umbrella when it is known that the day is cloudy. However, if A does not depend on B then the event A|B is equivalent to the ordinary event A.

6.3 CLASSICAL DEFINTION OF PROBABILITY AND ITS PROPERTIES

- If there are n, mutually exclusive, equally likely and exhaustive cases out of which m of them are favourable to an event A, then

$$P(A) = \frac{m}{n} = \frac{Favourable\ number\ of\ cases}{Exhaustive\ number\ of\ cases}$$

- Some properties of probability through the classical definition are listed below:
  - Probability of an event lies between 0 and 1.That is if A is an event, then

    $0 \leq P(A) \leq 1$.
  - Probability of an impossible event is 0.
  - Probability of a certain event is equal to 1.
  - If the occurrence of an event A implies the occurrence of another event B then $P(A) \leq P(B)$.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Probability of happening of a particular event and the probability of non-happening of that event when added results to unity i.e., $P(A) + P(\bar{A}) = 1$

- Some limitations of the classical approach to probability are:
  - It is based on the feasibility of subdividing the possible outcomes of the experiment into 'mutually exclusive', 'equally likely' and 'exhaustive cases'.
  - It fails when n, total number of outcomes, of the random experiment is infinite or actual value of n is not known.
  - The classical approach has limited applications, as in most cases it may not be possible to enumerate all the outcomes of a random experiment.
  - 'Equally likely' is synonymous to equally probable. Thus, the theory fails to compute the probability if the outcomes of a random experiment are not equally likely. In some cases it may be difficult to know whether outcomes of a random experiment are equally likely or not.

Example: What is the probability that a leap year selected at random will contain 53 Mondays?

Solution: We know that a leap year consists of 366 days, i.e. 52 complete weeks and two extra days. These two extra days may be any one of the following cases:

I. Sunday and Monday
II. Monday and Tuesday

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

III. Tuesday and Wednesday

IV. Wednesday and Thursday

V. Thursday and Friday

VI. Friday and Saturday

VII. Saturday and Sunday

Out of the above seven exhaustive cases (I) and (II) consist of Monday, which are the favourable cases.

Therefore, the required probability = 2/7.

Example: Two dice are thrown simultaneously. Find the probability of getting even numbers on both the dice.

Solution: The sample space is

S ={ 1,2,3,4,5,6 } x {1,2,3,4,5,6 }

={ (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) }

Therefore, n(S)= 36

Let, A be the event of getting even numbers on both the dice.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Therefore, n(A)=9

Therefore, required probability, P(A)= 9/36 = ¼.

6.4 BASIC LAWS OF PROBABILITY

- **Additive Law of Probability or The Theorem of Total Probability:** The probability that either of the two mutually exclusive events A and B will occur is given by the sum of their individual probabilities. In symbols we may write,

$$P(A + B) = P(A) + P(B)$$

- The theorem can be generalized for n mutually exclusive events. If A1, A2,…,An be n mutually exclusive events then

$$P(A_1+ A_2+…+A_n) = P(A_1)+P(A_2)+…+P(A_n)$$

- **General Additive Theorem:** The occurrence of at least one of the two events A and B, when A and B are independent events are given by

$$P(A+B) = P(A) + P(B) - P(AB)$$

- **Multiplicative Law of Probability or Theorem of Compound Probability:** The probability of the simultaneous occurrence of two events A and B is equal to the probability of A multiplied by the conditional probability of B given that A has already occurred. (Or it is equal to the probability of B multiplied by the conditional probability of A given that B has already occurred.) In symbols,

$$P(AB) = P(A)\, P(B|A)$$
$$= P(B)\, P(A|B)$$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Corollary:** If A and B are independent events P(B|A) is same as P(B). then for two independent events the theorem of compound probability becomes

$$P(AB) = P(A) \times P(B)$$

Example: Out of 100 students in a boys hostel 80 take tea, 40 take coffee and 25 take both. Find the probability that a student takes (a) either tea or coffee (b) neither tea nor coffee.

Solution: Let A denotes the event that the student takes tea and B denotes the event that the student takes coffee. Now, from the question we have,

$P(A)= 80/100$; $P(B)=40/100$; $P(A \cap B)= 25/100$

(a) The probability that the student selected at random takes either tea or coffee is given by:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$= 80/100 + 40/100 - 25/100$$
$$= 19/20$$

(b) The probability that the student selected at random takes neither tea nor coffee is given by:

$$1-P(A \cup B) = 1/20.$$

Example:A bag contains 4 red and 6 green balls. Two draws of one ball each are made without replacement. What is the probability that one is red and the other is green?

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Solution: Let A be the event that a red ball is drawn and let B be the event that a green ball is drawn.

So, P(A) = 4/10 and P(B)= 6/10

Probability of drawing a green ball in the second draw given that the first draw has given a red ball P(B|A) =6/9 (since only 9 balls are left out of which 6 are green).

Now, probability of the combined event A and B is given by,

P(AB) = P(A) x P(B|A) = 4/10 x 6/9 =24/90.

But it could also happen the first ball may be green and second ball is red.

Probability of drawing a green first, P(B) =6/10 and red next (given green has been drawn) is P(A|B) =4/9.

Therefore, P(AB) = P(B) x P(A|B) = 6/10 x 4/9 =24/90.

Now, when any one of the two situations (red and green or green and red) can happen and both of them are mutually exclusive.

Thus, the required probability, will be = 24/90 + 24/90 =48?90 =8/15.

Example: From a pack of 52 cards two cards are drawn at random. Find the probability that one is a king and the other a queen.

Solution: In a pack there are 52 cards. Two cards can be drawn in $^{52}C_2$ ways, which is the total number of cases.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Again, there are four kings and four queens of four different suits. A king can be drawn out of 4 kings in ${}^4C_1$ ways. Similarly, a queen can be drawn out of 4 queens in ${}^4C_1$ ways. Simultaneous occurrence of both the events will follow the multiplicative law giving the number of favourable cases as ${}^4C_1$ X ${}^4C_1$ ways.

Therefore, required probability $= \frac{{}^4C_1 . \ {}^4C_1}{{}^{52}C_2} = \frac{8}{663}$

6.5 PROBABILITY DISTRIBUTION

- The distribution obtained by taking the possible values of a random variable together with their respective probabilities is called a probability distribution.
- A probability distribution can be presented either with the help of a function or in tabular form where values of the random variable and corresponding probability are shown.
- The probability distribution for a discrete random variable is called as a discrete probability distribution or 'probability mass function' (pmf) and that of a continuous random variable is called a 'probability density function' or (pdf).
- Binomial distribution, Poisson distribution are some examples of discrete probability distribution whereas normal distribution is an example of continuous distribution.

6.6 BINOMIAL DISTRIBUTION

- Binomial distribution was discovered by James Bernoulli

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

(1654□1705) in the year 1700. But it was published in the year1713.

- Let a random experiment having only two outcomes either success or failure, be performed a number of times (n, say) under identical conditions. Let X be the random variable that represents the number of successes in n trials, with 'p' the probability of success which remains constant for each time the random experiment is performed. Thus, 'q =1-p' is the probability of failure in any trial. Under the above conditions the random variable X is said to follow binomial distribution if its probability mass function is given by

$$P(X = x)$$
$$= {}^{n}C_{x}p^{x}q^{n-x}, \quad x = 0,1,2,\dots,n \text{ and } 0 \leq p \leq 1, q = 1 - p$$

Where *n* and *p* or *q* are the parameters of the binomial distribution.

The binomial distribution is a discrete distribution as the random variable can take only the integral values i.e. 0,1,2,…,n. so, the probability function of the binomial distribution is also called as probability mass function (*p.m.f*). To notation used to denote that a random variable X follows binomial distribution with parameters n and p is *X ~ B(n, p).*

- **Condition for Binomial distribution:** We get the binomial distribution under the following experimentation conditions:
  - The number of trial n is finite.
  - The trials are independent of each other.
  - The probability of success p is constant for each trial.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Each trial must result in a success or failure.
- The events are discrete events.

- **Properties of Binomial distribution:**
  - The binomial distribution is a discrete distribution where the random variable X takes the values 0,1,2,…,n.
  - The binomial distribution has two parameters n and p or q.
  - The mean of the binomial distribution is np and variance is npq . So standard deviation is $\sqrt{npq}$
  - The mean is greater than the variance.
  - The binomial distribution may have either one or two modes.
  - The binomial distribution is a symmetrical distribution if p = q = 0.5. Otherwise it is a skewed distribution.
  - If X follows binomial distribution with parameters n_1 and p and Y follows binomial distribution with parameters n_2 and p , then X+Y follows binomial distribution with parameters n_1+n_2 and p. This property is known as additive property of binomial distribution.
- **Applications Of Binomial Distribution**: Binomial Distribution is used to determine the probability of -
  - Occurrence of heads or tails when a number of coins are tossed.
  - Number of males or females in a given population
  - Number of defective or non-defective items in a production process

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Number of successes or failures of a gambler in a particular game which is repeated a number of times.

Example: Five fair coins are tossed. Find the probability of (i) Exactly three heads. (ii) At least three heads.

Solution: In a fair coin we have probability of a head to occur =1/2.

Let X be a random variable that denotes the number of heads that occurs when five coins are tossed.

Here, n= no. of coins tossed =5.

So, using the binomial distribution we have,

(i) P[Exactly three heads] = P[X=3] = $^5C_3\ (1/2)^3\ (1-1/2)^{5-3}$ =5/16.

(ii) P[Atleast three heads] = P[X=3 or X=4 or X=5]

=P[X=3] + P[X=4] + P[X=5]

=$^5C_3\ (1/2)^3\ (1-1/2)^{5-3}$ + $^5C_4\ (1/2)^4\ (1-1/2)^{5-4}$ +

$^5C_5\ (1/2)^5\ (1-1/2)^{5-5}$

=1/2.

Example: The mean of a binomial distribution is 6 and the standard deviation is given by $\sqrt{3/2}$. Find the distribution.

Solution:We know that for a binomial distribution with parameters n and p. We have, mean= np and Variance= npq.

Also, variance = (Standard deviation)^2=3/2.

Thus, np=6 and npq =3/2.

Now, q=npq/np =1/4.

p=1-q=3/4

Also, np=6 =>n=8 (since p=3/4)

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

So, the required distribution is

$P(X=x)={}^{8}C_{x}\,(3/4)^{x}\,(1/4)^{8-x}$, where x = 0,1,2,…,8.

6.7 POISSON DISTRIBUTION

- The Poisson distribution, named after Simeon Denis Poisson (1781-1840). Poisson distribution is a discrete distribution. It describes random events that occurs rarely over a unit of time or space.
- It differs from the binomial distribution in the sense that we count the number of success and number of failures, while in Poisson distribution, the average number of success in given unit of time or space.
- A random variable X is said to follow Poisson distribution if it assumes only non-negative values and if its probability mass function is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0,1,2, \dots \text{and } \lambda > 0$$

where, λ is the parameter of the Poisson distribution.

The Poisson distribution is a discrete distribution as the random variable can take only the integral values i.e. 0,1,2,… so, the probability function of the Poisson distribution is also called as probability mass function (*p.m.f*). To notation used to denote that a random variable X follows Poisson distribution with parameter λ is $X \sim P(\lambda)$.

- **Condition for Poisson distribution:** We get the Poisson distribution under the following experimentation conditions:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The number of trials n should be indefinitely large i.e., $n->\infty$.
- The probability of success p for each trial is indefinitely small.
- $np = \lambda$, should be finite where λ is constant.

- **Properties of Poisson distribution:**
  - The Poisson distribution is a discrete distribution where the randomvariable X takes the values 0,1,2,… ∞
  - The Poisson distribution has one parameter i.e., λ
  - The mean and variance of the Poisson distribution are equal that is, mean = variance = λ.
  - The standard deviation is equal to$\sqrt{\lambda}$.
  - The Poisson distribution may have either one or two modes.
  - The Poisson distribution may be obtained as a limiting case of Binomial distribution.
  - If X and Y are two independent Poisson variants with parameters λ_1 and $\lambda_2$then X+Y is also a Poisson variate with parameter λ_1+λ_2.
- **Applications of Poisson distribution:** Poisson Distribution can be used in the following cases-
  - Number of cars passing through a certain point per minute during a busy hour of the day.
  - Number of suicides reported per week in a particular town.
  - Number of printing mistakes per page of a standard book.
  - Number of persons born blind per year in a particular village.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Deriving Poisson distribution from Binomial distribution:**The Poisson distribution is a limiting form of the binomial distribution. The conditions under which the binomial distribution tends to Poisson distribution are :
    - When the number of trials (n) is very large i.e. when $n \to \infty$
    - When the probability of success in each trial (p) is very small i.e. when p $\to$ 0, such that np (= λ) ,a reasonable finite quantity.

Example: In a factory manufacturing blades it is found that on an average 2% of the blades are defective. What is the probability that at most 5 defective blades will be found in a box of 200 blades?

Solution: Here, we have,

n= total number of blades =200

p = probability of a defective blade =2/100 = 0.02.

Thus, $\lambda = n\,.\,p$ = 200 x 0.02 = 4.

Let, X be the random variable which represents the number of defective blades in a packet of 200 blades.

Thus, by the question we are to find the probability of X ≤5.

So, P(X≤5) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)

$$= \sum_{x=0}^{5} P(X = x) = \sum_{x=0}^{5} \frac{e^{-4}4^{x}}{x!} = 0.785$$

Thus, required probability is 0.785.

Example: In a Poisson distribution we have the probability that X takes the value 0 is 0.1. Find the mean of the distribution.

Solution: We know that the mean of the Poisson distribution is equal to its parameter λ where, $P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Now, putting x=0, we have,

$$P(X=0)=\frac{e^{-\lambda}\lambda^{x}0}{0!}$$

$$=>0.1=e^{-\lambda}$$

$$=>e^{\lambda}=10$$

$$=>\lambda=\ log_e 10=2.3026$$

Thus, the mean of the Poisson distribution is 2.3026.

6.8 NORMAL DISTRIBUTION

- The Normal distribution also called as the Gaussian distribution, is a continuous probability distribution with two parameters μ and σ and is defined by the probability density function (p.d.f.)

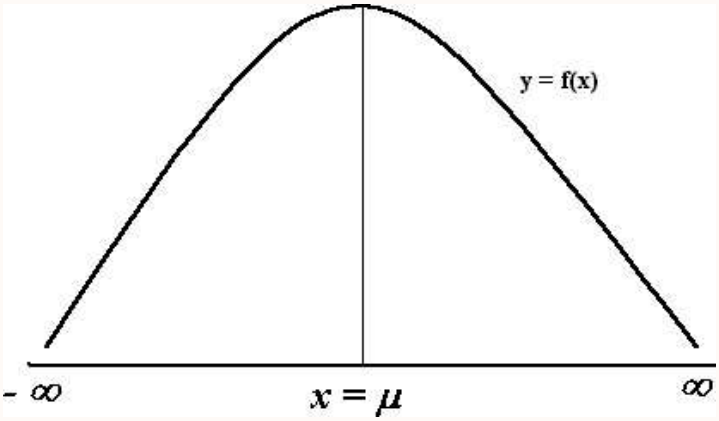
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\ ;\ -\infty<x<\infty;\ -\infty<\mu<\infty\ ;\ \sigma>0$$

  Here μ is the mean and σ is the standard deviation of the distribution. They are two parameters of the distribution as well. π and e are two mathematical constants having the approximate values 22/7 and 2.718 respectively.

- The history of the distribution is very interesting. It was discovered by an English Mathematician De-Moivre in 1733, used by Laplace, later in the year 1774 but the credit of the distribution was wrongly attributed to Gauss, who first made the reference of the distribution in 1809 to study the errors in the measurement of Astronomy.
- This is the most useful distribution in theoretical statistics because of its many important characteristics. Most of the probability distributions of statistics whether discrete or continuous tends to

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

normal distribution especially when the number of observations are large. The probability curve of normal distribution is known as



Normal Curve. The curve is symmetrical about its mean (μ), bell-shaped and the two tails extend to infinitely on either side.

- **Deriving Poisson distribution from Binomial and Poisson distribution:**

  Normal distribution is a limiting form of the binomial distribution under the following conditions:

  - n, the number of trials is indefinitely large i.e., $n \to \infty$ and
  - Neither p nor q is very small.

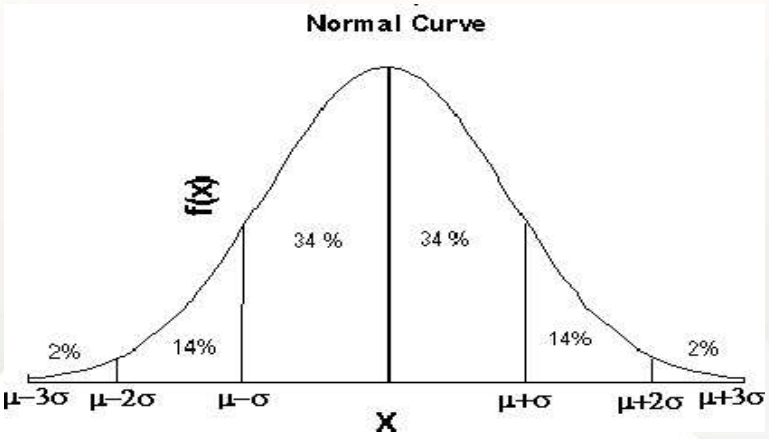
  Normal distribution can also be obtained as a limiting form of Poisson distribution with parameter $\lambda \to \infty$

- **Properties of Normal Distribution:** The normal distribution has the following properties:

  - The Normal Distribution has two parameters,μ (the mean of the distribution) and σ(the standard deviation of the distribution).
  - Normal distribution is a symmetrical distribution with mean = median = mode.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- It has only one mode at x = m (i.e., unimodal).
- The maximum ordinate occurs at x = μ and its value is $=\frac{1}{\sigma\sqrt{2\pi}}$.
- The quartiles are equidistant from the median, i.e. Q3 - Q2 = Q2 - Q1.
- The graph of the normal distribution is called the normal curve. It is a bell shaped curve and is symmetrical about the mean.
- The linear combination of independent normal variates is also a normal variate.
- The area under the normal curve can be distributed in the following manner:



- **Applications of Normal Distribution:** Normal distribution plays a very important role in statistical theory and its application becomes useful for the following reasons:
  - Most of the distribution occurring in practice, for example binomial, Poisson, hyper-geometric distribution are approximated by Normal Distribution.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Even if a variable is not normally distributed it can sometimes be brought to normal form by simple transformation of variable.
- Many of the distributions of sample statistics tend to normality and as such they can be best studied with the help of normal curve.
- The proof of all the test of significance in sampling are based upon the fundamental assumption that the population from which the sample has been drawn is normal.
- The theory of normal curve can be applied to the graduation of the curves, which are not normal.
- Normal distribution is extensively used in statistical quality control.

- **Standard Normal distribution:** Let X be random variable which follows normal distribution with mean μ and variance σ^2. The standard normal variate is defined as $Z = \frac{X-\mu}{\sigma}$ which follows.
- standardnormaldistributionwithmean0andstandarddeviation1i.e.,Z ~N(0,1).The standard normal distribution is given by:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}z^2};\ -\infty < z < \infty.$$

The advantage of the above function is that it does not contain any parameter. This enables us to compute the area under the normal probability curve.

****

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

GLOSSARY

- **Probability:**Probability may be defined as the science that deals with uncertainties and helps us to take decisions about actions even in the midst of uncertainties.
- **Favourable Cases:** The number of outcomes of a random experiment which result in the happening of a desired event are called as favourable cases.
- **Probability of an event:** If there are n, mutually exclusive, equally likely and exhaustive cases out of which m of them are favourable to an event A, then

$$P(A) = \frac{m}{n} = \frac{Favourable\ number\ of\ cases}{Exhaustive\ number\ of\ cases}$$

- **Probability distribution:** The distribution obtained by taking the possible values of a random variable together with their respective probabilities is called a probability distribution.
- **Probability mass function:**The probability distribution for a discrete random variable is called as a discrete probability distribution or 'probability mass function'.
- **Probability density function:**The probability distribution for a continuous random variable is called as a continuous probability distribution or 'probability density function'.

Distribution	Type	Functional Form	Rangeof Variable	Range of Parameter	Mean	Var.
Binomia	Discrete	$^nC_x p^x q^{n-x}$	*x = 0,*	*0 ≤p ≤1*	*np*	*npq*

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

I			*1,2,…,n*			
Poisson	Discrete	$\frac{e^{-\lambda}\lambda^{x}}{x!}$	*x= 0, 1, 2,…*	$\square > 0$	$\square$	$\square$
Normal	Continuous	$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\square\square < x < \square$	$\square\square < \square < \square$ $\square > 0$	$\square$	$\square^2$

REFERENCES

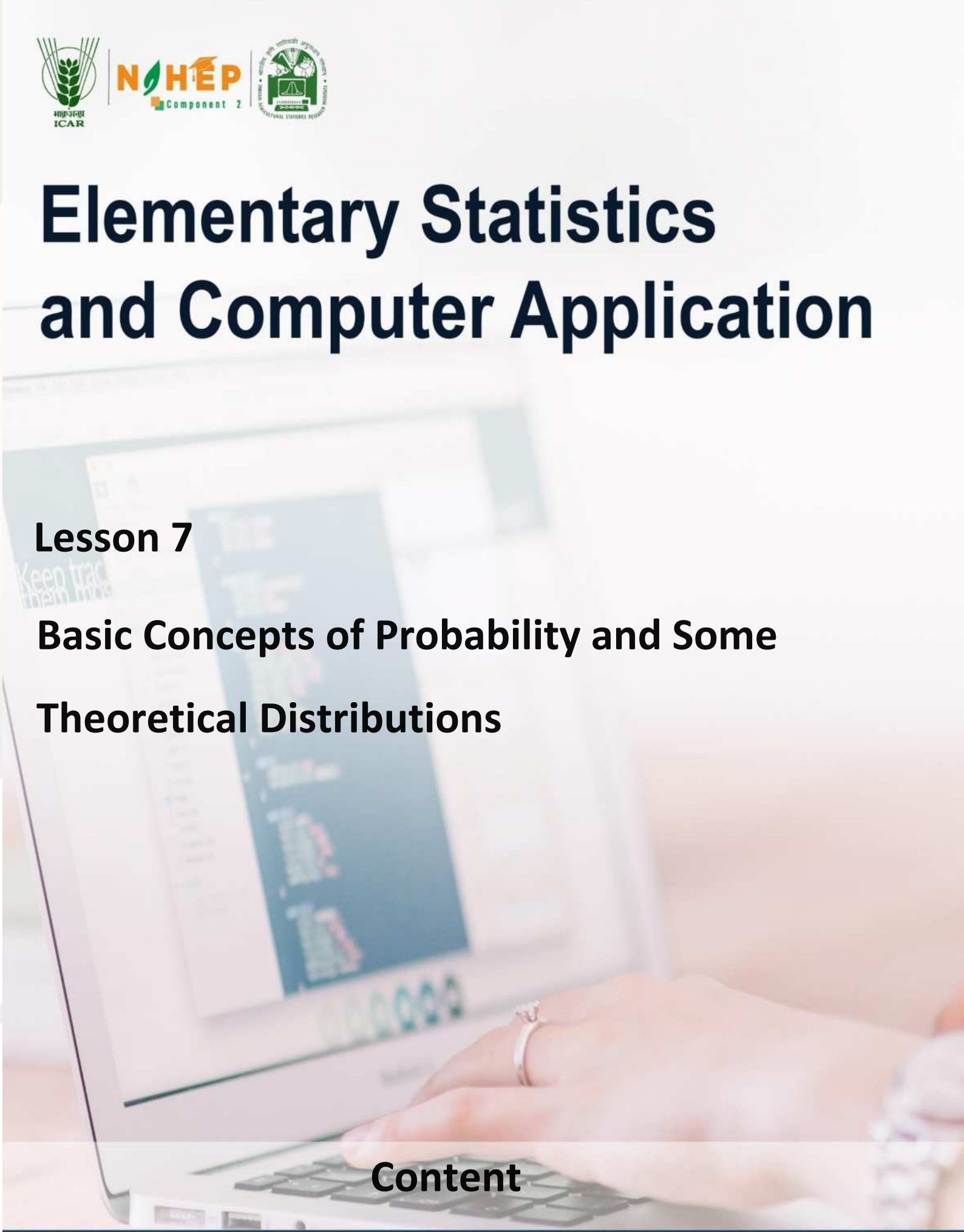
- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.
- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson 7

Basic Concepts of Probability and Some Theoretical Distributions



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 7	Sample Survey
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. Basic concepts of sample survey.
2. Sampling vs. Complete enumeration
3. Notion of random sampling vs. non-random sampling.
4. Basic idea of simple random sampling and stratified random sampling.

7.1 BASIC CONCEPTS

- **Population:** The dictionary meaning of the word population is the number of people residing in a particular geographical area. But in statistics, population may be defined as the aggregate of the entire group of items, units or subject which is under the reference of study. Population may consist of finite or infinite number of units which may be animate or inanimate. Population is also termed as universe. The total number of individuals in the population is called the population size. The in habitants of a region, no of wheat fields in a state, fruit plants in an orchard, and insects in a forest are examples of finite population. All real numbers, all stars in the night sky are examples of infinite population.
- **Sample:** A sample is that part of the population that is expected to exhibit the characteristic of the population. In sampling theory, instead of studying every unit of the population, only a part of the universe is studied and the conclusions are drawn on that basis of that sample for the entire universe.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Parameter:** A summary measure that describes any given characteristic of the population is known as parameter. Population are described in terms of certain measures like mean, standard deviation etc. These measures of the population are called parameter and are usually denoted by Greek letters. For example, population mean is denoted by μ, standard deviation by **σ and** variance by **σ^2.**

- **Statistic:** A summary measure that describes the characteristic of the sample is known as statistic. Thus sample mean, sample standard deviation etc. is statistic. For example, sample mean is denoted by $\bar{x}$, sample standard deviation is denoted by s. The statistic is a random variable because it varies from sample to sample.

- **Sampling unit:** An element or a group of elements on which the observations can be taken is called a sampling unit. The objective of the survey helps in determining the definition of sampling unit. For example, if the objective is to determine the total income of all the persons in the household, then the sampling unit is household. If the objective is to determine the income of any particular person in the household, then the sampling unit is the income of the particular person in the household. Before selecting a sample each individual must be divided into some sampling units.

- **Sampling frame:** The sampling frame is an exhaustive list of sampling units. Such a list may be either in tabular form or in the form of maps in case of area surveys. However, sometimes such

sampling frames are difficult to construct and even if it is constructed often it remains incomplete or out dated. The simplest example of sampling frame is the voter list, which provides the complete list of individuals along with their individuals and other details of a particular area.

- **Population size:** The total no. of individuals in the population is called its population size.
- **Sample size:** The total no. of individuals in the sample is called its sample size.

7.2 DIFFERENT TYPES OF POPULATION

- Population can be fundamentally classified into four categories:
    - **Finite population:** In this population the number of members of the population is fixed and limited. The population of babies born today in a particular health centre in the town is an example of finitepopulation.
    - **Infinite population:** Here the population consists of infinite number of items. For example the collection of all stars in a night sky, the real numbers lying in between 0 to 1.
    - **Real population:** A population consists of the items that are all present physically is termed as real population. For example, if we visit a health centre town and want to know the birth weights of all the new-born babies that were born in the last five years then we can get it from the Record Department of the health centre. Here each and every

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

member of the population actually exists.

- **Hypothetical population:** Let us consider another example in which one is interested in noting down that birth weights of new-born in the medical centrein the next five years. This is an example of hypothetical population because here the population does not exist in reality.

7.3 CENSUS AND SAMPLING METHODS

Statistical data can be collected either by census enquiry or by sample enquiry.

- A census enquiry is that in which all the items constituting the population are studied and conclusions are drawn therefrom. Thus, if we want to find out the average weight of all the ninety persons of a locality in a census enquiry, we will measure the actual weights of all the 90 persons and the total will be divided by 90 which will give us the average weight.
- On the other hand, in a sample enquiry, only a part of population is studied and from the results obtained by studying the sample, conclusions are drawn for the whole population. Sampling, i.e., the selection of a part or an aggregate of the material to represent the whole aggregate is a long established practice. Simple examples are provided by a handful of grain taken from a full bag or a piece of cloth cut off from a 'roll'. In the example of measurement of weight of 90 persons taken above, if we proceed by sample enquiry we will take a sample of suitable size (say 20 or 25 persons) and divide the aggregate by the respective number and thus, obtain

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

estimates for the population average.

- Thus, the basic difference between sample and census enquiry is that in a census enquiry all the items of the population are observed but in a sample enquiry, only a part of the population is observed and there from conclusions for the whole population are drawn.

7.4 OBJECTIVE OF SAMPLING THEORY:

Generally, sample is drawn from the population taking into consideration three important factors viz. time, labour and money. In case of study of samples we require lesser time, labour and money than of survey of population. The main objectives of sampling theory are-

- To obtain the characteristic of the population with the available resource at our disposal in terms of money and manpower by studying the sample values only.
- To obtain the best possible estimate of the population parameter.
- Again, there are some situation when sampling is the last resort.E.g.: To test the breaking strength of chalk pencil. (Otherwise it will be destructive); to estimate the no. of fishes in the river etc.

7.4 SAMPLING VS. COMPLETE ENUMERATION

- **Advantages of Sampling over Complete Enumeration**
    - In sample survey only a part of the population is to be

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

investigated and hence it takes less time, less money and lesslabour.

- In census method a large group of investigators are employed— a thorough and intensive training of such a large group of people can hardly be possible. But in sample survey method, only a small group of estimators are employed— they are given specialised training. Hence, though a little bit expensive, the output is much moreaccurate.
- Sample survey even with a limited budget can cover more geographicalarea.
- In sample survey, sampling error cannot be avoided. But in census there is no question of sampling errors. However, large magnitudes of non-sampling errors are involved in the census method—their magnitudes remain undetermined and they greatly influence the accuracy of the results. Sampling errors can be theoretically calculated and hence their magnitudes are known, but non-sampling errors do not follow any law of probability and hence, their magnitudes cannot be precisely determined.
- Sometimes complete enumeration is not feasible or practicable. This is especially the case in which testing is destructive. That is the subject under study gets destroyed or cannot be used any further on testing. For instance, to examine the quality of sweets in a sweet shop one has to restore to sampling methods, which is the only way out.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Similarly the blood test in the laboratory one has to resort to sample studyonly.

- **Limitations of sampling**

On the basis of the advantages of sampling enumerated above one should not come to the hasty conclusion that sampling is always to be preferred over a census enquiry. Sometimes, the sample must be so large in order to achieve the required accuracy that one might just do well to take a census. The following are the chief disadvantages or limitations ofsampling:

- In order to obtain accurate results it is indispensable that a sample survey has been properly planned and executed. Otherwise incorrect and misleading results may beobtained.
- Sampling requires the services of experts. If there is a scarcity of such people, sampling may give unsatisfactory results owing to the use of faulty methods of selection, inappropriate sampling design, or inefficient methods ofestimation.
- Complicated sampling plans require more labour than a complete count in the longrun.
- There are various sources of errors in sample surveys. Every attempt must be made to minimize the chances of errors, otherwise correct conclusions cannot beexpected.

7.5 BASIC PRINCIPLES OF SAMPLE SURVEY

- **Statistical Regularity:** This principle is based on the theory of

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

probability which may be considered as the basis for the selection of a random sample. The random sample is used in order to give each and every unit of the population equal chance of being included in the sample. In such a case the sample will start reflecting the heterogeneity present in the entire population and hence fourth the sample becomes a representative of the population.

- **Validity:** This ask the sample design to be obtained in such a manner that it can be used for the purpose obtaining valid estimates and tests about the parameter. The samples obtained by using a random technique satisfy this principle.

- **Optimization:** The word optimization means extreme case i.e. either maximization or minimization. Thus, the sampling technique should be developed in order to achieve both maximization and minimization. The maximization is to be done by increasing the efficiency for a fixed cost and minimization is attained by reducing the expenditure incurred in the survey for a fixed level of efficiency.

7.6 SAMPLING METHODS

The various methods of sampling can be grouped under: Probability sampling or random sampling and Non-probability sampling or non-random sampling.

- **Random sampling:** Under this method, every unit of the population at any stage has a pre-assigned chance of selection (or) each unit is

drawn with known probability. It helps to estimate the mean, variance etc. of the population parameters and efficiency of the concerning statistic.

Under probability sampling there are two procedures:

- Sampling with replacement (SWR)
- Sampling without replacement (SWOR)

When the successive draws are made with placing back the units selected in the preceding draws, it is known as sampling with replacement. When such replacement is not made it is known as sampling without replacement.

When the population is considerably small, sampling with replacement is adopted otherwise SWOR is preferred.

There are many kinds of random sampling. Some of the most commonly applied random sampling techniques are:

- Simple Random Sampling
- Systematic Random Sampling
- Stratified Random Sampling
- Cluster Sampling

- **Non-random sampling or purposive sampling:** Under this method, the selection of units in the sample from population is not governed by the probability laws. For example, the units are selected on the basis of personal judgment of the surveyor. The persons volunteering to take some medical test or to drink a new type of coffee also constitute the sample on non-random laws.

Another type of sampling is Quota Sampling. The survey in this case is continued until a predetermined number of units with the characteristic under study are picked up. For example, in order to conduct an experiment for rare type of disease, the survey is continued till the required number of patients with the disease are collected.

7.7 SIMPLE RANDOM SAMPLING (SRS)

- The basic probability sampling method is the simple random sampling. It is the simplest of all the probability sampling methods. It is used when the population is homogeneous.
- When the units of the sample are drawn independently with equal probabilities. The sampling method is known as Simple Random Sampling (SRS). Thus if the population consists of N units, the probability of selecting any unit is 1/N.
- A theoretical definition of SRS is as follows:

  Suppose we draw a sample of size n from a population of size N. There are NC possible samples of size n. If all possible samples have an equal probability 1/NC of being drawn, the sampling is said be simple random sampling.
- There are two methods in SRS: **Lottery method** and **Random no. table method**
    - **Lottery method:** This is most popular method and simplest method. In this method all the items of the universe are numbered on separate slips of paper of same size, shape and color. They are folded and mixed up in a drum or a box or a

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

container. A blindfold selection is made. Required number of slips is selected for the desired sample size. The selection of items thus depends on chance.

For example, if we want to select 5 plants out of 50 plants in a plot, we number the 50 plants first. We write the numbers from 1-50 on slips of the same size, role them and mix them. Then we make a blindfold selection of 5 plants. This method is also called unrestricted random sampling because units are selected from the population without any restriction. This method is mostly used in lottery draws. If the population is infinite, this method is inapplicable. There is a lot of possibility of personal prejudice if the size and shape of the slips are not identical.

- **Random number table method:** As the lottery method cannot be used when the population is infinite, the alternative method is using of table of random numbers.

  There are several standard tables of random numbers. But the credit for this technique goes to Prof. LHC. Tippet (1927). The random number table consists of 10,400 four-figured numbers. There are various other random numbers. They are fishers and Yates (1980 comprising of 15,000 digits arranged in twos. Kendall and B.B Smith (1939) consisting of 1,00,000 numbers grouped in 25,000 sets of 4 digit randomnumbers, Rand corporation (1955) consisting of 2,00,000 random numbers of 5 digits each etc.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Merits**
    - There is less chance for personal bias.
    - Sampling error can be measured.
    - This method is economical as it saves time, money and labour.
- **Demerits**
    - It cannot be applied if the population is heterogeneous.
    - This requires a complete list of the population but such up-to-date lists are not available in many enquires.
    - If the size of the sample is small, then it will not be a representative of the population.

7.8 STRATIFIED SAMPLING

- When the population is heterogeneous with respect to the characteristic in which we are interested, we adopt stratified sampling.
- When the heterogeneous population is divided into homogenous sub-population, the sub-populations are called strata. From each stratum a separate sample is selected using simple random sampling. This sampling method is known as stratified sampling.
- We may stratify by size of farm, type of crop, soil type, etc.The number of units to be selected may be uniform in all strata (or) may vary from stratum to stratum.

  There are four types of allocation of strata
    - **Equal allocation**: If the number of units to be selected is

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

uniform in all strata it is known as equal allocation of samples.

- **Proportional allocation:** If the number of units to be selected from a stratum is proportional to the size of the stratum, it is known as proportional allocation of samples.
- **Neyman's allocation:** When the cost per unit varies from stratum to stratum, it is known as optimum allocation.
- **Optimum allocation:**When the costs for different strata are equal, it is known as Neyman's allocation.

- **Merits**
  - It is more representative.
  - It ensures greater accuracy.
  - It is easy to administrate as the universe is sub-divided.
- **Demerits**
  - To divide the population into homogeneous strata, it requires more money, time and statistical experience which is a difficult one.
  - If proper stratification is not done, the sample will have an effect of bias.

****

GLOSSARY:

- **Population:** Population is the aggregate of the entire group of

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

items, units or subject which is under the reference of study.

- **Sample:** A sample is that part of the population that is expected to exhibit the characteristic of the population.
- **Parameter:**A summary measure that describes any given characteristic of the population is known as parameter.
- **Statistic:** A summary measure that describes the characteristic of the sample is known as statistic.
- **Sampling unit:** An element or a group of elements on which the observations can be taken is called a sampling unit.
- **Sampling frame:** The sampling frame is an exhaustive list of sampling units.
- **Census:** A census enquiry is that in which all the items constituting the population are studied and conclusions are drawn therefrom.
- **Random sampling:**Sampling method in which every unit of the population at any stage has a pre-assigned chance of selection.
- **Non-random sampling or purposive sampling:** Sampling method in which the selection of units in the sample from population is not governed by the probability laws, but intuition of the investigator.

REFERENCES

- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.

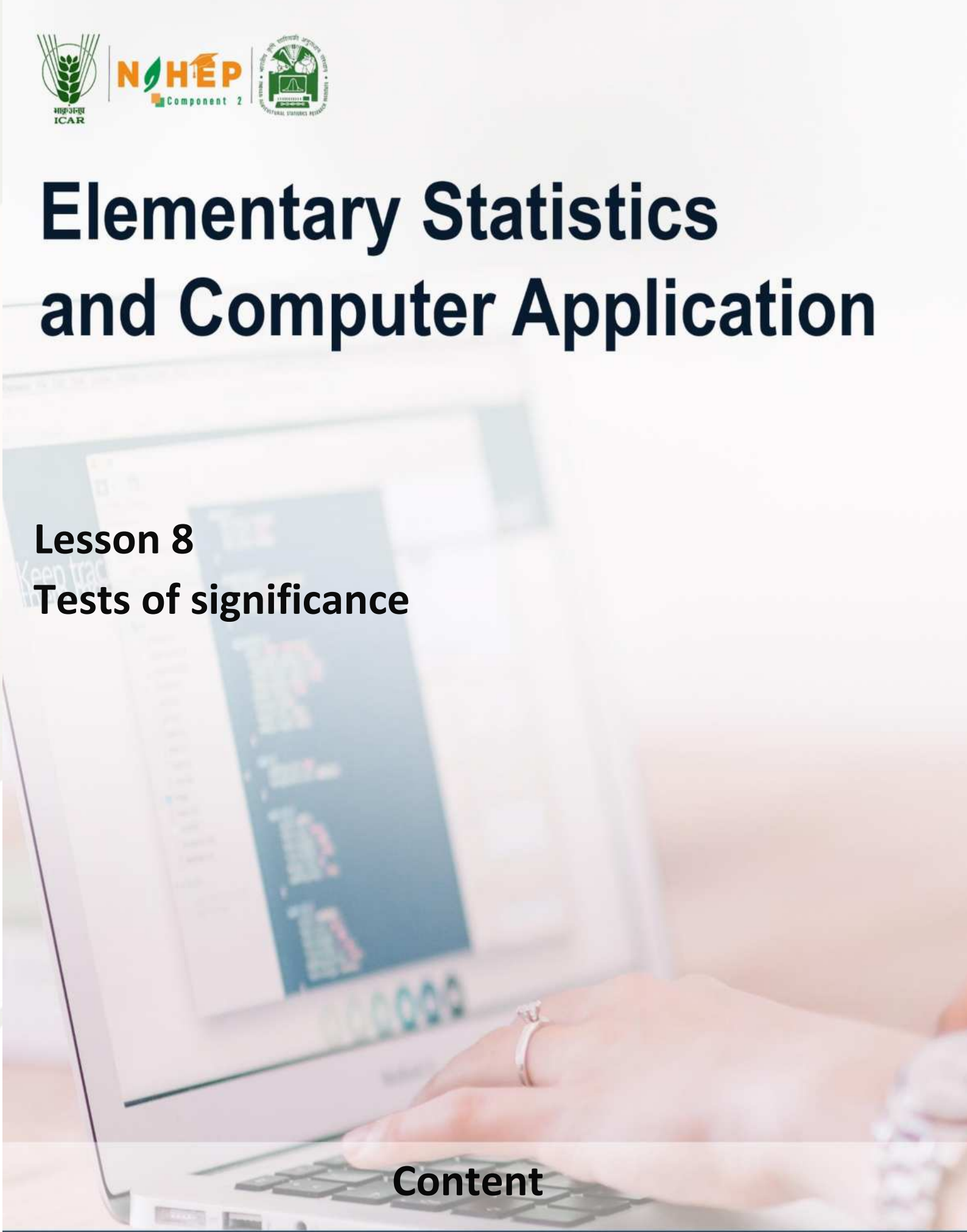
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson 8
Tests of significance



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 8	Tests of significance
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. To introduce the basic concept of tests of significance
2. To introduce the procedures of statistical hypothesis

Tests of significance

In applied investigations, one is often interested in comparing some characteristic (such as the mean, the variance or a measure of association between two characters) of a group with a specified value, or in comparing two or more groups with regard to the characteristic. For instance, one may want to compare two varieties of wheat with regard to the mean yield per hectare or to compare different lines of a crop in respect of variation between plants within lines. In making such comparisons one cannot rely on the numerical magnitudes of the index of comparison such as the mean, variance or measure of association. This is because each group is represented only by a sample of observations and if another sample were drawn the numerical value would change. This variation between samples from the same population can at best be reduced in a well-designed controlled experiment but can never be eliminated. One is forced to draw inference in the presence of the sampling fluctuations which affect the observed differences between groups, clouding the real differences. Statistical science provides an objective procedure for distinguishing whether the observed difference imply any real difference among groups. Such a procedure is called a test of significance. The test of significance is a method of making due allowance for the sampling fluctuation affecting the results of experiments or observations. The fact that the results of biological experiments are affected by a considerable amount of uncontrolled variation makes such tests necessary. These tests enable us to decide on the basis of the sample results, if

(i) The deviation between the observed sample statistic and the hypothetical value, or

(ii) The deviation between two independent sample statistics; is significant or might be attributed to chance or the fluctuations of sampling.

For applying the tests of significance, we first set up a hypothesis - a definite statement about the population parameters. In all such situations we set up an exact hypothesis such as, the treatments or variable in question do not differ in respect of the mean value, or the variability, or the association between the specified characters, as the case may be, and follow an objective procedure of analysis of data which leads to a conclusion of either of two kinds:

i) reject the hypothesis, or
ii) do not reject the hypothesis

Null Hypothesis:

A definite statement about the population parameter. Such a hypothesis, which is usually a hypothesis of no difference, is called null hypothesis and is usually denoted by H_0.

According to Prof. R. A. Fisher, null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true.

Alternative Hypothesis:

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H_1.

For example, if we want to test the null hypothesis that the population has a specified mean μ_0 (say), i.e., $H_0: \mu = \mu_0$

Then the alternative hypothesis could be:

(i) $H_1: \mu \neq \mu_0$ (i.e., $\mu > \mu_0$ Or $\mu < \mu_0$)

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

(ii) $H_1: \mu > \mu_0$

(iii) $H_1: \mu < \mu_0$

The alternative hypothesis in (i) is known as a two tailed alternative and the alternatives in (ii) and (iii) are known as right tailed and left tailed alternatives respectively. The setting of alternative hypothesis is very important since it enables us to decide whether we have to use a single tailed (right or left) or two tailed test.

Errors in sampling:

The main objective in sampling theory is to draw valid inferences about the population parameters on the basis of the sample results. As such we are liable to commit the following two types of errors:

Type I error: reject H_0 when it is true

Type II error: accept H_0 when it is wrong, i.e., accept H_0 when H_1 is true.

In practice, type I error amounts to rejecting a lot when it is good and type II error may be regarded as accepting the lot when it is bad.

Thus P (Reject a lot when it is good)= α

P (Accept a lot when it is bad)=β

Critical region:

A region which amounts to rejection of H_0 is termed as critical region of rejection.

Level of Significance:

The probability α is known as the level of significance. In other words, level of significance is the size of the type I error.

The level of significance usually employed in testing of hypothesis is 5% and 1%. The level of significance is always fixed in advance before

collecting the sample information. A significance level .05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

The Test Statistic: The test statistic is a value used in making a decision about the null hypothesis, and is found by converting the sample statistic to a score with the assumption that the null hypothesis is true.

$$Test\ \ statistic = \frac{Statistic - Parameter}{S\tan dard\ \ error\ \ of\ \ the\ \ statistic}$$

The test statistic often follows a well-known distribution (eg, normal, *t*, or chi-square).

$$\overline{X} \quad \text{or} \quad \text{equivalently,} \quad Z_c = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Critical values or Significant values:

The value of test statistic which separates the critical (or rejection) region and the acceptance region is called the critical value or significant value. It depends on

(i) The level of significance used and

(ii) The alternative hypothesis, whether it is a two tailed or single tailed.

Procedure for testing of hypothesis:

The various steps in testing of a statistical hypothesis are given below:

1. **Null hypothesis:** set up the Null Hypothesis H_0
2. **Alternative hypothesis:** set up the Alternative Hypothesis H_1

This will enable us to decide whether we have to use a single tailed test or two tailed test.

3. **Level of Significance:** Choose the appropriate level of significance (α) depending on the reliability of the estimates and permissible risk. This is to be decided before sample is drawn, i.e., α is fixed in advance.

4. **Test statistic:** compute the test statistic

5. **Conclusion:** compare the computed value of the test statistic with the statistical table value. If the calculated value is less than the table value then null hypothesis become not significant. Otherwise hypothesis is significant.

t distribution

W.S. Gosset, who wrote under pen name of student defined t distribution on his research paper "the probable error of the mean", published in 1908.

The probability density function of t distribution f (t) is given by

$$\frac{1}{\sqrt{\vartheta}\beta(\frac{1}{2},\frac{\vartheta}{2})}\frac{1}{(1+\frac{t^2}{\vartheta})^{\frac{(\vartheta+1)}{2}}}$$

$-\infty<t<\infty$

Applications of t distribution

1. To test if the sample mean differs significantly from the hypothetical value of the population mean.
2. To test the significance of the difference between two sample means
3. To test the significance of an observed sample correlation coefficient and sample regression coefficient
4. To test the significance of observed partial correlation coefficient.

Test of significance for single mean small sample ($n \leq 30$):

If a random sample of size n is selected from a population with mean μ then the sample mean $\bar{Y}$ is distributed with mean μ and variance s^2/n i.e. $\bar{Y}$ ~ N(μ, s^2/n), where $S^2 = \frac{1}{n-1}\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)$. Then tests on μ are summarized as follows:

Thus the hypotheses defined are

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

H_o: $\mu = \mu_o$ (some specified value)

H_1: $\mu \neq \mu_o$ or $\mu > \mu_o$ or $\mu < \mu_o$

(two-tailed) (one-tailed)

The test statistic *t* is given by

$$t = \frac{\bar{Y} - \mu_0}{\sqrt{s^2/n}}$$

where *t* has a Student's t-distribution with $\nu = n - 1$ degrees of freedom.

RR: $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$ or $t < -t_{\alpha}$ or $t > t_{\alpha}$

t test for difference of means:

Suppose we want to test if two independent samples x_i (i=1,2,…,n_1) and y_j(j=1,2,…,n_2) of sizes n_1 and n_2 have been drawn from two normal populations with means $\mu_x$and μ_y respectively.

