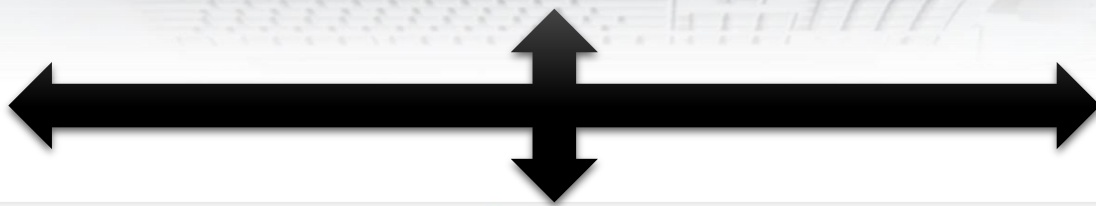


# *Elementary Statistics & Computer Application*



## 1. Elementary Statistics and Computer Application (HBS 100) 3 (2+ 1)

Basic concepts: Variable statistics, types and sources of data, classification and tabulation of Data, construction of frequency distribution, tables, graphic representation of data, simple, multiple component and percentage, bar diagram, pie diagram, histogram, frequency polygon and frequency curve average and measures of location, mean, mode, median, geometric mean, harmonic mean, percentiles and quadrilles, for raw and grouped data. Dispersion: Range, standard deviation, variance, coefficient of variation for raw and grouped data. Probability: Basic concept, additive and multiplicative laws. Theoretical distributions, binomial, poisson and normal distributions, sampling, basic concepts, sampling vs. complete enumeration parameter and statistic, sampling methods, simple random sampling and stratified random sampling. Tests of Significance: Basic concepts, tests for equality of means, and independent and paired t-tests, chi-square test for application of attributes and test for goodness of fit of mendelian ratios. Correlation: Scatter diagram, correlation co-efficient and its properties, regression, fitting of simple linear regression, test of significance of correlation and regression coefficient. Experimental Designs: Basic concepts, completely randomized design, randomized block design, latin square designs, factorial experiments, basic concepts, analysis of factorial experiments up to 3 factors – split plot design, strip plot design, long term experiments, plot size, guard rows. Computer application: Introduction to computers and personal computers, operating system, DOS and Windows 95, introduction to programming languages, BASIC language, concepts, basic and programming techniques, MS Office, Win Word, Excel, Power Point, introduction to Multi-Media and its application. VISUAL BASIC-concepts, basic and programming techniques, introduction to Internet.

*Practical: Construction of frequency distribution table and its graphical representation, histogram, frequency polygon, frequency curve, bar chart, simple, multiple, component and percentage bar charts, pie chart, mean, mode for row and grouped data, percentiles, quadrille, and median for row and grouped data, coefficient of variation, 't' test for independent, will equal and unequal variants, paired 't' test, chi-square test for contingency tables and theoretical ratios, correlation and linear regression. Studies on computer components –BASIC language, VISUAL BASIC, programming techniques, MS Office, Excel, Power P*

## **Lecture.1**

**Data – definition – Collection of data – Primary and secondary data – Classification of data – Qualitative and quantitative data.**

### **Basic Concepts**

#### **Statistics (Definition)**

Quantitative figures are known as data.

Statistics is the science which deals with the

- (i) Collection of data
- (ii) Organization of data or Classification of data
- (iii) Presentation of data
- (iv) Analysis of data
- (v) Interpretation of data

Data and statistics are not same as used commonly.

#### **Example for data**

1. No. of farmers in a block.
2. The rainfall over a period of time.
3. Area under paddy crop in a state.

#### **Functions of statistics**

Statistics simplifies complexity, presents facts in a definite form, helps in formulation of suitable policies, facilitates comparison and helps in forecasting.

#### **Uses of statistics**

Statistics has pervaded almost all spheres of human activities. Statistics is useful in the administration of various states, Industry, business, economics, research workers, banking, insurance companies etc.

### **Limitations of Statistics**

1. Statistical theories can be applied only when there is variability in the experimental material.
2. Statistics deals with only aggregates or groups and not with individual objects.
3. Statistical results are not exact.
4. Statistics can be misused.

### **Collection of data**

Data can be collected by using sampling methods or experiments.

### **Data**

The information collected through censuses and surveys or in a routine manner or other sources is called a raw data. When the raw data are grouped into groups or classes, they are known as grouped data.

There are two types of data

1. Primary data
2. Secondary data.

### **Primary data**

The data which is collected by actual observation or measurement or count is called primary data.

### **Methods of collection of primary data**

Primary data is collected in any one of the following methods

1. Direct personal interviews.
2. Indirect oral interviews
3. Information from correspondents.
4. Mailed questionnaire method.
5. Schedules sent through enumerators.



### **1. Direct personal interviews**

The persons from whom information are collected are known as informants or respondents. The investigator personally meets them and asks questions to gather the necessary information.

#### **Merits**

1. The collected informations are likely to be uniform and accurate. The investigator is there to clear the doubts of the informants.
2. People willingly supply information because they are approached personally. Hence more response is noticed in this method then in any other method.

#### **Limitations**

It is likely to be very costly and time consuming if the number of persons to be interviewed is large and the persons are spread over a wide area.

### **2. Indirect oral interviews**

Under this method, the investigator contacts witnesses or neighbors or friends or some other third parties who are capable of supplying the necessary information.

#### **Merits**

For almost all the surveys of this kind, the informants like within a closed area. Hence, the time and the cost are less. For certain surveys, this is the only method available.

#### **Limitations**

The information obtained by this method is not very reliable. The informants and the person who conducts a survey easily distort the truth.

### **3. Information from correspondents**

The investigator appoints local agents or correspondents in different places and compiles the information sent by them.

**Merits**

- For certain kinds of primary data collection, this is the only method available.
- This method is very cheap and expeditious.
- The quality of data collected is also good due to long experience of local representatives.

**Limitations**

Local agents and correspondents are not likely to be serious and careful.

**4. Mailed Questionnaire method**

Under this method a list of questions is prepared and is sent to all the informants by post. The list of questions is technically called questionnaire.

**Merits**

1. It is relatively cheap.
2. It is preferable when the informants are spread over a wide area.
3. It is fast if the informants respond duly.

**Limitations**

1. Were the informants are illiterate people, this method cannot be adopted.
2. It is possible that some of the persons who receive the questionnaires do not return them. Their action is known as non – response.

**5. Schedules sent through enumerators**

Under this method, enumerators or interviewers take the schedules, meet the informants and fill in their replies. A schedule is filled by the interviewer in a face to face situation with the informant.

**Merits**

1. It can be adopted even if the informants are illiterate.

2. Non-response is almost nil as the enumerators go personally and contact the informants.
3. The informations collected are reliable. The enumerators can be properly trained for the same.

### **Limitations**

1. It is costliest method.
2. Extensive training is to be given to the enumerators for collecting correct and uniform informations.

### **Secondary data**

The data which are compiled from the records of others is called secondary data.

The data collected by an individual or his agents is primary data for him and secondary data for all others. The secondary data are less expensive but it may not give all the necessary information.

Secondary data can be compiled either from published sources or from unpublished sources.

### **Sources of published data**

1. Official publications of the central, state and local governments.
2. Reports of committees and commissions.
3. Publications brought about by research workers and educational associations.
4. Trade and technical journals.
5. Report and publications of trade associations, chambers of commerce, bank etc.
6. Official publications of foreign governments or international bodies like U.N.O, UNESCO etc.

### **Sources of unpublished data**

All statistical data are not published. For example, village level officials maintain records regarding area under crop, crop production etc. They collect details for

administrative purposes. Similarly details collected by private organizations regarding persons, profit, sales etc become secondary data and are used in certain surveys.

### **Characteristics of secondary data**

The secondary data should possess the following characteristics. They should be reliable, adequate, suitable, accurate, complete and consistent.

### **Variables**

Variability is a common characteristic in biological Sciences. A quantitative or qualitative characteristic that varies from observation to observation in the same group is called a variable.

### **Quantitative data**

The basis of classification is according to differences in quantity. In case of quantitative variables the observations are made in terms of kgs, Lt, cm etc. Example weight of seeds, height of plants.

### **Qualitative data**

When the observations are made with respect to quality is called qualitative data.  
Eg: Crop varieties, Shape of seeds, soil type.  
The qualitative variables are termed as attributes.

### **Classification of data**

Classification is the process of arranging data into groups or classes according to the common characteristics possessed by the individual items.

- Data can be classified on the basis of one or more of the following kinds namely
1. Geography
  2. Chronology
  3. Quality
  4. Quantity.

### 1. Geographical classification (or) Spatial Classification

Some data can be classified area-wise, such as states, towns etc.

Data on area under crop in India can be classified as shown below

<b>Region</b>	<b>Area ( in hectares)</b>
Central India	-
West	-
North	-
East	-
South	-

### 2. Chronological or Temporal or Historical Classification

Some data can be classified on the basis of time and arranged chronologically or historically.

Data on Production of food grains in India can be classified as shown below

<b>Year</b>	<b>Tonnes</b>
1990-91	-
1991-92	-
1992-93	-
1993-94	-
1994-95	-

### 3. Qualitative Classification

Some data can be classified on the basis of attributes or characteristics. The number of farmers based on their land holdings can be given as follows

<b>Type of farmers</b>	<b>Number of farmers</b>
Marginal	907
Medium	1041
Large	1948
<b>Total</b>	<b>3896</b>

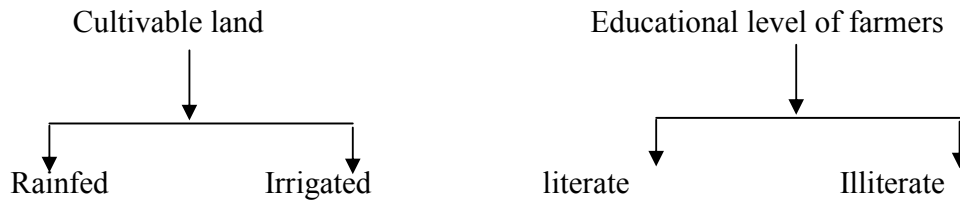
Qualitative classification can be of two types as follows

- (i) Simple classification
- (ii) Manifold classification

**(i) Simple Classification**

This is based on only one quality.

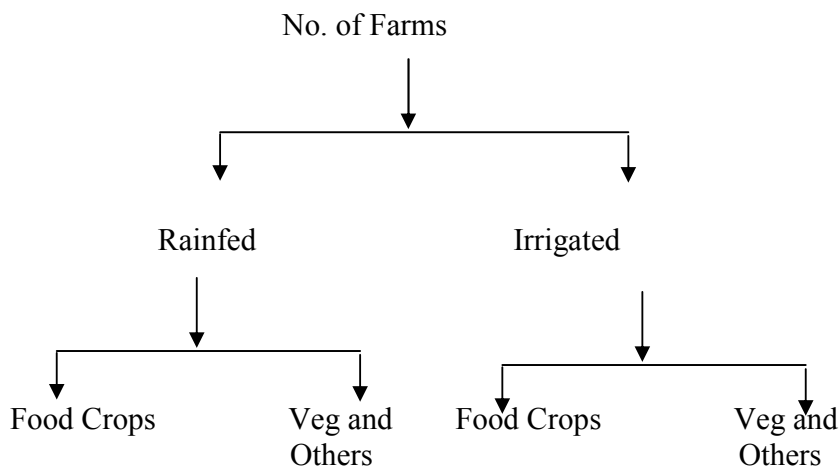
Eg:



**(ii) Manifold Classification**

This is based on more than one quality.

Eg:



**4. Quantitative classification**

Some data can be classified in terms of magnitude. The data on land holdings by farmers in a block. Quantitative classification is based the land holding which is the variable in this example.

Land holding ( hectare)	Number of Farmers
< 1	442
1-2	908
2-5	471
>5	124
<b>Total</b>	<b>1945</b>

### Difference between Primary and secondary data

	<b>Primary Data</b>	<b>Secondary Data</b>
1. Original data	Primary data are original because investigation himself collects them.	Secondary data are not original since investigator makes use of the other agencies.
2. Suitability	If these data are collected accurately and systematically their suitability will be very positive.	These might or might not suit the objectives of enquiry.
3. Time and labour	These data involve large expenses in terms of money, time and manpower	These data are relatively less costly.
4. Precaution	don't need any great precaution while using these data.	These should be used with great care and caution.

### Questions

1. A simple table contains data on

- a) Two characteristics
- b) Several characteristics
- c) One characteristic
- d) Three characteristics

**Ans: One characteristic**

2. When the collected data is grouped with reference to time, we have

- a) Quantitative classification
- b) Qualitative classification
- c) Geographical Classification
- d) Chorological Classification

**Ans: Chorological Classification**

3. Geographical classification means, classification of data according to Region.

**Ans: True**

4. An arrangement of data into rows and columns is known as Tabulation.

**Ans: True**

5. Data on yield is a quantitative variable

**Ans: True**



6. Qualitative variables are also called as attributes.

**Ans: True**

7. Define primary and secondary data

8. Give the advantages of tabulation.

9. Write a detail note on the types of classification

10. What are the essential characteristics of a good table?

## **Lecture.2**

### **Diagrammatic representation of data – uses and limitations – simple, Multiple, Component and percentage bar diagrams – pie chart.**

#### **Diagrams**

Diagrams are various geometrical shape such as bars, circles etc. Diagrams are based on scale but are not confined to points or lines. They are more attractive and easier to understand than graphs.

#### **Merits**

1. Most of the people are attracted by diagrams.
2. Technical Knowledge or education is not necessary.
3. Time and effort required are less.
4. Diagrams show the data in proper perspective.
5. Diagrams leave a lasting impression.
6. Language is not a barrier.
7. Widely used tool.

#### **Demerits (or) limitations**

1. Diagrams are approximations.
2. Minute differences in values cannot be represented properly in diagrams.
3. Large differences in values spoil the look of the diagram.
4. Some of the diagrams can be drawn by experts only. eg. Pie chart.
5. Different scales portray different pictures to laymen.

#### **Types of Diagrams**

The important diagrams are

1. Simple Bar diagram.
2. Multiple Bar diagram.
3. Component Bar diagram.
4. Percentage Bar diagram.

5. Pie chart
6. Pictogram
7. Statistical maps or cartograms.

In all the diagrams and graphs, the groups or classes are represented on the x-axis and the volumes or frequencies are represented in the y-axis.

### **Simple Bar diagram**

If the classification is based on attributes and if the attributes are to be compared with respect to a single character we use simple bar diagram.

### **Example**

1. The area under different crops in a state.
2. The food grain production of different years.
3. The yield performance of different varieties of a crop.
4. The effect of different treatments etc.

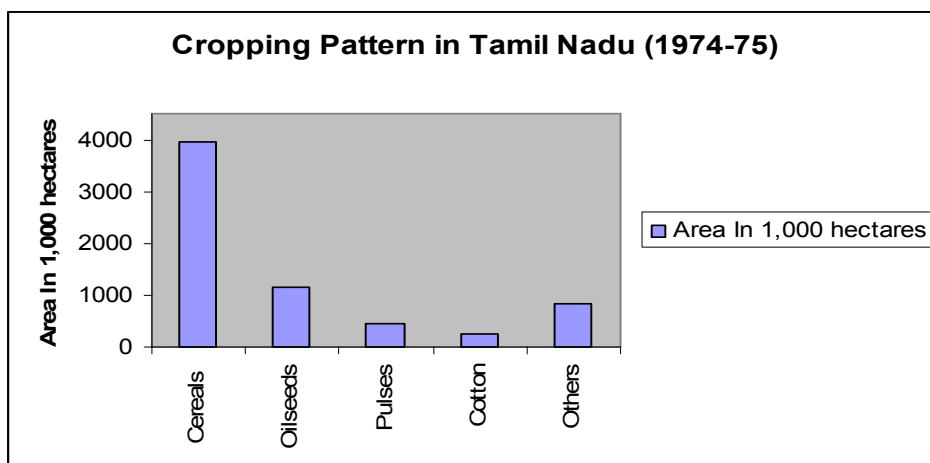
Simple bar diagrams Consists of vertical bars of equal width. The heights of these bars are proportional to the volume or magnitude of the attribute. All bars stand on the same baseline. The bars are separated from each others by equal intervals. The bars may be coloured or marked.

### **Example**

The cropping pattern in Tamil Nadu in the year 1974-75 was as follows.

<b>Crops</b>	<b>Area In 1,000 hectares</b>
Cereals	3940
Oilseeds	1165
Pulses	464
Cotton	249
Others	822

The simple bar diagram for this data is given below.



### Multiple bar diagram

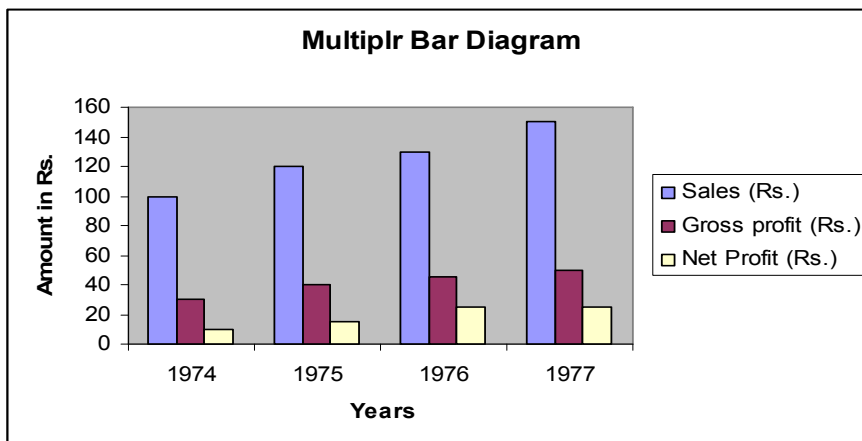
If the data is classified by attributes and if two or more characters or groups are to be compared within each attribute we use multiple bar diagrams. If only two characters are to be compared within each attribute, then the resultant bar diagram used is known as double bar diagram.

The multiple bar diagram is simply the extension of simple bar diagram. For each attribute two or more bars representing separate characters or groups are to be placed side by side. Each bar within an attribute will be marked or coloured differently in order to distinguish them. Same type of marking or colouring should be done under each attribute. A footnote has to be given explaining the markings or colourings.

### Example

Draw a multiple bar diagram for the following data which represented agricultural production for the period from 2000-2004

Year	Food grains (tones)	Vegetables (tones)	Others (tones)
2000	100	30	10
2001	120	40	15
2002	130	45	25
2003	150	50	25
2004			



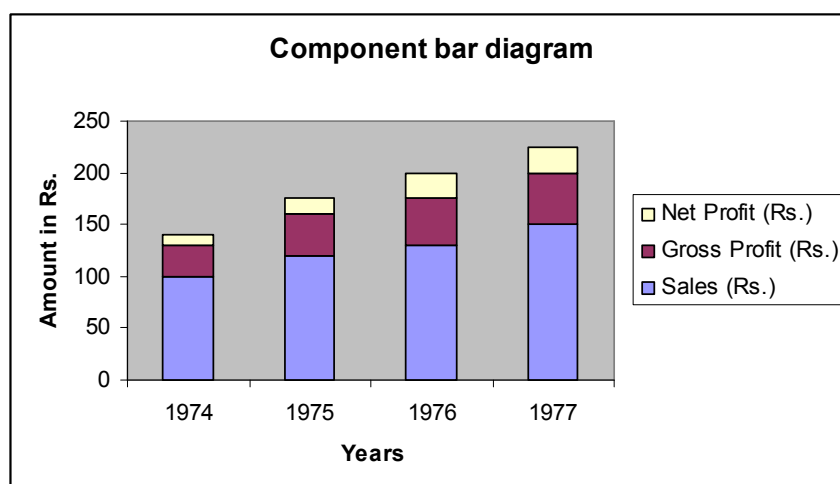
### Component bar diagram

This is also called sub – divided bar diagram. Instead of placing the bars for each component side by side we may place these one on top of the other. This will result in a component bar diagram.

Example:

Draw a component bar diagram for the following data

Year	Sales (Rs.)	Gross Profit (Rs.)	Net Profit (Rs.)
1974	100	30	10
1975	120	40	15
1976	130	45	25
1977	150	50	25



### Percentage bar diagram

Sometimes when the volumes of different attributes may be greatly different for making meaningful comparisons, the attributes are reduced to percentages. In that case each attribute will have 100 as its maximum volume. This sort of component bar chart is known as percentage bar diagram.

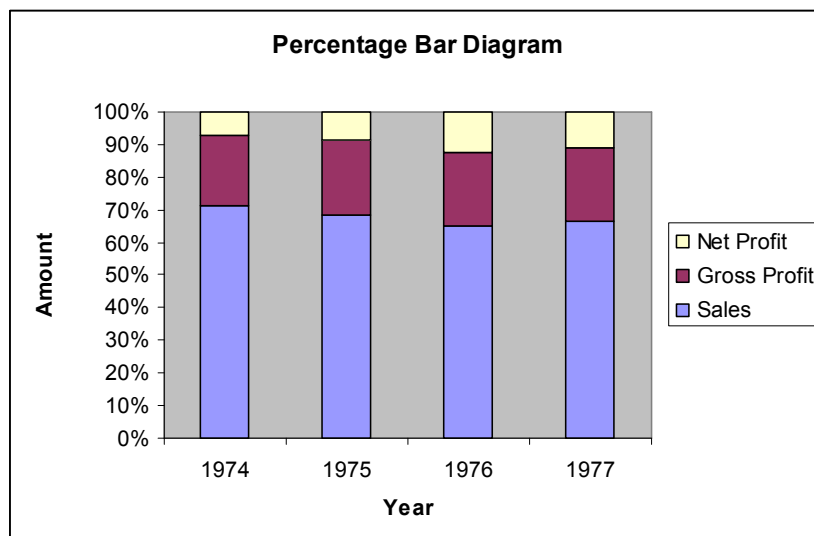
$$\text{Percentage} = \frac{\text{Actual value}}{\text{Total of the actual value}} \times 100,$$

Example:

Draw a Percentage bar diagram for the following data

Using the formula  $\text{Percentage} = \frac{\text{Actual value}}{\text{Total of the actual value}} \times 100$ , the above table is converted.

Year	Sales (Rs.)	Gross Profit (Rs.)	Net Profit (Rs.)
1974	71.43	21.43	7.14
1975	68.57	22.86	8.57
1976	65	22.5	12.5
1977	66.67	22.22	11.11



### **Pie chart / Pie Diagram**

Pie diagram is a circular diagram. It may be used in place of bar diagrams. It consists of one or more circles which are divided into a number of sectors. In the construction of pie diagram the following steps are involved.

Step 1:

Whenever one set of actual value or percentage are given, find the corresponding angles in degrees using the following formula

$$\text{Angle} = \frac{\text{Actual value}}{\text{Total of the actual value}} \times 360^\circ$$

$$\text{(or) Angle} = \frac{\text{Percentage}}{100} \times 360^\circ$$

Step 2:

Find the radius using the area of the circle  $\pi r^2$  where value of  $\pi$  is  $22/7$  or  $3.14$

### **Example**

Given the cultivable land area in four southern states of India. Construct a pie diagram for the following data.

<b>State</b>	<b>Cultivable area( in hectares)</b>
Andhra Pradesh	663
Karnataka	448
Kerala	290
Tamil Nadu	556
<b>Total</b>	<b>1957</b>

Using the formula

$$\text{Angle} = \frac{\text{Actual value}}{\text{Total of the actual value}} \times 360^\circ$$

(or)

$$\text{Angle} = \frac{\text{Percentage}}{100} \times 360^\circ$$

The table value becomes



State	Cultivable area
Andhra Pradesh	121.96
Karnataka	82.41
Kerala	53.35
Tamil Nadu	102.28

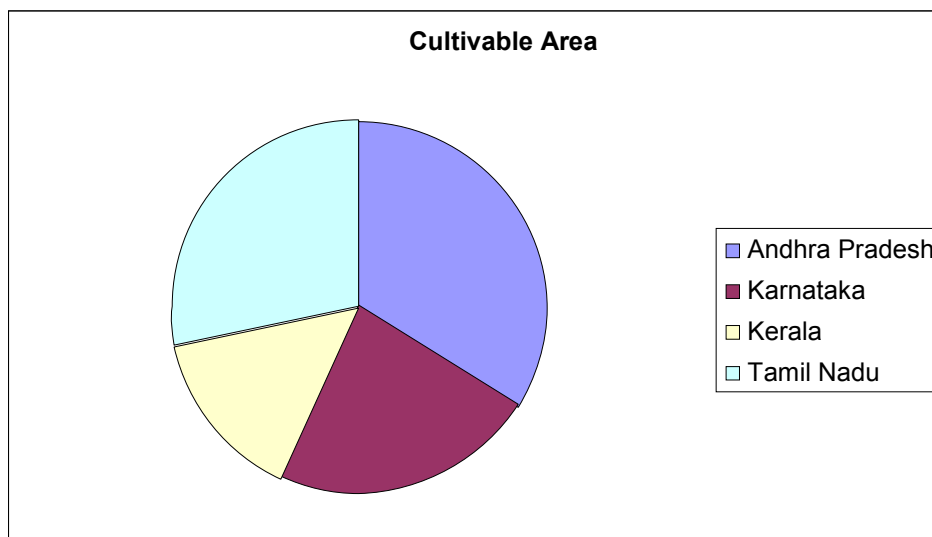
$$\text{Radius} = \pi r^2$$

$$\text{Here } \pi r^2 = 1957$$

$$r^2 = \frac{1957}{\pi} = \frac{1957}{3.14} = 623.24$$

$$r = 24.96$$

$$r = 25 \text{ (approx)}$$



### Questions

- In a component bar diagram the length of the bar
  - Will be same for all
  - Depends on the total
  - will not be same
  - none of these

**Ans: Depends on the total**

- The length of the bar will be same for all categories in
  - Multiple bar diagram
  - component bar diagram
  - Percentage bar diagram
  - none of these

**Ans: Percentage bar diagram**

3. Sub-divided bar diagram are also called Component bar diagram.

**Ans: True**

4. The multiple bar diagram is the extension of simple bar diagram.

**Ans: True**

5. In a bar the width of the bars should be equal.

**Ans: True**

6. In a percentage bar diagram the length of the bars will not be equal.

**Ans: False**

7. How diagrams are useful in representing statistical data?

8. How to draw a pie chart?

9. Explain how to draw simple and multiple bar diagrams.

10. Explain how to draw Component and percentage bar diagrams.

### Lecture.3

## Graphical representation – Histogram – Frequency polygon and Frequency curve

### Graphs

Graphs are charts consisting of points, lines and curves. Charts are drawn on graph sheets. Suitable scales are to be chosen for both x and y axes, so that the entire data can be presented in the graph sheet. Graphical representations are used for grouped quantitative data.

### Histogram

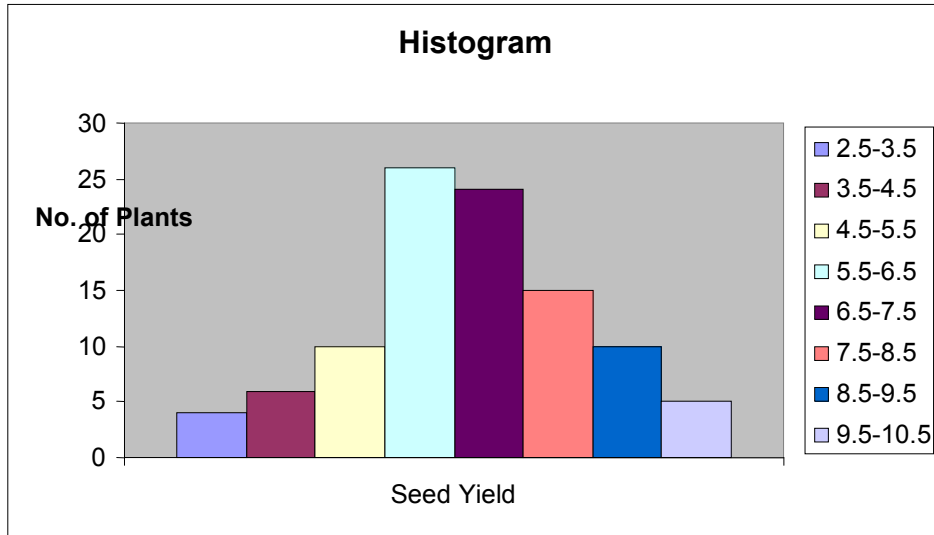
When the data are classified based on the class intervals it can be represented by a histogram. Histogram is just like a simple bar diagram with minor differences. There is no gap between the bars, since the classes are continuous. The bars are drawn only in outline without colouring or marking as in the case of simple bar diagrams. It is the suitable form to represent a frequency distribution.

Class intervals are to be presented in x axis and the bases of the bars are the respective class intervals. Frequencies are to be represented in y axis. The heights of the bars are equal to the corresponding frequencies.

### Example

Draw a histogram for the following data

Seed Yield (gms)	No. of Plants
2.5-3.5	4
3.5-4.5	6
4.5-5.5	10
5.5-6.5	26
6.5-7.5	24
7.5-8.5	15
8.5-9.5	10
9.5-10.5	5



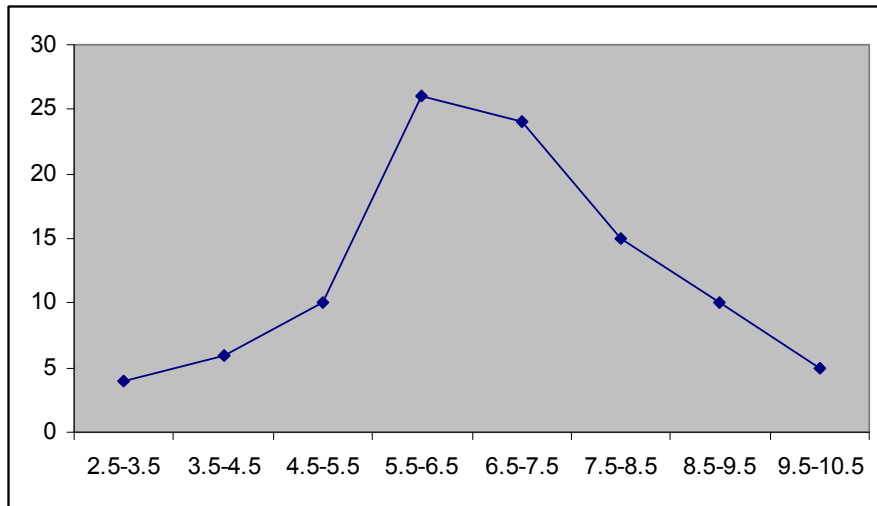
### Frequency Polygon

The frequencies of the classes are plotted by dots against the mid-points of each class. The adjacent dots are then joined by straight lines. The resulting graph is known as frequency polygon.

### Example

Draw frequency polygon for the following data

Seed Yield (gms)	No. of Plants
2.5-3.5	4
3.5-4.5	6
4.5-5.5	10
5.5-6.5	26
6.5-7.5	24
7.5-8.5	15
8.5-9.5	10
9.5-10.5	5



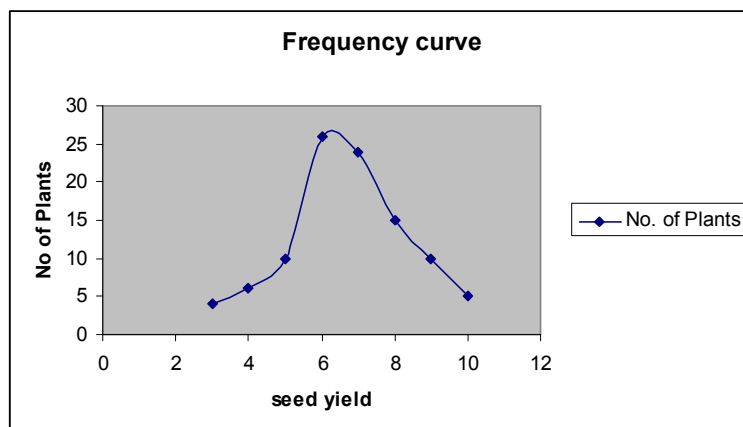
### Frequency curve

The procedure for drawing a frequency curve is same as for frequency polygon. But the points are joined by smooth or free hand curve.

### Example

Draw frequency curve for the following data

Seed Yield (gms)	No. of Plants
2.5-3.5	4
3.5-4.5	6
4.5-5.5	10
5.5-6.5	26
6.5-7.5	24
7.5-8.5	15
8.5-9.5	10
9.5-10.5	5



### Ogives

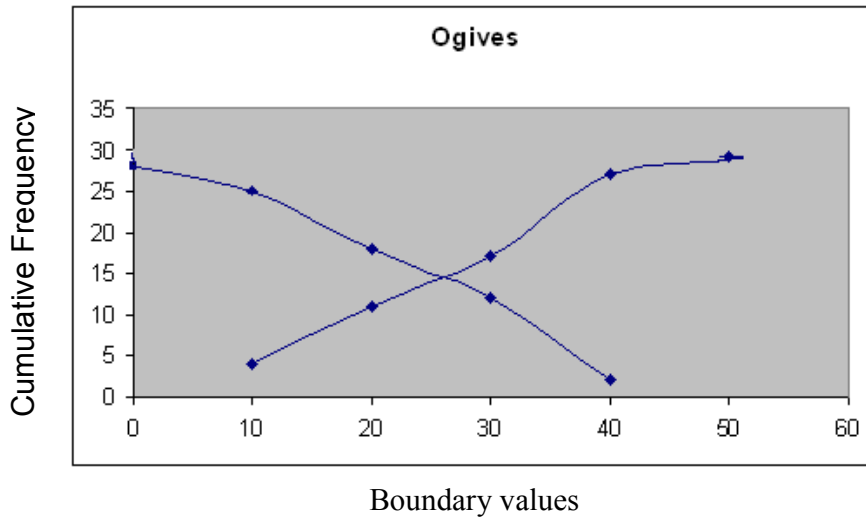
Ogives are known also as cumulative frequency curves and there are two kinds of ogives. One is less than ogive and the other is more than ogive.

**Less than ogive:** Here the cumulative frequencies are plotted against the upper boundary of respective class interval.

**Greater than ogive:** Here the cumulative frequencies are plotted against the lower boundaries of respective class intervals.

### Example

Continuous Interval	Mid Point	Frequency	< cumulative Frequency	> cumulative frequency
0-10	5	4	4	29
10-20	15	7	11	25
20-30	25	6	17	18
30-40	35	10	27	12
40-50	45	2	29	2



### Questions

- With the help of histogram we can draw
  - Frequency polygon
  - frequency curve
  - Frequency distribution
  - all the above

**Ans: all the above**

- Ogives for more than type and less than type distribution intersect at
  - Mean
  - median
  - Mode
  - origin

**Ans: median**

- To draw the frequency polygon we take the mid values in the X axis.
- To draw the frequency curve we take the mid values in the X axis.
- In a frequency curve the points are joined by bits of straight lines

**Ans: False**

- The histogram can be drawn for equal and unequal classes

**Ans: True**

- Explain how to draw frequency curve
- Explain how to draw histogram.
- Explain the diagrams that can be drawn for a frequency distribution table
- Explain how to draw less than and more than Ogives.



## Lecture.4

### **Measures of averages - Mean – median – mode – geometric mean – harmonic mean – computation of the above statistics for raw and grouped data - merits and demerits - measures of location – percentiles – quartiles - computation of the above statistics for raw and grouped data**

In the study of a population with respect to one in which we are interested we may get a large number of observations. It is not possible to grasp any idea about the characteristic when we look at all the observations. So it is better to get one number for one group. That number must be a good representative one for all the observations to give a clear picture of that characteristic. Such representative number can be a central value for all these observations. This central value is called a measure of central tendency or an average or a measure of locations. There are five averages. Among them mean, median and mode are called simple averages and the other two averages geometric mean and harmonic mean are called special averages.

#### **Arithmetic mean or mean**

Arithmetic mean or simply the mean of a variable is defined as the sum of the observations divided by the number of observations. It is denoted by the symbol  $\bar{x}$ . If the variable  $x$  assumes  $n$  values  $x_1, x_2, \dots, x_n$  then the mean is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$
$$= \frac{1}{n} \sum_{i=1}^n x_i$$

This formula is for the ungrouped or raw data.

#### **Example 1**

Calculate the mean for pH levels of soil 6.8, 6.6, 5.2, 5.6, 5.8

#### **Solution**

$$\bar{x} = \frac{6.8 + 6.6 + 5.2 + 5.6 + 5.8}{5} = \frac{30}{5} = 6$$

#### **Grouped Data**

The mean for grouped data is obtained from the following formula:

$$\bar{x} = \frac{\sum fx}{n}$$

Where  $x =$  the mid-point of individual class

$f =$  the frequency of individual class

$n =$  the sum of the frequencies or total frequencies in a sample.

### Short-cut method

$$\bar{x} = A + \frac{\sum fd}{n} \times c$$

Where  $d = \frac{x - A}{c}$

$A =$  any value in  $x$

$n =$  total frequency

$c =$  width of the class interval

### Example 2

Given the following frequency distribution, calculate the arithmetic mean

Marks : 64      63      62      61      60      59

Number of  
Students } : 8      18      12      9      7      6

### Solution

X	F	Fx	D=x-A	Fd
64	8	512	2	16
63	18	1134	1	18
<b>62</b>	12	744	0	0
61	9	549	-1	-9
60	7	420	-2	-14
59	6	354	-3	-18
	60	3713		-7

### Direct method

$$\bar{x} = \frac{\sum fx}{n}$$

$$\bar{x} = \frac{3713}{60} = 61.88$$

### Short-cut method

$$\bar{x} = A + \frac{\sum fd}{n} \times c$$

Here A = 62

$$\bar{x} = 62 - \frac{7}{60} \times 1 = 61.88$$

### Example 3

For the frequency distribution of seed yield of sesamum given in table, calculate the mean yield per plot.

Yield per plot in(in g)	64.5-84.5	84.5-104.5	104.5-124.5	124.5-144.5
No of plots	3	5	7	20

### Solution

Yield ( in g)	No of Plots (f)	Mid X	$d = \frac{x - A}{c}$	Fd
64.5-84.5	3	74.5	-1	-3
84.5-104.5	5	94.5	0	0
104.5-124.5	7	114.5	1	7
124.5-144.5	20	134.5	2	40
<b>Total</b>	<b>35</b>			<b>44</b>

A=94.5

The mean yield per plot is

Direct method:

$$\begin{aligned}\bar{x} &= \frac{\sum fx}{n} = \frac{(74.5 \times 3) + (94.5 \times 5) + (114.5 \times 7) + (134.5 \times 20)}{35} \\ &= \frac{4187.5}{35} = 119.64 \text{ gms}\end{aligned}$$

### Shortcut method

$$\bar{x} = A + \frac{\sum fd}{n} \times c$$

$$\bar{x} = 94.5 + \frac{44}{35} \times 20 = 119.64 \text{ g}$$

### **Merits and demerits of Arithmetic mean**

#### **Merits**

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.
3. If the number of items is sufficiently large, it is more accurate and more reliable.
4. It is a calculated value and is not based on its position in the series.
5. It is possible to calculate even if some of the details of the data are lacking.
6. Of all averages, it is affected least by fluctuations of sampling.
7. It provides a good basis for comparison.

#### **Demerits**

1. It cannot be obtained by inspection nor located through a frequency graph.
2. It cannot be in the study of qualitative phenomena not capable of numerical measurement i.e. Intelligence, beauty, honesty etc.,
3. It can ignore any single item only at the risk of losing its accuracy.
4. It is affected very much by extreme values.
5. It cannot be calculated for open-end classes.
6. It may lead to fallacious conclusions, if the details of the data from which it is computed are not given.

#### **Median**

The median is the middle most item that divides the group into two equal parts, one part comprising all values greater, and the other, all values less than that item.

#### **Ungrouped or Raw data**

Arrange the given values in the ascending order. If the number of values are odd, median is the middle value

If the number of values are even, median is the mean of middle two values.

By formula

When n is odd, Median = Md =  $\left(\frac{n+1}{2}\right)^{th}$  value

When n is even, Average of  $\left(\frac{n}{2}\right)$  and  $\left(\frac{n}{2}+1\right)^{th}$  value

#### Example 4

If the weights of sorghum ear heads are 45, 60, 48, 100, 65 gms, calculate the median

#### Solution

Here n = 5

First arrange it in ascending order

45, 48, 60, 65, 100

$$\begin{aligned}\text{Median} &= \left(\frac{n+1}{2}\right)^{th} \text{ value} \\ &= \left(\frac{5+1}{2}\right) = 3^{rd} \text{ value} = 60\end{aligned}$$

#### Example 5

If the sorghum ear- heads are 5, 48, 60, 65, 65, 100 gms, calculate the median.

#### Solution

Here n = 6

$$\text{Median} = \text{Average of } \left(\frac{n}{2}\right) \text{ and } \left(\frac{n}{2}+1\right)^{th} \text{ value}$$

$$\left(\frac{n}{2}\right) = \frac{6}{2} = 3^{rd} \text{ value} = 60 \quad \text{and} \quad \left(\frac{n}{2}+1\right) = \frac{6}{2} + 1 = 4^{th} \text{ value} = 65$$

$$\text{Median} = \frac{60+65}{2} = 62.5 \text{ g}$$

#### Grouped data

In a grouped distribution, values are associated with frequencies. Grouping can be in the form of a discrete frequency distribution or a continuous frequency distribution. Whatever may be the type of distribution, cumulative frequencies have to be calculated to know the total number of items.

### Cumulative frequency (cf)

Cumulative frequency of each class is the sum of the frequency of the class and the frequencies of the previous classes, ie adding the frequencies successively, so that the last cumulative frequency gives the total number of items.

### Discrete Series

Step1: Find cumulative frequencies.

Step2 : Find  $\left(\frac{n}{2} + 1\right)$

Step3: See in the cumulative frequencies the value just greater than  $\left(\frac{n}{2} + 1\right)$

Step4: Then the corresponding value of x is median.

### Example 6

The following data pertaining to the number of insects per plant. Find median number of insects per plant.

Number of insects per plant (x)	1	2	3	4	5	6	7	8	9	10	11	12
No. of plants(f)	1	3	5	6	10	13	9	5	3	2	2	1

### Solution

Form the cumulative frequency table

x	f	cf
1	1	1
2	3	4
3	5	9
4	6	15
5	10	25
6	13	38
7	9	47
8	5	52
9	3	55
10	2	57
11	2	59
12	1	60
	60	

Median = size of  $\left(\frac{n+1}{2}\right)^{th}$  item

Here the number of observations is even. Therefore median = average of (n/2)th item and (n/2+1)th item.

$$= (30^{\text{th}} \text{ item} + 31^{\text{st}} \text{ item}) / 2 = (6+6)/2 = 6$$

Hence the median size is 6 insects per plant.

### Continuous Series

The steps given below are followed for the calculation of median in continuous series.

Step1: Find cumulative frequencies.

Step2: Find  $\left(\frac{n}{2}\right)$

Step3: See in the cumulative frequency the value first greater than  $\left(\frac{n}{2}\right)$ , Then the corresponding class interval is called the Median class. Then apply the formula

$$\text{Median} = l + \frac{\frac{n}{2} - m}{f} \times c$$

where

$l$  = Lower limit of the medianal class

$m$  = cumulative frequency preceding the medianal class

$c$  = width of the class

$f$  = frequency in the median class.

$n$  = Total frequency.

### Example 7

For the frequency distribution of weights of sorghum ear-heads given in table below. Calculate the median.

Weights of ear heads ( in g)	No of ear heads (f)	Less than class	Cumulative frequency (m)
60-80	22	<80	22
80-100	38	<100	60
100-120	45	<120	105
120-140	35	<140	140
140-160	24	<160	164
Total	164		



### **Solution**

$$\text{Median} = l + \frac{\frac{n}{2} - m}{f} \times c$$

$$\left(\frac{n}{2}\right) = \left(\frac{164}{2}\right) = 82$$

It lies between 60 and 105. Corresponding to 60 the less than class is 100 and corresponding to 105 the less than class is 120. Therefore the medianal class is 100-120. Its lower limit is 100.

Here  $l = 100$ ,  $n=164$ ,  $f = 45$ ,  $c = 20$ ,  $m = 60$

$$\text{Median} = 100 + \frac{82 - 60}{45} \times 20 = 109.78 \text{ gms}$$

### **Merits of Median**

1. Median is not influenced by extreme values because it is a positional average.
2. Median can be calculated in case of distribution with open-end intervals.
3. Median can be located even if the data are incomplete.

### **Demerits of Median**

1. A slight change in the series may bring drastic change in median value.
2. In case of even number of items or continuous series, median is an estimated value other than any value in the series.
3. It is not suitable for further mathematical treatment except its use in calculating mean deviation.
4. It does not take into account all the observations.

### **Mode**

The mode refers to that value in a distribution, which occur most frequently. It is an actual value, which has the highest concentration of items in and around it. It shows the centre of concentration of the frequency in around a given value. Therefore, where the purpose is to know the point of the highest concentration it is preferred. It is, thus, a positional measure.

Its importance is very great in agriculture like to find typical height of a crop variety, maximum source of irrigation in a region, maximum disease prone paddy variety. Thus the mode is an important measure in case of qualitative data.

## Computation of the mode

### Ungrouped or Raw Data

For ungrouped data or a series of individual observations, mode is often found by mere inspection.

#### Example 8

Find the mode for the following seed weight

2, 7, 10, 15, 10, 17, 8, 10, 2 gms

∴ Mode = 10

In some cases the mode may be absent while in some cases there may be more than one mode.

#### Example 9

(1) 12, 10, 15, 24, 30 (no mode)

(2) 7, 10, 15, 12, 7, 14, 24, 10, 7, 20, 10

the modal values are 7 and 10 as both occur 3 times each.

### Grouped Data

For Discrete distribution, see the highest frequency and corresponding value of x is mode.

Example:

Find the mode for the following

Weight of sorghum in gms (x)	No. of ear head(f)
50	4
65	6
75	16
80	8
95	7
100	4

#### Solution

The maximum frequency is 16. The corresponding x value is 75.

∴ mode = 75 gms.

### Continuous distribution

Locate the highest frequency the class corresponding to that frequency is called the modal class.

Then apply the formula.

$$\text{Mode} = l + \frac{f_s}{f_p + f_s} \times c$$

Where  $l$  = lower limit of the model class

$f_p$  = the frequency of the class preceding the model class

$f_s$  = the frequency of the class succeeding the model class

and  $c$  = class interval

### Example 10

For the frequency distribution of weights of sorghum ear-heads given in table below. Calculate the mode

Weights of ear heads (g)	No of ear heads (f)	
60-80	22	
80-100	38	$f_p$
100-120	45	$f$
120-140	35	$f_s$
140-160	20	
<b>Total</b>	<b>160</b>	

### Solution

$$\text{Mode} = l + \frac{f_s}{f_p + f_s} \times c$$

Here  $l = 100$ ,  $f = 45$ ,  $c = 20$ ,  $m = 60$ ,  $f_p = 38$ ,  $f_s = 35$

$$\begin{aligned} \text{Mode} &= 100 + \frac{35_s}{38 + 35} \times 20 \\ &= 100 + \frac{35_s}{73} \times 20 \\ &= 109.589 \end{aligned}$$

### Geometric mean

The geometric mean of a series containing  $n$  observations is the  $n$ th root of the product of the values.

If  $x_1, x_2, \dots, x_n$  are observations then

$$G.M = \sqrt[n]{x_1, x_2 \dots x_n}$$

$$= (x_1, x_2 \dots x_n)^{1/n}$$

$$\text{Log GM} = \frac{1}{n} \log(x_1, x_2 \dots x_n)$$

$$= \frac{1}{n} (\log x_1 + \log x_2 + \dots + \log x_n)$$

$$= \frac{\sum \log x_i}{n}$$

$$GM = \text{Antilog} \frac{\sum \log x_i}{n}$$

For grouped data

$$GM = \text{Antilog} \left[ \frac{\sum f \log x_i}{n} \right]$$

GM is used in studies like bacterial growth, cell division, etc.

### Example 11

If the weights of sorghum ear heads are 45, 60, 48, 100, 65 gms. Find the Geometric mean for the following data

Weight of ear head x (g)	Log x
45	1.653
60	1.778
48	1.681
100	2.000
65	1.813
<b>Total</b>	<b>8.925</b>

### Solution

Here n = 5

$$GM = \text{Antilog} \frac{\sum \log x_i}{n}$$

$$= \text{Antilog} \frac{8.925}{5}$$

$$= \text{Antilog} 1.785$$

$$= 60.95$$

### Grouped Data

#### Example 12

Find the Geometric mean for the following

Weight of sorghum (x)	No. of ear head(f)
50	4
65	6
75	16
80	8
95	7
100	4

#### Solution

Weight of sorghum (x)	No. of ear head(f)	Log x	f x log x
50	5	1.699	8.495
63	10	10.799	17.99
65	5	1.813	9.065
130	15	2.114	31.71
135	15	2.130	31.95
<b>Total</b>	<b>50</b>	<b>9.555</b>	<b>99.21</b>

Here n= 50

$$\begin{aligned}
 \text{GM} &= \text{Antilog} \left[ \frac{\sum f \log x_i}{n} \right] \\
 &= \text{Antilog} \left[ \frac{99.21}{50} \right] \\
 &= \text{Antilog } 1.9842 = 96.43
 \end{aligned}$$

### Continuous distribution

#### Example 13

For the frequency distribution of weights of sorghum ear-heads given in table below. Calculate the Geometric mean

Weights of ear heads ( in g)	No of ear heads (f)
60-80	22
80-100	38
100-120	45

120-140	35
140-160	20
<b>Total</b>	<b>160</b>

**Solution**

<b>Weights of ear heads ( in g)</b>	<b>No of ear heads (f)</b>	<b>Mid x</b>	<b>Log x</b>	<b>f log x</b>
60-80	22	70	1.845	40 59
80-100	38	90	1.954	74.25
100-120	45	110	2.041	91.85
120-140	35	130	2.114	73.99
140-160	20	150	2.176	43.52
<b>Total</b>	<b>160</b>			<b>324.2</b>

Here  $n = 160$

$$\begin{aligned}
 \text{GM} &= \text{Antilog} \left[ \frac{\sum f \log x_i}{n} \right] \\
 &= \text{Antilog} \left[ \frac{324.2}{160} \right] \\
 &= \text{Antilog} [2.02625] \\
 &= 106.23
 \end{aligned}$$

### **Harmonic mean (H.M)**

Harmonic mean of a set of observations is defined as the reciprocal of the arithmetic average of the reciprocal of the given values. If  $x_1, x_2, \dots, x_n$  are  $n$  observations,

$$\text{H.M} = \frac{n}{\sum_{i=1}^n \left( \frac{1}{x_i} \right)}$$

For a frequency distribution

$$\text{H.M} = \frac{n}{\sum_{i=1}^n f \left( \frac{1}{x_i} \right)}$$

H.M is used when we are dealing with speed, rates, etc.

**Example 13**

From the given data 5, 10,17,24,30 calculate H.M.

X	$\frac{1}{x}$
5	0.2000
10	0.1000
17	0.0588
24	0.0417
30	0.4338

$$H.M = \frac{5}{0.4338} = 11.526$$

**Example 14**

Number of tomatoes per plant are given below. Calculate the harmonic mean.

Number of tomatoes per plant	20	21	22	23	24	25
Number of plants	4	2	7	1	3	1

**Solution**

Number of tomatoes per plant (x)	No of plants(f)	$\frac{1}{x}$	$f\left(\frac{1}{x}\right)$
20	4	0.0500	0.2000
21	2	0.0476	0.0952
22	7	0.0454	0.3178
23	1	0.0435	0.0435
24	3	0.0417	0.1251
25	1	0.0400	0.0400
	18		0.8216

$$H.M = \frac{n}{\sum f\left(\frac{1}{x_i}\right)} = \frac{18}{0.1968} = 21.91$$

**Merits of H.M**

1. It is rigidly defined.
2. It is defined on all observations.
3. It is amenable to further algebraic treatment.
4. It is the most suitable average when it is desired to give greater weight to smaller observations and less weight to the larger ones.

### Demerits of H.M

1. It is not easily understood.
2. It is difficult to compute.
3. It is only a summary figure and may not be the actual item in the series
4. It gives greater importance to small items and is therefore, useful only when small items have to be given greater weightage.
5. It is rarely used in grouped data.

### Percentiles

The percentile values divide the distribution into 100 parts each containing 1 percent of the cases. The  $x^{\text{th}}$  percentile is that value below which  $x$  percent of values in the distribution fall. It may be noted that the median is the  $50^{\text{th}}$  percentile.

For raw data, first arrange the  $n$  observations in increasing order. Then the  $x^{\text{th}}$  percentile is given by

$$P_x = \left( \frac{x(n+1)}{100} \right)^{\text{th}} \text{ item}$$

For a frequency distribution the  $x^{\text{th}}$  percentile is given by

$$P_x = l + \left( \frac{(x \cdot n / 100) - cf}{f} \times C \right)$$

Where

$l$  = lower limit of the percentile class which contains the  $x^{\text{th}}$  percentile value ( $x \cdot n / 100$ )

$cf$  = cumulative frequency upto  $l$

$f$  = frequency of the percentile class

$C$  = class interval

$N$  = total number of observations

### Percentile for Raw Data or Ungrouped Data

#### Example 15

The following are the paddy yields (kg/plot) from 14 plots:

30,32,35,38,40,42,48,49,52,55,58,60,62,and 65 ( after arranging in ascending order). The computation of  $25^{\text{th}}$  percentile ( $Q_1$ ) and  $75^{\text{th}}$  percentile ( $Q_3$ ) are given below:



$$\begin{aligned}
P_{25}(\text{or } Q_1) &= \left( \frac{25(14+1)}{100} \right)^{\text{th}} \text{ item} \\
&= \left( 3\frac{3}{4} \right)^{\text{th}} \text{ item} \\
&= 3^{\text{rd}} \text{ item} + (4^{\text{th}} \text{ item} - 3^{\text{rd}} \text{ item}) \left( \frac{3}{4} \right) \\
&= 35 + (38-35) \left( \frac{3}{4} \right) \\
&= 35 + 3 \left( \frac{3}{4} \right) = 37.25 \text{ kg}
\end{aligned}$$

$$\begin{aligned}
P_{75}(\text{or } Q_3) &= \left( \frac{75(14+1)}{100} \right)^{\text{th}} \text{ item} \\
&= \left( 11\frac{1}{4} \right)^{\text{th}} \text{ item} \\
&= 11^{\text{th}} \text{ item} + (12^{\text{th}} \text{ item} - 11^{\text{th}} \text{ item}) \left( \frac{1}{4} \right) \\
&= 55 + (58-55) \left( \frac{1}{4} \right) \\
&= 55 + 3 \left( \frac{1}{4} \right) = 55.75 \text{ kg}
\end{aligned}$$

### Example 16

The frequency distribution of weights of 190 sorghum ear-heads are given below. Compute 25<sup>th</sup> percentile and 75<sup>th</sup> percentile.

Weight of ear-heads (in g)	No of ear heads
40-60	6
60-80	28
80-100	35
100-120	55
120-140	30
140-160	15
160-180	12
180-200	9
<b>Total</b>	<b>190</b>

### Solution

Weight of ear-heads (in g)	No of ear heads	Less than class	Cumulative frequency
40-60	6	< 60	6
60-80	28	< 80	34
80-100	35	<100	69
100-120	55	<120	124
120-140	30	<140	154
140-160	15	<160	169
160-180	12	<180	181
180-200	9	<200	190
<b>Total</b>	<b>190</b>		

47.5

142.5

or  $P_{25}$ , first find out  $\left(\frac{25(190)}{100}\right)$ , and for  $P_{75}$ ,  $\left(\frac{75(190)}{100}\right)$ , and proceed as in the case of median.

For  $P_{25}$ , we have  $\left(\frac{25(190)}{100}\right) = 47.5$ .

The value 47.5 lies between 34 and 69. Therefore, the percentile class is 80-100. Hence,

$$\begin{aligned}P_{25} = Q_1 &= l + \left( \frac{(25 \cdot n / 100) - cf}{f} \times C \right) \\&= 80 + \left( \frac{(47.5) - 34}{35} \times 20 \right) \\&= 80 + \left( \frac{(13.5)}{35} \times 20 \right) \\&= 80 + 7.71 \text{ or } 87.71 \text{ g.}\end{aligned}$$

Similarly,

$$\begin{aligned}P_{75} &= l + \left( \frac{(75 \cdot n / 100) - cf}{f} \times C \right) \text{Class} \\&= 120 + \left( \frac{(142.5) - 121}{30} \times 20 \right) \\&= 120 + \left( \frac{(21.5)}{30} \times 20 \right) \\&= 120 + 14.33 = 134.33 \text{ g.}\end{aligned}$$

## Quartiles

The quartiles divide the distribution in four parts. There are three quartiles. The second quartile divides the distribution into two halves and therefore is the same as the median. The first (lower) quartile ( $Q_1$ ) marks off the first one-fourth, the third (upper) quartile ( $Q_3$ ) marks off the three-fourth. It may be noted that the second quartile is the value of the median and 50<sup>th</sup> percentile.

## Raw or ungrouped data

First arrange the given data in the increasing order and use the formula for  $Q_1$  and  $Q_3$  then quartile deviation, Q.D is given by

$$Q.D = \frac{Q_3 - Q_1}{2}$$

Where  $Q_1 = \left(\frac{n+1}{4}\right)^{th}$  item and  $Q_3 = 3\left(\frac{n+1}{4}\right)^{th}$  item

## Example 18

Compute quartiles for the data given below (grains/panicles) 25, 18, 30, 8, 15, 5, 10, 35, 40, 45

## Solution

5, 8, 10, 15, 18, 25, 30, 35, 40, 45

$$\begin{aligned} Q_1 &= \left(\frac{n+1}{4}\right)^{th} \\ &= \left(\frac{10+1}{4}\right)^{th} \\ &= (2.75)^{th} \text{ item} \\ &= 2^{nd} \text{ item} + \left(\frac{3}{4}\right)(3^{rd} \text{ item} - 2^{nd} \text{ item}) \\ &= 8 + \frac{3}{4}(10-8) \\ &= 8 + \frac{3}{4} \times 2 \end{aligned}$$

$$= 8+1.5$$

$$= 9.5$$

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{\text{th}}$$

$$= 3 \times (2.75)^{\text{th}} \text{ item}$$

$$= (8.75)^{\text{th}} \text{ item}$$

$$= 8^{\text{th}} \text{ item} + \left(\frac{1}{4}\right)(9^{\text{th}} \text{ item} - 8^{\text{th}} \text{ item})$$

$$= 35 + \frac{1}{4}(40-35)$$

$$= 35+1.25$$

$$= 36.25$$

### Discrete Series

Step1: Find cumulative frequencies.

Step2: Find  $\left(\frac{n+1}{4}\right)$

Step3: See in the cumulative frequencies, the value just greater than  $\left(\frac{n+1}{4}\right)$ , then the corresponding value of  $x$  is  $Q_1$

Step4: Find  $3\left(\frac{n+1}{4}\right)$

Step5: See in the cumulative frequencies, the value just greater than  $3\left(\frac{n+1}{4}\right)$ , then the corresponding value of  $x$  is  $Q_3$

### Example 19

Compute quartiles for the data given bellow (insects/plant).

X	5	8	12	15	19	24	30
f	4	3	2	4	5	2	4

### Solution

x	f	cf
5	4	4
8	3	7
12	2	9
15	4	13
19	5	18
24	2	20

$$Q_1 = \left(\frac{n+1}{4}\right)^{\text{th}} \text{ item} = \left(\frac{24+1}{4}\right) = \left(\frac{25}{4}\right) = 6.25^{\text{th}} \text{ item}$$

$$Q_3 = 3\left(\frac{n+1}{4}\right)^{\text{th}} \text{ item} = 3\left(\frac{24+1}{4}\right) = 18.75^{\text{th}} \text{ item} \therefore Q_1 = 8; Q_3 = 24$$

### Continuous series

Step1: Find cumulative frequencies

Step2: Find  $\left(\frac{n}{4}\right)$

Step3: See in the cumulative frequencies, the value just greater than  $\left(\frac{n}{4}\right)$ , then the corresponding class interval is called first quartile class.

Step4: Find  $3\left(\frac{n}{4}\right)$  See in the cumulative frequencies the value just greater than  $3\left(\frac{n}{4}\right)$  then the corresponding class interval is called 3<sup>rd</sup> quartile class. Then apply the respective formulae

$$Q_1 = l_1 + \frac{\frac{n}{4} - m_1}{f_1} \times c_1$$

$$Q_3 = l_3 + \frac{3\left(\frac{n}{4}\right) - m_3}{f_3} \times c_3$$

Where  $l_1$  = lower limit of the first quartile class

$f_1$  = frequency of the first quartile class

$c_1$  = width of the first quartile class

$m_1$  = c.f. preceding the first quartile class

$l_3$  = lower limit of the 3<sup>rd</sup> quartile class

$f_3$  = frequency of the 3<sup>rd</sup> quartile class

$c_3$  = width of the 3<sup>rd</sup> quartile class

$m_3$  = c.f. preceding the 3<sup>rd</sup> quartile class

**Example 20:** The following series relates to the marks secured by students in an examination.

Marks	No. of Students
0-10	11
10-20	18
20-30	25
30-40	28
40-50	30
50-60	33
60-70	22
70-80	15
80-90	12
90-100	10

Find the quartiles

**Solution**

C.I	f	cf
0-10	11	11
10-20	18	29
20-30	25	54
30-40	28	82
40-50	30	112
50-60	33	145
60-70	22	167
70-80	15	182
80-90	12	194
90-100	10	204
	204	

$$\left(\frac{n}{4}\right) = \left(\frac{204}{4}\right) = 51$$

$$3\left(\frac{n}{4}\right) = 153$$

$$Q_1 = l_1 + \frac{\frac{n}{4} - m_1}{f_1} \times c_1$$

$$= 20 + \frac{51 - 29}{25} \times 10 = 20 + 8.8 = 28.8$$

$$Q_3 = l_3 + \frac{3\left(\frac{n}{4}\right) - m_3}{f_3} \times c_3$$

$$= 60 + \frac{153 - 145}{22} \times 12 = 60 + 4.36 = 64.36$$

### Questions

1. The middle value of an ordered series is called
- a) 2nd quartile                      b) 5th decile
- c) 50th percentile                d) all the above

**Ans: all the above**

2. For a set of values the model value can be
- a) Unimodal                        b) bimodal
- c) Trimodal                         d) All of these

**d) Ans: all the above**

3. Mode is suitable for qualitative data.

**Ans: True**

4. Decile divides the group in to ten equal parts.

**Ans: True**

5. Mean is affected by extreme values.

**Ans: True**

6. Geometric mean can be calculated for negative values.

**Ans: False**

7. Define mean and median

8. For what type of data mode can be calculated.
9. Explain how to calculate the arithmetic mean for raw and grouped data.
10. Explain how to calculate median and mode for grouped data.



## Lecture.5

### Measures of dispersion - Range, Variance -Standard deviation – co-efficient of variation - computation of the above statistics for raw and grouped data

#### Measures of Dispersion

The averages are representatives of a frequency distribution. But they fail to give a complete picture of the distribution. They do not tell anything about the scatterness of observations within the distribution.

Suppose that we have the distribution of the yields (kg per plot) of two paddy varieties from 5 plots each. The distribution may be as follows

Variety I	45	42	42	41	40
Variety II	54	48	42	33	30

It can be seen that the mean yield for both varieties is 42 kg but cannot say that the performances of the two varieties are same. There is greater uniformity of yields in the first variety whereas there is more variability in the yields of the second variety. The first variety may be preferred since it is more consistent in yield performance.

Form the above example it is obvious that a measure of central tendency alone is not sufficient to describe a frequency distribution. In addition to it we should have a measure of scatterness of observations. The scatterness or variation of observations from their average are called the dispersion. There are different measures of dispersion like the range, the quartile deviation, the mean deviation and the standard deviation.

#### Characteristics of a good measure of dispersion

An ideal measure of dispersion is expected to possess the following properties

1. It should be rigidly defined
2. It should be based on all the items.
3. It should not be unduly affected by extreme items.
4. It should lend itself for algebraic manipulation.
5. It should be simple to understand and easy to calculate

## **Range**

This is the simplest possible measure of dispersion and is defined as the difference between the largest and smallest values of the variable.

- In symbols,  $\text{Range} = L - S$ .
- Where  $L =$  Largest value.
- $S =$  Smallest value.

In individual observations and discrete series,  $L$  and  $S$  are easily identified.

In continuous series, the following two methods are followed.

### **Method 1**

$L =$  Upper boundary of the highest class

$S =$  Lower boundary of the lowest class.

### **Method 2**

$L =$  Mid value of the highest class.

$S =$  Mid value of the lowest class.

### **Example 1**

The yields (kg per plot) of a cotton variety from five plots are 8, 9, 8, 10 and 11. Find the range

### **Solution**

$L=11, S = 8$ .

$\text{Range} = L - S = 11 - 8 = 3$

### **Example 2**

Calculate range from the following distribution.

Size:	60-63	63-66	66-69	69-72	72-75
Number:	5	18	42	27	8

### **Solution**

L = Upper boundary of the highest class = 75

S = Lower boundary of the lowest class = 60

Range = L – S = 75 – 60 = 15

### **Merits and Demerits of Range**

#### **Merits**

1. It is simple to understand.
2. It is easy to calculate.
3. In certain types of problems like quality control, weather forecasts, share price analysis, etc.,  
range is most widely used.

#### **Demerits**

1. It is very much affected by the extreme items.
2. It is based on only two extreme observations.
3. It cannot be calculated from open-end class intervals.
4. It is not suitable for mathematical treatment.
5. It is a very rarely used measure.

### **Standard Deviation**

It is defined as the positive square-root of the arithmetic mean of the Square of the deviations of the given observation from their arithmetic mean.

The standard deviation is denoted by s in case of sample and Greek letter  $\sigma$  (sigma) in case of population.

The formula for calculating standard deviation is as follows

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} \quad \text{for raw data}$$

And for grouped data the formulas are

$$s = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2} \text{ for discrete data}$$

$$s = C \times \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \text{ for continuous data}$$

Where  $d = \frac{x - A}{C}$

C = class interval

### Example 3

#### Raw Data

The weights of 5 ear-heads of sorghum are 100, 102, 118, 124, 126 gms. Find the standard deviation.

#### Solution

x	x <sup>2</sup>
100	10000
102	10404
118	13924
124	15376
126	15876
<b>570</b>	<b>65580</b>

$$\begin{aligned} \text{Standard deviation } s &= \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} \\ &= \sqrt{\frac{65580 - \frac{(570)^2}{5}}{5-1}} = \sqrt{150} = 12.25 \text{ gms} \end{aligned}$$

### Example 4

#### Discrete distribution

The frequency distributions of seed yield of 50 sesamum plants are given below. Find the standard deviation.

Seed yield in gms (x)	3	4	5	6	7
Frequency (f)	4	6	15	15	10

**Solution**

Seed yield in gms (x)	f	fx	fx <sup>2</sup>
3	4	12	36
4	6	24	96
5	15	75	375
6	15	90	540
7	10	70	490
<b>Total</b>	<b>50</b>	<b>271</b>	<b>1537</b>

Here  $n = 50$

$$\begin{aligned}
 \text{Standard deviation } s &= \sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2} \\
 &= \sqrt{\frac{1537}{50} - \left(\frac{271}{50}\right)^2} \\
 &= \sqrt{30.74 - 29.3764} \\
 &= 1.1677 \text{ gms}
 \end{aligned}$$

**Example 5**  
**Continuous distribution**

The Frequency distributions of seed yield of 50 sesamum plants are given below. Find the standard deviation.

Seed yield in gms (x)	2.5-3.5	3.5-4.5	4.5-5.5	5.5-6.5	6.5-7.5
No. of plants (f)	4	6	15	15	10

**Solution**

Seed yield in gms (x)	No. of Plants f	Mid x	$d = \frac{x - A}{C}$	df	d <sup>2</sup> f
2.5-3.5	4	3	-2	-8	16
3.5-4.5	6	4	-1	-6	6
4.5-5.5	15	5	0	0	0
5.5-6.5	15	6	1	15	15

6.5-7.5	10	7	2	20	40
<b>Total</b>	<b>50</b>	<b>25</b>	<b>0</b>	<b>21</b>	<b>77</b>

A=Assumed mean = 5

n=50, C=1

$$\begin{aligned}
 s &= C \times \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2} \\
 &= 1 \times \sqrt{\frac{77}{50} - \left(\frac{21}{50}\right)^2} \\
 &= \sqrt{1.54 - 0.1764} \\
 &= \sqrt{1.3636} = 1.1677
 \end{aligned}$$

### **Merits and Demerits of Standard Deviation**

#### **Merits**

1. It is rigidly defined and its value is always definite and based on all the observations and the actual signs of deviations are used.
2. As it is based on arithmetic mean, it has all the merits of arithmetic mean.
3. It is the most important and widely used measure of dispersion.
4. It is possible for further algebraic treatment.
5. It is less affected by the fluctuations of sampling and hence stable.
6. It is the basis for measuring the coefficient of correlation and sampling.

#### **Demerits**

1. It is not easy to understand and it is difficult to calculate.
2. It gives more weight to extreme values because the values are squared up.
3. As it is an absolute measure of variability, it cannot be used for the purpose of comparison.

#### **Variance**

The square of the standard deviation is called variance

(i.e.) variance = (SD)<sup>2</sup>.

### **Coefficient of Variation**

The Standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated. The standard deviation of heights of plants cannot be compared with the standard deviation of weights of the grains, as both are expressed in different units, i.e heights in centimeter and weights in kilograms. Therefore the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure is known as the coefficient of variation. The coefficient of variation is obtained by dividing the standard deviation by the mean and expressed in percentage. Symbolically, Coefficient of

$$\text{variation (C.V)} = \frac{SD}{\text{mean}} \times 100$$

If we want to compare the variability of two or more series, we can use C.V. The series or groups of data for which the C.V. is greater indicate that the group is more variable, less stable, less uniform, less consistent or less homogeneous. If the C.V. is less, it indicates that the group is less variable or more stable or more uniform or more consistent or more homogeneous.

### **Example 6**

Consider the measurement on yield and plant height of a paddy variety. The mean and standard deviation for yield are 50 kg and 10 kg respectively. The mean and standard deviation for plant height are 55 cm and 5 cm respectively.

Here the measurements for yield and plant height are in different units. Hence the variabilities can be compared only by using coefficient of variation.

$$\text{For yield, CV} = \frac{10}{50} \times 100 = 20\%$$

$$\text{For plant height, CV} = \frac{5}{55} \times 100 = 9.1\%$$

The yield is subject to more variation than the plant height.

### Questions

1. Which measure is affected most by the presence of extreme values.

- a) Range
- b) Standard Deviation
- b) Quartile Deviation
- d) Mean deviation

**Ans: Standard Deviation**

2. Variance is square of \_\_\_\_\_

- a) Range
- b) Standard Deviation
- c) Quartile Deviation
- d) Mean deviation

**Ans: Standard Deviation**

3. If the CV of variety I is 30% and variety II is 25% then Variety II is more consistent.

**Ans: True**

4. For the set of data 5, 5, 5,5,5,5 the Standard deviation value is zero.

**Ans: True**

5. The absolute measures of dispersion will have the original units.

**Ans: True**

6. The mean deviation value for a set of data can take even negative value.

**Ans: False**

7. Define dispersion.

8. Define C.V. What are its uses?

9. What are the differences between absolute measure and relative measure of dispersion?

10. How to calculate the standard deviation for raw and grouped data?



## Lecture.6

**Probability – Basic concepts-trial- event-equally likely- mutually exclusive – independent event, additive and multiplicative laws. Theoretical distributions- discrete and continuous distributions, Binomial distributions-properties**

### Probability

The concept of probability is difficult to define in precise terms. In ordinary language, the word probable means likely (or) chance. Generally the word, probability, is used to denote the happening of a certain event, and the likelihood of the occurrence of that event, based on past experiences. By looking at the clear sky, one will say that there will not be any rain today. On the other hand, by looking at the cloudy sky or overcast sky, one will say that there will be rain today. In the earlier sentence, we aim that there will not be rain and in the latter we expect rain. On the other hand a mathematician says that the probability of rain is '0' in the first case and that the probability of rain is '1' in the second case. In between 0 and 1, there are fractions denoting the chance of the event occurring. In ordinary language, the word probability means uncertainty about happenings. In Mathematics and Statistics, a numerical measure of uncertainty is provided by the important branch of statistics – called theory of probability. Thus we can say, that the theory of probability describes certainty by 1 (one), impossibility by 0 (zero) and uncertainties by the co-efficient which lies between 0 and 1.

**Trial and Event** An experiment which, though repeated under essentially identical (or) same conditions does not give unique results but may result in any one of the several possible outcomes. Performing an experiment is known as a trial and the outcomes of the experiment are known as events.

**Example 1:** Seed germination – either germinates or does not germinates are events.

2. In a lot of 5 seeds none may germinate (0), 1 or 2 or 3 or 4 or all 5 may germinate.

### Sample space (S)

A set of all possible outcomes from an experiment is called sample space. For example, a set of five seeds are sown in a plot, none may germinate, 1, 2, 3, 4 or all five may germinate. i.e the possible outcomes are {0, 1, 2, 3, 4, 5}. The set of numbers is called a sample space. Each possible outcome (or) element in a sample space is called sample point.

### Exhaustive Events

The total number of possible outcomes in any trial is known as exhaustive events (or) exhaustive cases.

### Example

1. When pesticide is applied a pest may survive or die. There are two exhaustive cases namely ( survival, death)
2. In throwing of a die, there are six exhaustive cases, since anyone of the 6 faces  
1, 2, 3, 4, 5, 6 may come uppermost.
3. In drawing 2 cards from a pack of cards the exhaustive number of cases is  $52C_2$ , since 2 cards can be drawn out of 52 cards in  $52C_2$  ways

Trial	Random Experiment	Total number of trials	Sample Space
(1)	One pest is exposed to pesticide	$2^1=2$	{S,D}
(2)	Two pests are exposed to pesticide	$2^2=4$	{SS, SD, DS, DD}
(3)	Three pests are exposed to pesticide	$2^3=8$	{SSS, SSD, SDS, DSS, SDD, DSD, DDS, DDD}
(4)	One set of three seeds	$4^1= 4$	{0,1,2,3}
(5)	Two sets of three seeds	$4^2=16$	{0,1}, {0,2}, {0,3} etc

### **Favourable Events**

The number of cases favourable to an event in a trial is the number of outcomes which entail the happening of the event.

#### **Example**

1. When a seed is sown if we observe non germination of a seed, it is a favourable event. If we are interested in germination of the seed then germination is the favourable event.

### **Mutually Exclusive Events**

Events are said to be mutually exclusive (or) incompatible if the happening of any one of the events excludes (or) precludes the happening of all the others i.e.) if no two or more of the events can happen simultaneously in the same trial. (i.e.) The joint occurrence is not possible.

#### **Example**

1. In observation of seed germination the seed may either germinate or it will not germinate. Germination and non germination are mutually exclusive events.

### **Equally Likely Events**

Outcomes of a trial are said to be equally likely if taking in to consideration all the relevant evidences, there is no reason to expect one in preference to the others. (i.e.) Two or more events are said to be equally likely if each one of them has an **equal chance of occurring**.

### **Independent Events**

Several events are said to be independent if the happening of an event is not affected by the happening of one or more events.

#### **Example**

1. When two seeds are sown in a pot, one seed germinates. It would not affect the germination or non germination of the second seed. One event does not affect the other event.

## Dependent Events

If the happening of one event is affected by the happening of one or more events, then the events are called dependent events.

### Example

If we draw a card from a pack of well shuffled cards, if the first card drawn is not replaced then the second draw is dependent on the first draw.

**Note:** In the case of independent (or) dependent events, the joint occurrence is possible.

## Definition of Probability

### Mathematical (or) Classical (or) a-priori Probability

If an experiment results in 'n' exhaustive cases which are mutually exclusive and equally likely cases out of which 'm' events are favourable to the happening of an event 'A', then the probability 'p' of happening of 'A' is given by

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

### Note

1. If  $m = 0 \Rightarrow P(A) = 0$ , then 'A' is called an impossible event. (i.e.) also by  $P(\phi) = 0$ .
2. If  $m = n \Rightarrow P(A) = 1$ , then 'A' is called assure (or) certain event.
3. The probability is a non-negative real number and cannot exceed unity (i.e.) lies between 0 to 1.
4. The probability of non-happening of the event 'A' (i.e.)  $P(\bar{A})$ . It is denoted by 'q'.

$$P(\bar{A}) = \frac{n - m}{n} = 1 - \frac{m}{n} = 1 - P(A)$$

$$\Rightarrow q = 1 - p$$

$$\Rightarrow p + q = 1$$

$$\text{(or) } P(A) + P(\bar{A}) = 1.$$

### Statistical (or) Empirical Probability (or) a-posteriori Probability

If an experiment is repeated a number (n) of times, an event 'A' happens 'm' times then the statistical probability of 'A' is given by

$$p = P(A) = \lim_{n \rightarrow \infty} \frac{m}{n}$$

### Axioms for Probability

1. The probability of an event ranges from 0 to 1. If the event cannot take place its probability shall be '0' if it certain, its probability shall be '1'.

Let  $E_1, E_2, \dots, E_n$  be any events, then  $P(E_i) \geq 0$ .

2. The probability of the entire sample space is '1'. (i.e.)  $P(S) = 1$ .

$$\text{Total Probability, } \sum_{i=1}^n P(E_i) = 1$$

3. If A and B are mutually exclusive (or) disjoint events then the probability of occurrence of either A (or) B denoted by  $P(A \cup B)$  shall be given by

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

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = P(E_1) + P(E_2) + \dots + P(E_n)$$

If  $E_1, E_2, \dots, E_n$  are mutually exclusive events.

**Example 1:** Two dice are tossed. What is the probability of getting (i) Sum 6 (ii) Sum 9?

### Solution

When 2 dice are tossed. The exhaustive number of cases is 36 ways.

**(i) Sum 6** =  $\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$

$\therefore$  Favourable number of cases = 5

$$P(\text{Sum 6}) = \frac{5}{36}$$

**(ii) Sum 9** =  $\{(3, 6), (4, 5), (5, 4), (6, 3)\}$

$\therefore$  Favourable number of cases = 4

$$P(\text{Sum 9}) = \frac{4}{36} = \frac{1}{9}$$

**Example 2:** A card is drawn from a pack of cards. What is a probability of getting (i) a king (ii) a spade (iii) a red card (iv) a numbered card?

**Solution**

There are 52 cards in a pack.

One can be selected in  $52C_1$  ways.

$\therefore$  Exhaustive number of cases is  $= 52C_1 = 52$ .

**(i) A king**

There are 4 kings in a pack.

One king can be selected in  $4C_1$  ways.

$\therefore$  Favourable number of cases is  $= 4C_1 = 4$

Hence the probability of getting a king  $= \frac{4}{52}$

**(ii) A spade**

There are 13 spades in a pack.

One spade can be selected in  $13C_1$  ways.

$\therefore$  Favourable number of cases is  $= 13C_1 = 13$

Hence the probability of getting a spade  $= \frac{13}{52}$

**(iii) A red card**

There are 26 red cards in a pack.

One red card can be selected in  $26C_1$  ways.

$\therefore$  Favourable number of cases is  $= 26C_1 = 26$

Hence the probability of getting a red card  $= \frac{26}{52}$

**(iv) A numbered card**

There are 36 numbered cards in a pack.

One numbered card can be selected in  $36C_1$  ways.

$\therefore$  Favourable number of cases is  $= 36C_1 = 36$

Hence the probability of getting a numbered card =  $\frac{36}{52}$

**Example 3:** What is the probability of getting 53 Sundays when a leap year selected at random?

**Solution**

A leap year consists of 366 days.  
This has 52 full weeks and 2 days remained.  
The remaining 2 days have the following possibilities.

- (i) Sun, Mon (ii) Mon, Tues (iii) Tues, Wed (iv) Wed, Thurs (v) Thurs, Fri (vi) Fri, Sat (vii) Sat, Sun.

In order that a lap year selected at random should contain 53 Sundays, one of the 2 over days must be Sunday.

- ∴ Exhaustive number of cases is = 7
- ∴ Favourable number of cases is = 2
- ∴ Required Probability is =  $\frac{2}{7}$

**Conditional Probability**

Two events A and B are said to be dependent, when B can occur only when A is known to have occurred (or vice versa). The probability attached to such an event is called the conditional probability and is denoted by P (A/B) (read it as: A given B) or, in other words, probability of A given that B has occurred.

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)}$$

If two events A and B are **dependent**, then the conditional probability of B given A is,

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(AB)}{P(A)}$$

## Theorems of Probability

There are two important theorems of probability namely,

1. The addition theorem on probability
2. The multiplication theorem on probability.

### I. Addition Theorem on Probability

(i) Let A and B be any two events which are **not mutually exclusive**

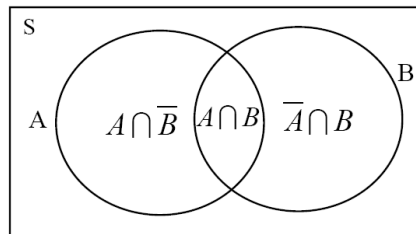
$$\begin{aligned}P(A \text{ or } B) &= P(A \cup B) = P(A + B) = P(A) + P(B) - P(A \cap B) \quad (\text{or}) \\ &= P(A) + P(B) - P(AB)\end{aligned}$$

#### Proof

Let us take a random experiment with a sample space S of N sample points.

Then by the definition of probability ,

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{n(A \cup B)}{N}$$



From the diagram, using the axiom for the mutually exclusive events, we write

$$P(A \cup B) = \frac{n(A) + n(\bar{A} \cap B)}{N}$$

Adding and subtracting  $n(A \cap B)$  in the numerator,

$$= \frac{n(A) + n(\bar{A} \cap B) + n(A \cap B) - n(A \cap B)}{N}$$

$$= \frac{n(A) + n(B) - n(A \cap B)}{N}$$

$$= \frac{n(A)}{N} + \frac{n(B)}{N} - \frac{n(A \cap B)}{N}$$

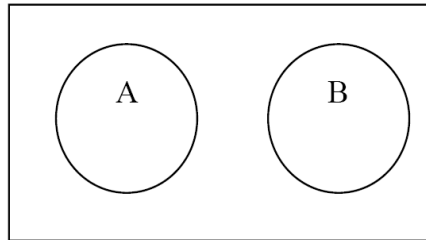
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



(ii) Let A and B be any two events which are **mutually exclusive**

$$P(A \text{ or } B) = P(A \cup B) = P(A + B) = P(A) + P(B)$$

**Proof**



We know that,  $n(A \cup B) = n(A) + n(B)$

$$\begin{aligned} P(A \cup B) &= \frac{n(A \cup B)}{n} \\ &= \frac{n(A) + n(B)}{n} \\ &= \frac{n(A)}{n} + \frac{n(B)}{n} \end{aligned}$$

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

**Note**

(i) In the case of 3 events, **(not mutually exclusive events)**

$$P(A \text{ or } B \text{ or } C) = P(A \cup B \cup C) = P(A + B + C)$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

(ii) In the case of 3 events, **(mutually exclusive events)**

$$P(A \text{ or } B \text{ or } C) = P(A \cup B \cup C) = P(A + B + C) = P(A) + P(B) + P(C)$$

**Example**

Using the additive law of probability we can find the probability that in one roll of a die, we will obtain either a one-spot or a six-spot. The probability of obtaining a one-spot is  $1/6$ . The probability of obtaining a six-spot is also  $1/6$ . The probability of rolling

a die and getting a side that has both a one-spot with a six-spot is 0. There is no side on a die that has both these events. So substituting these values into the equation gives the following result:

$$\frac{1}{6} + \frac{1}{6} - 0 = \frac{2}{6} = \frac{1}{3} = 0.3333$$

Finding the probability of drawing a 4 of hearts or a 6 of any suit using the additive law of probability would give the following:

$$\frac{1}{52} + \frac{4}{52} - 0 = \frac{5}{52} = 0.0962$$

There is only a single 4 of hearts, there are 4 sixes in the deck and there isn't a single card that is both the 4 of hearts and a six of any suit.

Now using the additive law of probability, you can find the probability of drawing either a king or any club from a deck of shuffled cards. The equation would be completed like this:

$$\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = 0.3077$$

There are 4 kings, 13 clubs, and obviously one card is both a king and a club. We don't want to count that card twice, so you must subtract one of its occurrences away to obtain the result.

## II. Multiplication Theorem on Probability

(i) If A and B be any two events which are **not independent**, then (i.e.) **dependent**.

$$\begin{aligned} P(A \text{ and } B) &= P(A \cap B) = P(AB) = P(A) \cdot P(B/A) \longrightarrow (I) \\ &= P(B) \cdot P(A/B) \longrightarrow (II) \end{aligned}$$

Where P (B/A) and P (A/B) are the conditional probability of B given A and A given B respectively.

### Proof

Let  $n$  is the total number of events

$n(A)$  is the number of events in  $A$

$n(B)$  is the number of events in  $B$

$n(A \cup B)$  is the number of events in  $(A \cup B)$

$n(A \cap B)$  is the number of events in  $(A \cap B)$

$$\begin{aligned}P(A \cap B) &= \frac{n(A \cap B)}{n} \\ &= \frac{n(A \cap B)}{n} \times \frac{n(A)}{n(A)} \\ &= \frac{n(A)}{n} \times \frac{n(A \cap B)}{n(A)}\end{aligned}$$

$$P(A \cap B) = P(A) \cdot P(B/A) \longrightarrow (I)$$

$$\begin{aligned}P(A \cap B) &= \frac{n(A \cap B)}{n} \\ &= \frac{n(A \cap B)}{n} \times \frac{n(B)}{n(B)} \\ &= \frac{n(B)}{n} \times \frac{n(A \cap B)}{n(B)}\end{aligned}$$

$$P(A \cap B) = P(B) \cdot P(A/B) \longrightarrow (II)$$

(ii) If  $A$  and  $B$  be any two events which are **independent**, then,

$$P(B/A) = P(B) \text{ and } P(A/B) = P(A)$$

$$P(A \text{ and } B) = P(A \cap B) = P(AB) = P(A) \cdot P(B)$$

### Note

(i) In the case of 3 events, **(dependent)**

$$P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/AB)$$

(ii) In the case of 3 events, **(independent)**

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

### Example

So in finding the probability of drawing a 4 and then a 7 from a well shuffled deck of cards, this law would state that we need to multiply those separate probabilities together. Completing the equation above gives:

$$p(4 \text{ and } 7) = \frac{4}{52} \times \frac{4}{52} = \frac{16}{2704} = 0.0059$$

Given a well shuffled deck of cards, what is the probability of drawing a Jack of Hearts, Queen of Hearts, King of Hearts, Ace of Hearts, and 10 of Hearts?

$$p(10, J, Q, K, A \text{ of hearts}) = \frac{1}{52} \times \frac{1}{52} \times \frac{1}{52} \times \frac{1}{52} \times \frac{1}{52} = 0.0000000026$$

In any case, given a well shuffled deck of cards, obtaining this assortment of cards, drawing one at a time and returning it to the deck would be highly unlikely (it has an exceedingly low probability).

### Questions

1. Probability is expressed as

- (a) Ratio (b) percentage  
(c) Proportion (d) all the above

**Ans: all the above**

2. Probability can take values from

- (a)  $-\infty$  to  $+\infty$  (b)  $-\infty$  to 1  
(c) 0 to +1 (d) -1 to +1

**Ans: 0 to +1**

3. The probability of a sure event is One.

**Ans: True**

4. If A and B are mutually exclusive events, then  $P(A \cup B) = \underline{P(A) + P(B)}$

5. An integer is chosen from 1 to 20. The probability that the number is divisible by 4 is  $\frac{1}{4}$ .

**Ans: True**

6. Mean of the Binomial Distribution is  $npq$ .

**Ans: False**

7. Define an independent event.

8. What is conditional probability?

9. State the addition and multiplication laws.

10. State the properties of Binomial distribution.

## Lecture.7

### Poisson Distributions - properties, Normal Distributions- properties

#### Theoretical Distributions

Theoretical distributions are

- |                          |   |                         |
|--------------------------|---|-------------------------|
| 1. Binomial distribution | } | Discrete distribution   |
| 2. Poisson distribution  |   |                         |
| 3. Normal distribution   | → | Continuous distribution |

#### Discrete Probability distribution

##### Bernoulli distribution

A random variable  $x$  takes two values 0 and 1, with probabilities  $q$  and  $p$  ie.,  $p(x=1) = p$  and  $p(x=0)=q$ ,  $q=1-p$  is called a Bernoulli variate and is said to be Bernoulli distribution where  $p$  and  $q$  are probability of success and failure. It was given by Swiss mathematician James Bernoulli (1654-1705)

##### Example

- Tossing a coin(head or tail)
- Germination of seed(germinate or not)

##### Binomial distribution

Binomial distribution was discovered by James Bernoulli (1654-1705). Let a random experiment be performed repeatedly and the occurrence of an event in a trial be called as success and its non-occurrence is failure. Consider a set of  $n$  independent trails ( $n$  being finite), in which the probability  $p$  of success in any trail is constant for each trial. Then  $q=1-p$  is the probability of failure in any trail.

The probability of  $x$  success and consequently  $n-x$  failures in  $n$  independent trials. But  $x$  successes in  $n$  trials can occur in  ${}^n C_x$  ways. Probability for each of these ways is  $p^x q^{n-x}$ .

$$\begin{aligned} P(\text{sss}\dots\text{ff}\dots\text{fsf}\dots\text{f}) &= p(s)p(s)\dots p(f)p(f)\dots \\ &= p,p\dots q,q\dots \\ &= (p,p\dots p)(q,q\dots q) \\ &\quad (x \text{ times}) (n-x \text{ times}) \end{aligned}$$

Hence the probability of  $x$  success in  $n$  trials is given by

$${}^n C_x p^x q^{n-x}$$

**Definition**

A random variable  $x$  is said to follow binomial distribution if it assumes non-negative values and its probability mass function is given by

$$P(X=x) = p(x) = \begin{cases} {}^n C_x p^x q^{n-x}, & x=0,1,2,\dots,n \\ q=1-p \\ 0, & \text{otherwise} \end{cases}$$

The two independent constants  $n$  and  $p$  in the distribution are known as the parameters of the distribution.

**Condition for Binomial distribution**

We get the binomial distribution under the following experimentation conditions

1. The number of trial  $n$  is finite
2. The trials are independent of each other.

3. The probability of success  $p$  is constant for each trial.
4. Each trial must result in a success or failure.
5. The events are discrete events.

### Properties

1. If  $p$  and  $q$  are equal, the given binomial distribution will be symmetrical. If  $p$  and  $q$  are not equal, the distribution will be skewed distribution.
2. Mean =  $E(x) = np$
3. Variance =  $V(x) = npq$  (mean > variance)

### Application

1. Quality control measures and sampling process in industries to classify items as defectives or non-defective.
2. Medical applications such as success or failure, cure or no-cure.

### Example 1

Eight coins are tossed simultaneously. Find the probability of getting atleast six heads.

### Solution

Here number of trials,  $n = 8$ ,  $p$  denotes the probability of getting a head.

$$\therefore p = \frac{1}{2} \text{ and } q = \frac{1}{2}$$

If the random variable  $X$  denotes the number of heads, then the probability of a success in  $n$  trials is given by

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

$$= 8C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{8-x} = 8C_x \left(\frac{1}{2}\right)^8$$

$$= \frac{1}{2^8} 8C_x$$



Probability of getting atleast six heads is given by

$$\begin{aligned}P(x \geq 6) &= P(x = 6) + P(x = 7) + P(x = 8) \\&= \frac{1}{2^8} 8C_6 + \frac{1}{2^8} 8C_7 + \frac{1}{2^8} 8C_8 \\&= \frac{1}{2^8} [8C_6 + 8C_7 + 8C_8] \\&= \frac{1}{2^8} [28 + 8 + 1] = \frac{37}{256}\end{aligned}$$

**Example 2** Ten coins are tossed simultaneously. Find the probability of getting (i) atleast seven heads (ii) exactly seven heads (iii) atleast seven heads

**Solution**

$$p = \text{Probability of getting a head} = \frac{1}{2}$$

$$q = \text{Probability of not getting a head} = \frac{1}{2}$$

The probability of getting  $x$  heads throwing 10 coins simultaneously is given by

$$\begin{aligned}P(X = x) &= {}^n C_x p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n \\&= {}^{10} C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = {}^{10} C_x \left(\frac{1}{2}\right)^{10} \\&= \frac{1}{2^{10}} {}^{10} C_x\end{aligned}$$

i) Probability of getting atleast seven heads

$$\begin{aligned}P(x \geq 7) &= P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10) \\&= \frac{1}{2^{10}} [{}^{10} C_7 + {}^{10} C_8 + {}^{10} C_9 + {}^{10} C_{10}] \\&= \frac{1}{1024} [120 + 45 + 10 + 1] = \frac{176}{1024}\end{aligned}$$

ii) Probability of getting exactly 7 heads

$$P(x = 7) = \frac{1}{2^{10}} {}^{10} C_7 = \frac{1}{2^{10}} (120) = \frac{120}{1024}$$

iii) Probability of getting atleast 7 heads

$$P(x \leq 7) = 1 - P(x > 7)$$

$$\begin{aligned}
&= 1 \text{ symbol } \{P(x = 8) + P(x = 9) + P(x = 10)\} \\
&= 1 - \frac{1}{2^{10}} [10C_8 + 10C_9 + 10C_{10}] \\
&= 1 - \frac{1}{2^{10}} [45 + 10 + 1] \\
&= 1 - \frac{56}{1024} \\
&= \frac{968}{1024}
\end{aligned}$$

**Example 3:** 20 wrist watches in a box of 100 are defective. If 10 watches are selected at random, find the probability that (i) 10 are defective (ii) 10 are good (iii) at least one watch is defective (iv) at most 3 are defective.

**Solution**

20 out of 100 wrist watches are defective

Probability of defective wrist watch,  $p = \frac{20}{100} = \frac{1}{5}$

$q = 1 - p = \frac{4}{5}$

Since 10 watches are selected at random,  $n = 10$

$P(X = x) = {}^n C_x p^x q^{n-x}$ ,  $x = 0, 1, 2, \dots, 10$

$$= {}^{10} C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{10-x}$$

i) Probability of selecting 10 defective watches

$$P(x = 10) = {}^{10} C_{10} \left(\frac{1}{5}\right)^{10} \left(\frac{4}{5}\right)^0 = 1 \cdot \frac{1}{5^{10}} \cdot 1 = \frac{1}{5^{10}}$$

ii) Probability of selecting 10 good watches (i.e. no defective)

$$\begin{aligned}
P(x = 0) &= {}^{10} C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10} \\
&= 1 \cdot 1 \left(\frac{4}{5}\right)^{10} = \left(\frac{4}{5}\right)^{10}
\end{aligned}$$

iii) Probability of selecting at least one defective watch

$$P(x \geq 1) = 1 - P(x < 1)$$

$$= 1 - P(x = 0)$$

$$= 1 - {}^{10}C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10}$$

$$= 1 - \left(\frac{4}{5}\right)^{10}$$

iv) Probability of selecting at most 3 defective watches

$$P(x \leq 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$$

$$= {}^{10}C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{10} + {}^{10}C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^9 + {}^{10}C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^8 + {}^{10}C_3 \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^7$$

□□□□□□□□□□□□□□□□

$$1.1 \left(\frac{4}{5}\right)^{10} + 10 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^9 + \frac{10.9}{1.2} \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^8 + \frac{10.9.8}{1.2.3} \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^7$$

$$= 1. (0.107) + 10 (0.026) + 45 (0.0062) + 120 (0.0016)$$

$$= 0.859 \text{ (approx)}$$

### 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.

### Definition

The probability that exactly x events will occur in a given time is as follows

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

called as probability mass function of Poisson distribution.

where  $\lambda$  is the average number of occurrences per unit of time

$$\lambda = np$$

### **Condition for Poisson distribution**

Poisson distribution is the limiting case of binomial distribution under the following assumptions.

1. The number of trials  $n$  should be indefinitely large ie.,  $n \rightarrow \infty$
2. The probability of success  $p$  for each trial is indefinitely small.
3.  $np = \lambda$ , should be finite where  $\lambda$  is constant.

### **Properties**

1. Poisson distribution is defined by single parameter  $\lambda$ .
2. Mean =  $\lambda$
3. Variance =  $\lambda$ . Mean and Variance are equal.

### **Application**

1. It is used in quality control statistics to count the number of defects of an item.
2. In biology, to count the number of bacteria.
3. In determining the number of deaths in a district in a given period, by rare disease.
4. The number of error per page in typed material.
5. The number of plants infected with a particular disease in a plot of field.
6. Number of weeds in particular species in different plots of a field.

**Example 4:** Suppose on an average 1 house in 1000 in a certain district has a fire during a year. If there are 2000 houses in that district, what is the probability that exactly 5 houses will have a fire during the year? [given that  $e^{-2} = 0.13534$ ]

**Solution:**

$$\begin{aligned} \text{Mean, } \bar{x} &= np, n = 2000 \text{ and } p = \frac{1}{1000} \\ &= 2000 \times \frac{1}{1000} \end{aligned}$$

$$\lambda = 2$$

The Poisson distribution is

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

$$\begin{aligned} P(X = 5) &= \frac{e^{-2} 2^5}{5!} \\ &= \frac{(0.13534) \times 32}{120} \\ &= \mathbf{0.036} \end{aligned}$$

**Example 5**

If 2% of electric bulbs manufactured by a certain company are defective. Find the probability that in a sample of 200 bulbs i) less than 2 bulbs ii) more than 3 bulbs are defective. [e<sup>-4</sup> = 0.0183]

**Solution**

$$\text{The probability of a defective bulb} = p = \frac{2}{100} = 0.02$$

Given that n = 200 since p is small and n is large

We use the Poisson distribution

$$\text{mean, } m = np = 200 \times 0.02 = 4$$

$$\text{Now, Poisson Probability function, } P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

i) Probability of less than 2 bulbs are defective

$$\begin{aligned}
&= P(X < 2) \\
&= P(x = 0) + P(x = 1) \\
&= e^{-4} + e^{-4} (4) \\
&= e^{-4} (1 + 4) = 0.0183 \times 5 \\
&= 0.0915
\end{aligned}$$

ii) Probability of getting more than 3 defective bulbs

$$\begin{aligned}
P(x > 3) &= 1 - P(x \leq 3) \\
&= 1 - \{P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)\} \\
&= 1 - e^{-4} \left\{ 1 + 4 + \frac{4^2}{2!} + \frac{4^3}{3!} \right\} \\
&= 1 - \{0.0183 \times (1 + 4 + 8 + 10.67)\} \\
&= 0.567
\end{aligned}$$

### Normal distribution

Continuous Probability distribution is normal distribution. It is also known as error law or Normal law or Laplacian law or Gaussian distribution. Many of the sampling distribution like student-t, f distribution and  $\chi^2$  distribution.

### Definition

A continuous random variable  $x$  is said to be a normal distribution with parameters  $\mu$  and  $\sigma^2$ , if the density function is given by the probability law

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

### Note

The mean  $\mu$  and standard deviation  $\sigma$  are called the parameters of Normal distribution. The normal distribution is expressed by  $X \sim N(\mu, \sigma^2)$

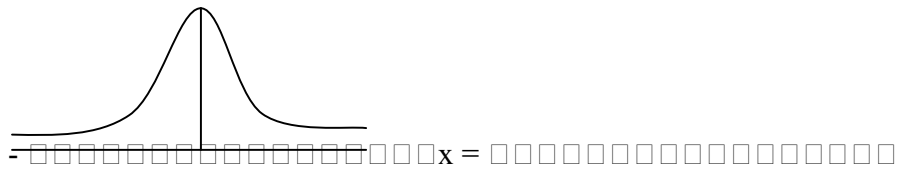
### Condition of Normal Distribution

i) Normal distribution is a limiting form of the binomial distribution under the following conditions.

- a)  $n$ , the number of trials is indefinitely large i.e.,  $n \rightarrow \infty$  and
  - b) Neither  $p$  nor  $q$  is very small.
- ii) Normal distribution can also be obtained as a limiting form of Poisson distribution with parameter  $m \rightarrow \infty$
- iii) Constants of normal distribution are mean =  $\mu$ , variation =  $\sigma^2$ , Standard deviation =  $\sigma$ .

**Normal probability curve**

The curve representing the normal distribution is called the normal probability curve. The curve is symmetrical about the mean ( $\mu$ ), bell-shaped and the two tails on the right and left sides of the mean extends to the infinity. The shape of the curve is shown in the following figure.



**Properties of normal distribution**

1. The normal curve is bell shaped and is symmetric at  $x = \mu$ .
2. Mean, median, and mode of the distribution are coincide  
i.e., Mean = Median = Mode =  $\mu$
3. It has only one mode at  $x = \mu$  (i.e., unimodal)
4. The points of inflection are at  $x = \mu \pm \sigma$
5. The maximum ordinate occurs at  $x = \mu$  and its value is  $= \frac{1}{\sigma\sqrt{2\pi}}$
6. Area Property  $P(\mu - \sigma < x < \mu + \sigma) = 0.6826$   
 $P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9544$   
 $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

### Standard Normal distribution

Let  $X$  be random variable which follows normal distribution with mean  $\mu$  and variance  $\sigma^2$ . The standard normal variate is defined as  $Z = \frac{X - \mu}{\sigma}$  which follows standard normal distribution with mean 0 and standard deviation 1 i.e.,  $Z \sim N(0,1)$ . The

standard normal distribution is given by  $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ ;  $-\infty < z < \infty$

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

### Note

#### Property of $\phi(z)$

1.  $\phi(-z) = 1 - \phi(z)$
2.  $P(a \leq Z \leq b) = \phi(b) - \phi(a)$

**Example 6:** In a normal distribution whose mean is 12 and standard deviation is 2. Find the probability for the interval from  $x = 9.6$  to  $x = 13.8$

### Solution

Given that  $Z \sim N(12, 4)$

$$\begin{aligned} P(9.6 \leq Z \leq 13.8) &= P\left(\frac{9.6 - 12}{2} \leq Z \leq \frac{13.8 - 12}{2}\right) \\ &= P(-1.2 \leq Z \leq 0) + P(0 \leq Z \leq 0.9) \\ &= P(0 \leq Z \leq 1.2) + P(0 \leq Z \leq 0.9) \quad [\text{by using symmetric property}] \\ &= 0.3849 + 0.3159 \\ &= 0.7008 \end{aligned}$$

When it is converted to percentage (ie) 70% of the observations are covered between 9.6 to 13.8.



**Example 7:** For a normal distribution whose mean is 2 and standard deviation 3. Find the value of the variate such that the probability of the variate from the mean to the value is 0.4115

**Solution:**

Given that  $Z \sim N(2, 9)$

To find  $X_1$ :

We have  $P(2 \leq Z \leq X_1) = 0.4115$

$$P\left(\frac{2-2}{3} \leq \frac{X-\mu}{\sigma} \leq \frac{X_1-2}{3}\right) = 0.4115$$

$$P(0 \leq Z \leq Z_1) = 0.4115 \text{ where } Z_1 = \frac{X_1 - 2}{3}$$

[From the normal table where 0.4115 lies is the value of  $Z_1$ ]

From the normal table we have  $Z_1 = 1.35$

$$\therefore 1.35 = \frac{X_1 - 2}{3}$$

$$\Rightarrow 3(1.35) + 2 = X_1$$

$$= X_1 = 6.05$$

(i.e) 41 % of the observation converged between 2 and 6.05

### Questions

1. For a Poisson distribution

- (a) mean > variance                      (b) mean = variance  
 (c) mean < variance                      (d) mean < variance

**Ans: mean = variance**

2. In normal distribution, skewness is

- (a) one    (b) zero  
 (c) greater than one                      (d) less than one

**Ans: zero**

3. Poisson distribution is a distribution for rare events

**Ans: True**

4. The total area under normal probability curve is one.

**Ans: True**

5. Poisson distribution is for continuous variable.

**Ans: False**

6. In a symmetrical curve mean, median and mode will coincide.

**Ans: True**

7. Give any two examples of Poisson distribution

8. The variance of a Poisson Distribution is 0.5. Find  $P(x = 3)$ .

**$[e^{-0.5} = 0.6065]$**

9. The customer accounts of a certain departmental store have an average balance of Rs.1200 and a standard deviation of Rs.400. Assuming that the account balances are normally distributed. (i) what percentage of the accounts is over Rs.1500? (ii) What percentage of the accounts is between Rs.1000 and Rs.1500? (iii) What percentage of the accounts is below Rs.1500?

10. State the Properties of normal distribution

## Lecture.8

### Sampling-basic concepts- sampling vs complete enumeration parameter and statistic-sampling methods-simple random sampling and stratified random sampling

## Sampling

### Population (Universe)

Population means aggregate of all possible units. It need not be human population. It may be population of plants, population of insects, population of fruits, etc.

### Finite population

When the number of observation can be counted and is definite, it is known as finite population

- No. of plants in a plot.
- No. of farmers in a village.
- All the fields under a specified crop.

### Infinite population

When the number of units in a population is innumerably large, that we cannot count all of them, it is known as infinite population.

- The plant population in a region.
- The population of insects in a region.

### Frame

A list of all units of a population is known as frame.

### 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  $\mu$ , standard deviation by  $\sigma$  and variance by  $\sigma^2$ .

## **Sample**

A portion or small number of unit of the total population is known as sample.

- All the farmers in a village(population) and a few farmers(sample)
- All plants in a plot is a population of plants.
- A small number of plants selected out of that population is a sample of plants.

## **Statistic**

A summary measure that describes the characteristic of the sample is known as statistic. Thus sample mean, sample standard deviation etc is statistic. The statistic is usually denoted by roman letter.

$\bar{x}$  - sample mean

s – standard deviation

The statistic is a random variable because it varies from sample to sample.

## **Sampling**

The method of selecting samples from a population is known as sampling.

### **Sampling technique**

There are two ways in which the information is collected during statistical survey.

They are

1. Census survey
2. Sampling survey

## **Census**

It is also known as population survey and complete enumeration survey. Under census survey the information are collected from each and every unit of the population or universe.

## **Sample survey**

A sample is a part of the population. Information are collected from only a few units of a population and not from all the units. Such a survey is known as sample survey.

Sampling technique is universal in nature, consciously or unconsciously it is adopted in every day life.

For eg.

1. A handful of rice is examined before buying a sack.
2. We taste one or two fruits before buying a bunch of grapes.
3. To measure root length of plants only a portion of plants are selected from a plot.

### **Need for sampling**

The sampling methods have been extensively used for a variety of purposes and in great diversity of situations.

In practice it may not be possible to collect information on all units of a population due to various reasons such as

1. Lack of resources in terms of money, personnel and equipment.
2. The experimentation may be destructive in nature. Eg- finding out the germination percentage of seed material or in evaluating the efficiency of an insecticide the experimentation is destructive.
3. The data may be wasteful if they are not collected within a time limit. The census survey will take longer time as compared to the sample survey. Hence for getting quick results sampling is preferred. Moreover a sample survey will be less costly than complete enumeration.
4. Sampling remains the only way when population contains infinitely many number of units.
5. Greater accuracy.

### **Sampling methods**

The various methods of sampling can be grouped under

- 1) Probability sampling or random sampling

## 2) Non-probability sampling or non random sampling

### **Random sampling**

Under this method, every unit of the population at any stage has equal chance (or) each unit is drawn with known probability. It helps to estimate the mean, variance etc of the population.

Under probability sampling there are two procedures

1. Sampling with replacement (SWR)
2. 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 finite sampling with replacement is adopted otherwise SWOR is adopted.

Mainly there are many kinds of random sampling. Some of them are.

1. Simple Random Sampling
2. Systematic Random Sampling
3. Stratified Random Sampling
4. Cluster Sampling

### **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  $NC_n$  possible samples of size  $n$ . If all possible samples have an equal probability  $1/NC_n$  of being drawn, the sampling is said to be simple random sampling.

There are two methods in SRS

1. Lottery method
2. Random no. table method

### **Lottery method**

This is the 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 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, roll 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 a table of random numbers.

There are several standard tables of random numbers. But the credit for this technique goes to Prof. LHC. Tippett (1927). The random number table consists of 10,400 four-figured numbers. There are various other random numbers. They are Fishers and Yates (1938) 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 random numbers, Rand corporation (1955) consisting of 2, 00,000 random numbers of 5 digits each etc.,

### **Merits**

1. There is less chance for personal bias.
2. Sampling error can be measured.
3. This method is economical as it saves time, money and labour.

### **Demerits**

1. It cannot be applied if the population is heterogeneous.
2. This requires a complete list of the population but such up-to-date lists are not available in many enquires.
3. If the size of the sample is small, then it will not be a representative of the population.

### **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

1. Equal allocation
2. Proportional allocation
3. Neyman's allocation
4. Optimum allocation

If the number of units to be selected is uniform in all strata it is known as equal allocation of samples.

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.

When the cost per unit varies from stratum to stratum, it is known as optimum allocation.

When the costs for different strata are equal, it is known as Neyman's allocation.

### **Merits**

1. It is more representative.
2. It ensures greater accuracy.
3. It is easy to administrate as the universe is sub-divided.

### **Demerits**

1. To divide the population into homogeneous strata, it requires more money, time and statistical experience which is a difficult one.
2. If proper stratification is not done, the sample will have an effect of bias.

### Questions

1. If each and every unit of population has equal chance of being included in the sample, it is known as

- (a) Restricted sampling                      (b) Purposive sampling  
(c) Simple random sampling              (d) None of the above

**Ans: Simple random sampling**

2. In a population of size 10 the possible number of samples of size 2 will be

- (a) 45                      (b) 40                      (c) 54                      (d) None of the above

**Ans: 45**

3. A population consisting of an unlimited number of units is called an infinite population.

**Ans: True**

4. If all the units of a population are surveyed it is called census.

**Ans: True**

5. Random numbers are used for selecting the samples in simple random sampling method.

**Ans: True**

6. The list of all units in a population is called as Frame.

**Ans: True**

7. What is sampling?

8. Explain the Lottery method.

9. Explain the method of selection of samples in simple random sampling.

10. Explain the method of selection of samples in Stratified random sampling.

## Lecture.9

**Test of significance – Basic concepts – null hypothesis – alternative hypothesis – level of significance – Standard error and its importance – steps in testing**

### Test of Significance

#### Objective

To familiarize the students about the concept of testing of any hypothesis, the different terminologies used in testing and application of different types of tests.

#### Sampling Distribution

By drawing all possible samples of same size from a population we can calculate the statistic, for example,  $\bar{x}$  for all samples. Based on this we can construct a frequency distribution and the probability distribution of  $\bar{x}$ . Such probability distribution of a statistic is known a sampling distribution of that statistic. In practice, the sampling distributions can be obtained theoretically from the properties of random samples.

#### Standard Error

As in the case of population distribution the characteristic of the sampling distributions are also described by some measurements like mean & standard deviation. Since a statistic is a random variable, the mean of the sampling distribution of a statistic is called the expected valued of the statistic. The SD of the sampling distributions of the statistic is called standard error of the Statistic. The square of the standard error is known as the variance of the statistic. It may be noted that the standard deviation is for units whereas the standard error is for the statistic.

#### Theory of Testing Hypothesis

##### Hypothesis

Hypothesis is a statement or assumption that is yet to be proved.

##### Statistical Hypothesis

When the assumption or statement that occurs under certain conditions is formulated as scientific hypothesis, we can construct criteria by which a scientific hypothesis is either rejected or provisionally accepted. For this purpose, the scientific hypothesis is translated into statistical language. If the hypothesis is given in a statistical language it is called a statistical hypothesis.

For eg:-

The yield of a new paddy variety will be 3500 kg per hectare – scientific hypothesis.

In Statistical language it may be stated as the random variable (yield of paddy) is distributed normally with mean 3500 kg/ha.

### **Simple Hypothesis**

When a hypothesis specifies all the parameters of a probability distribution, it is known as simple hypothesis. The hypothesis specifies all the parameters, i.e  $\mu$  and  $\sigma$  of a normal distribution.

Eg:-

The random variable  $x$  is distributed normally with mean  $\mu=0$  &  $SD=1$  is a simple hypothesis. The hypothesis specifies all the parameters ( $\mu$  &  $\sigma$ ) of a normal distributions.

### **Composite Hypothesis**

If the hypothesis specifies only some of the parameters of the probability distribution, it is known as composite hypothesis. In the above example if only the  $\mu$  is specified or only the  $\sigma$  is specified it is a composite hypothesis.

### **Null Hypothesis - $H_0$**

Consider for example, the hypothesis may be put in a form 'paddy variety A will give the same yield per hectare as that of variety B' or there is no difference between the average yields of paddy varieties A and B. These hypotheses are in definite terms. Thus these hypothesis form a basis to work with. Such a working hypothesis is known as null hypothesis. It is called null hypothesis because it nullifies the original hypothesis, that variety A will give more yield than variety B.

The null hypothesis is stated as 'there is no difference between the effect of two treatments or there is no association between two attributes (ie) the two attributes are independent. Null hypothesis is denoted by  $H_0$ .

Eg:-

There is no significant difference between the yields of two paddy varieties (or) they give same yield per unit area. Symbolically,  $H_0: \mu_1=\mu_2$ .

## **Alternative Hypothesis**

When the original hypothesis is  $\mu_1 > \mu_2$  stated as an alternative to the null hypothesis is known as alternative hypothesis. Any hypothesis which is complementary to null hypothesis is called alternative hypothesis, usually denoted by  $H_1$ .