Under the null hypothesis that the samples have been drawn from the normal populations with means $\mu_x$and $\mu_y$and under the assumption that the population variance are equal, i.e., $\sigma_x^2 = \sigma_y^2$

The t- test statistic is $\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{S\sqrt{(\frac{1}{n_1}+\frac{1}{n_2})}}$

Where $\bar{x}$ and $\bar{y}$ are both sample mean of x and y respectively

The sample variance S^2 is given by $\frac{1}{n_1+n_2-2}\left[\sum(x_i-\bar{x})^2+\sum(y_j-\bar{y})^2\right]$

is an unbiased estimate of the common population variance.

Paired t test for difference of Means

Let us now consider the case when the sample sizes are equal, i.e., $n_1 = n_2$ = n (say) and the two samples are not independent but the sample observations are paired together. The problem is to test if the sample means differ significant or not.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

For example, suppose we want to test the efficacy of a particular drug, say, for inducing sleep. Let x_i and y_i (i=1,2,…,n) be the readings, in hours of sleep, on the i^{th} individual, before and after the drug is given respectively.

Here we consider the increments, $d_i = x_i - y_i$

Under the null hypothesis, increments are due to fluctuations of sampling, i.e., the drug is not responsible for these increments,

The paired t test statistic

$$t = \frac{\bar{d} - (\mu_1 - \mu_2)}{\sqrt{s_d^2 / n}} \quad \text{Or} \quad \frac{\bar{d}}{s_d \sqrt{n}}$$

with $(n-1)$ degrees of freedom where $d = X - Y$, the difference between the paired observations, and s_d is the corresponding standard deviation.

Where the sample variance $S_d{}^2$ is given by $\frac{\sum_i (d_i - \bar{d})^2}{n-1}$. Follows t distribution with (n-1) d.f

CHI-SQUARE TEST FOR CATEGORIES OF DATA

Background: The Z and Student's t-test were used to analyse measurement data, which, in theory, are continuously variable. Between measurements of, say, 1m and 2m there is a continuous range from 1.0001 to 1.9999m.

But in some types of experiment we wish to record how many individuals fall into a particular category, such as blue eyes or brown eyes, motile or non-motile cells, etc. These counts, or **enumeration data**, are discontinuous (1, 2, 3 etc.) and must be treated differently from continuous data. Often the appropriate test is chi-square (χ^2), which we use to test whether the number of individuals in different categories fit a **null hypothesis** (an expectation of some sort).

Some assumptions:

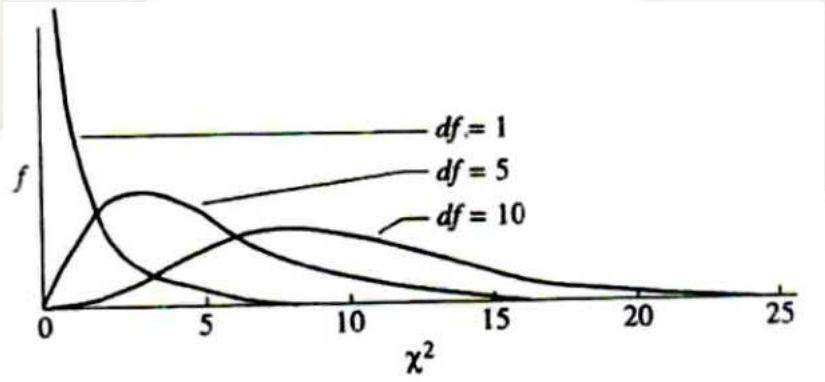
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The data are collected randomly (you just can't get away from this one!)
- The smallest expected # is at least 5. Remember, the chi-square test is an approximation, and approximations get better the bigger the sample size.

Conversely, they get worse, the smaller the sample size

The Chi-square distribution

- The value you calculate, Chi-square, will follow a Chi- squared distribution if n is moderately large.
- Just like many other distributions, the value of the Chi-square distribution depends on d.f. (degree of freedom) (or nu).
- Here is what it looks like for a couple of different values of nu:
- Just like before, we reject for values that windup in the tails (usually only the upper tail).
- Incidentally, what would a distribution look like composed of squared normally distributed variables?



Chi-square distributions for various *df* values

Chi-Square Test for Goodness of Fit :

A test of wide applicability to numerous problems of significance in frequency data is the

Chi-Square (χ^2) test of goodness of fit. It is a non-parametric test that is used to find out how the observed value of a given phenomenon is

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

significantly different from the expected value. The term goodness of fit is used to compare the observed sample distribution with the expected probability distribution and determines how well theoretical distribution (such as normal, binomial, or Poisson) fits the empirical distribution

Set up the hypothesis for Chi-Square goodness of fit test:

A. Null hypothesis: The fitted distribution is a good fit to the given data

B. Alternative hypothesis: The fitted distribution is not a good fit to the given data

Compute the value of Chi-Square goodness of fit test using the following formula:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where, χ^2= Chi-Square goodness of fit test O = observed value E= expected value

Degree of freedom: In Chi-Square goodness of fit test, the degree of freedom depends on the distribution of the sample. The following table shows the distribution and an associated degree of freedom:

Type of distribution	No of constraints	Degree of freedom
Binominal distribution	1	n-1
Poisson distribution	2	n-2
Normal distribution	3	n-3

Interpretation: The calculated value of Chi-Square goodness of fit test is compared with the table value. If the calculated value of Chi-Square goodness of fit test is greater than the table value, we will reject the null hypothesis and conclude that the fitted distribution is a good fit to the given data. If the calculated value of Chi-Square goodness of fit test is less than the table value, we will accept the null hypothesis and conclude that the fitted distribution is not a good fit to the given data.

Test for goodness of fit of Mendelian ratios.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Counting the numbers of various kinds of offspring, the likelihood of obtaining exactly a 1 : 1 or a 3 : 1 ratio is small due to random sampling effects. So one issue that needs to be addressed is how different from the expected ratio do our data need to be before we conclude something else is going on.

To address this question we use a commonly used statistical hypothesis test, called a goodness of fit test. Using the test, we test whether our observed data fit our expectation reasonably well, or do we have reason to reject the hypothesis that a particular Mendelian ratio occurs. One such goodness of fit test is often called the Chi-Square goodness of fit test. Equation for the goodness of fit test is: $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

Where the O_i are the observed values and E_i are the expected values of each of the 1, 2, ...k classes. So calculate the χ^2 statistic from the data set and the expected numbers of offspring. If the observed data match exactly the expected, then $\chi^2 = 0$. The more different the observed and expected, the bigger the χ^2 value you calculate becomes. At some point, we need to decide how big a χ^2value is too big. In other words, indicates a true departure of the observed from the expected. To know how big is too big, we can refer to the theoretical values of the χ^2 distribution. By convention in most sciences, we use the value of $\alpha = 0.05$. This value indicates that the probability of us observing a χ^2value as big or bigger than the one we obtained if the data actually were sampled from the expected ratio. $\alpha = 0.05$ means there is a 5% (small) chance of getting such a large χ^2 if the data were truly sampled from a population with the expected ratio. The larger the value of χ^2, greater is the discrepancy between observed and expected frequencies. **We reject the null hypothesis if.** $\chi^2 > \chi^2_{\alpha,K-1}$

Chi-Square Test of Independence of Attribute:

The Chi-Square test of independence is used to determine if there is a significant relationship between two nominal (categorical) variables. The

frequency of each category for one nominal variable is compared across the categories of the second nominal variable. The data can be displayed in a contingency table where each row represents a category for one variable and each column represents a category for the other variable.

For example, say a researcher wants to examine the relationship between gender (male vs. female) and empathy (high vs. low). The chi-square test of independence can be used to examine this relationship. The null hypothesis for this test is that there is no relationship between gender and empathy. The alternative hypothesis is that there is a relationship between gender and empathy (e.g. there are more high-empathy females than high-empathy males).

Independence of two variables of classification

Null Hypothesis: The null hypothesis asserts the independence of the variables under consideration. (for example, gender and voting behaviour are independent of each other)

For 2 x 2 Contingency table: Let where a, b, c and d are the observed frequencies

Attribute A	Attribute B		Total
	a	b	a + b
	c	d	c + d
Total	a + c	b + d	N

N = a + b + c + d. If all the expected frequencies are ≥ 5, then χ^2 can be calculated by using the formula

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \sim \chi_1^2(\alpha)$$ We reject the null hypothesis if $\chi^2 > \chi_1^2(\alpha)$

Yates correction: The **correction** is to prevent overestimation of calculated value, used when at least one cell of the table has an expected count

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

smaller than 5. Yates correction only applies when we have two categories (one degree of freedom), (e.g. male/ female), the *Chi sq* test should not, strictly, be used. There have been various attempts to correct this deficiency, but the simplest is to apply Yates correction to our data.

To do this, we simply subtract 0.5 from each calculated value of "O-E", ignoring the sign (plus or minus). In other words, an "O-E" value of +5 becomes +4.5, and an "O-E" value of -5 becomes -4.5. To signify that we are reducing the absolute value, ignoring the sign, we use vertical lines: |O-E|-0.5. Then we continue as usual but with these new (corrected) O-E values: we calculate (with the corrected values) $(O-E)^2$, $(O-E)^2/E$ and then sum the $(O-E)^2/E$ values to get χ^2. **OR** Another formula for Yates correction will be :

Adjusted $\chi^2 = \sum \frac{\left(|O-E|-0.5\right)^2}{E} \sim \chi_1^2(\alpha)$ ***OR***

For 2 × 2 table If any **expected cell frequency is < 5**, then applying Yates correction, we will be having the formula as $\chi^2 = \frac{N\left(|ad-bc|-\frac{N}{2}\right)^2}{(a+b)(a+c)(b+d)(c+d)} \sim \chi_1^2(\alpha)$.

We reject the null hypothesis if $\chi^2 > \chi_{1df}^2$ at 5% level of significance. If $\chi^2 = 0$, observed and expected frequencies agree exactly while if $\chi^2 > 0$, they do not agree exactly. The larger the value of χ^2 greater is the discrepancy between the observed and expected frequencies.

In two-way table with *r* rows and *c* columns contains r x c sample counts. Let N_{ij} denote the number of observations in the *i*th row and the *j*th column, $i = 1,2,...,r$,. $j = 1,2,.....c$ Let C_j are the column totals and R_i are the row totals.

Thus, $C_j = \sum_{i=1}^{r} N_{ij}$, $R_i = \sum_{j=1}^{c} N_{ij}$, and $n = \sum_{i=1}^{r} N_{ij} = \sum_{j=1}^{c} N_{ij}$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The expected count in the *ij*th cell of the table is denoted by E_{ij}. For an *r x c* table, the *expected counts/* are calculated from the marginal totals in the samples count table using the formula. $E_{ij} = \frac{R_i C_j}{n}$

To test H_o, that there is no relationship between the row and column classifications, a statistic called the *chi-square statistic* is used. This statistic compares the sample counts with their expected values. It is calculated from the following formula:

$$\chi^2 = \sum_{i-1}^{r} \sum_{j-1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where *observed* O_{ij}represents the sample counts, and *expected* E_{ij} represents the expected counts, and the sum is over all *r x c* entries in the sample or expected count tables.

The null hypothesis to be tested is that the row and column classifications are independent. The alternative hypothesis is that the null hypothesis is not true.

If H_o is true, the statistic χ_c^2 has approximately a χ^2 distribution with (*r* - 1)(*c* - 1) degrees of freedom.

Interpretation: If the observed chi-square test statistic is greater than the critical value, the null hypothesis can be rejected.

Glossary of the term:

Null Hypothesis: A definite statement about the population parameter. Such a hypothesis, which is usually a hypothesis of no difference, is called null hypothesis and is usually denoted by H_0.

Alternative Hypothesis:

Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis, usually denoted by H_1.

Type I error:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Reject the null hypothesis, When it is true

Type II error:

Accept the null hypothesis , when it is wrong, i.e., accept H_0 When H_1 is true.

Critical region:

A region which amounts to rejection of H_0 is termed as critical region of rejection.

Level of Significance:

The probability α is known as the level of significance. In other words, level of significance is the size of the type I error.

Critical values or significant values:

The value of test statistic which separates the critical (or rejection) region and the acceptance region is called the critical value or significant value.

Parameter:

Parameters are constant that summarize data for an entire population.

Statistics:

Statistics are constants that summarize data from a sample.

Degrees of freedom:

Degree of freedom is the number of independent observations in a sample of data that are available to estimate a parameter of the population from which that sample is drawn.

Population:

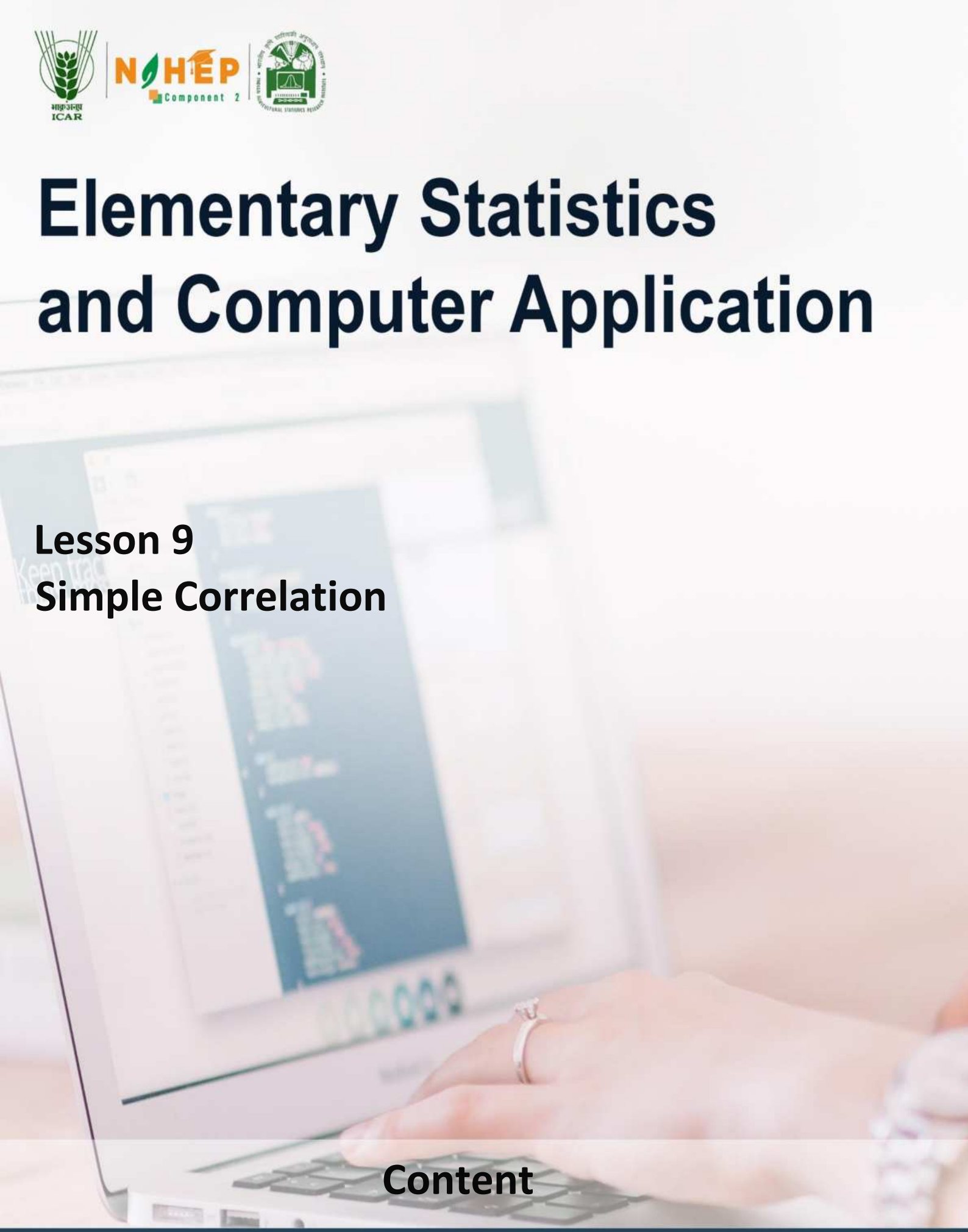
A population is the entire group that you want to draw conclusions about the population.

Sample:

A sample is the specific group that you will collect data from.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application



Lesson 9
Simple Correlation

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 9	Simple Correlation
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. Meaning of correlation and scatter diagram to study correlation.
2. Karl-Pearson's' Correlation coefficient.
3. Testing the significance of correlation coefficient.
4. Coefficient of determination and coefficient of alienation.

8.1 CORRELATION

- Correlation is the study of relationship between two or more variables. When there are two related variables their joint distribution is known as bivariate distribution and the type of correlation study is known as simple correlation.
- Suppose we have two continuous variables X and Y and if the change in X affects Y, the variables are said to be correlated. In other words, the systematic relationship between the variables is termed as correlation. When only 2 variables are involved the correlation is known as simple correlation and when more than 2 variables are involved the correlation is known as multiple correlation. When the variables move in the same direction, these variables are said to be correlated positively and if they move in the opposite direction they are said to be negatively correlated.

8.2 TYPES OF CORRELATION

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

On the basis of nature of relationship between two variables, correlation may be:

- **Positive correlation:** If the increase (or decrease) in one variable results a corresponding increase (or decrease) in the other variable i.e., two variables move in the same direction, variables are said to be positively correlated. E.g.: heights and weights of children, price and supply of a set of commodities etc.
- **Negative correlation:** If the increase (or decrease) in one variable results a corresponding decrease (or increase) in the other variable i.e., two variables deviate in the opposite direction, variables are said to be negatively correlated. E.g.: price and demand of a set of commodities, speed and time etc.
- **Perfect positive correlation:** When changes in two related are exactly proportional there is perfect correlation between them.

  If the increase (or decrease) in one variable results a corresponding and proportional increase (or decrease) in the other variable, variables are said to be perfect positively correlated. E.g.: radius and circumference of a circle etc.

- **Perfect negative correlation:** If the increase (or decrease) in one variable results a corresponding and proportional decrease (or increase) in the other variable, variables are said to be perfect negatively correlated. E.g.: pressure and volume etc.
- **Zero correlation:** If the change in one variable does not result any change in the other variable, variables are said to be uncorrelated or

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

independent or having zero correlation. E.g. Price of commodities and weight of individuals.

8.3 SCATTER DIAGRAM

- A diagrammatic way to investigate, whether there is any relation between the variables X and Y we use scatter diagram. Let (x_1,y_1), (x_2,y_2)….(x_n,y_n) be n pairs of observations. If the variables X and Y are plotted along the X-axis and Y-axis respectively in the x-y plane of a graph sheet the resultant diagram of dots is known as scatter diagram.
- From the scatter diagram we can say whether there is any correlation between X and Y and whether it is positive or negative or the correlation is linear or curvilinear.

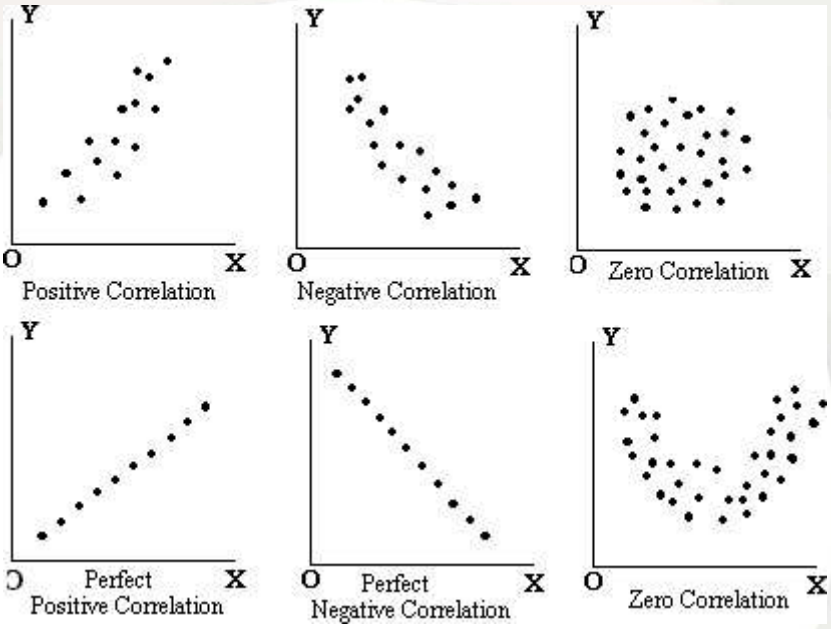


Fig. Different types of simple correlation between two variables.

8.4 PEARSON'S CORRELATION COEFFICIENT

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The measures of the degree of linear relationship between two quantitative variables is called Karl Pearson's correlation coefficient or product moment correlation coefficient. It is also sometimes simply called correlation coefficient and is denoted by r (in case of sample) and ρ (in case of population).
- The correlation coefficient r is given as the ratio of covariance of the variables X and Y to the product of the standard deviation of X and Y. Symbolically,

$$r = \frac{\sum(x - \bar{x})\,(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2}\sqrt{\sum(y - \bar{y})^2}}$$

which can be simplified as,

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}\sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

- The numerator is termed as sum of product of X and Y and abbreviated as SP(XY). In the denominator the first term is called sum off squares of X (i.e.) SS(X) and second term iscalled sum of squares of Y (i.e.) SS(Y).

$$r = \frac{SP(XY)}{\sqrt{SS(X)}\sqrt{SS(Y)}}$$

The denominator in the above formula is always positive. The numerator may be positive or negative making r to be either positive or negative.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

8.5ASSUMPTIONS OF KARL PEARSON'S CORRELATION COEFFICIENT

- Correlation coefficient r is used under certain assumptions, they are
  - The variables under study are continuous random variables and they are normally distributed.
  - The relationship between the variables is linear.
  - Each pair of observations is unconnected with other pair (independent).

8.6 PROPERTIES OF KARL PEARSON'S CORRELATION COEFFICIENT

- The correlation coefficient value ranges between –1 and +1.
- The correlation coefficient is not affected by change of origin or scale or both.
- If r > 0, it denotes positive correlation between the two variables x and y.
- If r< 0, it denotes negative correlation between the two variables x and y.
- If, r = 0 then the two variables x and y are not linearly correlated.
- If r = +1, then the correlation is perfect positive
- If r = -1, then the correlation is perfect negative.
- r is a pure number, independent of units of measurement.
- r is symmetric i.e. $r_{xy} = r_{yx}$

8.7 TESTING THE SIGNFICANCE OF CORRELATION COEFFICIENT

- The significance of r can be tested by Student's t test. The test statistic is given by,

$$t = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}}$$

This t is distributed as Student's t distribution with (n-2) degrees of freedom. The significance level is usually 5% or 1% as the investigator deems it fit to be.

Example: Compute correlation coefficient between plant height (cm) and yield (kgs) as per the data given below:

Plant height (cm)	39	65	62	90	82	75	25	98	36	78
Yield in Kgs	47	53	58	86	62	68	60	91	51	84

Also, test the significance of the correlation coefficient.

Solution:

H_0: The correlation coefficient r is not significant ρ= 0

H_1: The correlation coefficient r is significant. ρ≠ 0

Level of significance be 5%

From the data,

$$n = 10, \quad \sum x = 650, \quad \sum y = 660,$$

$$\sum xy = 45604, \quad \sum x^2 = 47648, \quad \sum y^2 = 45784$$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}} \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

$$= \frac{45704 - \frac{(650)(660)}{10}}{\sqrt{47648 - \frac{650^2}{10}}\sqrt{45784 - \frac{660^2}{10}}}$$

$$= 0.7804$$

Correlation coefficient is positively correlated.

Test statistic:

$$t = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} \sim (n-2)d.f$$

$$t = \frac{0.7804}{\sqrt{\frac{1-0.7804^2}{10-2}}} = 3.530$$

$$t_{tab} = t_{(10-2,5\% l.o.s)} = 2.306$$

Inference:

Since, t>t_{tab}, we therefore have evidence to not accept the null hypothesis and we concluded that the correlation coefficient r is significant (i.e.) there is a significant positive relationship between plant height and its yield.

8.8 COEFFICIENT OF DETERMINATION AND COEFFICIENT OF ALIENATION

- The relationship between the variables is interpreted by the square of the correlation coefficient (r^2) which is called the coefficient of determination.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The value 1-r^2 is called as coefficient of alienation.
- If r^2 is 0.72, it implies that on the basis of the samples, 72% of the variation in one variable is caused by the variation of the other variable. The coefficient of determination is used to compare two correlation coefficients.

****

GLOSSARY:

- **Correlation:** Correlation is the study of relationship between two or more variables.
- **Positive correlation:** If the increase (or decrease) in one variable results a corresponding increase (or decrease) in the other variable, the variables are said to have positive correlation.
- **Negative correlation:** If the increase (or decrease) in one variable results a corresponding decrease (or increase) in the other variable, the variables are said to have negative correlation.
- **Perfect correlation:** When changes in two related are exactly proportional there is perfect correlation between the variables.
- **Zero correlation:** If the change in one variable does not result any change in the other variable, variables are said to be uncorrelated or independent or having zero correlation.
- **Scatter diagram:** Scatter diagram is a diagrammatic way to investigate, whether there is any relation between the variables X and Y.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

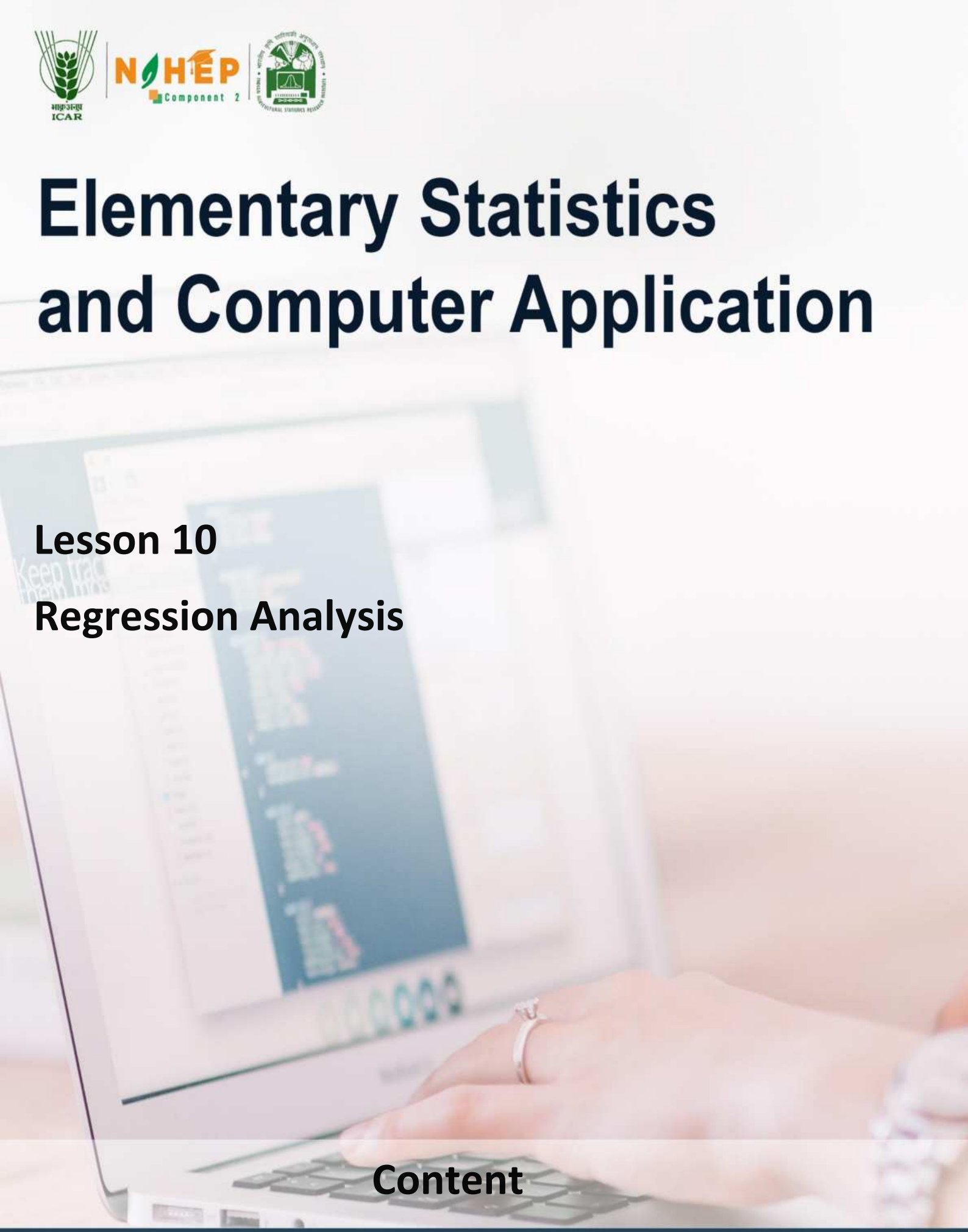
- **Karl Pearson's correlation coefficient:** It is a degree of linear relationship between two quantitative variables and is given as the ratio of covariance of the variables X and Y to the product of the standard deviation of X and Y.
- **Coefficient of determination:** The Square of the correlation coefficient is the coefficient of determination.

REFERENCES

- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.
- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application



Lesson 10

Regression Analysis

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 10	Regression Analysis
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

OBJECTIVES:

1. Basic knowledge of simple linear regression analysis.
2. Differences between correlation and regression analysis.
3. Idea about regression equations and regression coefficients.
4. Properties of regression coefficients.

9.1 Simple Linear Regression Analysis

- Regression analysis measures the nature and extent of the relationship between two or more variables, thus enabling us to make predictions.
- Regression is the functional relationship between two variables and of the two variables one may represent cause and the other may represent effect.
- The variable representing cause is known as independent variable and is denoted by X. The variable X is also known as *predictor variable* or *regressor*.
- The variable representing effect is known as dependent variable and is denoted by Y and is also known as *predicted variable*.
- The relationship between the dependent and the independent variable may be expressed as a function and such functional relationship is termed as regression.
- When there are only two variables the functional relationship is known as simple regression. When there are more than two

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

variables and one of the variables is dependent upon others, the functional relationship is known as multiple regression.

- The linear algebraic equation used for expressing a dependent variable in terms of independent variable is called linear regression equation.

9.2 Differences between Correlation and Regression Analysis

- Developing an algebraic equation between two variables from sample data and predicting thevalue of one variable, given the value of the other variable is referred to as regression analysis,while measuring the strength (or degree) of the relationship between two variables is referred as correlation analysis.
- Correlation analysis determines an association between two variables *X* and *Y*. Regression analysis, in contrast to correlation, determines the cause-and-effect relationship between *X* and *Y*.
- In linear regression analysis one variable is considered as dependent variable and other as independent variable, while in correlation analysis both variables are considered to be independent.
- Correlation doesn't help in making predictions while Regression enables us to make predictions using regression line.
- Correlation coefficients are symmetrical *i.e.,* $r_{XY} = r_{YX}$. Regression coefficients are not symmetrical *i.e.,* $b_{XY} \neq b_{YX}$.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Correlation is independent of the change of origin and scale. Regression coefficient is independent of change of origin but not of scale.

9.3 Regression Equations and Regression Coefficients

- The device used for estimating the values of one variable from the value of the other consists of a line through the points, drawn in such a manner as to represent the average relationship between the two variables. Such a line is called line of regression.
- The two variables X and Y which are correlated can be expressed in terms of each other in the form of straight line equations called regression equations. Such lines should be able to provide the best fit of sample data to the population data.
- The algebraic expression of regression lines is written as:
    - Regression equation of *y* on *x*

$$y = a + bx$$

is used for estimating the value of *y* for given values of *x*.

    - Regression equation of *x* on *y*

$$x = c + dy$$

is used for estimating the value of *x* for given values of *y*.

9.3.1 Nature of Regression Lines

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- When variables X and Y are correlated perfectly (either positive or negative), *i.e.*, if $r = \pm 1$, these lines coincide.
- Higher the degree of correlation, nearer the two regression lines are to each other.
- If $r = 0$, then the two regression lines intersect each other at 90°.
- Two linear regression lines intersect each other at the point of the average value of variables X and Y.

9.3.2 Regression Coefficients

- The fitted or estimated regression equation representing the straight line regression model is in the form

$$\hat{y} = a + bx$$

where $\hat{y}$ estimated average (mean) value of dependent variable *y* for a given value of independent variable *x*.

a = *y*-intercept that represents average value of $\hat{y}$

b = slope of regression line that represents the expected change in the value of *y* for unitchange in the value of *x*.

- To determine the value of $\hat{y}$ for a given value of *x*, this equation requires the determination of two unknown constants '*a*' (intercept) and '*b*' (also called regression coefficient). Once these constants are calculated, the regression line can be used to compute an estimated value of the dependent variable *y* for a given value of independent variable *x*.
- The regression coefficient '*b*' is also denoted as:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- b_{yx} (regression coefficient of *y* on *x*) in the regression line, *y* = *a* +$b_{yx}x$
- b_{xy}(regression coefficient of *x* on *y*) in the regression line, *x* = *c* + $b_{xy}y$

- The regression coefficient of *y* on *x* is given by

$$b_{yx} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{Cov(x,y)}{{\sigma_x}^2}$$

and $$a = \bar{y} - b_{yx}\bar{x}$$

- The regression coefficient of *x* on *y* is given by

$$b_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{(\sum y)^2}{n}} = \frac{Cov(x,y)}{{\sigma_y}^2}$$

and $$c = \bar{x} - b_{xy}\bar{y}$$

- Regression Coefficients in terms of Correlation Coefficient:

$$b_{yx} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{Cov(x,y)}{{\sigma_x}^2} = r\frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(y-\bar{y})^2} = \frac{Cov(x,y)}{{\sigma_y}^2} = r\frac{\sigma_x}{\sigma_y}$$

9.4 Properties of regression coefficients

- The correlation coefficient is the geometric mean of two regression coefficients, that is, , $r = \mp\sqrt{b_{xy}b_{yx}}$
-

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- If one regression coefficient is greater than one, then other regression coefficient must be less than one, because the value of correlation coefficient *r* cannot exceed one. However, both the regression coefficients may be less than one.
- Both regression coefficients must have the same algebraic sign.
- The correlation coefficient will have the same sign (either positive or negative) as that of the two regression coefficients.
- The arithmetic mean of regression coefficients b_{xy} and $b_{yx}$is more than or equal to the correlation coefficient *r*, that is, $\frac{b_{xy}+b_{yx}}{2} \geq r$.
- Regression coefficient is independent of change of origin but not of scale

9.5 Shortcut Method of Checking Regression Equations

- Suppose two regression equations are as follows:
  - $a_1x + b_1y + c_1 = 0$
  - $a_2x + b_2y + c_2 = 0$
- Case 1: If $a_1b_2 \leq a_2b_1$ (in magnitude, ignoring negative), then
  - $a_1x + b_1y + c_1 = 0$ is the regression of Y on X
  - $a_2x + b_2y + c_2 = 0$ is the regression of X on Y
- Case 2: If $a_1b_2 > a_2b_1$ (in magnitude, ignoring negative), then
  - $a_1x + b_1y + c_1 = 0$ is the regression of X on Y
  - $a_2x + b_2y + c_2 = 0$ is the regression of Y on X

Example: In trying to estimate the effectiveness of an advertising campaign, a firm compiled the following data:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Year	2008	2009	2010	2011	2012	2013	2014	2015
Adv. Expenditure ('000 Rs.)	12	15	15	23	24	38	42	48
Sales ('00000 Rs.)	5.0	5.6	5.8	7.0	7.2	8.8	9.2	9.5

Estimate the regression equation of advertising expenditure on sale and sale on advertising expenditure. Also estimate the probable sale when the expenditure is Rs. 60,000.

Solution: Consider the variables advertising expenditure as X and sale by Y.

To determine the regression co-efficients, the following table is prepared:

X	Y	X^2	Y^2	XY
12	5.0	144	25	60
15	5.6	225	31.36	84
15	5.8	225	33.64	87
23	7.0	529	49	161
24	7.2	576	51.84	172.8
38	8.8	1444	77.44	334.4

42	9.2	1764	84.64	386.4
48	9.5	2304	90.25	456
217	58.1	7211	443.17	1741.6

Now,

$$\bar{X} = \frac{\sum X}{n} = \frac{217}{8} = 27.125$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{58.1}{8} = 7.26$$

$$Cov(X,Y) = \frac{1}{n}\sum XY - \bar{X}\bar{Y} = \frac{1}{8} \times 1741.6 - 27.125 \times 7.26 = 20.78$$

$${\sigma_x}^2 = \frac{1}{n}\sum x^2 - \bar{x}^2 = \frac{1}{8} \times 7211 - 27.125^2 = 165.38$$

$${\sigma_y}^2 = \frac{1}{n}\sum y^2 - \bar{y}^2 = \frac{1}{8} \times 443.17 - 7.26^2 = 2.65$$

$$b_{yx} = \frac{Cov(x,y)}{{\sigma_x}^2} = \frac{20.78}{165.38} = 0.125$$

$$a = \bar{y} - b\bar{x} = 7.26 - 0.125 \times 27.125 = 3.87$$

Regression line of Y on X is:

$$y = 3.87 + 0.125x$$

Again,

$$b_{xy} = \frac{Cov(x,y)}{{\sigma_y}^2} = \frac{20.78}{2.65} = 7.84$$

$$c = 27.125 - 7.84 \times 7.26 = -29.79$$

Regression line of X on Y is:

$$x = -29.79 + 7.84y$$

When expenditure (x) is Rs. 60,000 then sales will be

$$y = 3.87 + 0.125 \times 60 = 11.37$$

****

GLOSSARY:

- **Regression analysis:** Regression analysis measures the nature and extent of the relationship between two or more variables, thus enabling us to make predictions
- **Regression:** Regression is the functional relationship between two variables and of the two variables where one represent cause and the other represent effect.
- **Predictor variable:** The variable representing cause is known as independent variable and is denoted by X. The variable X is also known as *predictor variable* or *regressor*.
- **Predicted variable:** The variable representing effect is known as dependent variable and is denoted by Y and is also known as *predicted variable*.
- **Regression equations:**The two variables X and Y which are correlated can be expressed in terms of each other in the form of straight line equations called regression equations.

REFERENCES

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- S. C. Gupta and V. K. Kapoor. Fundamentals of Mathematical Statistics. Sultan Chand & Sons, New Delhi.
- A. M. Gun, M. K. Gupta and B. Dasgupta. Fundamentals of Statistics (Volume One). The World Press Private Ltd, Kolkata.
- J. K. Sharma. Fundamentals of Business Statistics. Dorling Kindersley (India) Pvt. Ltd.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson 11

Experimental Designs

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 11	Experimental Designs
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

OBJECTIVES:

- To identify the experimental unit and the types of variables.
- To define the design structure and to test or establish a hypothesis
- To study the basic principles of experimental design
- To describe the important methods for increasing the accuracy of field experiments.
- To study the basic experimental designs like CRD, RBD and LSD with statistical models.

1.1 Basics of Experimental Design

- An absolute experiment is an experiment in which certain absolute values like average, correlation coefficient, median, mode etc are worked out to describe the population. In other type of experiment, where the researcher compares the effects of different objects/entities under consideration are known as comparative experiment.
- In a comparative experiment, for example, an experimental design is done in such a way that the experimenter can compare the objects or entities under identical conditions keeping the other sources of variations as minimum as possible.
- In agriculture, sometimes we are interested to study the varietal performance of different varieties of certain crop at a particular

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

place, so that a variety could be recommended for the specific crop for that area.

- The main idea in designing an experiment is to answer the question in mind of the researchers in a valid, efficient, economical way providing due consideration to the constraints and situations provided for the experiment.

1.2 Basic Principles of Design of Experiment

There are three basic principles of design of experiment. They are –

- **Replication**: The practice of assigning more than one plot to a treatment is called replication. Without the use of this principle, we cannot get an estimate of experimental error in a design of experiment.
- **Randomization**: When all the treatments have equal chance of being allocated to different experimental units it is known as randomization. If our conclusions are to be valid, treatment means and differences among treatment means should be estimated without any bias. For this purpose, we use the technique of randomization.
- **Local Control**: Local control is the technique which helps in reduction of experimental error, providing due consideration to the information available under the local condition where the actual experiment is supposed to be set.

1.3 Uniformity Trial

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The twin objectives in designing of experiment are the valid estimation of different effects along with the reduction in experimental error.
- To accomplish both the objectives, one should have a clear idea about the area/subject on which treatments are to be applied.
- In experiments, particularly in field experiments, in order to have an idea about the conditions of the proposed experimental area (with respect to soil fertility/productivity, soil conditions) a trial, known as uniformity trial is conducted.
- In Uniformity trial, generally a short duration crop is grown with uniform package of practices by dividing whole area into smallest unit plots.
- Sometimes division of plots into smallest unit is also done before recording the response. Responses are recorded from the basic unit plots. The overall mean response is also worked out.
- Uniformity trials also gives us some idea about the shape and size of the plots to be used.

1.4 Analysis of Variance

Why ANOVA?

- In real life things do not typically result in two groups being compared.
- Two-sample t-tests are problematic.
  - Increasing the risk of a Type I error.
  - At .05 level of significance, with 100 comparisons, 5 will show a difference when none exists (experiment-wise error).

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- So the more t-tests you run, the greater the risk of a type I error (rejecting the null when there is no difference).
- ANOVA allows us to see if there are differences between means with an COMPLETE test.

Analysis of variance is a technique that partitions the total sum of squares of deviations of the observations about their mean into portions associated with independent variables in the experiment and a portion associated with error.

- Depending upon the nature, type of data and classification of data, the ANOVA is developed for one-way classified data, two-way classified data with one observation per cell, two-way classified data with more than one observation per cell etc.
- The analysis of variance is based on certain basic assumptions:
    - The effects are additive in nature
    - The observations are independent
    - The variable concerned must be normally distributed
    - Variances of all populations from which samples have been drawn must be the same.

1.5 Completely Randomised Design (CRD)

- Completely Randomized Design (CRD) is the simplest of all designs where only two principles of design of experiments, *i.e.,* replication and randomization have been used.
- The principle of local control is not used in this design.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- In this design the treatments are assigned completely at random so that each experimental unit has the same chance of receiving any one treatment.
- But CRD is appropriate only when the experimental material is homogeneous.
- As there is generally large variation among experimental plots due to many factors CRD is not preferred in field experiments.
- In laboratory experiments and greenhouse studies it is easy to achieve homogeneity of experimental materials and therefore CRD is most useful in such experiments.

1.5.1 Layout of CRD: Completely randomized Design is the one in which all theexperimental units are taken in a single group which are homogeneous as far as possible. The randomization procedure for allotting the treatments to various units will be as follows.

Step 1: Determine the total number of experimental units.

Step 2: Assign a plot number to each of the experimental units starting from leftto right for all rows.

Step 3: Assign the treatments to the experimental units by using random numbers.

1.5.2 Statistical Model: The statistical model for CRD with one observation per unit

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, i = 1,2, \dots t, j = 1,2, \dots, r_i$$

y_{ij} = response corresponding to the j^{th} unit receiving i^{th} treatment

μ = overall mean effect

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

α_i= true effect of the i^{th} treatment

ε_{ij} = error term of the j^{th} unit receiving i^{th} treatment and are *i.i.d.*N(0, σ^2)

1.5.3 Hypothesis and ANOVA table for CRD:

H_0 : $\alpha_1 = \alpha_2 = \ldots\ldots\ldots = \alpha_t$ against the alternative hypothesis

H_1: all α's are not equal.

The different steps in forming the ANOVA table for a CRD are:

Total Sums of Squares (TSS) = $\sum_{i=1}^{t}\sum_{j=1}^{r_i} {y_{ij}}^2 - C.F.$

where $C.F. = \frac{G^2}{N}$, G = Grand Total and N = Total no. of observations

Treatment SS (TrSS) = $\sum_{i=1}^{t} \frac{{y_i}^2}{r_i} - C.F.$

Error SS (ESS) = By subtraction = TSS – Trss

Table 1: ANOVA table for CRD

Source of Variation	d.f.	SS	MSS	F
Treatments	t – 1	TrSS	TrMS = $\frac{TrSS}{t-1}$	$\frac{TrMS}{EMS}$
Error	n – t	ESS	EMS = $\frac{ESS}{n-t}$	
Total	n – 1	TSS		

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1.5.4 Decision Rule: Compare the calculated F with the critical value of F corresponding to treatmentdegrees of freedom and error degrees of freedom so that acceptance or rejection of the null hypothesis can be determined.

1.5.5 Critical Difference: When the test is significant, *i.e.,* when the null hypothesis is rejected then one should find out which pair of treatments are significantly different and which treatment is either the best or the worst with respect to the particular characters underconsideration.This procedure is simplified with the help of least significant difference (critical difference) value as per the given formula below:

$$\text{C.D.} = t_{errordf,(\alpha/2)} \times \left(S.E(d)\right)$$

$$\text{where S.E.(d)} = \sqrt{EMS(\frac{1}{r_i} + \frac{1}{r_j})}$$

r_i = number of replications for treatment *i*

r_j= number of replications for treatment *j*, and

t is the critical *t* value for error degrees of freedom at α level of significance.

1.5.6 Advantages of CRD

1. Its layout is very easy.
2. There is complete flexibility in this design *i.e.* any number of treatments and replications for each treatment can be tried.
3. Whole experimental material can be utilized in this design.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

4. This design yields maximum degrees of freedom for experimental error.
5. The analysis of data is simplest as compared to any other design.
6. Even if some values are missing the analysis can be done.

1.5.7 Disadvantages of CRD

1. It is difficult to find homogeneous experimental units in all respects and hence CRD is seldom suitable for field experiments as compared to other experimental designs.
2. It is less accurate than other designs.

1.6 Randomised Block Design (RBD)

- When the experimental material is heterogeneous, the experimental material is grouped into homogenous sub-groups called blocks.
- As each block consists of the entire set of treatments a block is equivalent to a replication.
- If the fertility gradient runs in one direction say from north to south or east to westthen the blocks are formed in the opposite direction.
- Such an arrangement of grouping theheterogeneous units into homogenous blocks is known as randomized blocks design.
- Each block consists of as many experimental units as the number of treatments.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Thetreatments are allocated randomly to the experimental units within each blockindependently such that each treatment occurs once. The number of blocks is chosen tobe equal to the number of replications for the treatments.

1.6.1 Layout of RBD:

Step 1: Depending upon the heterogeneity among the experimental units and thenumber of treatments to be included in the experiment, the whole experimental field is divided into number of blocks, perpendicular to the direction of heterogeneity, equal to the types/group of experimental units taking information from the uniformity trial or previous experience.

Step 2: Each block is divided into number of experimental units equal to the number of treatments (say t) across the fertility gradient.

Step 3: In each block consisting of t experimental units, allocate t treatments randomly so thatno treatment is repeated more than once in any block.

1.6.2 Statistical Model: Let us suppose that we have *t* number of treatments, each being replicated *r* number of times. The appropriate statistical model for RBD will be

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \qquad i = 1,2,\dots t, j = 1,2,\dots,r$$

where,y_{ij} = response corresponding to j^{th} replication/block of the i^{th} treatment

μ = the overall mean

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

α_i = the i^{th}treatment effect

β_j = the j^{th} replication effect

ε_{ij} = the error term for i^{th} treatment and j^{th} replication and are *i.i.d,* $N(0,\sigma^2)$.

1.6.3 Hypotheses and ANOVA table for RBD:

H_{01} : $\alpha_1 = \alpha_2 = \ldots\ldots\ldots\ldots = \alpha_t$; H_{02} : $\beta_1 = \beta_2 = \ldots\ldots\ldots\ldots = \beta_r$

Against the alternative hypotheses

H_{11}: all α's are not equal; H_{12}: all β's are not equal

The different steps in forming the ANOVA table for a RBD are

Total Sums of Squares (TSS) = $\sum_{i=1}^{t}\sum_{j=1}^{r} {y_{ij}}^2 - C.F.$

where $C.F. = \frac{G^2}{N}$, G = Grand Total and N = Total no. of observations

Treatment SS (TrSS) = $\frac{1}{r}\sum_{i=1}^{t} {y_{i0}}^2 - C.F.$

Replication SS (RSS) = $\frac{1}{t}\sum_{j=1}^{r} {y_{0j}}^2 - C.F.$

Error SS (ESS) = By subtraction = TSS – Trss – RSS

Table 2: ANOVA table for RBD

Sources of Variation	d.f.	SS	MSS	F value
Replication	r – 1	RSS	RMS	RMS/EMS
Treatment	t – 1	TrSS	TrMS	TrMS/EMS
Error	(r - 1)(t - 1)	ESS	EMS	
Total	rt – 1	TSS		

1.6.4 Decision Rule: The null hypotheses are rejected at α level of significance if the calculated valuesof F ratio corresponding to treatment and replication be greater than thecorresponding table value at the same level of significance with (t – 1), (t – 1)(r – 1)and (r – 1),(t -1)(r -1) degrees of freedom respectively.