Eg:-

There is a significance difference between the yields of two paddy varieties. Symbolically,

$$H_1: \mu_1 \neq \mu_2 \text{ (two sided or directionless alternative)}$$

If the statement is that A gives significantly less yield than B (or) A gives significantly more yield than B. Symbolically,

$$H_1: \mu_1 < \mu_2 \text{ (one sided alternative-left tailed)}$$

$$H_1: \mu_1 > \mu_2 \text{ (one sided alternative-right tailed)}$$

## **Testing of Hypothesis**

Once the hypothesis is formulated we have to make a decision on it. A statistical procedure by which we decide to accept or reject a statistical hypothesis is called testing of hypothesis.

### **Sampling Error**

From sample data, the statistic is computed and the parameter is estimated through the statistic. The difference between the parameter and the statistic is known as the sampling error.

### **Test of Significance**

Based on the sampling error the sampling distributions are derived. The observed results are then compared with the expected results on the basis of sampling distribution. If the difference between the observed and expected results is more than specified quantity of the standard error of the statistic, it is said to be significant at a specified probability level. The process up to this stage is known as test of significance.

### **Decision Errors**

By performing a test we make a decision on the hypothesis by accepting or rejecting the null hypothesis  $H_0$ . In the process we may make a correct decision on  $H_0$  or commit one of two kinds of error.

- We may reject  $H_0$  based on sample data when in fact it is true. This error in decisions is known as Type I error.

- We may accept  $H_0$  based on sample data when in fact it is not true. It is known as Type II error.

	<b>Accept <math>H_0</math></b>	<b>Reject <math>H_0</math></b>
Ho is true	Correct Decision	Type I error
Ho is false	Type II error	Correct Decision

The relationship between type I & type II errors is that if one increases the other will decrease.

The probability of type I error is denoted by  $\alpha$ . The probability of type II error is denoted by  $\beta$ .

The correct decision of rejecting the null hypothesis when it is false is known as the power of the test. The probability of the power is given by  $1-\beta$ .

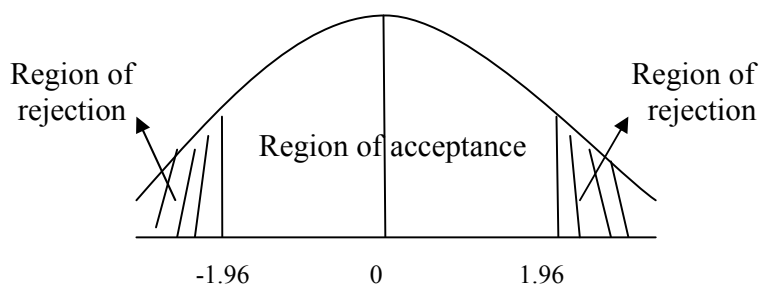
### Critical Region

The testing of statistical hypothesis involves the choice of a region on the sampling distribution of statistic. If the statistic falls within this region, the null hypothesis is rejected; otherwise it is accepted. This region is called critical region.

Let the null hypothesis be  $H_0: \mu_1 = \mu_2$  and its alternative be  $H_1: \mu_1 \neq \mu_2$ . Suppose  $H_0$  is true. Based on sample data it may be observed that statistic  $(\bar{x}_1 - \bar{x}_2)$  follows a normal distribution given by

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SE(\bar{x}_1 - \bar{x}_2)}$$

We know that 95% values of the statistic from repeated samples will fall in the range  $(\bar{x}_1 - \bar{x}_2) \pm 1.96$  times  $SE(\bar{x}_1 - \bar{x}_2)$ . This is represented by a diagram.



The border line value  $\pm 1.96$  is the critical value or tabular value of Z. The area beyond the critical values (shaded area) is known as critical region or region of rejection. The remaining area is known as region of acceptance.

If the statistic falls in the critical region we reject the null hypothesis and, if it falls in the region of acceptance we accept the null hypothesis.

In other words if the calculated value of a test statistic (Z, t,  $\chi^2$  etc) is more than the critical value in magnitude it is said to be significant and we reject  $H_0$  and otherwise we accept  $H_0$ . The critical values for the t and  $\chi^2$  are given in the form of readymade tables. Since the critical values are given in the form of table it is commonly referred as table value. The table value depends on the level of significance and degrees of freedom.

Example:  $Z_{cal} < Z_{tab}$  - We accept the  $H_0$  and conclude that there is no significant difference between the means

### **Test Statistic**

The sampling distribution of a statistic like Z, t, &  $\chi^2$  are known as test statistic. Generally, in case of quantitative data

$$\text{Test statistic} = \frac{\text{Statistic} - \text{Parameter}}{\text{Standard Error}(\text{Statistic})}$$

### **Note**

The choice of the test statistic depends on the nature of the variable (ie) qualitative or quantitative, the statistic involved (i.e) mean or variance and the sample size, (i.e) large or small.

### **Level of Significance**

The probability that the statistic will fall in the critical region is  $\frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$ . This  $\alpha$  is nothing but the probability of committing type I error. Technically the probability of committing type I error is known as level of Significance.

### **One and two tailed test**

The nature of the alternative hypothesis determines the position of the critical region. For example, if  $H_1$  is  $\mu_1 \neq \mu_2$  it does not show the direction and hence the critical region falls on either end of the sampling distribution. If  $H_1$  is  $\mu_1 < \mu_2$  or  $\mu_1 > \mu_2$  the direction is known. In the

first case the critical region falls on the left of the distribution whereas in the second case it falls on the right side.

**One tailed test** – When the critical region falls on one end of the sampling distribution, it is called one tailed test.

**Two tailed test** – When the critical region falls on either end of the sampling distribution, it is called two tailed test.

For example, consider the mean yield of new paddy variety ( $\mu_1$ ) is compared with that of a ruling variety ( $\mu_2$ ). Unless the new variety is more promising than the ruling variety in terms of yield we are not going to accept the new variety. In this case  $H_1 : \mu_1 > \mu_2$  for which one tailed test is used. If both the varieties are new our interest will be to choose the best of the two. In this case  $H_1 : \mu_1 \neq \mu_2$  for which we use two tailed test.

### **Degrees of freedom**

The number of degrees of freedom is the number of observations that are free to vary after certain restrictions have been placed on the data. If there are  $n$  observations in the sample, for each restriction imposed upon the original observation the number of degrees of freedom is reduced by one.

The number of independent variables which make up the statistic is known as the degrees of freedom and is denoted by  $\gamma$  (Nu)

### **Steps in testing of hypothesis**

The process of testing a hypothesis involves following steps.

1. Formulation of null & alternative hypothesis.
2. Specification of level of significance.
3. Selection of test statistic and its computation.
4. Finding out the critical value from tables using the level of significance, sampling distribution and its degrees of freedom.
5. Determination of the significance of the test statistic.
6. Decision about the null hypothesis based on the significance of the test statistic.
7. Writing the conclusion in such a way that it answers the question on hand.



### **Large sample theory**

The sample size  $n$  is greater than 30 ( $n \geq 30$ ) it is known as large sample. For large samples the sampling distributions of statistic are normal (Z test). A study of sampling distribution of statistic for large sample is known as large sample theory.

### **Small sample theory**

If the sample size  $n$  is less than 30 ( $n < 30$ ), it is known as small sample. For small samples the sampling distributions are  $t$ ,  $F$  and  $\chi^2$  distribution. A study of sampling distributions for small samples is known as small sample theory.

### **Test of Significance**

The theory of test of significance consists of various test statistic. The theory had been developed under two broad heading

1. Test of significance for large sample  
Large sample test or Asymptotic test or Z test ( $n \geq 30$ )
2. Test of significance for small samples ( $n < 30$ )  
Small sample test or Exact test- $t$ ,  $F$  and  $\chi^2$ .

It may be noted that small sample tests can be used in case of large samples also.

### **Large sample test**

Large sample test are

1. Sampling from attributes
2. Sampling from variables

### **Sampling from attributes**

There are two types of test for attributes

1. Test for single proportion
2. Test for equality of two proportions

### **Test for single proportion**

In a sample of large size  $n$ , we may examine whether the sample would have come from a population having a specified proportion  $P = P_0$ . For testing

We may proceed as follows

**1. Null Hypothesis (H<sub>0</sub>)**

H<sub>0</sub>: The given sample would have come from a population with specified proportion P=P<sub>0</sub>

**2. Alternative Hypothesis(H<sub>1</sub>)**

H<sub>1</sub> : The given sample may not be from a population with specified proportion

P≠P<sub>0</sub> (Two Sided)

P>P<sub>0</sub>(One sided-right sided)

P<P<sub>0</sub>(One sided-left sided)

**3. Test statistic**

$$Z = \frac{|p - P|}{\sqrt{\frac{PQ}{n}}}$$

It follows a standard normal distribution with μ=0 and σ<sup>2</sup>=1

**4. Level of Significance**

The level of significance may be fixed at either 5% or 1%

**5. Expected value or critical value**

In case of test statistic Z, the expected value is

$$Z_e = \left. \begin{array}{l} 1.96 \text{ at } 5\% \text{ level} \\ 2.58 \text{ at } 1\% \text{ level} \end{array} \right\} \longrightarrow \text{Two tailed test}$$

$$Z_e = \left. \begin{array}{l} 1.65 \text{ at } 5\% \text{ level} \\ 2.33 \text{ at } 1\% \text{ level} \end{array} \right\} \longrightarrow \text{One tailed test}$$

**6. Inference**

If the observed value of the test statistic Z<sub>0</sub> exceeds the table value Z<sub>e</sub> we reject the Null Hypothesis H<sub>0</sub> otherwise accept it.

**Test for equality of two proportions**

Given two sets of sample data of large size n<sub>1</sub> and n<sub>2</sub> from attributes. We may examine whether the two samples come from the populations having the same proportion. We may proceed as follows:

### 1. Null Hypothesis (H<sub>0</sub>)

H<sub>0</sub>: The given two sample would have come from a population having the same proportion

$$P_1 = P_2$$

### 2. Alternative Hypothesis (H<sub>1</sub>)

H<sub>1</sub>: The given two sample may not be from a population with specified proportion

$$P_1 \neq P_2 \text{ (Two Sided)}$$

$$P_1 > P_2 \text{ (One sided-right sided)}$$

$$P_1 < P_2 \text{ (One sided-left sided)}$$

### 3. Test statistic

$$Z = \frac{|(p_1 - p_2) - (P_1 - P_2)|}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

When P<sub>1</sub> and P<sub>2</sub> are not known, then

$$Z = \frac{|p_1 - p_2|}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \quad \text{for heterogeneous population}$$

Where q<sub>1</sub> = 1 - p<sub>1</sub> and q<sub>2</sub> = 1 - p<sub>2</sub>

$$Z = \frac{|p_1 - p_2|}{\sqrt{pq \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{for homogeneous population}$$

p = combined or pooled estimate.

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

### 4. Level of Significance

The level may be fixed at either 5% or 1%

### 5. Expected value

The expected value is given by

$$Z_e = \quad 1.96 \text{ at 5\% level } \quad ]$$

	2.58 at 1% level	→	Two tailed test
$Z_e =$	1.65 at 5% level	}	→ One tailed test
	2.33 at 1% level		

## 6. Inference

If the observed value of the test statistic  $Z$  exceeds the table value  $Z_e$  we may reject the Null Hypothesis  $H_0$  otherwise accept it.

### Sampling from variable

In sampling for variables, the test are as follows

1. Test for single Mean
2. Test for single Standard Deviation
3. Test for equality of two Means
4. Test for equality of two Standard Deviation

### Test for single Mean

In a sample of large size  $n$ , we examine whether the sample would have come from a population having a specified mean

#### 1. Null Hypothesis ( $H_0$ )

$H_0$ : There is no significance difference between the sample mean ie.,  $\mu = \mu_0$

or

The given sample would have come from a population having a specified mean

ie.,  $\mu = \mu_0$

#### 2. Alternative Hypothesis ( $H_1$ )

$H_1$  : There is significance difference between the sample mean

ie.,  $\mu \neq \mu_0$  or  $\mu > \mu_0$  or  $\mu < \mu_0$

### 3. Test statistic

$$Z = \frac{|\bar{x} - \mu|}{\frac{\sigma}{\sqrt{n}}}$$

When population variance is not known, it may be replaced by its estimate

$$Z = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}}$$

$$\text{where } s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

### 4. Level of Significance

The level may be fixed at either 5% or 1%

### 5. Expected value

The expected value is given by

$$Z_e = \begin{array}{l} 1.96 \text{ at } 5\% \text{ level} \\ 2.58 \text{ at } 1\% \text{ level} \end{array} \quad \left. \vphantom{\begin{array}{l} 1.96 \\ 2.58 \end{array}} \right\} \longrightarrow \text{Two tailed test}$$

$$Z_e = \begin{array}{l} 1.65 \text{ at } 5\% \text{ level} \\ 2.33 \text{ at } 1\% \text{ level} \end{array} \quad \left. \vphantom{\begin{array}{l} 1.65 \\ 2.33 \end{array}} \right\} \longrightarrow \text{One tailed test}$$

### 6. Inference

If the observed value of the test statistic  $Z$  exceeds the table value  $Z_e$  we may reject the Null Hypothesis  $H_0$  otherwise accept it.

## Test for equality of two Means

Given two sets of sample data of large size  $n_1$  and  $n_2$  from variables. We may examine whether the two samples come from the populations having the same mean. We may proceed as follows

### 1. Null Hypothesis ( $H_0$ )

$H_0$ : There is no significance difference between the sample mean i.e.,  $\mu = \mu_0$

or

The given sample would have come from a population having a specified mean

i.e.,  $\mu_1 = \mu_2$

### 2. Alternative Hypothesis ( $H_1$ )

$H_1$ : There is significance difference between the sample mean i.e.,  $\mu \neq \mu_0$

i.e.,  $\mu_1 \neq \mu_2$  or  $\mu_1 < \mu_2$  or  $\mu_1 > \mu_2$

### 3. Test statistic

When the population variances are known and unequal (i.e)  $\sigma_1^2 \neq \sigma_2^2$

$$Z = \frac{|\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

When  $\sigma_1^2 = \sigma_2^2$ ,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where  $\sigma = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2}$

The equality of variances can be tested by using F test.

When population variance is unknown, they may be replaced by their estimates  $s_1^2$  and  $s_2^2$

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{when } s_1^2 \neq s_2^2$$

when  $s_1^2 = s_2^2$

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{where } s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

#### 4. Level of Significance

The level may be fixed at either 5% or 1%

#### 5. Expected value

The expected value is given by

$Z_e =$	1.96 at 5% level	}	→ Two tailed test
	2.58 at 1% level		
$Z_e =$	1.65 at 5% level	}	→ One tailed test
	2.33 at 1% level		

#### 6. Inference

If the observed value of the test statistic  $Z$  exceeds the table value  $Z_e$  we may reject the Null Hypothesis  $H_0$  otherwise accept it.

### Questions

1. A hypothesis may be classified as

- |            |                   |
|------------|-------------------|
| (a) Simple | (b) Composite     |
| (c) Null   | (d) All the above |

**Ans: All the above**

2. Area of the critical region depends on

- |                             |                            |
|-----------------------------|----------------------------|
| (a) Size of type I error    | (b) Size of type II error  |
| (c) Value of the statistics | (d) Number of observations |

**Ans: Size of type I error**

3. Large sample test can be applied when the sample size exceeds 30.

**Ans: True**

4. If the calculated test statistic is greater than the critical value, the null hypothesis is rejected.

**Ans: True**

5. The standard error of mean is given by  $\frac{\sigma}{\sqrt{n}}$

**Ans: True**

6. If the alternative hypothesis is  $\mu_1 \neq \mu_2$  then the test is known as one tailed test.

**Ans: False**

7. Define standard error.

8. Define Type I and Type II error.

9. Describe the procedure of comparing two group means.

10. Describe the procedure of comparing two proportions.



## Lecture.10

### T-test – definition – assumptions – test for equality of two means-independent and paired t test

#### Student's t test

When the sample size is smaller, the ratio  $Z = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}}$  will follow t distribution

and not the standard normal distribution. Hence the test statistic is given as  $t = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}}$

which follows normal distribution with mean 0 and unit standard deviation. This follows a t distribution with (n-1) degrees of freedom which can be written as  $t_{(n-1)}$  d.f.

This fact was brought out by Sir William Gosset and Prof. R.A Fisher. Sir William Gosset published his discovery in 1905 under the pen name Student and later on developed and extended by Prof. R.A Fisher. He gave a test known as t-test.

#### Applications (or) uses

1. To test the single mean in single sample case.
2. To test the equality of two means in double sample case.
  - (i) Independent samples(Independent t test)
  - (ii) Dependent samples (Paired t test)
3. To test the significance of observed correlation coefficient.
4. To test the significance of observed partial correlation coefficient.
5. To test the significance of observed regression coefficient.

#### Test for single Mean

1. Form the null hypothesis

$$H_0: \mu = \mu_0$$

(i.e) There is no significance difference between the sample mean and the population mean

2. Form the Alternate hypothesis

$$H_1: \mu \neq \mu_0 \text{ (or } \mu > \mu_0 \text{ or } \mu < \mu_0)$$

ie., There is significance difference between the sample mean and the population mean

3. Level of Significance

The level may be fixed at either 5% or 1%

4. Test statistic

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \text{ which follows t distribution with (n-1) degrees of freedom}$$

$$\text{where } \bar{x} = \frac{\sum x_i}{n} \text{ and } s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

6. Find the table value of t corresponding to (n-1) d.f. and the specified level of significance.

7. Inference

If  $t < t_{tab}$  we accept the null hypothesis  $H_0$ . We conclude that there is no significant difference sample mean and population mean

(or) if  $t > t_{tab}$  we reject the null hypothesis  $H_0$ . (ie) we accept the alternative hypothesis and conclude that there is significant difference between the sample mean and the population mean.

**Example 1**

Based on field experiments, a new variety of green gram is expected to give a yield of 12.0 quintals per hectare. The variety was tested on 10 randomly selected farmer's fields. The yield (quintals/hectare) were recorded as 14.3, 12.6, 13.7, 10.9, 13.7, 12.0, 11.4, 12.0, 12.6, 13.1. Do the results conform to the expectation?

**Solution**

Null hypothesis  $H_0: \mu=12.0$

(i.e) the average yield of the new variety of green gram is 12.0 quintals/hectare.

Alternative Hypothesis:  $H_1: \mu \neq 12.0$

(i.e) the average yield is not 12.0 quintals/hectare, it may be less or more than 12 quintals / hectare

Level of significance: 5 %

Test statistic:

$$t = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right|$$

From the given data

$$\sum x = 126.3 \quad \sum x^2 = 1605.77$$

$$\bar{x} = \frac{\sum x}{n} = \frac{126.3}{10} = 12.63$$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{1605.77 - \frac{1595.169}{9}}{9}} = \sqrt{\frac{10.601}{9}}$$
$$= 1.0853$$

$$\frac{s}{\sqrt{n}} = \frac{1.0853}{\sqrt{10}} = 0.3432$$

$$\text{Now } t = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right|$$

$$t = \frac{12.63 - 12}{0.3432} = 1.836$$

Table value for t corresponding to 5% level of significance and 9 d.f. is 2.262 (two tailed test)

## Inference

$$t < t_{\text{tab}}$$

We accept the null hypothesis  $H_0$

We conclude that the new variety of green gram will give an average yield of 12 quintals/hectare.

## Note

Before applying t test in case of two samples the equality of their variances has to be tested by using F-test

$$F = \frac{s_1^2}{s_2^2} \sim F_{(n_1-1, n_2-1)} \text{ d.f. if } s_1^2 > s_2^2$$

or

$$F = \frac{s_2^2}{s_1^2} \sim F_{(n_2-1, n_1-1)} \text{ d.f. if } s_2^2 > s_1^2$$

where  $s_1^2$  is the variance of the first sample whose size is  $n_1$ .

$s_2^2$  is the variance of the second sample whose size is  $n_2$ .

It may be noted that the numerator is always the greater variance. The critical value for F is read from the F table corresponding to a specified d.f. and level of significance

## Inference

$$F < F_{\text{tab}}$$

We accept the null hypothesis  $H_0$ . (i.e) the variances are equal otherwise the variances are unequal.

## Test for equality of two Means (Independent Samples)

Given two sets of sample observation  $x_{11}, x_{12}, x_{13} \dots x_{1n}$ , and  $x_{21}, x_{22}, x_{23} \dots x_{2n}$  of sizes  $n_1$  and  $n_2$  respectively from the normal population.

1. Using F-Test, test their variances

### (i) Variances are Equal

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2 \text{ (or } \mu_1 < \mu_2 \text{ or } \mu_1 > \mu_2)$$

Test statistic

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where the combined variance

$$s^2 = \frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right]}{n_1 + n_2 - 2}$$

The test statistic t follows a t distribution with (n1+n2-2) d.f.

**(ii) Variances are unequal and n1=n2**

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

It follows a t distribution with  $\left( \frac{n_1 + n_2}{2} \right) - 1$  d.f.

**(i) Variances are unequal and n1≠n2**

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

This statistic follows neither t nor normal distribution but it follows Behrens-Fisher distribution. The Behrens – Fisher test is laborious one. An alternative simple method has been suggested by Cochran & Cox. In this method the critical value of t is altered as  $t_w$  (i.e) weighted t

$$t_w = \frac{t_1 \left( \frac{s_1^2}{n_1} \right) + t_2 \left( \frac{s_2^2}{n_2} \right)}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where  $t_1$  is the critical value for t with (n1-1) d.f. at a dspecified level of significance and  $t_2$  is the critical value for t with (n2-1) d.f. at a dspecified level of significance and

### Example 2

In a fertilizer trial the grain yield of paddy (Kg/plot) was observed as follows

Under ammonium chloride 42,39,38,60 &41 kgs

Under urea 38, 42, 56, 64, 68, 69,& 62 kgs.

Find whether there is any difference between the sources of nitrogen?

### Solution

$H_0: \mu_1 = \mu_2$  (i.e) there is no significant difference in effect between the sources of nitrogen.

$H_1: \mu_1 \neq \mu_2$  (i.e) there is a significant difference between the two sources

Level of significance = 5%

Before we go to test the means first we have to test their variances by using F-test.

F-test

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$s_1^2 = \frac{\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}}{n_1 - 1} = 82.5$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}}{n_2 - 1} = 154.33$$

$$\therefore F = \frac{s_2^2}{s_1^2} \sim F_{(n_2 - 1, n_1 - 1)} \text{ d.f. if } s_2^2 > s_1^2$$

$$F = \frac{154.33}{82.5} = 1.8707$$

$$F_{\text{tab}}(6,4) \text{ d.f.} = 6.16$$

$$\Rightarrow F < F_{\text{tab}}$$

We accept the null hypothesis  $H_0$ . (i.e) the variances are equal.

Use the test statistic

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where

$$s^2 = \frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} \right]}{n_1 + n_2 - 2} = \frac{330 + 926}{10} = 125.6$$

$$t = \frac{|44 - 57|}{\sqrt{125.6 \left( \frac{1}{7} + \frac{1}{75} \right)}} = 1.98$$

The degrees of freedom is  $5+7-2= 10$ . For 5 % level of significance, table value of t is 2.228

Inference:

$$t < t_{\text{tab}}$$

We accept the null hypothesis  $H_0$

We conclude that the two sources of nitrogen do not differ significantly with regard to the grain yield of paddy.

### Example 3

The summary of the results of an yield trial on onion with two methods of propagation is given below. Determine whether the methods differ with regard to onion yield. The onion yield is given in Kg/plot.

Method I	Method II
$n_1=12$	$n_2=12$
$\bar{x}_1 = 25.25$	$\bar{x}_2 = 28.83$
$SS_1=186.25$	$SS_2=737.6667$
$s_1^2 = 16.9318$	$s_2^2 = 67.0606$

### Solution

$H_0: \mu_1 = \mu_2$  (i.e) the two propagation methods do not differ with regard to onion yield.

$H_1: \mu_1 \neq \mu_2$  (i.e) the two propagation methods differ with regard to onion yield.

Level of significance = 5%

Before we go to test the means first we have to test their variability using F-test.

F-test

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

$$s_1^2 = \frac{\sum x_1^2 - \frac{(\sum x_1)^2}{n1}}{n1 - 1} = 16.9318$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{(\sum x_2)^2}{n2}}{n2 - 1} = 67.0606$$

$$\therefore F = \frac{s_2^2}{s_1^2} \sim F_{(n_2 - 1, n_1 - 1)} \text{ d.f. if } s_2^2 > s_1^2$$

$$F = \frac{67.0606}{16.9318} = 3.961$$

$$F_{\text{tab}}(11, 11) \text{ d.f.} = 2.82$$

$$\Rightarrow F > F_{\text{tab}}$$

We reject the null hypothesis  $H_0$ . we conclude that the variances are unequal.

Here the variances are unequal with equal sample size then the test statistic is

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where

$$s^2 = \frac{\left[ \sum x_1^2 - \frac{(\sum x_1)^2}{n1} \right] + \left[ \sum x_2^2 - \frac{(\sum x_2)^2}{n2} \right]}{n_1 + n_2 - 2}$$

$$s^2 = \frac{SS1 + SS2}{n_1 + n_2 - 2} = \frac{186.25 + 737.6667}{12 + 12 - 2} = 41.9962$$

$$t = \frac{25.25 - 28.83}{\sqrt{41.9962 \left( \frac{1}{12} + \frac{1}{12} \right)}} = \frac{3.58}{\sqrt{6.9994}} = 1.353$$

$$t = 1.353$$

The table value for  $\left( \frac{12+12}{2} - 1 \right) = 11$  d.f. at 5% level of significance is 2.201

Inference:

$$t < t_{\text{tab}}$$



We accept the null hypothesis  $H_0$

We conclude that the two propagation methods do not differ with regard to onion yield.

**Example 4**

The following data relate the rubber yield of two types of rubber plants, where the sample have been drawn independently. Test whether the two types of rubber plants differ in their yield.

Type I	6.21	5.70	6.04	4.47	5.22	4.45	4.84	5.84	5.88	5.82	6.09	5.59
	6.06	5.59	6.74	5.55								
Type II	4.28	7.71	6.48	7.71	7.37	7.20	7.06	6.40	8.93	5.91	5.51	6.36

**Solution**

$H_0$ :.,  $\mu_1 = \mu_2$  (i.e) there is no significant difference between the two rubber plants.

$H_1$   $\mu_1 \neq \mu_2$  (i.e) there is a significant difference between the two rubber plants.

Level of significance = 5%

Here

$n_1 = 16$	$n_2 = 12$
$\sum x_1 = 90.09$	$\sum x_2 = 80.92$
$\bar{x}_1 = 5.63$	$\bar{x}_2 = 6.7431$
$\sum x_1^2 = 513.085$	$\sum x_2^2 = 561.64$

Before we go to test the means first we have to test their variability using F-test.

F-test

$H_0$ :.,  $\sigma_1^2 = \sigma_2^2$

$H_1$ :.,  $\sigma_1^2 \neq \sigma_2^2$

$$s_1^2 = \frac{\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}}{n_1 - 1} = 0.388$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}}{n_2 - 1} = 1.452$$

$$\therefore F = \frac{s_2^2}{s_1^2} \sim F_{(n_2-1, n_1-1)} \text{ d.f. if } s_2^2 > s_1^2$$

$$F = \frac{1.452}{0.388} = 3.742$$

$$F_{\text{tab}}(11, 15) \text{ d.f.} = 2.51$$

$$\Rightarrow F > F_{\text{tab}}$$

We reject the null hypothesis  $H_0$ . Hence, the variances are unequal.

Here the variances are unequal with unequal sample size then the test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{(5.63 - 6.7431_2)}{\sqrt{\frac{0.388}{16} + \frac{1.452}{12}}} = 2.912$$

$$t_w = \frac{t_1 \left( \frac{S_1^2}{n1} \right) + t_2 \left( \frac{S_2^2}{n2} \right)}{\frac{S_1^2}{n1} + \frac{S_2^2}{n2}}$$

$$t_1 = t_{(16-1)} \text{ d.f.} = 2.131$$

$$t_2 = t_{(12-1)} \text{ d.f.} = 2.201$$

$$t_w = \frac{2.131 \left( \frac{0.388}{16} \right) + 2.201 \left( \frac{1.452}{12} \right)}{\frac{0.388}{16} + \frac{1.425}{12}} = 2.187$$

Inference:

$$t > t_w$$

We reject the null hypothesis  $H_0$ . We conclude that the second type of rubber plant yields more rubber than that of first type.

## Equality of two means (Dependant samples)

### Paired t test

In the t-test for difference between two means, the two samples were independent of each other. Let us now take particular situations where the samples are not independent.

In agricultural experiments it may not be possible to get required number of homogeneous experimental units. For example, required number of plots which are similar in all characteristics may not be available. In such cases each plot may be divided into two equal parts and one treatment is applied to one part and second treatment to another part of the plot. The results of the experiment will result in two correlated samples. In some other situations two observations may be taken on the same experimental unit. For example, the soil properties before and after the application of industrial effluents may be observed on number of plots. This will result in paired observation. In such situations we apply paired t test.

Suppose the observation before treatment is denoted by  $x$  and the observation after treatment is denoted by  $y$ . for each experimental unit we get a pair of observation  $(x,y)$ . In case of  $n$  experimental units we get  $n$  pairs of observations :  $(x_1,y_1), (x_2,y_2), \dots, (x_n,y_n)$ . In order to apply the paired t test we find out the differences  $(x_1 - y_1), (x_2 - y_2), \dots, (x_n - y_n)$  and denote them as  $d_1, d_2, \dots, d_n$ . Now  $d_1, d_2, \dots$  form a sample . we

apply the t test procedure for one sample (i.e)  $t = \frac{|\bar{d}|}{\sqrt{s^2/n}}$

$$\text{where } \bar{d} = \frac{\sum di}{n}, s^2 = \frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1}$$

the mean  $\bar{d}$  may be positive or negative. Hence we take the absolute value as  $|\bar{d}|$ . The

test statistic  $t$  follows a  $t$  distribution with  $(n-1)$  d.f.

### Example 5

In an experiment the plots were divided into two equal parts. One part received soil treatment A and the second part received soil treatment B. each plot was planted with sorghum. The sorghum yield (kg/plot) was recorded. The results are given below. Test the effectiveness of soil treatments on sorghum yield.

Soil treatment A	49	53	51	52	47	50	52	53
Soil treatment B	52	55	52	53	50	54	54	53

### Solution

H<sub>0</sub>:  $\mu_1 = \mu_2$ , there is no significant difference between the effects of the two soil treatments

H<sub>1</sub>:  $\mu_1 \neq \mu_2$ , there is significant difference between the effects of the two soil treatments

Level of significance = 5%

### Test statistic

$$t = \frac{|\bar{d}|}{\sqrt{s^2/n}}$$

x	y	d=x-y	d <sup>2</sup>
49	52	-3	9
53	55	-2	4
51	52	-1	1
51	52	-1	1
47	50	-3	16
50	54	-4	16
52	54	-2	4
53	53	0	0
Total		-16	44

$$\bar{d} = \frac{\sum di}{n} = \frac{-16}{8} = -2,$$

$$s^2 = \frac{\sum di^2 - \frac{(\sum di)^2}{n}}{n-1} = 1.7143$$

$$t = \frac{|-2|}{\sqrt{1.7143/8}} = 4.32$$

Table value of t for 7 d.f. at 5% l.o.s is 2.365

Inference:

$$t > t_{\text{tab}}$$

We reject the null hypothesis  $H_0$ . We conclude that there is a significant difference between the two soil treatments between A and B. Soil treatment B increases the yield of sorghum significantly,

### Questions

1. The test statistic  $F = \frac{s_1^2}{s_2^2}$  is used for testing

(a)  $H_0: \mu_1 = \mu_2$

(b)  $H_0: \sigma_1^2 = \sigma_2^2$

(c)  $H_0: \sigma_1 = \sigma_2$

(d)  $H_0: \sigma_2 = \sigma_0$

**Ans:  $H_0: \sigma_1^2 = \sigma_2^2$**

2. In paired t test with n observations in each group the degrees of freedom is

(a) n

(b) n-1

(c) n-2

(d) n+1

**Ans: n-1**

3. Student t- test is applicable in case of small samples.

**Ans: True**

4. F test is also known as variance ratio test.

**Ans: True**

5. In case of comparing the equality of two variances the greater variance should be taken in the numerator.

**Ans: True**

6. While comparing the means of two independent samples the variances of the two samples will be always equal.

**Ans: False**

7. Define t statistic.

8. Define F statistic.

9. Explain the procedure of testing the equality of two variances.

10. How to compare the means of two independent small samples.

## Lecture.11

### Attributes- Contingency table – 2x2 contingency table – Test for independence of attributes – test for goodness of fit of mendalian ratio

#### Test based on $\chi^2$ -distribution

In case of attributes we can not employ the parametric tests such as F and t. Instead we have to apply  $\chi^2$  test. When we want to test whether a set of observed values are in agreement with those expected on the basis of some theories or hypothesis. The  $\chi^2$  statistic provides a measure of agreement between such observed and expected frequencies.

The  $\chi^2$  test has a number of applications. It is used to

- (1) Test the independence of attributes
- (2) Test the goodness of fit
- (3) Test the homogeneity of variances
- (4) Test the homogeneity of correlation coefficients
- (5) Test the equalilty of several proportions.

In genetics it is applied to detect linkage.

#### Applications

##### $\chi^2$ – test for goodness of fit

A very powerful test for testing the significance of the discrepancy between theory and experiment was given by Prof. Karl Pearson in 1900 and is known as “chi-square test of goodness of fit “.

If  $O_i$ , ( $i=1,2,\dots,n$ ) is a set of observed (experimental frequencies) and  $E_i$  ( $i=1,2,\dots,n$ ) is the corresponding set of expected (theoretical or hypothetical) frequencies, then,

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

It follows a  $\chi^2$  distribution with n-1 d.f. In case of  $\chi^2$  only one tailed test is used.

### Example

In plant genetics, our interest may be to test whether the observed segregation ratios deviate significantly from the mendelian ratios. In such situations we want to test the agreement between the observed and theoretical frequency, such test is called as test of goodness of fit.

### Conditions for the validity of $\chi^2$ -test:

$\chi^2$  -test is an approximate test for large values of 'n' for the validity of  $\chi^2$  -test of goodness of fit between theory and experiment, the following conditions must be satisfied.

1. The sample observations should be independent.
2. Constraints on the cell frequency, if any, should be linear.  
Example:  $\sum O_i = \sum E_i$ .
3. N, the total frequency should be reasonably large, say greater than (>) 50.
4. No theoretical cell frequency should be less than (<)5. If any theoretical cell frequency is <5, then for the application of  $\chi^2$  - test, it is pooled with the preceding or succeeding frequency so that the pooled frequency is more than 5 and finally adjust for degree's of freedom lost in pooling.

### Example1

The number of yeast cells counted in a haemocytometer is compared to the theoretical value is given below. Does the experimental result support the theory?

No. of Yeast cells in the square	Observed Frequency	Expected Frequency
0	103	106
1	143	141
2	98	93
3	42	41
4	8	14
5	6	5

**Solution**

H<sub>0</sub>: the experimental results support the theory

H<sub>1</sub>: the experimental results does not support the theory.

Level of significance=5%

Test Statistic:

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

O <sub>i</sub>	E <sub>i</sub>	O <sub>i</sub> .E <sub>i</sub>	(O <sub>i</sub> .E <sub>i</sub> ) <sup>2</sup>	(O <sub>i</sub> .E <sub>i</sub> ) <sup>2</sup> /E <sub>i</sub>
103	106	-3	9	0.0849
143	141	2	4	0.0284
98	93	5	25	0.2688
42	41	1	1	0.0244
8	14	-6	36	2.5714
6	5	1	1	0.2000
400	400			<b>3.1779</b>

∴  $\chi^2 = 3.1779$

**Table value**

$\chi^2 (6-1=5 \text{ at } 5 \% \text{ l.os}) = 11.070$

Inference

$\chi^2 < \chi^2_{\text{tab}}$

We accept the null hypothesis.

(i.e) there is a good correspondence between theory and experiment.

**$\chi^2$  test for independence of attributes**

At times we may consider two characteristics on attributes simultaneously. Our interest will be to test the association between these two attributes

**Example:-** An entomologist may be interested to know the effectiveness of different concentrations of the chemical in killing the insects. The concentrations of chemical form one attribute. The state of insects ‘killed & not killed’ forms another attribute. The result of this experiment can be arranged in the form of a contingency table. In general one attribute may be divided into m classes as A<sub>1</sub>, A<sub>2</sub>, .....A<sub>m</sub> and the other attribute may be divided into n classes as B<sub>1</sub>, B<sub>2</sub>, .....B<sub>n</sub>. Then the contingency table will have m x n cells. It is termed as m x n contingency table

<del>A</del>	A1	A2	...	Aj	...	Am	Row Total
<del>B</del>							



B1	O11	O12	...	O1j		O1m	r1
B2	O21	O22	...	O2j		O2m	r2
⋮							
⋮							
Bi	Oij	Oi2	...	Oij		Oim	ri
⋮							
⋮							
Bn	On1	On2	...	Onj		Onm	rk
Column Total	c1	c2	...	cj	...	cm	$n = \sum ri = \sum cj$

where  $O_{ij}$ 's are observed frequencies.

The expected frequencies corresponding to  $O_{ij}$  is calculated as  $\frac{ri \cdot cj}{n}$ . The  $\chi^2$  is

computed as

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(o_{ij} - E_{ij})^2}{E_{ij}}$$

where

$O_{ij}$  – observed frequencies

$E_{ij}$  – Expected frequencies

$n$  = number of rows

$m$  = number of columns

It can be verified that  $\sum O_{ij} = \sum E_{ij}$

This  $\chi^2$  is distributed as  $\chi^2$  with  $(n-1)(m-1)$  d.f.

### 2x2 – contingency table

When the number of rows and number of columns are equal to 2 it is termed as 2 x 2 contingency table. It will be in the following form

	B <sub>1</sub>	B <sub>2</sub>	Row Total
A <sub>1</sub>	a	b	a+b    r <sub>1</sub>
A <sub>2</sub>	c	d	c+d    r <sub>2</sub>
Column Total	a+c	b+d	a+b+c+d =n
	c <sub>1</sub>	c <sub>2</sub>	

Where a, b, c and d are cell frequencies c1 and c2 are column totals, r1 and r2 are row totals and n is the total number of observations.

In case of 2 x 2 contingency table  $\chi^2$  can be directly found using the short cut formula,

$$\chi^2 = \frac{n(ad - bc)^2}{c1.c2.r1.r2}$$

The d.f associated with  $\chi^2$  is (2-1) (2-1) =1

### Yates correction for continuity

If anyone of the cell frequency is < 5, we use Yates correction to make  $\chi^2$  as continuous. The yares correction is made by adding 0.5 to the least cell frequency and adjusting the other cell frequencies so that the column and row totals remain same . suppose, the firat cell frequency is to be corrected then the consigency table will be as follows:

	B1	B2	Row Total
A1	a + 0.5	b - 0.5	a+b=r1
A2	c - 0.5	d + 0.5	c+d =r2
Column Total	a+c=c1	b+d=c2	n = a+b+c+d

Then use the  $\chi^2$  - statistic as

$$\chi^2 = \frac{n \left( \left| ad - bc \right| - \frac{n}{2} \right)^2}{c1.c2.r1.r2}$$

The d.f associated with  $\chi^2$  is (2-1) (2-1) =1

### Exapmle 2

The severity of a disease and blood group were studied in a research project. The findings sre given in the following table, knowmn as the m xn contingency table. Can this severity of the condition and blood group are associated.

Severity of a disease classified by blood group in 1500 patients.

Condition	Blood Groups				Total
	O	A	B	AB	

Severe	51	40	10	9	110
Moderate	105	103	25	17	250
Mild	384	527	125	104	1140
Total	540	670	160	130	1500

**Solution**

H<sub>0</sub>: The severity of the disease is not associated with blood group.

H<sub>1</sub>: The severity of the disease is associated with blood group.

Calculation of Expected frequencies

Condition	Blood Groups				Total
	O	A	B	AB	
Severe	39.6	49.1	11.7	9.5	110
Moderate	90.0	111.7	26.7	21.7	250
Mild	410.4	509.2	121.6	98.8	1140
Total	540	670	160	130	1500

Test statistic:

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^m \frac{(o_{ij} - E_{ij})^2}{E_{ij}}$$

The d.f. associated with the  $\chi^2$  is (3-1)(4-1) = 6

**Calculations**

O <sub>i</sub>	E <sub>i</sub>	O <sub>i</sub> ·E <sub>i</sub>	(O <sub>i</sub> ·E <sub>i</sub> ) <sup>2</sup>	(O <sub>i</sub> ·E <sub>i</sub> ) <sup>2</sup> /E <sub>i</sub>
51	39.6	11.4	129.96	3.2818
40	49.1	-9.1	82.81	1.6866
10	11.7	-1.7	2.89	0.2470
9	9.5	-0.5	0.25	0.0263
105	90.0	15	225.00	2.5000
103	111.7	-8.7	75.69	0.6776
25	26.7	-1.7	2.89	0.1082
17	21.7	-4.7	22.09	1.0180
384	410.4	-26.4	696.96	1.6982
527	509.2	17.8	316.84	0.6222
125	121.6	3.4	11.56	0.0951

104	98.8	5.2	27.04	0.2737
Total				12.2347

$$\therefore \chi^2 = 12.2347$$

Table value of  $\chi^2$  for 6 d.f. at 5% level of significance is 12.59

Inference

$$\chi^2 < \chi^2_{\text{tab}}$$

We accept the null hypothesis.

The severity of the disease has no association with blood group.

### Example 3

In order to determine the possible effect of a chemical treatment on the rate of germination of cotton seeds a pot culture experiment was conducted. The results are given below

Chemical treatment and germination of cotton seeds

	Germinated	Not germinated	Total
Chemically Treated	118	22	140
Untreated	120	40	160
Total	238	62	300

Does the chemical treatment improve the germination rate of cotton seeds?

### Solution

$H_0$ : The chemical treatment does not improve the germination rate of cotton seeds.

$H_1$ : The chemical treatment improves the germination rate of cotton seeds.

Level of significance = 1%

Test statistic

$$= \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \text{ with 1 d.f.}$$

$$\chi^2 = \frac{300(118 \times 40 - 22 \times 120)^2}{140 \times 160 \times 62 \times 238} = 3.927$$

**Table value**

$$\chi^2 (1) \text{ d.f. at } 1\% \text{ L.O.S} = 6.635$$

Inference

$$\chi^2 < \chi^2_{\text{tab}}$$

We accept the null hypothesis.

The chemical treatment will not improve the germination rate of cotton seeds significantly.

**Example 4**

In an experiment on the effect of a growth regulator on fruit setting in muskmelon the following results were obtained. Test whether the fruit setting in muskmelon and the application of growth regulator are independent at 1% level.

	<b>Fruit set</b>	<b>Fruit not set</b>	<b>Total</b>
Treated	16	9	25
Control	4	21	25
Total	20	30	50

**Solution**

$H_0$ : Fruit setting in muskmelon does not depend on the application of growth regulator.

$H_1$ : Fruit setting in muskmelon depend on the application of growth regulator.

Level of significance = 1%

After Yates correction we have

	Fruit set	Fruit not set	Total
Treated	15.5	9.5	25
Control	4.5	20.5	25
Total	20	30	50

Tet statistic

$$\chi^2 = \frac{n \left( |ad - bc| - \frac{n}{2} \right)^2}{(a+b)(c+d)(a+c)(b+d)}$$
$$\chi^2 = \frac{50 \left[ |15.5 \times 20.5 - 9.5 \times 4.5| - \frac{50}{2} \right]^2}{25 \times 25 \times 20 \times 30} = 8.33$$

**Table value**

$\chi^2$  (1) d.f. at 1 % level of significance is 6.635

Inference

$$\chi^2 > \chi^2_{\text{tab}}$$

We reject the null hypothesis.

Fruit setting in muskmelon is influenced by the growth regulator. Application of growth regulator will increase fruit setting in musk melon.

### Questions

1. The calculated value of  $\chi^2$  is  
(a) always positive (b) always negative  
(c) can be either positive or negative (d) none of these

**Ans: always positive**

2. Degrees of freedom for Chi-square in case of contingency table of order (4 × 3) are  
(a) 12 (b) 9 (c) 8 (d) 6

**Ans: 6**

3. One condition for application of  $\chi^2$  test is that no cell frequency should be less than five.

**Ans: True**

4. The distribution of the  $\chi^2$  depends on the degrees of freedom.

**Ans: True**

5. The greater the discrepancy between the observed and expected Frequency lesser the value of  $\chi^2$ .

**Ans: False**

6. When observed and expected frequencies completely coincide  $\chi^2$  will be zero.

**Ans: True**

7. What is a contingency table?

8. When and how to apply Yates correction?

9. Explain the  $\chi^2$  test of goodness of fit?

10. Explain how to test the independence of attributes?

## Lecture.12

### **Correlation – definition – Scatter diagram -Pearson’s correlation co-efficient – properties of correlation coefficient**

#### **Correlation**

Correlation is the study of relationship between two or more variables. Whenever we conduct any experiment we gather information on more related variables. When there are two related variables their joint distribution is known as bivariate normal distribution and if there are more than two variables their joint distribution is known as multivariate normal distribution.

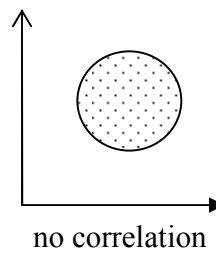
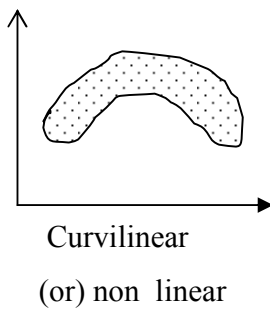
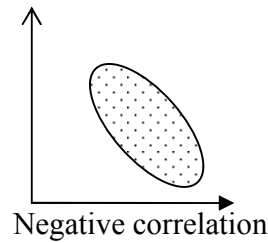
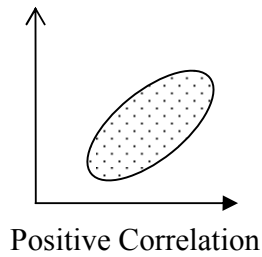
In case of bi-variate or multivariate normal distribution, we are interested in discovering and measuring the magnitude and direction of relationship between 2 or more variables. For this we use the tool known as correlation.

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.

#### **Scatter Diagram**

To investigate whether there is any relation between the variables X and Y we use scatter diagram. Let  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  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.





### Pearsons Correlation coefficient

The measures of the degree of relationship between two continuous variables is called correlation coefficient. It is denoted by  $r$  (in case of sample) and  $\rho$  (in case of population). The correlation coefficient  $r$  is known as Pearson's correlation coefficient as it was discovered by Karl Pearson. It is also called as product moment correlation.

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,

$$r = \frac{\frac{1}{n-1} (\sum (x-\bar{x})(y-\bar{y}))}{\sqrt{\frac{1}{n-1} \sum (x-\bar{x})^2} \sqrt{\frac{1}{n-1} \sum (y-\bar{y})^2}}$$

which can be simplified as

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}} \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

This correlation coefficient  $r$  is known as Pearson's Correlation coefficient. 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 of squares of  $X$  (i.e)  $SS(X)$  and second term is called sum of squares of  $Y$  (i.e)  $SS(Y)$

$$\therefore r = \frac{SP(XY)}{\sqrt{SS(X)}\sqrt{SS(Y)}}$$

The denominator in the above formula is always positive. The numerator may be positive or negative making  $r$  to be either positive or negative.

#### **Assumptions in correlation analysis:**

Correlation coefficient  $r$  is used under certain assumptions, they are

1. The variables under study are continuous random variables and they are normally distributed
2. The relationship between the variables is linear
3. Each pair of observations is unconnected with other pair (independent)

#### **Properties**

1. The correlation coefficient value ranges between  $-1$  and  $+1$ .
2. The correlation coefficient is not affected by change of origin or scale or both.
3. If  $r > 0$  it denotes positive correlation

$r < 0$  it denotes negative correlation between the two variables  $x$  and  $y$ .

$r = 0$  then the two variables  $x$  and  $y$  are not linearly correlated.(i.e)two variables are independent.

$r = +1$  then the correlation is perfect positive

$r = -1$  then the correlation is perfect negative.

#### **Testing the significance of $r$**

The significance of  $r$  can be tested by Student's  $t$  test. The test statistics 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 relationship between the variables is interpreted by the square of the correlation coefficient ( $r^2$ ) which is called coefficient of determination. 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 2 correlation coefficients.

### Problem

Compute Pearsons coefficient of correlation 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

### Solution

$H_0$ : The correlation coefficient r is not significant

$H_1$ : The correlation coefficient r is significant.

Level of significance 5%

From the data

$$n = 10$$

$$\sum x = 650 \quad \sum y = 660 \quad \sum xy = 45604 \quad \sum x^2 = 47648 \quad \sum y^2 = 45784$$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}} \sqrt{\sum y^2 - \frac{(\sum y)^2}{n}}}$$

$$\begin{aligned}
&= \frac{45604 - \frac{(650)(660)}{10}}{\sqrt{47648 - \frac{(650)^2}{10}} \sqrt{45784 - \frac{(660)^2}{10}}} \\
&= \frac{45604 - 42900}{(73.47)(47.1)} = 0.7804
\end{aligned}$$

Correlation coefficient is positively correlated.

### Test Statistic

$$t = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} \sim (n-2) \text{ d.f.}$$

$$t = \frac{0.7804}{\sqrt{\frac{1-(0.7804)^2}{10-2}}} = 3.530$$

$$t_{\text{tab}} = t_{(10-2, 5\% \text{los})} = 2.306$$

### Inference

$t > t_{\text{tab}}$ , we reject null hypothesis.

$\therefore$  The correlation coefficient  $r$  is significant. (i.e) there is a relation between plant height and yield.

### Questions

1. Limits for correlation coefficient.

- (a)  $-1 \leq r \leq 1$                       (b)  $0 \leq r \leq 1$   
(c)  $-1 \leq r \leq 0$                         (d)  $1 \leq r \leq 2$

**Ans:  $-1 \leq r \leq 1$**

2. The correlation coefficient is unaffected by change of

- (a) Origin                                      (b) scale  
(c) Scale & origin                            (d) None of these

**Ans: scale & origin**

3. When  $r = +1$ , there is Perfect positive correlation.

**Ans: True**

4. Karl pearsons correlation coefficient is calculated only when the two variables are continuous.

**Ans: True**

5. The correlation between two variables is symmetric

**Ans: True**

6. The correlation between two variables is known as multiple correlation.

**Ans: False**

7. What is a scatter diagram? Mention its uses

8. Define correlation.

9. Explain the method how to calculate the Karl pearsons correlation coefficient?

10. Mention the properties of the correlation coefficient?

### Lecture.13

#### Regression – definition – fitting of simple linear regression equation – testing the significance of the regression coefficient

##### Regression

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 repressor. The variable representing effect is known as dependent variable and is denoted by Y. Y 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 and if the relation between the two variables is a straight line it is known as simple linear regression. When there are more than two variables and one of the variables is dependent upon others, the functional relationship is known as multiple regression. The regression line is of the form  $y=a+bx$  where a is a constant or intercept and b is the regression coefficient or the slope. The values of 'a' and 'b' can be calculated by using the method of least squares. An alternate method of calculating the values of a and b are by using the formula:

The regression equation of y on x is given by  $y = a + bx$

The regression coefficient of y on x is given by

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

and  $a = \bar{y} - b \bar{x}$

The regression line indicates the average value of the dependent variable Y associated with a particular value of independent variable X.

### Assumptions

1. The x's are non-random or fixed constants
2. At each fixed value of X the corresponding values of Y have a normal distribution about a mean.
3. For any given x, the variance of Y is same.
4. The values of y observed at different levels of x are completely independent.

### Properties of Regression coefficients

1. The correlation coefficient is the geometric mean of the two regression coefficients
2. Regression coefficients are independent of change of origin but not of scale.
3. If one regression coefficient is greater than unit, then the other must be less than unit but not vice versa. ie. both the regression coefficients can be less than unity but both cannot be greater than unity, ie. if  $b_1 > 1$  then  $b_2 < 1$  and if  $b_2 > 1$ , then  $b_1 < 1$ .
4. Also if one regression coefficient is positive the other must be positive (in this case the correlation coefficient is the positive square root of the product of the two regression coefficients) and if one regression coefficient is negative the other must be negative (in this case the correlation coefficient is the negative square root of the product of the two regression coefficients). ie. if  $b_1 > 0$ , then  $b_2 > 0$  and if  $b_1 < 0$ , then  $b_2 < 0$ .
5. If  $\theta$  is the angle between the two regression lines then it is given by

$$\tan \theta = \frac{(1-r^2)\sigma_x\sigma_y}{r(\sigma_x^2 + \sigma_y^2)}$$

### Testing the significance of regression co-efficient

To test the significance of the regression coefficient we can apply either a t test or analysis of variance (F test). The ANOVA table for testing the regression coefficient will be as follows:

Sources of variation	d.f.	SS	MS	F
Due to regression	1	SS(b)	$S_b^2$	$S_b^2 / S_e^2$

Deviation from regression	n-2	SS(Y)-SS(b)	$S_e^2$	
Total	n-1	SS(Y)		

In case of t test the test statistic is given by

$$t = b / SE (b) \text{ where } SE (b) = s_e^2 / SS(X)$$

The regression analysis is useful in predicting the value of one variable from the given values of another variable. Another use of regression analysis is to find out the causal relationship between variables.

### Uses of Regression

The regression analysis is useful in predicting the value of one variable from the given value of another variable. Such predictions are useful when it is very difficult or expensive to measure the dependent variable, Y. The other use of the regression analysis is to find out the causal relationship between variables. Suppose we manipulate the variable X and obtain a significant regression of variables Y on the variable X. Thus we can say that there is a causal relationship between the variable X and Y. The causal relationship between nitrogen content of soil and growth rate in a plant, or the dose of an insecticide and mortality of the insect population may be established in this way.

### Example 1

From a paddy field, 36 plants were selected at random. The length of panicles(x) and the number of grains per panicle (y) of the selected plants were recorded. The results are given below. Fit a regression line y on x. Also test the significance (or) regression coefficient.

The length of panicles in cm (x) and the number of grains per panicle (y) of paddy plants.

S.No.	Y	X	S.No.	Y	X	S.No.	Y	X
1	95	22.4	13	143	24.5	25	112	22.9
2	109	23.3	14	127	23.6	26	131	23.9
3	133	24.1	15	92	21.1	27	147	24.8
4	132	24.3	16	88	21.4	28	90	21.2
5	136	23.5	17	99	23.4	29	110	22.2
6	116	22.3	18	129	23.4	30	106	22.7
7	126	23.9	19	91	21.6	31	127	23.0



8	124	24.0	20	103	21.4	32	145	24.0
9	137	24.9	21	114	23.3	33	85	20.6
10	90	20.0	22	124	24.4	34	94	21.0
11	107	19.8	23	143	24.4	35	142	24.0
12	108	22.0	24	108	22.5	36	111	23.1

Null Hypothesis  $H_0$ : regression coefficient is not significant.

Alternative Hypothesis  $H_1$ : regression coefficient is significant.

$$\sum y = 4174 \quad \sum y^2 = 496258 \quad \bar{y} = \frac{\sum y}{n} = 115.94$$

$$\sum x = 822.9 \quad \sum x^2 = 18876.83 \quad \bar{x} = \frac{\sum x}{n} = 22.86$$

$$\sum xy = 96183.4$$

$$SS(Y) = \sum y^2 - \frac{(\sum y)^2}{n} = 496258 - \frac{(4174)^2}{36} = 12305.8889$$

$$SS(X) = \sum x^2 - \frac{(\sum x)^2}{n} = 18876.83 - \frac{(822.9)^2}{36} = 66.7075$$

The regression line y on x is  $\bar{y} = a + b \bar{x}$

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{96183.4 - \frac{(822.9)(4174)}{36}}{66.7075} = 11.5837$$

$$\bar{y} = a + b \bar{x}$$

$$115.94 = a + (11.5837)(22.86)$$

$$a = 115.94 - 264.8034$$

$$a = -148.8633$$

The fitted regression line is  $y = -148.8633 + 11.5837x$

$$SS(b) = \frac{\left( \sum xy - \frac{\sum x \sum y}{n} \right)^2}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{(722.7167)^2}{66.7075} = 8950.8841$$

### Anova Table

Sources of Variation	d.f.	SS	MSS	F
Regression	1	8950.8841	8950.8841	90.7093
Error	36-2=34	3355.0048	98.6766	
Total	35	12305.8889		

For t-test

$$t = \frac{b}{SE(b)} \sim t_{(n-2)} d.f$$

$$SE(b) = \sqrt{\frac{Se^2}{SS(X)}} = \sqrt{\frac{98.6776}{66.7075}} = 1.2162$$

$$t = \frac{11.5837}{1.2162} = 9.5245$$

Table Value:

$t_{(n-2)}$  d.f.= $t_{34}$  d.f at 5% level=2.032

$t > t_{tab}$ . we reject  $H_0$ .

Hence t is significant.

### Questions

1. When the correlation coefficient  $r = +1$ , then the two regression lines

- a) are perpendicular to each other
- b) coincide
- c) are parallel to each other
- d) none of these

**Ans: coincide**

2. If one regression coefficient is greater than unity then the other must be
- a) greater than unity
  - b) equal to unity
  - c) less than unity
  - d) none of these

**Ans: less than unity**

3. If the correlation between the two variables is positive the regression coefficient will be positive.

**Ans: True**

4. The Dependent variable is also called as predicted variable.

**Ans: True**

5. Correlation coefficient is the geometric mean of two regression coefficients.

**Ans: True**

6. Regression gives the functional relationship between two variables.

**Ans: True**

7. What is meant by Cause and effect?

8. State the properties of regression coefficient.

9. From the following data, find the regression equation

$$\sum X = 21, \sum Y = 20, \sum X^2 = 91, \sum XY = 74, n = 7$$

10. Explain how to fit the regression equation of y on x and test the significance of the regression coefficient.

## Lecture.14

**Design of experiments – basic concepts – treatment – experimental unit – experimental error - basic principle – replication, randomization and local control.**

### Design of Experiments

Choice of treatments, method of assigning treatments to experimental units and arrangement of experimental units in different patterns are known as designing an experiment. We study the effect of changes in one variable on another variable. For example how the application of various doses of fertilizer affects the grain yield. Variable whose change we wish to study is known as **response variable**. Variable whose effect on the response variable we wish to study is known as **factor**.

**Treatment:** Objects of comparison in an experiment are defined as treatments. Examples are Varieties tried in a trail and different chemicals.

**Experimental unit:** The object to which treatments are applied or basic objects on which the experiment is conducted is known as experimental unit.

Example: piece of land, an animal, etc

**Experimental error:** Response from all experimental units receiving the same treatment may not be same even under similar conditions. These variations in responses may be due to various reasons. Other factors like heterogeneity of soil, climatic factors and genetic differences, etc also may cause variations (known as extraneous factors). The variations in response caused by extraneous factors are known as **experimental error**.

Our aim of designing an experiment will be to minimize the experimental error.

### Basic principles

To reduce the experimental error we adopt certain principles known as basic principles of experimental design.

The basic principles are 1) Replication, 2) Randomization and 3) Local control

## **Replication**

Repeated application of the treatments is known as replication.

When the treatment is applied only once we have no means of knowing about the variation in the results of a treatment. Only when we repeat several times we can estimate the experimental error.

With the help of experimental error we can determine whether the obtained differences between treatment means are real or not. When the number of replications is increased, experimental error reduces.

## **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**

Experimental error is based on the variations from experimental unit to experimental unit. This suggests that if we group the homogenous experimental units into blocks, the experimental error will be reduced considerably. Grouping of homogenous experimental units into blocks is known as local control of error.

In order to have valid estimate of experimental error the principles of replication and randomization are used.

In order to reduce the experimental error, the principles of replication and local control are used.

In general to have precise, valid and accurate result we adopt the basic principles.

## **Questions**

1. For valid conclusions we should have

- |                       |                     |
|-----------------------|---------------------|
| (a) Unbiased estimate | (b) biased estimate |
| (c) random estimate   | (d) none of these   |

**Ans: Unbiased estimate**

2. Response variable is also called as
- |                          |                        |
|--------------------------|------------------------|
| (a) Independent variable | (b) dependent variable |
| (c) treatment            | (d) error              |

**Ans: dependent variable**

3. The genetic differences of varieties are termed as extraneous factors.

**Ans: True**

4. Repetition of the treatment is known as replication.

**Ans: True**

5. Replication will increase the error.

**Ans: False**

6. Basic principles are adopted to reduce the experimental error.

**Ans: True**

7. What is experimental error?

8. Define treatment and experimental unit.

9. What is meant by designing an experiment?

10. Explain the basic principles and its uses?

## Lecture.15

### Completely randomized design – description – layout – analysis – advantages and disadvantages

#### Completely Randomized Design (CRD)

CRD is the basic single factor design. 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.

#### Layout of a CRD

Completely randomized Design is the one in which all the experimental 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 left to right for all rows.

**Step 3:** Assign the treatments to the experimental units by using random numbers.

The statistical model for CRD with one observation per unit

$$Y_{ij} = \mu + t_i + e_{ij}$$

$\mu$  = overall mean effect

$t_i$  = true effect of the  $i^{\text{th}}$  treatment

$e_{ij}$  = error term of the  $j^{\text{th}}$  unit receiving  $i^{\text{th}}$  treatment

The arrangement of data in CRD is as follows:

	Treatments				
	T <sub>1</sub>	T <sub>2</sub>	T <sub>i</sub>	T <sub>k</sub>	
	y <sub>11</sub>	y <sub>21</sub>	y <sub>i1</sub>	Y <sub>k1</sub>	
	y <sub>12</sub>	y <sub>22</sub>	y <sub>i2</sub>	Y <sub>k2</sub>	
	y <sub>1r1</sub>	y <sub>2r2</sub>	y <sub>iri</sub>	Y <sub>k rk</sub>	
Total	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>i</sub>	T <sub>k</sub>	GT

(GT – Grand total)

The null hypothesis will be

H<sub>0</sub>: μ<sub>1</sub> = μ<sub>2</sub> = ..... = μ<sub>k</sub> or There is no significant difference between the treatments

And the alternative hypothesis is

H<sub>1</sub>: μ<sub>1</sub> ≠ μ<sub>2</sub> ≠ ..... ≠ μ<sub>k</sub>. There is significant difference between the treatments

The different steps in forming the analysis of variance table for a CRD are:

1. 
$$C.F = \frac{(GT)^2}{n}$$

n = Total number of observations

2. 
$$\text{Total SS} = \text{TSS} = \sum_{i=1}^k \sum_{j=1}^v y_{ij}^2 - C.F$$

3. 
$$\begin{aligned} \text{Treatment SS} = \text{TrSS} &= \frac{Y_1^2}{r_1} + \frac{Y_2^2}{r_2} + \dots + \frac{Y_k^2}{r_k} - C.F \\ &= \sum_{i=1}^k \frac{Y_i^2}{r_i} - C.F \end{aligned}$$

4. 
$$\begin{aligned} \text{Error SS} = \text{ESS} &= \sum_{i=1}^k \sum_{j=1}^r y_{ij}^2 - \sum_{i=1}^k \frac{Y_i^2}{r_i} \\ &= \text{TSS} - \text{TrSS} \end{aligned}$$

5. Form the following ANOVA table and calculate F value.

Source of variation	d.f.	SS	MS	F
Treatments	t-1	TrSS	TrMS = $\frac{TrSS}{t-1}$	$\frac{TrMS}{EMS}$
Error	n-t	ESS	EMS = $\frac{ESS}{n-t}$	
<b>Total</b>	n-1	TSS		



6. Compare the calculated F with the critical value of F corresponding to treatment degrees of freedom and error degrees of freedom so that acceptance or rejection of the null hypothesis can be determined.

7. If null hypothesis is rejected that indicates there is significant differences between the different treatments.

8. Calculate C D value.

C.D. = SE(d). t

$$\text{where S.E(d)} = \sqrt{EMS\left(\frac{1}{r_i} + \frac{1}{r_j}\right)}$$

$r_i$  = 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 specified level of significance, either 5% or 1%.

### **Advantages of a 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.
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.

### **Disadvantages of a 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.

## Questions

1. CRD can be used with  
(a) Equal replication (b) unequal replication  
(c) Equal and unequal replication (d) single replication

**Ans: Equal and unequal replication**

2. When there are 5 treatments each replicated 4 times the total number of experimental plots will be

(a) 5 (b) 4 (c) 9 (d) 20

**Ans: 20**

3. In CRD the error degrees of freedom is  $rt-1$ .

**Ans: True**

4. CRD can be adopted only when the experimental material is homogenous.

**Ans: True**

5. CRD is a single factor experiment.

**Ans: True**

6. In CRD the total sum of squares is divided into treatment sum of squares, Replication sum of squares and error sum of squares.

**Ans: False**

7. Mention any two advantages of CRD?

8. When the treatments are large in a CRD what will happen to the precision of the experiment?

9. Explain the Layout of the CRD?

10. Explain the Layout of the CRD?

## Lecture.16

### Randomized blocks design – description – layout – analysis – advantages and disadvantages

#### Randomized Blocks 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 west then the blocks are formed in the opposite direction. Such an arrangement of grouping the heterogeneous units into homogenous blocks is known as randomized blocks design. Each block consists of as many experimental units as the number of treatments. The treatments are allocated randomly to the experimental units within each block independently such that each treatment occurs once. The number of blocks is chosen to be equal to the number of replications for the treatments.

The analysis of variance model for RBD is

$$Y_{ij} = \mu + t_i + r_j + e_{ij}$$

where

$\mu$  = the overall mean

$t_i$  = the  $i^{\text{th}}$  treatment effect

$r_j$  = the  $j^{\text{th}}$  replication effect

$e_{ij}$  = the error term for  $i^{\text{th}}$  treatment and  $j^{\text{th}}$  replication

#### Analysis of RBD

The results of RBD can be arranged in a two way table according to the replications (blocks) and treatments.

There will be  $r \times t$  observations in total where  $r$  stands for number of replications and  $t$  for number of treatments. .

The data are arranged in a two way table form by representing treatments in rows and replications in columns.

Treatment	Replication					Total
	1	2	3	.....	r	
1	y11	y12	y13	.....	y1r	T1
2	y21	y22	y23	.....	y2r	T2
3	y31	y32	y33	.....	y3r	T3
t	yt1	yt2	yt3	.....	ytr	Tt
Total	R1	R2	R3		Rr	G.T

In this design the total variance is divided into three sources of variation viz., between replications, between treatments and error

$$CF = \frac{(GT)^2}{rt}$$

$$\text{Total SS} = \text{TSS} = \sum \sum y_{ij}^2 - CF$$

$$\text{Replication SS} = \text{RSS} = \frac{1}{t} \sum R_j^2 - CF$$

$$\text{Treatments SS} = \text{TrSS} = \frac{1}{r} \sum T_i^2 - CF$$

$$\text{Error SS} = \text{ESS} = \text{Total SS} - \text{Replication SS} - \text{Treatment SS}$$

The skeleton ANOVA table for RBD with t treatments and r replications

Sources of variation	d.f.	SS	MS	F Value
Replication	r-1	RSS	RMS	RM S/ EM S
Treatment	t-1	TrSS	TrMS	TrMS/EMS
Error	(r-1)(t-1)	ESS	EMS	
Total	rt-1	TSS		

$$CD = SE(d) \cdot t \quad \text{where } S.E(d) = \sqrt{\frac{2EMS}{r}}$$

t = critical value of t for a specified level of significance and error degrees of freedom

Based on the CD value the bar chart can be drawn.

From the bar chart conclusion can be written.

### Advantages of RBD

The precision is more in RBD. The amount of information obtained in RBD is more as compared to CRD. RBD is more flexible. Statistical analysis is simple and easy. Even if some values are missing, still the analysis can be done by using missing plot technique.

### **Disadvantages of RBD**

When the number of treatments is increased, the block size will increase. If the block size is large maintaining homogeneity is difficult and hence when more number of treatments is present this design may not be suitable.

### **Questions**

1. RBD can be used with  
(a) Equal replication                      (b) unequal replication  
(c) Equal and unequal replication      (d) single replication

**Ans: Equal replication**

2. When there are 5 treatments each replicated 4 times the total number of experimental plots will be  
(a) 5      (b) 4      (c) 9      (d) 20

**Ans: 20**

3. In RBD the error degrees of freedom is  $(r-1)(t-1)$ .

**Ans: True**

4. RBD can be adopted when the experimental material is heterogeneous.

**Ans: True**

5. In RBD the blocking is done in one direction.

**Ans: True**

6. In RBD the total sum of squares is divided into treatment sum of squares, Replication sum of squares and error sum of squares.

**Ans: True**

7. Mention any two advantages of RBD?

8. Furnish the ANOVA model for RBD

9. Explain the Layout of the RBD?

10. Explain the computational procedure of RBD?

## Lecture.17

### Latin square design – description – layout – analysis – advantages and disadvantages.

#### Latin Square Design

When the experimental material is divided into rows and columns and the treatments are allocated such that each treatment occurs only once in each row and each column, the design is known as L S D.

In LSD the treatments are usually denoted by A B C D etc.

For a 5 x 5 LSD the arrangements may be

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>B</i>	<i>A</i>	<i>E</i>	<i>C</i>	<i>D</i>
<i>C</i>	<i>D</i>	<i>A</i>	<i>E</i>	<i>B</i>
<i>D</i>	<i>E</i>	<i>B</i>	<i>A</i>	<i>C</i>
<i>E</i>	<i>C</i>	<i>D</i>	<i>B</i>	<i>A</i>
<b>Square 1</b>				

	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>B</i>	<i>A</i>	<i>D</i>	<i>E</i>	<i>C</i>
<i>C</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>D</i>
<i>D</i>	<i>C</i>	<i>E</i>	<i>A</i>	<i>B</i>
<i>E</i>	<i>D</i>	<i>B</i>	<i>C</i>	<i>A</i>
<b>Square 2</b>				

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>
<i>C</i>	<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>
<i>D</i>	<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>
<i>E</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
<b>Square 3</b>				

#### Analysis

The ANOVA model for LSD is

$$Y_{ijk} = \mu + r_i + c_j + t_k + e_{ijk}$$

$r_i$  is the  $i^{\text{th}}$  row effect

$c_j$  is the  $j^{\text{th}}$  column effect

$t_k$  is the  $k^{\text{th}}$  treatment effect and

$e_{ijk}$  is the error term

The analysis of variance table for LSD is as follows:

Sources of Variation	d.f.	S S	M S	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		

## **F table value**

$F_{[t-1),(t-1)(t-2)]}$  degrees of freedom at 5% or 1% level of significance

Steps to calculate the above Sum of Squares are as follows:

$$\text{Correction Factor (CF)} = \frac{(GT)^2}{(t)^2}$$

$$\text{Total Sum of Squares (TSS)} = \sum (y_{ijk})^2 - CF$$

$$\text{Row sum of squares (RSS)} = \frac{1}{t} \sum_{i=1}^t (R_i)^2 - CF$$

$$\text{Column sum of squares (CSS)} = \frac{1}{t} \sum_{j=1}^t (C_j)^2 - CF$$

$$\text{Treatment sum of squares (TrSS)} = \frac{1}{t} \sum_{k=1}^t (T_k)^2 - CF$$

$$\text{Error Sum of Squares} = \text{TSS} - \text{RSS} - \text{CSS} - \text{TrSS}$$

These results can be summarized in the form of analysis of variance table.

Calculation of SE, SE (d) and CD values

$$SE = \sqrt{\frac{EMS}{r}}$$

where r is the number of rows

$$SE(d) = \sqrt{2} SE .$$

$$CD = SE (d). t$$

where t = table value of t for a specified level of significance and error degrees of freedom

Using CD value the bar chart can be drawn and the conclusion may be written.

### **Advantages**

- LSD is more efficient than RBD or CRD. This is because of double grouping that will result in small experimental error.
- When missing values are present, missing plot technique can be used and analysed.

### **Disadvantages**

- This design is not as flexible as RBD or CRD as the number of treatments is limited to the number of rows and columns. LSD is seldom used when the number of treatments is more than 12. LSD is not suitable for treatments less than five.

**Because of the limitations on the number of treatments, LSD is not widely used in agricultural experiments.**

**Note: The number of sources of variation is two for CRD, three for RBD and four for LSD.**

### **Questions**

1. In a Latin Square design the number of rows will be equal to
- |                        |  |
|------------------------|--|
| a) No. of columns      | b) No. of Treatments                     |
| c) No. of Replications | d) No. of Columns & Number of Treatments |

**Ans: No. of Columns & Number of Treatments**

2. In a Latin Square design with 5 treatments the number of experimental units will be equal to
- |       |       |       |       |
|-------|-------|-------|-------|
| a) 25 | b) 20 | c) 24 | d) 36 |
|-------|-------|-------|-------|

**Ans: 25**

3. If the number of experimental units is 36 then the number of rows will be equal to 6.

**Ans: True**

4. The error degrees of freedom in LSD with t treatments will be  $(t-1)(t-2)$ .

**Ans: True**

5. If the experimental material is homogeneous then LSD can be adopted.

**Ans: False**

6. In a LSD each treatment should occur only once in each row and each column.

**Ans: True**



7. Furnish the ANOVA model for LSD.
8. What is a Latin Square Design?
9. State the advantages and disadvantages of LSD.
10. Explain the computational procedure of LSD?

## Lecture.18

### **Factorial experiments – factor and levels – types – symmetrical and asymmetrical – simple, main and interaction effects – advantages and disadvantages**

#### **Factorial Experiments**

When two or more number of factors are investigated simultaneously in a single experiment such experiments are called as factorial experiments.

#### **Terminologies**

1. **Factor:** Factor refers to a set of related treatments. We may apply of different doses of nitrogen to a crop. Hence nitrogen irrespective of doses is a factor.
2. **Levels of a factor:** Different states or components making up a factor are known as the levels of that factor. eg different doses of nitrogen.

#### **Types of factorial Experiment**

A factorial experiment is named based on the number of factors and levels of factors. For example, when there are 3 factors each at 2 levels the experiment is known as  $2 \times 2 \times 2$  or  $2^3$  factorial experiments.

If there are 2 factors each at 3 levels then it is known as  $3 \times 3$  or  $3^2$  factorial experiment.

- In general if there are n factors each with p levels then it is known as  $p^n$  factorial experiment.
- For varying number of levels the arrangement is described by the product. For example, an experiment with 3 factors each at 2 levels, 3 levels and 4 levels respectively then it is known as  $2 \times 3 \times 4$  factorial experiment.
- If all the factors have the same number of levels the experiment is known as symmetrical factorial otherwise it is called as mixed factorial.
- Factors are represented by capital letters. Treatment combinations are usually by small letters.
- For example, if there are 2 varieties v0 and v1 and 2 dates of sowing d0 and d1 the treatment combinations will be
- v0d0, v1d0, v0d1 and v1d1.

### Simple and Main Effects

**Simple effect** of a factor is the difference between its responses for a fixed level of other factors.

**Main effect** is defined as the average of the simple effects.

**Interaction** is defined as the dependence of factors in their responses. Interaction is measured as the mean of the differences between simple effects.

### Advantages

1. In such type of experiments we study the individual effects of each factor and their interactions.
2. In factorial experiments a wide range of factor combinations are used.
3. Factorial approach will result in considerable saving of the experimental resources, experimental material and time.

### Disadvantages

1. When number of factors or levels of factors or both are increased, the number of treatment combinations increases. Consequently block size increases. If block size increases it may be difficult to maintain homogeneity of experimental material. This will lead to increase in experimental error and loss of precision in the experiment.
2. All treatment combinations are to be included for the experiment irrespective of its importance and hence this results in wastage of experimental material and time.
3. When many treatment combinations are included the execution of the experiment and statistical analysis become difficult.

### Questions

1. In a factorial experiment the minimum number of factors will be  
a) One      b) two      c) three      d) none of the above

**Ans: two**

2. With factor A which has two levels and factor B which has three levels the number of combinations will be  
a) 5      b) 6      c) 2      d) 3

**Ans: 6**

3. In a factorial experiments Factor refers to a set of related treatments.

**Ans: True**

4. A three 3 x3 factorial experiment can also be written as  $3^2$  factorial experiment.

**Ans: True**

5. In a factorial experiment factors are always represented by small letters.

**Ans: False**

6. If all the factors have the same number of levels the experiment is known as symmetrical factorial.

**Ans: True**

7. What is a main effect?

8. What is an interaction effect?

9. What are the advantages and disadvantages of Factorial experiments?

10. Write down the treatment combinations for a factorial experiment with varieties as one factor and n levels as the second factor. The levels for the varieties are 4 ( $V_0, V_1, V_2$  &  $V_3$ ) for n levels 4 ( $n_0, n_1, n_2$  &  $n_3$ )

## Lecture.19

### 2<sup>2</sup> Factorial Experiments in RBD – lay out – analysis

#### 2<sup>2</sup> Factorial Experiments in RBD

2<sup>2</sup> factorial experiment means two factors each at two levels. Suppose the two factors are A and B and both are tried with two levels the total number of treatment combinations will be four i.e. a<sub>0</sub>b<sub>0</sub>, a<sub>0</sub>b<sub>1</sub>, a<sub>1</sub>b<sub>0</sub> and a<sub>1</sub>b<sub>1</sub>.

The allotment of these four treatment combinations will be as allotted in RBD. That is each block is divided into four experimental units. By using the random numbers these four combinations are allotted at random for each block separately.

The analysis of variance table for two factors A with a levels and B with b levels with r replications tried in RBD will be as follows:

Sources of Variation	d.f.	SS	MS	F
Replications	r-1	RSS	RMS	
Factor A	a-1	ASS	AMS	AMS / EMS
Factor B	b-1	BSS	BMS	BMS / EMS
AB (interaction)	(a-1)(b-1)	ABSS	ABMS	ABMS / EMS
Error	(r-1)(ab-1)	ESS	EMS	
Total	rab-1	TSS		

As in the previous designs calculate the replication totals to calculate the RSS, TSS in the usual way. To calculate ASS, BSS and ABSS, form a two way table A X B by taking the levels of A in rows and levels of B in the columns. To get the values in this table the missing factor is replication. That is by adding over replication we can form this table.

$$CF = \frac{(GT)^2}{r \times a \times b}$$

$$RSS = \frac{\sum_{i=1}^r R_i^2}{a \times b} - CF$$

A X B Two way table

B A	b <sub>0</sub>	b <sub>1</sub>	Total
a <sub>0</sub>	a <sub>0</sub> b <sub>0</sub>	a <sub>0</sub> b <sub>1</sub>	A <sub>0</sub>
a <sub>1</sub>	a <sub>1</sub> b <sub>0</sub>	a <sub>1</sub> b <sub>1</sub>	A <sub>1</sub>
Total	B <sub>0</sub>	B <sub>1</sub>	Grand Total

$$ASS = \frac{A_0^2 + A_1^2}{b \times r} - CF$$

$$BSS = \frac{B_0^2 + B_1^2}{a \times r} - CF$$

$$ABSS = \frac{(a_0 b_0)^2 + (a_0 b_1)^2 + (a_1 b_0)^2 + (a_1 b_1)^2}{r} - CF - ASS - BSS$$

$$ESS = TSS - RSS - ASS - BSS - ABSS$$

By substituting the above values in the ANOVA table corresponding to the columns sum of squares, the mean squares and F value can be calculated.

### Questions

1. 2<sup>2</sup> factorial experiment means two factors each at  
 a) two levels    b) three levels    c) four levels    d) one level

**Ans: two levels**

2. If the total number of combinations are four then each block is divided into \_\_\_\_\_  
 experimental units

- a) Two    b) three    c) four    d) none of the above

**Ans: four**

3. The two factors are A and B and both are tried with three levels the total number of combinations will be nine.

**Ans: True**

4. The error degrees of freedom for an experiment with one factor at 2 levels and another at three levels and 3 replications will be 10.

**Ans: True**

5. In a two factor experiment the highest order interaction will be the two factor interaction.

**Ans: True**

6. In a factorial experiment with two factors the treatment sum of squares will be split up into factor 1, factor 2 and factor 1 x factor 2 interaction sum of squares.

**Ans: True**

7. How to calculate the replication sum of squares in a two factor experiment.

8. In a factorial experiment what is the total number of experimental units when there are 3 replications, 4 levels for factor A and 3 levels for factor B

9. Furnish the ANOVA table for a two factor experiment with factor A at a levels, Factor B at b levels and the number of replication r.

10. Explain the procedure of forming a A X B two way table and calculating the factor ASS, BSS and A X BSS.

## Lecture.20

### 2<sup>3</sup> factorial experiments in RBD – lay out – analysis.

#### 2<sup>3</sup> Factorial Experiment in RBD

2<sup>3</sup> factorial experiment mean three factors each at two levels. Suppose the three factors are A, B and C are tried with two levels the total number of combinations will be eight i.e. a<sub>0</sub>b<sub>0</sub>c<sub>0</sub>, a<sub>0</sub>b<sub>0</sub>c<sub>1</sub>, a<sub>0</sub>b<sub>1</sub>c<sub>0</sub>, a<sub>0</sub>b<sub>1</sub>c<sub>1</sub>, a<sub>1</sub>b<sub>0</sub>c<sub>0</sub>, a<sub>1</sub>b<sub>0</sub>c<sub>1</sub>, a<sub>1</sub>b<sub>1</sub>c<sub>0</sub> and a<sub>1</sub>b<sub>1</sub>c<sub>1</sub>.

The allotment of these eight treatment combinations will be as allotted in RBD. That is each block is divided into eight experimental units. By using the random numbers these eight combinations are allotted at random for each block separately.

**The analysis of variance table for three factors A with a levels, B with b levels and C with c levels with r replications tried in RBD will be as follows:**

Sources of Variation	d.f.	SS	MS	F
Replications	r-1	RSS	RMS	
Factor A	a-1	ASS	AMS	AMS / EMS
Factor B	b-1	BSS	BMS	BMS / EMS
Factor C	c-1	CSS	CMS	CMS / EMS
AB	(a-1)(b-1)	ABSS	ABMS	ABMS / EMS
AC	(a-1)(c-1)	ACSS	ACMS	ACMS / EMS
BC	(b-1)(c-1)	BCSS	BCMS	BCMS / EMS
ABC	(a-1)(b-1)(c-1)	ABCSS	ABCMS	ABCMS / EMS
Error	(r-1)(abc-1)	ESS	EMS	
Total	rabc-1	TSS		



## Analysis

1. Arrange the results as per treatment combinations and replications.

Treatment combination	Replication				Treatment Total
	R1	R2	R3	...	
a0b0c0					T1
a0b0c1					T2
a0b1c0					T3
a0b1c1					T4
a1b0c0					T5
a1b0c1					T6
a1b1c0					T7
a1b1c1					T8

As in the previous designs calculate the replication totals to calculate the CF, RSS, TSS, overall TrSS in the usual way. To calculate ASS, BSS, CSS, ABSS, ACSS, BCSS and ABCSS, form three two way tables A X B, AXC and BXC.

AXB two way table can be formed by taking the levels of A in rows and levels of B in the columns. To get the values in this table the missing factor is replication. That is by adding over replication we can form this table.

A X B Two way table

B A	$b_0$	$b_1$	Total
$a_0$	$a_0 b_0$	$a_0 b_1$	$A_0$
$a_1$	$a_1 b_0$	$a_1 b_1$	$A_1$
Total	$B_0$	$B_1$	Grand

			Total
--	--	--	-------

$$ASS = \frac{A_0^2 + A_1^2}{b \times c \times r} - CF$$

$$BSS = \frac{B_0^2 + B_1^2}{a \times c \times r} - CF$$

$$ABSS = \frac{(a_0 b_0)^2 + (a_0 b_1)^2 + (a_1 b_0)^2 + (a_1 b_1)^2}{c \times r} - CF - ASS - BSS$$

A X C two way table can be formed by taking the levels of A in rows and levels of C in the columns

A X C Two way table

C A	c <sub>0</sub>	c <sub>1</sub>	Total
a <sub>0</sub>	a <sub>0</sub> c <sub>0</sub>	a <sub>0</sub> c <sub>1</sub>	A <sub>0</sub>
a <sub>1</sub>	a <sub>1</sub> c <sub>0</sub>	a <sub>1</sub> c <sub>1</sub>	A <sub>1</sub>
Total	C <sub>0</sub>	C <sub>1</sub>	Grand Total

$$CSS = \frac{C_0^2 + C_1^2}{a \times b \times r} - CF$$

$$ACSS = \frac{(a_0 c_0)^2 + (a_0 c_1)^2 + (a_1 c_0)^2 + (a_1 c_1)^2}{b \times r} - CF - ASS - CSS$$

B X C two way table can be formed by taking the levels of B in rows and levels of C in the columns

B X C Two way table

C B	c <sub>0</sub>	c <sub>1</sub>	Total
--------	----------------	----------------	-------

$b_0$	$b_0 c_0$	$b_0 c_1$	$B_0$
$b_1$	$b_1 c_0$	$b_1 c_1$	$B_1$
Total	$C_0$	$C_1$	Grand Total

$$BCSS = \frac{(b_0 c_0)^2 + (b_0 c_1)^2 + (b_1 c_0)^2 + (b_1 c_1)^2}{a \times r} - CF - BSS - CSS$$

$$ABCSS = \frac{(a_0 b_0 c_0)^2 + (a_0 b_0 c_1)^2 + (a_0 b_1 c_0)^2 + (a_0 b_1 c_1)^2 + (a_1 b_0 c_0)^2 + (a_1 b_0 c_1)^2 + (a_1 b_1 c_0)^2 + (a_1 b_1 c_1)^2}{r}$$

-CF-ASS-BSS-CSS-ABSS-ACSS-BCSS

$$ESS = TSS - RSS - ASS - BSS - CSS - ABSS - ACSS - BCSS - ABCSS$$

By substituting the above values in the ANOVA table corresponding to the columns sum of squares, the mean squares and F value can be calculated.

### Questions

1.  $2^3$  factorial experiment means two factors each at

- a) two levels      b) three levels      c) four levels      d) one level

**Ans: three levels**

2. If the total number of combinations are eight then each block is divided into \_\_\_\_\_ experimental units

- a) Two      b) three      c) four      d) eight

**Ans: eight**

3. If the three factors are A, B and C are tried with three levels the total number of combinations will be twenty seven.

**Ans: True**

4. The error degrees of freedom for an experiment with one factor at 2 levels, second factor at 3 levels and the third at 3 levels and 3 replications will be 34.

**Ans: True**

5. In a three factor experiment the highest order interaction will be the two factor interaction.

**Ans: False**

6. In a factorial experiment with three factors the treatment sum of squares will be split up into factor 1, factor 2 and factor 1 x factor 2 interaction sum of squares.

**Ans: False**

7. How to calculate the three factor interaction sum of squares in a three factor experiment.

8. In a factorial experiment what is the total number of experimental units when there are 3 replications, 4 levels for factor A and 3 levels for factor B and C.

9. Furnish the ANOVA table for a three factor experiment with factor A at a levels, Factor B at b levels and factor C at c levels and the number of replication r.

10. Explain the procedure of forming a B X C two way table and calculating the factor BSS, CSS and B X CSS.

## **Lecture.21**

### **Split plot design – layout – ANOVA Table**

#### **Split-plot Design**

In field experiments certain factors may require larger plots than for others. For example, experiments on irrigation, tillage, etc requires larger areas. On the other hand experiments on fertilizers, etc may not require larger areas. To accommodate factors which require different sizes of experimental plots in the same experiment, split plot design has been evolved.

In this design, larger plots are taken for the factor which requires larger plots. Next each of the larger plots is split into smaller plots to accommodate the other factor. The different treatments are allotted at random to their respective plots. Such arrangement is called split plot design.

In split plot design the larger plots are called main plots and smaller plots within the larger plots are called as sub plots. The factor levels allotted to the main plots are main plot treatments and the factor levels allotted to sub plots are called as sub plot treatments.

#### **Layout and analysis of variance table**

First the main plot treatment and sub plot treatment are usually decided based on the needed precision. The factor for which greater precision is required is assigned to the sub plots.

The replication is then divided into number of main plots equivalent to main plot treatments. Each main plot is divided into subplots depending on the number of sub plot treatments. The main plot treatments are allocated at random to the main plots as in the case of RBD. Within each main plot the sub plot treatments are allocated at random as in the case of RBD. Thus randomization is done in two stages. The same procedure is followed for all the replications independently.

The analysis of variance will have two parts, which correspond to the main plots and sub-plots. For the main plot analysis, replication X main plot treatments table is

formed. From this two-way table sum of squares for replication, main plot treatments and error (a) are computed. For the analysis of sub-plot treatments, main plot X sub-plot treatments table is formed. From this table the sums of squares for sub-plot treatments and interaction between main plot and sub-plot treatments are computed. Error (b) sum of squares is found out by residual method. The analysis of variance table for a split plot design with  $m$  main plot treatments and  $s$  sub-plot treatments is given below.

**Analysis of variance for split plot with factor A with  $m$  levels in main plots and factor B with  $s$  levels in sub-plots will be as follows:**

Sources of Variation	d.f.	SS	MS	F
Replication	$r-1$	RSS	RMS	RMS/EMS (a)
A	$m-1$	ASS	AMS	AMS/EMS (a)
Error (a)	$(r-1)(m-1)$	ESS (a)	EMS (a)	
B	$s-1$	BSS	BMS	BMS/EMS (b)
AB	$(m-1)(s-1)$	ABSS	ABMS	ABMS/EMS (b)
Error (b)	$m(r-1)(s-1)$	ESS (b)	EMS (b)	
Total	$rms - 1$	TSS		

### Analysis

Arrange the results as follows

Treatment Combination	Replication				Total
	R1	R2	R3	...	
A0B0	a0b0	a0b0	a0b0	...	T00
A0B1	a0b1	a0b1	a0b1	...	T01
A0B2	a0b2	a0b2	a0b2	...	T02
Sub Total	A01	A02	A03	...	T0
A1B0	a1b0	a1b0	a1b0	...	T10
A1B1	a1b1	a1b1	a1b1	...	T11
A1B2	a1b2	a1b2	a1b2	...	T12
Sub Total	A11	A12	A13	...	T1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Total	R1	R2	R3	...	G.T

$$\text{Compute CF} = \frac{(G T)^2}{r \times m \times s}$$

$$\text{TSS} = [(a_0b_0)^2 + (a_0b_1)^2 + (a_0b_2)^2 + \dots] - \text{CF}$$

Form A x R Table and calculate RSS, ASS and Error (a) SS

Treatment	Replication				Total
	R1	R2	R3	...	
A0	A01	A02	A03	...	T0
A1	A11	A12	A13	...	T1
A2	A21	A22	A23	...	T2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Total	R1	R2	R3	...	GT

$$\text{RSS} = \left( \frac{R1^2 + R2^2 + R3^2 + \dots}{m \cdot s} \right) - \text{CF}$$

$$\text{ASS} = \left( \frac{T0^2 + T1^2 + T2^2 + \dots}{r \cdot s} \right) - \text{CF}$$

$$\text{A x R table SS} = \left( \frac{A01^2 + A02^2 + A03^2 + \dots}{b} \right) - \text{CF}$$

Error (a) SS = A x R TSS - RSS - ASS.

Form A x B Table and calculate BSS, Ax B SSS and Error (b) SS

Treatment	Replication				Total
	B0	B1	B2	...	
A0	T00	T01	T02	...	T0
A1	T10	T11	T12	...	T1
A2	T20	T21	T22	...	T2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Total	C0	C1	C2	...	GT

$$\text{BSS} = \left( \frac{C0^2 + C1^2 + C2^2 + \dots}{r \cdot m} \right) - \text{CF}$$

$$A \times B \text{ table SS} = \left( \frac{T_0^2 + T_1^2 + T_3^2 + \dots}{r} \right) - CF$$

ABSS= A x B Table SS – ASS- ABSS

Error (b) SS= Table SS-ASS-BSS-ABSS –Error (a) SS.

Then complete the ANOVA table.

### Questions

1. To accommodate factors which require different sizes of experimental plots in the same experiment \_\_\_\_\_ design has been evolved

- a) Split plot      b) CRD      c) RBD      d) LSD

**Ans: Split plot**

2. The number of error terms in a split plot design is

- a) One      b) two      c) three      d) none of these

**Ans: two**

3. The plot size for the subplot treatment will be small when compared to the plot size of the main plot treatments.

**Ans: True**

4. The precision for the sub plot treatments is more when compared to the main plot treatments.

**Ans: True**

5. The plot sizes for the main plot treatment and subplot treatment are not same.

**Ans: True**

6. The plot size for subplot and interaction are same in split plot design.

**Ans: True**

7. When will you adopt split plot design?

8. What is error (a) in a split plot design?

9. Furnish the skeleton ANOVA table with 3 replications, 4 main plot treatments and 3 subplot treatments.

10. Furnish the layout of a split plot design.



## Lecture.22

### Strip plot design – layout – ANOVA Table

#### Strip Plot Design

This design is also known as **split block design**. When there are two factors in an experiment and both the factors require large plot sizes it is difficult to carry out the experiment in split plot design. Also the precision for measuring the interaction effect between the two factors is higher than that for measuring the main effect of either one of the two factors. Strip plot design is suitable for such experiments.

In strip plot design each block or replication is divided into number of vertical and horizontal strips depending on the levels of the respective factors.

Replication 1				Replication 2					
	a <sub>0</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>1</sub>		a <sub>3</sub>	a <sub>0</sub>	a <sub>2</sub>	a <sub>1</sub>
b <sub>1</sub>					b <sub>1</sub>				
b <sub>0</sub>					b <sub>2</sub>				
b <sub>2</sub>					b <sub>0</sub>				

In this design there are plot sizes.

1. Vertical strip plot for the first factor – vertical factor
2. Horizontal strip plot for the second factor – horizontal factor
3. Interaction plot for the interaction between 2 factors

The vertical strip and the horizontal strip are always perpendicular to each other. The interaction plot is the smallest and provides information on the interaction of the 2 factors. Thus we say that interaction is tested with more precision in strip plot design.

#### Analysis

The analysis is carried out in 3 parts.

1. Vertical strip analysis
2. Horizontal strip analysis

### 3. Interaction analysis

Suppose that A and B are the vertical and horizontal strips respectively. The following two way tables, viz., A X Rep table, B X Rep table and A X B table are formed. From A X Rep table, SS for Rep, A and Error (a) are computed. From B X Rep table, SS for B and Error (b) are computed. From A X B table, A X B SS is calculated.

When there are r replications, a levels for factor A and b levels for factor B, then the ANOVA table is

<b>X</b>	<b>d.f.</b>	<b>SS</b>	<b>MS</b>	<b>F</b>
Replication	(r-1)	RSS	RMS	RMS/EMS (a)
A	(a-1)	ASS	AMS	AMS/EMS (a)
Error (a)	(r-1) (a-1)	ESS (a)	EMS (a)	
B	(b-1)	BSS	BMS	BMS/EMS (b)
Error (b)	(r-1) (b-1)	ESS (b)	EMS (b)	
AB	(a-1) (b-1)	ABSS	ABMS	ABMS/EMS (c)
Error (c)	(r-1) (a-1) (b-1)	E SS (c)	EMS (c)	
Total		(rab - 1)	TSS	

### Analysis

Arrange the results as follows:

<b>Treatment Combination</b>	<b>Replication</b>				<b>Total</b>
	<b>R1</b>	<b>R2</b>	<b>R3</b>	<b>...</b>	
A0B0	a0b0	a0b0	a0b0	...	T00
A0B1	a0b1	a0b1	a0b1	...	T01
A0B2	a0b2	a0b2	a0b2	...	T02
Sub Total	A01	A02	A03	...	T0
A1B0	a1b0	a1b0	a1b0	...	T10
A1B1	a1b1	a1b1	a1b1	...	T11
A1B2	a1b2	a1b2	a1b2	...	T12
Sub Total	A11	A12	A13	...	T1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Total	R1	R2	R3	...	G.T

$$\text{Compute CF} = \frac{(G T)^2}{r \times m \times s}$$

TSS = [

(a0b0)<sup>2</sup> +

Treatment	Replication				Total
	R1	R2	R3	...	
B0	B01	B02	B03	...	T0
B1	B11	B12	B13	...	T1
B2	B21	B22	B23	...	T2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Total	R1	R2	R3	...	GT

(a0b1)<sup>2</sup>+(a0b2)<sup>2</sup>+...]-CF

1) Vertical Strip Analysis

Form A x R Table and calculate RSS, ASS and Error(a) SS

Treatment	Replication				Total
	R1	R2	R3	...	
A0	A01	A02	A03	...	T0
A1	A11	A12	A13	...	T1
A2	A21	A22	A23	...	T2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Total	R1	R2	R3	...	GT

$$RSS = \left( \frac{R1^2 + R2^2 + R3^2 + \dots}{m \cdot s} \right) - CF$$

$$ASS = \left( \frac{T0^2 + T1^2 + T2^2 + \dots}{r \cdot s} \right) - CF$$

$$A \times R \text{ table SS} = \left( \frac{A01^2 + A02^2 + A03^2 + \dots}{b} \right) - CF$$

Error (a) SS= A x R TSS-RASS-ASS.

2) Horizontal Strip Analysis

Form B x R Table and calculate RSS, BSS and Error(b) SS

$$3) \text{ BSS} = \left( \frac{T0^2 + T1^2 + T3^2 + \dots}{r \cdot s} \right) - CF$$

$$4) \text{ B x R table SS} = \left( \frac{B01^2 + B02^2 + B03^2 + \dots}{a} \right) - CF$$

$$5) \text{ Error (b) SS} = \text{B x R TSS} - \text{RSS} - \text{BSS}$$

3) Interaction Analysis

Form A x B Table and calculate BSS, Ax B SSS and Error (b) SS

Treatment	Replication				Total
	B0	B1	B2	...	
A0	T00	T01	T02	...	T0
A1	T10	T11	T12	...	T1
A2	T20	T21	T22	...	T2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Total	C0	C1	C2	...	GT

$$\text{A x B table SS} = \left( \frac{T0o^2 + T01^2 + T03^2 + \dots}{r} \right) - CF$$

$$\text{ABSS} = \text{A x B Table SS} - \text{ASS} - \text{ABSS}$$

$$\text{Error (c) SS} = \text{TSS} - \text{ASS} - \text{BSS} - \text{ABSS} - \text{Error (a) SS} - \text{Error (a) SS}$$

Then complete the ANOVA table.

### Questions

1. To accommodate factors which require different sizes of experimental plots in the same experiment \_\_\_\_\_ design has been evolved

- a) Strip plot      b) CRD      c) RBD      d) LSD

**Ans: Strip plot**

2. The number of error terms in a strip plot design is

- a) One      b) two      c) three      d) none of these

**Ans: three**

4. The plot size for the treatments allotted in vertical strips will not be equal when compared to the treatments allotted in horizontal strips.

**Ans: True**

5. The degrees of freedom for Error (b) in a strip plot design is  $(r-1)(a-1)$ .

**Ans: False**

6. The analysis of a strip plot design is carried out in three stages, viz, horizontal strip analysis, vertical strip analysis and interaction analysis.

**Ans: True**

7. When will you adopt strip plot design?

8. What is error (c) in a strip plot design?

9. Furnish the skeleton ANOVA table with 3 replications, 3 treatments in horizontal strip and 3 treatments in vertical strip.

10. Furnish the layout of a strip plot design.

## **Lecture.23**

### **Long term experiments – ANOVA table – guard rows – optimum plot size – determination methods.**

#### **Long Term Experiments**

A long term experiment is an experimental procedure that runs through a long period of time, in order to test a hypothesis or observe a phenomenon that takes place at an extremely slow rate. Several agricultural field experiments have run for more than 100 years. Experiments that are conducted at several sites or repeated over different seasons can also be classified as long term experiments. Performance of crops varies considerably from location to location as well as season to season. This is because of the influence of environmental factors such as rainfall, temperature etc. In order to determine the effects, the experiments have to be repeated at different locations and seasons. With such repetition of experiments practical recommendations may be made with greater confidence especially with new crop varieties or new techniques are introduced. Here we discuss the experiments that are conducted over different locations or different seasons.

#### **Layout of experiment**

Once the locations or seasons are decided upon the next step is to select the appropriate design of experiment. The individual experiments may be designed as CRD, RBD, split plot etc. The same design is adopted for all the locations or seasons. However randomization of treatments should be done afresh for each experiment.

#### **Analysis**

The results of repeated experiments are analysed using combined analysis of variance method.

The combined analysis is aimed at

1. to test whether there are significant differences between the treatments at various environments or loc or seasons etc.

2. test the consistency of the treatment at different environments. i.e. to test the presence or absence of interaction of the treatment with environments.

The presence of interaction will indicate that the responses change with environment.

In the first stage of the combined analysis the results of the individual locations are analysed based on the basic experimental design tried. In the second stage of the analysis various SS are computed by combining all the data.

**If the basic design adopted is RBD with  $t$  treatments and  $r$  replications and  $p$  locations the ANOVA table will be**

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F-ratio
Replication within locations	$p(r-1)$	RSS	RMS	
Locations	$p-1$	LSS	LMS	
Treatments	$t-1$	TrSS	TrMS	$\frac{\text{TrMS}}{\text{LXTMS}}$
Location x Treatments	$(p-1)(t-1)$	LXTSS	LXTMS	$\frac{\text{LXTMS}}{\text{EMS}}$
Combined error	$p(r-1)(t-1)$	ESS	EMS	
Total	$rtp-1$	TSS		

But before proceeding with the combined analysis it is necessary to test whether the EMS of the individual experiments are homogenous and the heterogeneity of EMS can be tested by either Bartlett's test or Hartley's test.

When the EMS are homogenous the analysis is done as follows:

Rep within location SS = Sum of replication SS of all locations

Pooled error SS = sum of error SS of all locations

The treatment X location two-way table is formed. From this two way table treatment SS, locations SS and treatment X location SS are computed.

The significance of treatment X location interaction is tested and if it is found to be significant then the interaction mean square is used for calculating the F value for treatments.

### **Optimum plot size**

Size and shape of experimental units will affect the accuracy of the experimental units. Select a plot with optimum plot size for this purpose. **Minimum size of experimental plot for a given degree of precision is known as optimum plot size.** Optimum plot size depends on crop, available land area, number of treatments etc.

To determine the optimum plot size two methods are available. They are (1) Maximum curvature method and (2) Fairfield Smith's variance law. For determining the optimum plot size in either method data are to be collected by conducting an Uniformity trial.

An uniformity trial is a trial conducted over an experimental material by selecting a particular variety of a crop and for the entire experimental unit uniform treatments are given. At harvest, the experimental unit is divided into small basic units (depending on the crop) and yield recorded. Then to find the optimum plot size, the basic units are combined by adding the basic units in rows or columns. But while combining rows or columns no row or column should be left out. Then for the new units formed we calculate coefficient of variation and based on the CV values the optimum plot size is determined.

### **Questions**

1. A long term experiment is an experiment conducted in  
a) one season    b) more than one season    c) more than one year    d) **both b and c**

**Ans: both b and c**

2. The homogeneity of the error variances of the individual seasons or locations is tested by

a) t test    b) F test    c) Bartlett's test    d) none of these

**Ans: Bartlett's test**



3. The significant interaction indicated that the responses change with the environment.

**Ans: True**

4. Minimum size of experimental plot for a given degree of precision is known as optimum plot size.

**Ans: True**

5. The designs adopted in two seasons for the same experiment need not be the same design.

**Ans: False**

6. The combined analysis is used to test whether there are significant differences between the treatments at various environments or loc or seasons etc.

**Ans: True**

7. What is a uniformity trail?

8. Mention the methods of determining the optimum plot size.

9. How to determine the optimum plot size?

10. Furnish the ANOVA table of an experiment conducted in RBD in  $s$  seasons.

## Lecture.24

### Introduction to computers – components of CPU – A/L unit – memory unit and control unit and their functions – application of computer in Agricultural Research

#### Introduction to Computers

A **computer** is an electronic device that manipulates information or "data." It has the ability to **store**, **retrieve**, and **process** data.

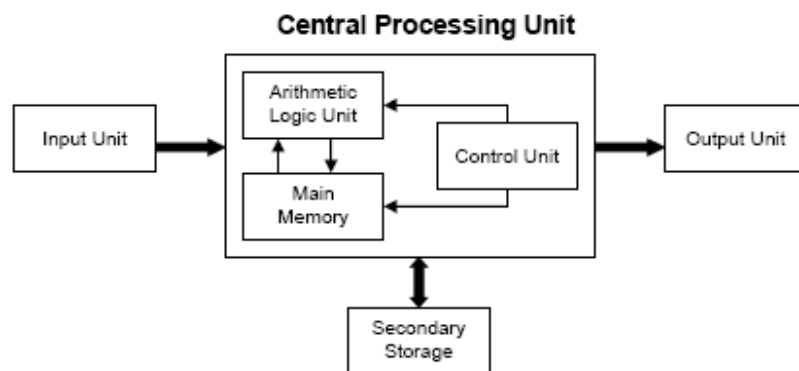
All types of computers consist of two basic parts – **hardware** and **software**.

**Hardware** is any part of your computer that has a **physical structure**, such as the computer monitor or keyboard.

**Software** is any **set of instructions** that tells the hardware what to do. It is what guides the hardware and tells it how to accomplish each task.

#### Components of CPU

The computer system essentially comprises three important parts – input device, central processing unit (CPU) and the output device. The CPU itself is made up of many components. Among the various components the three important components namely, the arithmetic logic unit (ALU), memory unit, and the control unit are important. In addition to these, auxiliary storage/secondary storage devices are used to store data and instructions on a long-term basis.

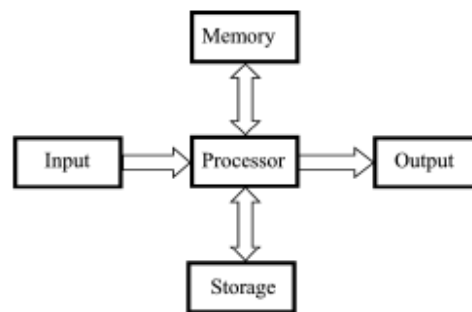


Schematic Representation of a Computer

## Arithmetic logic unit

An **arithmetic logic unit (ALU)** is a digital circuit that performs arithmetic and logical operations. The ALU is a fundamental building block of the central processing unit (CPU) of a computer, and even the simplest microprocessors contain one for purposes such as maintaining timers.

The Memory unit is linked with other parts of the computer



Linking Memory with other units

Computer's memory can be classified into two types – RAM and ROM. RAM or Random Access Memory is the central storage unit in a computer system. It is the place in a computer where the operating system, application programs and the data in current use are kept temporarily so that they can be accessed by the computer's processor. The more RAM a computer has, the more data a computer can manipulate. This memory is also known as user memory.

Random access memory, also called the Read/Write memory, is the temporary memory of a computer. It is said to be 'volatile' since its contents are accessible only as long as the computer is on. The contents of RAM are cleared once the computer is turned off.

ROM or Read Only Memory is a special type of memory which can only be read and contents of which are not lost even when the computer is switched off. It typically contains manufacturer's instructions. Among other things, ROM also stores an initial program called the 'bootstrap loader' whose function is to start the computer software operating, once the power is turned on.

Read-only memories can be manufacturer-programmed or user-programmed. While manufacturer-programmed ROMs have data burnt into the circuitry, user programmed ROMs can have the user load and then store read-only programs. PROM or Programmable ROM is the name given to such ROMs.

*Input devices* allows the user to enter the program and data and send it to the processing unit. The common input devices are keyboard, mouse and scanners.

*Output devices* show the processed data – information – the result of processing. The devices are normally a monitor and printers.

### **Commonly available hardware and their functions**

#### **The monitor**

The monitor is the video display we will be looking at, most of the time, to evaluate our work, find out whether the assignments are being carried out satisfactorily. A monitor is largely controlled by some pieces of hardware inside of the computer. But the monitor itself is mainly used to display our work in a graphical setting we can easily interpret. To display what is going on with the computer, the monitor is connected to the computer using a cable. The connection is usually done from the back of both machines. Monitors come in different sizes. The size of the monitor is measured diagonally on the screen and is given in inches. Based on this, monitors range in sizes of 14", 15", 17", 19", 21", 24", or 29", etc. Monitors are also characterized by the flatness of their screen. The flatter and the wider screens are usually the better. Whatever may be the size of the monitor the number of lines of text displayed will be same in all.

#### **The Keyboard**

A computer keyboard is a wide object that is equipped with buttons on which there are letters and numbers. To provide a better management, the keys on a keyboard are divided in sections. This arrangement is by convention so the users would be familiar with them and be able to use any keyboard they come in contact with.

## **The mouse**

A mouse is an object that is meant to fit the proportions of a hand and is positioned on the table so the user can move it easily. Like the other parts, a mouse is connected to the computer, usually to the back, by a cable. Nowadays, it is not unusual to have a wireless mouse so that it doesn't need a cable. The mouse is one of several pieces of hardware you will be using when interacting with the computer. A mouse is primarily made of three parts: the buttons, the handling area, and the rolling object. By default, a mouse has two buttons: left and right. Most mice nowadays are also equipped with a wheel on top known as scroll mouse.

## **Printer**

A printer is an output device that produces text and graphics on paper.

### **Types of Printer**

Printers can be divided into two main groups, **impact printer** and **non-impact printer**. Impact printer produces text and images when tiny wire pins on print head strike the ink ribbon by physically contacting the paper. Non-impact printer produces text and graphics on paper without actually striking the paper. Dot matrix is an example of an impact printer. Laser and inkjet are the examples of non-impact printers

## **Software**

A set of instructions that tells the computer what to do.

### **Two types of software**

#### **Systems software**

*System software* is computer *software* designed to operate the computer hardware and to provide and maintain a platform for running application *software*.

#### **Applications Software**

Application software is computer software designed to help the user perform a particular task. Such programs are also called software applications, applications or apps.

Typical examples are word processors, spreadsheets, media players and database applications.

## **Operating system**

An **operating system (OS)** is an interface between hardware and user, which is responsible for the management and coordination of activities and the sharing of the resources of a computer, that acts as a host for computing applications run on the machine. One of the purposes of an operating system is to handle the resource allocation and access protection of the hardware.

## **Ms-dos**

MS-DOS is the most well known operating system. MS-DOS was created in 1981 when it was used on an IBM PC. DOS, as with any operating system, controls computer activity. It manages operations such as data flow, display, data entry amongst other various elements that make up a system.

The role of DOS is to interpret commands that the user enters via the keyboard. These commands allow the following tasks to be executed:

- file and folder management
- disk upgrades
- hardware configuration
- memory optimisation
- program execution

## **Windows Os**

- **Windows** is the operating system sold by the Seattle-based company Microsoft. Microsoft, originally christened "Traf-O-Data" in 1972, was renamed "Microsoft" in November 1975, then "Microsoft" on November 26, 1976.

- Microsoft entered the marketplace in August 1981 by releasing version 1.0 of the operating system Microsoft DOS (MS-DOS), a 16-bit command-line operating system
- The first version of Microsoft Windows (**Microsoft Windows 1.0**) came out in November 1985. It had a graphical user interface, inspired by the user interface of the Apple computers of the time. Windows 1.0 was not successful with the public, and Microsoft Windows 2.0, launched December 9, 1987, did not do much better.
- It was on May 22, 1990 that Microsoft Windows became a success, with **Windows 3.0**, then **Windows 3.1** in 1992, and finally **Microsoft Windows for Workgroups**, later renamed **Windows 3.11**, which included network capabilities. Windows 3.1 cannot be considered an entirely separate operating system because it was only a graphical user interface running on top of MS-DOS.
- On August 24, 1995, Microsoft launched the operating system **Microsoft Windows 95**.
- Windows 98 natively supported features other than those of MS-DOS but was still based upon it.
- On September 14, 2000, Microsoft released **Windows Me** (for *Millennium Edition*), also called **Windows Millennium**. Windows Millennium was based largely on Windows 98 (and therefore on MS-DOS), but added additional multimedia and software capabilities.
- On May 24, 1993, the first version of Windows NT was released. It was called **Windows NT 3.1**, and was followed by **Windows NT 3.5** in September 1994 and **Windows 3.51** in June 1995. With **Windows NT 4.0**, launched for sale on August 24, 1996, Windows NT finally became a true success.
- In July 1998, Microsoft released **Windows NT 4.0 TSE** (*Terminal Server Emulation*), the first Windows system that allowed terminals to be plugged into a server, i.e. use thin clients to open a session on the server.
- On February 17, 2000, the next version of NT 4.0 was renamed **Windows 2000** (instead of *Windows NT 5.0*) in order to highlight the unification of "NT" with the "Windows 9x" systems. Windows 2000 is an entirely 32-bit system with

characteristics of Windows NT, as well as an improved task manager and full compatibility with USB and FireWire peripherals.

- Then, on October 25, 2001, **Windows XP** arrived on the scene. This was a merger of the preceding operating systems.
- Finally, on April 24, 2003, a server operating system was released by Microsoft: **Windows Server 2003**.

## **Linux Operating System**

Linux is a free open-source operating system based on UNIX. Linux was originally created by Linus Torvalds with the assistance of developers from around the globe. Linux is free to download, edit and distribute. Linux is a very powerful operating system and it is gradually becoming popular throughout the world.

### **Advantages of Linux**

#### **Low cost**

There is no need to spend time and huge amount money to obtain licenses since Linux and much of it's software come with the GNU General Public License. There is no need to worry about any software's that you use in Linux.

#### **Stability**

Linux has high stability compared with other operating systems. There is no need to reboot the Linux system to maintain performance levels. Rarely it freeze up or slow down. It has a continuous up-times of hundreds of days or more.

#### **Performance**

Linux provides high performance on various networks. It has the ability to handle large numbers of users simultaneously.

#### **Networking**

Linux provides a strong support for network functionality; client and server



systems can be easily set up on any computer running Linux. It can perform tasks like network backup more faster than other operating systems.