1.6.5 Least significant difference (LSD) or **Critical Difference:**

CD for Replication

$$\sqrt{\frac{2\,EMS}{t}} \times t_{(t-1)(r-1),\frac{\alpha}{2},}$$

where t is the number of treatments and $t_{(t-1)(r-1)\,,\frac{\alpha}{2},}$ is the table value of t at α level of significance and (t-1)(r-1) degrees of freedom.

CD for Treatment

$$\sqrt{\frac{2\,EMS}{r}} \times t_{\frac{\alpha}{2},(t-1)(r-1)}$$

where r is the number of treatments and $t_{(t-1)(r-1)\,,\frac{\alpha}{2},}$ is the table value of t at α level of significance and (t-1)(r-1) degrees of freedom.

1.6.6 Advantages of RBD:

1. RBD is the simplest of all block design.
2. RBD uses all the three principles of the design of experiments.
3. RBD takes care of soil heterogeneity.
4. The layout is very simple.
5. It is more efficient compared to CRD.

1.6.7 Disadvantages of RBD:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1. Number of treatments cannot be very large. If number of treatments is very large then it is very difficult to have a greater homogenous block to accommodate all the treatments. In practice, the number of treatments in RBD should not exceed twelve.
2. Like CRD, flexibility of using variable replication for different treatments isnot possible.
3. Missing observation, if any is to be estimated first and then analysis of data tobe taken.
4. It takes care of heterogeneity of experimental area in one direction only.

1.7 Latin Square Design (LSD)

- A Latin square is an arrangement of treatments in such a way that each treatment occurs once and only once in each row and each column.
- This type of allocation of treatments helps in estimating the variation among row blocks as well as column blocks.
- Subsequently the total variations among the experimental units are partitioned into different sources viz, row, column, treatments and errors.
- If t is the number of treatments which is also equal to the number of replications for each treatment then the total number of experimental units needed for this design is t x t. These t^2 units are arranged in t rows and t columns.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1.7.1 Layout of LSD: The layout of the Latinsquare design starts with a standard Latin square.A standard Latin square is an arrangement inwhich the treatments are arranged in natural/alphabetical order or systematically. Then in thenext step, columns are randomized, and in the laststep keeping the first row intact, the rest of therows are randomized. As such we shall get thelayout of Latin square of different orders.

1.7.2 Statistical Model: Let there be *t* treatments, so there should be *t* rows and *t* columns, and we need a field of *t* x *t* =t^2 experimental units.

As such the statistical model and analysis would be as follows:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \qquad i,j,k = 1,2,\dots t$$

where, y_{ijk} = response corresponding to the i^{th} treatment in j^{th} block and k^{th} column

μ = the overall mean

α_i = the i^{th} treatment effect

β_j = the j^{th} row effect

γ_k = the k^{th} column effect

ε_{ij} = the error term for i^{th} treatment in j^{th}block and k^{th} column and are *i.i.d,* $N(0,\sigma^2)$.

1.7.3 Hypotheses and ANOVA table for LSD:

$H_{01} : \alpha_1 = \alpha_2 = \dots\dots\dots\dots = \alpha_t$

$H_{02} : \beta_1 = \beta_2 = \dots\dots\dots\dots = \beta_t$

$H_{03} : \gamma_1 = \gamma_2 = \dots\dots\dots\dots = \gamma_t$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Against the alternative hypotheses

H_{11}: all α's are not equal

H_{12}: all β's are not equal

H_{13}: all γ'sare not equal

The different Sums of Squares (SS) for constructing ANOVA table are

Total Sums of Squares (TSS) = $\sum_{i=1}^{t}\sum_{j=1}^{t}\sum_{k=1}^{t} {y_{ijk}}^2 - C.F.$

where $C.F. = \frac{G^2}{N}$, G = Grand Total and N = Total no. of observations

Treatment SS (TrSS) = $\frac{1}{t}\sum_{i=1}^{t} {y_{i00}}^2 - C.F.$

Row SS (RSS) = $\frac{1}{t}\sum_{j=1}^{t} {y_{0j0}}^2 - C.F.$

Column SS (CSS) = $\frac{1}{t}\sum_{k=1}^{t} {y_{00k}}^2 - C.F.$

Error SS (ESS) = By subtraction = TSS – Trss – RSS – CSS

Table 3: ANOVA table for LSD

Sources of Variation	d.f.	SS	MSS	F
Rows	t – 1	RSS	RMS	RMS/EMS
Columns	t – 1	CSS	CMS	CMS/EMS
Treatments	t – 1	TrSS	TrMS	TrMS/EMS
Error	(t - 1)(t - 2)	ESS	EMS	
Total	$t^2 – 1$	TSS		

1.7.4 Decision Rule: The null hypothesis are rejected at a level of significance if the calculated values of F ratio corresponding to

treatment or row or column be greater than the corresponding table values at the same level of significance with same (t-1), (t-1)(t-2) degrees of freedom respectively.

1.7.5 Least significant difference (LSD) or **Critical Difference**: The formula for calculation of CD value is same for row/column/treatment because all these degrees of freedom are same. The critical difference for rows/columns/treatments at α level of significance is given by

$$\sqrt{\frac{2EMS}{t}} \times t_{(t-1)(t-2),\frac{\alpha}{2}}$$

where $t_{(t-1)(t-2),\frac{\alpha}{2}}$ is the table value of t at α level of significance with (t-1)(t-2) degrees of freedom.

1.7.6 Advantages of LSD:

1. Latin square design is improved design over the other two basic designs CRD and RBD, as it takes care of the heterogeneity or the variations in two perpendicular directions.
2. LSD is more efficient than RBD or CRD. This is because of double grouping that will result in small experimental error.
3. When missing values are present, missing plot technique can be used and analysed.

1.7.7 Disadvantages of LSD:

1. The layout of the design is not so simple as was in the case of CRD or RBD.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

2. The LSD design is heavily disadvantageous because of the fact that number of replications equal to the number of treatments.
3. The analysis assumes that there are no interactions.
4. It requires mainly square shaped plots.

GLOSSARY:

- **Treatment:** The different objects under comparison in a comparative experiment are known as treatment.
- **Experimental unit**: The smallest part/unit of the experimental area in which one applies treatments and from which observations on study variables are recorded is called experimental area.
- **Block:** The concept of block is associated mainly with field experiment. In field experiment the area selected for experimental purpose may not be of homogeneous in nature (with respect to fertility, soil structure, texture and other conditions). Blocking is the technique by virtue of which the whole experimental area is subdivided into number of small parts, each having homogeneous nature.
- **Experimental Error:** The variations in response among the different experimental units may be partitioned into two components –
    - The systematic part/the assignable part and
    - The non-systematic part/non assignable part.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Variations in experimental units due to different in treatments, blocking etc., which are known to the experimenter, constitute the assignable part. On the other hand, the part of variations which cannot be assigned to specific reasons or causes are termed as the experimental error.

- **Precision:** The precision of an experiment is defined as the reciprocal of the variance of mean.
- **Efficiency:** It is the ratio of the variance of the difference between two treatment means in two different experiments.

REFERENCES:

- Fisher RA. (1953). Design and Analysis of Experiments. Oliver & Boyd.
- Nigam A.K.and Gupta V.K. (1979). Handbook on Analysis of Agricultural Experiments. IASRI Publ.
- Sahu P.K. (2016). Applied Statistics for Agriculture, Veterinary,Fishery, Dairy and AlliedFields. Springer, India.

Singh S.P. and Verma R.P.S. Agricultural Statistics for M.Sc. Ag. Classes. Rama Publishing

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson 12

Factorial Experiments

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 12	Factorial Experiments
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

OBJECTIVES:

1. To discuss the basic concepts on Factorial Experiments.
2. To study the statistical analysis of 2^2 factorial experiment.
3. To study the statistical analysis of 2^3 factorial experiment.
4. To study the different methods of computing factorial effect totals.

1.1 Basic Concept of Factorial Experiment

- Factorial experiments involve simultaneously more than one factor each at two or more levels.
- Several factors affect simultaneously the characteristic under study in factorial experiments and the experimenter is interested in the main effects and the interaction effects among different factors.
- Through the factorial experiments, we can study the
    - individual effect of each factor and
    - interaction effect.
- A general notation for representing the factors is to use capital letters, *e.g.*, A, B, C etc. and levels of a factor are represented in small letters.

1.2 Symmetrical and Asymmetrical Factorial Experiment

- If the number of levels of various factors are equal then the factorial experiment is called a symmetrical factorial experiment.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- In general, S^n factorial experiment means an experiment with '*n*' factors, each at '*S*' levels, where '*n*' is any positive integer ≥ 2.
- If the number of levels of the different factors are unequal then the experiment is termed as asymmetrical or mixed factorial experiment.

1.3 2^2 Factorial Experiment

- Two factors, say A and B each at two levels.
- The two levels of A, they are denoted as a_0 and a_1. Similarly, the two levels of B are represented as b_o and b_1. Other alternative representation to indicate the two levels of A is 0 (for a_0) and 1 (for a_1). The factors of B are then 0 (for b_0) and 1 (for b_1).
- Thus, for a 2^2 factorial experiment with factors A and B each at two levels 0 and 1, the four treatment combinations may be represented as a_0b_0, a_1b_0, a_0b_1 and a_1b_1 or (1), a, b and ab.
- Sometimes 0 is referred to as 'low level' and 1 is referred to 'high level'. (1) denote that both factors are at lower levels 00 (or a_0b_0). This is called as control treatment.

1.3.1 Main Effects and Interactions

The simple effect of A at level $b_0 = a_1b_0 – a_0b_0 = a – (1)$

The simple effect of A at level $b_1 = a_1b_1 – a_0b_1 = ab – b$

The simple effect of B at level $a_0 = a_0b_1 – a_0b_0 = b – (1)$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The simple effect of B at level a_1 = $a_1b_1 - a_1b_0 = ab - a$

The main effect of A = average of the simple effects of A.

$$= \frac{1}{2}[a - (1) + ab - b]$$

The main effect of B = $\frac{1}{2}[b - (1) + ab - a]$

The interaction effect of AB = $\frac{1}{2}$[simple effect of A at b_1 – simple effect of A at b_0]

Or, = $\frac{1}{2}$[simple effect of B at a_1 – simple effect of B at a_0].

$$= \frac{1}{2}[ab - a - b + (1)]$$

1.3.2 Model of 2^2 design

If $y_{ijk}$is the response observed at i^{th} level of A, j^{th} level of B in the k^{th}replication, then the linear model for a 2^2 experiment is given as:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \rho_k + e_{ijk}; \quad i,j = 0,1; \quad k = 1,2,\dots,r$$

where μ is the general mean, $\alpha_i$and β_j are the effects of the i^{th} level of A and j^{th} level ofB respectively, $(\alpha\beta)_{ij}$is the interaction effect of i^{th} level of A with j^{th} level of B, ρ_k is the effect due to the k^{th} replication, e_{ijk} is the error effect due to chance factor or random error which is N(0, σ_e^2).

1.3.3 Statistical Analysis

- Hypotheses to be tested are:

  H_{01}: $\alpha_1 = \alpha_2 = 0$ against the alternative hypothesis H_{11}: $\alpha_1 \neq \alpha_2$

  H_{02}: $\beta_1 = \beta_2 = 0$ against the alternative hypothesis H_{12}: $\beta_1 \neq \beta_2$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

H_{03}: $\alpha_1\beta_1 = \alpha_1\beta_2 = \alpha_2\beta_1 = \alpha_2\beta_2 = 0$ against the alternative hypothesis H_{13}: all the interaction effects are not equal.

- For a 2^2 factorial experiment, the various factorial effect totals can be expressed as mutually orthogonal contrasts of the 4 treatment totals.

Thus, $$[A] = [a] - [1] + [ab] - [b]$$

$$[B] = [b] - [1] + [ab] - [a]$$

and $$[AB] = [ab] - [a] - [b] + [1].$$

- The ANOVA can be carried out as given in the following table.

Table 1: ANOVA table for 2^2 factorial experiment

Sources of Variation	d.f.	S.S.	M.S.S.	Variance Ratio (F)
Replication (Blocks)	r – 1	$S_R^2 = \frac{1}{t}\sum_{i=1}^{r} R_i^2 - C.F.$	$s_R^2 = \frac{S_R^2}{r-1}$	$F_R = \frac{s_R^2}{s_E^2}$
Treatments	t – 1 = 3	$S_t^2 = \frac{1}{r}\sum_{j=1}^{t} T_j^2 - C.F.$	$s_t^2 = \frac{S_t^2}{t-1}$	$F_t = \frac{s_t^2}{s_E^2}$
Main Effects:				
A	1	$S_A^2 = \frac{[A]^2}{4r}$	$s_A^2 = S_A^2$	$F_A = \frac{s_A^2}{s_E^2}$
B	1	$S_B^2 = \frac{[B]^2}{4r}$	$s_B^2 = S_B^2$	$F_B = \frac{s_B^2}{s_E^2}$
Interaction Effect				

AB	1	$S_{AB}^2 = \frac{[AB]^2}{4r}$	$s_{AB}^2 = S_{AB}^2$	$F_{AB} = \frac{s_{AB}^2}{s_E^2}$
Error	3(r - 1)	S_E^2 = by subtraction	$s_E^2 = \frac{S_E^2}{3(r-1)}$	-
Total	2^2r – 1	$S_T^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{r} y_{ijk}^2 - C.F.$	-	-

* C.F. = $\frac{G^2}{2^2 r}$, where G is the grand total of all the observations.

1.3.4 Decision Rule

The decision rule to reject the concerned null hypothesis if the corresponding calculated F-statistic in the table is greater than critical value of $F_{\alpha;1,3(r-1)}$ otherwise the hypothesis may be accepted at α% level of significance.

1.3.5 Critical Difference

If the Cal. F > $F_{\alpha;1,3(r-1)}$ corresponding to interaction effect then we need to find out the CD (Critical Difference) value at specified level of significance and error degrees of freedom using the following formula:

$$CD(Main\ effect) = \sqrt{\frac{2MSE}{2r}}\, t_{\alpha,error\ d.f.}$$

$$CD(Interaction\ effect) = \sqrt{\frac{2MSE}{r}}\, t_{\alpha,error\ d.f.}$$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

If the difference between means of any pair of treatment combination is greater than the corresponding CD value, then these two means under comparison are said to be significantly different from each other.

1.4 2^3 Factorial Experiment

- Three factors, say, A, B and C, each at two levels.
- Consider 3 factors, say, A, B, C each at two levels, say (a_0, a_1), (b_0, b_1) and (c_0, c_1) respectively, so that there are $2^3 = 8$ treatment combinations which are given as:$a_0b_0c_0$ or *(1)*, $a_1b_0c_0$ or *a*, $a_0b_1c_0$ or *b*, $a_1b_1c_0$ or *ab*, $a_0b_0c_1$ or *c*, $a_1b_0c_1$ or *ac*, $a_0b_1c_1$ or *bc*, $a_1b_1c_1$ or *abc*.

1.4.1 Main Effects and Interactions

$$A = \frac{1}{4}[(a-1)(b+1)(c+1)]$$

$$= \frac{1}{4}[abc - bc + ac - c + ab - b + a - (1)]$$

$$B = \frac{1}{4}[(a+1)(b-1)(c+1)]$$

$$= \frac{1}{4}[abc + bc - ac - c + ab + b - a - (1)]$$

$$C = \frac{1}{4}[(a+1)(b+1)(c-1)]$$

$$= \frac{1}{4}[abc + bc + ac + c - ab - b - a - (1)]$$

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

$$AB = \frac{1}{4}[(a-1)(b-1)(c+1)]$$

$$= \frac{1}{4}[abc - bc - ac + c + ab - b - a + (1)]$$

$$AC = \frac{1}{4}[(a-1)(b+1)(c-1)]$$

$$= \frac{1}{4}[abc - bc + ac - c - ab + b - a + (1)]$$

$$BC = \frac{1}{4}[(a+1)(b-1)(c-1)]$$

$$= \frac{1}{4}[abc + bc - ac - c - ab - b + a + (1)]$$

$$ABC = \frac{1}{4}[(a-1)(b-1)(c-1)]$$

$$= \frac{1}{4}[abc - bc - ac + c - ab + b + a - (1)]$$

1.4.2 Model of 2^3 Design

If $y_{ijkl}$is the response observed at i^{th} level of A, j^{th} level of B, k^{th}level of C inthe l^{th} replication, then the linear model for a 2^3 experiment is given as:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \rho_l + e_{ijkl};$$

$$i, j, k = 0,1; l = 1,2, \dots, r$$

where μ is the general mean, α_i, β_j and γ_k are the effects of the i^{th} level of A, j^{th} level of B and kth level of C respectively, $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$ and $(\beta\gamma)_{jk}$are the 1st order interaction effects with the different levels of

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

the factors, $(\alpha\beta\gamma)_{ijk}$ is the 2nd order interaction effect and ρ_l is the effect due to the l^{th} replication, e_{ijk} is the error effect due to chance factor or random error which is $N(0, \sigma_e^2)$.

1.4.3 Statistical Analysis

- Hypotheses to be tested:

  H_{01}: $\alpha_1 = \alpha_2$ against H_{11}: $\alpha_1 \neq \alpha_2$

  H_{02}: $\beta_1 = \beta_2$ against H_{12}: $\beta_1 \neq \beta_2$

  H_{03}: $\gamma_1 = \gamma_2$ against H_{13}: $\gamma_1 \neq \gamma_2$

  H_{04}: $\alpha_1\beta_1 = \alpha_1\beta_2 = \alpha_2\beta_1 = \alpha_2\beta_2$ against H_{14}: all $(\alpha\beta)_{ij}$'s are not equal.

  H_{05}: $\alpha_1\gamma_1 = \alpha_1\gamma_2 = \alpha_2\gamma_1 = \alpha_2\gamma_2$ against H_{15}: all $(\alpha\gamma)_{ik}$'s are not equal.

  H_{06}: $\beta_1\gamma_1 = \beta_1\gamma_2 = \beta_2\gamma_1 = \beta_2\gamma_2$ against H_{16}: all $(\beta\gamma)_{jk}$'s are not equal.

  H_{07}: all $(\alpha\beta\gamma)_{ijk}$'s are equal against H_{17}: all $(\alpha\beta\gamma)_{ijk}$'s are not equal.

- The various factorial effect totals can be expressed as mutually orthogonal contrasts of the 8 treatment totals given as:

$$A = [abc] - [bc] + [ac] - [c] + [ab] - [b] + [a] - [1]$$

$$B = [abc] + [bc] - [ac] - [c] + [ab] + [b] - [a] - [1]$$

and so on.

- The ANOVA can be carried out as given in the following table.

Table 2: ANOVA table for 2^3 factorial experiment

Sources of Variation	d.f.	S.S.	M.S.S.	Variance Ratio (F)
Replication (Blocks)	r – 1	$S_R^2 = \frac{1}{t}\sum_{i=1}^{r} R_i^2 - C.F.$	$s_R^2 = \frac{S_R^2}{r-1}$	$F_R = \frac{s_R^2}{s_E^2}$
Treatments	8 – 1 = 7	$S_t^2 = \frac{1}{r}\sum_{j=1}^{t} T_j^2 - C.F.$	$s_t^2 = \frac{S_t^2}{t-1}$	$F_t = \frac{s_t^2}{s_E^2}$
Main Effects:				
A	1	$S_A^2 = \frac{[A]^2}{8r}$	$s_A^2 = S_A^2$	$F_A = \frac{s_A^2}{s_E^2}$
B	1	$S_B^2 = \frac{[B]^2}{8r}$	$s_B^2 = S_B^2$	$F_B = \frac{s_B^2}{s_E^2}$
C	1	$S_C^2 = \frac{[C]^2}{8r}$	$s_C^2 = S_C^2$	$F_C = \frac{s_C^2}{s_E^2}$
1st order Interaction Effects				
AB	1	$S_{AB}^2 = \frac{[AB]^2}{8r}$	$s_{AB}^2 = S_{AB}^2$	$F_{AB} = \frac{s_{AB}^2}{s_E^2}$
AC	1	$S_{AC}^2 = \frac{[AC]^2}{8r}$	$s_{AC}^2 = S_{AC}^2$	$F_{AC} = \frac{s_{AC}^2}{s_E^2}$
BC	1	$S_{BC}^2 = \frac{[BC]^2}{8r}$	$s_{BC}^2 = S_{BC}^2$	$F_{BC} = \frac{s_{BC}^2}{s_E^2}$
2nd order Interaction Effect				
ABC	1	$S_{ABC}^2 = \frac{[ABC]^2}{8r}$	$s_{ABC}^2 = S_{ABC}^2$	$F_{ABC} = \frac{s_{ABC}^2}{s_E^2}$
Error	7(r - 1)	S_E^2 = by subtraction	$s_E^2 = \frac{S_E^2}{7(r-1)}$	-

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Total	$2^2r - 1$	$S_T^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{r} y_{ijk}^2 - C.F.$	-	-

1.4.4 Decision Rule

If the Cal. F > $F_{\alpha;1,7(r-1)}$ then the concerned hypothesis is rejected at α% level of significance.

1.4.5 Critical Difference

For pairwise treatment comparison, the CD values are calculated as

CD (Main Effects) = $\sqrt{\frac{2MSE}{4r}}\, t_{\alpha,error\ d.f.}$

CD (1st order Interaction Effects) = $\sqrt{\frac{2MSE}{2r}}\, t_{\alpha,error\ d.f.}$

CD (2nd order Interaction Effect) = $\sqrt{\frac{2MSE}{r}}\, t_{\alpha,error\ d.f.}$

1.5 Computation of Factorial Effect Totals

- The main effects and interaction effect totals in factorial experiment can be computed by many methods. The commonly used methods are:

1.5.1 Fisher's Algebraic Method:

- In this method a table of divisors and the signs giving the general mean effect M and the $2^n - 1$ factorial effects in terms of treatment means for a 2^n design is prepared.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- In this table, give a plus to the treatment combinations which contain the small letter corresponding to the factorial effects. The signs for interactions can be obtained on multiplying the signs of corresponding factors.
- For e.g., for a 2^3 design, the table of divisors and signs is obtained as:

Table 3: Table of divisors and signs for a 2^3 design

Factorial Effect	(1)	(a)	(b)	(ab)	(c)	(ac)	(bc)	(abc)
M	+	+	+	+	+	+	+	+
A	-	+	-	+	-	+	-	+
B	-	-	+	+	-	-	+	+
AB	+	-	-	+	+	-	-	+
C	-	-	-	-	+	+	+	+
AC	+	-	+	-	-	+	-	+
BC	+	+	-	-	-	-	+	+
ABC	-	+	+	-	+	-	-	+

- The factorial effect totals are given as:

$$A = [abc] - [bc] + [ac] - [c] + [ab] - [b] + [a] - [1]$$

$$B = [abc] + [bc] - [ac] - [c] + [ab] + [b] - [a] - [1]$$

$$AB = [abc] - [bc] - [ac] + [c] + [ab] - [b] - [a] + [1]$$ and so on.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- This method of computation of main effects and interaction effects is known as Fisher's algebraic method. This method can be used for experiments involving several factors at two levels.

1.5.2 Yate's Method (Algebraic method using tabular form):

- Yate's method consists in the following steps:
  - In the first column write the treatment combinations. It is an essential part of the procedure that the treatment combinations be written in a standard systematic order.
  - Against each treatment combinations, write the corresponding total yields from all the replicates.
  - The entries in the third column can be split into two halves. The first half is obtained by writing down in order, the pairwise sums of the values in column 2 and the second half the pairwise differences of the values in second column. It is to be remembered that the first number is to be subtracted from the second number of a pair.
  - To complete the fourth column the whole of the procedure as explained in step 3 is repeated on column 4 and the fifth column is derived from fourth in a similar manner.
- Thus, for a 2^n factorial experiment there will be '*n*' cycles of this 'sum and differences' procedure.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The first term in the last, viz., (n+1)th column always gives the grand total (G) while the others entries in the last column are the totals of the main effects or the interactions corresponding to the treatment combinations in the first column of the table.
- For e.g., for a 2^3 factorial experiment the Yate's method of computing factorial effect totals is given in the following table:

Table 4: Yate's method of computing factorial effect totals for a 2^3 design

Treat. Comb. (Col.1)	Treat. Totals (Col.2)	(Col.3)	(Col.4)	(Col.5)	Effect Totals (Col.6)
(1)	[1]	[1] + [a]=u_1	$u_1+u_2=v_1$	$v_1+v_2=w_1$	G
a	[a]	[b] + [ab]=u_2	$u_3+u_4=v_2$	$v_3+v_4=w_2$	[A]
b	[b]	[c] + [ac]=u_3	$u_5+u_6=v_3$	$v_5+v_6=w_3$	[B]
ab	[ab]	[bc] + [abc]=u_4	$u_7+u_8=v_4$	$v_7+v_8=w_4$	[AB]
c	[c]	[a] – [1]=u_5	$u_2 – u_1=v_5$	$v_2 – v_1=w_5$	[C]
ac	[ac]	[ab] – [b]=u_6	$u_4 – u_3=v_6$	$v_4 – v_3=w_6$	[AC]

bc	[bc]	[ac] – [c]=u_7	u_6 – u_5=v_7	v_6 – v_5=w_7	[BC]
abc	[abc]	[abc] – [bc]=u_8	u_8 – u_7=v_8	v_8 – v_7=w_8	[ABC]

GLOSSARY:

- **Factor**: Factor refers to a set of related treatments.
- **Levels of a factor**: Different states or components making up a factor are known as the levels of that factor.
- **Simple Effect**:Simple effect of a factor is the difference between its responses for a fixed level of other factors.
- **Main Effect**:Main effect is defined as the average of the simple effects.
- **Interaction:**Interaction is defined as the dependence of factors in their responses. Interaction is measured as the mean of the differences between simple effects.
- **Contrast**: A linear combination $\sum_{i=1}^{k} c_i t_i$ of k treatment means t_i(i = *1,2,….,k*) is called a contrast of treatment means if $\sum_{i=1}^{k} c_i = 0$. In other words, contrast is a linear combination of treatment means such that the sum of the co-efficient is zero.
- **Orthogonal Contrast**: Two contrast are said to be orthogonal if sum of product of two co-efficient is zero.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

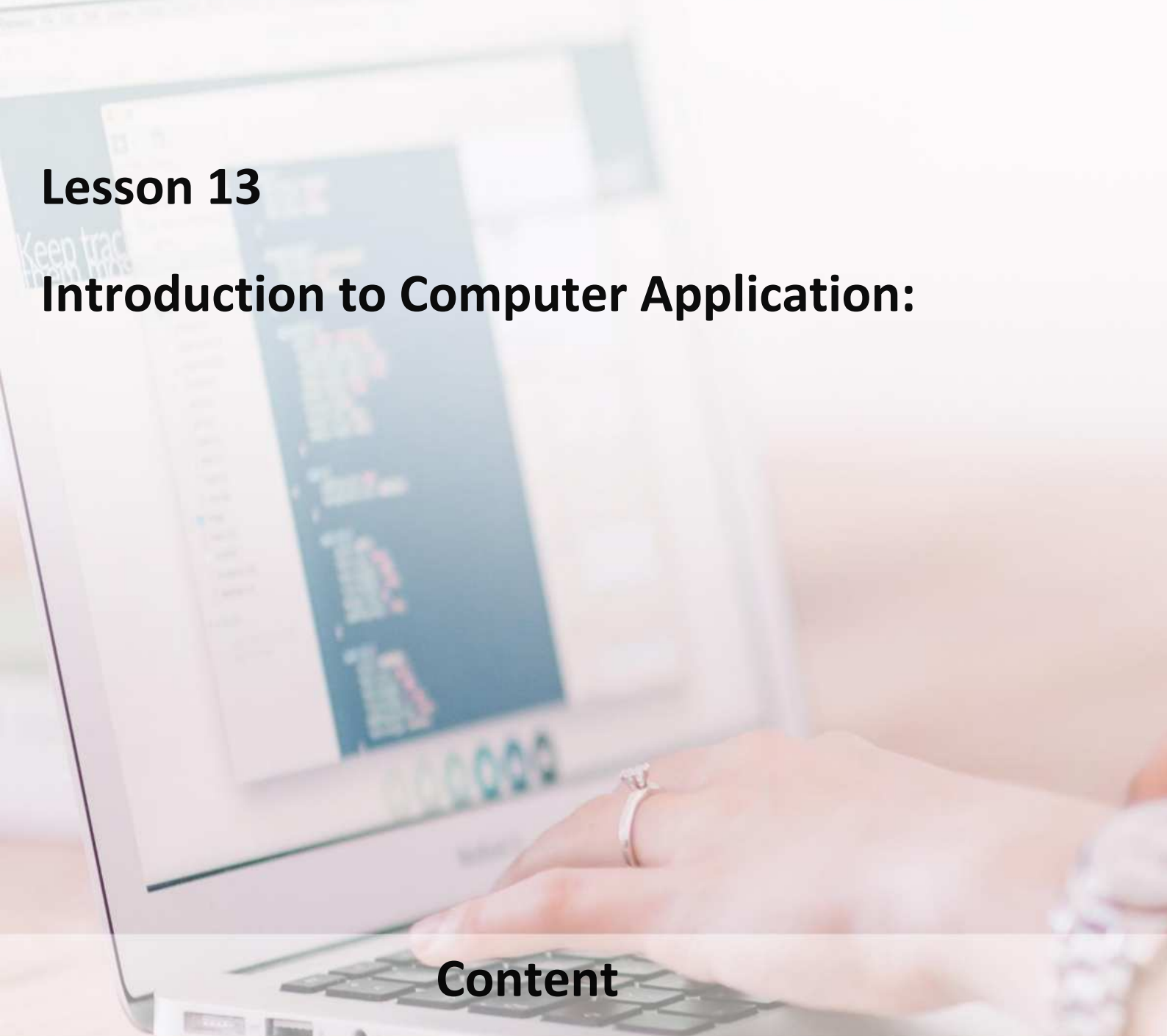
- **Mutually Orthogonal Contrast**: Let there be '*n*' orthogonal contrast. If all are pairwise orthogonal contrasts then these contrasts are said to be mutually orthogonal contrast.

REFERENCES:

- Fisher RA. (1953). Design and Analysis of Experiments. Oliver & Boyd.
- Nigam A.K. and Gupta V.K. (1979). Handbook on Analysis of Agricultural Experiments. IASRI Publ.
- Sahu P.K. (2016). Applied Statistics for Agriculture, Veterinary, Fishery, Dairy and Allied Fields. Springer, India.
- Singh S.P. and Verma R.P.S. Agricultural Statistics for M.Sc. Ag. Classes. Rama Publishing House, Meerut.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application



Lesson 13

Introduction to Computer Application:

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 13	Introduction to Computer Application
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

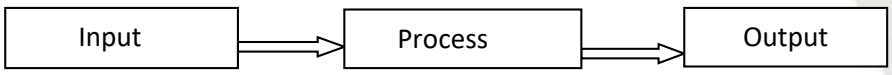
1. To introduce the basics of computer.
2. To explain the characteristics of computer.
3. To introduce with different components of computer with their functions.
4. To explain the various types of memories of computer.

1. Brief introduction to Computers

A **computer** is a high speed electronic device that receives input, stores or processes the input as per user instructions and provides output in desired format. Computers have become an integral part of our lives because they can accomplish easy tasks repeatedly without getting bored and complex ones repeatedly without committing errors.

1.1 Input-Process-Output Model

Computer input is called **data** and the output obtained after processing it, based on user's instructions is called **information**. Raw facts and figures which can be processed using arithmetic and logical operations to obtain information are called **data**.



Work Flow of Computer

1.2 Characteristics of Computer

To understand why computers are such an important part of our lives, let us look at some of its characteristics −

- **Speed :**
  A computer works with much higher speed and accuracy compared to humans while performing mathematical calculations. Computers can process millions (1,000,000) of instructions per second. The time

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

taken by computers for their operations is microseconds and nanoseconds.

- **Accuracy :**
  Computers perform calculations with 100% accuracy. Errors may occur due to wrong data, wrong instructions or data inconsistency or inaccuracy.
- **Diligence:**
  A computer can perform millions of tasks or calculations with the same consistency and accuracy. It doesn't feel any fatigue or lack of concentration. Its memory also makes it superior to that of human beings.
- **Versatility :**
  Versatility refers to the capability of a computer to perform different kinds of works with same accuracy and efficiency.
- **Reliability :**
  A computer is reliable as it gives consistent result for similar set of data i.e., if we give same set of input any number of times, we will get the same result.
- **Storage Capacity:** A computer has built-in memory called primary memory where it stores data. Secondary storage are removable devices such as CDs, pen drives, etc., which are also used to store data. It can store a very large amount of data at a fraction of cost of traditional storage of files. Also, data is safe from normal wear and tear associated with paper.

1.3 Advantages of using Computer

- Computers can do the same task repetitively with same accuracy.
- Computers do not get tired or bored.
- Computers can take up routine tasks while releasing human resource for more intelligent functions.

1.4 Disadvantages of using Computer

- Computers have no intelligence; they follow the instructions blindly without considering the outcome.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Regular electric supply is necessary to make computers work, which could prove difficult everywhere especially in developing nations.

2. Functional units of computer system

A computer consists of three main components *viz.* Input unit, Central Processing Unit (CPU), and Output unit.

Again CPU consists of three more units viz. Memory unit, Control unit and Arthmetic and Logic unit. So, overall any computer system consists of five units. These are shown in Fig. 1 and their functions are explained briefly.

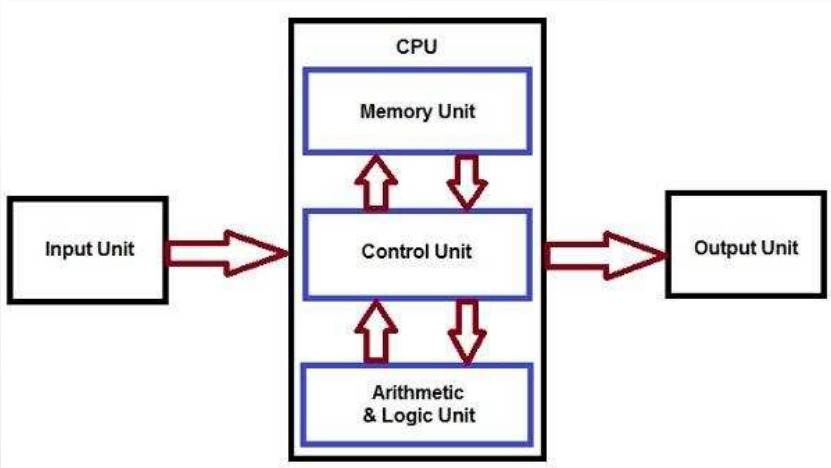


Fig. 1 Block Diagram of Computer

2.1 Input Unit :

Input units are used by the computer to read the data. The most commonly used input devices are keyboards, mouse, joysticks, trackballs, microphones, etc.

However, the most well-known input device is a keyboard. Whenever a key is pressed, the corresponding letter or digit is automatically translated into its corresponding binary code and transmitted over a cable to either the memory or the processor.

2.2 Memory or Storage Unit

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

This unit can store instructions, data, and intermediate results. This unit supplies information to other units of the computer when needed. It is also known as an internal storage unit or the main memory or the primary storage or Random Access Memory (RAM).

- The Memory unit can be referred to as the storage area in which programs are kept which are running, and that contains data needed by the running programs.
- The Memory unit can be categorized in two ways namely, primary memory and secondary memory.
- It enables a processor to access running execution applications and services that are temporarily stored in a specific memory location.
- Primary storage is the fastest memory that operates at electronic speeds. Primary memory contains a large number of semiconductor storage cells, capable of storing a bit of information. The word length of a computer is between 16-64 bits.
- It is also known as the volatile form of memory, means when the computer is shut down, anything contained in RAM is lost.
- Cache memory is also a kind of memory which is used to fetch the data very soon. They are highly coupled with the processor.
- The most common examples of primary memory are RAM and ROM.
- Secondary memory is used when a large amount of data and programs have to be stored for a long-term basis.
- It is also known as the Non-volatile memory form of memory, means the data is stored permanently irrespective of shut down.

2.3 Control unit

This unit controls the operations of all parts of the computer but does not carry out any actual data processing operations. Functions of this unit are:–

- It is responsible for controlling the transfer of data and instructions among other units of a computer.
- It manages and coordinates all the units of the computer.
- It is also known as central nerve of a computer.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- It obtains the instructions from the memory, interprets them, and directs the operation of the computer.
- It communicates with Input/output devices for transfer of data or results from storage.
- It does not process or store data.

2.4 ALU (Arithmetic Logic Unit)

This unit consists of two subsections namely,

- Arithmetic section
- Logic Section

2.4.1 Arithmetic section

- The function of arithmetic section is to perform arithmetic operations like addition, subtraction, multiplication, and division. All complex operations are done by making repetitive use of the above operations.

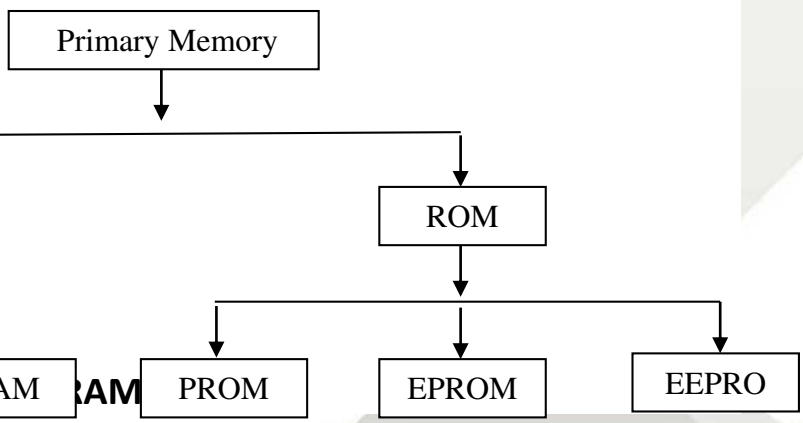
2.4.2 Logic Section

- The function of the logic section is to perform logic operations such as comparing, selecting, matching, and merging of data by using the logical operator AND, OR, NOT etc.

2.5 Types of Memory :

Computer memory is of two basic type – Primary memory or Main memory (RAM and ROM) and Secondary memory (hard drive, CD, etc.).

Random Access Memory (RAM) is primary-volatile memory and Read Only Memory (ROM) is primary-non-volatile memory.



2.5.1 R cess M AM

- It is also called as *read write memory* or the *main memory* or the *primary memory*.
- The programs and data that the CPU requires during execution of a program are stored in this memory.
- It is a volatile memory as the data loses when the power is turned off.
- RAM is further classified into two types- *SRAM (Static Random Access Memory)* and *DRAM (Dynamic Random Access Memory)*.

SRAM VS DRAM

SRAM	DRAM
SRAM has lower access time, so it is faster compared to DRAM.	DRAM has higher access time, so it is slower than SRAM.
SRAM is costlier than DRAM.	DRAM costs less compared to SRAM.
SRAM requires a constant power supply, which means this type of memory which consumes more power.	DRAM offers reduced power consumption because the information is stored in the capacitor.
SRAM has a low packaging density.	DRAM has a high packaging density.

Uses of RAM

Here, are important uses of RAM:

- RAM is utilized in the computer as a scratchpad, buffer, and main memory.
- It offers a fast operating speed.
- It is also popular for its compatibility
- It offers low power dissipation

i) Read Only Memory (ROM)

- It stores crucial information essential to operate the system, like the program essential (mathematical functions) to boot the computer.
- It is not volatile.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Always retains its data.
- Used in embedded systems or where the programming needs no change.
- Used in calculators and peripheral devices.
- ROM is further classified into 4 types- *ROM*, *PROM*, *EPROM*, and *EEPROM*.

Types of Read Only Memory (ROM)

- PROM (Programmable read-only memory) – It can be programmed by user. Once programmed, the data and instructions in it cannot be changed.
- EPROM (Erasable Programmable read only memory) – It can be reprogrammed. To erase data from it, expose it to ultra violet light. To reprogram it, erase all the previous data.
- EEPROM (Electrically erasable programmable read only memory) – The data can be erased by applying electric field, no need of ultra violet light. We can erase only portions of the chip.

2.6 Output Unit

- The primary function of the output unit is to send the processed results to the user. Output devices display information in a way that the user can understand.
- Output devices are pieces of equipment that are used to generate information or any other response processed by the computer. These devices display information that has been held or generated within a computer.
- The most common example of an output device is a monitor.

2.7 Secondary Memory

This memory is permanent in nature. It is used to store the different programs and the information permanently (which were temporarily stored in RAM). It holds the information till we erase it.

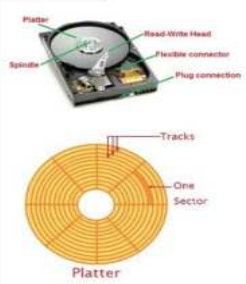
2.7.1 Different types of secondary storage devices are:

Hard Disc, Compact Disc, DVD, Pen Drive, Flash Drive, etc.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Hard Disc

this is the main storage device of the computer which is fixed inside the CPU box. Its storage capacity is very high that varies from 200 GB to 3 TB. As it is fixed inside the CPU box, it is not easy to move the hard disc from one computer to another.



- A hard disc contains a number of metallic discs which are called platters. Information is recorded on the surface of the platters in a series of concentric circles. These circles are called Tracks. For the purpose of addressing information, the surface is considered to be divided into segments called Sectors. This division helps in the proper organization of data on the platter and helps in maximum utilization of the storage space..

Compact Disc (CD)

It is a thin plastic disc coated with metal. Computer can read and write data stored on it. This is an optical storage device with a storage capacity of up to 700 MB and it can store varieties of data like pictures, sounds, movies, texts etc.

CD-ROM

CD-ROM refers to Compact Disc-Read Only Memory. Data or information is recorded at the time of manufacturing and it can only be read. A CD-ROM cannot be used to record fresh data by the computer.

CD-R

CD-R is the short form of Compact Disc-Recordable. Data can be written on it once and can be read whenever required. The data written once cannot be erased.

CD-RW

CD-RW stands for Compact Disc Re-writable. CD-RW can be used to write information over and over again, i.e previous information can be erased and new information can be written on it using a CD writer fixed inside the CPU box.
CDs are slow in comparison to hard discs to read or write the information on them.They are portable storage devices.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

DVD

DVD stands for Digital Versatile Disc. it is an optical storage device which reads data faster than a CD.A single layer, single sided DVD can store data up to 4.7 GB, i.e, around 6 times than that of CD and a double layer DVD can store data up to 17.08 GB ,i.e., around 25 times that of CD. Though DVDs look just like CDs, they can hold much more data, for example, a full length movie.

Flash Drive:

It is an electronic memory device popularly known as pen drive in which data can be stored permanently and erased when not needed. It is a portable storage device that can be easily connected and removed from the CPU to store data in it. Its capacity can vary form 2 GB to 256 GB.

Blue-ray Disc

This is a newly invented optical data storage device whose storage capacity can be form 25 GB up to 200 GB. It is mainly used to store high quality sound and movie data. They are the scratch resistant discs, that's why, storing data on these is much safer than a CD OR DVD.

2.8 Measuring units of Memory

Data in the computer's memory is represented by the two digits 0 and 1.These two digit are called **Binary Digits** or **Bits**. A bit is the smallest unit of computer's memory. To represent each character in memory, a set of 8 binary digits is used. This set of 8 bit is called a Byte.

So, one Byte is used to represent one character of data.
Bits=0,1
1Byte=8bits (e.g,11001011)

To represent a large amount of data in memory, higher data storage units are used like KB (Kilobyte), MB(megabyte), GB(Gigabyte), TB(terabyte), etc.

But all these unites are formed with the set of bytes like,

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1 KB(kilobyte) = 2^{10} Bytes=1024 Bytes
1 MB(megabyte) = 2^{10} KB=1024 KB = 1024*1024 Bytes= 1048576 Bytes
1 GB(Gigabyte) = 2^{10} MB=1024 MB
1 TB(Terabyte) = 2^{10} GB= 1024 GB

GLOSSARY :

Computer: A computer is an electronic device which can produce out put by processing data .

CPU: Central processing unit which consists of three units.

INPUT UNIT: These are used by the computer to read the data. The most commonly used input devices are keyboards, mouse, joysticks, trackballs, microphones, etc.

Control unit: This unit controls the operations of all parts of the computer but does not carry out any actual data processing operations. Also known as the central nerve of computer.

Primary memory: Also known as the main memory of the computer system used for storing internal data. RAM is primary-volatile memory and ROM is primary-non-volatile memory.

Secondary memory: Also known as the external or auxiliary memory of the computer system used for storing data externally. Examples are : Pen Drives, CDs, Magnetic Tapes etc

Input Devices : Used for entering data in the computer. Eg. –Key boards, Scanner, pointing devices etc.

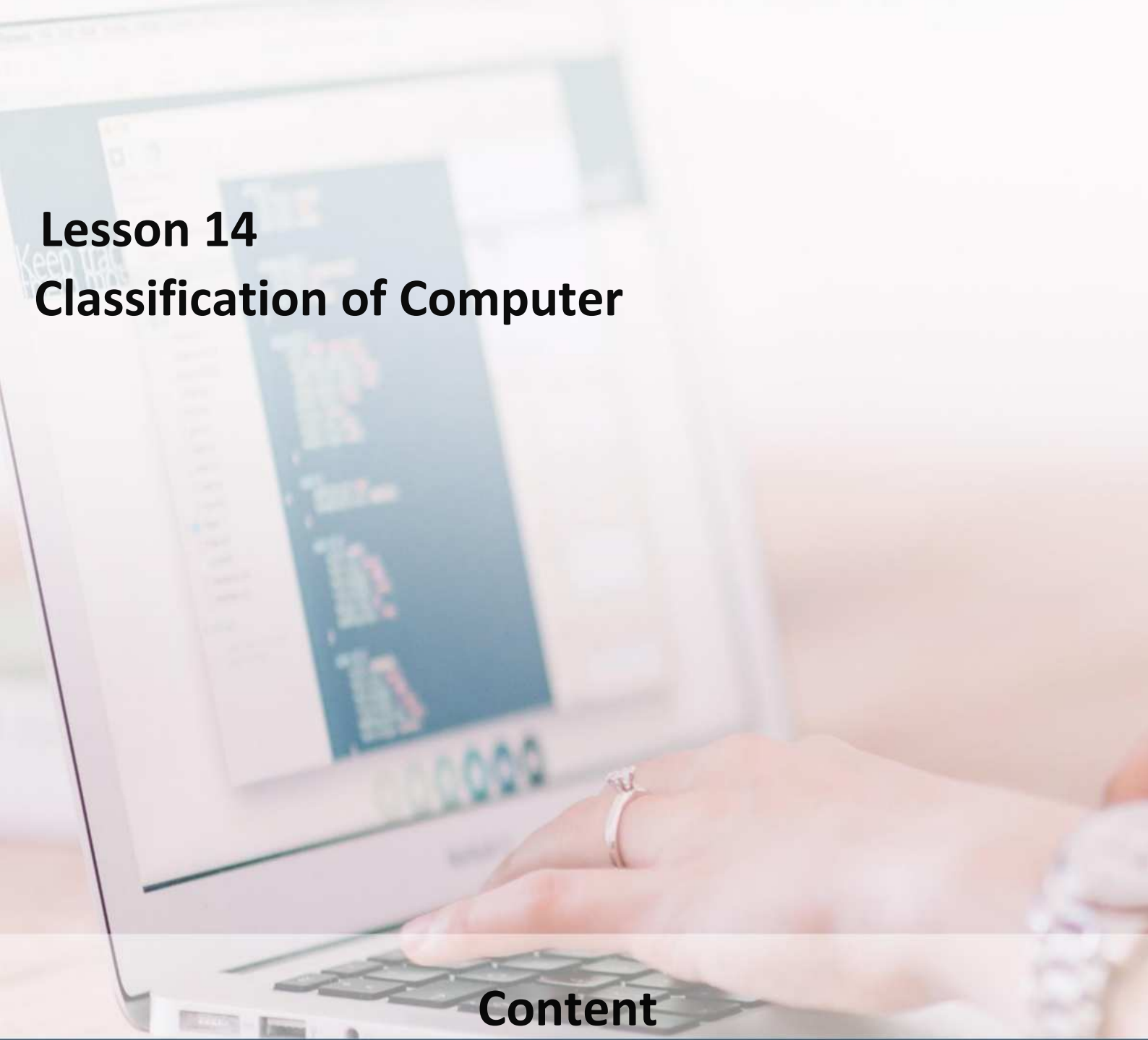
DVD : DVD stands for Digital Versatile Disc. It is an optical storage device which reads data faster than a CD

REFERENCES

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson.
2. Fundamentals of Computers by V Rajaraman.
3. https://tutorialpoints.com/
4. Fundamentals of Computer Science by N Subramaniam.