### **Flexibility**

Linux is very flexible. Linux can be used for high performance server applications, desktop applications, and embedded systems. You can install only the needed components for a particular use. You can also restrict the use of specific computers.

### **Compatibility**

It runs all common Unix software packages and can process all common file formats.

### **WiderChoice**

There is a large number of Linux distributions which gives you a wider choice. Each organization develop and support different distribution. You can pick the one you like best; the core function's are the same.

### **Fast and easy installation**

Linux distributions come with user-friendly installation. Better use of hard disk: Linux uses its resources well enough even when the hard disk is almost full.

### **Multitasking**

Linux is a multitasking operating system. It can handle many things at the same time.

### **Security**

Linux is one of the most secure operating systems. File ownership and permissions make linux more secure.

### **Open source**

Linux is an Open source operating systems. You can easily get the source code for linux and edit it to develop your personal operating system.

## Questions

1. RAM is a temporary memory.

**Ans: True**

2. Windows allow opening more than one window at a time.

**Ans: True**

3. Windows was developed by

- a) Sun Microsystems
- b) Microsoft
- c) Intel
- d) None of the above

**Ans: Microsoft**

4. CD can be used to

- a) Read
- b) Write
- c) Read and write
- d) None of the above

**Ans: read and write**

5. Recycle bin stores the deleted files.

**Ans: True**

6. Monitor is an input device.

**Ans: True**

7. GUI means Graphical User Interface.

**Ans: True**

8. LCD is an input device.

**Ans: True**

9. McAfee is an

- a) Anti virus
- b) An operating system
- c) Neuro
- d) Search engine

**Ans: anti virus**

10. Mozilla Firefox is a \_\_\_\_\_

- a) Search engine
- b) Browser
- c) Operating system
- d) Office software

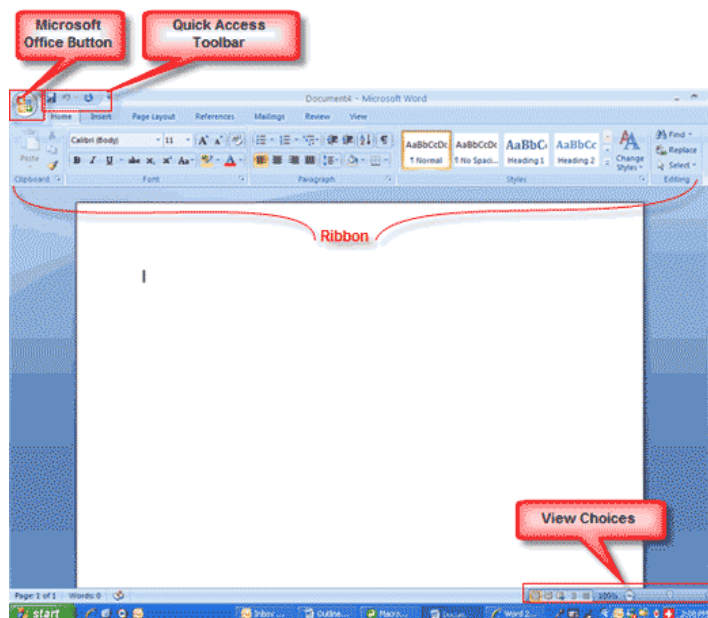
**Ans: browser**

## Lecture.25

**Hardware – commonly available hardware and their functions – software – types – system software and application software – an introduction to DOS, Windows and Linux OS**

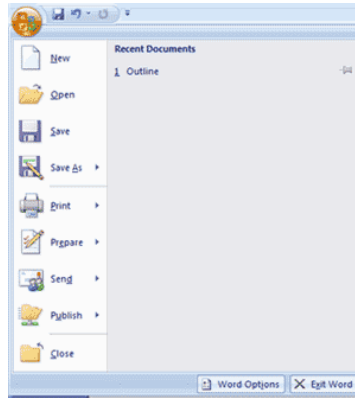
### Microsoft Word

#### Screen Layout



#### Menus

When we begin to explore Word 2007 you will notice a new look to the menu bar. There are three features that we should remember as we work within Word 2007: the Microsoft Office Button, the Quick Access Toolbar, and the Ribbon. These three features contain many of the functions that were in the menu of previous versions of Word. The functions of these three features will be more fully explored below.



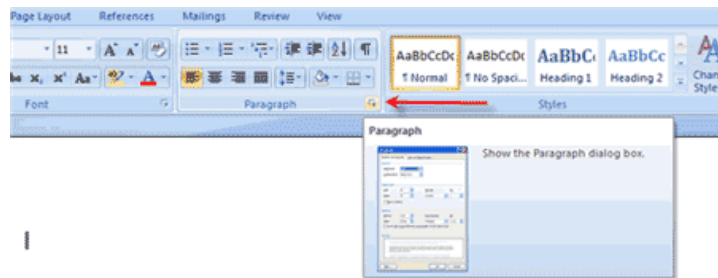
### **The Microsoft Office Button**

The Microsoft Office button performs many of the functions that were located in the File menu of older versions of Word. This button allows you to create a new document, open an existing document, save or save as, print, send (through email or fax), publish or close.

### **The Ribbon**



The Ribbon is the panel at the top portion of the document. It has seven tabs: Home, Insert, Page Layout, References, Mailings, Review, and View that contain many new and existing features of Word. Each tab is divided into groups. The groups are logical collections of features designed to perform functions that you will utilize in developing or editing your Word document. Commonly used features are displayed on the Ribbon, to view additional features within each group, click on the arrow at the bottom right of each group.



Each of the tabs contains the following tools:

**Home:** Clipboard, Fonts, Paragraph, Styles, and Editing.

**Insert:** Pages, Tables, Illustrations, Links, Header & Footer, Text, and Symbols

**Page Layout:** Themes, Page Setup, Page Background, Paragraph, Arrange

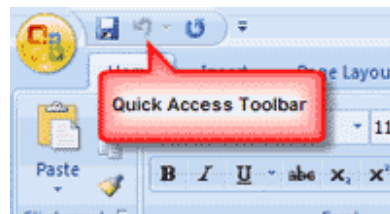
**References:** Table of Contents, Footnote, Citation & Bibliography, Captions, Index, and Table of Authorities

**Mailings:** Create, Start Mail Merge, Write & Insert Fields, Preview Results, Finish

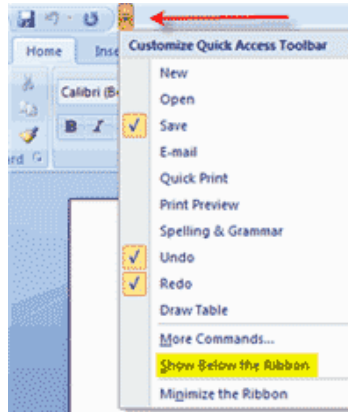
**Review:** Proofing, Comments, Tracking, Changes, Compare, Protect

**View:** Document Views, Show/Hide, Zoom, Window, Macros

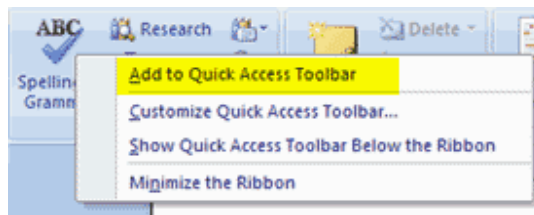
### Quick Access Toolbar



The quick access toolbar is a customizable toolbar that contains commands that you may want to use. You can place the quick access toolbar above or below the ribbon. To change the location of the quick access toolbar, click on the arrow at the end of the toolbar and click on Show Below the Ribbon.



You can also add items to the quick access toolbar. Right click on any item in the Office Button or the Ribbon and click on Add to Quick Access Toolbar and a shortcut will be added to the Quick Access Toolbar.



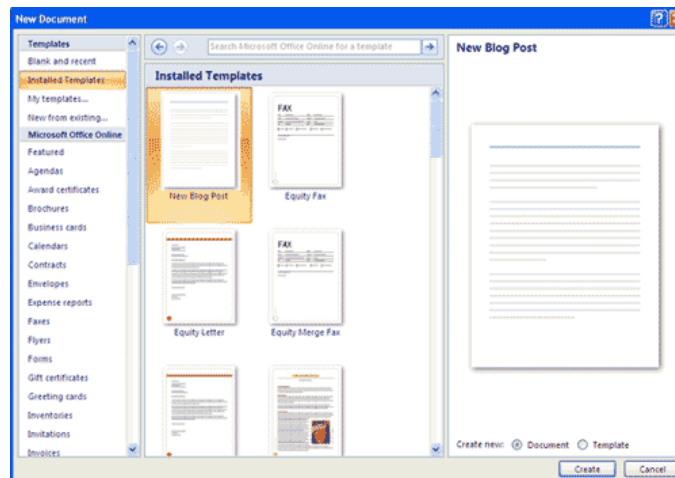
### Create a New Document

There are several ways to create new documents, open existing documents, and save documents in Word:



- Click the **Microsoft Office Button** and Click **New** or
- Press CTRL+N (Depress the CTRL key while pressing the “N”) on the keyboard

We will notice that when we click on the Microsoft Office Button and Click **New**, we have many choices about the types of documents we can create. If we wish to start from a blank document, click **Blank**. If we wish to start from a template we can browse through our choices on the left, see the choices on center screen, and preview the selection on the right screen.



## Opening an Existing Document

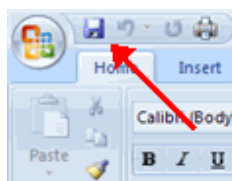


- Click the **Microsoft Office Button** and Click **Open**, or
- Press CTRL+O (Depress the CTRL key while pressing the “O”) on the keyboard, or
- If we have recently used the document you can click the **Microsoft Office Button** and click the name of the document in the **Recent Documents** section of the window Insert picture of recent docs

## Saving a Document



- Click the **Microsoft Office Button** and Click **Save** or **Save As** (remember, if we’re sending the document to someone who does not have Office 2007, you will need to click the **Office Button**, click **Save As**, and Click **Word 97-2003 Document**), or
- Press CTRL+S (Depress the CTRL key while pressing the “S”) on the keyboard, or
- Click the **File** icon on the Quick Access Toolbar



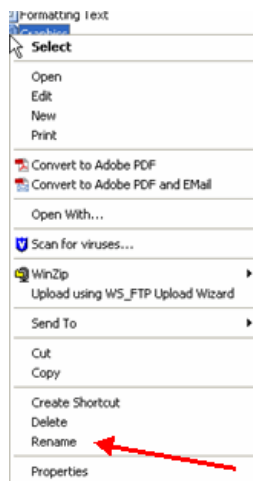


## Renaming Documents

To rename a Word document while using the program:

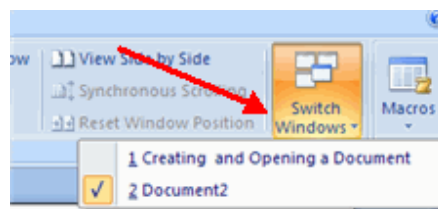


- Click the **Office Button** and find the file we want to rename.
- Right-click the document name with the mouse and select **Rename** from the shortcut menu.
- Type the new name for the file and press the **ENTER** key.



## Working on Multiple Documents

Several documents can be opened simultaneously if you are typing or editing multiple documents at once. All open documents will be listed in the **View Tab** of the Ribbon when you click on Switch Windows. The current document has a checkmark beside the file name. Select another open document to view it.



## Document Views

There are many ways to view a document in Word.

- **Print Layout:** This is a view of the document as it would appear when printed. It includes all tables, text, graphics, and images.

- **Full Screen Reading:** This is a full view length view of a document. Good for viewing two pages at a time.
- **Web Layout:** This is a view of the document as it would appear in a web browser.
- **Outline:** This is an outline form of the document in the form of bullets.
- **Draft:** This view does not display pictures or layouts, just text.

To view a document in different forms, click the document views shortcuts at the bottom of the screen or:



- Click the **View Tab** on the Ribbon
- Click on the appropriate document view.



### Close a Document

To close a document:

- Click the **Office Button**
- Click **Close**

### Editing a Document

### Formatting Text »

#### Typing and inserting Text

**To enter text, just start typing! The text will appear where the blinking cursor is located. Move the cursor by using the arrow buttons on the keyboard or positioning the mouse and clicking the left button. The keyboard shortcuts listed below are also helpful when moving through the text of a document:**

Move Action	Keystroke
Beginning of the line	<b>HOME</b>
End of the line	<b>END</b>
Top of the document	<b>CTRL+HOME</b>
End of the document	<b>CTRL+END</b>

### Selecting Text

To change any attributes of text it must be highlighted first. Select the text by dragging the mouse over the desired text while keeping the left mouse button depressed,

or hold down the **SHIFT** key on the keyboard while using the arrow buttons to highlight the text. The following table contains shortcuts for selecting a portion of the text:

<b>Selection</b>	<b>Technique</b>
Whole word	double-click within the word
Whole paragraph	triple-click within the paragraph
Several words or lines	drag the mouse over the words, or hold down <b>SHIFT</b> while using the arrow keys
Entire document	choose <b>Editing   Select   Select All</b> from the Ribbon, or press <b>CTRL+A</b>

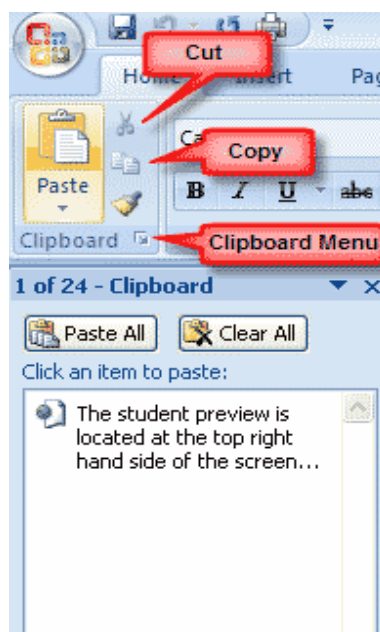
Deselect the text by clicking anywhere outside of the selection on the page or press an arrow key on the keyboard.

### **Inserting Additional Text**

Text can be inserted in a document at any point using any of the following methods:

- **Type Text:** Put your cursor where you want to add the text and begin typing
- **Copy and Paste Text:** Highlight the text you wish to copy and right click and click **Copy**, put your cursor where you want the text in the document and right click and click **Paste**.
- **Cut and Paste Text:** Highlight the text you wish to copy and right click and click **Cut**, put your cursor where you want the text in the document and right click and click **Paste**.
- **Drag Text:** Highlight the text you wish to move, click on it and drag it to the place where you want the text in the document.

You will notice that you can also use the Clipboard group on the Ribbon.

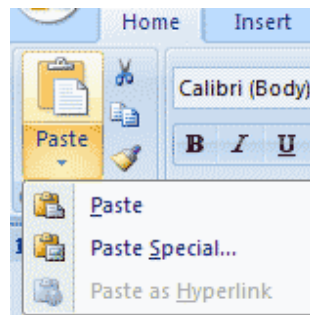


## Rearranging Blocks of Text

To rearrange text within a document, you can utilize the **Clipboard Group** on the **Home Tab** of the Ribbon.

Insert picture of clipboard group labeled

- **Move text:** Cut and Paste or Drag as shown above
- **Copy Text:** Copy and Paste as above or use the Clipboard group on the Ribbon
- **Paste Text:** Ctrl + V (hold down the CTRL and the “V” key at the same time) or use the Clipboard group to Paste, Paste Special, or Paste as Hyperlink



## Deleting Blocks of Text

Use the **BACKSPACE** and **DELETE** keys on the keyboard to delete text.

Backspace will delete text to the left of the cursor and Delete will erase text to the right.

To delete a large selection of text, highlight it using any of the methods outlined above and press the **DELETE** key.

## Search and Replace Text

To find a particular word or phrase in a document:

- Click **Find** on the **Editing Group** on the Ribbon
- To find and replace a word or phrase in the document, click **Replace** on the **Editing Group** of the Ribbon.



## Undo Changes

To undo changes:

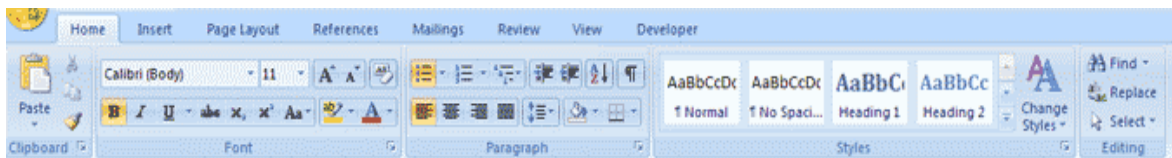
- Click the **Undo Button** on the Quick Access Toolbar



## Formatting Text

### Styles

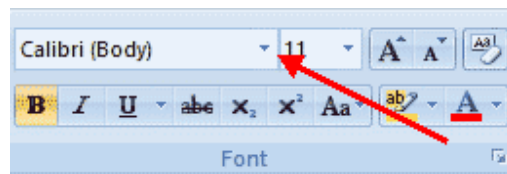
A style is a format enhancing tool that includes font typefaces, font size, effects (bold, italics, underline, etc.), colors and more. You will notice that on the Home Tab of the Ribbon, that you have several areas that will control the style of your document: Font, Paragraph, and Styles.



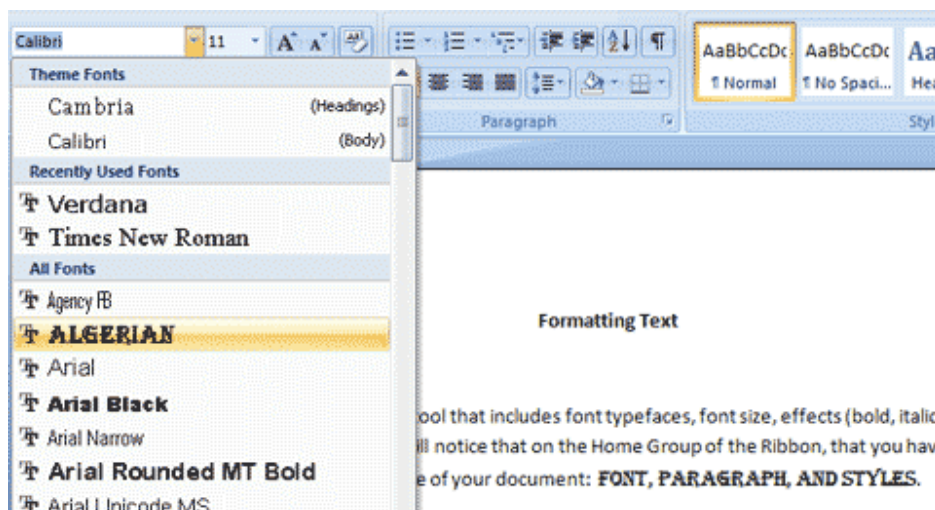
### Change Font Typeface and Size

To change the font typeface:

- Click the **arrow** next to the font name and choose a font.



- Remember that you can preview how the new font will look by highlighting the text, and hovering over the new font typeface.

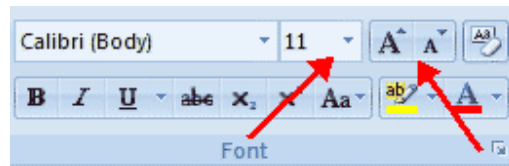


### Formatting Text

ool that includes font typefaces, font size, effects (bold, italics:  
 ill notice that on the Home Group of the Ribbon, that you have  
 e of your document: **FONT, PARAGRAPH, AND STYLES.**

### To change the font size

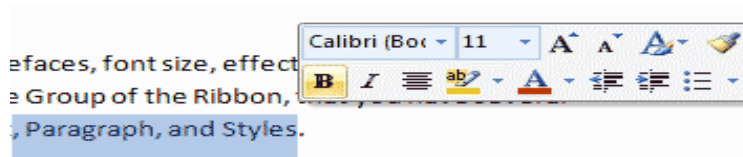
- Click the **arrow** next to the font size and choose the appropriate size, or
- Click the **increase** or **decrease** font size buttons.



### Font Styles and Effects

Font styles are predefined formatting options that are used to emphasize text. They include: Bold, Italic, and Underline. To add these to text:

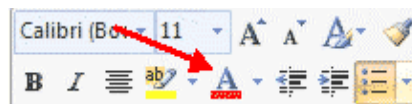
- Select the text and click the **Font Styles** included on the Font Group of the Ribbon, or
- Select the text and right click to display the font tools



### Change Text Color

To change the text color:

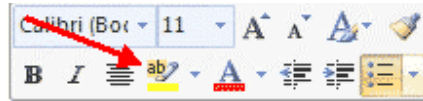
- Select the text and click the **Colors** button included on the Font Group of the Ribbon, or
- Highlight the text and right click and choose the colors tool.
- Select the color by clicking the down arrow next to the font color button.



### Highlight Text

Highlighting text allows you to use emphasize text as you would if you had a marker. To highlight text:

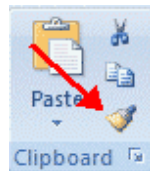
- Select the text
- Click the **Highlight Button** on the Font Group of the Ribbon, or
- Select the text and right click and select the highlight tool
- To change the color of the highlighter click on down arrow next to the highlight button.



## Copy Formatting

If you have already formatted text the way you want it and would like another portion of the document to have the same formatting, you can copy the formatting. To copy the formatting, do the following:

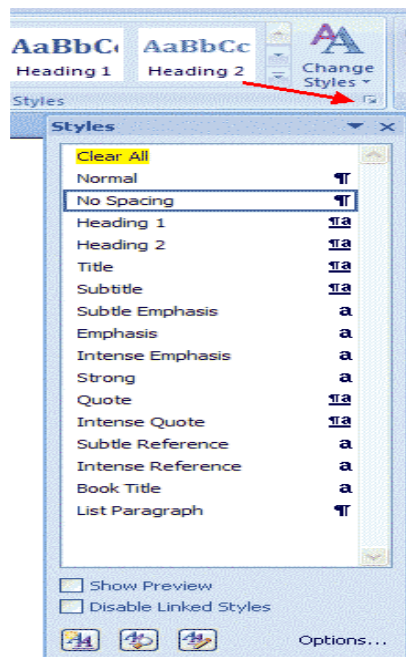
- Select the text with the formatting you want to copy.
- Copy the format of the text selected by clicking the **Format Painter** button on the Clipboard Group of the Home Tab
- Apply the copied format by selecting the text and clicking on it.



## Clear Formatting

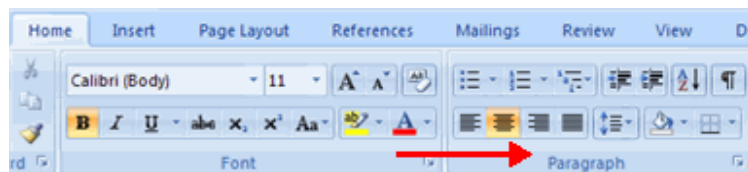
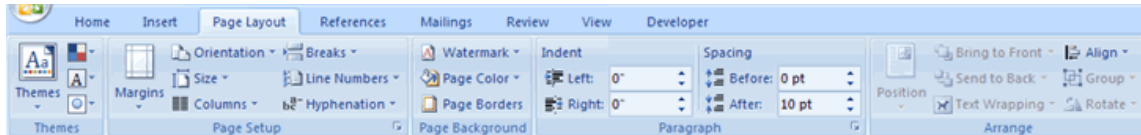
To clear text formatting:

- Select the text you wish to clear the formatting
- Click the **Styles** dialogue box on the Styles Group on the Home Tab
- Click **Clear All**



## Formatting Paragraph

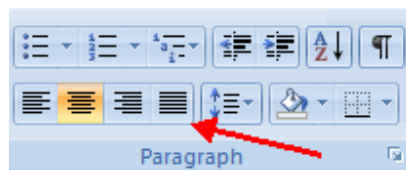
Formatting paragraphs allows you to change the look of the overall document. You can access many of the tools of paragraph formatting by clicking the **Page Layout** Tab of the Ribbon or the **Paragraph** Group on the Home Tab of the Ribbon.



## Change Paragraph Alignment

The paragraph alignment allows you to set how you want text to appear. To change the alignment:

- Click the **Home Tab**
- Choose the appropriate button for alignment on the Paragraph Group.
  - **Align Left:** the text is aligned with your left margin
  - **Center:** The text is centered within your margins
  - **Align Right:** Aligns text with the right margin
  - **Justify:** Aligns text to both the left and right margins.



## Indent Paragraphs

Indenting paragraphs allows you set text within a paragraph at different margins. There are several options for indenting:

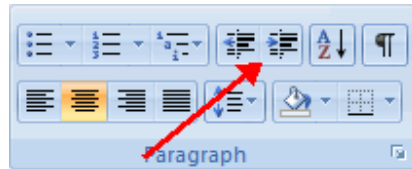
- **First Line:** Controls the left boundary for the first line of a paragraph
- **Hanging:** Controls the left boundary of every line in a paragraph except the first one
- **Left:** Controls the left boundary for every line in a paragraph
- **Right:** Controls the right boundary for every line in a paragraph

To indent paragraphs, you can do the following:

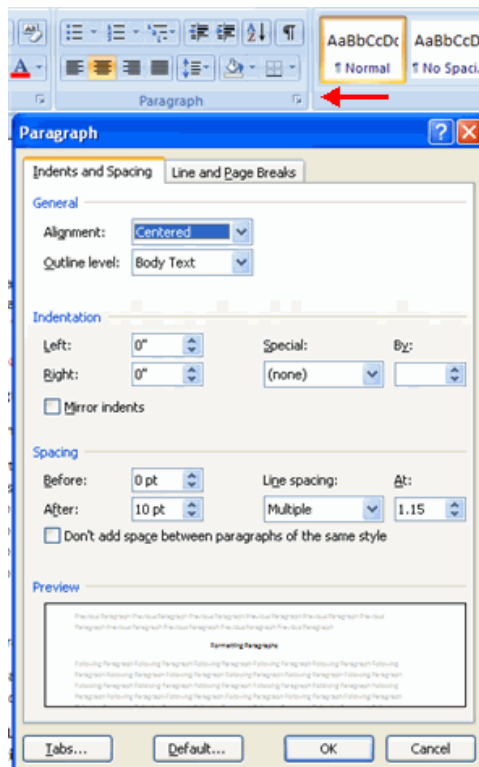
- Click the **Indent** buttons to control the indent.



- Click the **Indent** button repeated times to increase the size of the indent.



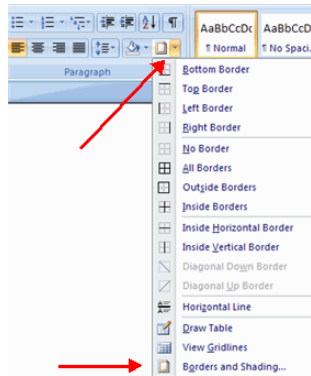
- Click the dialog box of the **Paragraph** Group
- Click the **Indents and Spacing** Tab
- Select your indents



### Add Borders and Shading

You can add borders and shading to paragraphs and entire pages. To create a border around a paragraph or paragraphs:

- Select the area of text where you want the border or shading.
- Click the **Borders** Button on the Paragraph Group on the Home Tab
- Choose the **Border and Shading**

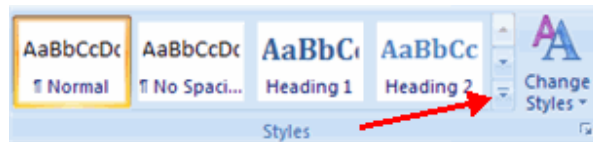


- Choose the appropriate options

### Apply Styles

Styles are a present collection of formatting that you can apply to text. To utilize **Quick Styles**:

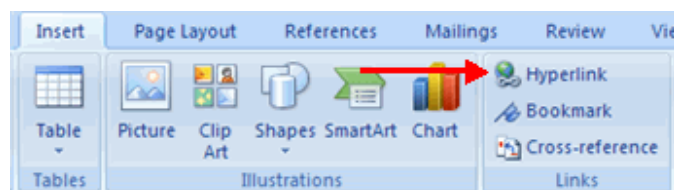
- Select the text you wish to format.
- Click the dialog box next to the **Styles Group** on the Home Tab.
- Click the style you wish to apply.

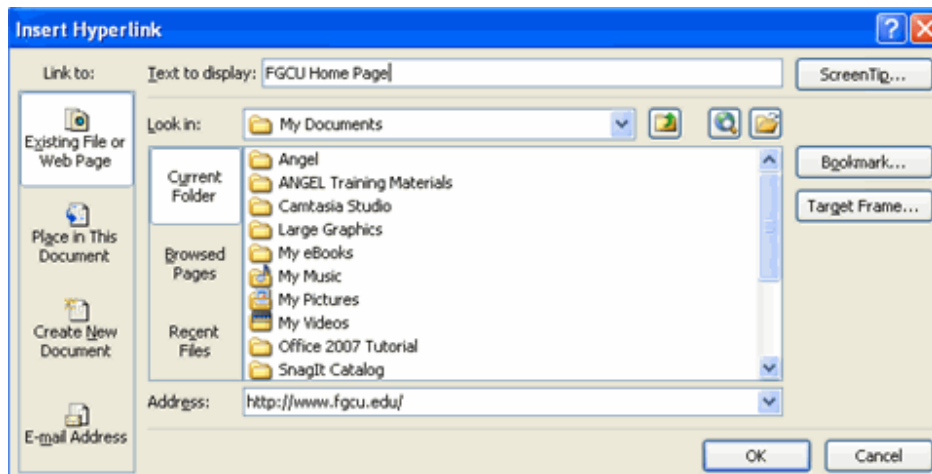


### Create Links

Creating links in a word document allows you to put in a URL that readers can click on to visit a web page. To insert a link:

- Click the **Hyperlink Button** on the Links Group of the Insert Tab.
- Type in the text in the “Text to Display” box and the web address in the “Address” box.

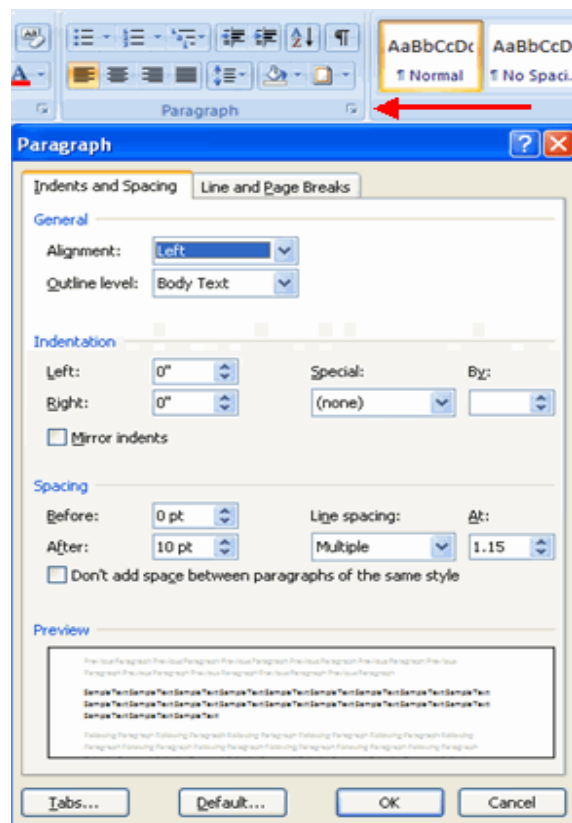




## Change Spacing between Paragraphs and Lines

You can change the space between lines and paragraphs by doing the following:

- Select the paragraph or paragraphs you wish to change.
- On the Home Tab, Click the **Paragraph** Dialog Box
- Click the **Indents and Spacing** Tab
- In the **Spacing** section, adjust your spacing accordingly



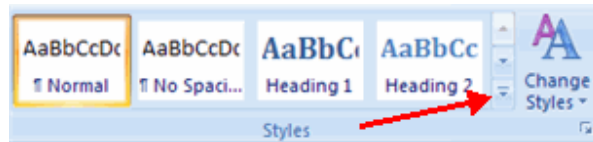
## STYLES

The use of Styles in Word will allow you to quickly format a document with a consistent and professional look. Styles can be saved for use in many documents.

### Apply Styles

There are many styles that are already in Word ready for you to use. To view the available styles click the **Styles** dialog box on the Styles Group in the Home Tab. To apply a style:

- Select the text
- Click the **Styles** Dialog Box
- Click the **Style** you choose



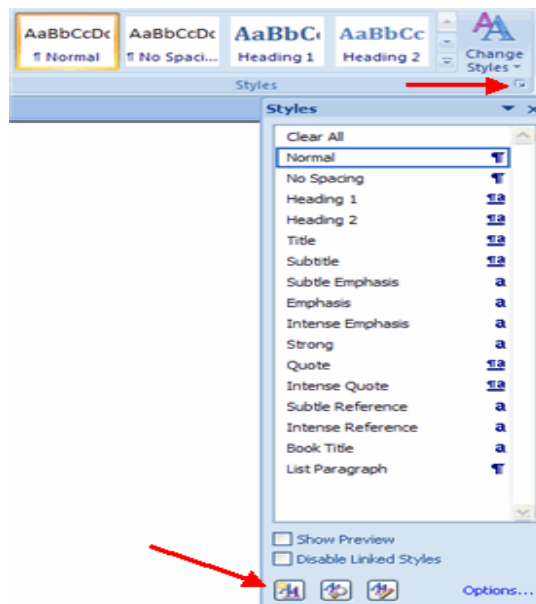
### Creating New Styles

You can create styles for formatting that you use regularly. There are two ways to do this: New Styles or New Quick Styles.

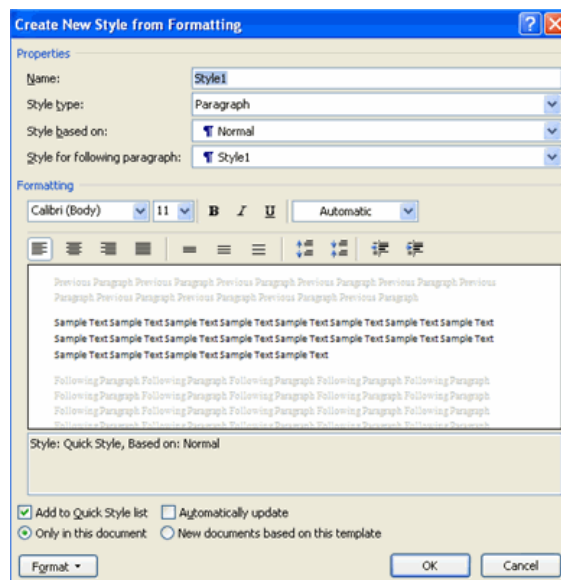
### New Styles

To create a new style:

- Click the **Styles** Dialog Box
- Click the **New Style** Button



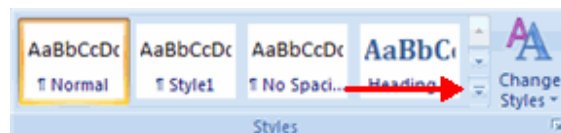
- Complete the **New Style** dialog box.
- At the bottom of that dialog box, you can choose to add this to the **Quick Style List** or to make it available only in this document.



## New Quick Style

To create a style easily:

- Insert your cursor anywhere in the chosen style
- Click the **Styles** dialog box



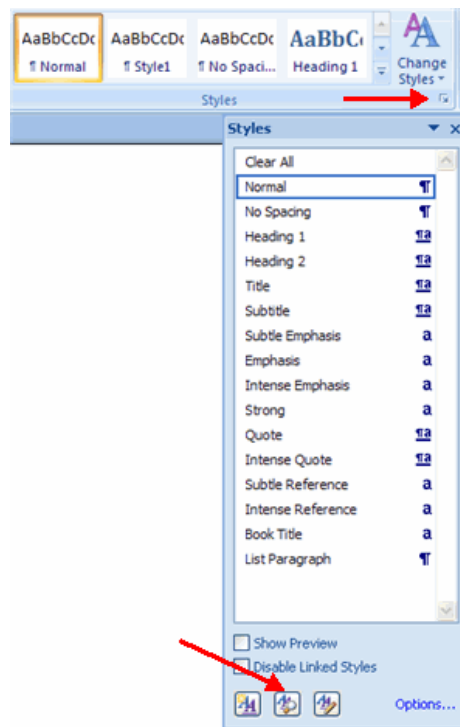
- Click **Save Selection as New Quick Style**



## Style Inspector

To determine the style of a particular section of a document:

- Insert cursor anywhere in the text that you want to explain the style
- Click the **Styles Drop Down Menu**
- Click the **Style Inspector** Button



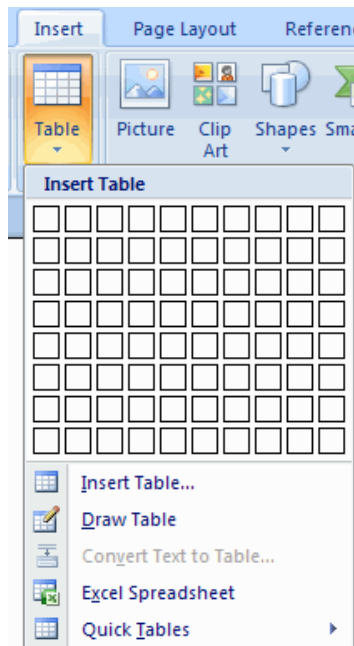
## TABLE

Tables are used to display data in a table format.

### Create a Table

To create a table:

- Place the cursor on the page where you want the new table
- Click the **Insert** Tab of the Ribbon
- Click the **Tables** Button on the Tables Group. You can create a table one of four ways:
  - Highlight the number of row and columns
  - Click **Insert Table** and enter the number of rows and columns
  - Click the **Draw Table**, create your table by clicking and entering the rows and columns
  - Click **Quick Tables** and choose a table



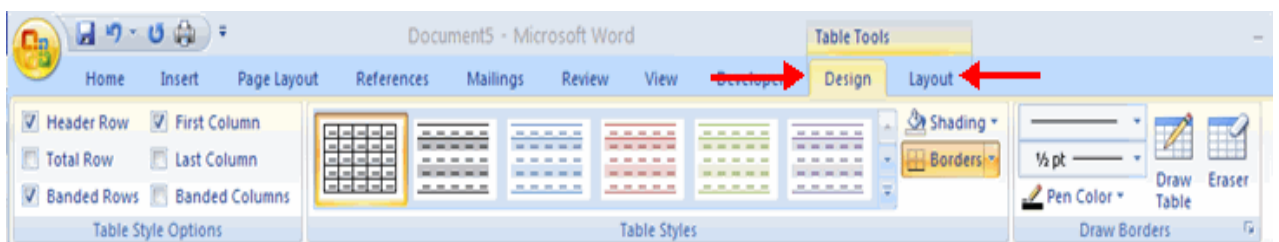
### Enter Data in a Table

Place the cursor in the cell where you wish to enter the information. Begin typing.

### Modify the Table Structure and Format a Table

To **modify** the structure of a table:

- Click the table and notice that you have two new tabs on the Ribbon: **Design** and **Layout**. These pertain to the table design and layout.



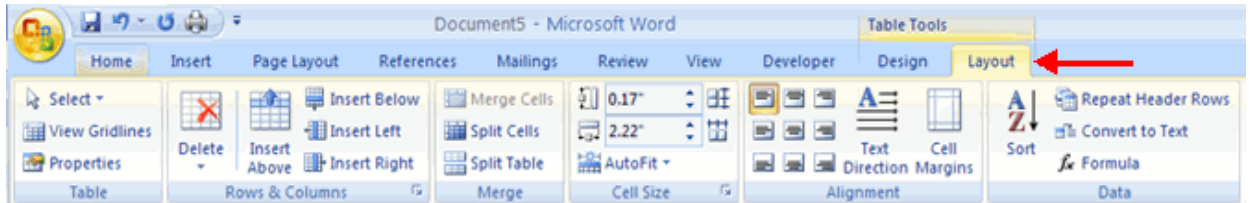
On the Design Tab, you can choose:

- **Table Style Options**
- **Table Styles**
- **Draw Borders**

To **format** a table, click the table and then click the **Layout** Tab on the Ribbon. This Layout tab allows you to:

- **View Gridlines and Properties** (from the Table Group)
- **Insert Rows and Columns** (from the Rows & Columns Group)

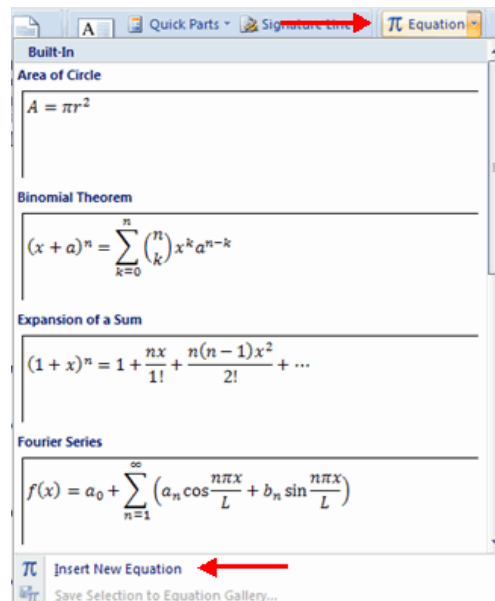
- **Delete the Table, Rows and/or Columns** (from the Rows & Columns Group)
- **Merge or Split Cells** (from the Merge Group)
- **Increase and Decrease cell size** (Cell Size Group)
- **Align text within the cells and change text directions** (Alignment Group)



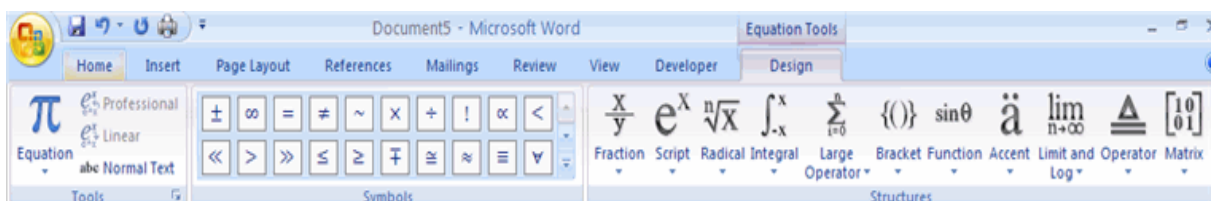
## Equations

**Word 2007 also allows you to insert mathematical equations. To access the mathematical equations tool:**

- Place your cursor in the document where you want the symbol
- Click the **Insert** Tab on the Ribbon
- Click the **Equation** Button on the Symbols Group
- Choose the appropriate equation and structure or click Insert New Equation



- To edit the equation click the equation and the **Design** Tab will be available in the Ribbon

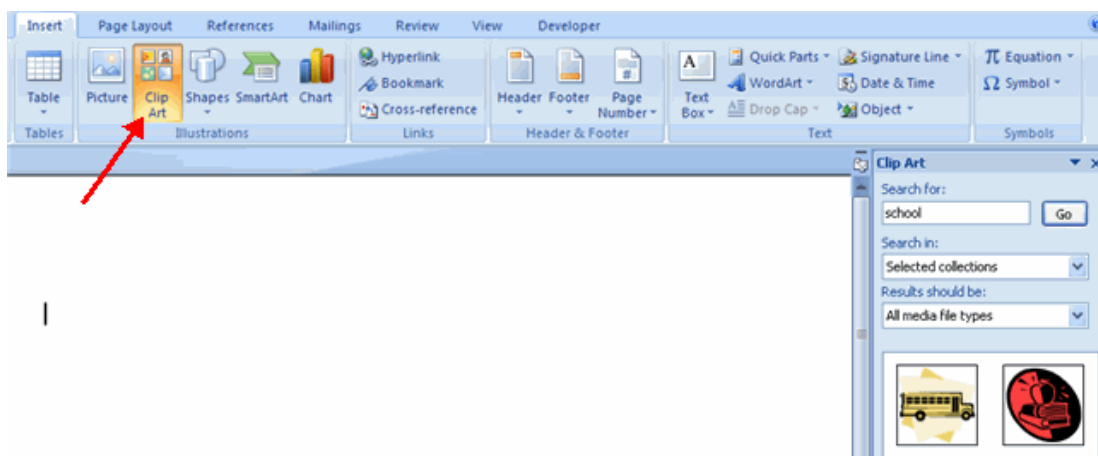




## Illustrations, Pictures, and SmartArt

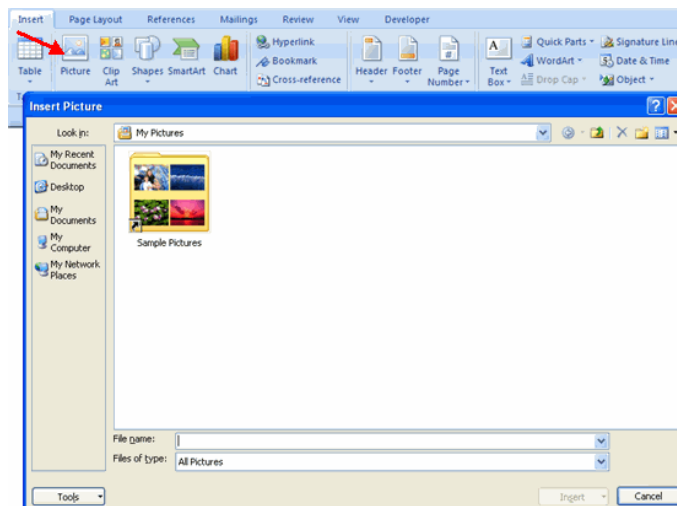
Word 2007 allows you to insert illustrations and pictures into a document. To insert **illustrations**:

- Place your cursor in the document where you want the illustration/picture
- Click the **Insert** Tab on the Ribbon
- Click the **Clip Art** Button
- The dialog box will open on the screen and you can search for clip art.
- Choose the illustration you wish to include



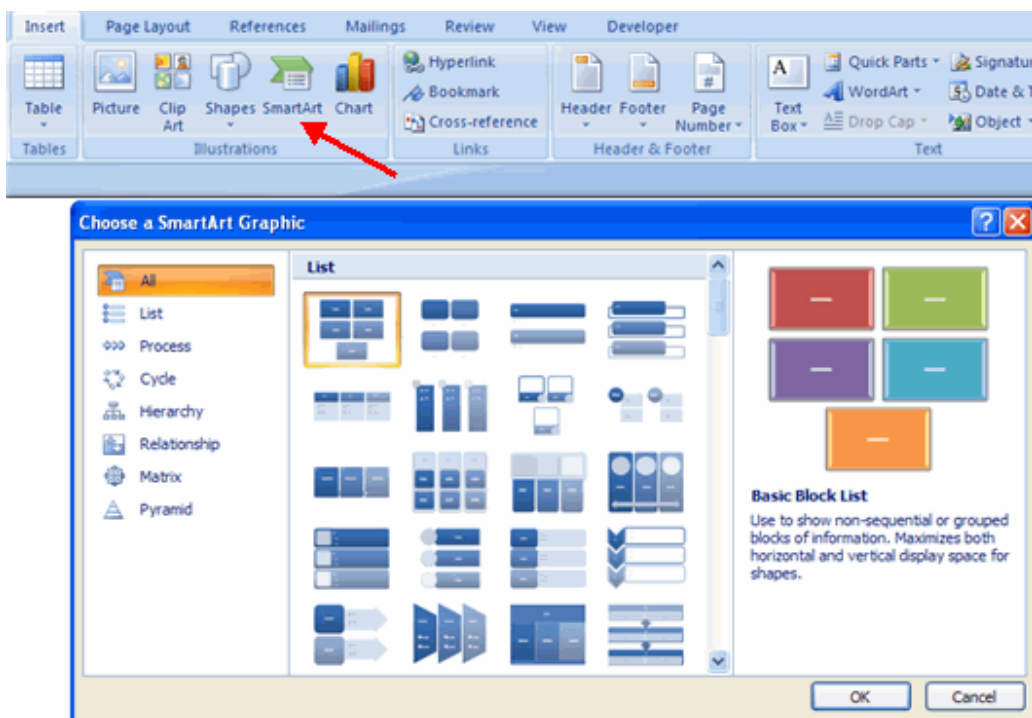
To insert a **picture**

- Place your cursor in the document where you want the illustration/picture
- Click the **Insert** Tab on the Ribbon
- Click the **Picture** Button
- Browse to the picture you wish to include
- Click the **Picture**
- Click **Insert**



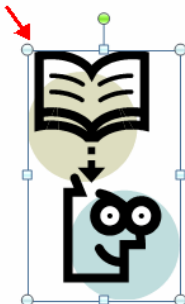
**Smart Art** is a collection of graphics you can utilize to organize information within your document. It includes timelines, processes, or workflow. To insert SmartArt

- Place your cursor in the document where you want the illustration/picture
- Click the **Insert** Tab on the Ribbon
- Click the **SmartArt** button
- Click the **SmartArt** you wish to include in your document
- Click the arrow on the left side of the graphic to insert text or type the text in the graphic.



### Resize Graphics

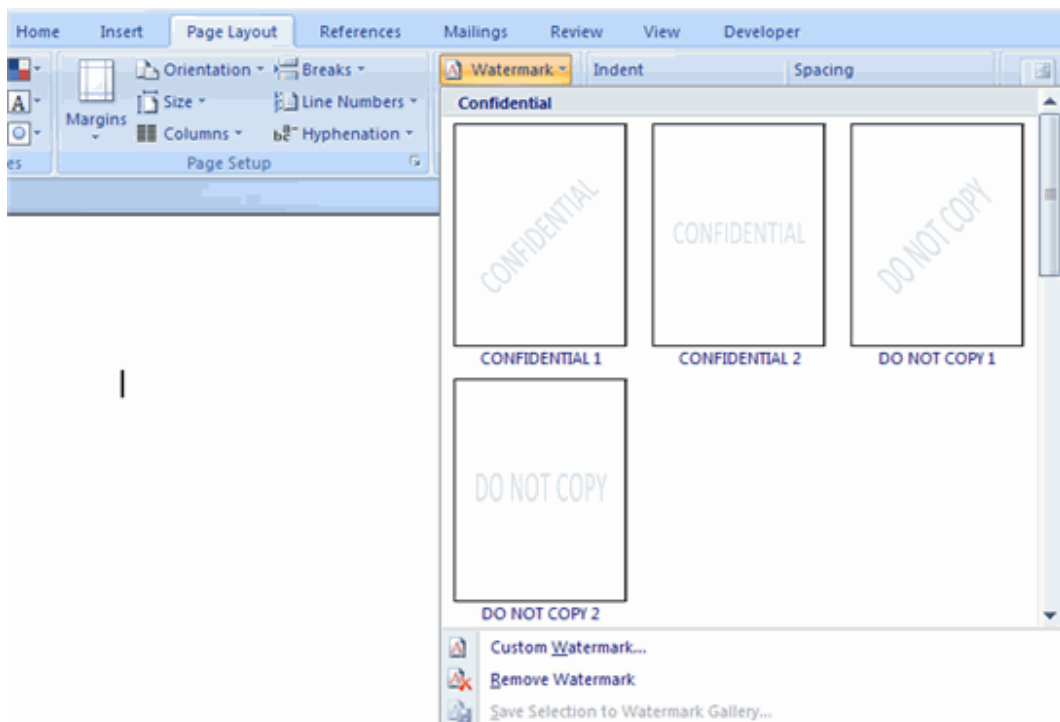
All graphics can be resized by clicking the image and clicking one corner of the image and dragging the cursor to the size you want the picture.



## Watermarks

A watermark is a translucent image that appears behind the primary text in a document. To insert a watermark:

- Click the **Page Layout** Tab in the Ribbon
- Click the **Watermark** Button in the Page Background Group
- Click the **Watermark** you want for the document or click **Custom Watermark** and create your own watermark.
- To remove a watermark, follow the steps above, but click **Remove Watermark**

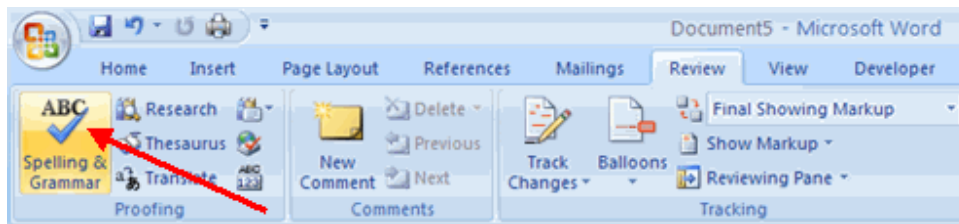


There are many features to help you proofread your document. These include: Spelling and Grammar, Thesaurus, AutoCorrect, Default Dictionary, and Word Count.

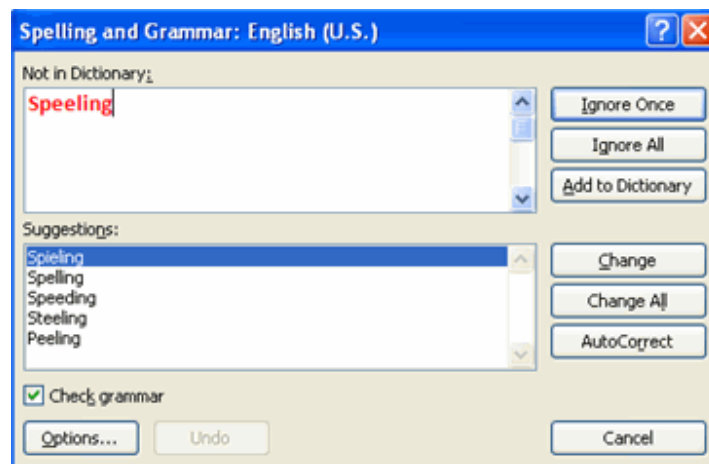
## Spelling and Grammar

To check the spelling and grammar of a document

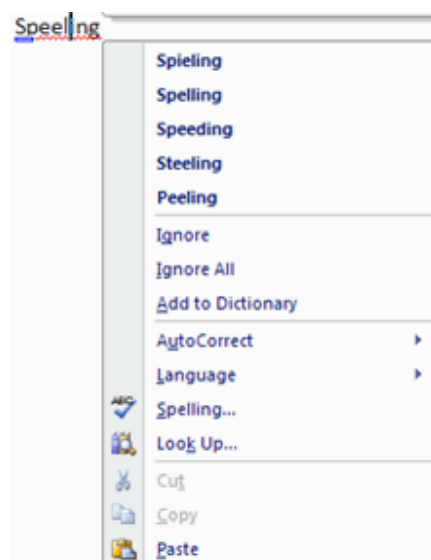
- Place the cursor at the beginning of the document or the beginning of the section that you want to check
- Click the **Review** Tab on the Ribbon
- Click **Spelling & Grammar** on the Proofing Group.



- Any errors will display a dialog box that allows you to choose a more appropriate spelling or phrasing.



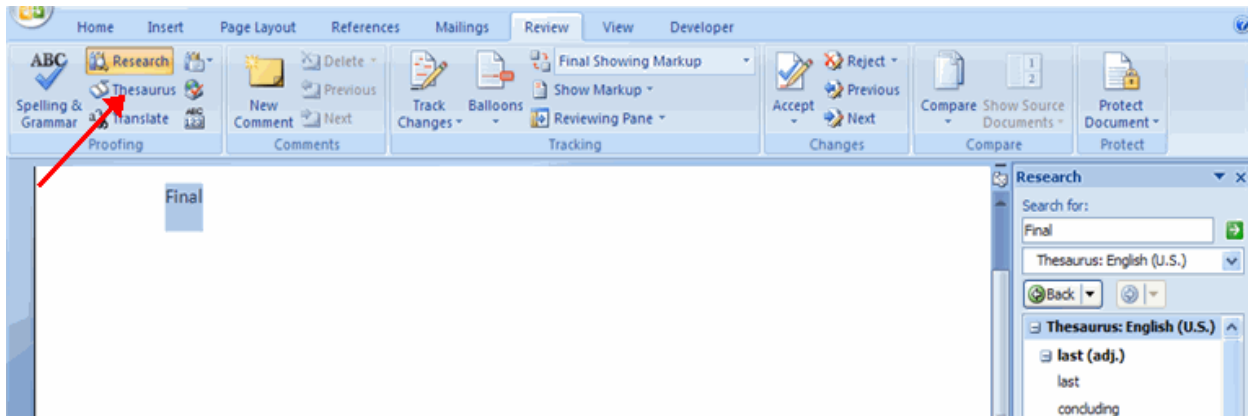
If you wish to check the spelling of an individual word, you can right click any word that has been underlined by Word and choose a substitution.



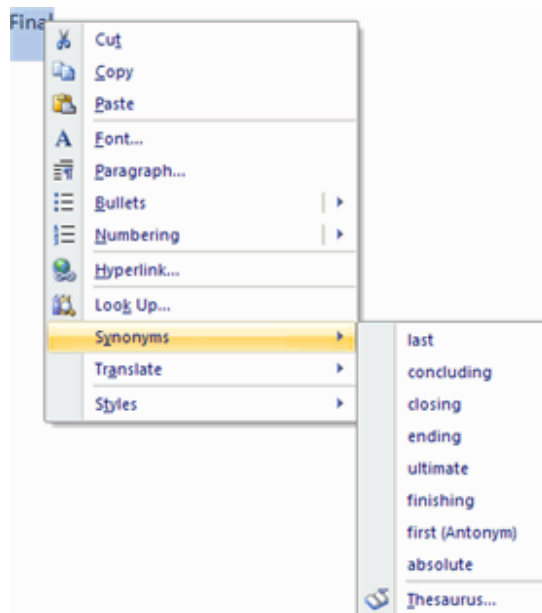
### Thesaurus

The Thesaurus allows you to view synonyms. To use the thesaurus:

- Click the **Review** Tab of the Ribbon
- Click the **Thesaurus** Button on the Proofing Group.
- The thesaurus tool will appear on the right side of the screen and you can view word options.



You can also access the thesaurus by right-clicking any word and choosing **Synonyms** on the menu.

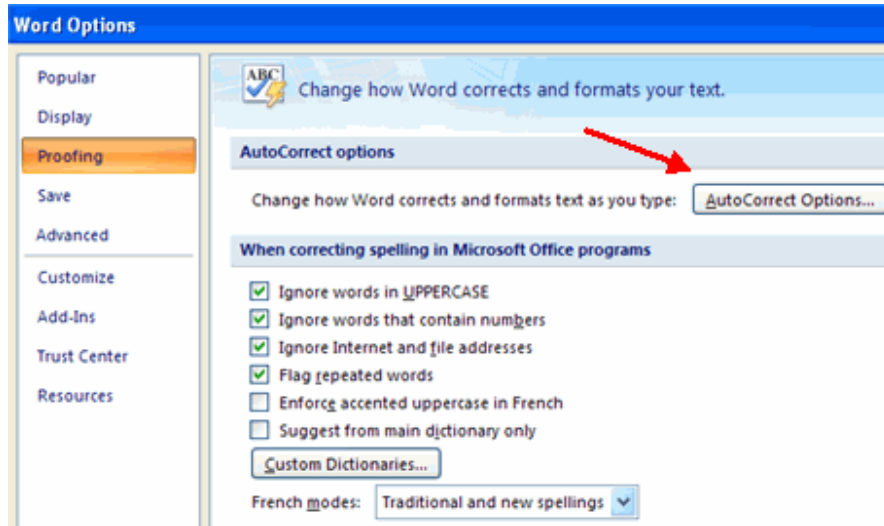


### Customize AutoCorrect

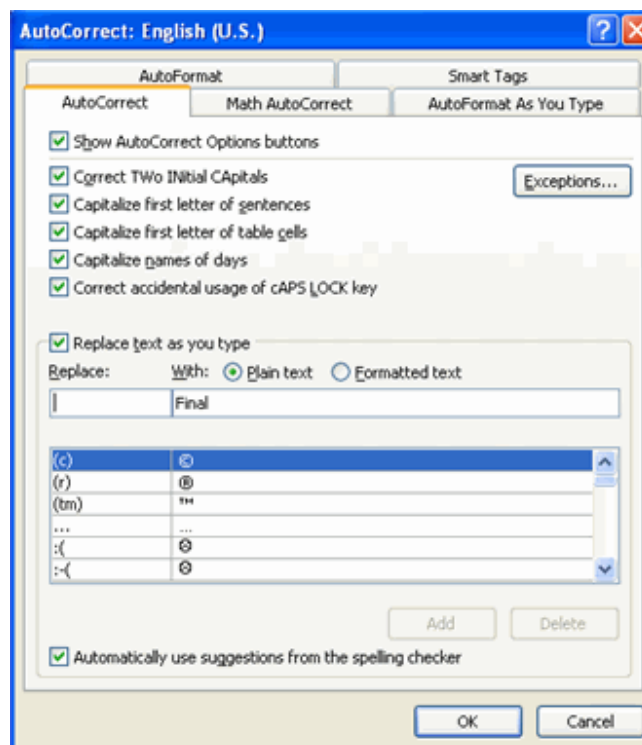
You can set up the AutoCorrect tool in Word to retain certain text the way it is. To customize AutoCorrect:

- Click the **Microsoft Office** button
- Click the **Word Options** Button
- Click the **Proofing** tab

- Click **AutoCorrect Options** button



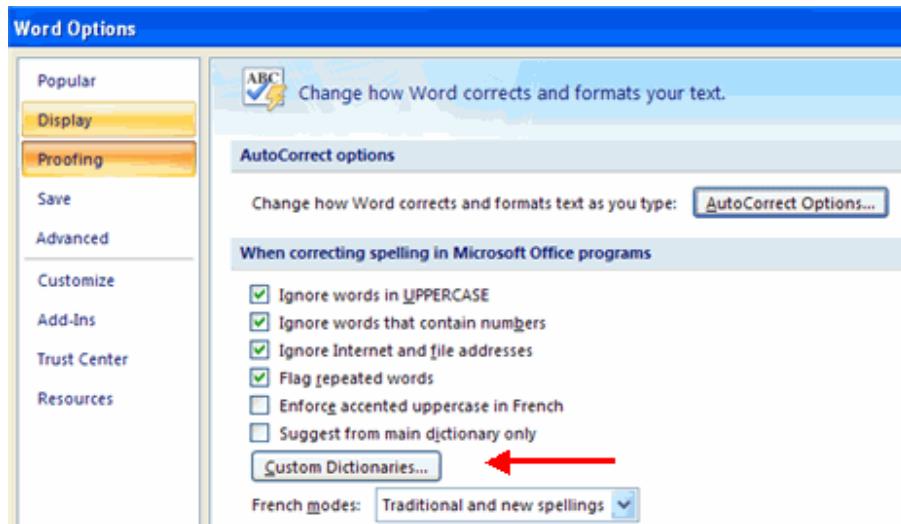
- On the **AutoCorrect Tab**, you can specify words you want to replace as you type



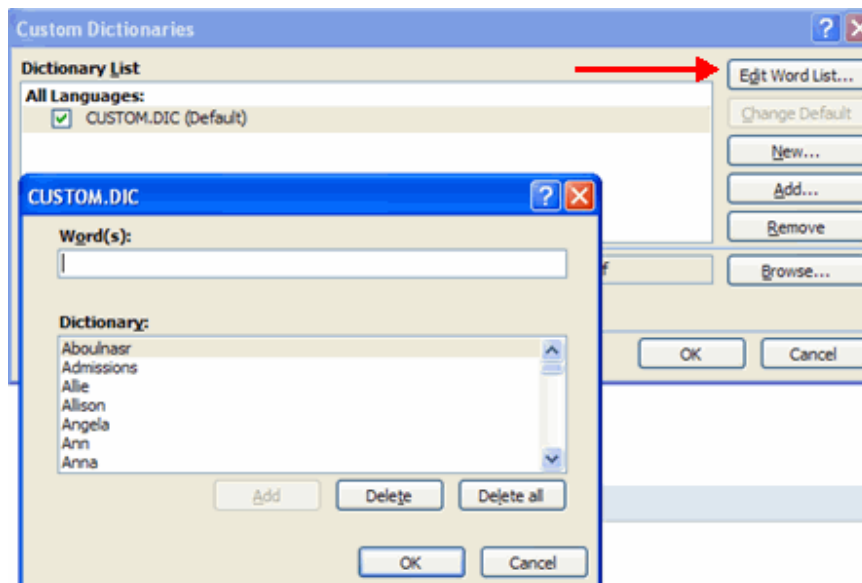
### Create a New Default Dictionary

Often you will have business or educational jargon that may not be recognized by the spelling and/or grammar check in Word. You can customize the dictionary to recognize these words.

- Click the **Microsoft Office** button
- Click the **Word Options** Button
- Click the **Proofing** tab
- Click the **When Correcting Spelling** tab
- Click **Custom Dictionaries**



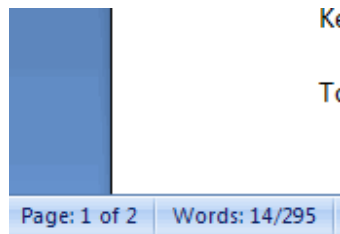
- Click **Edit Word List**
- Type in any words that you may use that are not recognized by the current dictionary.



### Check Word Count

To check the word count in Word 2007 look at the bottom left corner of the screen.

It will give you a total word count or if you have text highlighted it will tell you how many words are highlighted out of the total.

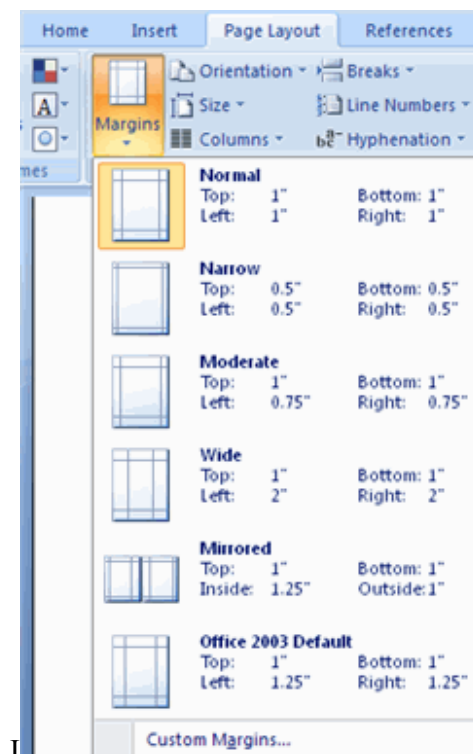


## PAGE MARGINS

### Modify Page Margins and Orientations

The page margins can be modified through the following steps:

- Click the **Page Layout** Tab on the Ribbon
- On the **Page Setup** Group, Click **Margins**
- Click a **Default Margin**, or
- Click **Custom Margins** and complete the dialog box.

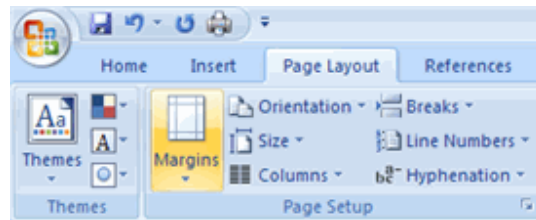


To change the Orientation, Size of the Page, or Columns:

- Click the **Page Layout** Tab on the Ribbon



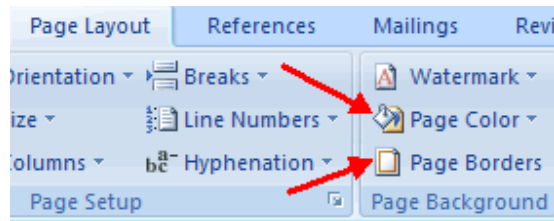
- On the Page Setup Group, Click the **Orientation**, **Size**, or **Columns** drop down menus
- Click the appropriate choice



### Apply a Page Border and Color

To apply a page border or color:

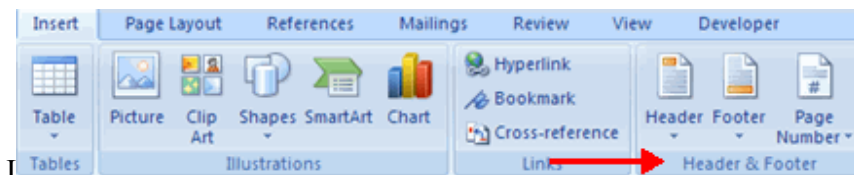
- Click the **Page Layout** Tab on the Ribbon
- On the Page Background Group, click the **Page Colors** or **Page Borders** drop down menus



### Insert Common Header and Footer Information

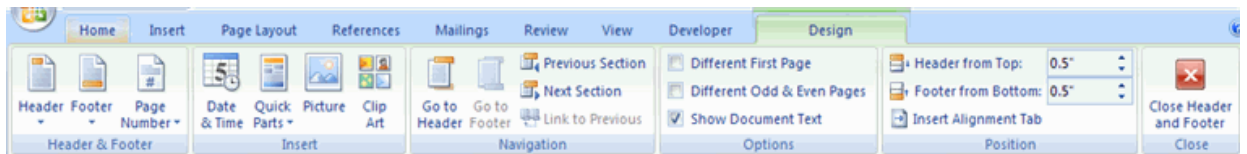
To insert Header and Footer information such as page numbers, date, or title, first, decide if you want the information in the header (at the top of the page) or in the Footer (at the bottom of the page), then:

- Click the **Insert** Tab on the Ribbon
- Click **Header or Footer**
- Choose a style



- The **Header/Footer Design** Tab will display on the Ribbon

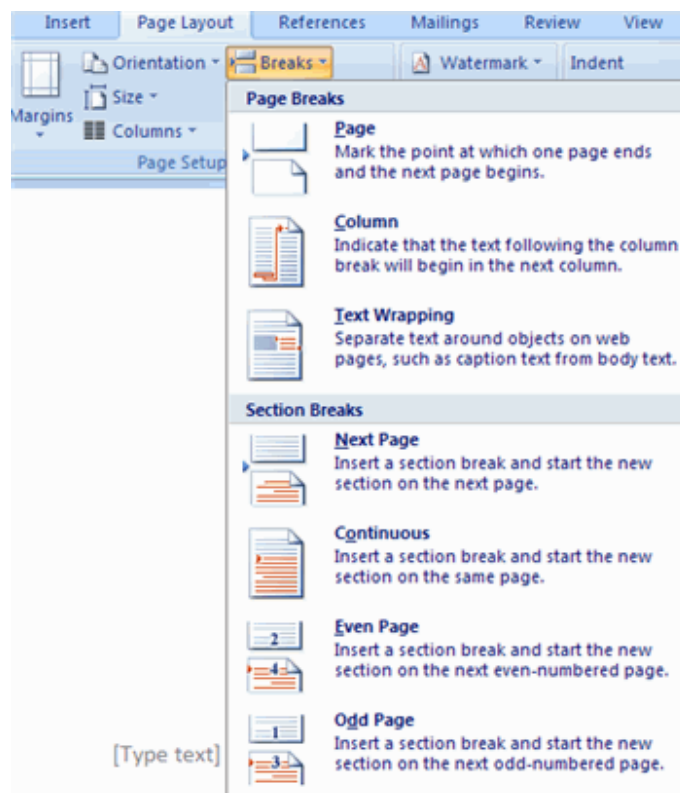
- Choose the information that you would like to have in the header or footer (date, time, page numbers, etc.) or type in the information you would like to have in the header or footer



### Create a Page Break

To insert a page break:

- Click the **Page Layout** Tab on the Ribbon
- On the **Page Setup** Group, click the **Breaks Drop Down Menu**
- Click **Page Break**

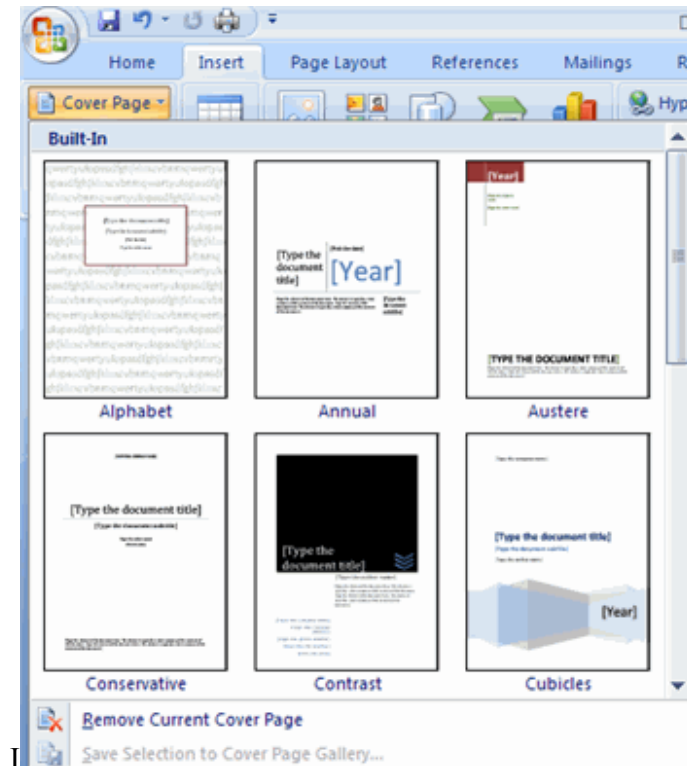


### Insert a Cover Page

To insert a cover page:

- Click the **Insert** Tab on the Ribbon
- Click the **Cover Page** Button on the Pages Group

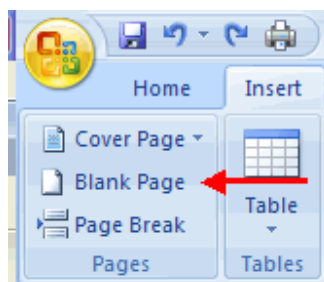
- Choose a style for the cover page



### Insert a Blank Page

To insert a blank page:

- Click the **Insert** Tab on the Ribbon
- Click the **Blank Page** Button on the Page Group



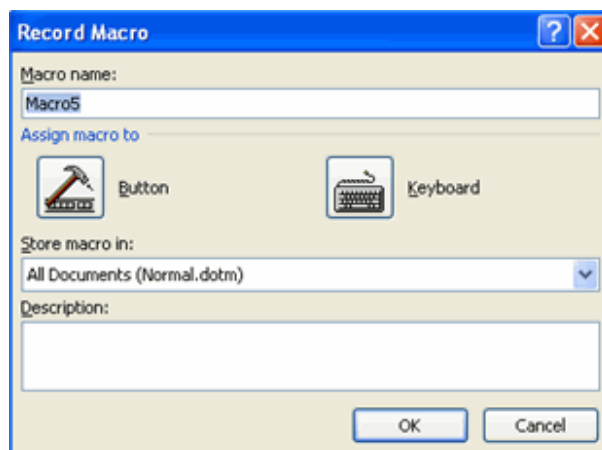
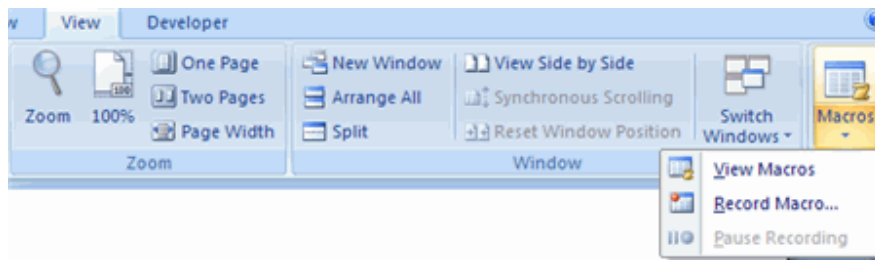
### MACROS

Macros are advanced features that can speed up editing or formatting you may perform often in a Word document. They record sequences of menu selections that we choose so that a series of actions can be completed in one step.

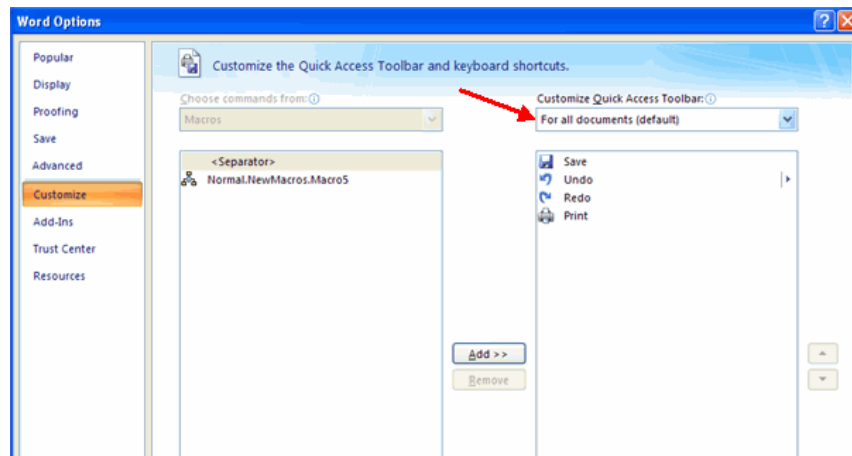
## Recording a Macro

To record a Macro:

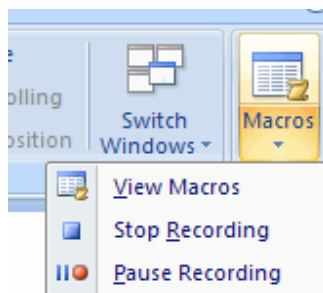
- Click the **View** Tab on the Ribbon
- Click **Macros**
- Click **Record Macro**



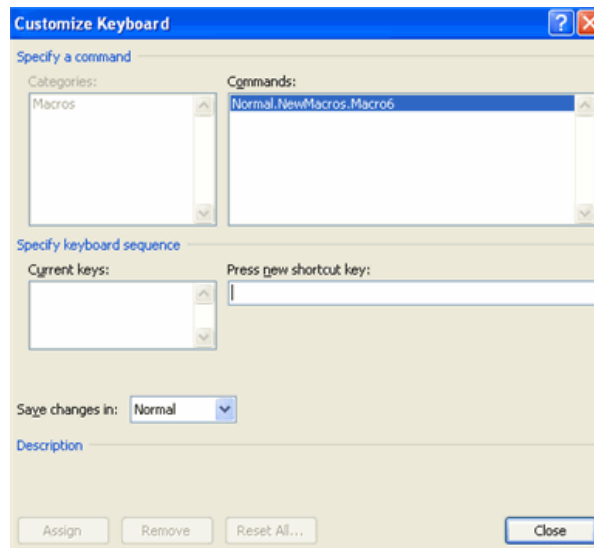
- Enter a name (without spaces)
- Click whether you want it assigned to a **button** (on the Quick Access Toolbar) or the **keyboard** (a sequence of keys)
- To assign the macro a **button** on the Quick Access Toolbar:
  - Click **Button**
  - Under the **Customize Quick Access Toolbar**, select the document for which you want the Macro available



- Under Choose Commands: Click the **Macro** that you are recording
- Click **Add**
- Click **OK** to begin Recording the Macro
- Perform the actions you want recorded in the Macro
- Click on **Macros**
- Click on **Stop Recording Macros**



- To assign a macro button to a **keyboard shortcut**:
  - Click **Keyboard**
  - In the **Press New Shortcut Key** box, type the key sequence that you want and click **Assign**



- Click **Close** to begin recording the Macro
- Perform the actions you want recorded in the Macro
- Click on **Macros**
- Click on **Stop Recording Macros**

### Running a Macro

Running a macro depends on whether it's been added to the Quick Access Toolbar or if it's been given a Keyboard Shortcut.

- To run a Macro from the Quick Access Toolbar, simply click the **Macro Icon**



- To run a Macro from the Keyboard shortcut, simply **press the keys** that you have programmed to run the Macro.

### Table of Contents

### Creating Web Pages »

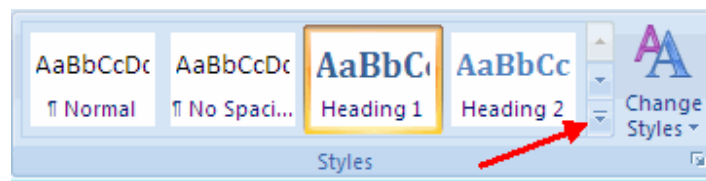
The easiest way to create a **Table of Contents** is to utilize the **Heading Styles** that you want to include in the Table of Contents. For example: Heading 1, Heading 2, etc. based on the content of your document. When you add or delete headings from your document, Word updates your **Table of Contents**. Word also updates the page number in the table of contents when information in the document is added or deleted. When you create a Table of Contents, the first thing you want to do is mark the entries in your document. The Table of Contents is formatted based on levels of headings. Level 1 will include any text identified with the style Heading 1.

## Mark Table of Contents Entries

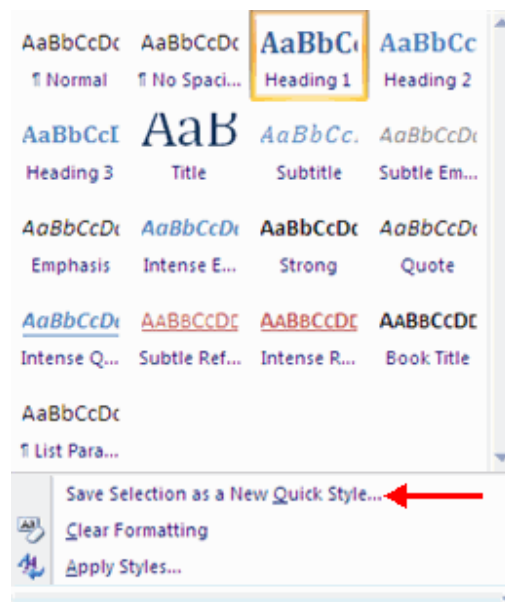
You can mark the Table of Contents entries in one of two ways: by using built-in heading styles or by marking individual text entries.

### To Use Built-In Heading Styles

- Select the text that you wish to be the heading
- Click the **Home** Tab
- In the Styles Group, click **Heading 1** (or the appropriate heading)

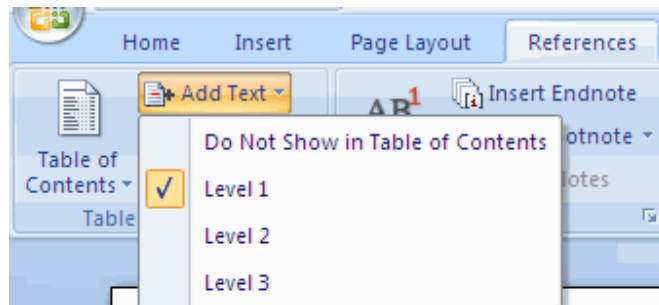


- If you don't see the style you want, click the arrow to expand the **Quick Styles Gallery**
- If the style you want does not appear click **Save Selection** as New Quick Style



### To Mark Individual Entries

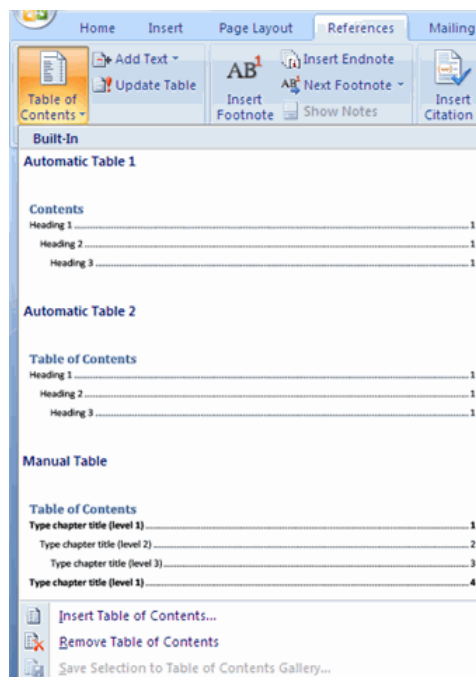
- Select the text you wish to make a heading
- Click the **References** Tab
- Click **Add Text** in the Table of Contents Group
- Click the **Level** that you want to label your selection



### Create a Table of Contents

To create the table of contents:

- Put your cursor in the document where you want the Table of Contents
- Click the **References** Tab
- Click the **Table of Contents** button

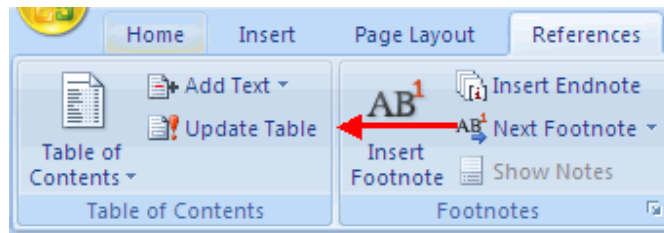


### Update Table of Contents

If you have added or removed headings or other table of contents entries you can update by:

- Apply headings or mark individual entries as directed above
- Click the **References** Tab in the Ribbon
- Click **Update Table**

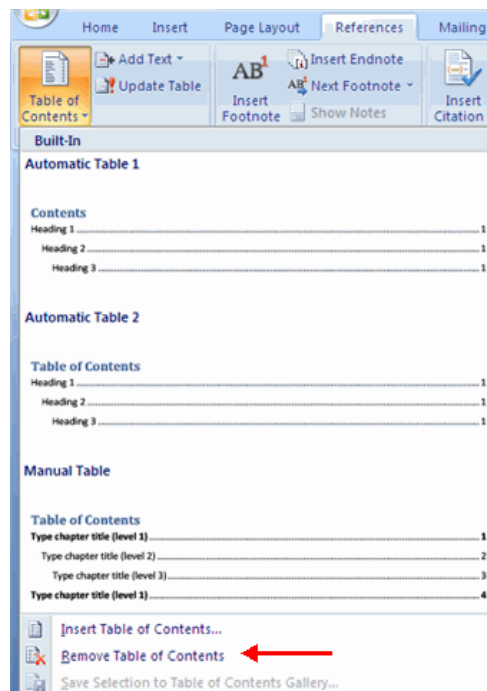




### Delete Table of Contents

To delete a table of contents:

- Click the **References** Tab on the Ribbon
- Click **Table of Contents**
- Click **Remove Table of Contents**

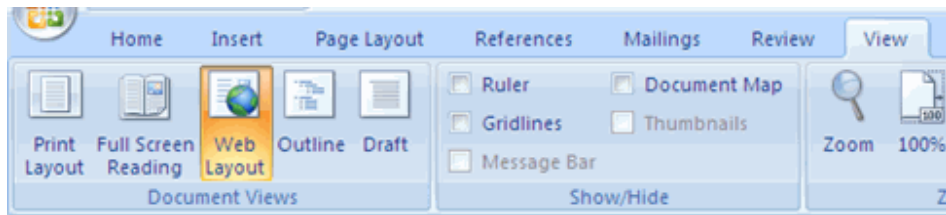


### Creating Web Pages

### Lists »

Simple web pages can be created in Word using the Save as Feature. In a web document, you can insert pictures and hyperlinks. To view the document as you would a web page:

- Click the **View** Tab on the Ribbon
- Click the **Web Layout** Button in the Document Views Group



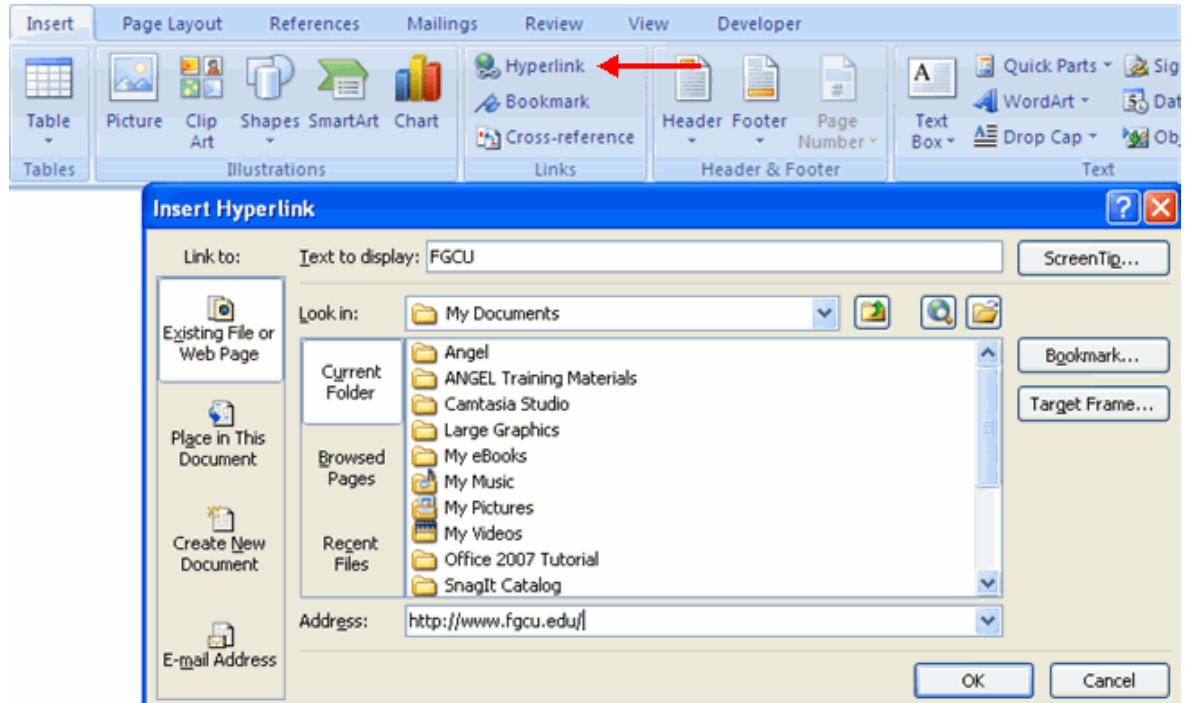
## Entering Text

To enter text into the document, simply begin typing. If you want to adjust the layout of the page and text, you should use tables to format the page properly.

## Hyperlinks

Hyperlinks, or links, allow the reader to click on text and go to another web site. To create a hyperlink:

- Select the text that will be the link
- Click the **Insert** Tab of the Ribbon
- Click the **Hyperlink** Button on the Links Group
- Type in the web address, or URL, of the link
- Click **OK**

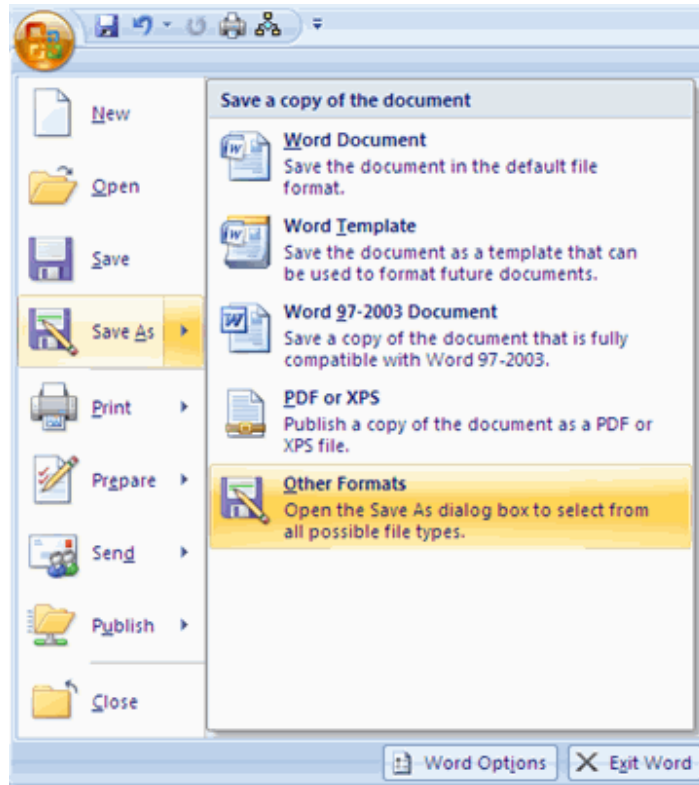


## Saving Web Pages

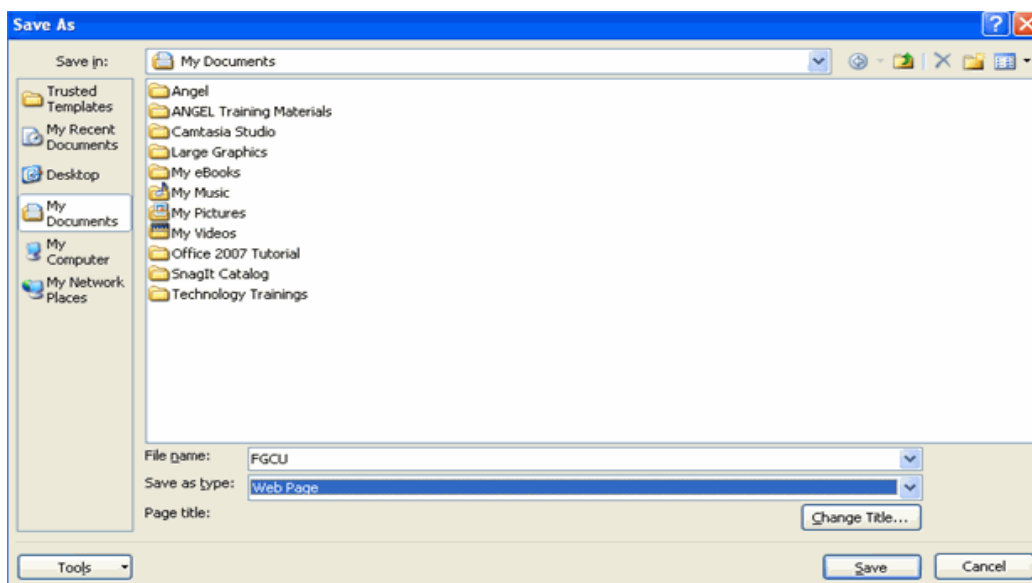
To save a web page:

- Click the **Office Button**
- Move the cursor over **Save As**

- Click **Other Formats**



- Under Save as Type, click **Web Page**
- Type in the name of the document (without spaces)



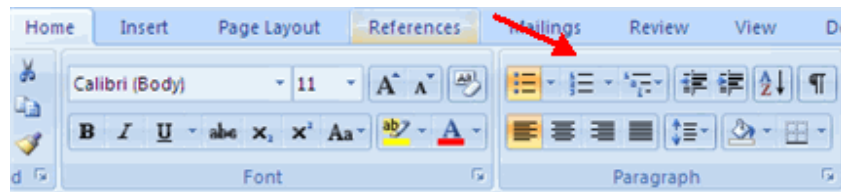
Lists allow you to format and organize text with numbers, bullets, or in an outline.

### Bulleted and Numbered Lists

Bulleted lists have bullet points, numbered lists have numbers, and outline lists combine numbers and letters depending on the organization of the list.

To **add** a list to existing text:

- Select the text you wish to make a list
- From the Paragraph Group on the Home Tab, Click the **Bulleted or Numbered Lists** button



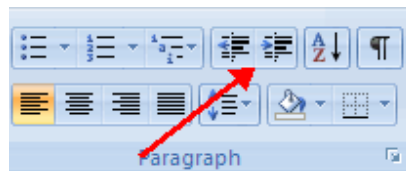
To **create** a new list:

- Place your cursor where you want the list in the document
- Click the **Bulleted or Numbered Lists** button
- Begin typing

### Nested Lists

A nested list is list with several levels of indented text. To create a nested list:

- Create your list following the directions above
- Click the **Increase or Decrease Indent** button

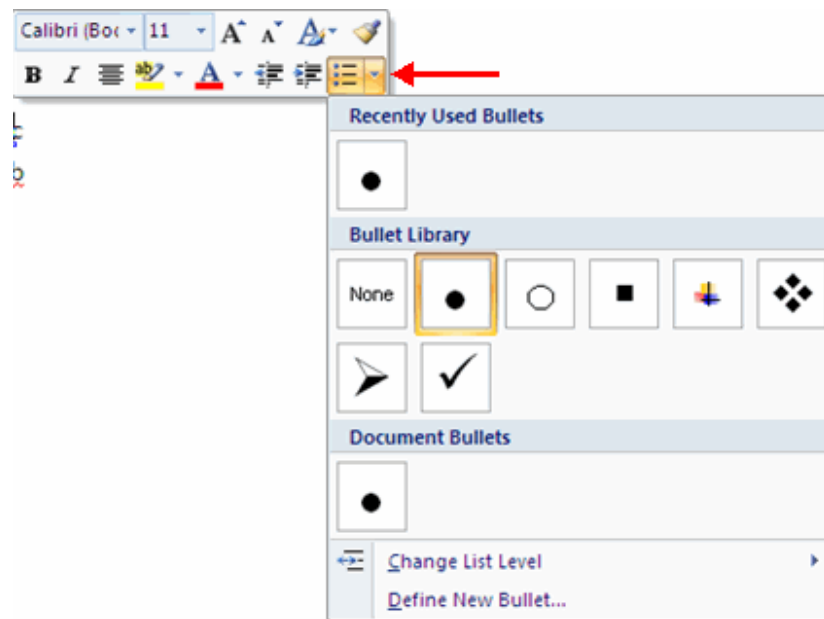


### Formatting Lists

The bullet image and numbering format can be changed by using the **Bullets or Numbering** dialog box.

- Select the entire list to change all the bullets or numbers, or  
Place the cursor on one line within the list to change a single bullet

- Right click
- Click the **arrow** next to the bulleted or numbered list and choose a bullet or numbering style.



## References and Citations

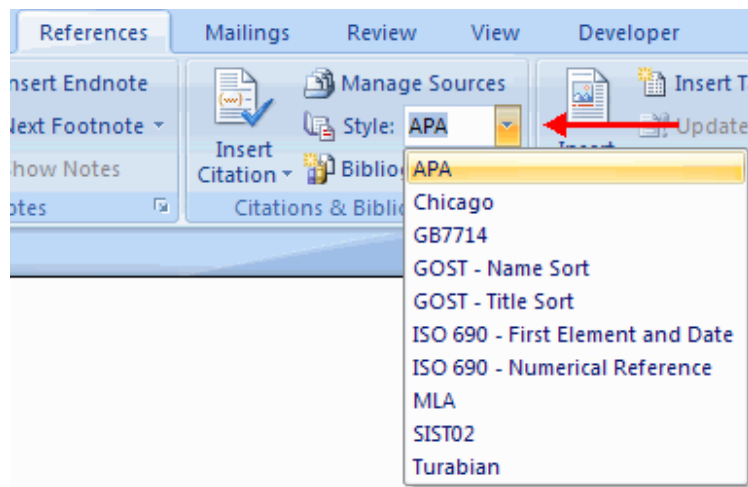
## Track Changes »

Word 2007 offers great tools for citing sources, creating a bibliography, and managing the sources. The first step to creating a reference list and citations in a document is to choose the appropriate style that you will be using for formatting the citations and references.

### Style

To choose a publishing style:

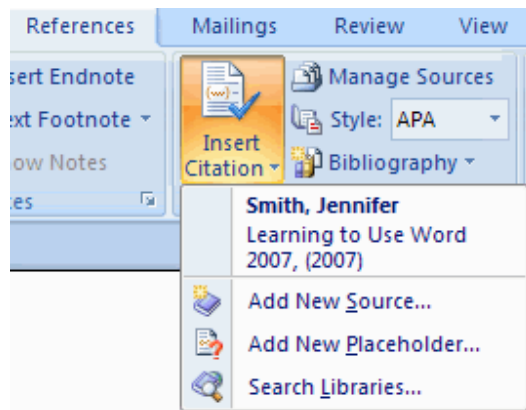
- Click the **References** Tab on the Ribbon
- Click the drop down box next to **Style** in the Citations & Bibliography Group
- Choose the appropriate style.



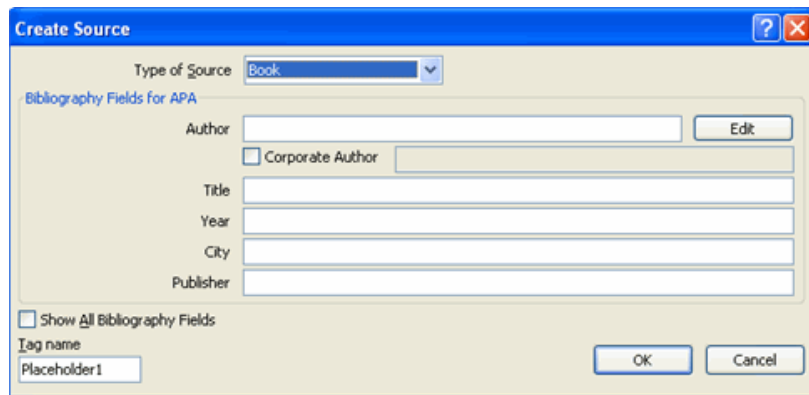
## Citations

To insert a citation in the text portion of your document:

- Click the **References** Tab on the Ribbon
- Click the **Insert Citation** Button on the Citations & Bibliography Group
- If this is a new source, click **New Source**
- If you have already created this source, it will in the drop down list and you can click on it



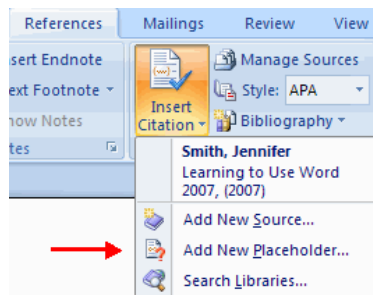
- If you are creating a **New Source**, choose the **type of source** (book, article, etc.)
- Complete the **Create Source Form**
- If you need additional fields, be sure to click the **Show All Bibliography Fields** check box
- Click **OK**



## Placeholders

Placeholders can be utilized when there is a reference to be cited, but you do not have all of the information on the source. To insert a Placeholder:

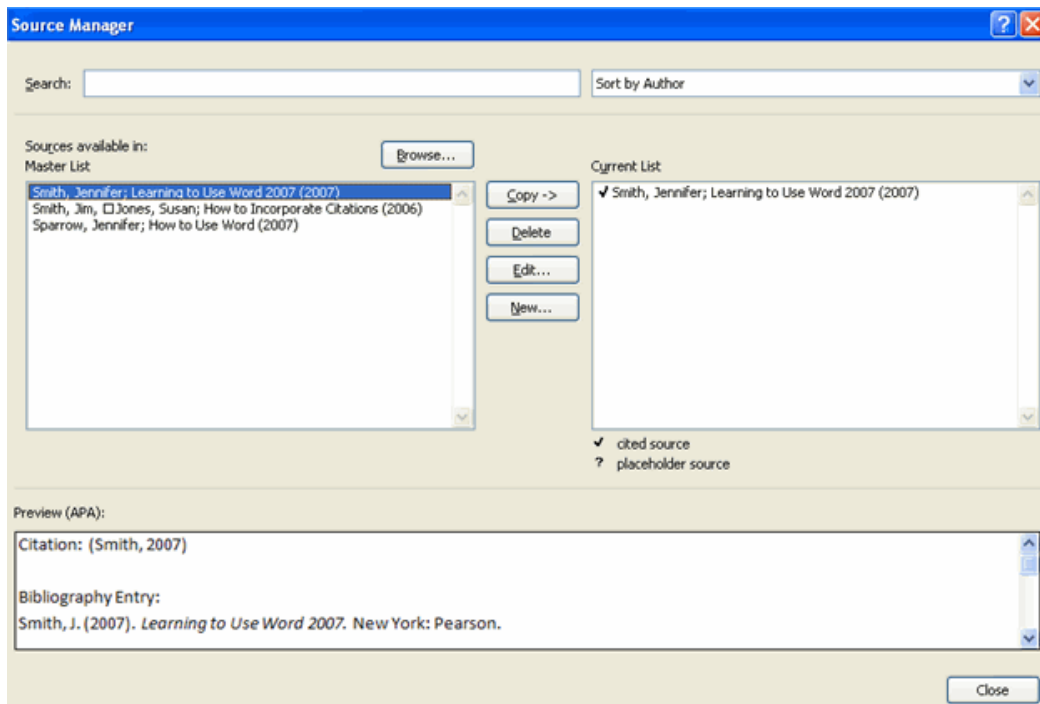
- Click **Insert Citation**
- Click **Add New Placeholder**



## Manage Sources

Once you have completed a document you may need to add or delete sources, modify existing sources, or complete the information for the placeholders. To Manage Sources:

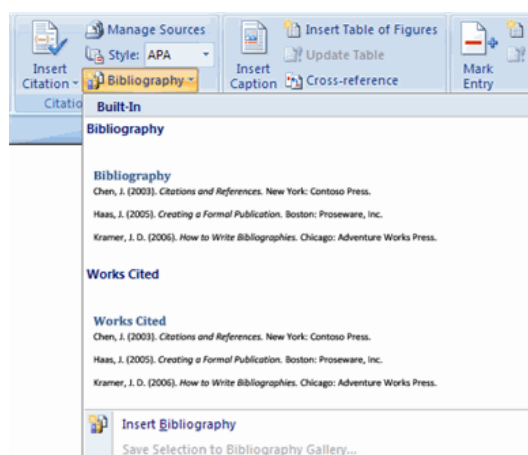
- Click the **References** Tab on the Ribbon
- Click the **Manage Sources** Button on the Citations & Bibliography Group
- From this menu you can **Add, Delete, and Edit Sources** (note, you can preview the source in the bottom pane of the window)



## Bibliography

To add a Bibliography to the document:

- Place the cursor in the document where you want the bibliography
- Click the **References** Tab on the Ribbon
- Click the **Bibliography** Button on the Citations & Bibliography Group
- Choose **Insert Built-in Bibliography/Works Cited** or **Insert Bibliography**



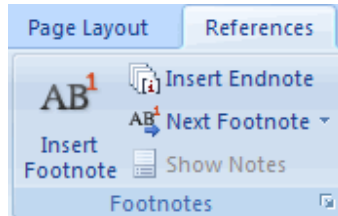
## Insert Footnote

Some types of academic writing utilize footnotes. To insert a footnote:

- Click the **References** Tab on the Ribbon



- Click **Insert Footnote** (or Insert Endnote depending on your needs)
- Begin typing the footnote



### Questions

1. Ctrl+N is a shortcut key to open a new document.

**Ans: True**

2. Spelling mistakes are underlined in red colour in MSWord.

**Ans: True**

3. Shortcut key for copy option is ctrl+v.

**Ans: False**

4. Cut-paste and copy-paste does the same operation.

**Ans: False**

5. Bullets and numbering are is used for formatting.

**Ans: True**

6. Grammatical mistakes are underlined in red color in MSWord.

**Ans: False**

7. Save and saveAs does the same operation.

**Ans: False**

8. Give two font names.

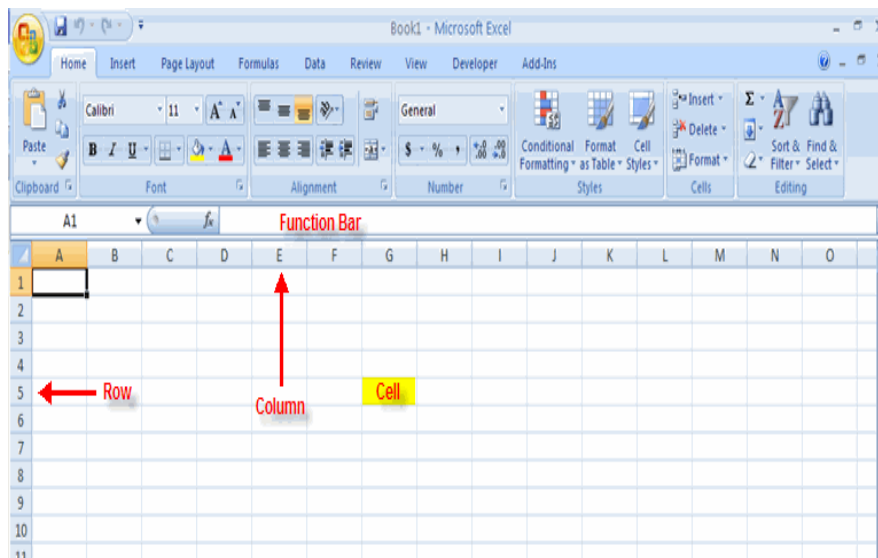
9. What are the justifications available in word?

10. How to enter equations in a word document?

## Lecture.26

**Programming languages – types – low level and high level Programming languages.  
BASIC language – input/output statements – READ, DATA, INPUT and PRINT  
statements**

### EXCEL



### Spreadsheets

A spreadsheet is an electronic document that stores various types of data. There are vertical columns and horizontal rows. A cell is where the column and row intersect. A cell can contain data and can be used in calculations of data within the spreadsheet. An Excel spreadsheet can contain workbooks and worksheets. The workbook is the holder for related worksheets.

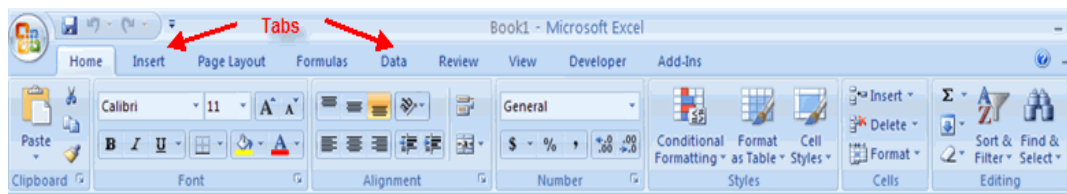
### Microsoft Office Button

The Microsoft Office Button performs many of the functions that were located in the File menu of older versions of Excel. This button allows you to create a new workbook, Open an existing workbook, save and save as, print, send, or close.

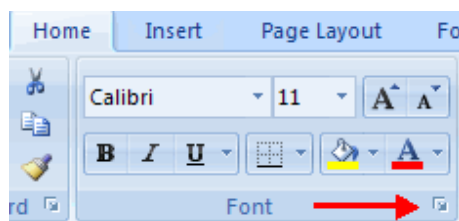


## Ribbon

The ribbon is the panel at the top portion of the document. It has seven tabs: Home, Insert, Page Layouts, Formulas, Data, Review, and View. Each tab is divided into groups. The groups are logical collections of features designed to perform function that we will utilize in developing or editing our Excel spreadsheets.



Commonly utilized features are displayed on the Ribbon. To view additional features within each group, click the arrow at the bottom right corner of each group.



**Home:** Clipboard, Fonts, Alignment, Number, Styles, Cells, Editing

**Insert:** Tables, Illustrations, Charts, Links, Text

**Page Layouts:** Themes, Page Setup, Scale to Fit, Sheet Options, Arrange

**Formulas:** Function Library, Defined Names, Formula Auditing, Calculation

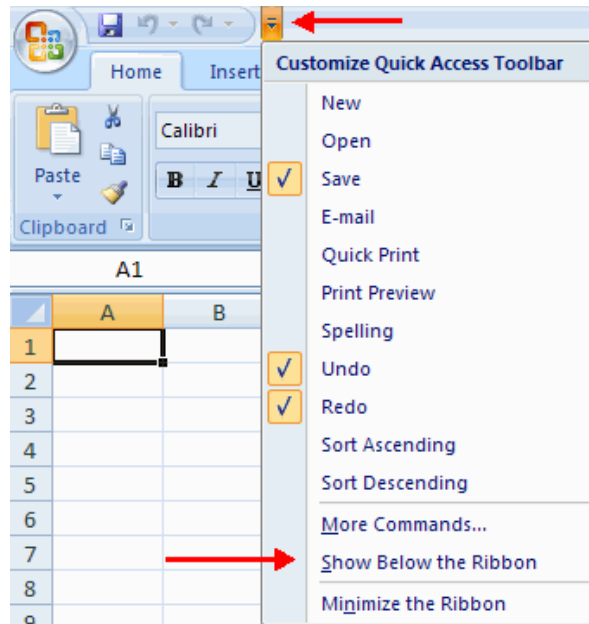
**Data:** Get External Data, Connections, Sort & Filter, Data Tools, Outline

**Review:** Proofing, Comments, Changes

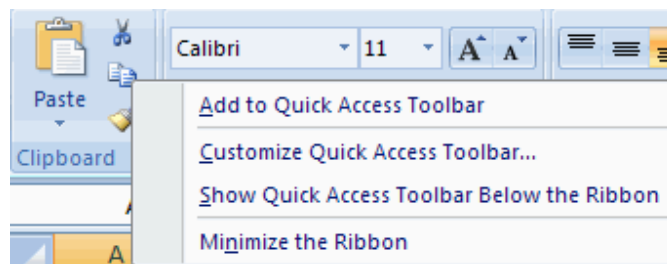
**View:** Workbook Views, Show/Hide, Zoom, Window, Macros

## Quick Access Toolbar

The **quick access toolbar** is a customizable toolbar that contains commands that we may want to use. We can place the quick access toolbar above or below the ribbon. To change the location of the quick access toolbar, click on the arrow at the end of the toolbar and click **Show Below the Ribbon**.

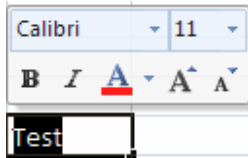


We can also add items to the quick access toolbar. Right click on any item in the Office Button or the Ribbon and click Add to Quick Access Toolbar and a shortcut will be added.



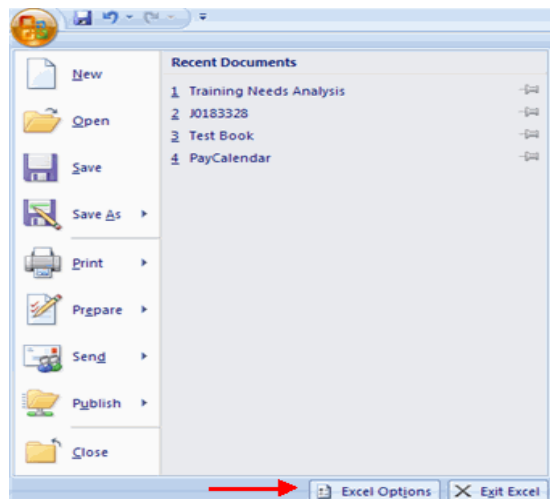
## Mini Toolbar

A new feature in Office 2007 is the Mini Toolbar. This is a floating toolbar that is displayed when we select text or right-click text. It displays common formatting tools, such as Bold, Italics, Fonts, Font Size and Font Color.