Elementary Statistics and Computer Application



Lesson 14
Classification of Computer

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 14	Classification of Computer
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. To introduce the classification of computers on based on different factors.
2. To introduce the different types of software of computer.
3. To introduce the different types of hardware of computer.

1. Classification of Computer:

Broadly, computers can classify based on two factors -

(a) **Based on the type of input they accept and**

(b) Size, in terms of capacities and speed of operation.

These are discussed below

1.1 Based on the type of input they accept, the computer is of three types:

Analog Computers

Analog computers are used to process analog data. Analog data is of continuous nature and which is not discrete or separate. Such type of data includes temperature, pressure, speed weight, voltage, depth etc. These quantities are continuous and having an infinite variety of values.

It measures continuous changes in some physical quantity e.g. The Speedometer of a car measures speed, the change of temperature is measured by a Thermometer, the weight is measured by Weights machine.

These computers are ideal in situations where data can be accepted directly from measuring instrument without having to convert it into numbers or codes.

Analog computers are widely used for certain specialized engineering and scientific applications, for calculation and measurement of analog quantities. They are frequently used to control process such as those found in oil refinery where flow and temperature measurements are important.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Digital Computers

A Digital Computer, as its name implies, works with digits to represent numerals, letters or other special symbols. Digital Computers operate on inputs which are ON-OFF (1 - 0) type and its output is also in the form of ON-OFF signal.

A digital computer can be used to process numeric as well as non-numeric data. It can perform arithmetic operations like addition, subtraction, multiplication and division and also logical operations. Most of the computers available today are digital computers. The most common examples of digital computers are accounting machines and calculators.

The results of digital computers are more accurate than the results of analog computers. Analog computers are faster than digital. Analog computers lack memory whereas digital computers store information. We can say that digital computers count and analog computers measures.

Hybrid Computers

A hybrid computer is a combination of digital and analog computers. It combines the best features of both types of computers, i.e. it has the speed of analog computer and the memory and accuracy of digital computer. Hybrid computers are used mainly in specialized applications where both kinds of data need to be processed. Therefore, they help the user, to process both continuous and discrete data. For example a petrol pump contains a processor that converts fuel flow measurements into quantity and price values. In hospital Intensive Care Unit (ICU), an analog device is used which measures patient's blood pressure and temperature etc, which are then converted and displayed in the form of digits. Hybrid computers for example are used for scientific calculations, in defence and radar systems.

1.2 Based on size, in terms of capacities and speed of operation they can be classified as:

- Desktop
- Laptop

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Tablet
- Server
- Mainframe
- Supercomputer

Desktop

Desktop computers are **personal computers (PCs)** designed for use by an individual at a fixed location. IBM was the first computer to introduce and popularize use of desktops. A desktop unit typically has a CPU (Central Processing Unit), monitor, keyboard and mouse. Introduction of desktops popularized use of computers among common people as it was compact and affordable.

Laptop

Despite its huge popularity, desktops gave way to a more compact and portable personal computer called laptop in 2000s. Laptops are also called **notebook computers** or simply **notebooks**. Laptops run using batteries and connect to networks using Wi-Fi (Wireless Fidelity) chips. They also have chips for energy efficiency so that they can conserve power whenever possible and have a longer life.

Modern laptops have enough processing power and storage capacity to be used for all office work, website designing, software development and even audio/video editing.

Tablet

After laptops computers were further miniaturized to develop machines that have processing power of a desktop but are small enough to be held in one's palm. Tablets have touch sensitive screen of typically 5 to 10 inches where one finger is used to touch icons and invoke applications.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Keyboard is also displayed virtually whenever required and used with touch strokes. Applications that run on tablets are called **apps**. They use operating systems by Microsoft (Windows 8 and later versions) or Google (Android). Apple computers have developed their own tablet called **iPad** which uses a proprietary OS called **iOS**.

Server

Servers are computers with high processing speeds that provide one or more services to other systems on the **network**. They may or may not have screens attached to them. A group of computers or digital devices connected together to share resources is called a **network**.

Servers have high processing powers and can handle multiple requests simultaneously. Most commonly found servers on networks include −

- File or storage server
- Game server
- Application server
- Database server
- Mail server
- Print server

Mainframe

Mainframes are computers used by organizations like banks, airlines and railways to handle millions and trillions of online transactions per second. Important features of mainframes are-

- Big in size
- Hundreds times Faster than servers, typically hundred megabytes per second
- Very expensive
- Use proprietary OS provided by the manufacturers
- In-built hardware, software and firmware security features

Supercomputer

Supercomputers are the fastest computers on Earth. Generally these computers consist of number of CPUs which work in parallel. They are used for carrying out complex, fast and time intensive calculations for scientific and engineering applications. Supercomputer speed or performance is measured in teraflops, i.e. 1012 floating point operations per second.

Fugaku, the Japanese computing cluster is the world's fastest supercomputer with a rating of maximum sustained performance level of 442,010 teraflops per second on the Linpack benchmark. Chinese supercomputer Sunway TaihuLight is 93 petaflops per second, i.e. 93 quadrillion floating point operations per second. It has been announced that AMD has joined the US Department of Energy (DOE), Oak Ridge National Laboratory (ORNL) and Cray, in announcing what is expected to be the **world's fastest** exascale-class **supercomputer**, scheduled to be delivered to ORNL in **2021**, aiming to deliver what is expected to be more than 1.5 exaflops of **expected processing performance.**

Most common uses of supercomputers include −

- Molecular mapping and research
- Weather forecasting
- Environmental research
- Oil and gas exploration
- Data mining

2. Software

Software is a set of programs, which is designed to perform a well-defined function. A program is a sequence of instructions written to solve a particular problem.

There are three types of software −

- System Software

- Application Software
- Utility Software

2.1 System Software

The system software is a collection of programs designed to operate, control, and extend the processing capabilities of the computer itself. System software is generally prepared by the computer manufacturers. These software products comprise of programs written in low-level languages, which interact with the hardware at a very basic level. System software serves as the interface between the hardware and the end users.

Some examples of system software are Operating System, Compilers, Interpreter, Assemblers, etc.

Features of a system software −

- Close to the system
- Fast in speed
- Difficult to design
- Difficult to understand
- Less interactive
- Smaller in size
- Difficult to manipulate
- Generally written in low-level language

Based on its function, system software is of three types −

- Operating System
- Language Processor
- Device Drivers

Operating System

System software that is responsible for functioning of all hardware parts and their interoperability to carry out tasks successfully is called operating system (OS). OS is the first software to be loaded into computer memory when the computer is switched on and this is called **booting**.

Language Processor

An important function of system software is to convert all user instructions into machine understandable language. When we talk of human machine interactions, languages are of three types −

- **Machine-level language** − This language is nothing but a string of 0s and 1s that the machines can understand. It is completely machine dependent.
- **Assembly-level language** − This language introduces a layer of abstraction by defining **mnemonics**. **Mnemonics** are English like words or symbols used to denote a long string of 0s and 1s. The complete **instruction** will also tell the memory address. Assembly level language is **machine dependent**.
- **High level language** − This language uses English like statements and is completely independent of machines. Programs written using high level languages are easy to create, read and understand.

Program written in high level programming languages like Java, C++, etc. is called **source code**. Set of instructions in machine readable form is called **object code** or **machine code**. **System software** that converts source code to object code is called **language processor**. There are three types of language interpreters−

- **Assembler** − Converts assembly level program into machine level program.
- **Interpreter** − Converts high level programs into machine level program line by line.
- **Compiler** − Converts high level programs into machine level programs at one go rather than line by line.

Device Drivers

System software that controls and monitors functioning of a specific device on computer is called **device driver**. Each device like printer, scanner, microphone, speaker, etc. that needs to be attached externally to the system has a specific driver associated with it. When we attach a new device, we need to install its driver so that the OS knows how it needs to be managed.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

2.2 Application Software

Application software products are designed to satisfy a particular need of a particular environment. All software applications prepared in the computer lab can come under the category of Application software.

Application software may consist of a single program, such as Microsoft's notepad for writing and editing a simple text. It may also consist of a collection of programs, often called a software package, which work together to accomplish a task, such as a spreadsheet package.

Examples of Application software are the following −

- Payroll Software
- Student Record Software
- Inventory Management Software
- Income Tax Software
- Railways Reservation Software
- Microsoft Office Suite Software
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

2.3 Utility Software

Application software that assists system software in doing their work is called **utility software**. Thus utility software is actually a cross between system software and application software. Examples of utility software include −

- Antivirus software
- Disk management tools
- File management tools
- Compression tools
- Backup tools

3. Hardware:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Hardware represents the physical and tangible components of a computer, i.e. the components that can be seen and touched.

Examples of Hardware are the following −

- **Input devices** − keyboard, mouse, etc.
- **Output devices** − printer, monitor, etc.
- **Secondary storage devices** − Hard disk, CD, DVD, etc.
- **Internal components** − CPU, motherboard, RAM, etc.

3.1 Relationship between Hardware and Software:

- Hardware and software are mutually dependent on each other. Both of them must work together to make a computer produce a useful output.
- Software cannot be utilized without supporting hardware.
- Hardware without a set of programs to operate upon cannot be utilized and is useless.
- To get a particular job done on the computer, relevant software should be loaded into the hardware.
- Hardware is a one-time expense.
- Software development is very expensive and is a continuing expense.
- Different software applications can be loaded on hardware to run different jobs.
- Software acts as an interface between the user and the hardware.
- If the hardware is the 'heart' of a computer system, then the software is its 'soul'. Both are complementary to each other.

Glossary

Hardware: Physical components of computer system.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Software: It is a set of programs, which is designed to perform a well-defined function.

Program: A program is a sequence of instructions written to solve a particular problem.

Continuous Data: Data obtained by measurement are called continuous data. Examples are, temperature measured by a thermometer, speed of car shown by speedometer.

Discrete Data: Data obtained by counting are called continuous data. Examples are, number of students in a class, number mangoes in a basket etc..

Hybrid Computers: A hybrid computer is a combination of digital and analog computers. It combines the best features of both types of computers.

Servers: Servers are computers with high processing speeds that provide one or more services to other systems on the network.

System software: The system software is a collection of programs designed to operate, control, and extend the processing capabilities of the computer itself.

Operating System: System software that is responsible for functioning of all hardware parts and their interoperability to carry out tasks successfully is called operating system.

REFERENCES

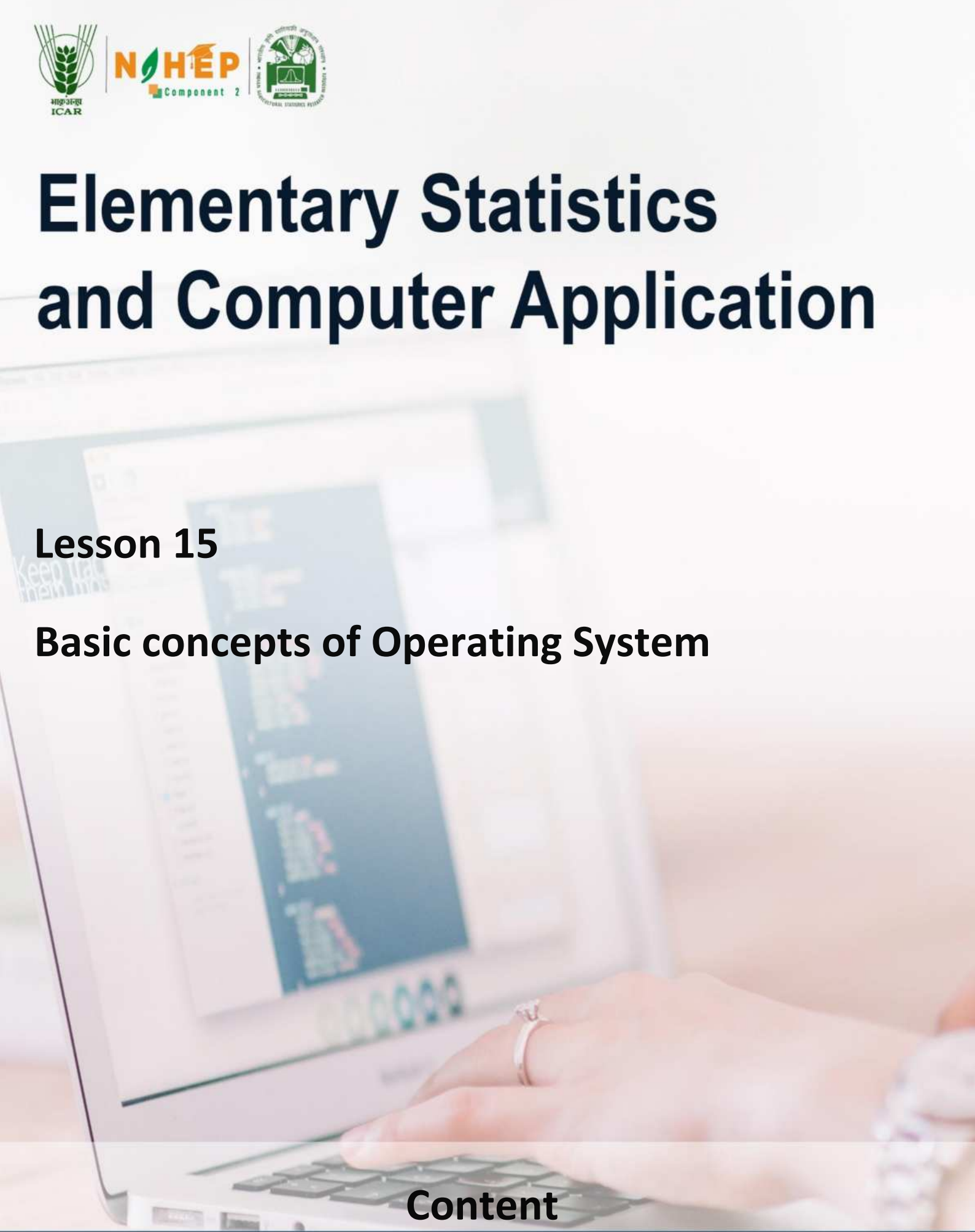
1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson.
2. Fundamentals of Computers by V Rajaraman.
3. https://tutorialpoints.com/
4. Fundamentals of Computer Science by N Subramaniam.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson 15

Basic concepts of Operating System



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 15	Basic concepts of Operating System
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. To introduce the with basic concepts of Operating system.
2. To classify the different types of Operating system
3. To explain the functions of different kinds of OS.

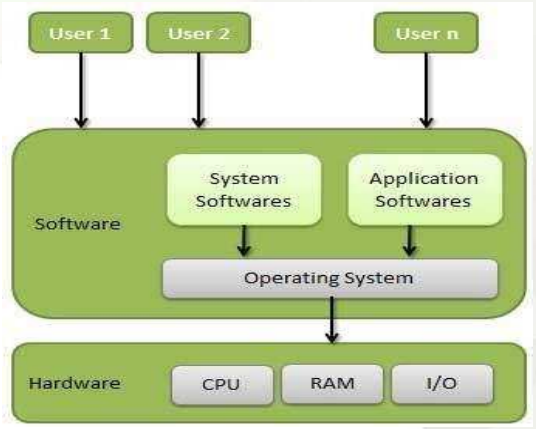
Basic concepts of Operating System (OS):

An Operating System (OS) is an interface between computer user and computer hardware. An operating system is software which performs all the basic tasks like file management, memory management, process management, handling input and output, and controlling peripheral devices such as disk drives and printers.

Some popular Operating Systems are MS-DOS, Linux Operating System, Windows Operating System, UNIX, VMS, OS/400, AIX, z/OS, etc.

1. Definition

An operating system is a program that acts as an interface between the user and the computer hardware and controls the execution of all kinds of programs. The following figure describes about OS.



1.2 Types of OS:

Operating systems are there from the very first computer generation and they keep evolving with time. In this chapter, we will discuss some of the important types of operating systems which are most commonly used.

1.2.1 Batch operating system

This type of OS was first to evolve. It allows one program to run at a time. These kind of OS can still be found on some mainframe computer running

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

batches of jobs. Batch processing OS works on a series of program that are held in a queue. The OS is responsible scheduling the jobs according priority and the resources required.

The problems with Batch Systems are as follows :−

- Lack of interaction between the user and the job.
- CPU is often idle, because the speed of the mechanical I/O devices is slower than the CPU.
- Difficult to provide the desired priority.

1.2.2 Multiuser or Time-sharing operating systems:

Time-sharing is a technique which enables many people, located at various terminals, to use a particular computer system at the same time. This system is used in computer network which allow different users to access same data and application programs on the same network. Multi user OS build a user database account which defends the right that user can have on a particular recourse of the system.

1.2.2.1 Advantages of Timesharing operating systems are as follows –

- Provides the advantage of quick response.
- Avoids duplication of software.
- Reduces CPU idle time.

1.2.2.2 Disadvantages of Time-sharing operating systems are as follows –

- Problem of reliability.
- Question of security and integrity of user programs and data.
- Problem of data communication.

1.2.3 Multiprogramming OS:

In this OS, more than one process (task) can be executed concurrently. The process is switched rapidly between the processes. Hence, a user can have more than one process running at a time. For example, a user on his computer can gave a word processor and an audio CD player running at the same time.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1.2.4 Distributed operating System

Distributed systems use multiple central processors to serve multiple real-time applications and multiple users. Data processing jobs are distributed among the processors accordingly.

The processors communicate with one another through various communication lines (such as high-speed buses or telephone lines). These are referred as **loosely coupled systems** or distributed systems. Processors in a distributed system may vary in size and function. These processors are referred as sites, nodes, computers, and so on.

1.2.4.1 Advantages of distributed systems:

- With resource sharing facility, a user at one site may be able to use the resources available at another.
- Speedup the exchange of data with one another via electronic mail.
- If one site fails in a distributed system, the remaining sites can potentially continue operating.
- Better service to the customers.
- Reduction of the load on the host computer.
- Reduction of delays in data processing.

1.2.5 Network operating System

A Network Operating System runs on a server and provides the server the capability to manage data, users, groups, security, applications, and other networking functions. The primary purpose of the network operating system is to allow shared file and printer access among multiple computers in a network, typically a local area network (LAN), a private network or to other networks.

Examples of network operating systems include Microsoft Windows Server 2003, Microsoft Windows Server 2008, UNIX, Linux, Mac OS X, Novell NetWare, and BSD.

1.2.5.1 Advantages of network operating systems

- Centralized servers are highly stable.
- Security is server managed.
- Upgrades to new technologies and hardware can be easily integrated into the system.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Remote access to servers is possible from different locations and types of systems.

1.2.5.2 Disadvantages of network operating systems

- High cost of buying and running a server.
- Dependency on a central location for most operations.
- Regular maintenance and updates are required.

1.2.6 Real Time operating System

A real-time system is defined as a data processing system in which the time interval required to process and respond to inputs is so small that it controls the environment. The time taken by the system to respond to an input and display of required updated information is termed as the **response time**. So in this method, the response time is very less as compared to online processing.

Real-time systems are used when there are rigid time requirements on the operation of a processor or the flow of data and real-time systems can be used as a control device in a dedicated application. A real-time operating system must have well-defined, fixed time constraints, otherwise the system will fail. For example, Scientific experiments, medical imaging systems, industrial control systems, weapon systems, robots, air traffic control systems, etc.

2. Functions of an operating System

- Memory Management
- Processor Management
- Device Management
- File Management
- Security
- Control over system performance
- Job accounting
- Error detecting aids
- Coordination between other software and users

2.1 Memory Management

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Memory management refers to management of Primary Memory or Main Memory. Main memory is a large array of words or bytes where each word or byte has its own address.

Main memory provides a fast storage that can be accessed directly by the CPU. For a program to be executed, it must in the main memory. An Operating System does the following activities for memory management –

- Keeps tracks of primary memory, i.e., what part of it are in use by whom, what part are not in use.
- In multiprogramming, the OS decides which process will get memory when and how much.
- Allocates the memory when a process requests it to do so.
- De-allocates the memory when a process no longer needs it or has been terminated.

2.2 Processor Management

In multiprogramming environment, the OS decides which process gets the processor when and for how much time. This function is called **process scheduling**. An Operating System does the following activities for processor management −

- Keeps tracks of processor and status of process. The program responsible for this task is known as **traffic controller**.
- Allocates the processor (CPU) to a process.
- De-allocates processor when a process is no longer required.

2.3 Device Management

An Operating System manages device communication via their respective drivers. It does the following activities for device management −

- Keeps tracks of all devices. Program responsible for this task is known as the **I/O controller**.
- Decides which process gets the device when and for how much time.
- Allocates the device in the efficient way.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- De-allocates devices.

2.4 File Management

A file system is normally organized into directories for easy navigation and usage. These directories may contain files and other directions.

An Operating System does the following activities for file management −

- Keeps track of information, location, uses, status etc. The collective facilities are often known as **file system**.
- Decides who gets the resources.
- Allocates the resources.
- De-allocates the resources.

2.5 Other Important Activities

Following are some of the important activities that an Operating System performs −

- **Security** − By means of password and similar other techniques, it prevents unauthorized access to programs and data.
- **Control over system performance** − Recording delays between request for a service and response from the system.
- **Job accounting** − Keeping track of time and resources used by various jobs and users.
- **Error detecting aids** − Production of dumps, traces, error messages, and other debugging and error detecting aids.
- **Coordination between other softwares and users** − Coordination and assignment of compilers, interpreters, assemblers and other software to the various users of the computer systems.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Glossary :

Operating System: An operating system is a program that acts as an interface between the user and the computer hardware and controls the execution of all kinds of programs. The following figure describes about OS.

System Software : It is kind of software which controls the all the operation of a computer. Eg. Any type of OS.

Batch operating system Batch processing OS works on a series of program that are held in a queue. The OS is responsible scheduling the jobs according priority and the resources required.

Multiuser OS : Many users can work at a time on this type of OS**.**

Memory Management: Memory management refers to management of Primary Memory or Main Memory. Main memory is a large array of words or bytes where each word or byte has its own address

Security – By means of password and similar other techniques, it prevents unauthorized access to programs and data.

Control over system performance – Recording delays between request for a service and response from the system.

Job accounting – Keeping track of time and resources used by various jobs and users.

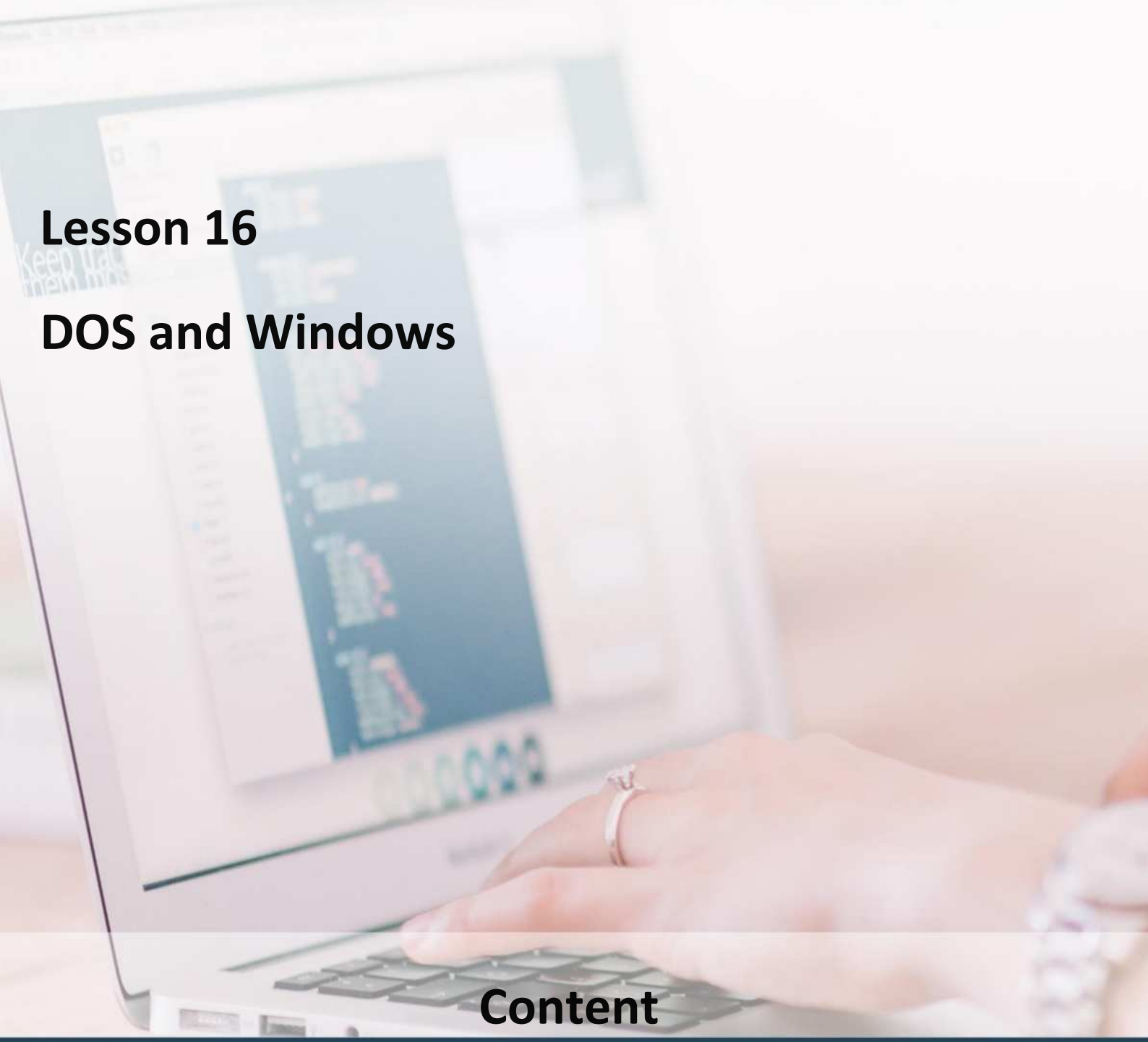
REFERENCES

1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson.
2. Fundamentals of Computers by V Rajaraman.
3. https://tutorialpoints.com/
4. Fundamentals of Computer Science by N Subramaniam.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application



Lesson 16

DOS and Windows

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 16	DOS and Windows
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

1. To introduce with the basic concepts of DOS.
2. To introduce with the basic of Windows
3. To explain the various command of DOS and Windows

1. DOS (Disk Operating System)

1.1 Introduction of DOS

DOS is one of the oldest and widely used operating system. Any operating system which runs through the hard disk drive is termed as Disk Operating System (D.O.S). It was a Command Line Interface (CLI) OS that revolutionized the PC market. DOS was difficult to use because of its interface. The users needed to remember instructions to do their tasks.

It was the first operating system used by IBM-compatible computers. It was first available in two different versions that were typically the same, but marketed and controlled under two different brands. MS-DOS was the framework behind Windows operating systems until Windows XP.

“PC-DOS” was the version of DOS developed by IBM and sold to the first IBM-compatible manufactured computers. “MS-DOS” was the version of dos that Microsoft bought the rights and patents, and was merged with the first versions of Windows.

Command line was used by DOS, or text-based interface, that typed command allowed by the users. By giving simple instructions such as pwd (print working directory) and cd (change directory), the user can open files or run the program or browse the files on the hard drive.

Written originally by Tim Patterson (considered as the father of DOS) and owned by Seattle Computer Products, Microsoft takes over 86-DOS for $75,000, licensed the same software and released it with an IBM PC as MS-DOS 1.0 in 1982 with IBM and Microsoft joint venture.

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1.2 Features of DOS

Following are the significant features of DOS −

- It is a 16-bit operating system
- The mouse cannot be used to operate it, Input in it is through basic system commands.
- The maximum space available is 2 GB.
- It is a free OS.
- It uses a text-based interface and requires text and codes to operate
- It does not support graphical interface
- It is a single user operating system.
- It is a Character Based interface system.
- It is very helpful in making file management e.g., creating, editing, deleting files, etc.
- It is machine independence.
- It manages (computer) files, input and output units and computer memory.

1.3 List of DOS made from 1981 – 1998 are as follows:

1. IBM PC DOS – 1981
2. DR-DOS – 1988
3. ROM-DOS – 1989
4. PTS-DOS – 1993
5. FREE-DOS – 1998

It was rebranded version under the title IBM PC DOS, both of which came in the year 1981. DOS other than Microsoft in the market are:

1. Apple DOS
2. Apple Pro DOS
3. Atari DOS
4. Commodore DOS
5. TRSDOS

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

6. Amiga DOS

1.4 Advantages and Disadvantages of DOS

3.1.3.1 Advantages

- Direct access is can be made to the Basic Input Output System (BIOS) and its underlying hardware.
- Due to its size, it will “boot” much faster than any windows version, thus it will run in a smaller system.
- It is very lightweight so it does not have the overhead of the multitasking operating system.
- It is good for making workarounds for managing/administering an MS system, and for combining programs.

1.5 Disadvantages

- Difficulty in memory access when addressing more than 640 MB of RAM.
- Interrupt levels for hardware needs to be managed by our self.
- Automatic IRQ ordering is not supported by the OS.

1.6 Files, Directory, Drives of DOS

1.6.1 MS-DOS Files and Filenames

File: One of the primary functions of the OS is to handle disk files. A file is a set of information stored on a disk under a specific name. A file can contain only data or it can contain a set of instructions, called a program, telling the computer how to perform a particular task.

Naming a file : In MS-DOS the file name follows 8dot3 format and is divided into two parts – primary name and secondary name. The primary name is up to 8 characters long and the secondary name is up to 3 characters and both the names are separated by a dot (.).

For example, in the file-name Logo.jpg, Logo is the primary name and .jpg is the secondary name. Secondary names are fixed for a particular type of file, meaning for system files the secondary name is designated as .sys, for text files it is .txt and so on.

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

MS-DOS allows the following characters to be used in a filename and extension:

- uppercase and lowercase case letters A through Z
- numbers 0 through 9
- special characters $ # & @ ( ) ! ^ ` ~ { }

To name a file or directory special characters like < > , . / * ? | & Space are not allowed. Here is the list of some of the important types of files with their default secondary names:

Text file	.txt	Database file	.dbm
Command file	.com	Library file	.lib
System file	.sys	Batch file	.bat
Program file	.prg	Executable file	.exe

2. DOS Commands

DOS commands are divided into 2 types: Internal Commands and External Commands.

2.1 Internal Commands

Internal Commands are built into the operating system as the part of a file called CMMAND.COM. It is a command processor, which works as an interface between the user and DOS. It basically interprets what user has typed at the DOS prompt and processes them.

When an Internal Command is typed, MS-DOS will perform it immediately. All of the internal commands are part of the shell which could be command.com or cmd.exe (depending on version of MS-DOS or Windows) and are not separate files on the hard drive. As long as we can open a command line we can run any of the internal commands included with the version of MS-DOS

Example Of MS-DOS Internal Command Are:

1. CLS – It is a command that allows to clear the complete contents of the screen and leave only a prompt.

2. BREAK – Break can be used to enable or disable the braking capability of the computer.
3. REN – It is used to rename files and directories from the original name to a new name.
4. CHDIR – Chdir (change directory) is a command used to switch directories in MS-DOS.
5. EXIT – The exit command is used to withdrawal from the currently running application and the MS-DOS session.
6. RMDIR – Removes an empty directory in MS-DOS.
7. DEL- Del is a command used to delete files from the computer.
8. COPY – Allows to copy one or more files to an alternate location.
9. VOL – Displays the volume of information about the designated drive.
10. TYPE- Display the contents of a text file.

2.2 External Commands

These external commands are for performing advanced tasks and they do need some external file support as they are not stored in COMMAND.COM. There are also batch commands or batch files which are text files that contain a list of internal and/or external commands which are executed in sequence when the batch file is executed. AUTOEXEC.BAT gets executed automatically on booting.

Examples of External Commands are:-

1. DELTREE- Short for delete tree, deltree is a command used to delete files and directories permanently from the computer.
2. TREE- Allows the user to view a listing of files and folders in an easy to read the listing.
3. PRINT – The print command allows users to print a text file to a line printer, in the background.
4. FIND – Allows to search for text within a file.
5. XCOPY – Xcopy is a powerful version of the copy command with additional features; has the capability of moving files, directories, and even whole drives from one location to another.
6. DISK COMP- Compares the contents of a floppy disk in the source drive to the contents of a floppy disk in the target drive.
7. FORMAT – Format is used to erase information off of a computer diskette or fixed drive.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

3.3 Drives and directories. Drives and directories are places where we can put files. Directories are contained within drives, and a drive can have many directories. A drive is denoted by letter followed by a colon, as in "**a:**" or "**c:**". Typically (but not always), drive letters are assigned as follows:
a: or **b:** are used to denote floppy disk drive

c: is always used to denote hard disk

Drives are locations where we can be. To go to given drive, type its name at the DOS prompt and press enter. The drive we are currently "in" is called the current or default drive.
Directories are locations within drives. Directories may themselves be hierarchically nested within each other. The top level directory on any drive is called "the root". All other directories are subdirectories of the root. The root is denoted by the drive letter and colon followed by a backslash ("\"), as in **c:**

Path : Every file has path starting from root through subdirectories reaching a file.

Path of file paint.exe is c:\Programs\Accessories\paint.exe

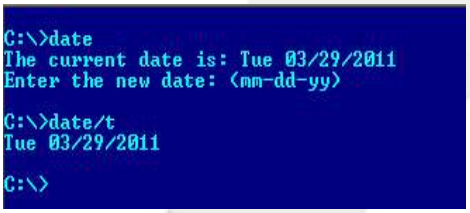
We can execute DOS command on Windows OS by selecting the following path. (This varies as per Windows OS version)

Start → Programs → Accessories Command Prompt

It will display command prompt as C:\> by default. The DOS commands can be typed at this prompt.

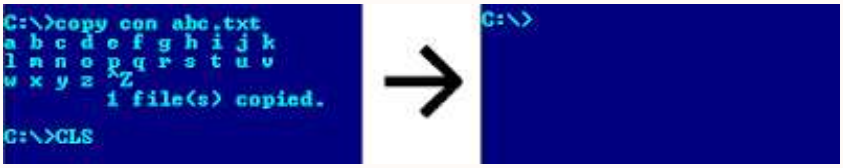
In higher version of windows we can move to command prompt by clicking at start button and typing “cmd” at search option.

For example, to see the today’s date, if we type date and press enter key, today’s date will be display as shown in the figure.



CLS It is used to clear the screen. Syntax is **CLS**

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute



Dir It displays the list of directories and files on the screen.
Syntax:- C : / > dir

a. C : / > dir/p – It displays the list of directories or files page wise
b. C: / > dir/w- It displays the list of directories or files width wise
c. C : / > dir/d: –It display list of directories or files in drive D
d. C : / > dir filename **.** extension – It displays the information of specified file.
e. C : / > dir file name with wild cards.

Wild cards It is the set of special characters. Wild are used with some commonly used DOS commands there are two types of wild cards.

1. Asterisk (*)
2. Question mark (?)

1. Asterisk:- (*) The wild word will match all characters.
C : / > dir *.* will display list of all files and directories.
2. C : / > dir R*.* will display all files stored with first character R.

TYPE
This command allows the user to see the contents of a file.

Syntax :- C :/ >Type path
Eg: C:/ > Type D:/> ramu
It will show the content of “ramu” file on drive D

REN
The purpose of this command is to rename the old file name with new file name.

Syntax : - C : / > ren old filename newf ilename
C : / > ren ramu somu

DEL

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The purpose of this command is to delete file. The user can also delete multiple files by busing this command and long with while cards.

Syntax : - C : / > Del file name **.** extension
C : / > Del ramu
C : Del x **.** prg.

MD

The purpose of this command is to create a new directly or sub directly i.e sub ordinate to the currently logged directly.

Syntax : - C : /> MD directory
C : /> MD sub directory

Ex : C : / > MD college
Now user wants to create a sub directory first year in college directory then
C : / > cd college
C : / > college > Md first year

CD

The purpose of this command is to change from one director to another directory or sub – directory.

Syntax : - C : / > CD directory name

Ex : C: / > cd college
C : / > college > CD first year
C : / > college > first year >
If the user wants to move to the parents directory then use CD command as
C : / > college > first year > cd ..
C : / > college >

RD

The purpose of this command is to remove a directory or sub directory. If the user wants to remove a directory or sub – directory then first delete all the files in the sub – directory and then remove sub directory and remove empty main directory.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

COPY

The purpose of this command is to copy one or more specified files to another disk with same file name or with different file name.

Syntax : - C : / > copy source file target file

Some more examples of External commands

These commands are not permanent part of the memory. To execute or run this commands an external file is required. These commands are situated on hard disk under a particular file. The absence of external commands does affect the OS.

Example : [.] Dot exe, bat.

Some commonly used DOS external commands are .

CHKDSK

The command CHSDK returns the configuration status of the selected disk. It returns the information about the volume, serial number, total disk space, space in directories, space in each allocation unit, total memory and free memory.

Syntax : - C : / > CHKDSK drive name

Eg:- C : / > CHKDSK e :

If drive name is not mentioned by default current drive is considered.

DISKCOPY

Disk copy command is used to make duplicate copy of the disk like xerox copy. It first formats the target disk and then copies the files by collection from the source disk and copied to the target disk.

Syntax : - C : / > disk copy < source path > < destination path >

Ex: - c : / > diskcopy A : B :

NOTE: - This command is used after diskcopy command to ensure that disk is copied successfully.

FORMAT

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Format is used to erase information off of a computer diskette or fixed drive.

Syntax : - C : / > format drive name

Ex : C : / > format A:

LABEL

This command is used to see volume label and to change volume label.

Syntax : C : / > label drive name

Ex : C : / > label A:

SCANDISK

This utility is used to repair and check various disk errors. It also defects various physical disk errors and surface errors.

Syntax : - C : / > scandisk < drive names >

C : / > Scandisk A :

MOVE

The purpose of move is move to files from one place to another place.

Syntax: C : / > Move < source path > < target path >

TREE

This command displays the list of directories and files on specified path using graphical display. It displays directories of files like a tree.

3. Windows OS :

Windows is a general name for Microsoft Windows. It is developed and marketed by an American multinational company Microsoft. Microsoft Windows is a collection of several proprietary graphical **operating systems** that provide a simple method to store files, run the software, play games, watch videos, and connect to the Internet.

3.1 History of Windows

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Bill Gates is known as the founder of Windows. Microsoft was founded by Bill Gates and Paul Allen, the childhood friends on 4 April 1975 in Albuquerque, New Mexico, U.S.

The first project towards the making of Windows was **Interface Manager**. Microsoft was started to work on this program in 1981, and in November 1983, it was announced under the name "Windows," but Windows 1.0 was not released until November 1985. It was the time of Apple's Macintosh, and that's the reason Windows 1.0 was not capable of competing with Apple's operating system, but it achieved little popularity. Windows 1.0 was just an extension of MS-DOS (an already released Microsoft's product), not a complete operating system. The first Microsoft Windows was a graphical user interface for MS-DOS. But, in the later 1990s, this product was evolved as a fully complete and modern operating system.

3.2 Windows Versions

The versions of Microsoft Windows are categorized as follows:

Early versions of Windows

The first version of Windows was Windows 1.0. It cannot be called a complete operating system because it was just an extension of MS-DOS, which was already developed by Microsoft.

The shell of Windows 1.0 was a program named MS-DOS Executive. Windows 1.0 had introduced some components like Clock, Calculator, Calendar, Clipboard viewer, Control Panel, Notepad, Paint, Terminal, and Write, etc.

In December 1987, Microsoft released its second Windows version as Windows 2.0. It got more popularity than its previous version Windows 2.0. Windows 2.0 has some improved features in user interface and memory management.

The early versions of Windows acted as graphical shells because they ran on top of MS-DOS and used it for file system services.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Windows 3.x

The third major version of Windows was Windows 3.0. It was released in 1990 and had an improved design. Two other upgrades were released as Windows 3.1 and Windows 3.2 in 1992 and 1994, respectively. Microsoft tasted its first broad commercial success after the release of Windows 3.x and sold 2 million copies in just the first six months of release.

Windows 9x (Windows 95, Windows 98)

Windows 9x was the next release of Windows. Windows 95 was released on 24 August 1995. It was also the MS-DOS-based Windows but introduced support for native 32-bit applications. It provided increased stability over its predecessors, added plug and play hardware, preemptive multitasking, and also long file names of up to 255 characters.

It had two major versions Windows 95 and Windows 98

Windows NT (3.1/3.5/3.51/4.0/2000)

Windows NT was developed by a new development team of Microsoft to make it a secure, multi-user operating system with POSIX compatibility. It was designed with a modular, portable kernel with preemptive multitasking and support for multiple processor architectures.

Windows XP

Windows XP was the next major version of Windows NT. It was first released on 25 October 2001. It was introduced to add security and networking features.

It was the first Windows version that was marketed in two main editions: the "Home" edition and the "Professional" edition.

The "Home" edition was targeted towards consumers for personal computer use, while the "Professional" edition was targeted towards business environments and power users. It included the "Media Center" edition later, which was designed for home theater PCs and provided

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

support for DVD playback, TV tuner cards, DVR functionality, and remote controls, etc.

Windows XP was one of the most successful versions of Windows.

Windows Vista

After Windows XP's great success, Windows Vista was released on 30 November 2006 for volume licensing and 30 January 2007 for consumers. It had included a lot of new features such as a redesigned shell and user interface to significant technical changes. It extended some security features also.

Windows 7

Windows 7 and its Server edition Windows Server 2008 R2 were released as RTM on 22 July 2009. Three months later, Windows 7 was released to the public. Windows 7 had introduced a large number of new features, such as a redesigned Windows shell with an updated taskbar, multi-touch support, a home networking system called HomeGroup, and many performance improvements.

Windows 7 was supposed to be the most popular version of Windows to date.

Windows 8 and 8.1

Windows 8 was released as the successor to Windows 7. It was released on 26 October, 2012. It had introduced a number of significant changes such as the introduction of a user interface based around Microsoft's Metro design language with optimizations for touch-based devices such as tablets and all-in-one PCs. It was more convenient for touch-screen devices and laptops.

Microsoft released its newer version Windows 8.1 on 17 October 2013 and includes features such as new live tile sizes, deeper One Drive integration, and many other revisions.

Windows 8 and Windows 8.1 were criticized for the removal of the Start menu.

Windows 10

Microsoft announced Windows 10 as the successor to Windows 8.1 on 30 September 2014. Windows 10 was released on 29 July 2015. Windows 10 is the part of the Windows NT family of operating systems.

Microsoft has not announced any newer version of Windows after Windows 10.

3.3 Difference between DOS and Windows OS

DOS and Windows both are operating systems. DOS is a single tasking, single user and CLI based OS whereas Windows is a multitasking, multiuser and GUI based OS.

Following are the important differences between DOS and Windows.

Sl. No	DOS	Windows
1	DOS stands for Disk Operating System.	Windows stands for simply Windows, no specific form.
2	DOS is single tasking OS.	Windows is multi-tasking OS.
3	DOS consumes quite low power.	Windows consumes high power.
4	DOS memory requirements are quite low.	Windows memory requirements are quite high as compared to DOS.
5	DOS has no support for networking.	Windows supports networking.
6	DOS is complex in usage. We need to remember commands to use DOS properly.	Windows usages is user-friendly and is quite simple to use.
7	DOS is command line based OS.	Windows is GUI based OS
8	Multimedia is not supported in DOS.	Windows supports multimedia likes games, videos, audios etc.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

9	DOS command execution is faster than Windows.	Windows operations are slower as compared to DOS.
10	DOS supports single window at a time.	Windows supports multiple window at a time.

3.4 Components of Windows OS :

The main components of the Windows Operating System are the following:

- Configuration and maintenance
- User interface
- Applications and utilities
- Windows Server components
- File systems
- Networking
- Scripting and command-line
- Kernel
- NET Framework
- Security
- Deprecated components and apps
- APIs

3.5 GUI Components of Windows

Desktop

It is the very first screen that we will see once the windows start. Here we see “My Computer”, “My Documents”, “Start Menu”, “Recycle Bin”, and the shortcuts of any applications that we have created.

Taskbar

Taskbar appears at bottom of Desktop. It has the currently running applications, we can also pin applications that we frequently use by using an option Pin to Taskbar”.

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Start Menu

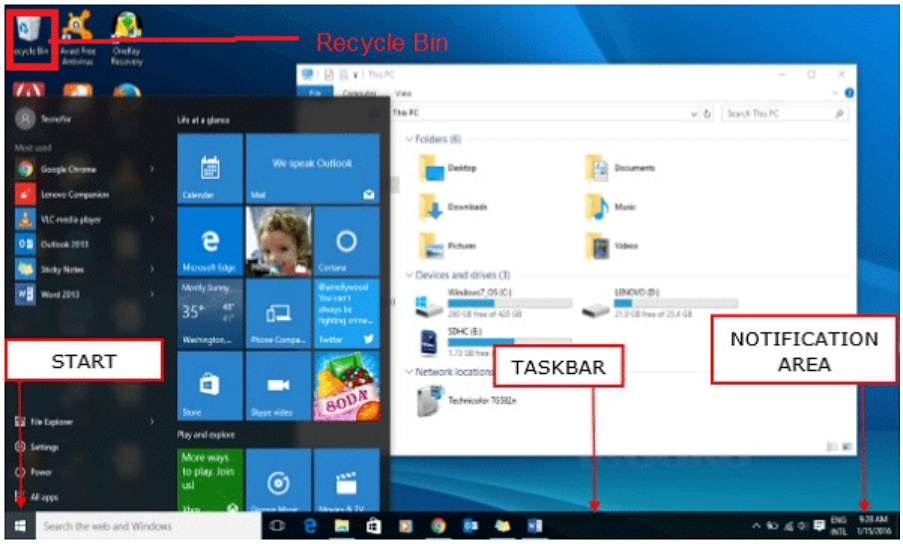
This is located in the bottom left corner of Windows OS GUI. This is the place where the user can search for any setting and for any application for their use. Users can uninstall or repair applications from the control panel. The user can do a lot of activities just by searching through the start menu.

My Computer

It contains the disk drives and other hardware components of computer system. When we double click on “My Computer” menu, it will let us navigate between different computer drives and the control panel tools. We can see and manage the contents that are inside our drive.

Recycle Bin

When we delete an item from any of our drives by making use of “delete” button or even by simply right clicking and selecting “delete” option, it is not deleted completely, instead, it is moved to “Recycle Bin” folder of Windows. We can recover our content if we have deleted it by mistake from here or if we choose to delete the items from here, it will get deleted permanently. If we wish to delete the item in first go itself without moving it to recycle bin, we can use the key “Shift+Del”



3.8 Features of Windows

Main Features of Windows:

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Windows Search:** We can have numerous files and contents located on our system and sometimes we may run out of memory about the exact location of our file. Windows Search is a search function included with Windows that allows the user to search their entire computer

- **Windows File Transfer:** We may have the need to transfer in or transfer out the files and contents from our machine to other devices such as other computers or mobiles and tablets. We can do this by using an Easy Transfer Cable, CDs or DVDs, a USB flash drive, wireless Bluetooth, a network folder, or an external hard disk.