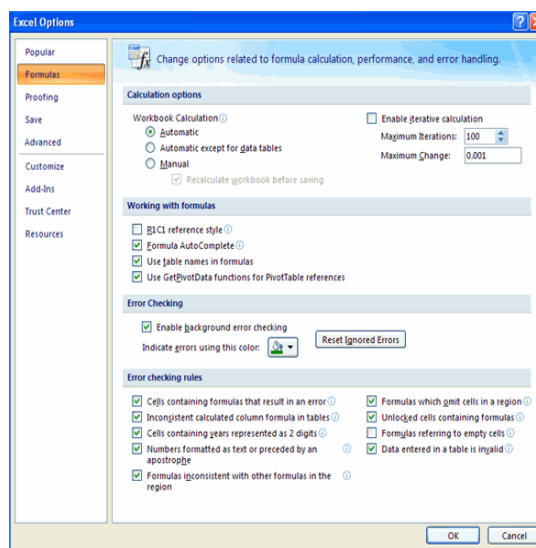
Excel 2007 offers a wide range of customizable options that allow us to make Excel work the best for us. To access these customizable options:

- Click the **Office Button**
- Click **Excel Options**



## Formulas

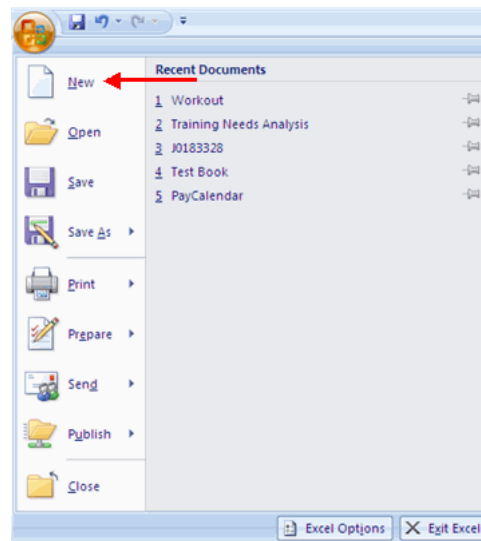
This feature allows you to modify calculation options, working with formulas, error checking, and error checking rules.



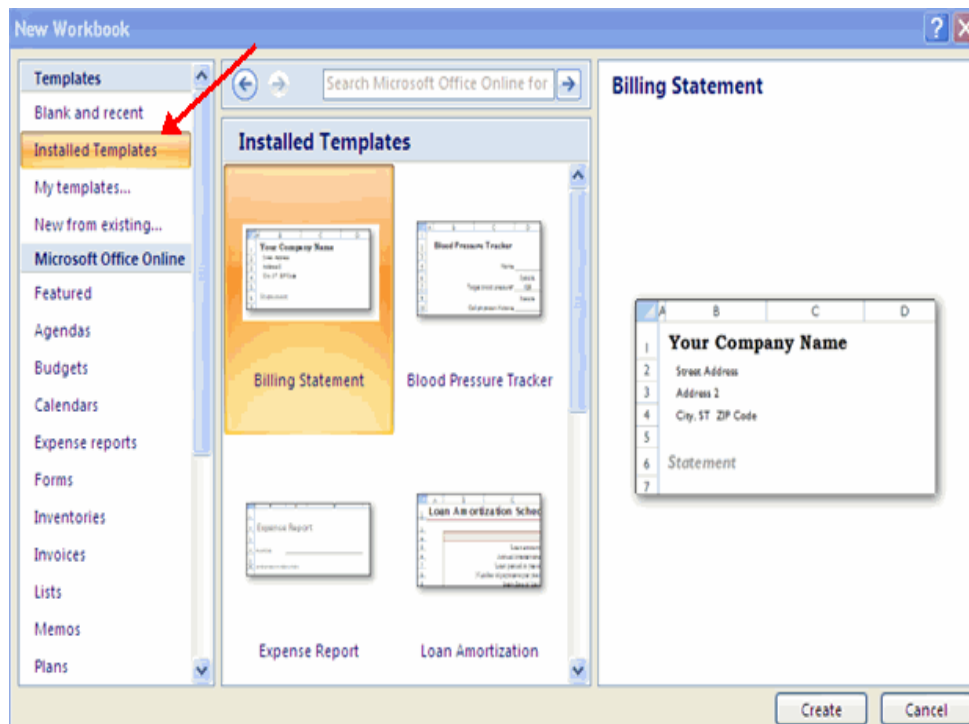
## Create a Workbook

To create a new Workbook:

- Click the **Microsoft Office Toolbar**
- Click **New**
- Choose **Blank Document**



If you want to create a new document from a template, explore the templates and choose one that fits your needs.

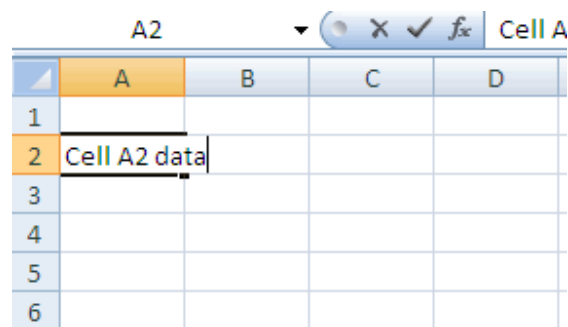


## Entering Data

There are different ways to enter data in Excel: in an active cell or in the formula bar.

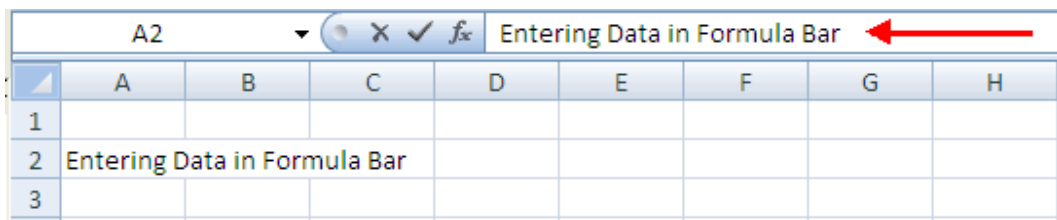
To enter data in an active cell:

- Click in the **cell** where you want the data
- Begin typing



To enter data into the **formula bar**

- Click the cell where you would like the data
- Place the cursor in the **Formula Bar**
- Type in the data



Excel allows us to move, copy, and paste cells and cell content through cutting and pasting and copying and pasting.

## Select Data

To select a cell or data to be copied or cut:

- Click the **cell**

	A	B	C
1			
2	2-Jun		
3	4-Jun		
4	6-Jun		
5			
6			
7			

- Click and drag the cursor to select many cells in a range

	A	B	C	D	E	F
1		Widgets	Customers	Sales	Price	
2	2-Jun					
3	4-Jun	2	4	2	5	
4	6-Jun					
5						
6						
7						

### Select a Row or Column

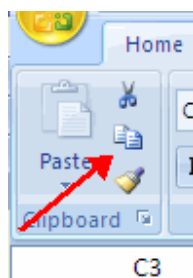
To select a row or column click on the **row** or **column header**.

	A	B	C	D	E	F
1		Widgets	Customers	Sales	Price	
2	2-Jun					
3	4-Jun	2	4	2	5	
4	6-Jun					
5						

### Copy and Paste

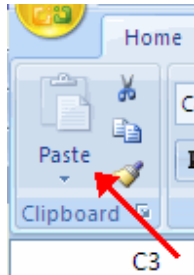
To copy and paste data:

- Select the cell(s) that you wish to copy
- On the **Clipboard** group of the **Home** tab, click **Copy**





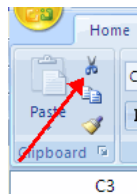
- Select the cell(s) where you would like to copy the data
- On the **Clipboard** group of the **Home** tab, click **Paste**



### Cut and Paste

To cut and paste data:

- Select the cell(s) that you wish to copy
- On the **Clipboard** group of the **Home** tab, click **Cut**

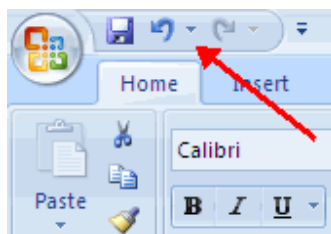


- Select the cell(s) where you would like to copy the data
- On the **Clipboard** group of the **Home** tab, click **Paste**

### Undo and Redo

To undo or redo your most recent actions:

- On the **Quick Access Toolbar**
- Click **Undo** or **Redo**



### Auto Fill

The Auto Fill feature fills cell data or series of data in a worksheet into a selected range of cells. If we want the same data copied into the other cells, we only need to complete one cell. If we want to have a series of data (for example, days of the week) fill in the first two cells in the series and then use the auto fill feature. To use the Auto Fill feature:

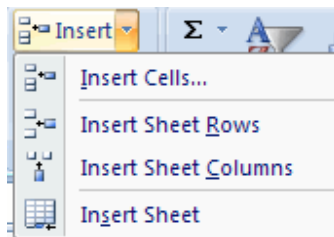
- Click the **Fill Handle**
- Drag the **Fill Handle** to complete the cells

	A	B	C	D
1		Widgets	Customers	Sales
2	2-Jun			
3	4-Jun	2	4	
4	6-Jun			
5				

### Insert Cells, Rows, and Columns

To insert cells, rows, and columns in Excel:

- Place the cursor in the row below where you want the new row, or in the column to the left of where you want the new column
- Click the **Insert** button on the **Cells** group of the **Home** tab
- Click the appropriate choice: **Cell, Row, or Column**



### Delete Cells, Rows and Columns

To delete cells, rows, and columns:

- Place the cursor in the cell, row, or column that you want to delete
- Click the **Delete** button on the **Cells** group of the **Home** tab
- Click the appropriate choice: **Cell, Row, or Column**

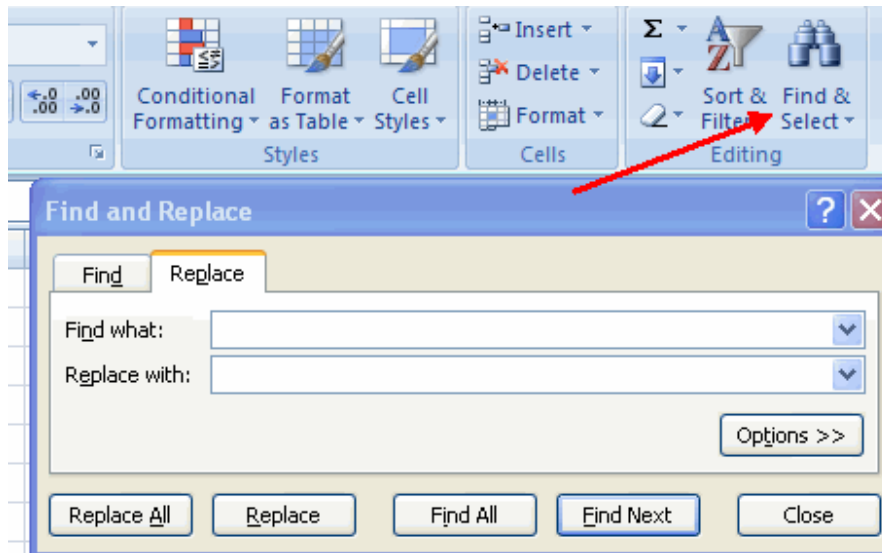


### Find and Replace

To find data or find and replace data:

- Click the **Find & Select** button on the **Editing** group of the **Home** tab
- Choose **Find** or **Replace**

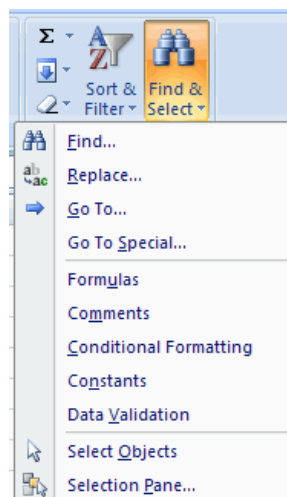
- Complete the **Find What** text box
- Click on **Options** for more search options



### Go To Command

The Go To command takes you to a specific cell either by cell reference (the Column Letter and the Row Number) or cell name.

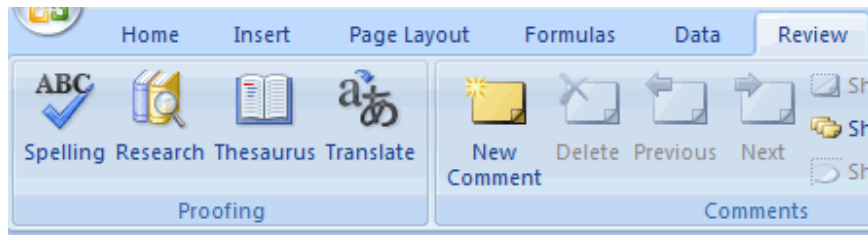
- Click the **Find & Select** button on the **Editing** group of the **Home** tab
- Click **Go To**



### Spell Check

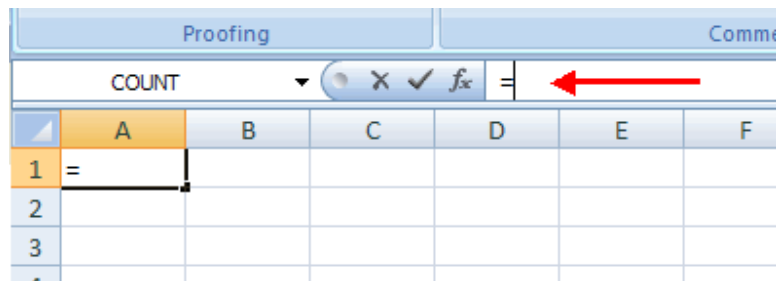
To check the spelling:

- On the **Review** tab click the **Spelling** button



## Excel Formulas

A formula is a set of mathematical instructions that can be used in Excel to perform calculations. Formals are started in the formula box with an = sign.



There are many elements to and excel formula.

**References:** The cell or range of cells that you want to use in your calculation

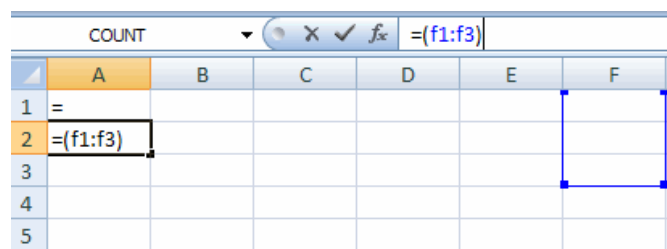
**Operators:** Symbols (+, -, \*, /, etc.) that specify the calculation to be performed

**Constants:** Numbers or text values that do not change

**Functions:** Predefined formulas in Excel

To create a basic formula in Excel:

- Select the **cell** for the formula
- Type = (the equal sign) and the **formula**
- Click **Enter**



## Calculate with Functions

A function is a built in formula in Excel. A function has a name and arguments (the mathematical function) in parentheses. Common functions in Excel:

**Sum:** Adds all cells in the argument

**Average:** Calculates the average of the cells in the argument

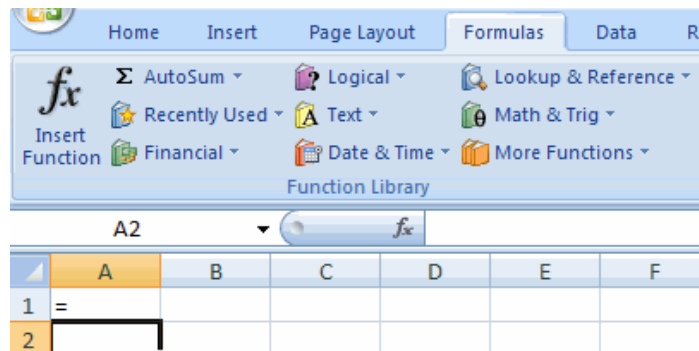
**Min:** Finds the minimum value

**Max:** Finds the maximum value

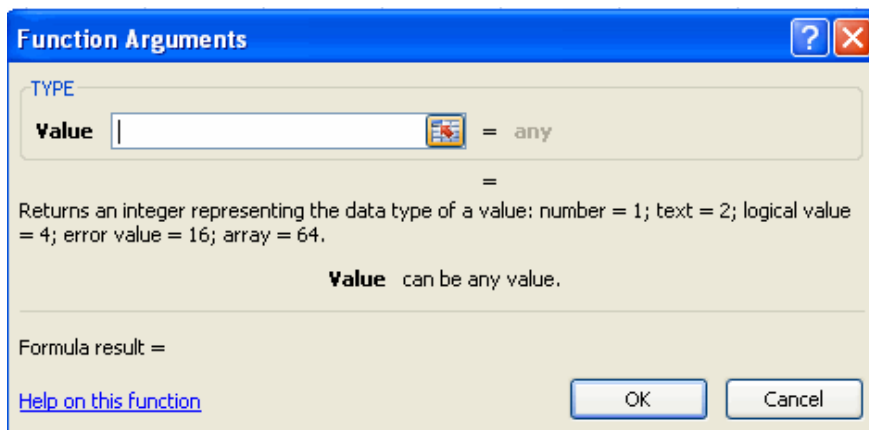
**Count:** Finds the number of cells that contain a numerical value within a range of the argument

To calculate a function:

- Click the **cell** where you want the function applied
- Click the **Insert Function** button
- Choose the function
- Click **OK**



- Complete the Number 1 box with the first cell in the range that you want calculated
- Complete the Number 2 box with the last cell in the range that you want calculated



### Function Library

The function library is a large group of functions on the Formula Tab of the Ribbon. These functions include:

**AutoSum:** Easily calculates the sum of a range

**Recently Used:** All recently used functions

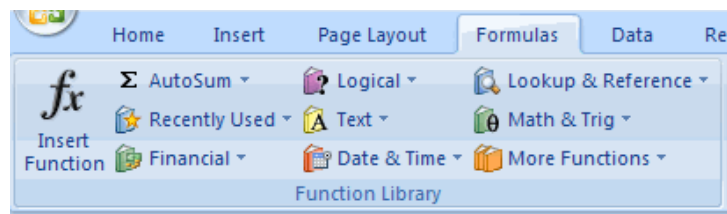
**Financial:** Accrued interest, cash flow return rates and additional financial functions

**Logical:** And, If, True, False, etc.

**Text:** Text based functions

**Date & Time:** Functions calculated on date and time

**Math & Trig:** Mathematical Functions



## Relative, Absolute and Mixed References

Calling cells by just their column and row labels (such as "A1") is called **relative referencing**. When a formula contains relative referencing and it is copied from one cell to another, Excel does not create an exact copy of the formula. It will change cell addresses relative to the row and column they are moved to. For example, if a simple addition formula in cell C1 " $=A1+B1$ " is copied to cell C2, the formula would change to " $=A2+B2$ " to reflect the new row. To prevent this change, cells must be called by **absolute referencing** and this is accomplished by placing dollar signs "\$" within the cell addresses in the formula. Continuing the previous example, the formula in cell C1 would read " $=A\$1+B\$1$ " if the value of cell C2 should be the sum of cells A1 and B1. Both the column and row of both cells are absolute and will not change when copied. **Mixed referencing** can also be used where only the row OR column fixed. For example, in the formula " $=A\$1+B2$ ", the row of cell A1 is fixed and the column of cell B2 is fixed.

## Linking Worksheets

You may want to use the value from a cell in another worksheet within the same workbook in a formula. For example, the value of cell A1 in the current worksheet and cell A2 in the second worksheet can be added using the format "sheetname!celladdress". The formula for this example would be " $=A1+Sheet2!A2$ " where the value of cell A1 in the current worksheet is added to the value of cell A2 in the worksheet named "Sheet2".

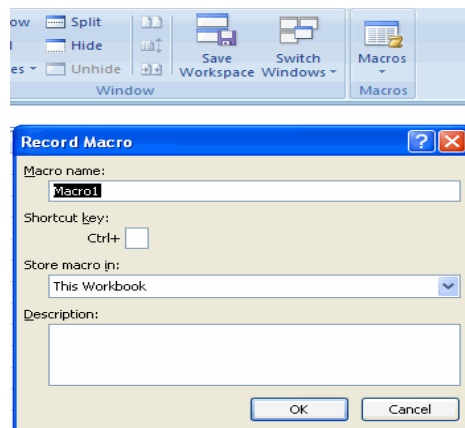
## MACROS

Macros are advanced features that can speed up editing or formatting we may perform often in an Excel worksheet. They record sequences of menu selections that we choose so that a series of actions can be completed in one step.

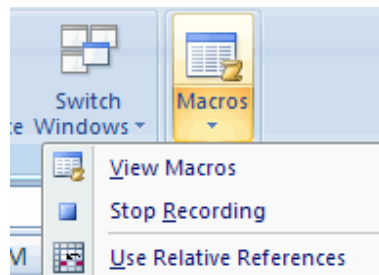
### Recording a Macro

To record a Macro:

- Click the **View** tab on the Ribbon
- Click **Macros**
- Click **Record Macro**
- Enter a name (without spaces)
- Enter a **Shortcut Key**
- Enter a **Description**



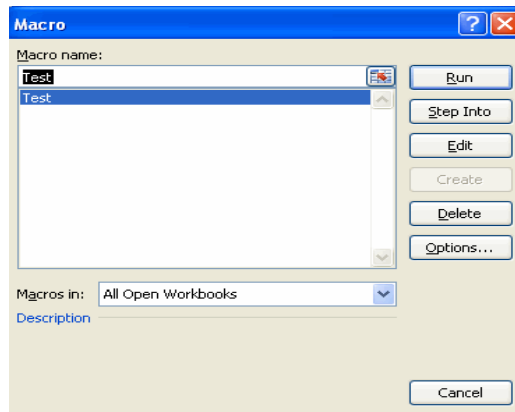
- Perform the **Macro**
- Click **Macros**
- Click **Stop Recording**



## Running a Macro

To run a Macro from the Keyboard shortcut, simply **press the keys** that you have programmed to run the Macro. Or we can view all macros and run by:

- Click **Macros**
- Click **View Macros**
- Choose the **Macro** and click **Run**



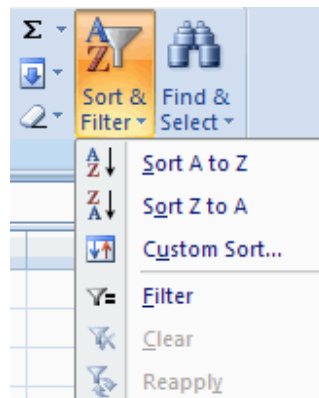
## SORT AND FILTER

Sorting and Filtering allow us to manipulate data in a worksheet based on given set of criteria.

### Basic Sorts

To execute a basic descending or ascending sort based on one column:

- Highlight the cells that will be sorted
- Click the **Sort & Filter** button on the **Home** tab
- Click the **Sort Ascending (A-Z)** button or **Sort Descending (Z-A)** button

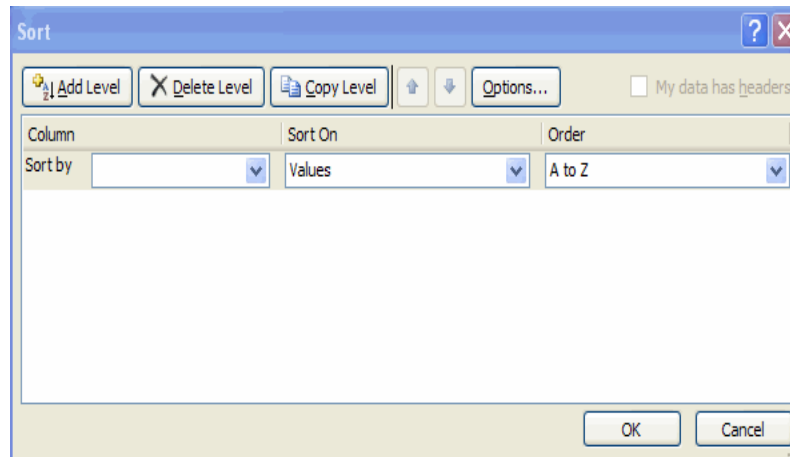


### Custom Sorts

To sort on the basis of more than one column:



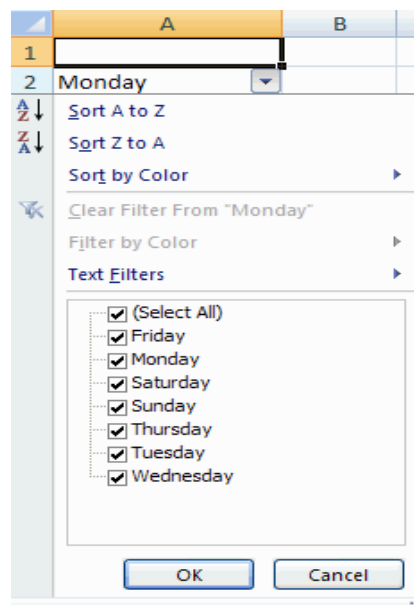
- Click the **Sort & Filter** button on the **Home** tab
- Choose which column we want to sort by first
- Click **Add Level**
- Choose the next column we want to sort
- Click **OK**



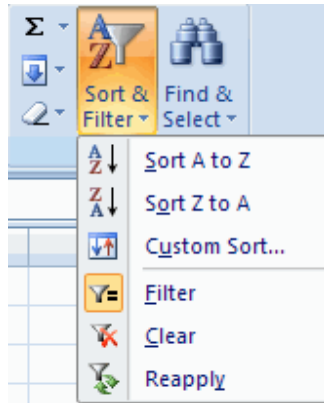
## Filtering

Filtering allows us to display only data that meets certain criteria. To filter:

- Click the column or columns that contain the data we wish to filter
- On the **Home** tab, click on **Sort & Filter**
- Click **Filter** button
- Click the **Arrow** at the bottom of the first cell
- Click the **Text Filter**
- Click the **Words** we wish to Filter



- To clear the filter click the **Sort & Filter** button
- Click **Clear**



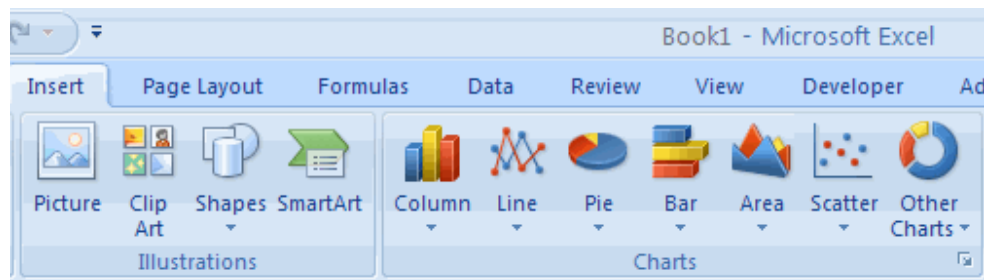
## CHARTS

Charts allow us to present information contained in the worksheet in a graphic format. Excel offers many types of charts including: Column, Line, Pie, Bar, Area, Scatter and more. To view the charts available click the Insert Tab on the Ribbon.

### Create a Chart

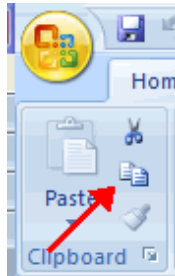
To create a chart:

- Select the **cells** that contain the data we want to use in the chart
- Click the **Insert** tab on the Ribbon
- Click the type of **Chart** we want to create



### Copy a Chart to Word

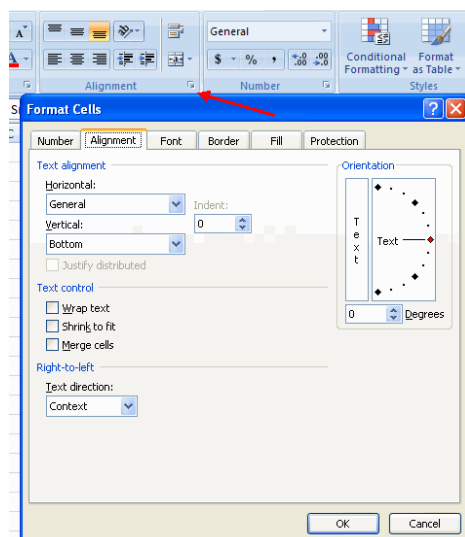
- Select the **chart**
- Click **Copy** on the **Home** tab
- Go to the **Word** document where you want the chart located
- Click **Paste** on the **Home** tab



### Format Cells Dialog Box

In Excel, you can also apply specific formatting to a cell. To apply formatting to a cell or group of cells:

- Select the cell or cells that will have the formatting
- Click the **Dialog Box** arrow on the **Alignment** group of the **Home** tab



There are several tabs on this dialog box that allow you to modify properties of the cell or cells.

**Number:** Allows for the display of different number types and decimal places

**Alignment:** Allows for the horizontal and vertical alignment of text, wrap text, shrink text, merge cells and the direction of the text.

**Font:** Allows for control of font, font style, size, color, and additional features

**Border:** Border styles and colors

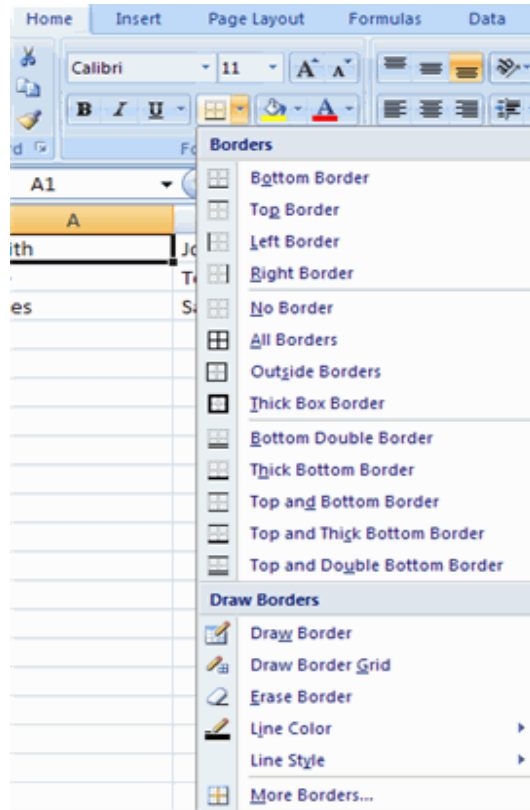
**Fill:** Cell fill colors and styles

### Add Borders and Colors to Cells

Borders and colors can be added to cells manually or through the use of styles. To add borders manually:

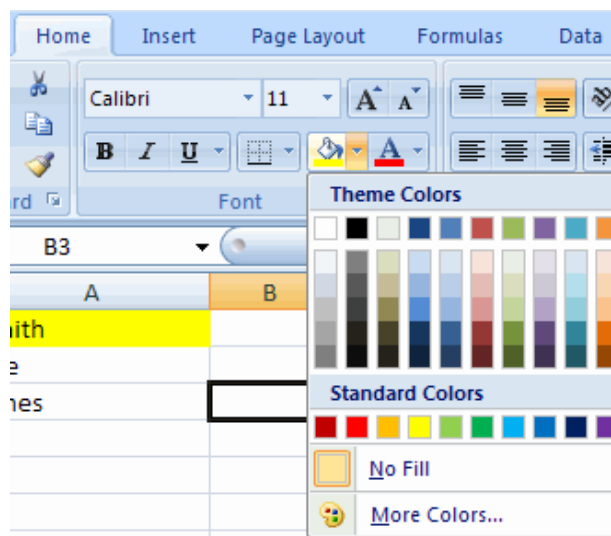
- Click the **Borders** drop down menu on the **Font** group of the **Home** tab

- Choose the appropriate border



To apply colors manually:

- Click the **Fill** drop down menu on the **Font** group of the **Home** tab
- Choose the appropriate color



To apply borders and colors using styles:

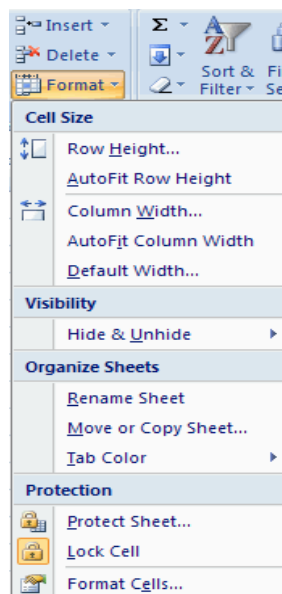
- Click **Cell Styles** on the **Home** tab
- Choose a style or click **New Cell Style**



### Change Column Width and Row Height

To change the width of a column or the height of a row:

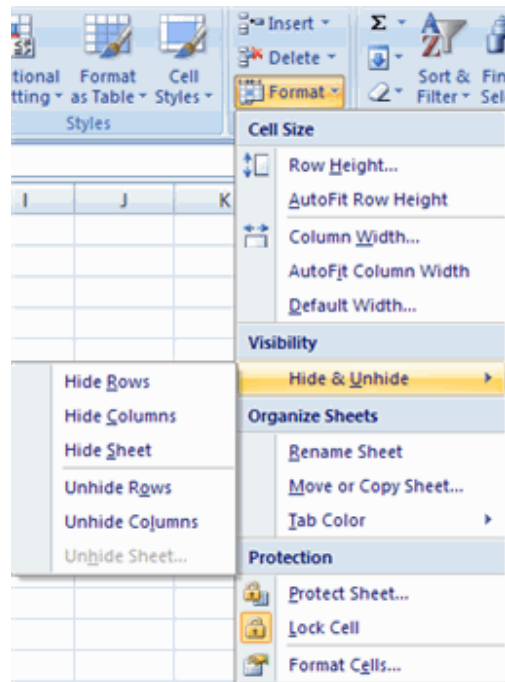
- Click the **Format** button on the **Cells** group of the **Home** tab
- Manually adjust the height and width by clicking **Row Height** or **Column Width**
- To use **AutoFit** click **AutoFit Row Height** or **AutoFit Column Width**



## Hide or Unhide Rows or Columns

To hide or unhide rows or columns:

- Select the row or column you wish to hide or unhide
- Click the **Format** button on the **Cells** group of the **Home** tab
- Click **Hide & Unhide**



## Merge Cells

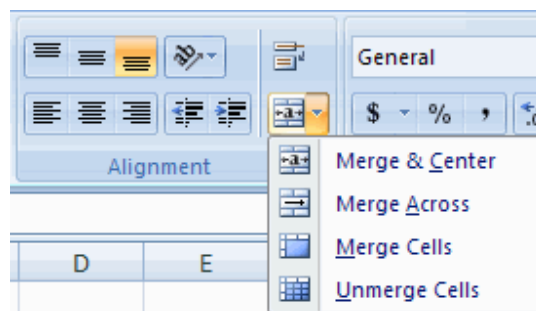
To merge cells select the cells you want to merge and click the **Merge & Center** button on the **Alignment** group of the **Home** tab. The four choices for merging cells are:

**Merge & Center:** Combines the cells and centers the contents in the new, larger cell

**Merge Across:** Combines the cells across columns without centering data

**Merge Cells:** Combines the cells in a range without centering

**Unmerge Cells:** Splits the cell that has been merged



## Align Cell Contents

To align cell contents, click the cell or cells you want to align and click on the options within the **Alignment** group on the **Home** tab. There are several options for alignment of cell contents:

**Top Align:** Aligns text to the top of the cell

**Middle Align:** Aligns text between the top and bottom of the cell

**Bottom Align:** Aligns text to the bottom of the cell

**Align Text Left:** Aligns text to the left of the cell

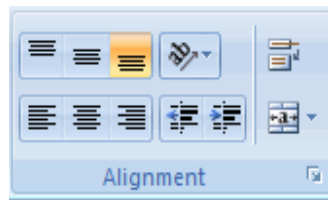
**Center:** Centers the text from left to right in the cell

**Align Text Right:** Aligns text to the right of the cell

**Decrease Indent:** Decreases the indent between the left border and the text

**Increase Indent:** Increase the indent between the left border and the text

**Orientation:** Rotate the text diagonally or vertically



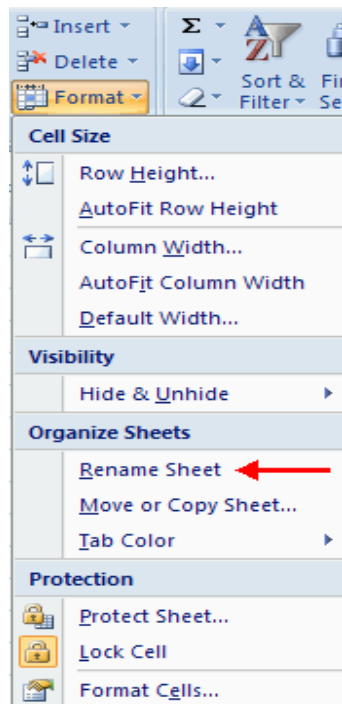
**Developing a Workbook**

**Printing »**

## Format Worksheet Tab

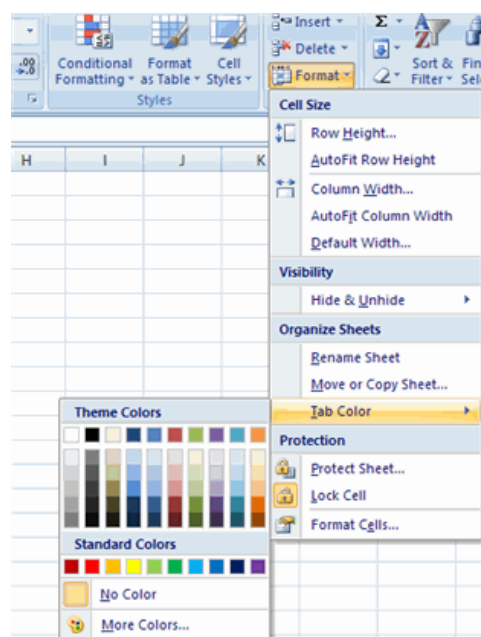
You can rename a worksheet or change the color of the tabs to meet your needs. To rename a worksheet:

- Open the sheet to be renamed
- Click the **Format** button on the **Home** tab
- Click **Rename** sheet
- Type in a new name
- Press **Enter**



To change the color of a worksheet tab:

- Open the sheet to be renamed
- Click the **Format** button on the **Home** tab
- Click **Tab Color**
- Click the color

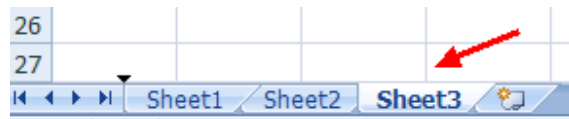




## Reposition Worksheets in a Workbook

To move worksheets in a workbook:

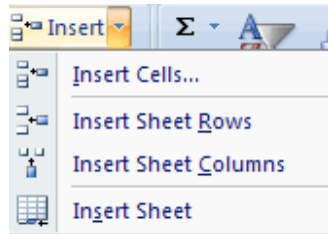
- Open the workbook that contains the sheets you want to rearrange
- **Click and hold** the worksheet tab that will be moved until an arrow appears in the left corner of the sheet
- **Drag** the worksheet to the desired location



## Insert and Delete Worksheets

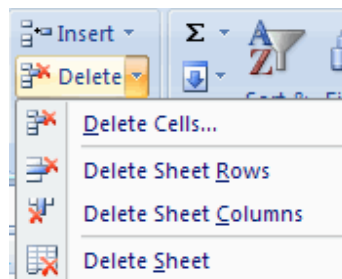
To insert a worksheet

- Open the workbook
- Click the **Insert** button on the **Cells** group of the **Home** tab
- Click **Insert Sheet**



To delete a worksheet

- Open the workbook
- Click the **Delete** button on the **Cells** group of the **Home** tab
- Click **Delete Sheet**

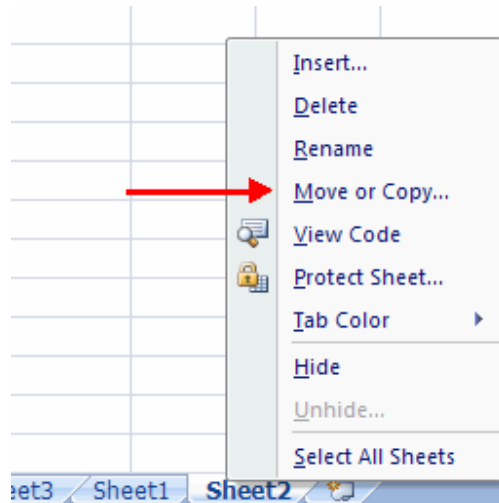


## Copy and Paste Worksheets

To copy and paste a worksheet:

- Click the tab of the worksheet to be copied
- **Right click** and choose **Move** or **Copy**

- Choose the desired position of the sheet
- Click the check box next to **Create a Copy**
- Click **OK**



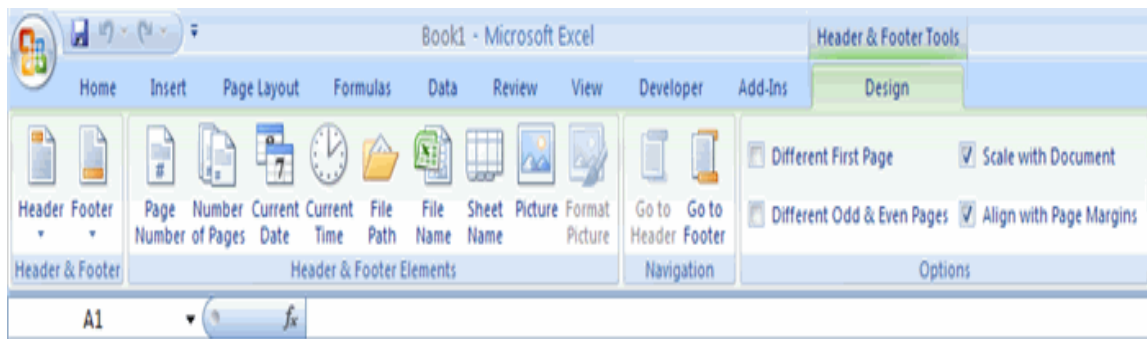
## Page Properties and Printing

## Layout »

### Create a Header or Footer

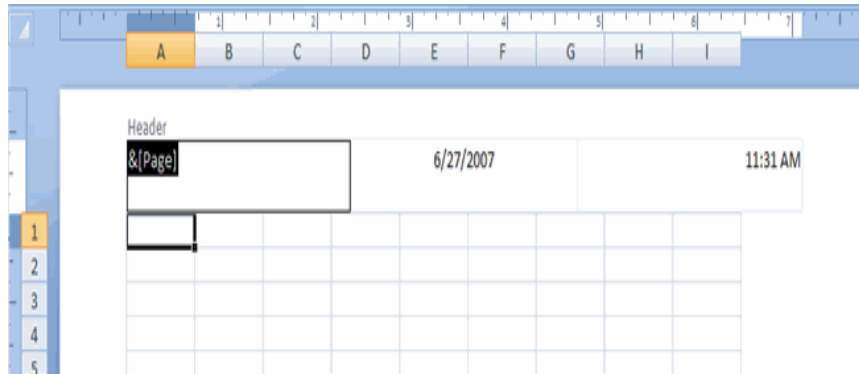
To create a header or footer:

- Click the **Header & Footer** button on the **Insert** tab
- This will display the **Header & Footer Design Tools Tab**
- To switch between the Header and Footer, click the **Go to Header** or **Go to Footer** button



- To insert text, enter the text in the header or footer
- To enter preprogrammed data such as page numbers, date, time, file name or sheet name, click the appropriate button

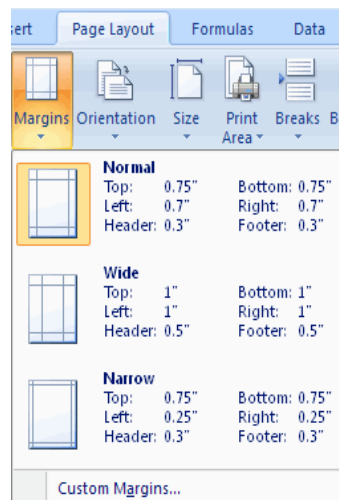
- To change the location of data, click the desired cell



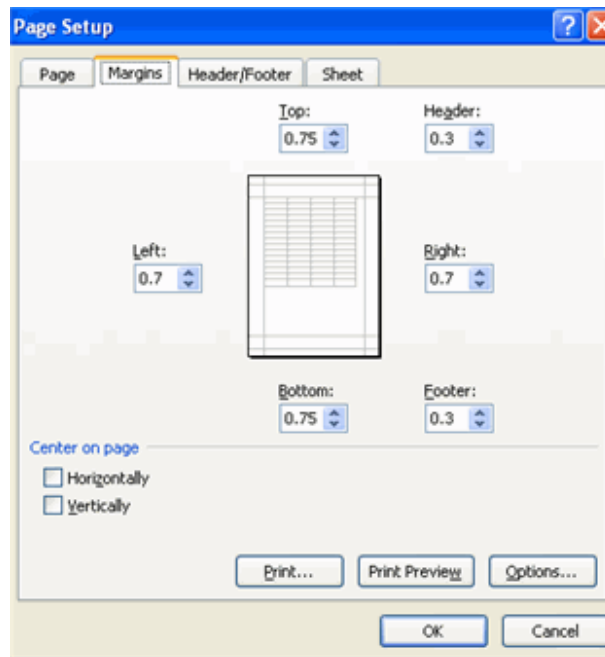
### Set Page Margins

To set the page margins:

- Click the **Margins** button on the **Page Layout** tab
- Select one of the give choices, or



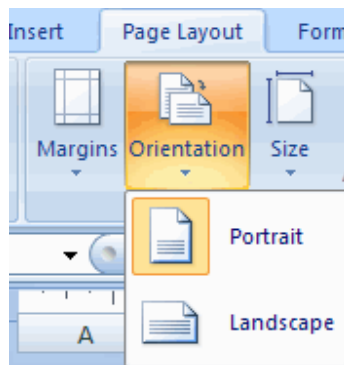
- Click **Custom Margins**
- Complete the boxes to set margins
- Click **Ok**



### Change Page Orientation

To change the page orientation from portrait to landscape:

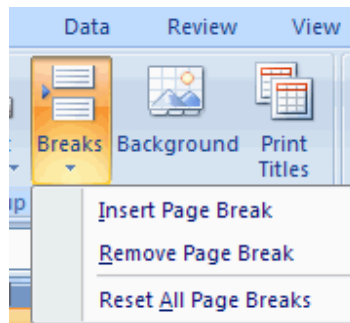
- Click the **Orientation** button on the **Page Layout** tab
- Choose **Portrait** or **Landscape**



### Set Page Breaks

You can manually set up page breaks in a worksheet for ease of reading when the sheet is printed. To set a page break:

- Click the **Breaks** button on the **Page Layout** tab
- Click **Insert Page Break**



### Print a Range

There may be times when you only want to print a portion of a worksheet. This is easily done through the Print Range function. To print a range:

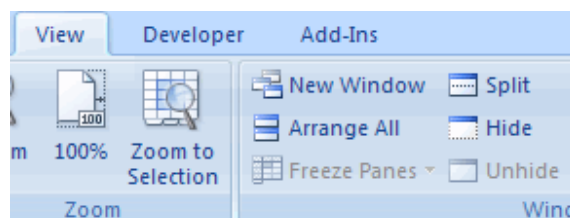
- Select the area to be printed
- Click the **Print Area** button on the **Page Layout** tab
- Click **Select Print Area**



### Split a Worksheet

You can split a worksheet into multiple resizable panes for easier viewing of parts of a worksheet. To split a worksheet:

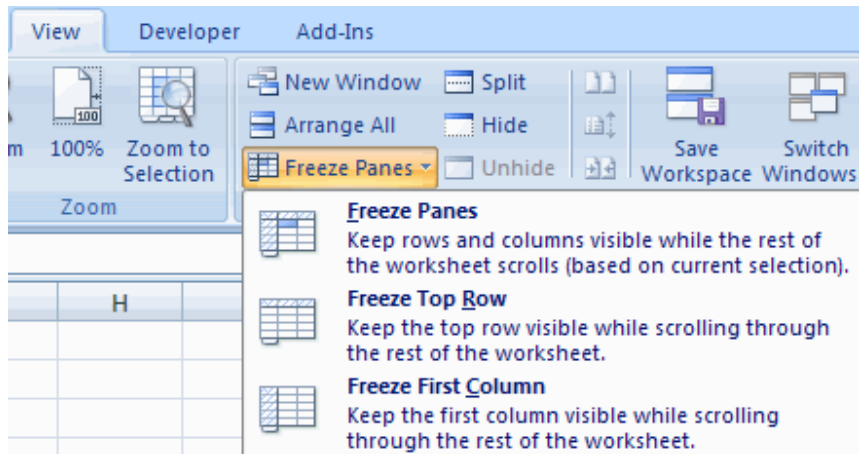
- Select any cell in center of the worksheet you want to split
- Click the **Split** button on the **View** tab
- Notice the split in the screen, you can manipulate each part separately



### Freeze Rows and Columns

You can select a particular portion of a worksheet to stay static while you work on other parts of the sheet. This is accomplished through the Freeze Rows and Columns Function. To Freeze a row or column:

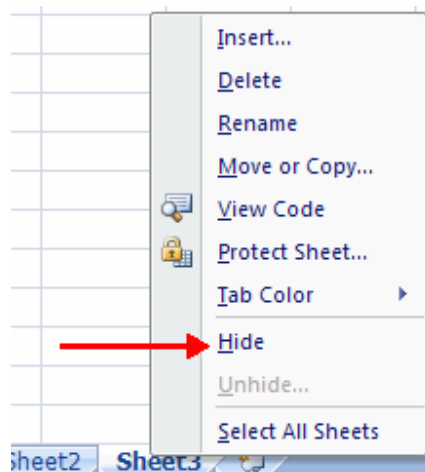
- Click the **Freeze Panes** button on the **View** tab
- Either select a section to be frozen or click the defaults of top row or left column
- To unfreeze, click the **Freeze Panes** button
- Click **Unfreeze**



### Hide Worksheets

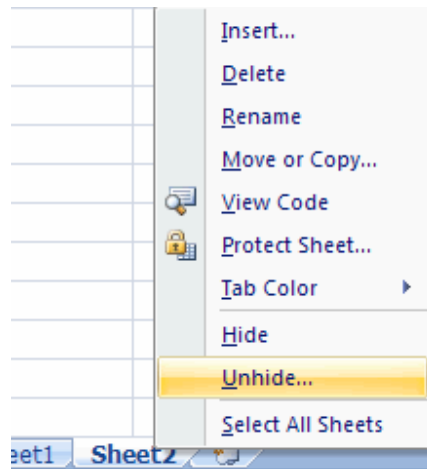
To hide a worksheet:

- Select the tab of the sheet you wish to hide
- **Right-click** on the tab
- Click **Hide**



To unhide a worksheet:

- **Right-click** on any worksheet tab
- Click **Unhide**
- Choose the worksheet to unhide



### Questions

1. A row in Excel is identified by \_\_\_\_\_

**Ans: numbers**

2. Excel has \_\_\_\_\_ sheets by default

**Ans: 3**

3. The columns in excel are named using the numbers.

**Ans: False**

4. An excel file is saved with xls extension in Office 2007.

**Ans: False**

5. Excel is a spread sheet software.

**Ans: True**

6. The intersection of a row and column is called a cell in Excel.

**Ans: True**

7. Name two charts in Excel.

8. What is the use of Wrap text.

9. Manipulate the following data using the Mathematical functions (MAX, COUNT, STDEV, AVERAGE, SIN) in MS-EXCEL

169 129 186 143 214 170 136 156

10. Represent the following data using charts in MS EXCEL

i) Bar Chart ii) Pie Chart

Cropping pattern in TAMILNADU

<b>CROPS</b>	<b>AREA (in 1,000 Hectares)</b>
Paddy	5432
Coconut	1365
Pulses	6856
Cotton	1239
Others	629



## Lecture.27

**Control statements – GOTO, IF statements– Loop Statement – FOR and NEXT statement - subscripted variable – simple programmes to calculate mean, standard deviation, simple correlation and simple linear regression 1.equation using above statement**

### POWERPOINT

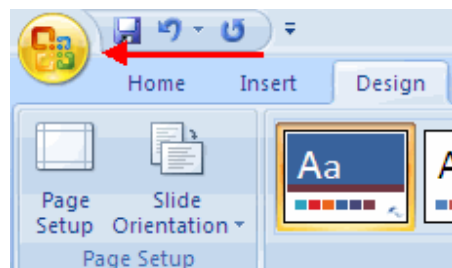
There are three features that you should remember as you work within PowerPoint 2007: the Microsoft Office Button, the Quick Access Toolbar, and the Ribbon. The function of these features will be more fully explored below.

#### Presentations

A presentation is a collection of data and information that is to be delivered to a specific audience. A PowerPoint presentation is a collection of electronic slides that can have text, pictures, graphics, tables, sound and video. This collection can run automatically or can be controlled by a presenter.

#### Microsoft Office Button

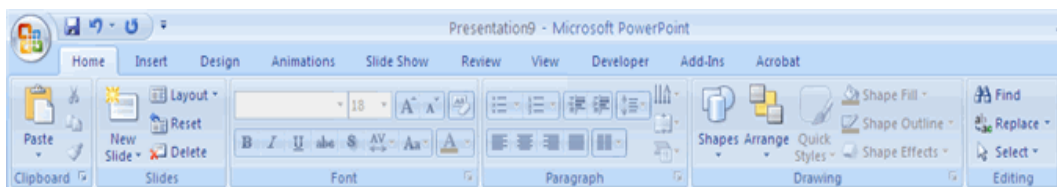
The Microsoft Office Button performs many of the functions that were located in the File menu of older versions of PowerPoint. This button allows you to create a new presentation, Open an existing presentation, save and save as, print, send, or close.



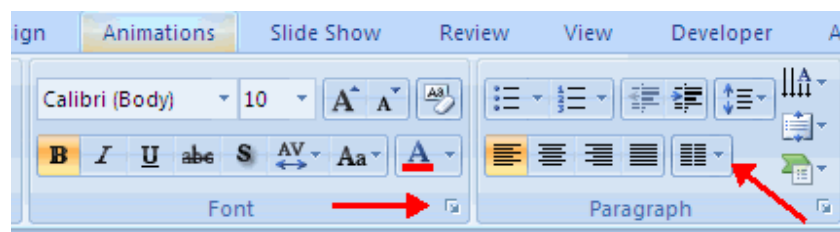
#### Ribbon

The ribbon is the panel at the top portion of the document. It has seven tabs: Home, Insert, Design, Animations, Slide Show, Review and View. Each tab is divided

into groups. The groups are logical collections of features designed to perform function that you will utilize in developing or editing your PowerPoint slides.



Commonly utilized features are displayed on the Ribbon. To view additional features within each group, click the arrow at the bottom right corner of each group.



**Home:** Clipboard, Slides, Font, Paragraph, Drawing, and Editing

**Insert:** Tables, Illustrations, Links, Text, and Media Clips

**Design:** Page Setup, Themes, Background

**Animations:** Preview, Animations, Transition to this Slide

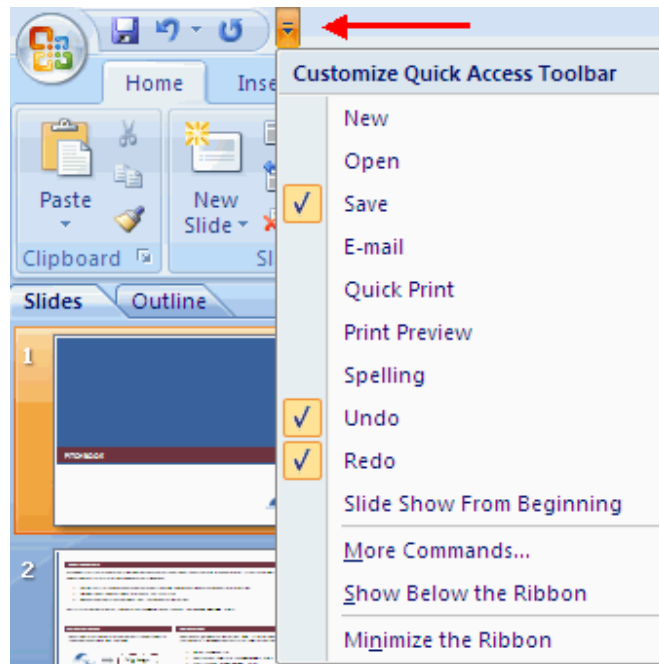
**Slide Show:** Start Slide Show, Set Up, Monitors

**Review:** Proofing, Comments, Protect

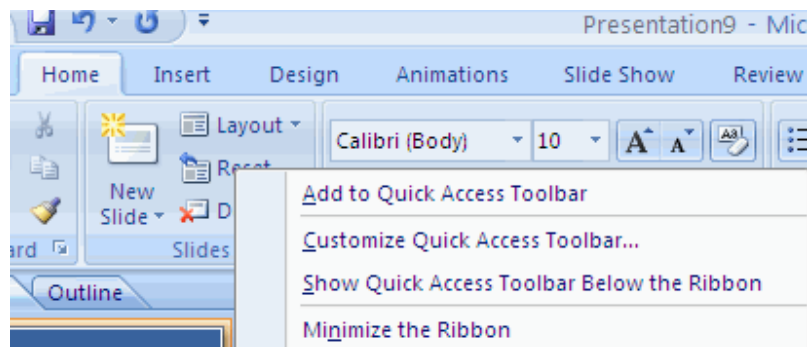
**View:** Presentation Views, Show/Hide, Zoom, Window, Macros

### Quick Access Toolbar

The **quick access toolbar** is a customizable toolbar that contains commands that you may want to use. You can place the quick access toolbar above or below the ribbon. To change the location of the quick access toolbar, click on the arrow at the end of the toolbar and click **Show Below the Ribbon**.