- **Windows Updates:** Windows includes an automatic update feature with the intended purpose of keeping its operating system safe and up-to-date.

- **Windows taskbar:** At the bottom most part of our windows, We will see a row which is known as the taskbar. It has the currently running applications, We can also pin applications that we frequently use by using an option Pin to Taskbar". The taskbar is the main navigation tool for Windows

- **Remote Desktop Connection:** This feature of windows allows us to connect to another system and work remotely on another system.

Glossary :

Disk Operating System: Any operating system which runs through the hard disk drive is termed as Disk Operating System (D.O.S).

Command Line Interface DOS is Command Line Interface (CLI) OS. That commands are typed from the keyboard DOS interprets those commands. The users needed to remember instructions to do their tasks.

File: A file is a set of information stored on a disk under a specific name. A file can contain only data or it can contain a set of instructions, called a program, telling the computer how to perform a particular task.

Directory: A directory may contain flies or its sub directories.

Path: Every file has path starting from root through subdirectories reaching a file. It gives the location of any file or directory from the root directory.

Commands: Instructions given to the computer

Internal commands: Commands of DO S which are resident part of command.com. Ex. Cls, mem, date etc

External commands: Commands of DO S which are not resident part of command.com. Ex. CHKDSK, FORMAT etc

GUI: Graphical users interface. Commands are stored in graphical pictures. By clicking these picture commands can be executed. Eg. Windows, Linux OS.

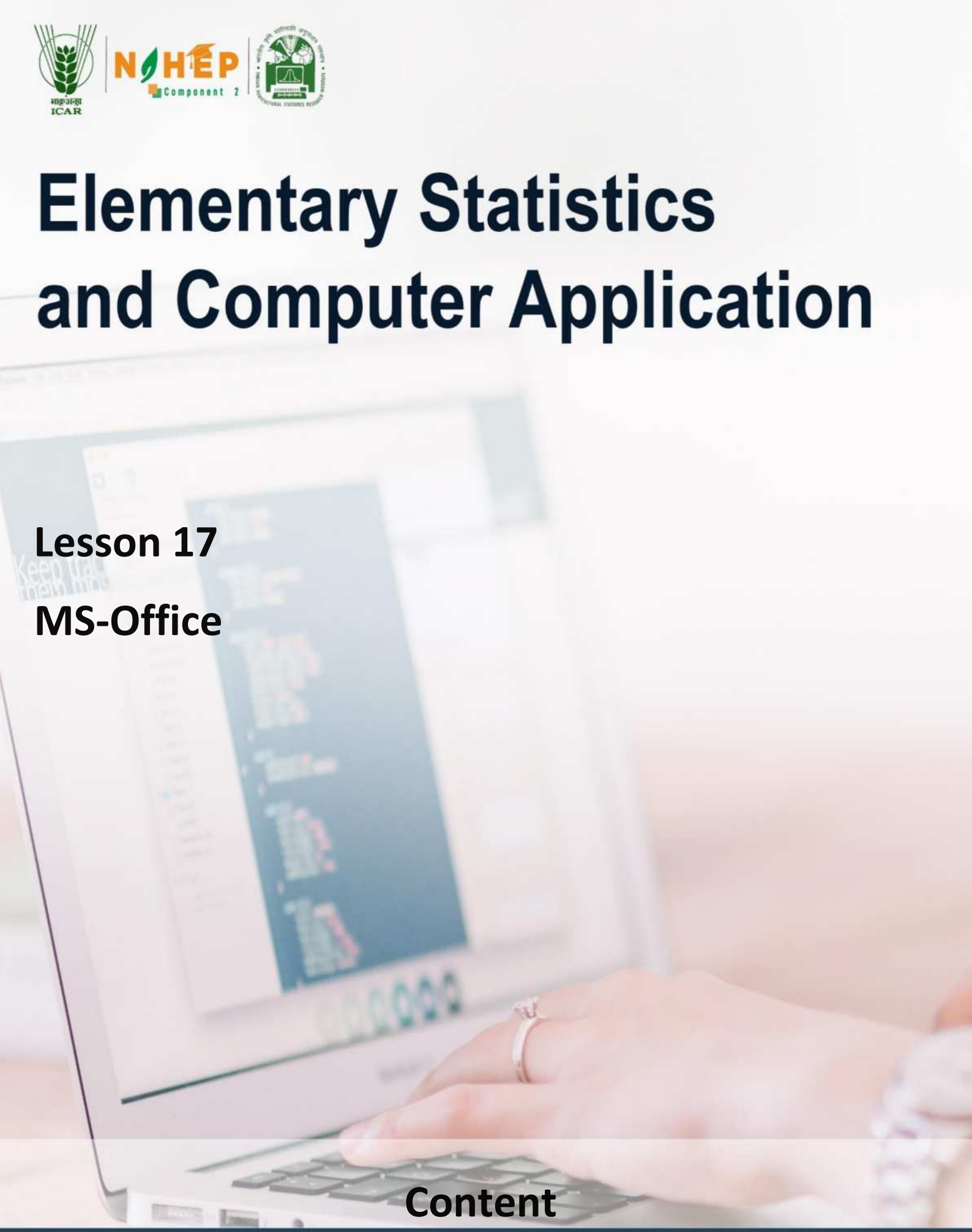
Recycle Bin: Used to store deleted items.

REFERENCES

1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson
2. "O" Level, Introduction to Information Technology, Satish Jain
3. https://www.javatpoint.com/what-is-windows
4. https://www.computerhope.com/msdos.htm

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application



Lesson 17

MS-Office

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 17	MS-Office
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objective:

1. To introduce the basic idea of MS-Office.
2. To discuss the basic features of MS-Word
3. To introduce the basic features of MS-Excel
4. To introduce the basic features of MS-Powerpoint.

Introduction to MS Office

Microsoft Office (or simply **Office**) is a family of server software, and services developed by Microsoft. It was first announced by Bill Gates on August 1, 1988, in Las Vegas.

The first version of Office contained Microsoft Word, Microsoft Excel, and Microsoft Power Point. Over the years, Office applications have grown substantially closer with shared features such as a common spell checker, data integration etc.

Office is produced in several versions targeted towards different end-users and computing environments. The original, and most widely used version, is the desktop version, available for PCs running the Windows, Linux and Mac OS operating systems. Office Online is a version of the software that runs within a web browser, while Microsoft also maintains Office apps for Android and iOS.

Microsoft Office is a suite of desktop productivity applications that is designed specifically to be used for office or business use. It is a proprietary product of Microsoft Corporation and was first released in 1990. Microsoft Office is available in 35 different languages and is supported by Windows, Mac and most Linux variants. It was primarily created to automate the manual office work with a collection of purpose-built applications.

It mainly consists of Word, Excel, Power Point, Access, OneNote, Outlook and Publisher applications.

Each of the applications in Microsoft Office serves as specific knowledge or office domain such as:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1. Microsoft Word: Helps users in creating text documents.
2. Microsoft Excel: Creates simple to complex data/numerical spreadsheets.
3. Microsoft PowerPoint: Stand-alone application for creating professional multimedia
presentations.
4. Microsoft Access: Database management application.
5. Microsoft Publisher: Introductory application for creating and publishing marketing
materials.
6.Microsoft OneNote: Alternate to a paper notebook, it enables an user to neatly organize their notes.

As of 2016, Microsoft Office 2016 is the latest version, available in 4 different variants including Office Home Student 2016, Office Home Business 2016 and Office Professional 2 and the online/cloud Office 365 Home Premium.

Word Processor

A software for creating, storing and manipulating text documents is called word processor. Some common word processors are MS-Word, WordPad, WordPerfect, Google docs, etc.

A word processor allows us to −

- Create, save and edit documents
- Format text properties like font, alignment, font color, background color, etc.
- Check spelling and grammar
- Add images
- Add header and footer, set page margins and insert watermarks

MS-Word :

Microsoft Word is a word-processing software, designed to create professional-quality documents with the finest document- formatting tools, Word helps to organize and write our documents more efficiently. Word also includes powerful editing and revising tools so that we can collaborate with others easily. It is developed by Microsoft and is part of Microsoft Office Suite. It enables us to create, edit and save professional documents like letters and reports.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Important Features of Ms-Word

Ms-Word not only supports word processing features but also DTP features. Some of the important features of Ms-Word are listed below:

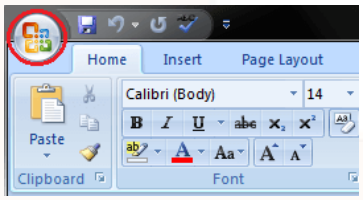
- Using word we can create the document and edit them later, as and when required, by adding more text, modifying the existing text, deleting/moving some part of it.
- Changing the size of the margins can reformat complete document or part of text.
- Font size and type of fonts can also be changed. Page numbers and Header and Footer can be included.
- Spelling can be checked and correction can be made automatically in the entire document. Word count and other statistics can be generated
- Text can be formatted in columnar style as we see in the newspaper. Text boxes can be made.
- Tables can be made and included in the text.
- Word also allows the user to mix the graphical pictures with the text. Graphical pictures can either be created in word itself or can be imported from outside like from Clip Art Gallery.
- Word also has the facility of macros. Macros can be either attached to some function/special keys or to a tool bar or to a menu.
- It also provides online help of any option.

Opening MS-Word :

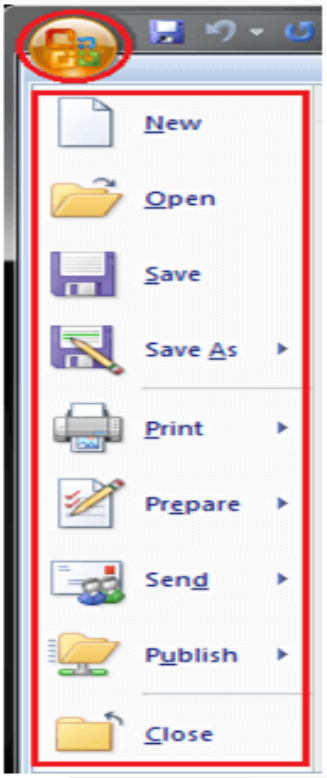
1. Click the **Start** button - the Start menu appears
2. Point to the entry for **All Programs**
3. Click on the entry for **Microsoft Office – Word 2007/2010(whatever version installed)**

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Microsoft Office Button



Microsoft Office Button is located on the top left corner of the window. It is a new user interface feature that replaced the traditional "File" menu. When we click the button it offers a list of commands to perform different tasks which are New, Open, Save, Save As, Print, Prepare, Send, Publish and Close. These commands are described below along with the following image.



New: This command enables us to create a new file, i.e. Word document.

Open: This command allows us to open an existing file on the computer.

Save: This command is used to save a file after completing the work. we can also save the changes made to the currently open file.

Save As: This command helps us to save a new file with a desired file name to a desired location on the hard drive.

Print: This command is used to print a hard copy of the currently open document.

Prepare: This command allows us to prepare the document for distribution, i.e. we can view and edit the document properties and inspect the hidden metadata.

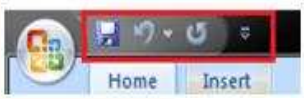
Send: This command allows us to share the document with other users, i.e. we can send a document through e-mail or by posting to a blog.

Publish: This command allows us to distribute the document to other people, i.e. we can create a blog with the content of the document.

Close: This command is used to close the currently open file.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Quick Access Toolbar



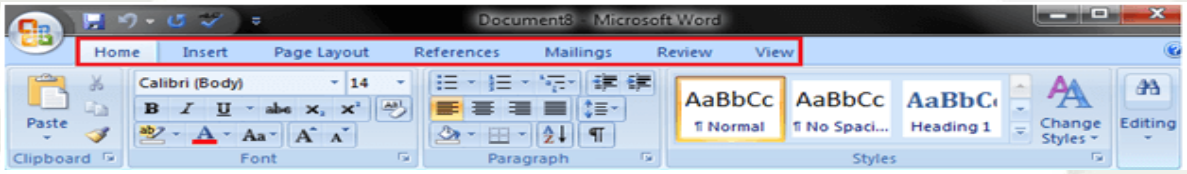
Quick Access Toolbar lies next to the Microsoft Office Button. It is a customizable toolbar that comes with a set of independent commands. It gives us quick access to commonly used commands such as Save, Undo, Redo, etc.

Title Bar

It lies next to the Quick Access Toolbar. It displays the title of the currently open document or application. It is present on almost all windows displayed on our computer. So, if there are several windows across the screen, we can identify each window by looking at the title bar. In many graphical user interfaces, we can also move a window by dragging the title bar.

Ribbon and Tabs

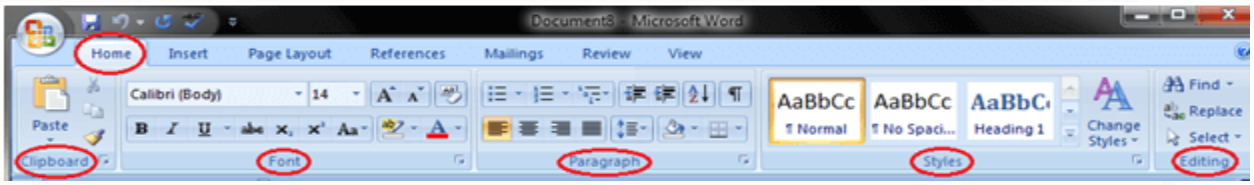
The Ribbon is a user interface element which was introduced by Microsoft in Microsoft Office 2007. It is located below the Quick Access Toolbar and the Title Bar. It comprises seven tabs; Home, Insert, Page layout, References, Mailing, Review and View. Each tab has specific groups of related commands. It gives us quick access to the commonly used commands that need to complete a task.



Home tab:

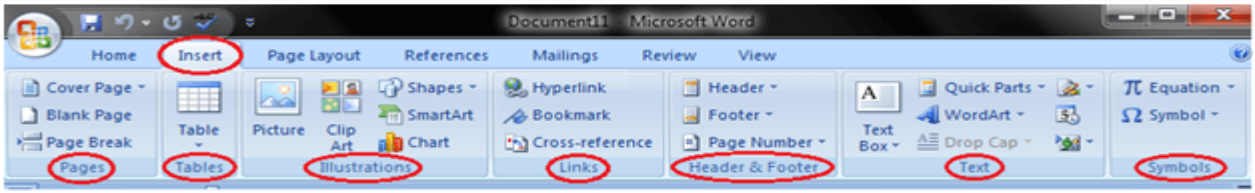
The Home tab is the default tab in Microsoft Word. It has five groups of related commands; Clipboard, Font, Paragraph, Styles and Editing. It helps us change document settings like font size, adding bullets, adjusting styles and many other common features. It also helps us to return to the home section of the document.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute



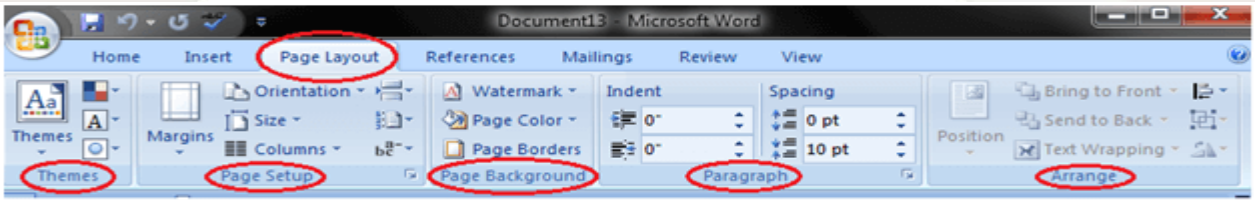
Insert tab:

Insert Tab is the second tab in the Ribbon. As the name suggests, it is used to insert or add extra features in the document. It is commonly used to add tables, pictures, clip art, shapes, page number, etc. The Insert tab has seven groups of related commands; Pages, Tables, Illustrations, Links, Header & Footer, Text and Symbols.



Page Layout tab:

It is the third tab in the Ribbon. This tab allows us to control the look and feel of our document, i.e. we can change the page size, margins, line spacing, indentation, documentation orientation, etc. The Page Layout tab has five groups of related commands; Themes, Page Setup, Page Background, Paragraph and Arrange.



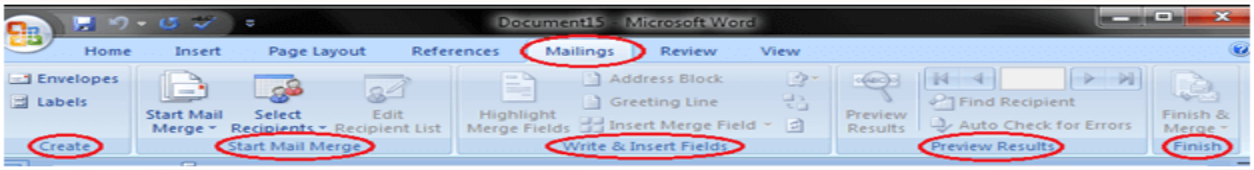
References tab:

It is the fourth tab in the Ribbon. It allows us to enter document sources, citations, bibliography commands, etc. It also offers commands to create a table of contents, an index, table of contents and table of authorities. The References tab has six groups of related commands; Table of Contents, Footnotes, Citations & Bibliography, Captions, Index and Table of Authorities.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

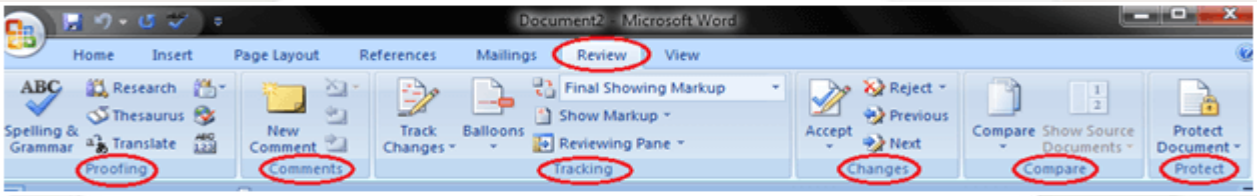
Mailings tab:

It is the fifth tab in the ribbon. It is the least-often used tab of all the tabs available in the Ribbon. It allows us to merge emails, writing and inserting different fields, preview results and convert a file into a PDF format. The Mailings tab has five groups of related commands; Create, Start Mail Merge, Write & Insert Fields, Preview Results and Finish.



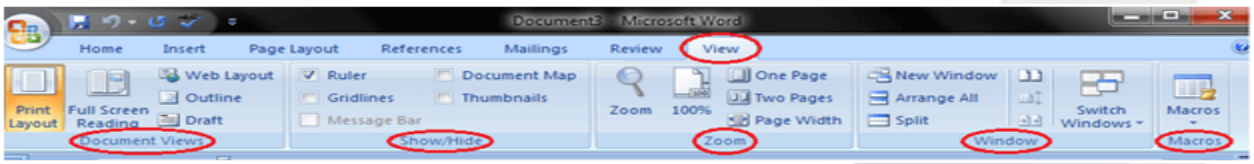
Review tab:

It is the sixth tab in the Ribbon. This tab offers some important commands to modify our document. It helps us proofread our content, to add or remove comments, track changes, etc. The Review tab has six groups of related commands; Proofing, Comments, Tracking, Changes, Compare and Protect.



View tab:

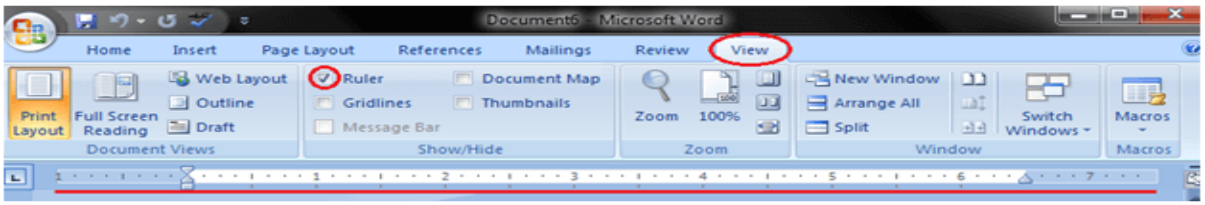
The View tab is located next to the Review tab. This tab allows us to switch between Single Page and Two Page views. It also enables us to control various layout tools like boundaries, guides, rulers. Its primary purpose is to offers us different ways to view our document. The View tab has five groups of related commands; Document Views, Show/Hide, Zoom, Window and Macros.



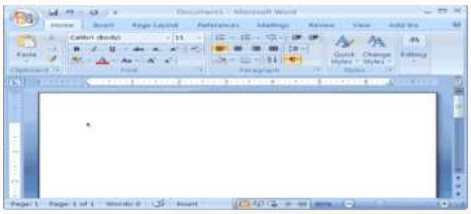
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Ruler

The Ruler is located below the Ribbon around the edge of the document. It is used to change the format of the document, i.e. it helps us align the text, tables, graphics and other elements of our document. It uses inches or centimeters as the measurements unit and gives us an idea about the size of the document.

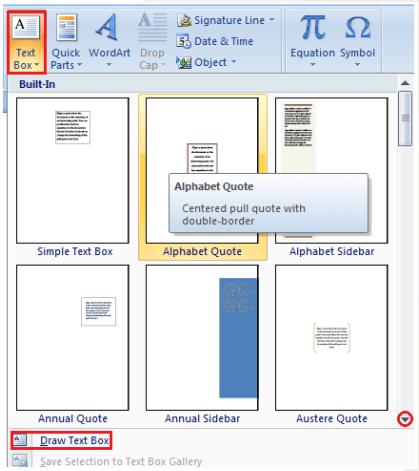


Create a New Document



1. Click the **File (Office Button)** tab and then click **New**.
2. Under A**vailable Templates**, click **Blank Document**.
3. Click **Create**.
4. Screen will appear as shown in the figure with a blinking line (Cursor)

Formatting Documents



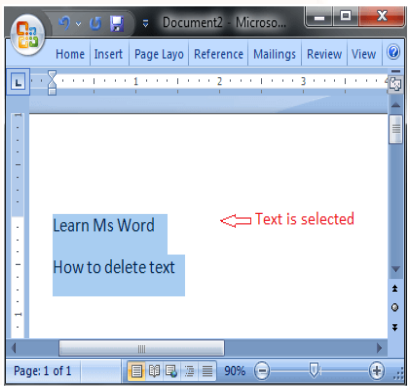
To Insert Text in MS Word

The basic steps to insert text or to create a new document in Word are listed below;

- Go to the start menu and look for Microsoft Word icon
- Click the icon to open the Microsoft Word
- We will see a blinking cursor or insertion point in the text area below the ribbon
- Now, as we start typing, the words will appear on the screen in the text area
- To change the location of insertion point press spacebar, Enter or Tab keys

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

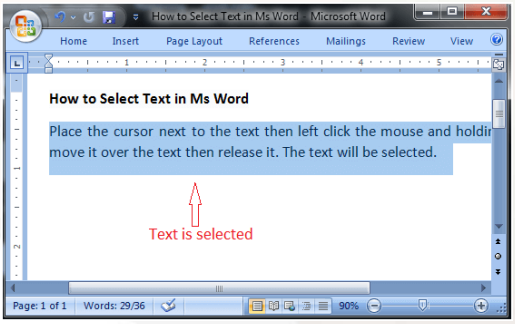
To Delete Text in MS Word



We can easily delete the text in Word including characters, paragraphs or all of the content of our document. Word offers us different methods to delete the text; some of the commonly used methods are given below;

- Place the cursor next to the text then press Backspace key
- Place the cursor to the left of the text then press Delete key
- Select the text and press the Backspace or Delete key
- Select the text and type over it the new text.

To Select Text in MS Word



Place the cursor next to the text then left click the mouse and holding it down move it over the text then release it. The text will be selected.

Some shortcuts for selecting text are:

- To select a single word double click within the word
- To select the entire paragraph triple click within the paragraph
- To select entire document, in Home tab, in Editing group click Select then choose Select All option or press CTRL+A
- o Shift + Arrow; hold down the shift key then press the arrow key, the word will select the text in the direction of the arrow key. There are three arrow keys, so we can select the text in three different directions.

How to Copy and Paste Text in MS Word

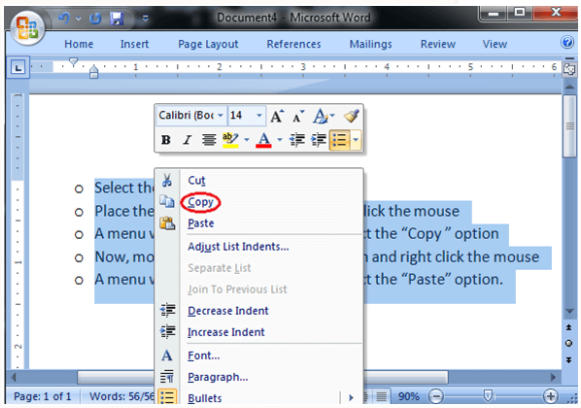
Word offers different methods to copy and paste text. Some of the popular methods are given below;

Method 1;

- Select the text we want to copy

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Select the Home tab and click the Copy command
- Place the cursor where we want to paste the text
- Click the Paste command in Home tab



Method 2;

- Select the text
- Place the cursor over the text and right click the mouse
- A menu will appear; with a left click select the "Copy" option
- Now, move the cursor to a desired location and right click the mouse
- A menu will appear; with a left click select the 'Paste" option.

To Save the Document in MS Word

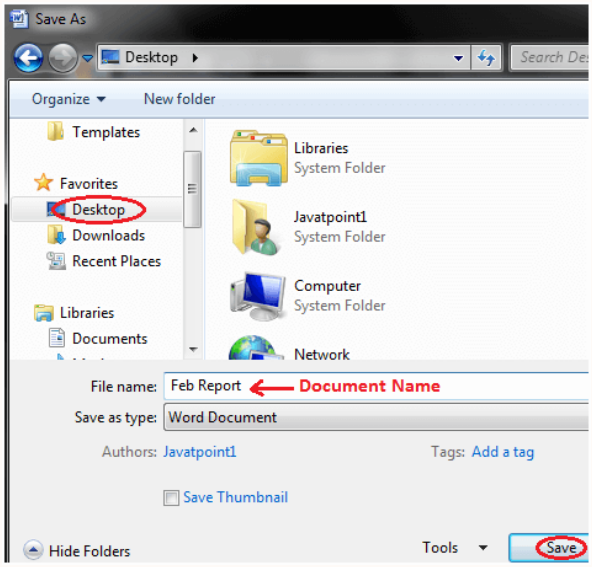
When we create a document it is important to save the document so that it can be viewed or reused later. The basic steps to save a document are listed below;

- Click the Microsoft Office Button
- A list of different commands appears
- Click the 'Save As' command
- it displays 'Save As' Dialogue Box
- Save the document to desired location with a desired name

We can also choose 'Save' command from the list to save the document to its current location with same title. If we are saving a fresh document it displays 'Save As' dialogue box.

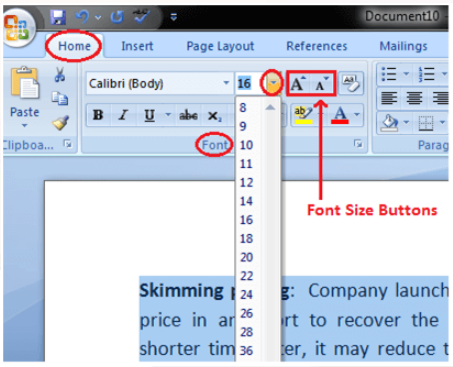
The shortcut method to save a document is to press "Ctrl+S" keys. It opens the 'Save As' dialogue box where we can name document and save it to a desired location.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute



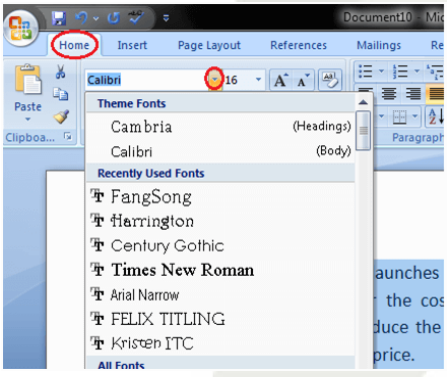
Change Font Size

We can easily change the font size of our text in the document. The basic steps to change the Font size are listed below;



- Select the text that we want to modify
- In Home tab locate the Font group
- In Font group click the drop-down arrow next to font size box
- Font size menu appears
- Select the desired font size with a left click
- Select the text and click the increase or decrease font size buttons

Change Font Style



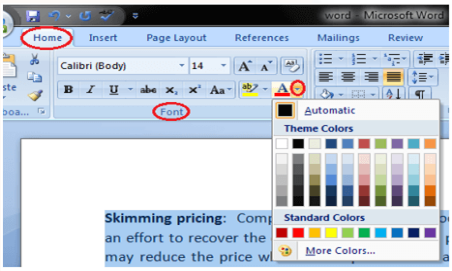
The basic steps to change the font of a text in a document are given below;

- Select the text we want to modify
- Select the Home tab and locate the Font group
- Click the drop-down arrow next to font style box
- Font style menu appears
- With a left click select the desired font style

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- If we want to change the font to bold or italic, click the 'B' or 'I' icons on the format bar.

Change Font Color

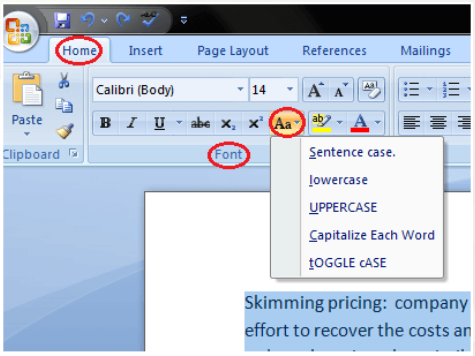


MS Word allows us to change the Font color of our text. If we want to emphasize a particular word or phrase, we can change its font color. The basic steps to change the Font color are given below;

- Select the text we want to modify
- In Home tab locate the Font group
- Click the drop-down arrow next to Font color button
- Font color menu appears
- Select the desired font color with a left click
- Word will change the Font color of the selected text.

Change Text Case

The case menu offers four options;



Sentence case: It capitalizes the first letter of each sentence.

Lowercase: It changes the text from uppercase to lowercase.

Uppercase: It capitalizes all the all letters of our text.

Capitalize Each Word: It capitalizes the first letter of each word.

Toggle Case: It allows us to shift between two case views, e.g. to shift between Capitalize Each Word and cAPITALIZE eACH wORD .

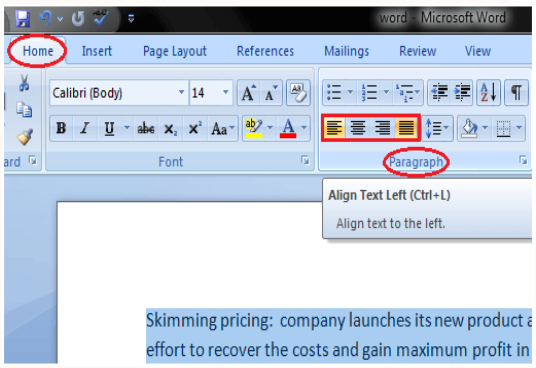
Change Text Alignment

We can change the text alignment in our document to make it more presentable and readable. The basic steps to change the text alignment are given below;

- Select the content we want to modify

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- In Home tab locate the Paragraph group
- It has four alignment options ;



Align Text Left: Aligns the text towards left margin

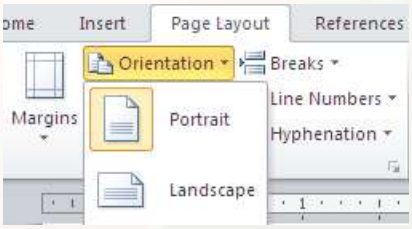
Center: Brings the text at centre

Align Text Right: Aligns the text towards right margin

Justify: Aligns the text to both left and right margins

- Select the desired alignment option with a left click

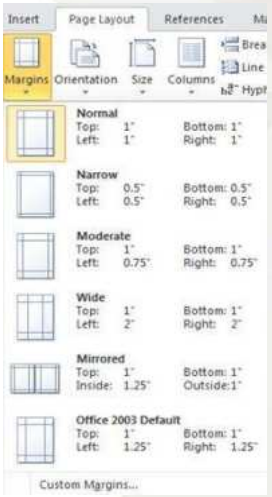
Page Orientation : We can choose either portrait (vertical) or landscape (horizontal) orientation for all or part of our document.



Change Page Orientation

1. On the **Page Layout** tab, in the **Page Setup** group, click **Orientation**.
2. Click **Portrait** or **Landscape**

Page Margins

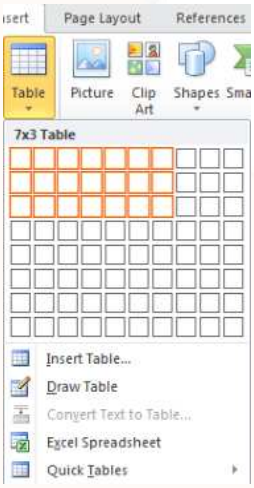


Page margins are the blank space around the edges of the page. In general, we insert text and graphics in the printable area inside the margins When we change a document's page margins, we change where text and graphics appear on each page. We can change the page margins either by choosing from one of Word's predefined settings in the Margins gallery or by creating custom margins.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Creating table in MS-Word

Using tables in Word can provide we with additional elements to any document. Tables can be used to create lists or format text in an organized fashion.



Inserting a Table

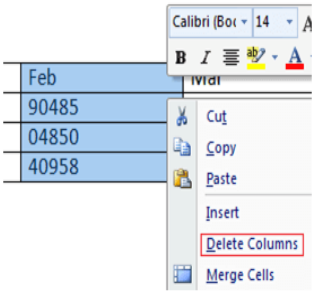
1. Click where we want to insert a table.
2. On the **Insert** tab, in the **Tables** group, click **Table**
3. A drop down box will appear; click and hold our mouse then drag to select the number of rows and columns that we want to insert into our document. We will see our table appearing in our document as we drag on the grid.
4. Once we have highlighted the rows and columns we would like let go of our mouse and the table will be in our document

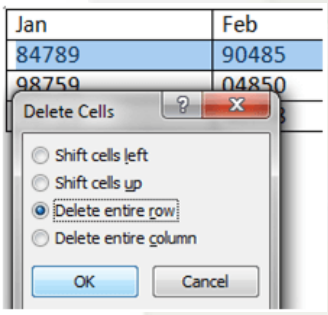
Add Row/Column to Table :

1. Click on the table.
2. Under **Table Tools**, go to the **Layout** tab
3. Click on the **Insert Above** or **Insert Below** to add a row, Click on **Insert Left** or **Insert Right** to insert a column.

Delete Column or Row in Table

The table command also allows us to delete a column or row in our table. We can delete the unwanted columns or rows by following these steps;





- Select the column or row of the table
- Right click the mouse
- A menu appears
- As required select 'Delete Columns' or 'Delete Rows'

Print Preview

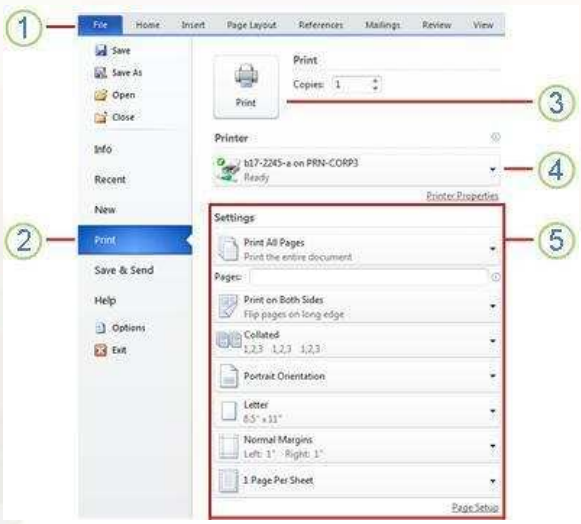
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Print Preview automatically displays when we click on the **Print** tab. Whenever we make a change to a print-related setting, the preview is automatically updated.

1. Click the **File** tab, and then click **Print**. To go back to our document, click the **File** tab.
2. A preview of our document automatically appears. To view each page, click the arrows below the preview.

Printing of Document :

The **Print** tab is the place to go to make sure we are printing what we want.

1. 

1.Click the **File** tab.

2. Click the **Print** command to print a document.

3.Click the **Print** button to print our document.

4. This dropdown shows the currently selected printer. Clicking the dropdown will display other available printers.

5.These dropdown menus show currently selected **Settings**. Rather than just showing we the name of a feature, these dropdown menus show us what the status of a feature is and describes it. This can help figure out if we want to change the setting from what we have.

Basic Computer Shortcut Keys:

The table contains a list of some commonly used basic shortcut keys

Shortcut Keys	Explanation

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Ctrl+A	It allows to **select** the entire content of a page, including images and other objects.
Ctrl+B	It offers users with the option to **bold the selected text** of a page. It also has the various uses in different internet browsers, like in **Firefox** and **Netscape**, it is used to view the bookmarks, and in **Internet Explorer,** used to display the **favorites**.
Ctrl+C	It is used to **copy** the selected content, including other objects of a page.
Ctrl+V	It offers users with the option to **paste** the copied data. We need to copy data once, and then we can paste it any number of times.
Ctrl+F	It provides users with the option to **find** or **search** text in the current document or window.
Ctrl+I	It allows the user to **italicize** and un-italicize the selected **text**.
Ctrl+N	It allows the users to create a **new** or blank **document** in Microsoft applications and some other software. It is also used in internet browsers to open a **new tab**.
Ctrl+O	It is widely used to open a file in the current software.
Ctrl+P	It is used to open the **print** preview window for the current page or document. For example, if we press **Ctrl+P** when a browser or any other document window is open, we will see a print preview window of this page.
Ctrl+S	It is used to **save** the document or a file. we can also use **Shift+F12** to save the file in Microsoft Word.
Ctrl+Y	Its use is to **redo** any undo text and other objects, and it is also used to repeat the last performed action.
Ctrl+Z	It is used to **undo** the content and other objects. For example, if we have deleted the data by mistake, we can retrieve this data by pressing Ctrl+Z immediately.

Spreadsheet

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Spreadsheet is a software that assists users in processing and analyzing tabular data. It is a computerized accounting tool. Data is always entered in a **cell** (intersection of a **row** and a **column**) and formulas and functions to process a group of cells is easily available. Some of the popular spreadsheet software include MS-Excel, G-numeric, Google Sheets, etc. Here is a list of activities that can be done within a spread sheet software −

- Simple calculations like addition, average, counting, etc.
- Preparing charts and graphs on a group of related data
- Data entry
- Data formatting
- Cell formatting
- Calculations based on logical comparisons

MS-Excel

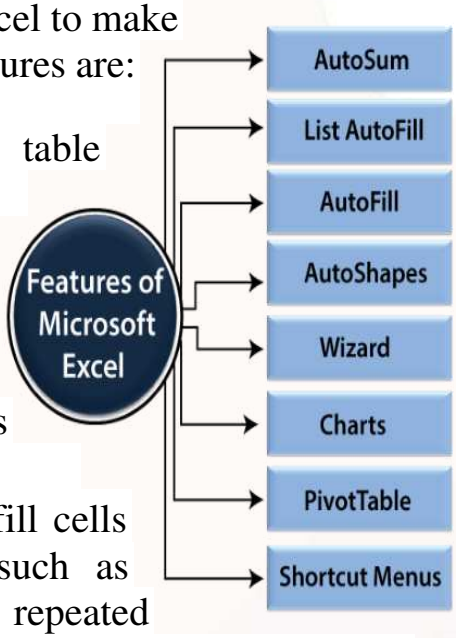
Excel is one of the early software tools developed by Microsoft. The program has been widely adopted by the industry and it is considered the standard by many businesses and organization for spreadsheet applications. Excel is usually packaged in with a software suit developed by Microsoft called Microsoft Office.

Microsoft Excel is a computer application program written by Microsoft. It mainly comprises tabs, groups of commands and worksheet. It is mainly used to store tabular data.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Microsoft Excel Features

There are several features that are available in Excel to make our task more manageable. Some of the main features are:



AutoFormat - lets us choose many preset table formatting options.

1. **AutoSum:** It helps us to add the contents of a cluster of adjacent cells.
2. **List AutoFill:** It automatically develops cell formatting when a new component is added to the end of a list.
3. **AutoFill:** It feature allows us to quickly fill cells with a repetitive or sequential record such as chronological dates or numbers, and repeated document. AutoFill can also be used to copy function. We can also alter text and numbers with this feature.
4. **AutoShapes:** Autoshapes toolbar will allow us to draw some geometrical shapes, arrows, flowchart items, stars, and more. With these shapes, we can draw our graphs.
5. **Wizard:** It guides us to work effectively while we work by displaying several helpful tips and techniques based on what we are doing. Drag and Drop feature will help us to reposition the record and text by simply dragging the data with the help of the mouse.
6. **Charts:** These features will help us in presenting a graphical representation of our data in the form of Pie, Bar, Line charts, and more.
7. **PivotTable:** It flips and sums data in seconds and allows us to execute data analysis and generating documents like periodic financial statements, statistical documents, etc. We can also analyze complex data relationships graphically.
8. **Shortcut Menus:** These commands that are appropriate to the function that we are doing occur by clicking the right mouse button.

Opening MS-Excel :

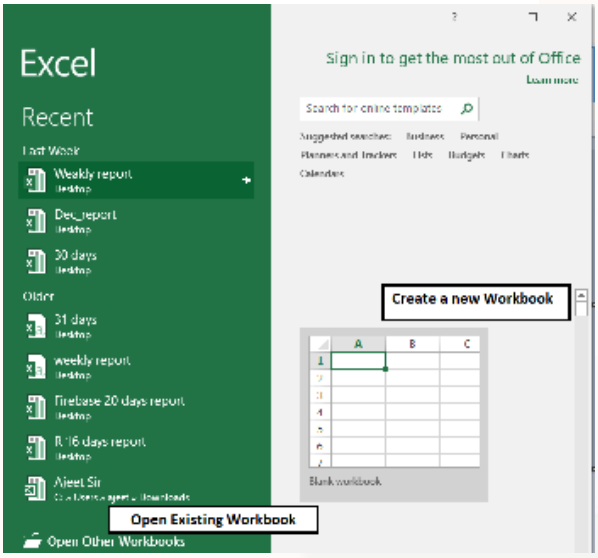
1. Click the **Start** button - the Start menu appears
2. Point to the entry for **All Programs**

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

3. Click on the entry for **Microsoft Office – Excel 2007/2010(whatever version installed)**

To Open Microsoft Excel

When we open Excel 2016 for the first time, the Excel Start Screen will occur. From here, we'll be able to create a new workbook, choose a template, and access our recently edited workbooks.



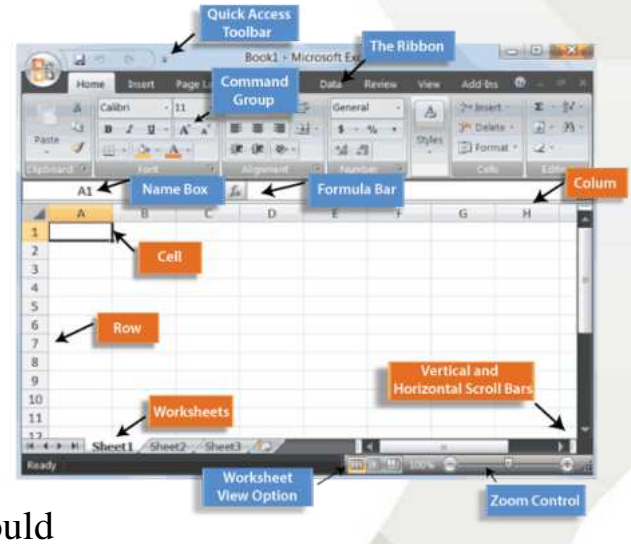
1. From the Excel Start Screen, locate and select the Blank workbook to create the Excel interface.
2. To click Open Other Workbooks to work on an existing workbook.

To set up Excel, so it automatically accessible a new workbook

1. Click File then Options.
2. On the General tab, under the Startup option, uncheck the display the Start screen when this program starts box.
3. The next time we start Excel, it opens a blank workbook automatically same to older versions of Excel.

Excel Interface

After starting Excel, we will see two windows - one within the other. The outer window is the **Application Window**, and the inner window is a **Workbook Window**. When maximized, the Excel Workbook Window composite in with the Application Window.



After completing this module, we should be able to:

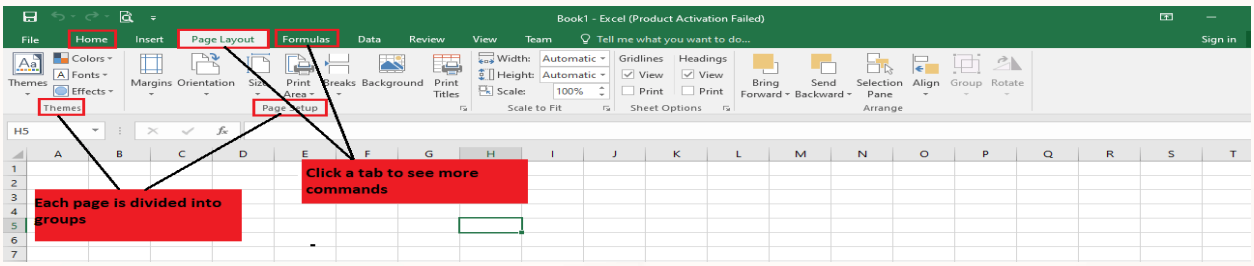
- Identify the components of the Application Window.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Identify the components of the Workbook Window.

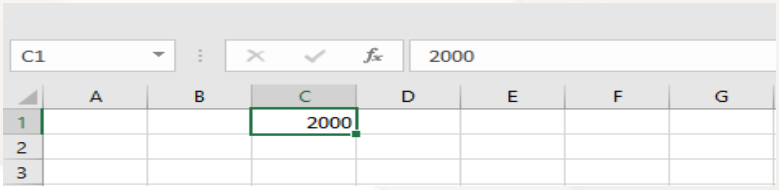
The Ribbon

Excel 2016 utilizes a **tabbed Ribbon system** instead of traditional menus. The **Ribbon** includes multiple tabs, each with several **groups of commands**. We will use these tabs to perform the most common function in Excel.



The Formula Bar

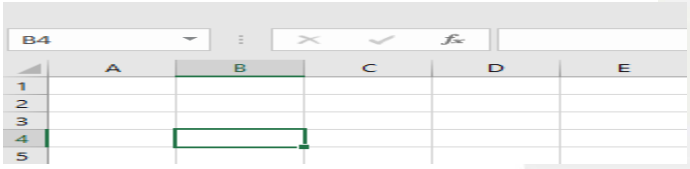
In the **formula bar**, we can enter or edit data, a formula, or a function that will occur in a specific cell.



In the image, cell C1 is selected, and 2000 is entered into the formula bar. Note how the data contains in both the formula bar and in cell C1.

The Name Box

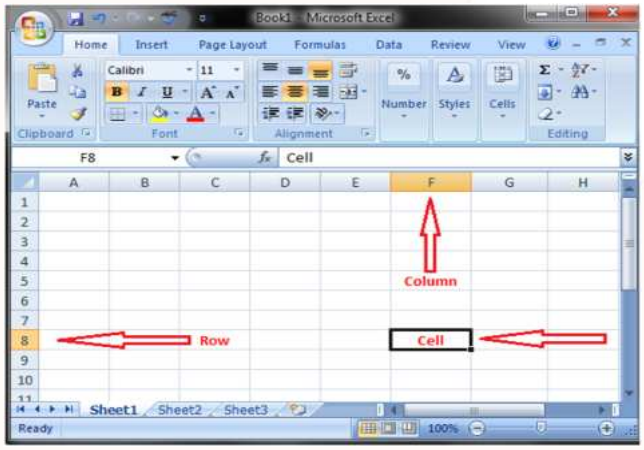
The Name box present the location or **"name"** of a **selected cell**.



In the image, cell B4 is selected. Noted that cell B4 is where column B and row 4 intersect.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Worksheet



Excel files are known as **workbooks**. Each workbook hold one or more worksheets (also called a "**spread sheets**").

Whenever we create a new Excel workbook, it will include **one worksheet** named **Sheet1**.

It is made up of rows, columns and cells.

Rows

Rows run horizontally across the worksheet. A row is identified by the number that is on left side of the row, from where the row originates.

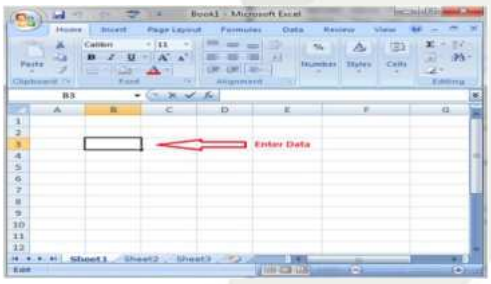
Columns

Columns run vertically downward across the worksheet . A column is identified by a column header that is on the top of the column, from where the column originates.

Cells

Cells are small boxes in the worksheet where we enter data. A cell is the intersection of a row and column. It is identified by row number and column header such as A1, A2.

To enter data in Excel



Select a cell with a single click where we want to enter data; cell B3 is selected in the image given below. Then double click in the cell to enter data. We can enter text, numbers and formulas in the cell.

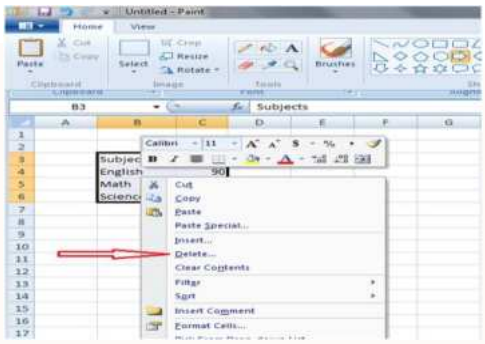
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

After entering data, we can press Tab key to move to next column and can press Enter key to move to next row. We can press arrow keys for more options to move to other cells.

To delete data, rows and columns in Excel

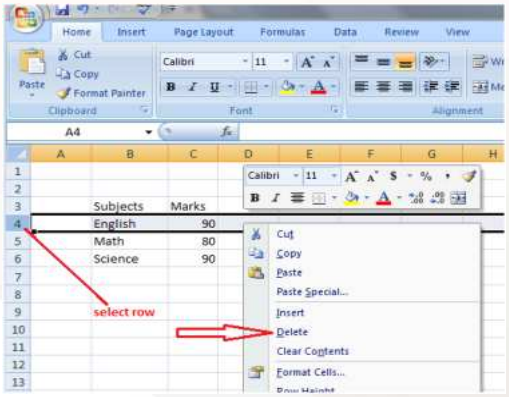
How to Delete Data

Select the data we want to delete, right click on it then select delete option from the menu. We can also delete it by pressing Delete key on the keyboard.



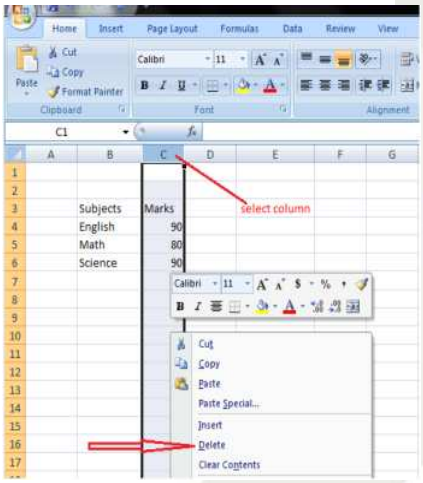
How to Delete a Row

Select the row by a left click on the row number then right click and select the Delete option. If we want to delete more rows drag the mouse downward to select more rows.



How to Delete a Column

Select the column by a left click on the column header then right click and select Delete option. To delete more columns drag the mouse horizontally left or right to select more columns.



Formulas in MS Excel

Formulas are the Bread and butter of worksheet. Without formula, worksheet will be just simple tabular representation of data. A formula

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

consists of special code, which is entered into a cell. It performs some calculations and returns a result, which is displayed in the cell.

Elements of Formulas

A formula can consist of any of these elements −

- **Mathematical operators, such as +(for addition) and *(for multiplication)**

  Example −

  - =A1+A2 Adds the values in cells A1 and A2.
- **Values or text**

  Example −

  - =200*0.5 Multiplies 200 times 0.15. This formula uses only values, and it always returns the same result as 100.
- **Cell references (including named cells and ranges)**

  Example −

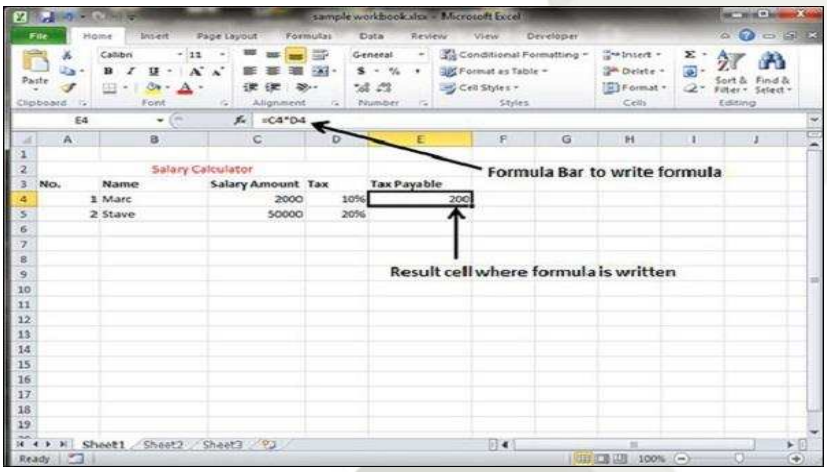
  - =A1=C12 Compares cell A1 with cell C12. If the cells are identical, the formula returns TRUE; otherwise, it returns FALSE.
- **Worksheet functions (such as SUMor AVERAGE)**

  Example −

  - =SUM(A1:A12) Adds the values in the range A1:A12.

Creating Formula

For creating a formula we need to type in the Formula Bar. Formula begins with '=' sign. When building formulas manually, we can either type in the cell addresses or we can point to them in the worksheet. Using the **Pointing method** to supply the cell addresses for formulas is often easier and more powerful method of



Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

formula building. When we using built-in functions, we click the cell or drag through the cell range that we want to use when defining the function's arguments in the Function Arguments dialog box. See the screen shot.

As soon as we complete a formula entry, Excel calculates the result, which is then displayed inside the cell within the worksheet (the contents of the formula, however, continue to be visible on the Formula bar anytime the cell is active). If we make an error in the formula that prevents Excel from being able to calculate the formula at all, Excel displays an Alert dialog box suggesting how to fix the problem.

Functions:-

A function is a predefined program which gives the result to specified values of needed calculations. It reduces the complex task to a simple one. A function consists of a formula, which in turn consists of cell references. There are different types of functions available in Ms-Excel. They are as follows:

1. Mathematical & Trigonometric functions.
2. Date & Time functions
3. Text functions
4. Logical Functions.
5. Database functions
6. Statistical Functions
7. Lookup & reference
8. Information
9. Financial Functions.

Parts of Function:

Every function is Excel comprises of two parts:

1. Function name
2. Function arguments

Function Arguments:

Arguments are data received by the function. Each function receives a specific kind of arguments. It can be numbers, text, date or logical values such as True or False

Syntax:

=<Function name> (<List of arguments>)

The following example illustrates the usage of the ROUND function.

=ROUND(6,3)

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Here, ROUND is the name of the function, which rounds the decimal part of a number to the specified number digits. In this example, the ROUND function will round off the value in the cell C6 up to three decimal places.

Rules for Using Functions:

Just like a formula, functions also follow certain rules.

1. All functions must begin with sign.
2. The arguments of a function must be enclosed within brackets.
3. The arguments should be separated by a comma.
4. The cell range should be mentioned using a colon.

When entering a functions, we should start it with an '=' (equal to) symbol. Function is identified with a parenthesis. Eg: Sum( ), avg( ). Whenever we select any function, a function wizard appears which will show the next step to execute the given function. When we select the function command, a dialog box appears where to the left side we can see the different types of functions. When we select a particular category, the functions in that category are shown to its right side.

Mathematical & Trigonometric Functions:-

All the mathematical calculations like addition, multiplication, power etc can be done easily by using these functions. We can also find the trigonometric values of sin, cos, tan etc., using these functions.

a) Sum( ):- This function is used to add the given values that we pass as the parameters. The parameters can be the values or may be the cell addresses containing the values.

Syntax: =sum (number1, number2,……)

b) Fact( ):- It will return the factorial of a given number.

Syntax: =Fact(number)

Date & Time functions:-

These functions are related to system date and time. In Excel, every date and time is identified with a number.

a) DateValue( ):- This function returns the numbers that represents a date in Excel. The parameter we have to specify is the date in text i.e., in single quotes. When entering the date, we have to give month, day and then the year.

Syntax: =datevalue(date_text)

Eg: =datevalue('12/22/2007')

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

b) Today( ):- It will not take any parameters but will display the system date.
Eg: =today( )

Text functions:-
These functions deals with text or the string we enter in the cells.
a) Lower( ):- It is used to convert the given string into lower case.

Syntax: =lower('string')
Eg: =lower('AURORA')
Result: aurora
b) Upper( ):- It converts the given string into upper case.

Syntax: =upper('string')
Eg: =upper('madhu')
Result: MADHU

Logical Functions:-
These functions are used to display the appropriate value by checking the given condition.

a) If( ):- This function checks for a condition specified as the first parameter, and when it is true displays the second parameter in the cell otherwise third parameter value is displayed.

Syntax: =if(condition, true value, false value)
Eg: =if(3>5, "good", "best")
Result: best
b) And( ):- It will return true when all the conditions passed as the parameters are true. Otherwise it will return false.

Syntax: =and(conditiona 1, condition 2)
Eg: =and(2>1, 5>4,7>3)
Result: True
c) Or( ):- It will return true when any of the condition is true.

Syntax: =or(conditiona 1, condition 2)
Eg: =or(3>5,9>2,45>23)
Result: True

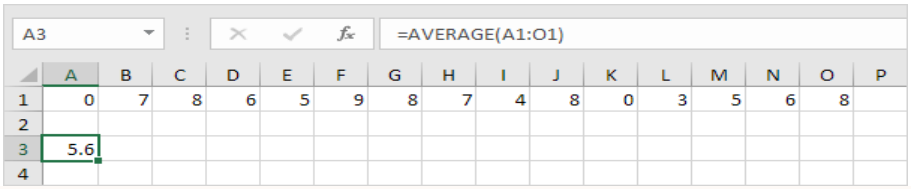
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Statistical Functions:-

Here, some very useful **statistical functions** in **Excel**.

Average

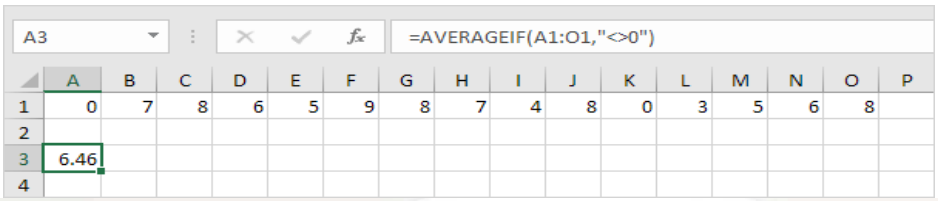
To calculate the average of a group of numbers, use the AVERAGE function.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	7	8	6	5	9	8	7	4	8	0	3	5	6	8	
2																
3	5.6															
4																

Average

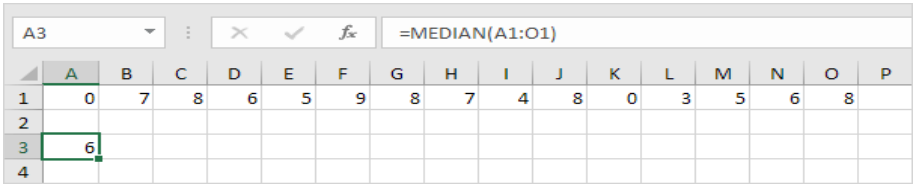
To average cells based on one criteria, use the AVERAGEIF function. For example, to calculate the average excluding zeros.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	7	8	6	5	9	8	7	4	8	0	3	5	6	8	
2																
3	6.46															
4																

Median

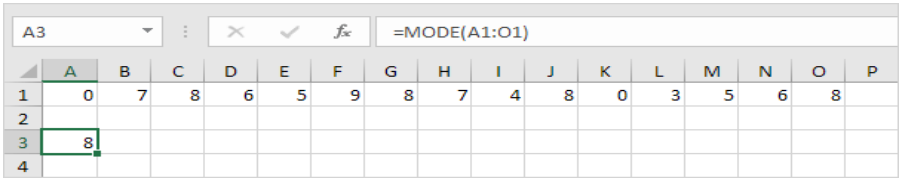
To find the median (or middle number), use the MEDIAN function.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	7	8	6	5	9	8	7	4	8	0	3	5	6	8	
2																
3	6															
4																

Mode

To find the most frequently occurring number, use the MODE function.

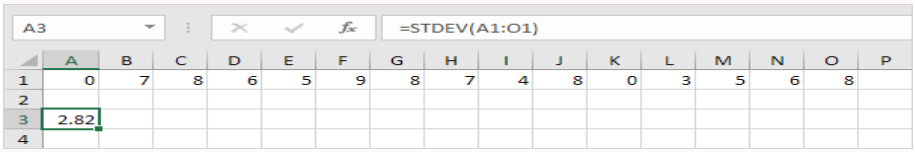


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	0	7	8	6	5	9	8	7	4	8	0	3	5	6	8	
2																
3	8															
4																

Standard Deviation

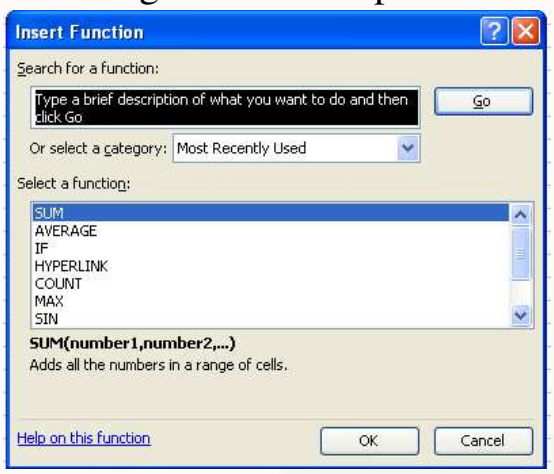
To calculate the standard deviation, use the STEDV function.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute



The Insert Function Feature:

By now, we must be familiar with various functions available in Excel. Working with these functions are as easy as learning them. Excel provides a feature called Insert Function, which helps us to work with these functions more easily.



To invoke the Insert Function dialog box,

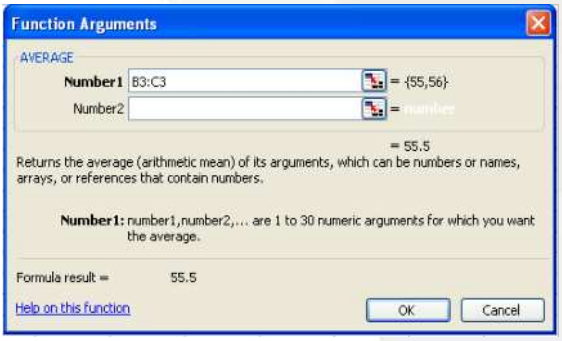
1. Click the Insert menu.
2. Choose the Function option.

The Insert Function dialog box allows us to search for a function text box.

1. Type the function in the Search for function text box.
2. Click the Go button.

Inserting Functions:

The insert Function dialog box enables we to insert a specific function for the data in the worksheet. Assume that we have the marks of a student in five subjects. We can find the average mark of the student using the Average function under the Statistical category.



3. Enter the cell range in the Number1 text box
4. Click the OK button.

Saving and Sharing Workbooks

Whenever we create a new workbook in Excel, we'll need to know how to store it to access and edit it later. As with previous versions of Excel, we can save files locally to our computer. But unlike older versions, Excel 2016 also lets us save a workbook to the cloud using OneDrive. We can also export and share workbooks with others directly from Excel.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

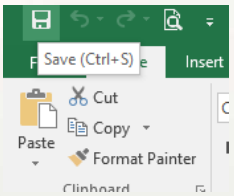
Save and Save As

Excel offers two methods to save a file: Save and Save As. These options work in similar methods, with a few crucial differences:

- **Save:** When we create or edit a workbook, we'll use the Save command to save our changes. We'll use this command most of the time. When we save a file, we'll only need to select a file name and location the first time. After that, we can just click the Save command to save it with a similar name and location.
- **Save As:** We'll use this command to create a copy of a workbook while keeping the original. When we use Save As, we'll need to select a different name and/or location for the copied version.

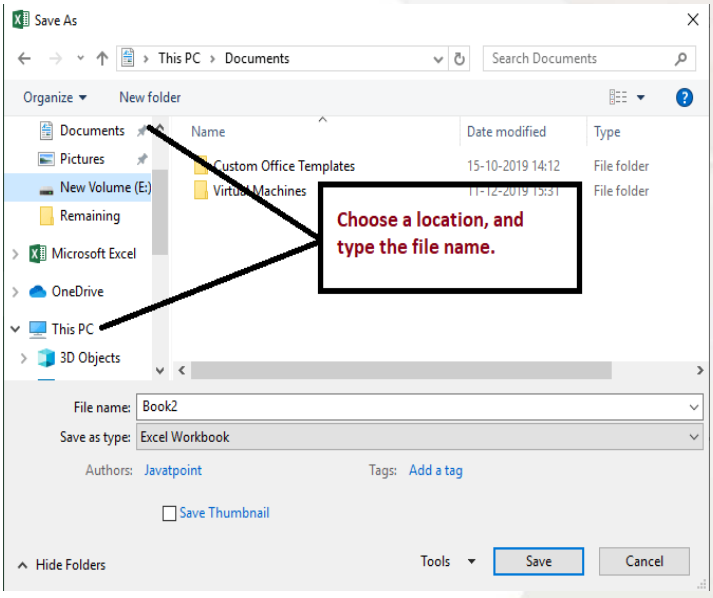
To save a workbook

It's essential to save our workbook whenever we start a new project or make changes to an existing one. Saving early and often can prevent our work from being lost.



1. To locate and select the Save command on the Quick Access Toolbar.

2. If we're saving the file for the first time, the Save As pane will occur in the Backstage view.



3. We'll then need to choose where to save the file and give it a file name. To save the workbook to our computer, select the Computer, then click Browse. Alternatively, we can click OneDrive to save the file to our OneDrive.

4. The **Save As** dialog box will emerge. Select the location where we want to save the workbook.

5. Enter the **file name** for the workbook, then click Save.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

6. The workbook will be **saved**. We can click the Save command again to save our changes as we modify the workbook.

Presentation Tool

Presentation tool enables user to demonstrate information broken down into small chunks and arranged on pages called **slides**. A series of slides that present a coherent idea to an audience is called a **presentation**. The slides can have text, images, tables, audio, video or other multimedia information arranged on them. MS-Power Point, Open Office Impress, Lotus Freelance, etc. are some popular presentation tools.

Introduction to MS-Power Point

A Power Point presentation is a presentation created using Microsoft Power Point software. The presentation is a collection of individual slides that contain information on a topic.

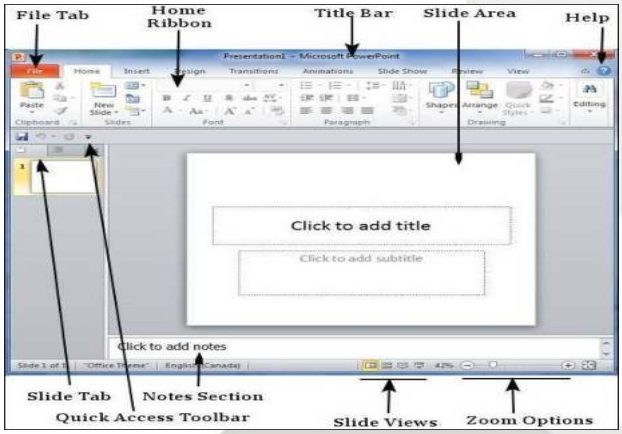
Power Point presentations are commonly used in business meetings and for training and educational purposes.

Microsoft Power Point is a software product used to perform computer based presentations. There are various circumstances in which a presentation is made: such as teaching a class, introducing a product to sell, explaining an organizational structure, etc.

Microsoft PowerPoint is a software program developed by Microsoft to produce effective presentations. It is a part of Microsoft Office suite. The program comprises slides and various tools like word processing, drawing, graphing and outlining. Thus it can display text, table, chart, graphics and media in the slides.

Features

PowerPoint software features and formatting options include a wizard that walks we through the presentation creation process.



Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

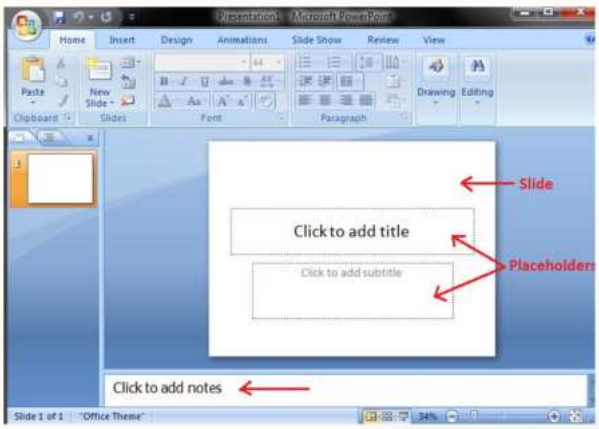
Design templates---pre packaged background designs and font styles that will be applied to all slides in a presentation. When viewing a presentation, slide progression can be manual, using the computer mouse or keyboard to progress to the next slide, or slides can be set up to progress after a specified length of time.

Slide, Placeholder and Notes

Slide: Presentation is created on slides. It lies in the centre of the PowerPoint window.

Placeholder: By default two placeholders appear in the slide when we open the PowerPoint.

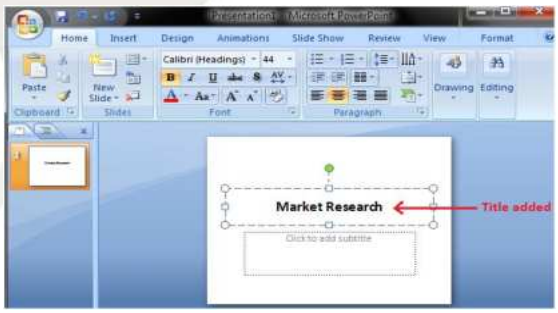
Click to add notes: This space is provided to create notes if needed.

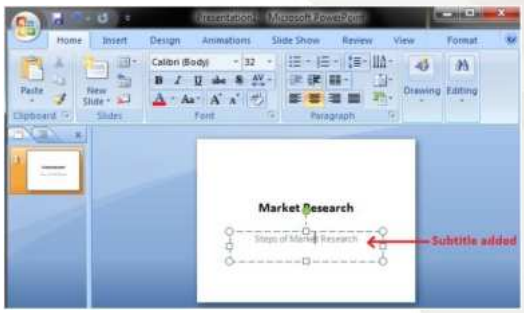


Creating a Presentation

When we open PowerPoint by default a slide appears. The slide has two placeholders or text boxes. Additional text boxes can be added from the Insert tab.

To start creating presentation click on the placeholder or text box a





blinking cursor will appear. Then type the title and click outside the box. The text box will disappear.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

How to Save a Presentation

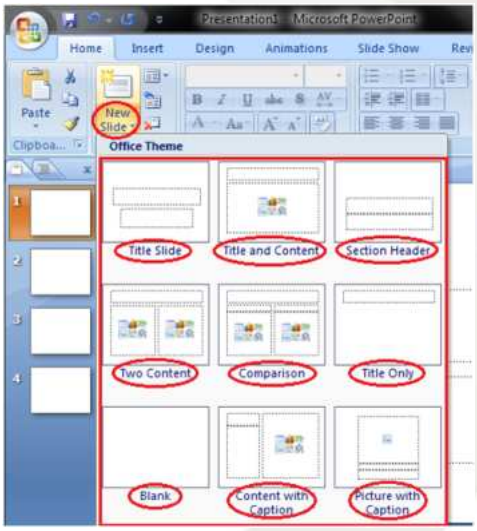
There are multiple options to save a presentation. The frequently used options are:

Click on the Microsoft Office Button then select Save or Save As from the menu.

Add Slide

There are multiple ways to add slide in PowerPoint presentation. The frequently used option is to click the New Slide button.

- Select the slide next to which we want the new slide to appear
- In Home tab, click the drop-down arrow on the New Slide button
- It will display the office themes
- Select the slide choice that suits our requirement



Apply Themes in Slide

Themes are design templates that make the presentation colorful and stylist. With a single click we can apply a theme to the entire presentation.

- Open the Design tab
- Locate the Themes group
- Click the desired theme

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Theme will be added to the entire presentation

To see all available Themes click the drop- down arrow on the right bottom corner of the Themes group.

Apply or Change Color in Themes

- Open the Design tab
- Click the drop-down arrow next to Colors in the Themes group
- With a left click select the desired color set
- To create new color set click the Create New Theme Colors

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

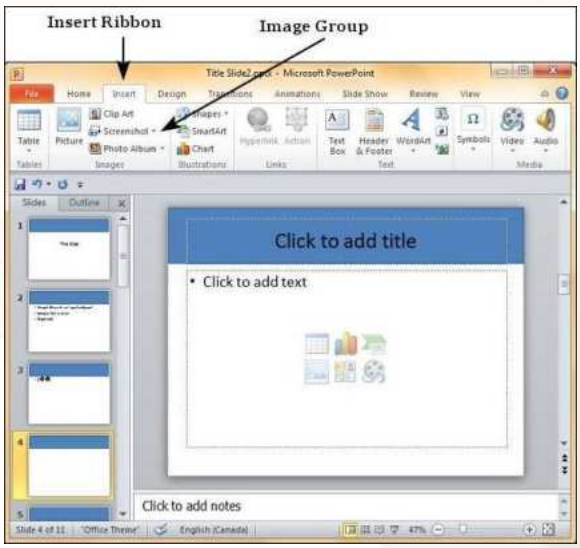
Add Pictures to Slide

PowerPoint supports multiple content types including images or pictures. With regards to pictures PowerPoint classifies them into two categories −

- **Picture** − Images and photos that are available on our computer or hard drive
- **Clip Art** − Online picture collection that we can search from the clip art sidebar

Although their sources are different, both these types can be added and edited in similar fashion. Given below are the steps to add picture to a slide.

Step 1 − Go to the **Images** group in the **Insert** ribbon.

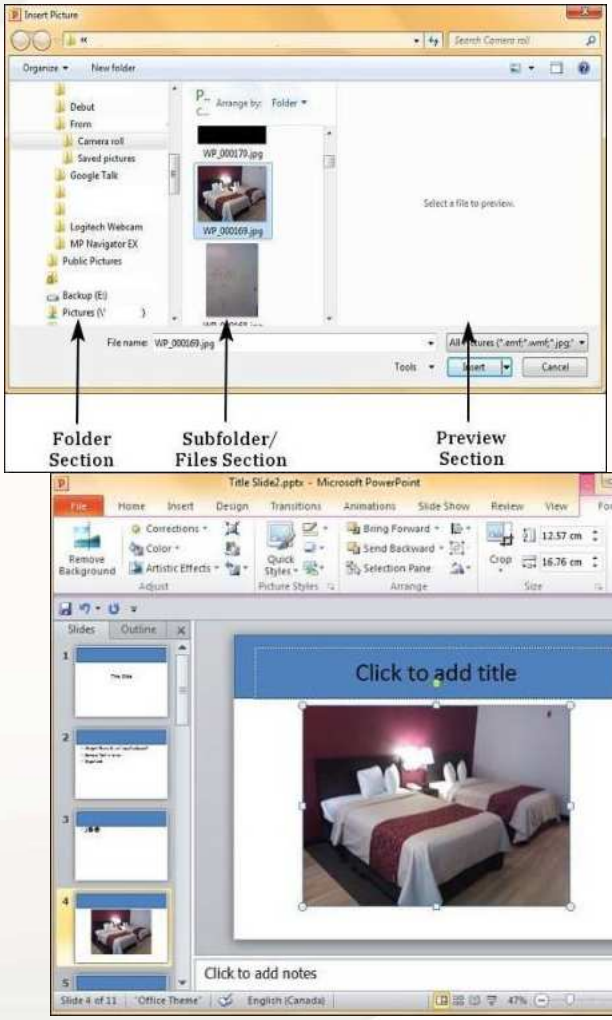


Step 2 − Click on Picture to open the **Insert Picture** dialog and add a picture to the slide.

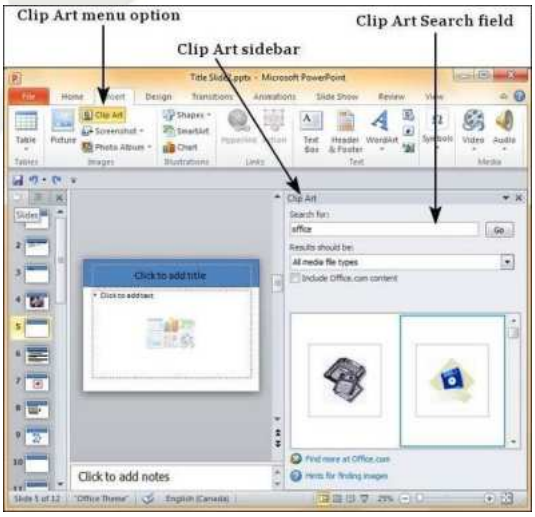
Step 3 − In this dialog, we have three sections: to the left corner, we have folders that can be browsed, the section in the center shows the subfolders and files in the selected folder and to the right, we can have a preview of the selected image.

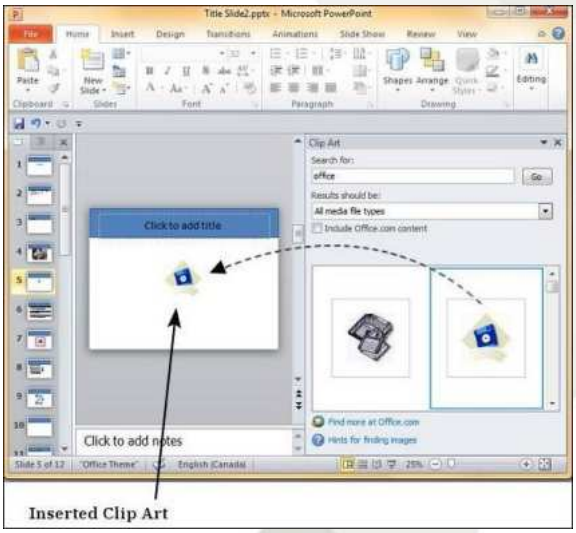
Step 4 − Select the image we want and click **Open** to add the picture to the slide.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute



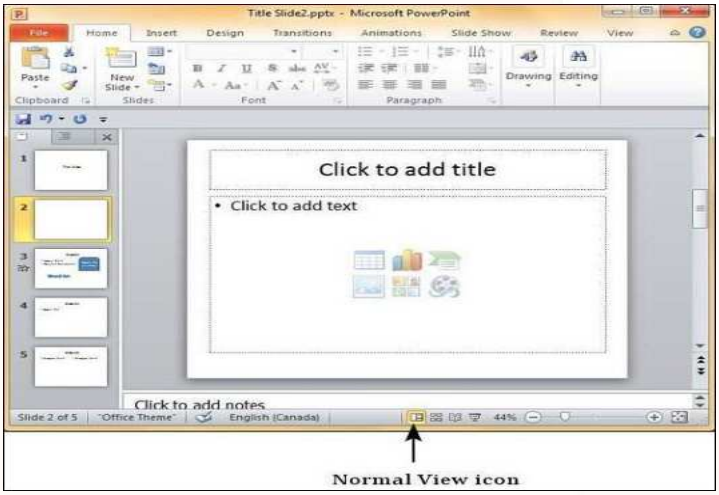
Step 5 − To add online pictures, click on **Clip Art** and search for keywords in the **Clip Art sidebar**.





Step 6 − Once we have the clipart we want to use, double-click on the image to add it to the slide.

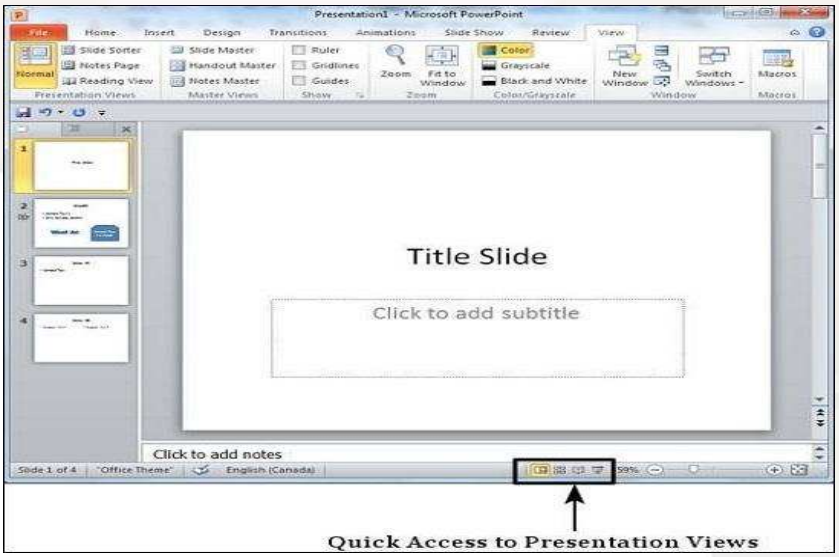
We can view the created slide with the following slide view options:



1. Presentation View :

PowerPoint supports multiple views to allow users to gain the maximum from the features available in the program. Each view supports a different set of functions and is designed accordingly PowerPoint views can be accessed from two locations.

- Views can be accessed quickly from the bottom bar just to the left of the zoom settings.

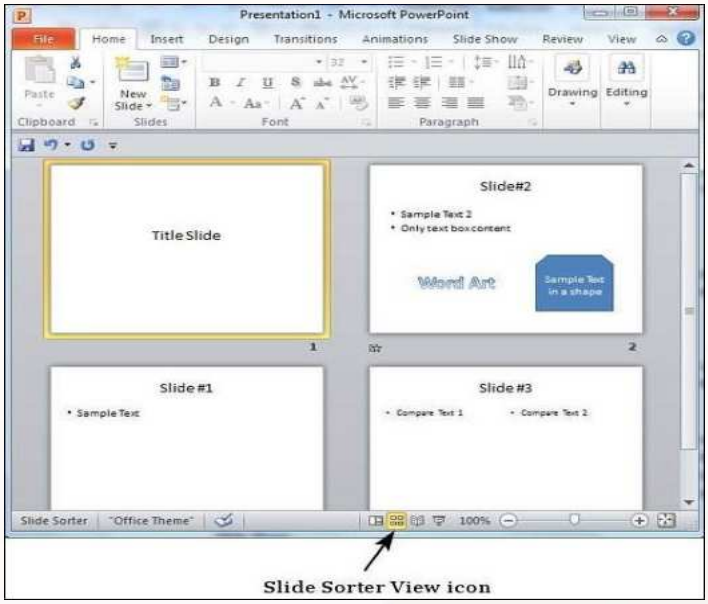


Views can also be accessed from the **Presentation Views** section in the View ribbon

Here is a short description of the various views and their features.

2. Normal View

This is the default view in PowerPoint and this is primarily used to create and edit slides. We can create/ delete/ edit/ rearrange slides, add/ remove/ modify content and manipulate sections from this view.

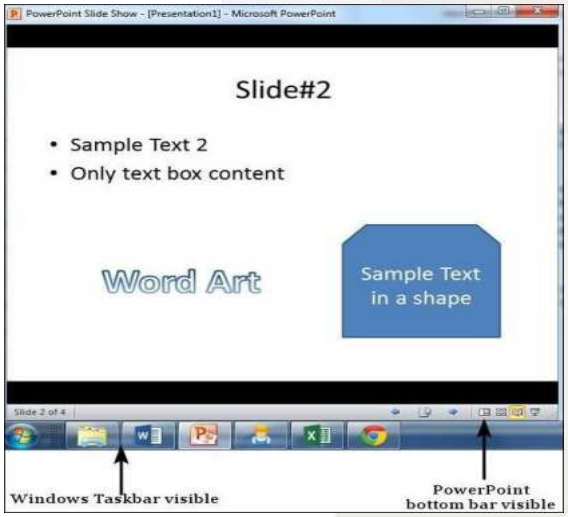


3. Slide Sorter View

This view is primarily used to sort slides and rearrange them. This view is also ideal to add or remove sections as it presents the slides in a more compact manner making it easier to rearrange them.

4. Reading View

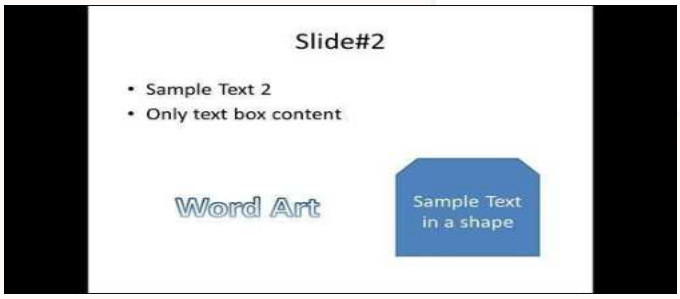
This view is new to PowerPoint 2010 and it was created mainly to review the slideshow without losing access to rest of the Windows applications. Typically, when we run the slideshow, the presentation takes up the entire screen so other applications cannot be accessed from the taskbar. In the reading view the taskbar is still available while viewing the slideshow which is convenient. We cannot make any modifications when on this view.



Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

5. Slides Show

This is the traditional slideshow view available in all the earlier versions of PowerPoint. This view is used to run the slideshow during presentation.



Glossary :

Microsoft Office: Microsoft Office is a suite of desktop productivity applications that is designed specifically to be used for office or business use.

Microsoft Excel: Creates simple to complex data/numerical spreadsheets.

Microsoft PowerPoint: Stand-alone application for creating professional multimedia

presentations.

Microsoft Access: Database management application.

Microsoft Publisher: Introductory application for creating and publishing marketing.

Word Processor A software for creating, storing and manipulating text documents is called word processor. Some common word processors are MS-Word, WordPad, WordPerfect, Google docs, etc.

Spreadsheet : Spreadsheet is a software that assists users in processing and analyzing tabular data. It is a computerized accounting tool.

REFERENCES

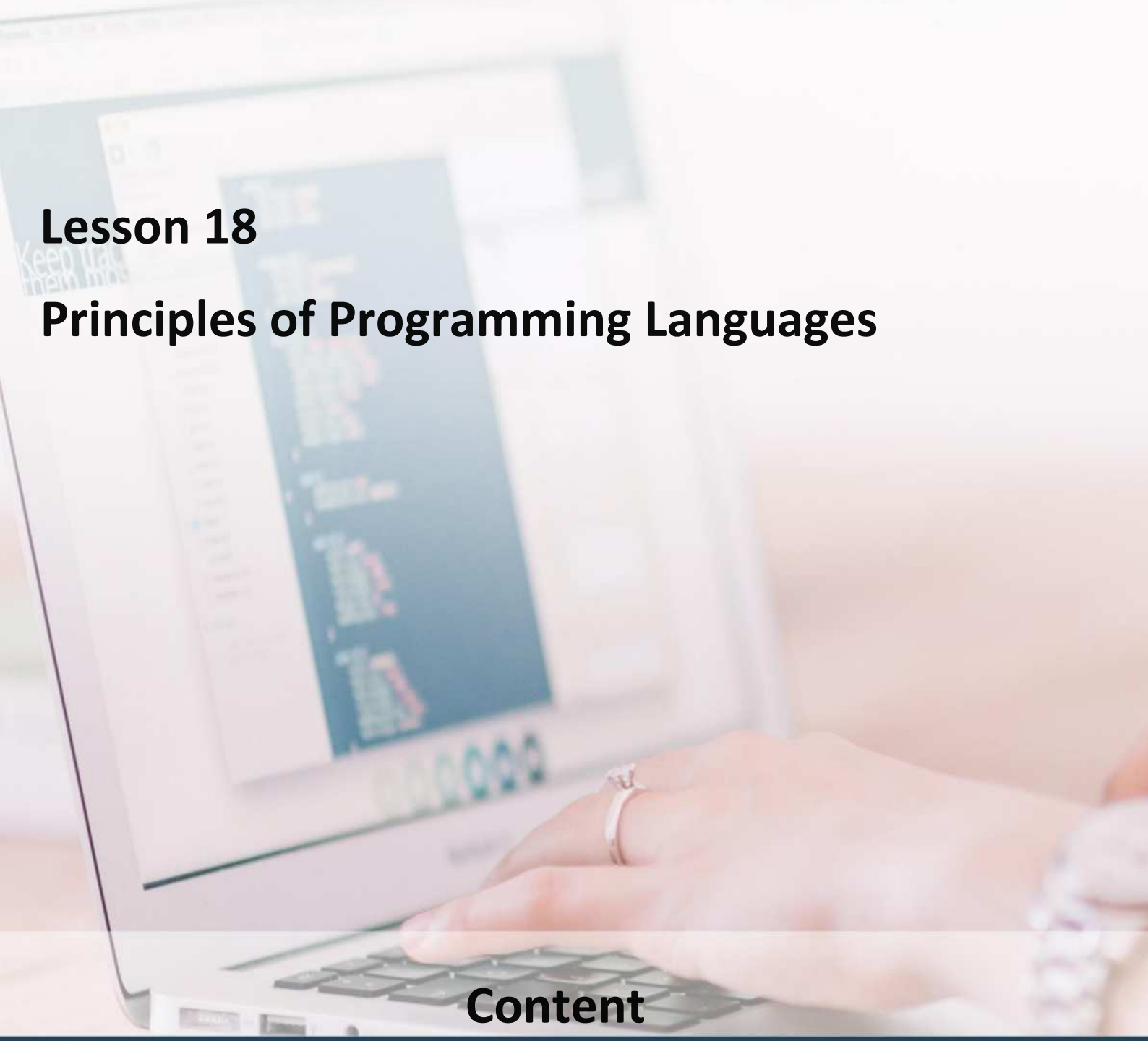
1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson.
2. Fundamentals of Computers by V Rajaraman.
3. https://tutorialpoints.com/

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Fundamentals of Computer Science by N

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application



Lesson 18

Principles of Programming Languages

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 18	
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

- To introduce the basic idea of Program Development
- To understand the evolution of programming languages.
- To introduce the principles and techniques involved in design and implementation of modern programming languages.
- To introduce the notations to describe the syntax and semantics of programming languages.

2.1 Program Development Life Cycle (PDLC)

Program Development Life Cycle (PDLC) is a systematic way of developing quality software. It provides an organized plan for breaking down the task of program development into manageable chunks, each of which must be successfully completed before moving on to the next phase.

The program development process is divided into the steps discussed below:

1. **Defining the Problem**

   The first step is to define the problem. In major software projects, this is a job for system analyst, who provides the results of their work to programmers in the form of a *program specification*. The program specification defines the data used in program, the processing that should take place while finding a solution, the format of the output and the user interface.

2. **Designing the Program**

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Program design starts by focusing on the main goal that the program is trying to achieve and then breaking the program into manageable components, each of which contributes to this goal. This approach of program design is called *top-bottom program design* or *modular programming*. The first step involve identifying *main routine*, which is the one of program's major activity. From that point, programmers try to divide the various components of the main routine into smaller parts called *modules*. For each module, programmer draws a conceptual plan using an appropriate program design tool to visualize how the module will do its assign job.

Program Design Tools:

The various program design tools are described below:

- **Structure Charts** – A *structure chart*, also called *Hierarchy chart*, show top-down design of program. Each box in the structure chart indicates a task that program must accomplish. dule, called the *Ma* or *Control module*. For exa

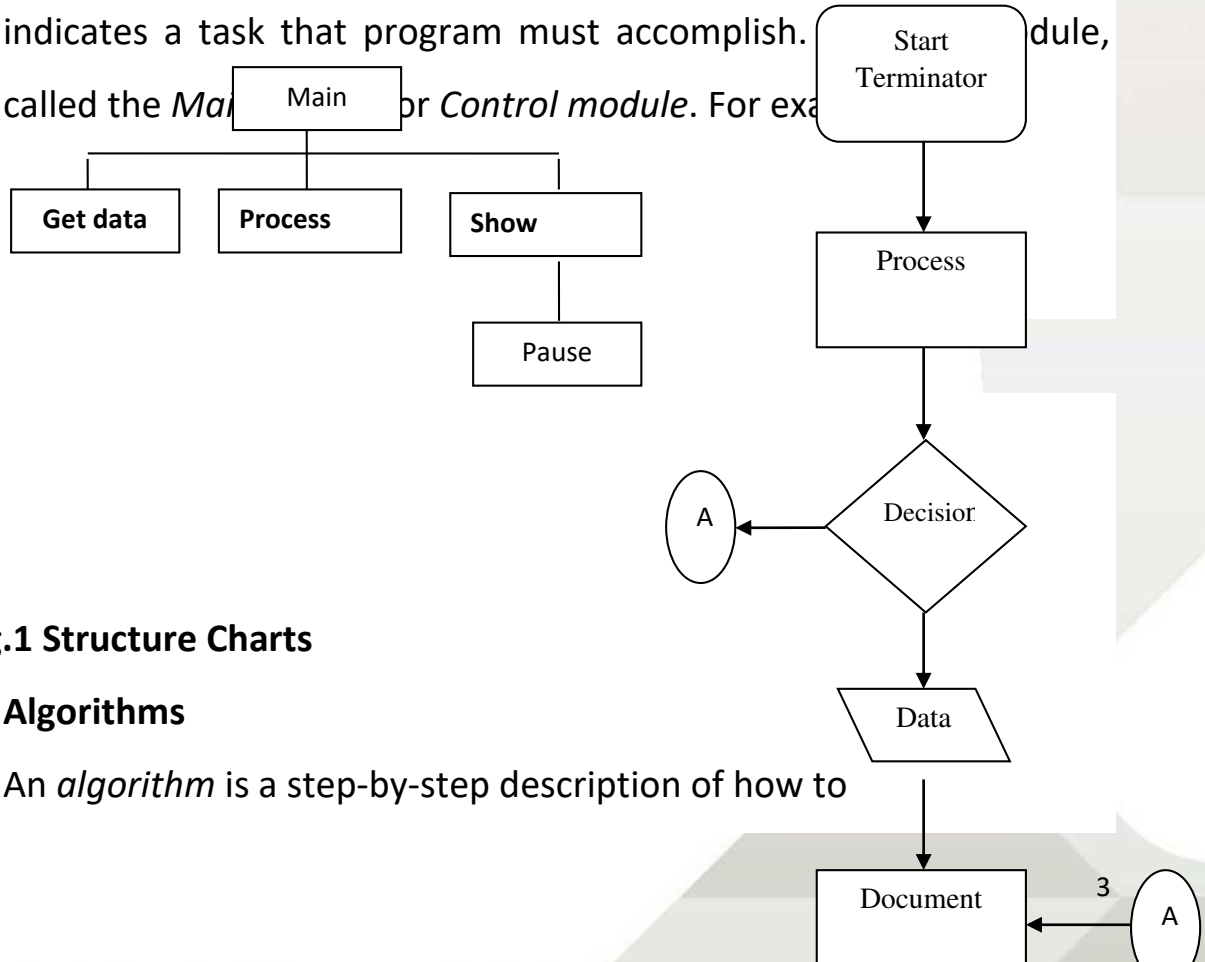


Fig.1 Structure Charts

- **Algorithms**

  An *algorithm* is a step-by-step description of how to

arrive at a solution in the easiest way. Algorithms are not restricted to computer world only. In fact, we use them in everyday life.

- **Flowcharts –**

  A *flowchart* is a diagram that shows the logic of the program. Flowchart is a graphical or symbolic representation of an algorithm. It is the diagrammatic representation of the step-by-step solution to a given problem. Program Design consists of the steps a programmer should do before they start coding the program in a specific language. Proper program design helps other programmers to maintain the program in the future.

- **Decision tables**

  A *Decision table* is a special kind of table, which is divided into four parts by a pair of horizontal and vertical lines.

- **Pseudocode**

  A *pseudocode* is another tool to describe the way to arrive at a solution. They are different from algorithm by the fact that they are expressed in program language like constructs.

3. **Coding the Program**

   Coding the program means translating an algorithm into specific programming language. The technique of programming using only well-defined control structures is known as *Structured programming*. Programmer must follow the language rules, violation of any rule causes *error*. These errors must be eliminated before going to the next step.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

4. **Testing and Debugging the Program**

   After removal of syntax errors, the program will execute. However, the output of the program may not be correct. This is because of logical error in the program. A logical error is a mistake that the programmer made while designing the solution to a problem. So the programmer must find and correct logical errors by carefully examining the program output using *Test data*. Syntax error and Logical error are collectively known as *Bugs*. The process of identifying errors and eliminating them is known as *Debugging*.

5. **Documenting the Program**

   After testing, the software project is almost complete. The *structure charts, pseudocodes, flowcharts* and *decision tables* developed during the design phase become documentation for others who are associated with the software project. This phase ends by writing a manual that provides an overview of the program's functionality, tutorials for the beginner, in-depth explanations of major program features, reference documentation of all program commands and a thorough description of the error messages generated by the program.

6. **Deploying and Maintaining the Program**

   In the final phase, the program is deployed (installed) at the user's site. Here also, the program is kept under watch till the user gives a green signal to it.

   Even after the software is completed, it needs to be maintained and evaluated regularly. In software maintenance, the programming team fixes program errors and updates the software.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

2.2 About Programming Languages:

A programming language is a set of commands, instructions, and other syntax use to create a software program. Languages that programmers use to write code are called "high-level languages." This code can be compiled into a "low-level language," which is recognized directly by the computer hardware.

High-level languages are designed to be easy to read and understand. This allows programmers to write source code in a natural fashion, using logical words and symbols. For example, reserved words like function, while, if, and else are used in most major programming languages. Symbols like <, >, ==, and!= are common operators. Many high-level languages are similar enough that programmers can easily understand source code written in multiple languages.

Examples of high-level languages include Fortran, BASIC, COBOL, C++, Java, Perl, and PHP. Languages like C++ and Java are called "compiled languages" since the source code must first be compiled in order to run. Languages like Perl and PHP are called "interpreted languages" since the source code can be run through an interpreter without being compiled. Generally, compiled languages are used to create software applications, while interpreted languages are used for running scripts, such as those used to generate content for dynamic websites.

Low-level languages include assembly and machine languages. An assembly language contains a list of basic instructions and is much more difficult to read than a high-level language. In rare cases, a programmer

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

may decide to code a basic program in an assembly language to ensure it operates as efficiently as possible. An assembler can be used to translate the assembly code into machine code. The machine code, or machine language, contains a series of binary codes that are understood directly by a computer's CPU. Needless to say, machine language is not designed to be human readable.

2.3 Classification of Programming Languages

Computer programming language can be classified into two major categories:

- Low Level
- High Level

Low Level Languages

The languages which use only primitive operations of the computer are known as low language. In these languages, programs are written by means of the memory and registers available on the computer. As we all know that the architecture of computer differs from one machine to another, so far each type of computer there is a separate low level programming language. In the other words, Programs written in one low level language of one, architectural can't be ported on any other machine dependent languages. Examples are **Machine Language** and **Assembly Language**.

Machine Language

In machine language program, the computation is based on binary numbers. All the instructions including operations, registers, data and

memory locations are given in there binary equivalent. The machine directly understands this language by virtue of its circuitry design so these programs are directly executable on the computer without any translations. This makes the program execution very fast. Machine languages are also known as first generation languages.

A typical low level instruction consists essentially of two parts:

Operation part Address Part

and

•**An Operation Part :** Specifies operation to be performed by the computer, also known as Opcode.

•**An Address Part :** Specifies location of the data on which operation is to be performed.

Advantages

Machine language makes most efficient use of computer system resources like storage, registers, etc. the instruction of a machine language program are directly executable so there is no need of translators. Machine language instruction can be used to manipulate the individual bits in a computer system with high execution speed due to direct manipulation of memory and registers.

Drawbacks

Machine languages are machine dependent and, therefore, programs are not portable from one computer to other. Programming in machine language usually results in poor programmer productivity. Machine languages require programmers to control the use of each register in the computer's Arithmetic Logic Unit and

computer storage locations must be addressed directly, not symbolically. Machine language requires a high level of programming skill which increases programmer training costs. Programs written in machine language are more error prone and difficult to debug because it is very difficult to remember all binary equivalent of register, opcode, memory location, etc. program size is comparatively very big due to non-use of reusable codes and use of very basic operations to do a complex computation.

Assembly language

Assembly languages are also known as second generation languages. These languages substitutes alphabetic or numeric symbols for the binary codes of machine language. That is, we can use mnemonics for all opcodes, registers and for the memory locations which provide us with a facility to write reusable code in the form of macros. Has two parts, one is macro name and the other is macro body which contains the line of instructions. A macro can be called at any point of the program by its name to use the instruction. A macro can be called at any point of the program by its name to use the instructions given in the macro repetitively.

These language require a translator known as “Assembler” for translating the program code written in assembly language to machine language. Because computer can interpret only the machine code instruction, once the translation is completed the program can be executed.

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Assembly language is the most basic programming language available for any processor. With assembly language, a programmer works only with operations implemented directly on the physical CPU. Assembly language lacks high-level conveniences such as variables and functions, and it is not portable between various families of processors. Nevertheless, assembly language is the most powerful computer programming language available, and it gives programmers the insight required to write effective code in high-level languages. Machine code for displaying $ sign on lower right corner of screen:

10111000, 00000000, 10111000, 10001110, 11011000, 11000110, 00000110, 10011110, 00001111, 00100100, 11001101, 00011111.

The program above, written in assembly language, looks like this:

```
MOV AX, 47104
MOV DS, AX
MOV [3998], 36
INT 32
```

Advantages

Assembly language provide optimal use of computer resources like registers and memory because of direct use of these resources within the programs. Assembly language is easier to use than machine language because there is no need to remember or calculate the binary equivalents for opcode and registers. An assembler is useful for detecting programming errors. Assembly language encourages modular programming which provides the facility of reusable code, using macro.

Drawbacks

Assembly language programs are not directly executable due to the need of translation. Also, these languages are machine dependent and, therefore, not portable from one machine to another. Programming in assembly language requires a high level of programming skills and knowledge of computer architecture of the particular machine.

High level languages (HLL)

All high level language are procedure-oriented language and are intended to be machine independent. Programs are written in statements akin to English language, a great advantage over mnemonics of assembly languages require languages use mnemonics of assembly language. That is, the high level languages use natural language like structures. These languages require translators (compilers and interpreters) for execution. The programs written in a high level language can be ported on any computer that is why they are known as machine independent. The early high level language come in third generation of languages, COBOL, BASIC, APL, etc. These languages enable the programmer to write instruction using English words and familiar mathematical symbols which makes it easier than technical details of the computer. It makes the programs more readable too.

Procedures

Procedures are the reusable code which can be called at any point of the program. Each procedure is defined by a name and set of instructions accomplishing a particular task. The procedure can be called by its name with the list of required parameters which should pass to that procedure.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Advantages of High Level Languages

These are the third generation languages. These are procedure-oriented languages and are machine independent. Programs are written in English like statements. As high level languages are not directly executable, translators (compilers and interpreters) are used to convert them in machine language equivalent.

Advantages

1) These are easier to learn than assembly language.
2) Les time is required to write programs.
3) These provide better documentation.
4) These are easier to maintain.
5) These have an extensive vocabulary.

Limitation of Programming language

1) A long sequence statement is to be written for every program.
2) Additional memory space is required for storing compiler or interpreter.
3) Execution time is very high as the HLL programs are not directly executable.

Compiler:

It is a program which translates a high level language program into a machine language program. A compiler is more intelligent than an assembler. It checks all kinds of limits, ranges, errors etc. But its program run time is more and occupies a larger part of the memory. It has slow speed. Because a compiler goes through the entire program and then

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

translates the entire program into machine codes. If a compiler runs on a computer and produces the machine codes for the same computer then it is known as a self compiler or resident compiler. On the other hand, if a compiler runs on a computer and produces the machine codes for other computer then it is known as a cross compiler.

When a user writes a code in a high level language such as Java and wants it to execute, a specific compiler which is designed for Java is used before it will be executed. The compiler scans the entire program first and then translates it into machine code which will be executed by the computer processor and the corresponding tasks will be performed.

Assembler

A computer will not understand any program written in a language, other than its machine language. The programs written in other languages must be translated into the machine language. Such translation is performed with the help of software. A program which translates an assembly language program into a machine language program is called an assembler. If an assembler which runs on a computer and produces the machine codes for the same computer then it is called self assembler or resident assembler. If an assembler that runs on a computer and produces the machine codes for other computer then it is called Cross Assembler.

Assemblers are further divided into two types:

- One Pass Assembler and
- Two Pass Assembler.

One pass assembler is the assembler which assigns the memory addresses to the variables and translates the source code into machine code in the first pass simultaneously.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

A Two Pass Assembler is the assembler which reads the source code twice. In the first pass, it reads all the variables and assigns them memory addresses. In the second pass, it reads the source code and translates the code into object code.

Interpreter

An interpreter is a program which translates statements of a program into machine code. It translates only one statement of the program at a time. It reads only one statement of program, translates it and executes it. Then it reads the next statement of the program again translates it and executes it. In this way it proceeds further till all the statements are translated and executed. On the other hand, a compiler goes through the entire program and then translates the entire program into machine codes. A compiler is 5 to 25 times faster than an interpreter.

By the compiler, the machine codes are saved permanently for future reference. On the other hand, the machine codes produced by interpreter are not saved. An interpreter is a small program as compared to compiler. It occupies less memory space, so it can be used in a smaller system which has limited memory space.

Difference between Compiler and Assembler

(a) Compiler is a computer program that reads a program written in one language and translates it in to another language, while an assembler can be considered a special type of compiler which translates only Assembly language to machine code.

(b) Compilers usually produce the machine executable code directly from a high level language, but assemblers produce an object code

which might have to be linked using linker programs in order to run on a machine. Because Assembly language has a one to one mapping with machine code, an assembler may be used for producing code that runs very efficiently for occasions in which performance is very important (for e.g. graphics engines, embedded systems with limited hardware resources compared to a personal computer like microwaves, washing machines, etc.).

The main differences between compiler and interpreter are listed below:

a) The interpreter takes one statement then translates it and executes it and then takes another statement. While the compiler translates the entire program in one go and then executes it.

b) Compiler generates the error report after the translation of the entire page while an interpreter will stop the translation after it gets the first error.

c) Compiler takes a larger amount of time in analyzing and processing the high level language code comparatively interpreter takes lesser time in the same process.

d) Besides the processing and analyzing time the overall execution time of a code is faster for compiler relative to the interpreter.

Algorithms

A computer is a useful tool for solving a great variety of problems. To make a computer do anything (i.e. solve a problem), we have to write a computer program. In a computer program, we tell a computer, step by step, exactly what we want it to do. The computer then executes the program, following each step mechanically, to accomplish the end goal.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The sequence of steps to be performed in order to solve a problem by the computer is known as an algorithm.

In mathematics, computer science, and related subjects, an algorithm is a finite sequence of steps expressed for solving a problem. An algorithm can be defined as “a process that performs some sequence of operations in order to solve a given problem”. Algorithms are used for calculation, data processing, and many other fields. In computing, algorithms are essential because they serve as the systematic procedures that computers require. A good algorithm is like using the right tool in the workshop. It does the job with the right amount of effort. Using the wrong algorithm or one that is not clearly defined is like trying to cut a piece of plywood with a pair of scissors: although the job may get done, we have to wonder how effective we were in completing it..

In computer programming, there are often many different algorithms to accomplish any given task. Each algorithm has advantages and disadvantages in different situations. Sorting is one place where a lot of research has been done, because computers spend a lot of time sorting lists. Three reasons for using algorithms are efficiency, abstraction and reusability.

Efficiency: Certain types of problems, like sorting, occur often in computing. Efficient algorithms must be used to solve such problems considering the time and cost factor involved in each algorithm.

Abstraction: Algorithms provide a level of abstraction in solving problems because many seemingly complicated problems can be distilled into simpler ones for which well known algorithms exist. Once we see a

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

more complicated problem in a simpler light, we can think of the simpler problem as just an abstraction of the more complicated one. For example, imagine trying to find the shortest way to route a packet between two gateways in an internet. Once we realize that this problem is just a variation of the more general shortest path problem, we can solve it using the generalised approach.

Reusability: Algorithms are often reusable in many different situations. Since many well-known algorithms are the generalizations of more complicated ones, and since many complicated problems can be distilled into simpler ones, an efficient means of solving certain simpler problems potentially lets us solve many complicated problems.

Expressing Algorithms:

Algorithms can be expressed in many different notations, including natural languages, pseudo code, flowcharts and programming languages. Natural language expressions of algorithms tend to be verbose and ambiguous, and are rarely used for complex or technical algorithms. Pseudo code and flowcharts are structured ways to express algorithms that avoid many ambiguities common in natural language statements, while remaining independent of a particular implementation language. Programming languages are primarily intended for expressing algorithms in a form that can be executed by a computer, but are often used to define or document algorithms. Sometimes it is helpful in the description of an algorithm to supplement small flowcharts with natural language and/or arithmetic expressions written inside block diagrams to summarize what

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

the flowcharts are accomplishing. Consider an example for finding the largest number in an unsorted list of numbers.

The solution for this problem requires looking at every number in the list, but only once at each.

1) *Algorithm using natural language statements*:

a) Assume the first item is largest.

b) Look at each of the remaining items in the list and if it is larger than the largest item so far, make a note of it.

c) The last noted item is the largest item in the list when the process is complete.

2) Algorithm using pseudocode:

largest= L_0

for each *item* in the list (*Length(L)≥1*), **do**

ifthe *item* ≥largest, **then**

largest = the *item*

return*largest*

Flowcharts

A Flowchart is a type of diagram (graphical or symbolic) that represents an algorithm or process. Each step in the process is represented by a different symbol and contains a short description of the process step. The flow chart symbols are linked together with arrows showing the process flow direction. A flowchart typically shows the flow of data in a process, detailing the operations/steps in a Pictorial format which is easier to understand than reading it in a textual format.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

A flowchart describes what operations (and in what sequence)are required to solve a given problem. A flowchart can be likened to the blueprint of a building. As we know a designer draws a blue print before starting construction on a building. Similarly, a programmer prefers to draw a flowchart prior to writing a computer program. Flowcharts are a pictorial or graphical representation of a process. Flowcharts are used in analyzing, designing, documenting or managing a process or program in various fields. Flowcharts are generally drawn in the early stages of formulating computer solutions. Often we see how flowcharts are helpful in explaining the program to others. Hence, it is correct to say that a flowchart is a must for the better documentation of a complex program.

For example, consider that we need to find the sum, average and product of 3 numbers given by the user.

Algorithm for the given problem is as follows:

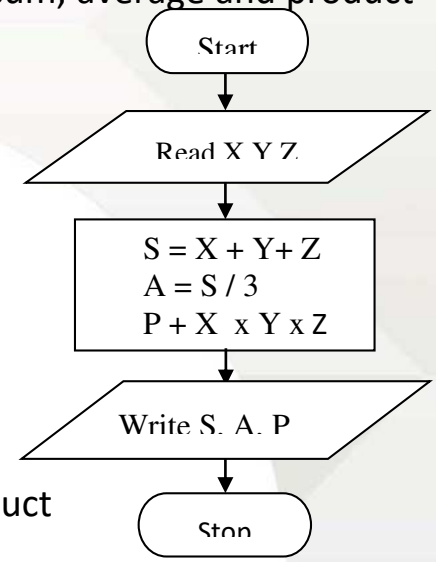
Read X, Y, Z

Compute Sum (S) as X + Y + Z

Compute Average (A) as S / 3

Compute Product (P) as X × Y × Z

Write (Display) the Sum, Average and Product



Flowchart for the above problem

Advantages of using flowcharts

The *benefits of flowcharts* are as follows:

1. **Communication:** Flowcharts are better way of communicating the logic of a system to all concerned.
2. **Effective analysis:** With the help of flowchart, problem can be analysed in more effective way.
3. **Proper documentation:** Program flowcharts serve as a good program documentation, which is needed for various purposes.
4. **Efficient Coding:** The flowcharts act as a guide or blueprint during the systems analysis and program development phase.
5. **Proper Debugging**: The flowchart helps in debugging process.
6. **Efficient Program Maintenance:** The maintenance of operating program becomes easy with the help of flowchart. It helps the programmer to put efforts more efficiently on that part.

Limitations of using flowcharts

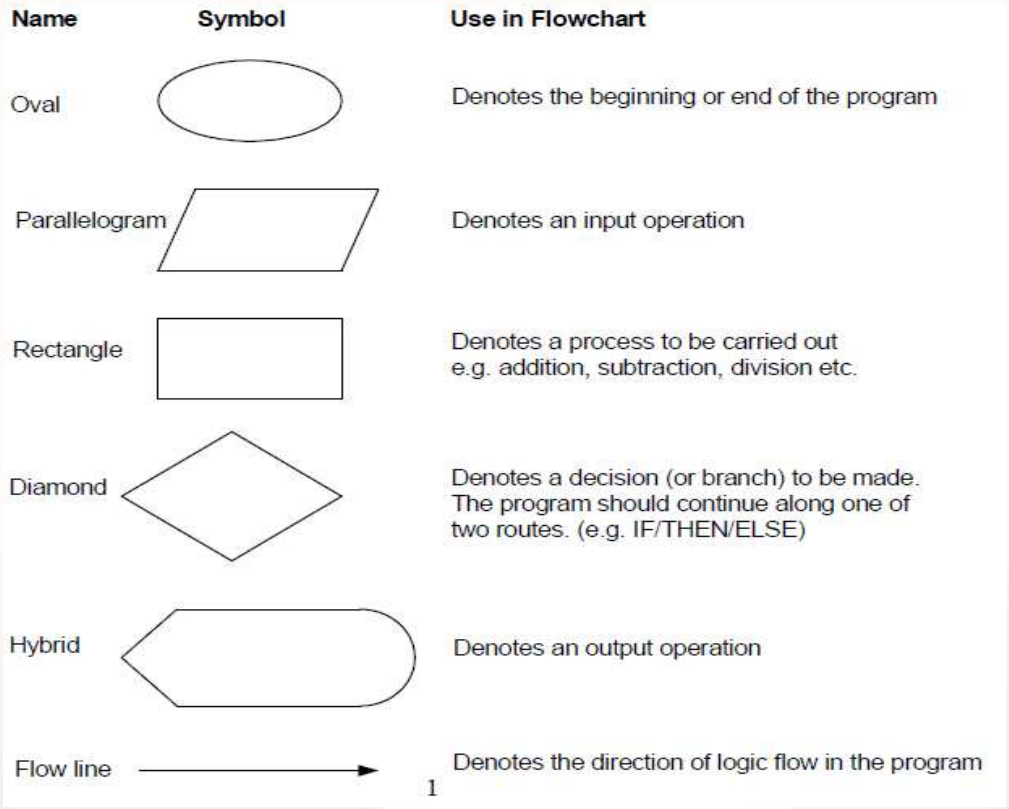
Although a flowchart is a very useful tool, there are a few limitations in using flowcharts which are listed below:

1. **Complex logic:** Sometimes, the program logic is quite complicated. In that case, flowchart becomes complex and clumsy.
2. **Alterations and Modifications**: If alterations are required the flowchart may require re-drawing completely.
3. **Reproduction:** As the flowchart symbols cannot be typed, reproduction of flowchart becomes a problem.
4. The essentials of what is done can easily be lost in the technical details of how it is done.

Flowchart symbols & guidelines:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Flowcharts are usually drawn using some standard symbols; however, some special symbols can also be developed when required. Some standard symbols, which are frequently required for flowcharting



many computer programs are shown.

Program for to calculate equation

$ax^2 + bx+c$

hint: d = sqt (), and the roots are

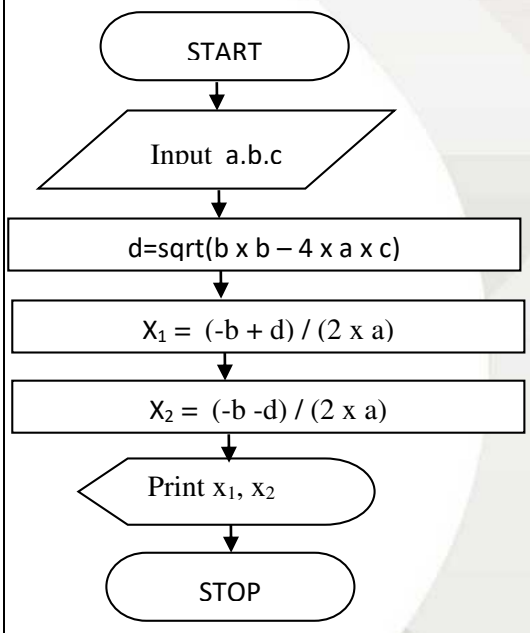
$x_1 = (b+d)/2a$

and $x_2 = (-b-d)/2a$

Algorithm:

Step1: Input the coefficients (a,b,c) of the quadratic equation

Step 2 calculate d using sqrt()



Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Step 3 calculate x_1 using x_1 (-b+d) / (2 x a) Step 4 calculate x_2 using x_2 (-b-d) / (2 x a) Step 5 Print x_1 and x_2	

Flow chart for Quadratic Equation:

For Following Algorithms draw the flowcharts:

1. **Algorithm to swap the contents of two variables**.

   Step 1: [Input 2 numbers]

   Read a, b;

   Step 2: [Exchange a & b]

   Temp=a;

   a=b;

   b=temp;

   Step 3: [Output the result]

   Display a b.

   Step 4: [Terminate]

   Stop.

2. **Algorithm to check whether a given number is even or odd**.

   Step 1: [Input the number].

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Read n.

Step 2: [check for remainder when divided by 2]

Check condition (n%2==0) then

If true display “The number is even”.

Else

If false display “the number is odd”.

Step 3: [Terminate]

Stop.

3. **Algorithm to check whether a given number is positive or negative**.

Step 1: [Input the number].

Read n.

Step 2: [Check whether the number is ‘+’ veor’ –‘ ve].

Check condition (n<=0) then

If true Display “The number is negative”

Else

If false Display “the number is positive”.

Step 3: [Terminate]

Stop.

4. **Algorithm to find the area of the triangle**.

Step 1: [Input 3 sides of the triangle].

Read a,b,c.

Step 2: [Compute the value of s].

S = (a+b+c)/2.

Step 3: [Compute the area of the triangle].

Area= (s*$\sqrt{(s-a)*(s-b)*(s-c)}$)

Designed and developed under the aegis of NAHEP Component-2 Project “Investments In ICAR Leadership In Agricultural Higher Education”
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Step 4: [Output the result].

Display area.

Step 5: [Terminate].

Stop.

5. **Algorithm to find the factorial of a given number**.

Step 1: [Read the number]

Read n.

Step 2: [Initialization]

Factorial =1.

Step 3: [Find the factorial of n]

For i<- 1 to n in steps of 1 do

factorial = factorial * i.

End for.

Step 4: [Output the result]

Display factorial.

Step 5: [Terminate]

Stop.

6. **Algorithm to find the sum of first N natural numbers**.

Step 1: [Input the number of terms]

Read n.

Step 2: [Initialization]

Sum=0

Step 3: [Find the sum of all terms]

For i<- 1 to n in steps of 1 do

sum = sum + i

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

end for.

Step 4: [output the result]

Display sum.

Step 5: [Terminate]

Stop.

7. **Algorithm to find the biggest of 3 numbers**.

Step 1: [Read the numbers]

Read a,b,c.

Step 2: [Comparision]

Check condition (a>b) and (a>c) then

display a is greatest.

else check if (b>c) then

display b is greatest.

else display c is greatest.

Step 3: [Terminate]

Stop.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

GLOSSARY

Program; A set of instructions is called a program.

Algorithm: Step by step procedure for solving a problem using computer.

Flowchart: Graphical or pictorial representation of algorithm is called flow chart.

Coding: Coding means translating an algorithm into specific programming language.

Low level language: Low-level languages include assembly and machine languages.

The machine language: Only language understood by the computer. It contains a series of binary codes that are understood directly by a computer's CPU. Needless to say, machine language is not designed to be human readable.

High level languages (HLL): All high level language are procedure-oriented language and are intended to be machine independent. Programs are written in statements like English language,

Compiler: A kind of system software which translate HLL to machine language.

Assembler: A kind of system software which translate assembly language to machine

language.

Interpreter: A kind of system software which translate HLL language to machine

language.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

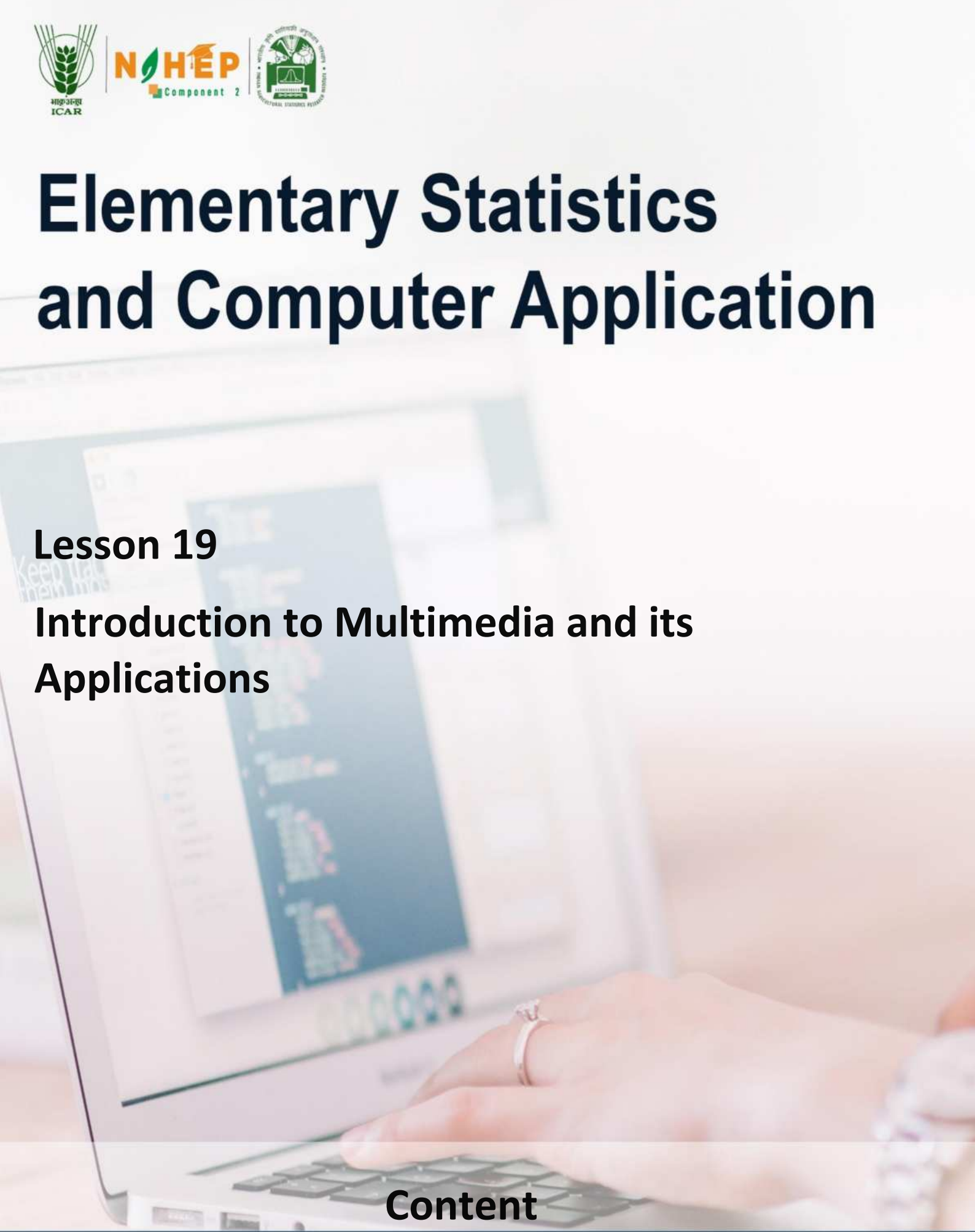
REFERENCES

1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson.
2. Introduction to Information Technology, O-Level by Satish Jain.
3. Fundamentals of Computers by V Rajaraman.
4. https://tutorialpoints.com/

Elementary Statistics and Computer Application

Lesson 19

Introduction to Multimedia and its Applications



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 19	Introduction to Multimedia and its Applications
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

- To introduce the students the basics of multimedia.
- To help the students to develop higher-order thinking skills.
- To help the students to identify and solve problems more easily compared to the scenario where teaching is made possible only by textbooks.

Definition of Multimedia

Multimedia as name suggests is the combination of Multi and Media that is many types of media (hardware/software) used for communication of information.

Definition: Multimedia is a representation of information in an attractive and interactive manner with the use of a combination of text, audio, video, graphics and animation.

In other words, it is a computerized method of presenting information combining textual data, audio, visuals (video), graphics and animations. For examples: E-Mail, Yahoo Messenger, Video Conferencing, and Multimedia Message Service (MMS).

Multimedia Hardware

Most of the computers now-a-days come equipped with the hardware components required to develop/view multimedia applications. Following are the various categories in which we can define the various types of hardware required for multimedia applications.

- **Processor** The heart of any multimedia computer is its processor. Today Core i5 or higher processor is recommended for a multimedia computer.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Memory and Storage Devices** – It is needed for storing various files used during production, original audio and video clips, edited pieces and final mined pieces. It is also needed for backup of any project files.

There are three types memory

- **Primary memory**
- **Flash Memory**-
- **Secondary Memory:**

- **Input Devices** - Following are the various types of input devices which are used in multimedia systems.
  - **Keyboard**- Most common and very popular input device is keyboard. The keyboard helps in inputting the data to the computer. Keyboards are of two sizes 84 keys or 101/102 keys, but now 104 keys or 108 keys keyboard is also available for Windows and Internet.

  - **Mouse** - Mouse is most popular Pointing device. It is a very famous cursor-control device. Generally, it has two buttons called left and right button and scroll bar is present at the mid. Mouse can be used to control the position of cursor on screen, but it cannot be used to enter text into the computer.

  - **Joystick** - Joystick is also a pointing device, which is used to move cursor position on a monitor screen. It is a stick having a spherical ball at its both lower and upper ends. The joystick can be moved in all four directions. The function of joystick is similar to that of a mouse. It is mainly used in Computer Aided Designing (CAD) and playing computer games.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Light Pen** - Light pen is a pointing device, which is similar to a pen. It is used to select a displayed menu item or draw pictures on the monitor screen.
- **Track Ball** - Track ball is an input device that is mostly used in notebook or laptop computer, instead of a mouse. A track ball comes in various shapes like a ball, a button and a square.

- **Scanner** - Scanner is an input device, which works more like a photocopy machine. It is used when some information is available on a paper and it is to be transferred to the hard disc of the computer for further manipulation.

- **Digitizer** - Digitizer is an input device, which converts analog information into a digital form. Digitizer can convert a signal from the television camera into a series of numbers that could be stored in a computer. They can be used by the computer to create a picture of whatever the camera had been pointed at.

- **Magnetic Ink Card Reader (MICR)** - MICR input device is generally used in banks because of a large number of cheques to be processed every day.
- **Optical Character Reader (OCR)** - OCR is an input device used to read a printed text. OCR scans text optically character by character, converts them into a machine readable code and stores the text on the system memory.

- **Bar Code Readers** - Bar Code Reader is a device used for reading bar coded data (data in form of light and dark lines). Bar coded data is generally used in labelling goods, numbering the books, etc.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Optical Mark Reader (OMR)** - OMR is a special type of optical scanner used to recognize the type of mark made by pen or pencil. It is used where one out of a few alternatives is to be selected and marked. It is specially used for checking the answer sheets of examinations having multiple choice questions.

- **Voice Systems** - Following are the various types of input devices which are used in multimedia systems.
    - **Microphone**- The microphone is used for various applications like adding sound to a multimedia presentation or for mixing music.

    - **Speaker**- Speaker is an output device to produce sound which is stored in digital form. The speaker is used for various applications like adding sound to a multimedia presentation or for movies displays etc.

- **Digital Camera** - Digital camera is an input device to input images that is then stored in digital form. The digital camera is used for various applications like adding images to a multimedia presentation or for personal purposes.

- **Digital Video Camera** - Digital Video camera is an input device to input images/video that is then stored in digital form. The digital video camera is used for various applications like adding videos to a multimedia presentation or for personal purposes.

Output Devices - Following are few of the important output devices, which are used in Computer Systems:

- **Monitors** - Monitor commonly called as Visual Display Unit (VDU) is the main output device of a computer. It forms images from tiny dots, called pixels, which are arranged in a rectangular form.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The sharpness of the image depends upon the number of the pixels. There are two kinds of viewing screen used for monitors:

- Cathode-Ray Tube (CRT) Monitor-
- Flat-Panel Display Monitor-
- Example is LCD (Liquid-Crystal Device)

- **Printers** - Printer is the most important output device, which is used to print information on paper.

- **Dot Matrix Printer**-.
- **Daisy Wheel**-

- **Line Printers**- Line printers are printers, which print one line at a time.

- **Laser Printers**- These are non-impact page printers. They use laser lights to produce the dots needed to form the characters to be printed on a page.
- **Inkjet Printers**- Inkjet printers are non-impact character printers based on a relatively new technology. They print characters by spraying small drops of ink onto paper. Inkjet printers produce high quality output with presentable features.

- **Screen Image Projector** - Screen image projector or simply projector is an output device used to project information from a computer on a large screen so that a group of people can see it simultaneously. A presenter first makes a PowerPoint presentation on the computer. Now a screen image projector is plugged to a computer system and presenter can make a presentation to a group of

people by projecting the information on a large screen. Projector makes the presentation more understandable.

- Speakers **and Sound Card** - Computers need both a sound card and speakers to hear audio, such as music, speech and sound effects.

Multimedia Software

Multimedia software tells the hardware what to do. For example, multimedia software tells the hardware to display the color blue, play the sound of cymbals crashing etc.

To produce these media elements(movies, sound, text, animation, graphics etc.) there are various software available in the market such as Paint Brush, Photo Finish, Animator, Photo Shop, 3D Studio, Corel D raw, Sound Blaster, IMAGINET, Apple Hyper Card, Photo Magic, Picture Publisher.

Multimedia Software Categories

Following are the various categories of Multimedia software

- **Device Driver Software**- These software's are used to install and configure the multimedia peripherals.
- **Media Players**- Media players are applications that can play one or more kind of multimedia file format.
- **Media Conversion Tools**- These tools are used for encoding / decoding multimedia contexts and for converting one file format to another.
- **Multimedia Editing Tools**- These tools are used for creating and editing digital multimedia data.
- **Multimedia Authoring Tools**- These tools are used for combing different kinds of media formats and deliver them as multimedia contents.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Tools for Creation of Multimedia Application:

Multimedia applications are created with the help of following mentioned tools and packages.

The sound, text, graphics, animation and video are the integral part of multimedia software. To produce and edit these media elements, there are various software tools available in the market. The categories of basic software tools are:

- **Text Editing Tools**- These tools are used to create letters, resumes, invoices, purchase orders, user manual for a project and other documents. MS-Word is a good example of text tool. It has following features:
    - Creating new file, opening existing file, saving file and printing it.
    - Insert symbol, formula and equation in the file.
    - Correct spelling mistakes and grammatical errors.
    - Align text within margins.
    - Insert page numbers on the top or bottom of the page.
    - Mail-merge the document and making letters and envelopes.
    - Making tables with variable number of columns and rows.
- **Painting and Drawing Tools**- These tools generally come with a graphical user interface with pull down menus for quick selection. We can create almost all kinds of possible shapes and resize them using these tools. Drawing file can be imported or exported in many image formats like .gif, .tif, .jpg, .bmp, etc. Some examples of drawing software are Corel Draw, Freehand, Designer, Photoshop, Fireworks, Point etc.

  These software have following features:
    - Tools to draw a straight line, rectangular area, circle etc.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Different colour selection option.
- Pencil tool to draw a shape freehand.
- Eraser tool to erase part of the image.
- Zooming for magnified pixel editing.

- **Image Editing Tools**- Image editing tools are used to edit or reshape the existing images and pictures. These tools can be used to create an image from scratch as well as images from scanners, digital cameras, clipart files or original artwork files created with painting and drawing tools. Examples of Image editing or processing software are Adobe Photoshop and Paint Shop Pro.
- **Sound Editing Tools**- These tools are used to integrate sound into multimedia project very easily. We can cut, copy, paste and edit segments of a sound file by using these tools. The presence of sound greatly enhances the effect of a mostly graphic presentation, especially in a video. Examples of sound editing software tools are: Cool Edit Pro, Sound Forge and Pro Tools. This software has following features:
  - Record our own music, voice or any other audio.
  - Record sound from CD, DVD, Radio or any other sound player.
  - We can edit, mix the sound with any other audio.
  - Apply special effects such as fade, equalizer, echo, reverse and more.
- **Video Editing Tools**- These tools are used to edit, cut, copy, and paste video and audio files. Video editing used to require expensive, specialized equipment and a great deal of knowledge. The artistic process of video editing consists of deciding what elements to retain, delete or combine from various sources so that they come together in an organized, logical and visually planning manner. Today computers are powerful enough to handle this job, disk space

is cheap and storing and distributing our finished work on DVD is very easy. Examples of video editing software are Adobe Premiere and Adobe After Effects.

- **Animation and Modelling Tools**- An animation is to show the still images at a certain rate to give it visual effect with the help of Animation and modelling tools. These tools have features like multiple windows that allow us to view our model in each dimension, ability to drag and drop primitive shapes into a scene, colour and texture mapping, ability to add realistic effects such as transparency, shadowing and fog etc. Examples of Animations and modelling tools are 3D studio max and Maya.

Components of Multimedia

Following are the common components of multimedia:

- **Text**- All multimedia productions contain some amount of text. The text can have various types of fonts and sizes to suit the profession presentation of the multimedia software.
- **Graphics**- Graphics makes the multimedia application attractive. In many cases people do not like reading large amount of textual matter on the screen. Therefore, graphics are used more often than text to explain a concept, present background information etc. There are two types of Graphics:
    - **Bitmap images**- Bitmap images are real images that can be captured from devices such as digital cameras or scanners. Generally bitmap images are not editable. Bitmap images require a large amount of memory.
    - **Vector Graphics**- Vector graphics are drawn on the computer and only require a small amount of memory. These graphics are editable.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- **Audio**- A multimedia application may require the use of speech, music and sound effects. These are called audio or sound element of multimedia. Speech is also a perfect way for teaching. Audio are of analog and digital types. Analog audio or sound refers to the original sound signal. Computer stores the sound in digital form. Therefore, the sound used in multimedia application is digital audio.
- **Video**- The term video refers to the moving picture, accompanied by sound such as a picture in television. Video element of multimedia application gives a lot of information in small duration of time. Digital video is useful in multimedia application for showing real life objects. Video have highest performance demand on the computer memory and on the bandwidth if placed on the internet. Digital video files can be stored like any other files in the computer and the quality of the video can still be maintained. The digital video files can be transferred within a computer network. The digital video clips can be edited easily.
- **Animation**- Animation is a process of making a static image look like it is moving. An animation is just a continuous series of still images that are displayed in a sequence. The animation can be used effectively for attracting attention. Animation also makes a presentation light and attractive. Animation is very popular in multimedia application

Applications of Multimedia

Following are the common areas of applications of multimedia.

- **Multimedia in Business**- Multimedia can be used in many applications in a business. The multimedia technology along with communication technology has opened the door for information of global wok groups. Today the team members may be working anywhere and can work for various companies. Thus the work place

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

will become global. The multimedia network should support the following facilities:

- Voice Mail
- Electronic Mail
- Multimedia based FAX
- Office Needs
- Employee Training
- Sales and Other types of Group Presentation
- Records Management

- **Multimedia in Marketing and Advertising**- By using multimedia marketing of new products can be greatly enhanced. Multimedia boost communication on an affordable cost opened the way for the marketing and advertising personnel. Presentation that have flying banners, video transitions, animations, and sound effects are some of the elements used in composing a multimedia based advertisement to appeal to the consumer in a way never used before and promote the sale of the products.
- **Multimedia in Entertainment**- By using multimedia marketing of new products can be greatly enhanced. Multimedia boost communication on an affordable cost opened the way for the marketing and advertising personnel. Presentation that have flying banners, video transitions, animations, and sound effects are some of the elements used in composing a multimedia based advertisement to appeal to the consumer in a way never used before and promote the sale of the products.
- **Multimedia in Education**- Many computer games with focus on education are now available. Consider an example of an educational game which plays various rhymes for kids. The child can paint the pictures, increase reduce size of various objects etc apart from just

playing the rhymes. Several other multimedia packages are available in the market which provides a lot of detailed information and playing capabilities to kids.

- **Multimedia in Bank**- Every bank has a lot of information which it wants to impart to in customers. For this purpose, it can use multimedia in many ways. Bank also displays information about its various schemes on a PC monitor placed in the rest area for customers. Today on-line and internet banking have become very popular. These use multimedia extensively. Multimedia is thus helping banks give service to their customers and also in educating them about banks attractive finance schemes.
- **Multimedia in Hospital**- Multimedia best use in hospitals is for real time monitoring of conditions of patients in critical illness or accident. The conditions are displayed continuously on a computer screen and can alert the doctor/nurse on duty if any changes are observed on the screen. Multimedia makes it possible to consult a surgeon or an expert who can watch an ongoing surgery line on his PC monitor and give online advice at any crucial juncture.
- **Multimedia Pedagogues**- Pedagogues are useful teaching aids only if they stimulate and motivate the students. The audio-visual support to a pedagogue can actually help in doing so.
- **Communication Technology and Multimedia Services**- These services may include:
    - Basic Television Services
    - Interactive entertainment
    - Digital Audio
    - Video on demand
    - Home shopping
    - Financial Transactions
    - Interactive multiplayer or single player games
    - Digital multimedia libraries

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- E-Newspapers, e-magazines

Glossary :

Multimedia : Multimedia is a representation of information in an attractive and interactive manner with the use of a combination of text, audio, video, graphics and animation.

Processor: The heart of any multimedia computer is its processor.

Digital Camera: Digital camera is an input device to input images that is then stored in digital form.

Text: All multimedia productions contain some amount of text. The text can have various types of fonts and sizes to suit the profession presentation of the multimedia software.

Graphics: Graphics are used more often than text to explain a concept, present background information etc. There are two types of Graphics.

Animation and Modeling Tools: An animation is to show the still images at a certain rate to give it visual effect with the help of Animation and modeling to

REFERENCES

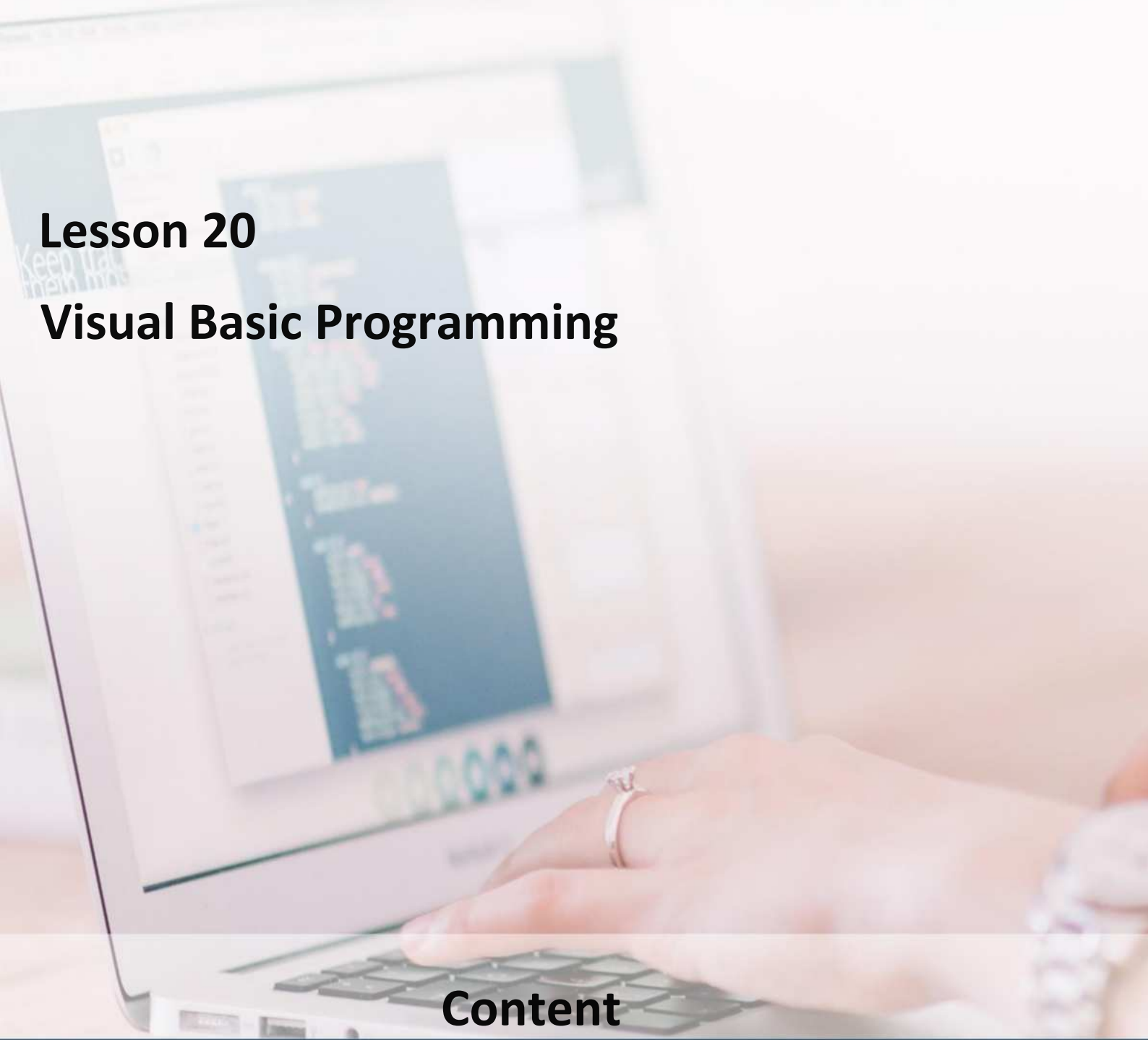
1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson.
2. Fundamentals of Computers by V Rajaraman.
3. https://tutorialpoints.com/
4. Fundamentals of Computer Science by N Subramaniam.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Elementary Statistics and Computer Application

Lesson 20

Visual Basic Programming



Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 20	Visual Basic Programming
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

- An overview of visual basic programming.
- Features of visual basic programming.
- History of visual basic programming.
- How visual basic works with demonstration.

Introduction:

Visual Basic (VB) is a simple, modern, object-oriented and type-safe programming language. Visual Basic language has its roots from the family of C languages such as C, C++ and it is mostly similar to Java programming.

VB Programming language will allow developers to build a variety of secure and robust applications such as windows applications, web applications, database applications, etc. which will run on **.NET Framework**.

.NET Framework is a development platform for building apps for windows, web, azure, etc. by using programming languages such as C#, F# and Visual Basic. It consists of two major components such as **Common Language Runtime** (CLR), it's an execution engine that handles running apps and **.NET Framework Class Library**, which provides a library of tested and reusable code that developer, can use it in their applications.

Overview of Visual Basic

- VB is an object-oriented programming language and it supports the concepts of encapsulation, abstraction, polymorphism, etc.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- In VB all the variables, methods and application's entry point are encapsulated within the class definitions.
- VB is developed specifically for .NET Framework and it enables programmers to migrate from C/C++ and Java easily.
- VB is a fully Event-driven and visual programming language.
- Microsoft provided an IDE (Integrated Development Environment) tool called Visual Studio to implement VB programs easily.

Features of Visual Basic)

- VB contains various features that make it similar to other programming languages such as C, C++, and Java.
- VB is a modern programming language and it is very powerful, simple for building the applications
- VB is useful in developing windows, web and device applications.
- VB provides automatic memory management by clearing unused objects
- VB is a type-safe programming language and it makes impossible to perform unchecked type casts.
- VB provides a structured and extensible approach for error detection and recovery. It is a structure-oriented programming language and the compilation, execution of VB applications are faster due to automatic scalability.

History of Visual Basic

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The initial release of VB programming language is in 2002 with .NET Framework 1.0 and it's more like Java programming. The latest version of VB is **15.8** and it was released with Visual Studio 2017 with a lot of new features.

Visual Basic Setup Development Environment

VB programming language has built on **.NET Framework** to build a variety of secure and robust applications such as windows, web or database applications based on our requirements.

To run VB applications, we require to install a **.NET Framework** component on our machines. In case, if we are using a Windows operating system, then by default **.NET Framework** installed on our machine.

Visual Studio IDE

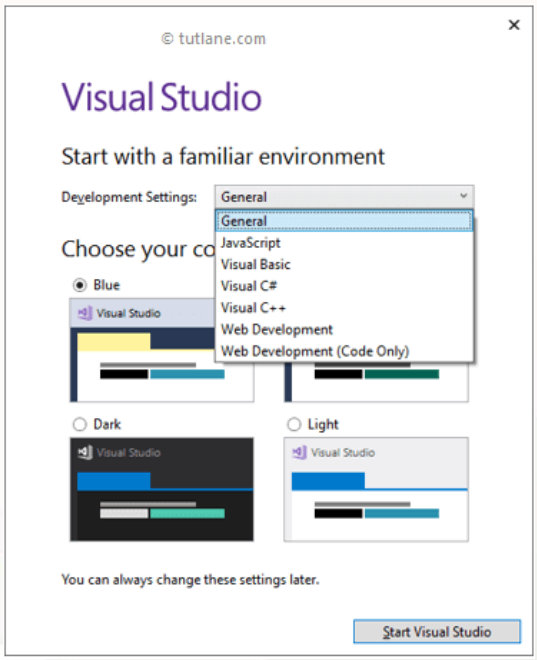
Microsoft has provided an **IDE** (Integrated Development Environment) tool called **Visual Studio** to build applications using programming languages such as C#, F#, Visual Basic, etc. based on our requirements.

The **Visual Studio IDE** will provide a common user interface with a collection of different development tools to build applications with different programming languages such as C#, F#, Visual Basic, etc.

To install and use Visual Studio for commercial purpose we need to buy a license from Microsoft. In case, if we want to use Visual Studio for learning (non-commercial) purposes, Microsoft provided a free Visual Studio Community version.

We can download and install a Visual Studio Community version from visualstudio.com. In our Visual Basic (VB) tutorial, we will use the Visual Studio 2017 community version.

Now we will see how to create a console application using visual studio in a visual basic programming language.

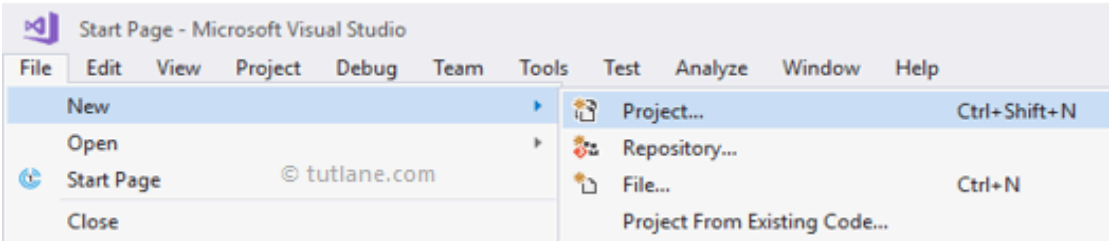


Create a Project in Visual Studio

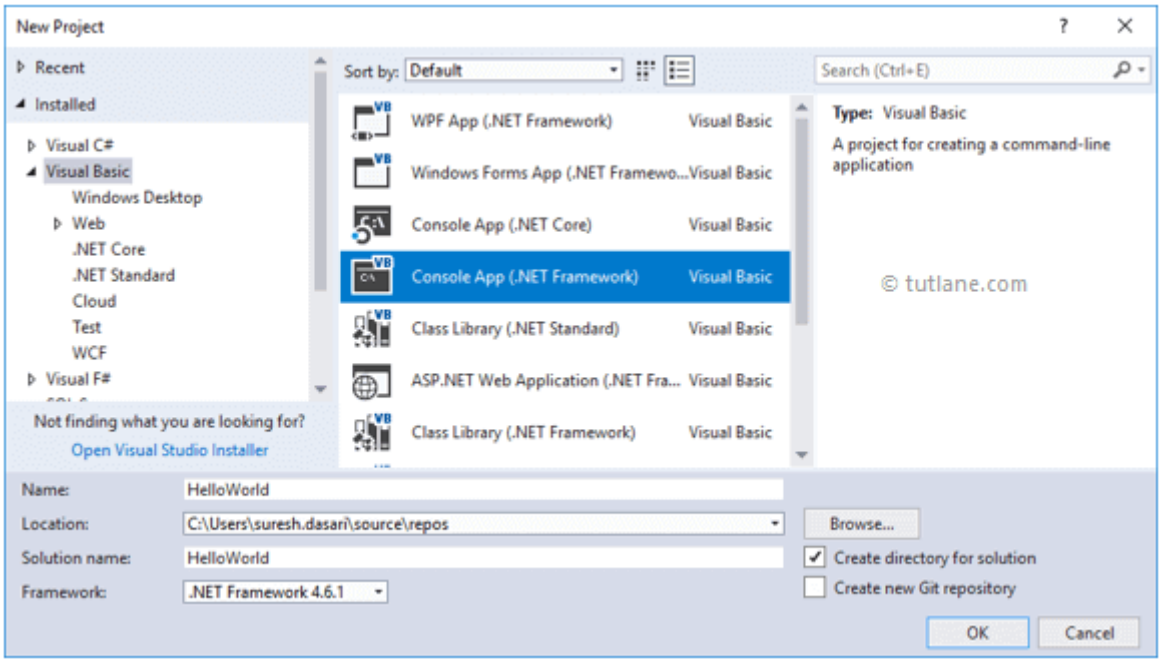
Once we are done with visual studio installation, open visual studio and it will be prompted to sign in for the first time. The **sign-in** step is optional so we can skip it and in the next step, the dialog box will appear and ask us to choose our **Development Settings** and **color theme**.

Once we select the required options, click on **Start Visual Studio** option like as shown.

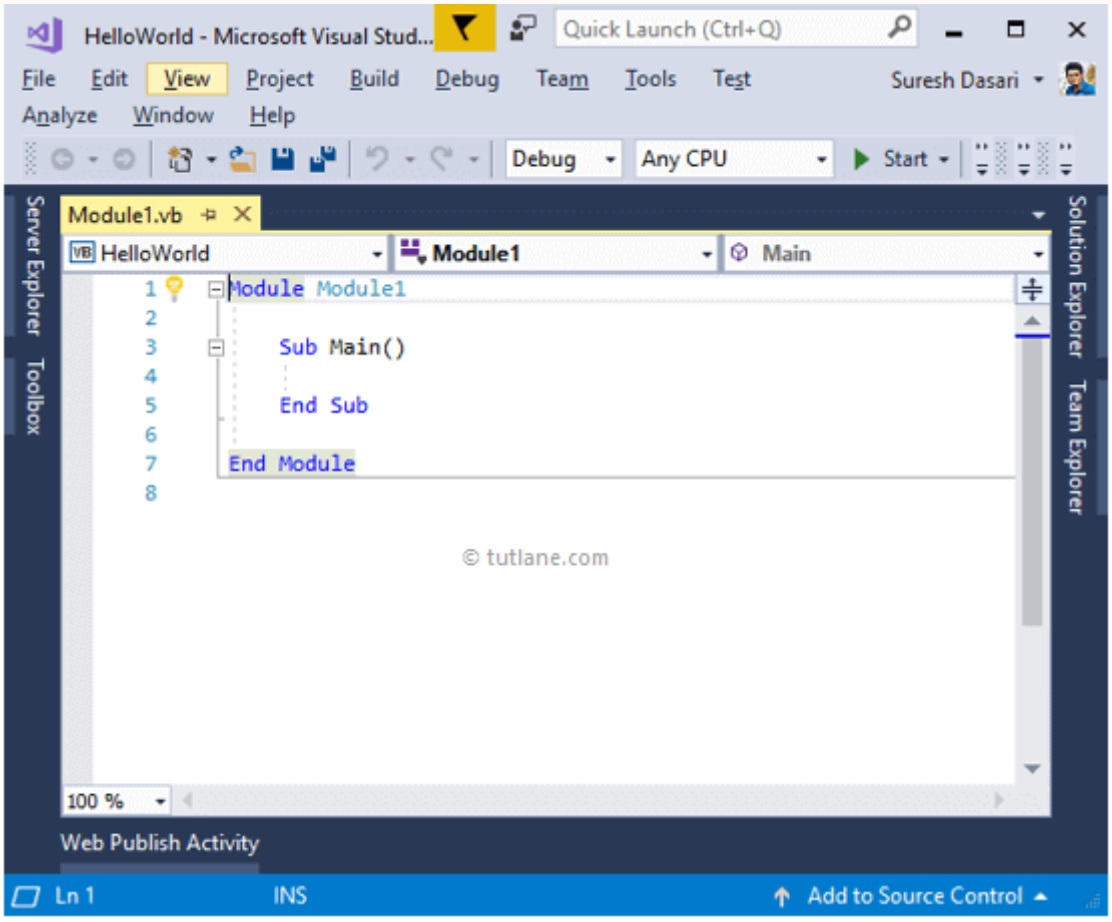
After visual studio launch, create a new console application using a visual basic programming language for that, Go to ***File*** > ***New*** > select ***Project*** like as shown below.



Once we click on **Project**, a new popup will open in that select **Visual Basic** from the left pane and choose **Console App**. In the **Name** section give any name for our project and select an appropriate **Location** path to save our project files and click **OK** like as shown below.



Once we click on the **OK** button, a new console application will be created like as shown below:



This is how we can create a console application in a visual basic programming language using visual studio based on our requirements.

Visual Basic Hello World Program

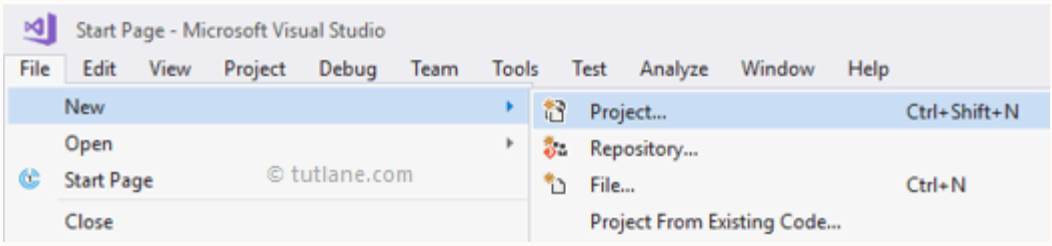
By using Visual Studio, we can easily create a Hello World Program or Console Application in Visual Basic based on our requirements.

In the previous chapter, we learned how to download and Install Visual Studio on Windows Machine. In case, if we are not installed a visual studio, and then follow the instructions to install visual studio otherwise open our visual studio.

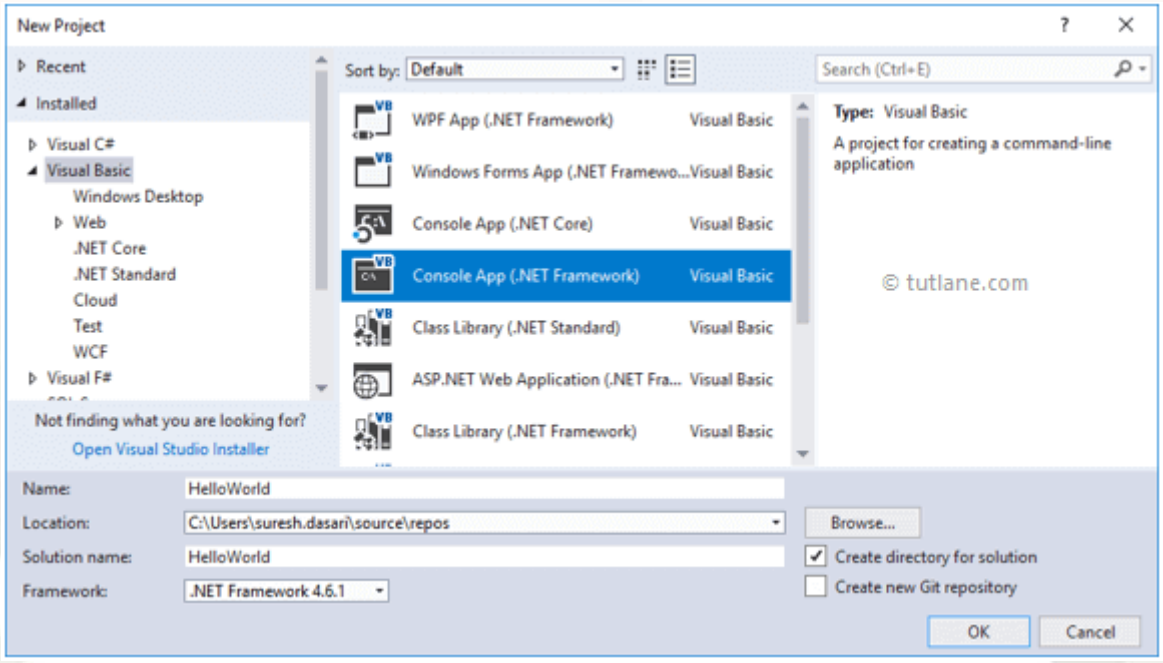
Create Visual Basic Console Application

To create a new application in visual studio, go to Menu bar, select ***File*** > ***New*** > select a **Project** like as shown below.

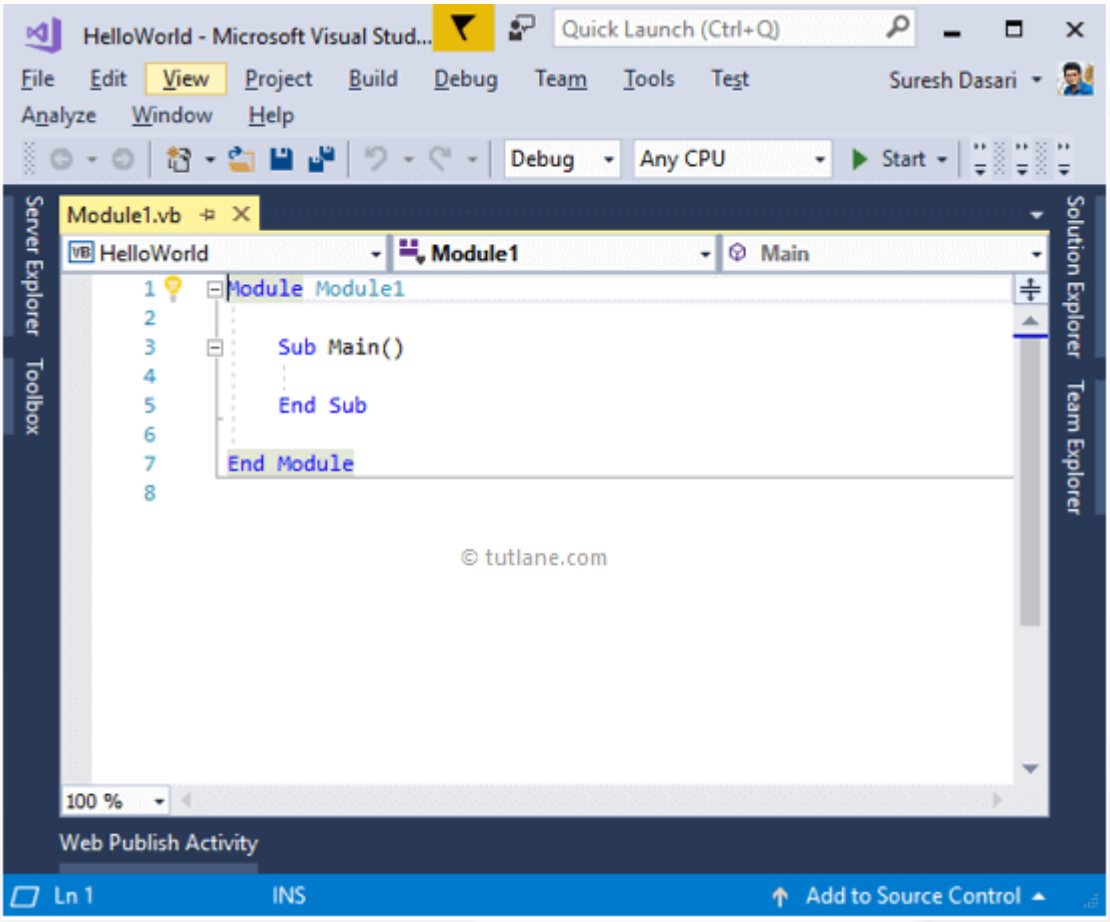
Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute



Once we click on **Project**, a new popup will open in that select **Visual Basic** from the left pane and choose **Console App**. In the **Name** section give any name for our project and select appropriate **Location** path to save our project files and click **OK** like as shown below.



Once we click on **OK** button, a new console application will be created like as shown below. In case **Module1.vb** file not opened in our code editor, open the **Solution Explorer** menu in right side and double click on our **Module1.vb** file.



If we observe the above image, by default the application contains a **Main()** method because the console applications in visual basic will always start from the **Main()** method of program class.

Visual Basic Hello World Program Example

Now, replace our **Module1.vb** file code like as shown following to display the "**Hello World**" message.

```
Imports System
Module Module1
  Sub Main()
    Console.WriteLine("Hello World!")
    Console.WriteLine("Press Enter Key to Exit.")
    Console.ReadLine()
```

```
End Sub
End Module
```

If we observe the above code, we used a lot of parameters to implement "**Hello World**" program in visual basic. In next section, we will learn all the parameters in detailed manner.

Visual Basic Hello World Program Structure

Now we will go through each step of our visual basic program and learn each parameter in detailed manner.

Imports System

Here, Imports System is the .NET Framework library namespaces and we used Imports keyword to import system namespace to use existing class methods such WriteLine(), ReadLine(), etc. By default the .NET Framework provides a lot of namespaces to make the application implementation easily.

The name space is a collection of classes and **classes** are the collection of objects and methods.

Module Module1

Here, Module Module1 is used to define a module (**Module1**). The module (**Module1**) will contain all the variables, methods, etc. based on our requirements.

Sub Main()

Here, Sub Main () is used to define a method in our module (Module1).

The keyword Sub is a procedure and it is useful to write a series of Visual Basic statements within Sub and End Sub statements.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- The name **Main** will refer the name of our method in module (**Module1**). The **Main()** method is the entry point of our console application.

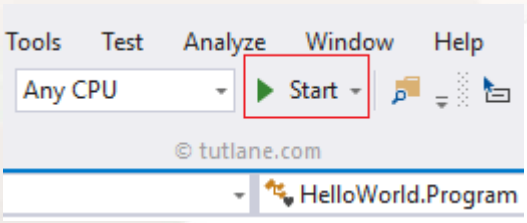
Console.WriteLine() / ReadLine()

Here, Console.WriteLine() and Console.ReadLine() methods are used to write a text to console and read the input from console.

The Console is a class of .NET Framework namespace System and WriteLine() and ReadLine() are the methods of Console class.

Compile & Run VB Hello World Program

To see the output of our Visual Basic Hello World Program, we need to compile and run the application by pressing either Ctrl + F5 or click on **Start** option in the menu bar like as shown below.



Once we click on **Start** option or Ctrl + F5, our program will get compiled and show the result like as shown below.

This is how we can create and execute the applications in visual basic (vb) programming language using visual studio based on our requirements.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Visual Basic Data Types

In Visual Basic, **Data Types** are useful to define a type of data the variable can hold such as integer, float, string, etc. in our application.

Visual Basic is a **Strongly Typed** programming language so before we perform any operation on a variable, it's mandatory to define a variable with a required data type to indicate what type of data the variable can hold in our application.

Syntax of Defining Visual Basic Data Types

Following is the syntax of defining data types in visual basic.

Dim [Variable Name] As [Data Type]

Dim [Variable Name] As [Data Type] = [Value]

If we observe the above syntax, we added a required data type after the variable name to tell the compiler about the type of data the variable can hold.

ITEM	DESCRIPTION
Dim	It is useful to declare and allocate the storage space for one or more variables.
[Variable Name]	It's the name of the variable to hold the values in our application.
As	The As clause in the declaration statement allows to define the data type.
[Data Type]	t's a type of data the variable can hold such as integer, string, decimal, etc.
[Value]	Assigning a required value to the variable.

Data Types in Visual Basic

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

The following table shows the list of available data types in a visual basic programming language with memory size and range of values.

Data type	Size	Range
Boolean	It depends on the Platform.	True or False
Byte	1 byte	0 to 255
Char	2 bytes	0 to 65535
Date	8 bytes	0:00:00am 1/1/01 to 11:59:59pm 12/31/9999
Decimal	16 bytes	(+ or -)1.0 x 10e-28 to 7.9 x 10e28
Double	8 bytes	-1.79769313486232e308 to 1.79769313486232e308
Integer	4 bytes	-2,147,483,648 to 2,147,483,647
Long	8 bytes	-9,223,372,036,854,775,808 to 9,223,372,036,854,775,807
Object	4 bytes on a 32-bit platform, 8 bytes on a 64-bit platform	Any type can be stored in a variable of type Object
SByte	1 byte	-128 to 127
Short	2 bytes	-32,768 to 32,767
Single	4 bytes	-3.4028235E+38 through -1.401298E-45 † for negative values; 1.401298E-45 through 3.4028235E+38 † for positive values
String	Depend on platform	0 to approximately 2 billion Unicode characters

Visual Basic Variables

In Visual Basic, **Variables** will represent storage locations and each variable will have a particular data type to determine the type of values the variable can hold.

Visual Basic is a **Strongly Typed** programming language so before we perform any operation on variables, it's mandatory to define the variable with a required data type to indicate that the type of data the variable can hold in our application.

Syntax of Visual Basic Variables Declaration

Following is the syntax of declaring and initializing variables in visual basic.

Dim [Variable Name] As [Data Type]

Dim [Variable Name] As [Data Type] = [Value]

If we observe the above variable declarations, we added a required data type after the variable name to tell the compiler about the type of data the variable can hold.

Item	Description
Dim	It is useful to declare and allocate the storage space for one or more variables.
[Variable Name]	It's the name of the variable to hold the values in our application.
As	The As clause in the declaration statement allows to define the data type.
[Data Type]	t's a type of data the variable can hold such as integer, string, decimal, etc.
[Value]	Assigning a required value to the variable.

Now, we will see how to declare and initialize the values to the variables in visual basic applications with examples.

Visual Basic data type and Variables Example

Following is the example of using the data type and variables in visual basic.

```
Module Module1
```

```
    Sub Main()
        Dim id As Integer
        Dim name As String = "Mr. X"
        Dim percentage As Double = 10.23
        Dim gender As Char = "M"c
        Dim isVerified As Boolean
        id = 10
        isVerified = True
          Console.WriteLine("Id:{0}", id)
          Console.WriteLine("Name:{0}", name)
           Console.WriteLine("Percentage:{0}", percentage)
           Console.WriteLine("Gender:{0}", gender)
          Console.WriteLine("Verfied:{0}", isVerified)
          Console.ReadLine()
    End Sub
End Module
```

If we observe the above example, we defined multiple variables with different data types and assigned the values based on our requirements.

When we execute the above example, we will get the result as shown below.

```
Id:10
Name: Mr. X
Percentage:10.23
Gender:M
Verfied:True
```


Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

This is how we can declare and initialize the data type and variables in Visual Basic applications based on our requirements.

Visual Basic Operators

In Visual Basic, **Operator** is a programming element that specifies what kind of operation needs to perform on operands or variables. For example, an addition (+) operator in Visual Basic is used to perform the sum operation on operands.

Visual Basic Operator Types

In Visual Basic different type of operators available, those are:

- Arithmetic Operators
- Assignment Operators
- Logical/Bitwise Operators
- Comparison Operators
- Concatenation Operators

Now, we will learn each operator in a detailed manner with examples in Visual Basic programming language.

In Visual Basic, **Arithmetic Operators** are useful to perform the basic arithmetic calculations like addition, subtraction, division, etc. based on our requirements.

The following table lists the different types of arithmetic operators available in Visual Basic.

Operator	Description	Example(a=6,b=3)
+	It will add two operands.	a + b = 9

-	It will subtract two operands.	a - b = 3
*	It will multiply two operands.	a * b = 18
/	It divides two numbers and returns a floating-point result.	a / b = 2
\	It divides two numbers and returns an integer result.	a \ b = 2
Mod	It divides two numbers and returns only the remainder.	a Mod b = 0
^	It raises a number to the power of another number.	a ^ b = 216

In Visual Basic, **Assignment Operators** are useful to assign a new value to the operand.

The following table lists the different types of assignment operators available in Visual Basic.

Operator	**Description**	**Example**
=	It will assign a value to a variable or property.	a = 10
+=	It will perform the addition of left and right operands and assign a result to the left operand.	a += 10 equals to a = a + 10
-=	It will perform a subtraction of left and right operands and assign a result to the left operand.	a -= 10 equals to a = a - 10
*=	It will perform a multiplication of left and right operands and assign a result to the left operand.	a *= 10 equals to a = a * 10

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

/=	It will perform a division of left and right operands and assign the floating-point result to the left operand.	a /= 10 equals to a = a / 10
\=	It will perform a division of left and right operands and assign the integer result to the left operand.	a \= 10 equals to a = a \ 10
^=	It will raise the value of a variable to the power of expression and assigns the result back to the variable.	a ^= 10 equals to a = a ^ 10
&=	It will concatenate a String expression to a String variable and assigns the result to the variable.	a &= "World" equals to a = a & "World"
>>=	It will move the left operand bit values to the right based on the number of positions specified by the second operand.	a >>= 2 equals to a = a >> 2
<<=	It will move the left operand bit values to the left based on the number of positions specified by the second operand.	a <<= 2 equals to a = a << 2

In Visual Basic, **Logical / Bitwise** Operators are useful to perform the logical operation between two operands like AND, OR, etc. based on our requirements. The Logical / Bitwise Operators will always work with Boolean expressions (**true** or **false**) and return Boolean values.

The following table lists the different types of logical/bitwise operators available in Visual Basic.

Operator	**Description**	**Example**

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

		(a=True, b=False)
And	It will return true if both operands are non zero.	a And b = False
Or	It will return true if any one operand becomes a non zero.	a Or b = True
Not	It will return the reverse of a logical state that means if both operands are non zero then it will return false.	Not(a And b) = True
Xor	It will return true if any one of expression1 and expression2 evaluates to true.	a Xor b = True
AndAlso	It will perform the short-circuiting logical operation and return true if both operands evaluate to true.	a AndAlso b = False
OrElse	It will perform the short-circuiting logical operation and return true if any one of operand evaluates to true.	a OrElse b = True
IsFalse	It will determine whether an expression is False.	
IsTrue	It will determine whether an expression is True.	

In Visual Basic, **Comparison Operators** are useful to determine whether the defined two operands are equal, greater than or less than, etc. based on our requirements.

The following table lists the different types of comparison operators available in Visual Basic.

Operator	Description	Example(a=10,b=5)
<	It will return true if right operand greater than left operand.	a < b = False
<=	It will return true if right operand greater than or equal to the left operand.	a <= b = False
>	It will return true if left operand greater than the right operand.	a > b = True
>=	It will return true if left operand greater than or equal to the right operand.	a >= b = True
=	It will return true if both operands are equal.	a = b = False
<>	It will return true if both operands are not equal.	a <> b = True
Is	It will return true if two object references refer to the same object.	
IsNot	It will return true if two object references refer to different objects.	

In Visual Basic, **Concatenation Operators** are useful to concatenate defined operands based on our requirements.

The following table lists the different types of concatenation operators available in Visual Basic.

Operator	Description	Example (a=Hello,b=World)
&	It will concatenate given two expressions.	a & b = HelloWorld

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

+	It can be used to add two numbers or concatenate two string expressions.	a + b = HelloWorld

Visual Basic If Statement

If statement is useful to execute the block of code or statements conditionally based on the value of an expression. Generally, in Visual Basic the statement that needs to be executed based on the condition is known as a "**Conditional Statement**" and the statement is more likely a block of code.

Syntax of if Statement

Following is the syntax of defining the *if statement* in Visual Basic programming language.

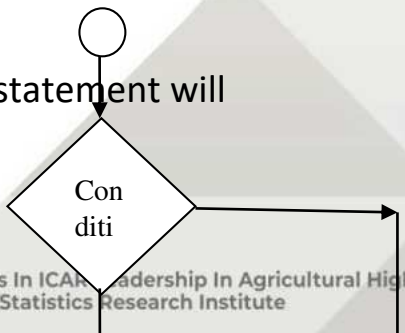
```
If ( bool_expression)Then
    // Statements to Execute if condition is true
End If
```

If we observe the above If statement syntax, the statements inside of If condition will be executed only when the "**bool_expression**" returns **true** otherwise those statements will be ignored for execution.

Following is the sample example of using the If statement in Visual Basic programming language.

```
Dim x As Integer = 20
If x >= 10 Then
    Console.WriteLine("Number Greater than 10")
End If
```

If we observe the above example, the Console statement will



be executed only when the defined condition (**x >= 10**) returns **true**.

Visual Basic If Statement Flow Chart Diagram

Following is the flow chart diagram which will represent the process flow of **If statement** in Visual Basic programming language.

If we observe the above If statement flow chart diagram, when the defined condition is true, then the statements within the If condition will be executed otherwise the If condition will come to an end without executing the statements.

If Statement Example

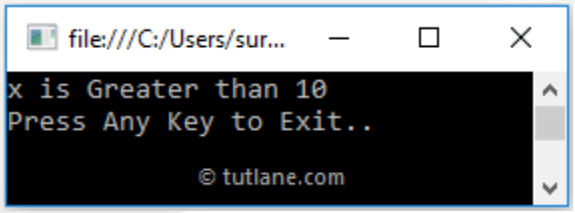
Following is the example of defining the If statement in Visual Basic programming language to execute the block of code or statements based on a Boolean expression.

```
Module Module1
    Sub Main()
        Dim x As Integer = 20, y As Integer = 10
        If x >= 10 Then
            Console.WriteLine("x is Greater than 10")
```

```
                End If
                If y <= 5 Then
                        Console.WriteLine("y is less than or equals to 5")
                End If
            Console.WriteLine("Press Enter Key to Exit..")
            Console.ReadLine()
        End Sub
    End Module
```

If we observe the above example, we defined two If conditions to execute the statements based on the defined variables (**x**, **y**) values.

When we execute the above Visual Basic program, we will get the result as shown below.



If we observe the above result, only one If condition is true that's the reason only one statement printed on the console window.

This is how we can use the If statement in Visual Basic programming language to execute the block of code or statements based on our requirements.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Glossary

- **Visual Basic (VB)**: It is a simple, modern, object-oriented and type-safe programming language. Visual Basic language has its roots from the family of C languages such as C, C++ and it is mostly similar to Java programming.
- **.NET Framework**: **.**NET Framework is a development platform for building apps for windows, web, azure, etc. by using programming languages such as C#, F# and Visual Basic (VB).
- **Visual Studio IDE:** Microsoft has provided an **IDE** (Integrated Development Environment) tool called **Visual Studio** to build applications using programming languages such as C#, F#, Visual Basic, etc. based on our requirements.
- **Visual Basic Data Types: In Visual Basic, Data Types are useful to define a type of data the variable can hold such as integer, float, string, etc. in our application.**
- **Visual Basic Variables: In Visual Basic, Variables will represent storage locations and each variable will have a particular data type to determine the type of values the variable can hold.**
- **Visual Basic Operators: In Visual Basic, Operator is a programming element that specifies what kind of operation needs to perform on operands or variables.**

REFERENCES:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Teach Yourself Visual Basic by Sams
- Visual Basic Essentials (Neil Smyth)

Elementary Statistics and Computer Application

Lesson 21

INTERNET BASICS

Content

DESIGNED AND DEVELOPED UNDER THE AEGIS OF
NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications
ICAR-Indian Agricultural Statistics Research Institute

Course Name	Elementary Statistics And Computer Application
Lesson No. 21	INTERNET BASICS
Content Creator	Dr. Raju Prasad Paswan
University Name	Assam Agricultural University,Jorhat
Course Reviewer	Dr.S.S.Sidhu
University Name	Punjab Agricultural University, Ludhiana

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Content

Objectives:

- To introduce the basics of internet.
- To introduce the history of internet.
- To introduce with the functioning of internet
- To discuss about advantages of internet.
- Study about different key terms associated with internet

1.1 **Introduction**

The Internet, sometimes called simply "the Net," is a worldwide system of computer networks - a network of networks in which users at any one computer can, if they have permission, get information from any other computer (and sometimes talk directly to users at other computers). The U.S. Department of Defense laid the foundation of the Internet roughly 50 years ago with a network called ARPANET in 1969. But the general public didn't use the Internet much until after the development of the World Wide Web in the early 1990s.

ARPANET was a network that connected major computers at the University of California at Los Angeles, the University of California at Santa Barbara, Stanford Research Institute, and the University of Utah. Within a couple of years, several other educational and research institutions joined the network.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

And now a day's millions of people have access to the Internet from home, work, or their public library.

1.2 World Wide Web:

The World Wide Web came into being in 1991, thanks to developer Tim Berners-Lee and others at the European Laboratory for Particle Physics, also known as Conseil European pour la Recherche Nucleure (CERN). The CERN team created the protocol based on hypertext that makes it possible to connect content on the Web with hyperlinks. Berners-Lee now directs the World Wide Web Consortium (W3C), a group of industry and university representatives that oversees the standards of Web technology.

Early on, the Internet was limited to noncommercial uses because its backbone was provided largely by the National Science Foundation, the National Aeronautics and Space Administration, and the U.S. Department of Energy, and funding came from the government. But as independent networks began to spring up, users could access commercial Web sites without using the government-funded network. By the end of 1992, the first commercial online service provider, Delphi, offered full Internet access to its subscribers, and several other providers followed. In June 1993, the Web boasted just 130 sites. By a year later, the number had risen to nearly 3,000. By April 1998, there were more than 2.2 million sites on the Web.

Today, the Internet is a public, cooperative, and self-sustaining facility accessible to hundreds of millions of people worldwide. Physically, the Internet uses a portion of the total resources of the currently existing

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

public telecommunication networks. Technically, what distinguishes the Internet is its use of a set of protocols called **TCP/IP** (for Transmission Control Protocol/Internet Protocol). Two recent adaptations of Internet technology, the intranet and the extranet, also make use of the TCP/IP protocol.

1.3 Why is the Internet Called a Network?

Internet is called a network as it creates a network by connecting computers and servers across the world using routers, switches and telephone lines, and other communication devices and channels. So, it can be considered a global network of physical cables such as copper telephone wires, fiber optic cables, TV cables, etc. Furthermore, even wireless connections like 3G, 4G, or Wi-Fi make use of these cables to access the Internet.

Internet is different from the World Wide Web as the World Wide Web is a network of computers and servers created by connecting them through the internet. So, the internet is the backbone of the web as it provides the technical infrastructure to establish the WWW and acts as a medium to transmit information from one computer to another computer. It uses web browsers to display the information on the client, which it fetches from web servers.

The internet is not owned by a single person or organization entirely. It is a concept based on physical infrastructure that connects networks with other networks to create a global network of billions of computers. As of 12 August 2016, there were more than 300 crores of internet users across the world.

1.4 Internet Connection Methods

- **Dial-up:** Dial-up is a method that uses a telephone line, which we connect to a phone jack, just as we would connect our telephone to

the wall. Dial-up is the slowest connection method and it requires our computer to have a dial-up modem.

- **Broadband:** Broadband is a high-speed connection method which can utilize cable, DSL, or satellite. Each of these methods requires different types of hardware.

- **Fiber-optic:** Fiber-optic communication transmits data by sending pulses of light through ultra-thin optical fiber. Because light travels so quickly, this technology can transmit Internet data at super-fast speeds.

1.5 How does the Internet work?

The internet works with the help of clients and servers. A device such as a laptop/desktop, which is connected to the internet, is called a client, not a server as it is not directly connected to the internet. However, it is indirectly connected to the internet through an Internet Service Provider (ISP) and is identified by an IP address, which is a string of numbers. Our home that uniquely identifies our home, likewise an IP address acts as the shipping address of our device. The IP address is provided by our ISP, and we can see what IP address our ISP has given to our system.

A server is a large computer that stores websites. It also has an IP address. A place where a large number of servers are stored is called a data centre. The server accepts requests send by the client through a browser over a network (internet) and responds accordingly.

To access the internet we need a domain name, which represents an IP address number, i.e., each IP address has been assigned a domain name. For example, youtube.com, facebook.com, paypal.com are used to represent the IP addresses. Domain names are created as it is difficult for a person to remember a long string of numbers. However, internet does not understand the domain name, it understands the IP address, so when we enter the domain name in the browser search bar, the internet has to

get the IP addresses of this domain name from a huge phone book, which is known as DNS (Domain Name Server). For example, if we have a person's name, we can find his phone number in a phone book by searching his name. The internet uses the DNS server in the same way to find the IP address of the domain name. DNS servers are managed by ISPs or similar organizations.

Now with these basics, let us explain how the internet works –

When we turn on our computer and type a domain name in the browser search bar, our browser sends a request to the DNS server to get the corresponding IP address.

After getting the IP address, the browser forwards the request to the respective server. Once the server gets the request to provide information about a particular website, the data starts flowing.

The data is transferred through the optical fiber cables in digital format or in the form of light pulses. As the servers are placed at distant places, the data may have to travel thousands of miles through optical fiber cable to reach our computer. The optical fiber is connected to a router, which converts the light signals into electrical signals. These electrical signals are transmitted to our laptop using an Ethernet cable.

Thus, we receive the desired information through the internet, which is actually a cable that connects us with the server. Furthermore, if we are using wireless internet using wi fi or mobile data, the signals from the optical cable are first sent to a cell tower and from where it reaches to our cell phone in the form of electromagnetic waves.

1.6 Who Manages Internet ?

The internet is managed by ICANN (Internet Corporation for Assigned Names and Numbers) located in the USA. It manages IP addresses assignment, domain name registration, etc. The data transfer is very fast

on the internet. The moment we press enter we get the information from a server located thousands of miles away from us.
The reason for this speed is that the data is sent in the binary form (0, 1), and these zeros and ones are divided into small pieces called packets, which can be sent at high speed.

1.7 Advantages of the Internet:

Instant Messaging: We can send messages or communicate to anyone using internet, such as email, voice chat, video conferencing, etc.

Get directions: Using GPS technology, we can get directions to almost every place in a city, country, etc. We can find restaurants, malls, or any other service near our location.

Online Shopping: It allows us to shop online such as clothes, shoes, book movie tickets, railway tickets, flight tickets, and many more.

Pay Bills: We can pay our bills online, such as electricity bills, gas bills, college fees, etc.

Online Banking: It allows us to use internet banking in which we can check our balance, receive or transfer money, get a statement, request cheque-book, etc.

Online Selling: We can sell our products or services online. It helps us reach more customers and thus increases our sales and profit.

Work from Home: In case we need to work from home, we can do it using a system with internet access. Today, many companies allow their employees to work from home.

Entertainment: We can listen to online music, watch videos or movies, play online games.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Cloud computing: It enables us to connect our computers and internet enabled devices to cloud services such as cloud storage, cloud computing, etc.

Career building: We can search for jobs online on different job portals and send our CV through email if required.

1.8 Important Terms:

Browser- Contains the basic software we need in order to find, retrieve, view, and send information over the Internet.
Download- To copy data from a remote computer to a local computer.

Upload- To send data from a local computer to a remote computer.

E-mail (**electronic mail**)- It is the exchange of computer-stored messages by telecommunication. E-mail can be distributed to lists of people as well as to individuals. However, we can also send non-text files, such as graphic images and sound files, as attachments sent in binary streams.

Filter- Software that allows targeted sites to be blocked from view.

Home Page- The beginning "page" of any site.

HTML (HyperText Markup Language) - The coding language used to create documents for use on the World Wide Web.

HTTP (HyperText Transport Protocol) - The set of rules for exchanging files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web. Relative to the TCP/IP suite of protocols (which are the basis for information exchange on the Internet), HTTP is an application protocol.

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

Hypertext - Generally any text that contains "links" to other text.

Search Engine - A web server that collects data from other web servers and puts it into a database (much like an index), it provides links to pages that contain the object of our search.

TCP/IP-TCP/IP **(Transmission Control Protocol/Internet Protocol)** is the basic communication language or protocol of the Internet. It can also be used as a communications protocol in a private network (either an intranet or an extranet). When we are set up with direct access to the Internet, our computer is provided with a copy of the TCP/IP program just as every other computer that we may send messages to or get information from also has a copy of TCP/IP.

URL (Uniform Resource Locator) - The Internet address. The prefix of a URL indicates which area of the Internet will be accessed. URLs look differently depending on the Internet resource we are seeking.

WWW (World Wide Web)- It is commonly known as the Web. It is a system of interlinked hypertext documents that are accessed via the Internet. With a web browser, one can view web pages that may contain text, images, videos, and other multimedia and navigate between them via hyperlinks.

Web Browser:

A Web browser contains the basic software we need in order to find, retrieve, view, and send information over the Internet. This includes software that lets us:

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

- Send and receive electronic-mail (or e-mail) messages worldwide nearly instantaneously.
- Read messages from newsgroups (or forums) about thousands of topics in which users share information and opinions.
- Browse the World Wide Web (or Web) where we can find a rich variety of text, graphics, and interactive information.

The five most popular desktop web browsers are; **Microsoft Internet Explorer**, **Mozilla Firefox**, **Google Chrome**, Apple's **Safari**, and **Opera**. etc. **Google Chrome** has been the most popular browser in the United States since December 2013. In other countries, Google Chrome has also taken up a dominating role.

URL:

Every user of Internet has an IP number, a unique number consisting of 4 parts separated by dots. The IP number is the server's or a user's name in the internet.

165.113.245.2

128.143.22.55

However, it is harder for people to remember numbers than to remember word combinations. So, addresses are given "word-based" addresses called URLs. The URL and the IP number are one and the same.

The standard way to give the address of any resource on the Internet that is part of the World Wide Web (WWW). A URL looks like this:

http://www.matisse.net/seminars.html

telnet://well.sf.ca.us

gopher://gopher.ed.gov/

The URL is divided into sections:

transfer/transport protocol:// server (or domain). generic top level domain/path/filename

The first part of a URL defines the transport protocol.

http:// (HyperText Transport Protocol) moves graphical, hypertext files

ftp:// (File Transfer Protocol) moves a file between 2 computers

gopher:// (Gopher client) moves text-based files

news: (News group reader) accesses a discussion group

telnet:// (Telnet client) allows remote login to another computer

Here's an example:

http://www.vrml.k12.la.us/tltc/mainmenu.htm

- **http** is the protocol
- **www.vrml.k12.la.us** is the server
- **tltc/** is the path
- **mainmenu.htm** is the filename of the page on the site

Designed and developed under the aegis of NAHEP Component-2 Project "Investments In ICAR Leadership In Agricultural Higher Education"
Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute

1. we do not have to enter **http://** , most browsers will add that information when we press **Enter** or click the  button at the end of the Address Bar.
2. To view recently visited Web sites, click the down arrow at the end of the address field.
3. When we start typing a frequently used Web address in the Address bar, a list of similar addresses appears that we can choose from. And if a Web-page address is wrong, Internet Explorer can search for similar addresses to try to find a match.
4. The URL **must** be typed correctly. If we get a "Server Does Not Have A DNS Entry" message, this message tells that our browser can't locate the server (i.e. the computer that hosts the Web page). It could mean that the network is busy or that the server has been removed or taken down for maintenance. Check our spelling and try again later.

DOMAINS

Domains divide World Wide Web sites into categories based on the nature of their owner, and they form part of a site's address, or uniform resource locator (URL). Common top-level domains are:

.com—commercial enterprises	**.mil**—military site
.org—organization site (non-profits, etc.)	**.int**—organizations established by international treaty
.net—network	**.biz**—commercial and personal
.edu—educational site (universities, schools, etc.)	**.info**—commercial and personal

.gov—government organizations	**.name**—personal sites

Additional three-letter, four-letter, and longer top-level domains are frequently added. Each country linked to the Web has a two-letter top-level domain, for example .fr is France, .ie is Ireland, .In in India.

DNS - Domain Name System

Short for **D**omain **N**ame **S**ystem (or **S**ervice or **S**erver), an Internet service that translates domain names into IP addresses. Because domain names are alphabetic, they're easier to remember. The Internet however, is really based on IP addresses. Every time we use a domain name, a DNS service must translate the name into the corresponding IP address. For example, the domain name www.example.com might translate to198.105.232.4.

The DNS system is, in fact, its own network. If one DNS server doesn't know how to translate a particular domain name, it asks another one, and so on, until the correct IP address is returned.

E-Mail

Electronic mail, most commonly referred to as email or e-mail is a method of exchanging digital messages from an author to one or more recipients. Modern email operates across the Internet or other computer networks. Some early email systems required that the author and the recipient both be online at the same time, in common with instant messaging**.** Today's email systems are based on a store-and-

forward model. Email servers accept, forward, deliver, and store messages. Neither the users nor their computers are required to be online simultaneously. They need connect only briefly typically to a mail server for as long as it takes to send or receive messages.

An Internet email message consists of three components, the message envelope, the message header and the message body. The message header contains control information including minimally an originator's email address and one or more recipient addresses. Usually descriptive information is also added such as a subject header field and a message submission date/time stamp.

Email is an information and communications technology. It uses technology to communicate a digital message over the Internet. Users use email differently, based on how they think about it. There are many software platforms available to send and receive. Popular email platforms include Gmail, Hotmail, Yahoo! Mail, Outlook, and many others.

Glossary

- **Client :** A device such as a laptop/desktop, which is connected to the internet is called a client.
- **Server :** Server accepts requests send by the client through a browser over a network (internet) and responds accordingly
- **Browser**--Contains the basic software we need in order to find, retrieve, view, and send information over the Internet.
- **Download**--To copy data from a remote computer to a local computer.
- **Upload**—To send data from a local computer to a remote computer.

- **E-mail** - E-mail (electronic mail) is the exchange of computer-stored messages by telecommunication. E-mail can be distributed to lists of people as well as to individuals.
- **Filter** - Software that allows targeted sites to be blocked from view.
- **Home Page** - The beginning "page" of any site.
- **HTML (HyperText Markup Language)** - The coding language used to create documents for use on the World Wide Web.
- **HTTP (HyperText Transport Protocol)** - The set of rules for exchanging files (text, graphic images, sound, video, and other multimedia files) on the World Wide Web.
- **Hypertext** - Generally any text that contains "links" to other text.
- **Search Engine** - A web server that collects data from other web servers and puts it into a database (much like an index), it provides links to pages that contain the object of our search.
- **TCP/IP --** TCP/IP (Transmission Control Protocol/Internet Protocol) is the basic communication language or protocol of the Internet.
- **URL (Uniform Resource Locator)** - The Internet address.
- **WWW (World Wide Web)** - The World Wide Web (abbreviated as WWW commonly known as the Web) is a system of interlinked hypertext documents that are accessed via the Internet.
- **Domains** - Domains divide World Wide Web sites into categories based on the nature of their owner, and they form part of a site's address, or uniform resource locator (URL).
- **DNS -** Short for **D**omain **N**ame **S**ystem (or **S**ervice or **S**erver), an Internet service that translates *domain names* into IP addresses.

REFERENCES

1. Introduction to Information Technology, ITL Education Solution Limited, by Pearson.
2. Introduction to Information Technology, O-Level by Satish Jain
3. The Internet Book by Douglas E. Comer
4. www.javatpoint.com