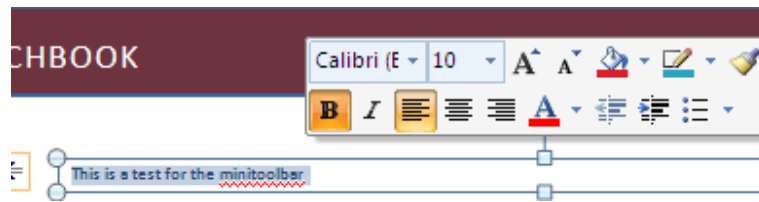


You can also add items to the quick access toolbar. Right click on any item in the Office Button or the Ribbon and click Add to Quick Access Toolbar and a shortcut will be added.



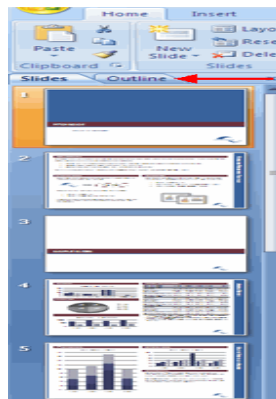
### **Mini Toolbar**

A new feature in Office 2007 is the Mini Toolbar. This is a floating toolbar that is displayed when you select text or right-click text. It displays common formatting tools, such as Bold, Italics, Fonts, Font Size and Font Color.



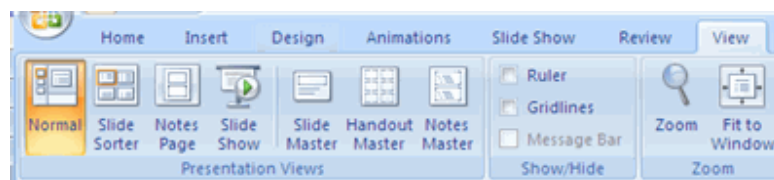
## Navigation

Navigation through the slides can be accomplished through the Slide Navigation menu on the left side of the screen. Also, an outline appears from materials that have been entered in the presentation. To access the outline, click the outline tab.



## Slide Views

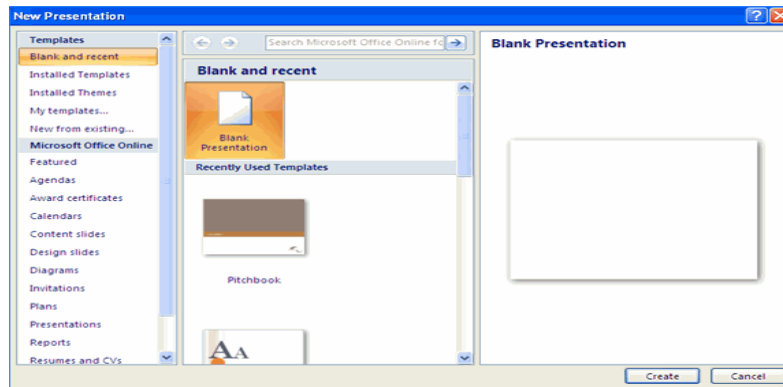
Presentations can be viewed in a variety of manners. On the View tab, the Presentation Views group allows you to view the slides as Normal, Slide Sorter, Notes Page, Slide Show, Slide Master, Handout Master, and Notes Master.



## New Presentation

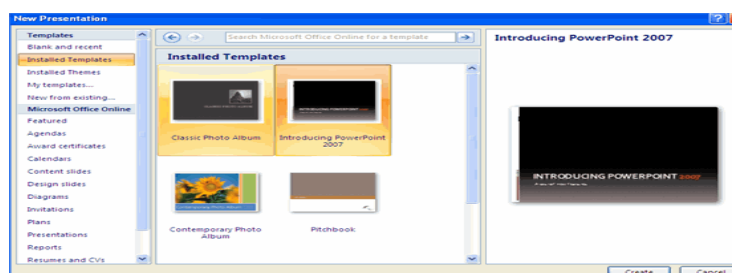
You can start a new presentation from a blank slide, a template, existing presentations, or a Word outline. To create a new presentation from a blank slide:

- Click the **Microsoft Office Button**
- Click **New**
- Click **Blank Presentation**



To create a new presentation from a template:

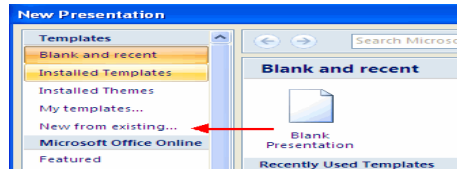
- Click the **Microsoft Office Button**
- Click **New**
- Click **Installed Templates** or **Browse through Microsoft Office Online Templates**
- Click the template you choose



To create a new presentation from an existing presentation:

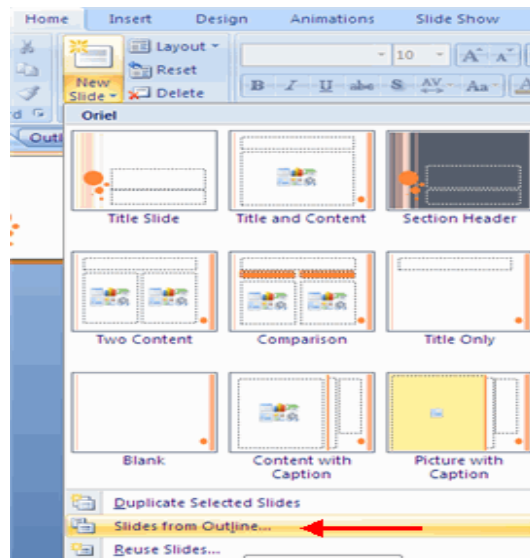
- Click the **Microsoft Office Button**
- Click **New**

- Click **New from Existing**
- Browse to and click the presentation



To create a new presentation from a Word outline:

- Click the slide where you would like the outline to begin
- Click **New Slide** on the **Home** tab
- Click **Slides from Outline**
- Browse and click the Word Document that contains the outline

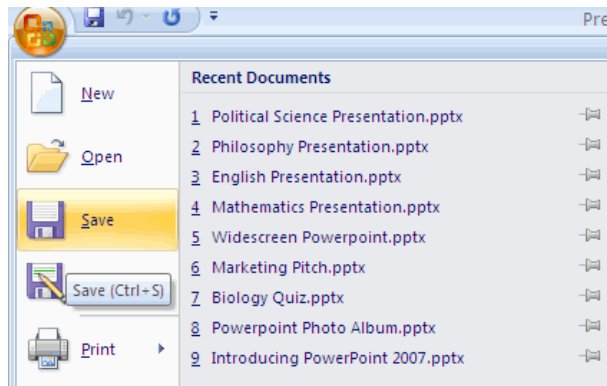


### Save a Presentation

When you save a presentation, you have two choices: **Save** or **Save As**.

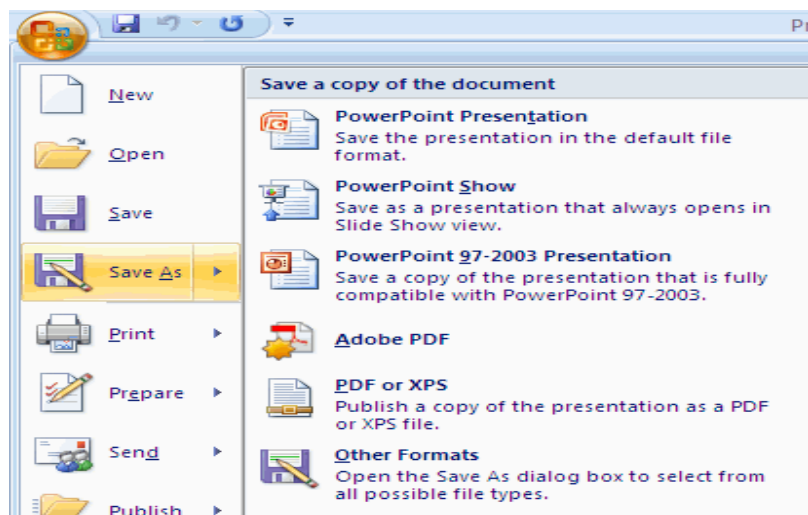
To save a document:

- Click the **Microsoft Office Button**
- Click **Save**



You may need to use the **Save As** feature when you need to save a presentation under a different name or to save it for earlier versions of PowerPoint. Remember that older versions of PowerPoint will not be able to open PowerPoint 2007 presentation unless you save it as a PowerPoint 97-2003 Format. To use the **Save As** feature:

- Click the **Microsoft Office Button**
- Click **Save As**
- Type in the name for the Presentation
- In the **Save as Type** box, choose **Excel 97-2003 Presentation**



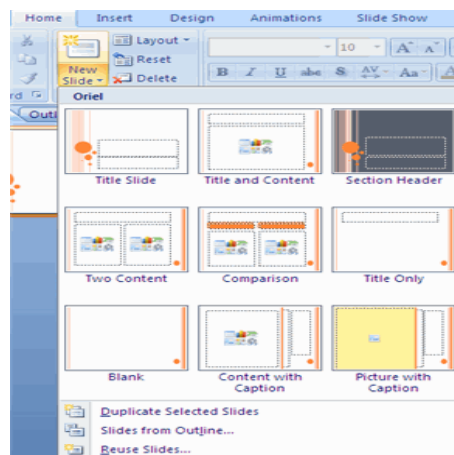
## Add Slides

There are several choices when you want to add a new slide to the presentation:

Office Themes, Duplicate Selected Slide, or Reuse Slides.

To create a new slide from Office Themes:

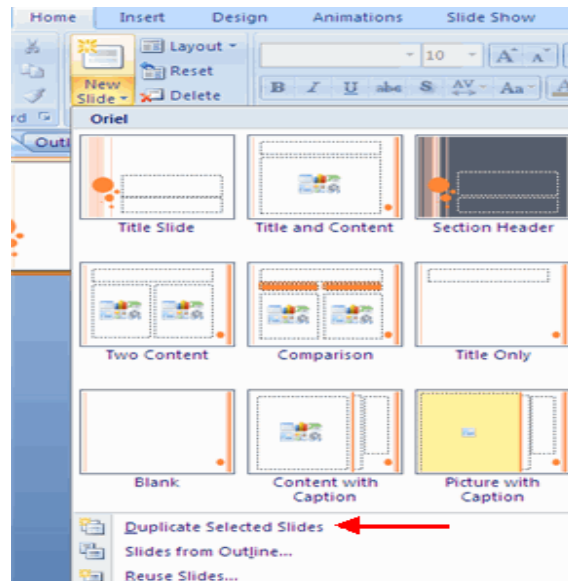
- Select the slide immediately **BEFORE** where you want the new slide
- Click the **New Slide** button on the **Home** tab
- Click the slide choice that fits your material



To create a slide as a duplicate of a slide in the presentation:

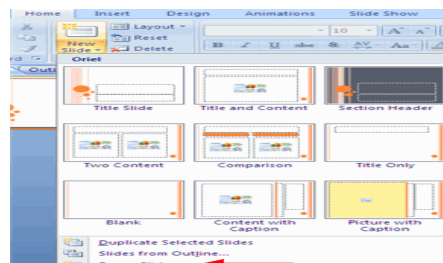
- Select the slide to duplicate
- Click the **New Slide** button on the **Home** tab
- Click **Duplicate Selected Slides**





To create a new slide from another presentation:

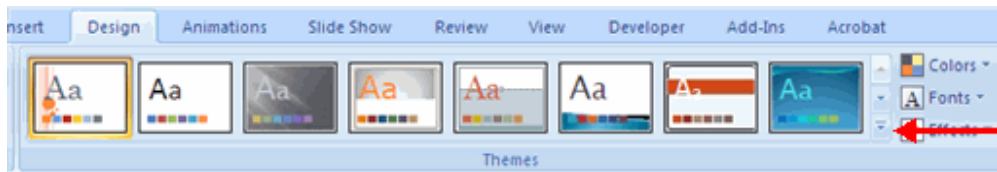
- Select the slide immediately BEFORE where you want the new slide
- Click the **New Slide** button on the Home tab
- Click **Reuse Slides**
- Click **Browse**
- Click **Browse File**
- Locate the slide show and click on the slide to import



## Themes

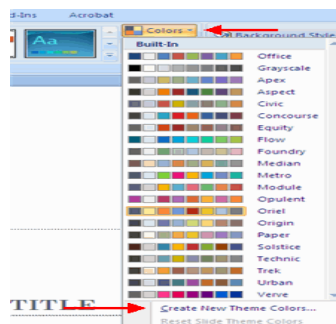
Themes are design templates that can be applied to an entire presentation that allows for consistency throughout the presentation. To add a theme to a presentation:

- Click the **Design** tab
- Choose one of the displayed **Themes** or click the **Galleries** button



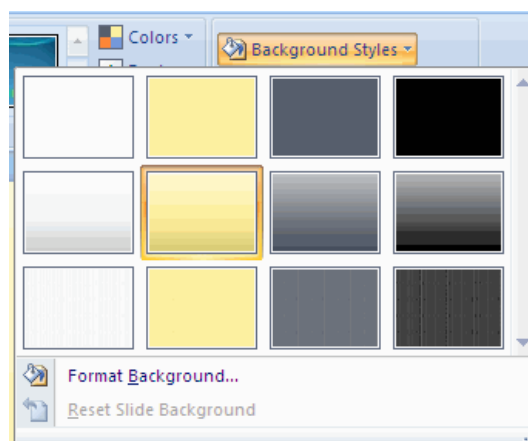
To apply new colors to a theme:

- Click the **Colors** drop down arrow
- Choose a color set or click **Create New Theme Colors**



To change the background style of a theme

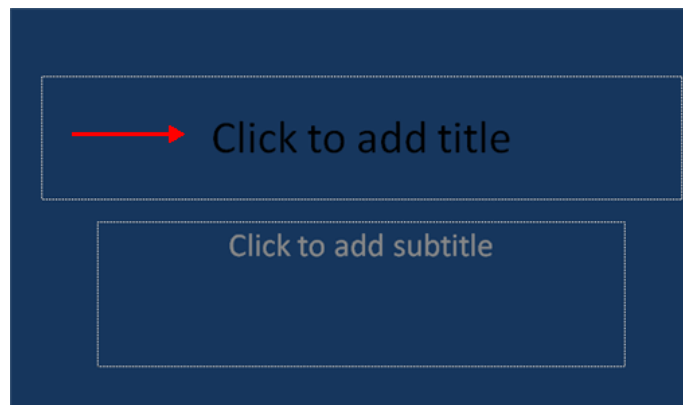
- Click the **Background Styles** button on the **Design** tab



## Enter Text

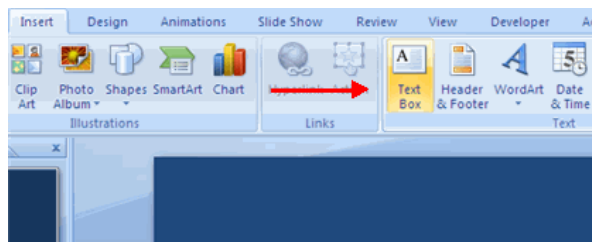
To enter text:

- Select the **slide** where you want the text
- Click in a **Textbox** to add text



To add a text box:

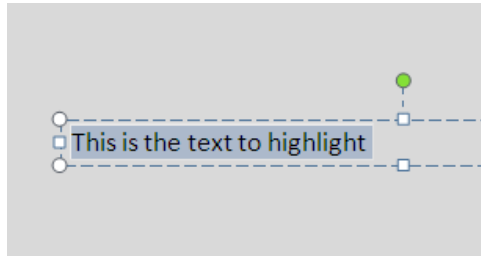
- Select the **slide** where you want to place the text box
- On the **Insert** tab, click **Text Box**
- Click on the slide and drag the cursor to expand the text box
- Type in the text



## Select Text

To select the text:

- Highlight the text



## Copy and Paste

To copy and paste data:

- Select the item(s) that you wish to copy
- On the **Clipboard Group** of the **Home Tab**, click **Copy**
- Select the item(s) where you would like to copy the data
- On the **Clipboard Group** of the **Home Tab**, click **Paste**



## Cut and Paste

To cut and paste data:

- Select the item(s) that you wish to copy
- On the **Clipboard Group** of the **Home Tab**, click **Cut**
- Select the items(s) where you would like to copy the data
- On the **Clipboard Group** of the **Home Tab**, click **Paste**



## Undo and Redo

To undo or redo your most recent actions:

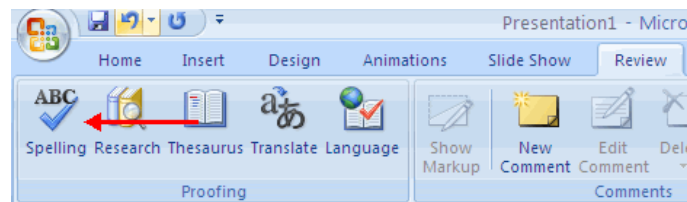
- On the **Quick Access Toolbar**
- Click **Undo** or **Redo**



## Spell Check

To check the spelling in a presentation:

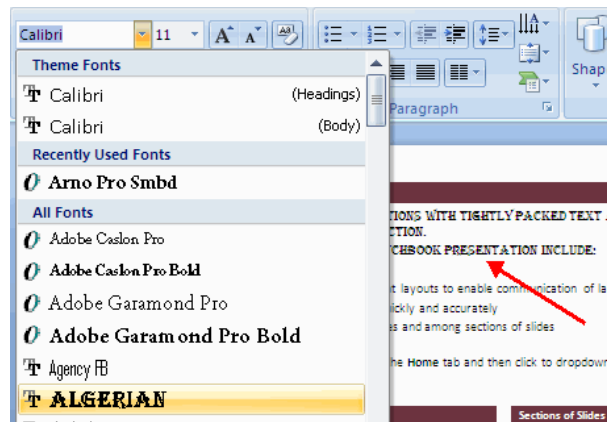
- Click the **Review** tab
- Click the **Spelling** button



## Change Font Typeface and Size

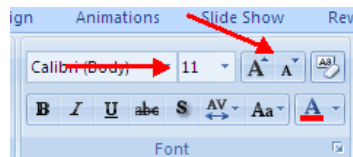
To change the font typeface:

- Click the **arrow** next to the font name and choose a font.
- Remember that you can preview how the new font will look by highlighting the text, and hovering over the new font typeface.



To change the font size:

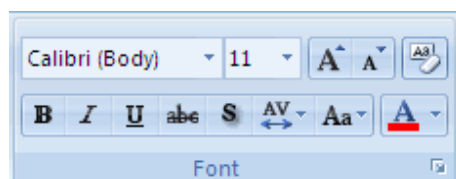
- Click the **arrow** next to the font size and choose the appropriate size, or
- Click the **increase or decrease** font size buttons.



### Font Styles and Effects

Font styles are predefined formatting options that are used to emphasize text. They include: Bold, Italic, and Underline. To add these to text:

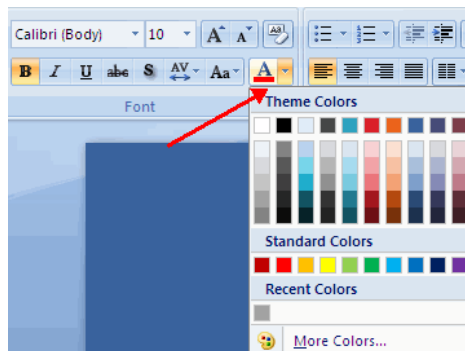
- Select the text and click the **Font Styles** included on the Font group of the Home tab or
- Select the text and right click to display the font tools



## Change Text Color

To change the text color:

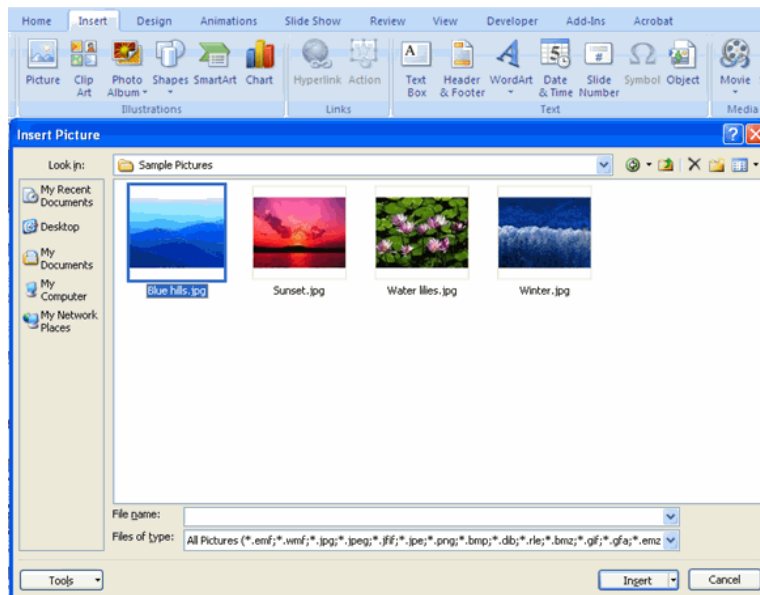
- Select the text and click the **Colors** button included on the Font Group of the Ribbon, or
- Highlight the text and right click and choose the colors tool.
- Select the color by clicking the down arrow next to the font color button.



## Adding Picture

To add a picture:

- Click the **Insert** Tab
- Click the **Picture** Button
- Browse to the picture from your files
- Click the **name** of the picture
- Click **insert**
- To move the graphic, click it and drag it to where you want it

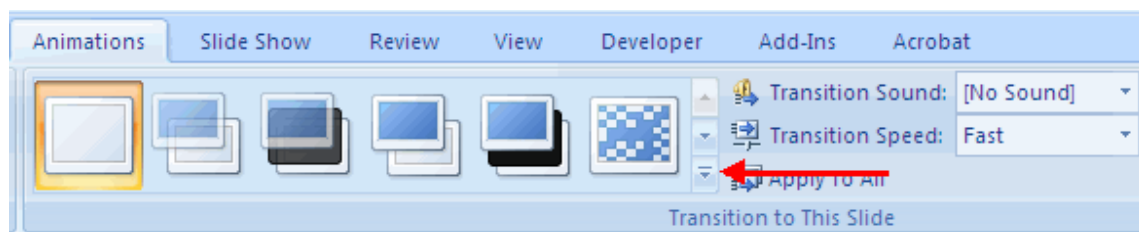


## Slide Effects

### Slide Transitions

Transitions are effects that are in place when you switch from one slide to the next. To add slide transitions:

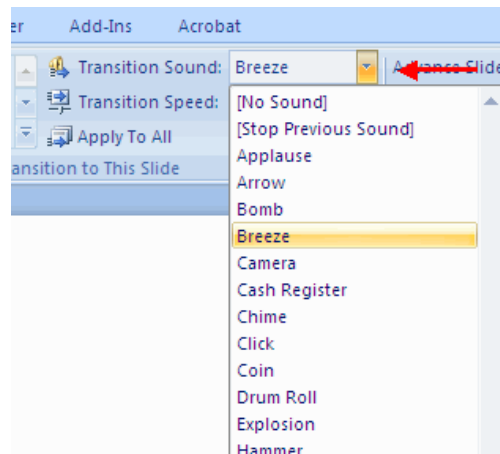
- Select the slide that you want to transition
- Click the **Animations** tab
- Choose the appropriate animation or click the **Transition** dialog box



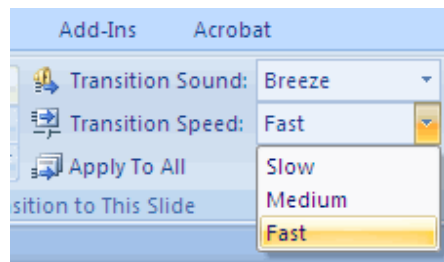
To adjust slide transitions:

- Add sound by clicking the arrow next to **Transition Sound**



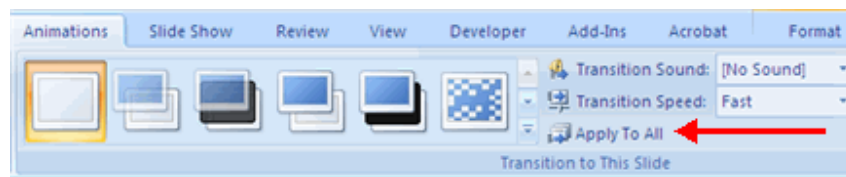


- Modify the transition speed by clicking the arrow next to **Transition Speed**



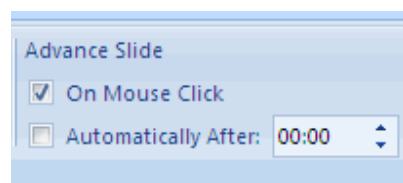
To apply the transition to all slides

- Click the **Apply to All** button on the **Animations** tab



To select how to **advance a slide**

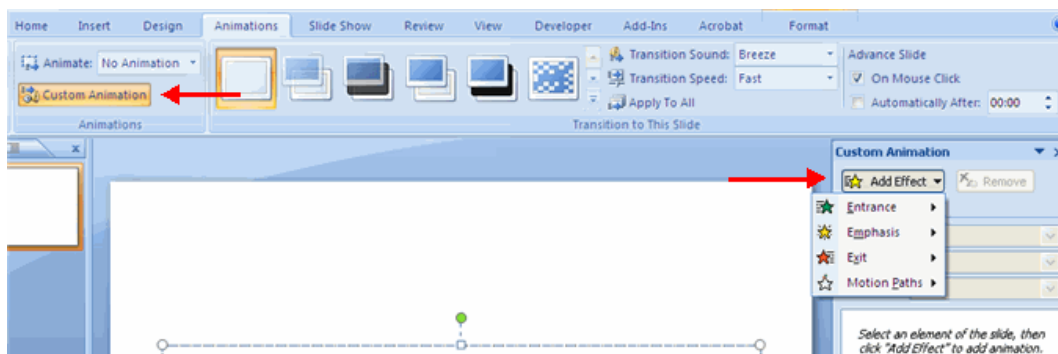
- Choose to **Advance on Mouse Click**, or
- Automatically after a set number of seconds



## Slide Animation

Slide animation effects are predefined special effects that you can add to objects on a slide. To apply an animation effect:

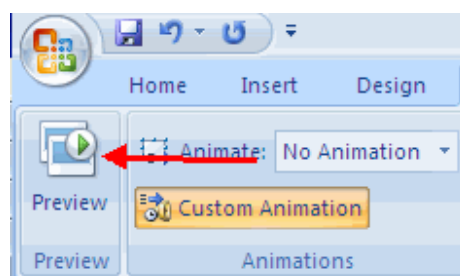
- Select the object
- Click the **Animations** tab on the Ribbon
- Click **Custom Animation**
- Click **Add Effect**
- Choose the appropriate effect



## Animation Preview

To preview the animation on a slide:

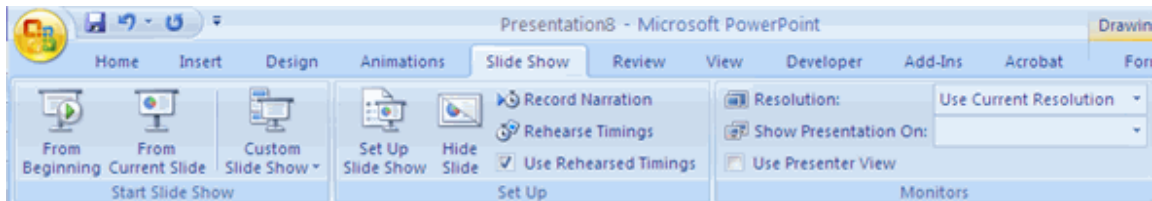
- Click the **Preview** button on the **Animations** tab



## Slide Show Options

The Slide Show tab of the ribbon contains many options for the slide show. These options include:

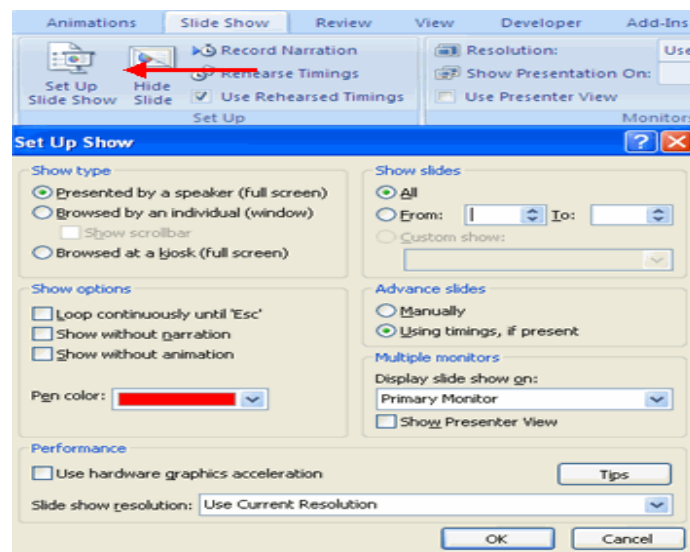
- Preview the slide show from the beginning
- Preview the slide show from the current slide
- Set up Slide Show



### Set Up Slide Show

This option allows you to set preferences for how the slide show will be presented. The options include:

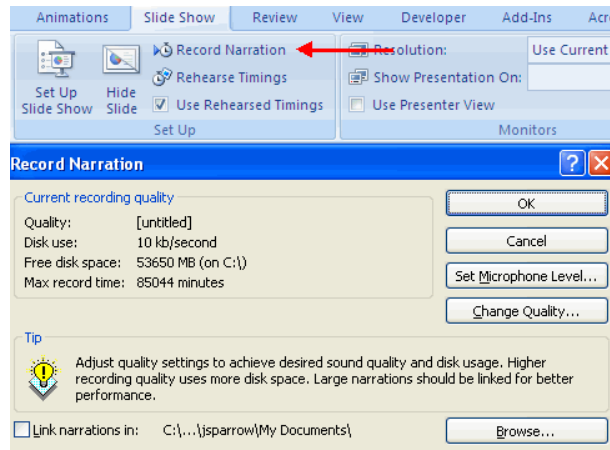
- Whether the show will run automatically or will be presented by a speaker
- The looping options
- Narration options
- Monitor resolutions



### Record Narration

When you want to record narration for the slides:

- Click the **Record Narration** button
- Click **Set Microphone Level** to check the levels of audio input
- Click **OK** to record the narration



## Rehearse Timings

Use Rehearsed Timings to rehearse the timings of slide with audio.

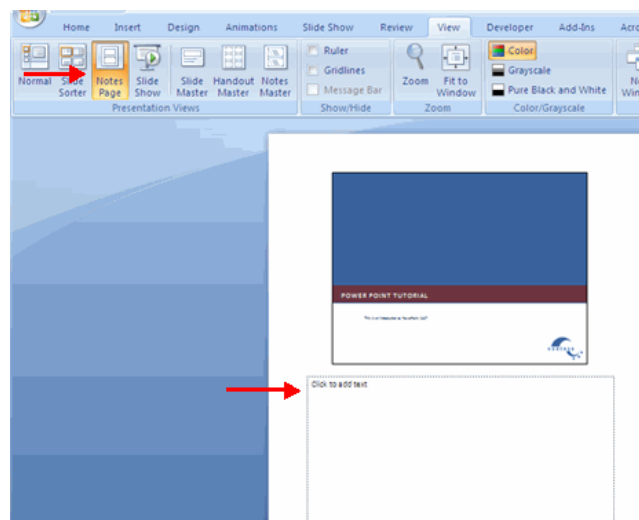
- Click the Rehearse Timings button
- Practice speaking and advance the slides as you would in the presentation
- When you have completed this click through the end of the slide
- Choose whether or not to keep this timing or to retry



## Create Speaker Notes

Speaker Notes can be added to allow you to create notes for each slide. To add speaker notes:

- Select the slide
- Click **View**
- Click **Note Pages**
- Click the **Click to add Notes** section of the screen
- Type in the **Notes** for that slide



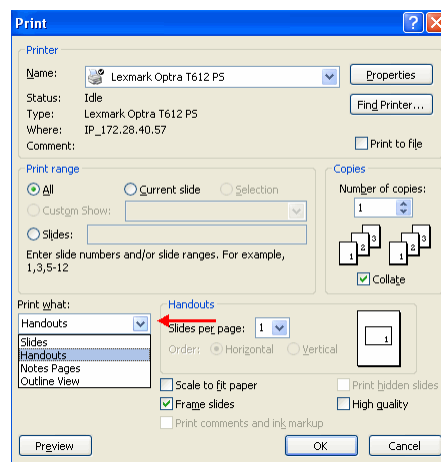
## Print a Presentation

There are many options for printing a presentation. They are:

- **Slides:** These are slides that you would see if you were showing the presentation, one slide per page
- **Handouts:** 1, 2, 3, 4, 6 or 9 per page, this option allows for more slides per page
- **Notes Page:** This includes the slides and the speaker notes
- **Outline View:** This will print the outline of the presentation

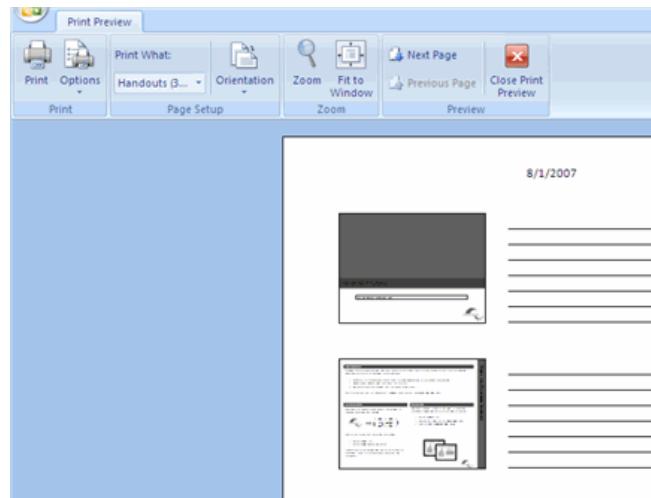
To access the print options

- Click the **Microsoft Office Button**
- Click **Print**
- In the Print Dialog Box, click the **arrow** next to **Print what**
- Choose the format and click **OK** to print



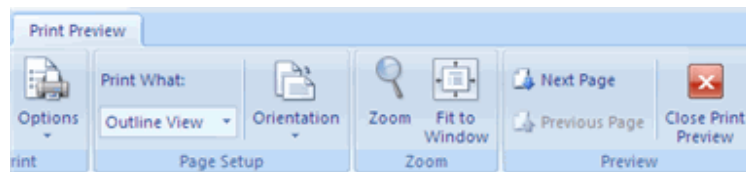
To print preview

- Click the **Microsoft Office Button**
- Place the cursor over **Print**
- Click **Print Preview**
- Click the arrow next to **Print What** to change print options
- To print from Print Preview, click **Print**



To Exit Print Preview

- Click the **Close Print Preview** button



## Questions

1. \_\_\_\_\_ is the short cut key to add a new slide in Mspowerpoint

**Ans: ctrl+M**

2. \_\_\_\_\_ key is used for slide show in MSPowerpoint

**Ans: F5**

3. The view that displays the slides on a presentation as miniature representations of the slides is called

- slide show
- slide sorter view
- notes page view

d. outline view

**Ans: slide sorter view**

4. What are symbols used to identify items in a list?

a. Icons

b. Markers

c. Bullets

d. Graphics

**Ans: Bullets**

5. Which command brings you to the first slide in your presentation?

a. Next slide button

b. Page up

c. Ctrl + home

d. Ctrl + end

**Ans: Ctrl + home**

6. Presentation designs regulate the formatting and layout for the slide and are commonly called

a. Design templates

b. Templates

c. Placeholders

d. Blueprints

**Ans: Templates**

7. Which of the following format options should be used to display dollars

a. Normal

b. Percentage



c. Currency

d. Comma

**Ans: Currency**

8. Prepare a power point presentation with a minimum of five slides.

9. Insert a picture and chart in a presentation.

10. How to prepare a presentation and set animations.

## **Lecture.28**

### **MS Office – introduction – MS WORD – creating a simple document – editing and printing of a document**

#### **BASIC**

##### **Beginners' All Purpose Symbolic Instruction Code**

The computer needs instructions to perform an operation. Instructions to the computer are provided with the help of a programming language.

The language whose design is governed by the circuitry and structure of the machine is known as Machine Language.

The computer cannot understand source language (source code) and they need to be translated into the machine language (machine code) with the help of other programs known as compiler or interpreter.

##### **Compiler**

A compiler is a program which can translate the BASIC program to machine code. It reads statement by statement and checks for the syntax. If the program has no mistake then it converts the program to machine code and stores it separately.

##### **Interpreter**

A program that translates and executes source language statements one line at a time. Compiled programs generally run faster than interpreted programs whereas interpreter translates every time. BASIC is especially designed to be executed by an interpreter.

##### **BASIC Language**

BASIC is one of the easiest and simplest high level programming language. BASIC was first designed and developed in the mid of 1960's at Dartmouth college, U.S.A. under the direction of J.G.Kemeny and T.E.Kurtz. Due to its simplicity and flexibility, its use has spread widely among users.

## Statement and Keywords

A statement specifies a step of a program. Every statement consists of some specific BASIC word(s). Such a word is called keyword. For eg. DATA, READ, LET, PRINT etc. A keyword specifies specific information.

A statement is always formed by using a keyword. Generally, the format of any BASIC statement is

l	Keyword	Specification
---	---------	---------------

Hence l indicates the statement or the line number. It is also referred to as the label of the statement. The value of label must be unsigned integer number and can range from 1 to 99,999.

Eg: 10 PRINT A,B,C,D

## Elements of BASIC Language

### 1. Character set

A BASIC language, as any other language, has its own alphabet or character set. All the quantities defined and used in BASIC are constructed by using the character set.

The character set in BASIC are

Letter → A, B, C... Z

a, b, c ... z

Digits → 0, 1, 2 ... 9

Special Characters → + - \* / ( ) = . , \$ < > ;

Blanks → In the text, we shall denote a blank by the character ^.

Eg. 30 PRINT "GOVIND"; "^"; "SINGH"

Quote → '(Single Quote) or "(Double Quote)

[; → different quantities appearing in the list of the PRINT statement can be separated by semicolons]

## **2. Constants and Variables**

### **Constants**

The quantities whose value cannot be changed in a program are referred to as constants.

There are two types of constants.

- a) Numeric Constants.
- b) Non-numeric or Literal or String constants.

#### **a) Numeric constant**

A constant consists of the number 0, 1, 2 ... 9

There are two types

- i) Integer Constant
- ii) Real Constant

#### **i) Integer Constant**

A number without an explicit decimal is called integer constant (i.e. whole number).

Eg. 1, 20, 300.

#### **ii) Real Constant**

The number includes a decimal point.

Eg. 22.33.

.33 is called its fractional part and 22 is its integer part. The number can be positive or negative. The sign always precedes the number. If the value is a positive number, it need not be specified as + 12.12. By default if no sign is appearing, then it is taken as a positive quantity.

#### **b) String Constant**

Letters and Digits enclosed with quotes.

Eg. "345", "TNAU"

## **Variables**

Its value changes during execution. A variable must have a name. The name refers to the memory location where the value of the variable is stored. The name of the variable may consist of one to eight characters. The first character must be an alphabet and may be followed by alphabets or numerals.

There are two types

- i) Numeric variables
- ii) String or Non-numeric variables.

When the value is numeric, it is called numeric or Arithmetic variable and can be used in arithmetic computations. When the value is non-numeric, it is called string or non-numeric variable and cannot be used in arithmetic calculations. A string variable must consist of a name followed by a dollar sign (\$).

Eg. LET A\$ = "PAUL".

The string variable should be assigned a string constant only, while the arithmetic variable is assigned a numeric constant only.

## **Some BASIC functions**

There are two types of functions

- a) Pre-defined functions
- b) User-defined functions

## **Pre-defined functions**

Pre-defined functions are also referred to as Library functions and are supplied as a part of the BASIC system.

The form is

name (argument)

where the name is pre-defined and must be specified as given.

The argument may be a constant, a variable, another function or a combination of these. The argument must be enclosed in parenthesis.

Function	Definition
SQR(X)	$\sqrt{x}$
EXP(X)	$e^x$
LOG(X)	$\log(x)$
SIN(X)	$\sin(x)$
COS(X)	$\cos(x)$
TAN(X)	$\tan(x)$

### User defined functions

The user can define his own functions with the help of a DEF (DEFINE) statement. The form of this statement is

---

1 DEF FNname(argument) =expression

---

The name is a user define name, FN indicates that the name refers to function and is to be specified.

Eg : 32 DEF FNA(R) = 3.1415 \*R  $\uparrow$  +1.5

33 ....

34 ....

35 X= FNA(3)

36 ....

### Priority rules to arithmetic operators or hierarchy of arithmetic operations

1. The exponential is performed first. When more than one operator is used in an expression and no parentheses are specified, evaluation is performed form left to right.

Thus  $3 + 2 \cdot 6/3$  is first reduced to  $3 + 6 \cdot 4/3$  and then to 24.33

2. Multiplication and division are done after exponentiation. When two or more operators appear in sequence and parentheses are not used, the evaluation of the expression proceeds from left to right. Thus the expression  $4 * 5 / 2 * 8$  is first reduced to  $20 / 2 * 8$  then to  $10 * 8$  finally to 80.

3. Addition and subtraction are performed last of all.

Eg :

$$\begin{array}{r} 1) \quad 4 \uparrow 3 / 8 * 4 \\ \quad \quad 64 / 8 * 4 \\ \quad \quad \quad 8 * 4 \\ \quad \quad \quad \quad =32 \end{array}$$

$$2) \quad 8 / 4 / 2 * 6 + 4 - 2.5$$

$$\quad \quad \square$$

$$\quad \quad 2 / 2 * 6 + 4 - 2.5$$

$$\quad \quad \square$$

$$\quad \quad 1 * 6 + 4 - 2.5$$

$$\quad \quad \square$$

$$\quad \quad 6 + 4 - 2.5$$

$$\quad \quad \square$$
$$10 - 2.5 \quad = 7.5$$

When parenthesis is used, the number of opening parentheses must be equal to the number of the closing parentheses. Parentheses may be contained within the other. This is called **Nesting**.

For eg:

$$\begin{array}{c} ( X + ( ( A + B ) * C / D ) - D * Q ) \\ \quad \quad \quad (1) \\ \quad \quad \quad \quad (2) \\ \quad \quad \quad \quad \quad (3) \end{array}$$

Parentheses (1) is nested within parentheses (2) while (2) is nested in (3). First the expression in (1) is evaluated, then the expression in (2) is computed and last of all the contents of (3) are calculated.

### Relational Expressions

BASIC allows the construction and use of relational expressions and are formed by using the relational operators. These operators are

<b>Relational Operator</b>	<b>Meaning</b>
=	Equal
<>	Not equal
>	Greater than
<	less than
>=	Greater than or equal
<=	less than or equal

Relational expressions are constructed by combining two arithmetic expressions with relational operators.

The general form is

$$\underline{\text{Arithmetic\_expression\_1} \oplus \text{arithmetic\_expression\_2}}$$

Where  $\oplus$  indicates a relational operator.

Eg; A=2, B=3

A = B ,      A > B      A < B  
False      False      True

Relational expression are commonly used to test conditions in IF–THEN ELSE statements.

### **Logical Expressions**

Logical expressions are designed using logical operators and relational expressions. Commonly used logical operators are NOT, AND, OR, XOR, EQV. The following tables explain their action. Such tables are called **truth tables**.

#### **NOT Table**

<b>A</b>	<b>NOT A</b>
True	False
False	True

#### **AND Table**

<b>A</b>	<b>B</b>	<b>A AND B</b>
True	True	True
True	False	False
False	True	False
False	False	False



### OR Table

A	B	A OR B
True	True	True
True	False	True
False	True	True
False	False	False

### XOR table

A	B	A XOR B
True	True	False
True	False	True
False	True	True
False	False	False

### EQV table

A	B	A EQV B
True	True	True
True	False	False
False	True	False
False	False	True

### Example

1.  $20 < 15 \text{ OR } 5 > 4$

F OR T  $\longrightarrow$  True

2.  $6 * 5 <> 10 * 3 \text{ AND } 7 < 5$

F AND F  $\longrightarrow$  False

Logical expressions may be used in the IF –THEN & IF –THEN –ELSE statements as test conditions like relational expressions.

The order of evaluation of the operators is as follows

	Operator	Priority
1.	$\uparrow$	Highest
2.	$*$ , $/$	
3.	$+$ , $-$	
4.	$=$	
5.	$>$	
6.	$<$	
7.	$<>$	
8.	$>=$	

9.	<=	
10.	NOT	
11.	AND	
12.	OR	
13.	XOR	
14.	EQV	Lowest

## Lecture.29

### Spreadsheet – MSEXCEL – rows, columns and cell identification – workbook and a worksheet – using simple statistical functions like average, median, mode, variance and standard deviation – creation of bar and pie charts in Excel

#### Assignment and Input / Output statements

We know that data must be assigned to the variable before they are used in computations. This can be done with LET or READ statements. After the program is run, the results of calculations are obtained with a PRINT statement. The LET statement is also called Assignment statement (in mode versions of BASIC LET need not be used in all the assignment statements) and the READ and the PRINT statements are the Input and the Output statements respectively.

#### The LET Statement

This statement is used to assign a value to a variable. Its format is

$$\frac{1 \text{ LET } v \text{ expr}}{\quad}$$

Where

- l        – a statement number (label)
- v        – Variable name
- expr    – an expression

eg :

```
10 LET A = 15
```

This will assign the value 15 to the variable A.

```
10 LET B = 4 + 5 * 6
```

First the expression  $4 + 5 * 6$  is evaluated according to the rules of evaluation of expressions. The value is 34. This value is assigned to the variable B.

Instead the above statement it can also be written as LET statement is optional in BASIC.

```
10 A=15 and
```

```
10 B=4+5*6
```

### **The READ statement**

Variables may be assigned values with or without the LET statement. Another statement that can be used to give values to variables is the READ statement. Its format is

l READ list

where

list --- indicates the names of variables—simple or subscripted—separated from each other by a comma.

Eg ;

26 READ A, B

27 READ P, Q, R, S, T (10)

### **The DATA statement**

The DATA statement gets the data values from the input device and stores them in the computer memory. The format of this statement is

l DATA d1,d2,d3...

where

d1, d2, d3... are the data values.

Eg:

28 DATA 2, 13, -4, 20.5

29 READ A, B, C, D

The assignment of values to the variables is: to start with, the pointer is at the value 2; 2 becomes the value of A. then the pointer moves to 13 and this becomes the value of B. After this, the pointer points to -4 and this is assigned to variable C. Next, the pointer goes to 20.5 and this becomes the value of D. the values 2, 13, -4, 20.5 stay unchanged in the computer memory.

There may be as many DATA statement in a program as are desired. They may be placed anywhere but always before the END statement.

Eg:

10 DATA 2.2, -3.0, 5

20 READ P,Q

```
30 DATA 8, 9.67
```

```
40 READ R
```

```
50 READ S, T
```

To start with, the pointer is at value 2.2. When the READ statement at line 20 is executed, P is assigned the value 2.2 and Q is assigned the value -3.0. Now the pointer is at value 5. When the READ statement at line 40 is executed, R gets the value 5 and the pointer moves to the next value 8. Values 8 and 9.67 are assigned to the variables S and T after the execution of the READ statement at line 50.

### **The RESTORE statement**

This statement transfers the pointer, regardless of where it is, to the first value of the data statement in the program.

Its form is

```
I RESTORE
```

eg :

```
28 RESTORE
```

Nothing should be specified after the word RESTORE. This statement may be used anywhere in the program.

Eg :

```
10 DATA 25, -10, 3.6
```

```
20 READ J, K
```

```
30 RESTORE
```

```
40 READ M, N, P
```

```
50 END
```

assigns the value 25 to J, -10 to K when the READ statement 20 is executed. After this, the RESTORE statement at line 30 transfers the pointer to the start of the data and so the data list can be read from the beginning again. The value 25 is assigned to M, -10 to N, 3.6 to P.

### **The PRINT statement**

This statement writes the results of computations performed by the computer on the printer.

The form of the PRINT statement is

l PRINT list

where

list --- may be constants, variables, expressions or message separated by commas or semicolons

eg :

```
90 PRINT A, B, C
```

This will print the value stored in A, B and C

```
92 PRINT "PRITAM"
```

This will print the word PRITAM.

```
94 PRINT " P =",P
```

Outputs the value of variable P which is preceded by the name of the variable, that is, first will be printed P= and then the value of P.

```
96 PRINT 6 +7*2, A
```

the value of expression  $6+7*2$  is calculated first. It is 20. Then 20 and the value of 'A' are printed, in that order, on the same line.

```
98 PRINT
```

When this statement is executed, a line is skipped on the printer. Thus, this form of PRINT statement is useful to insert spaces in the output data.

### **Use of Semicolon in the PRINT statement**

The different quantities appearing in the list of the PRINT statement can be separated by semicolons. For instance, we can specify

```
33 PRINT A; B; -8; D/E
```

When the semicolon is used as the delimiter, there is less space between the printed values. i.e values are 'squeezed' and we get results in a 'packed' format.

Eg:

```
34 PRINT 2116; 126; -67
```

The printed output may appear as

2116 126 -67

```
35 PRINT "GOBIND" ; "SINGH"
```

The output generated appear as

GOBINDSINGH

If you wish to have a space between the words GOBIND and SINGH, specify statement

35 as

```
35 PRINT "GOBIND"; "^"; "SINGH"
```

Or as

```
35 PRINT "GOBIND^" ; "SINGH"
```

semicolons and commas may be mixed in the PRINT list items

Eg:

```
38 PRINT "P="; -20, "Q="; 7777
```

This generates the output as

P = -20    Q= 7777

### **Program**

Let  $X = 1.11$ ,  $Y = 2.22$ ,  $Z = 3.33$ . Develop a program to compute  $X^2+Y^2+Z^2$  and  $X^3+Y^3+Z^3$ .

```
10 LET X=1.11
20 LET Y=2.22
30 LET Z=3.33
40 LET S =X^2+Y^2+Z^2
50 LET C= X^3+Y^3+Z^3
60 PRINT X, Y, Z, S, C
70 STOP
80 END
```

### **The WRITE Statement**

This statement is used to obtain output on the screen.

Its format is

```
1 WRITE list
```

where

list --- may be constants, variables, expressions or message separated by commas or semicolons

The difference between PRINT and WRITE statement is that with the WRITE statement comma appears after each output value and the literal value is accompanied by quotes.

### **The INPUT statement**

We have seen that there are two ways of assigning values to variables:

i) by the assignment (LET) and (ii) by the statement pair DATA/READ. When these statements are used, values are specified in the program itself. No input device need be used for entering data values. However, there is another way by which variables may be allotted values. This is by the INPUT statement. The form of this statement is

INPUT list

where

list --- represents the names of variables, simple or subscripted, separated by commas.

Eg :

44 INPUT X, Y, Z

Now the values of the variables X, Y, Z are entered through the input device (i.e) through the keyboard. Whenever an INPUT statement is executed in the program, the computer writes a question mark (?) It is a signal to the user that he should now enter the data values for variables. The user then types the constants that are to become the values of variables appearing in the list of the INPUT statement. RESTORE statement can not be used with INPUT statement.

**Input “no of observations”;n**

**Print “no of observations”**

**Input n**



## Program

Develop a program to convert temperatures, given in degrees centigrade(C), to degrees Fahrenheit (F) using the formula

$$F=(9/5)C +32$$

```
10 REM TO CONVERT TEMPERATUER IN
20 REM CENTIGRADE TO FAHRENHEIT
30 PRINT "TEMP IN CENTIGRADE";
40 INPUT C
50 LET F=(9/5)*C +32
60 PRINT
70 PRINT "TEMP IN FAHRENHEIT=";F
80 END
```

When the program is executed, the INPUT statement of line 40 writes ? after the following message:

TEMP IN CENTIGRADE ?

You type the number 37 after the question mark and the line on the screen appear as :

TEMP IN CENTIGRADE ? 37

After this, press key ENTER and the output appear as

TEMP IN FAHRENHEIT =98.6

Some versions of BASIC allow specify the output message with the INPUT statement as well. The format appears as:

```
1 INPUT "message", list
```

eg:

```
49 INPUT "ENTER VALUE OF RADIUS", R
```

Now the following appears on the screen

ENTER VALUE OF RADIUS ?

### **The REM statement**

This statement is used to insert remarks (REMARKS) or identification comments in a program.

The form of the REM statement is

1 REM comments

The comments may be any message not enclosed between quotes.

Eg:

```
29 REM THIS PROGRAM CALCULATES  
30 REM THE SUM OF THREE NUMBERS
```

when the comment cannot be contained in one line, it may be continued on the next line with new line number and the word REM.

### **Program**

One foot = 12.0 inches, one inch = 2.54 centimeters. Develop a program to compute inches and feet of 76.2 centimeters. Print the result as

CM =      INCHES =      FEET =

```
10 REM PROGRAM TO COMPUTE INCHES AND  
20 REM FEET WHEN GIVEN DATUM IS IN CM  
30 READ A, B, C  
40 DATA 12.0, 2.54, 76.2  
50 J=C/B  
60 REM VARIABLE J STORES THE NUMBER OF  
70 REM INCHES IN C CENTIMETERS  
80 F=J/A  
90 REM VARIABLE F STORES THE NUMBER OF  
100 REM FEET IN J INCHES  
110 PRINT "CM =", C,"INCHES=",J,"FEET =", F  
120 STOP  
130 END
```

### **The STOP and END statements**

The form of STOP statement is

1 STOP

eg :

500 STOP

When this statement is encountered in a program, the execution of the program stops. There may be more than one STOP statement or none at all in a program. It may be placed in the program wherever necessary. STOP is also known as logical end of the program.

The end of a program is always indicated by the END statement. The form of this statement is

1 END

This must be the last statement in a program. This indicates the physical end of the program.

eg :

1000 END

The computer will stop the execution of the program when the END statement is executed. Even if some more statements are available after END it will not be included for execution.

### **Program**

Let  $X = 1.11$ ,  $Y = 2.22$ ,  $Z = 3.33$ . Develop a program to compute  $X^2+Y^2+Z^2$  and  $X^3+Y^3+Z^3$ .

```
10 X=1.11
20 Y=2.22
30 Z=3.33
40 S =X^2+Y^2+Z^2
50 C= X^3+Y^3+Z^3
60 PRINT X, Y, Z, S, C
70 STOP
```

80 END

### Program

One foot = 12.0 inches, one inch = 2.54 centimeters. Develop a program to compute inches and feet of 76.2 centimeters. Print the result as

CM =      INCHES =              FEET =

```
10 REM PROGRAM TO COMPUTE INCHES AND
20 REM FEET WHEN GIVEN DATUM IS IN CM
30 READ A, B, C
40 DATA 12.0, 2.54, 76.2
50 J=C/B
60 REM VARIABLE J STORES THE NUMBER OF
70 REM INCHES IN C CENTIMETERS
80 F=J/A
90 REM VARIABLE F STORES THE NUMBER OF
100 REM FEET IN J INCHES
110 PRINT "CM =", C,"INCHES=",J,"FEET =", F
120 STOP
130 END
```

### Program

Develop a program to convert temperatures, given in degrees centigrade(C), to degrees Fahrenheit (F) using the formula

$$F=(9/5)C +32$$

```
10 REM TO CONVERT TEMPERATUER IN
20 REM CENTIGRADE TO FAHRENHEIT
30 PRINT "TEMP IN CENTIGRADE";
40 INPUT C
50 F=(9/5)*C +32
60 PRINT
```

```
70 PRINT "TEMP IN FAHRENHEIT=";F  
80 END
```

## Lecture.30

### Presentation software – Power Point – creation of simple slides – slide show

#### Transfer and Control Statements

Normally the execution of the program is based on the statement number. It reads the statements in the ascending order of the statement numbers. Sometimes we may want to change the flow of the program from one part to another part of the program. For this purpose we use transfer and control statements. The statements in BASIC used for this purpose are GOTO, If then, if then else and FOR and NEXT.

#### The Go To Statement

This statement causes the transfer of program execution control from one part of the program to another unconditionally

Its form is

$$\underline{\underline{l \text{ GoTo } n}}$$

Where

n -- is the statement label which must be in the same program, n is different from l.

Eg:

1 GoTo 10

2 GoTo 100

#### The ON-GoTo Statement

This statement is a conditional branch transfer type of statement. The transfer of control depends on the value of the expression which is a part of this statement.

The general form is

$$\underline{\underline{l \text{ ON } V \text{ GoTo } n_1, n_2, \dots, n_p}}$$

Where

v -- May be a constant, a variable, or an expression

n<sub>1</sub>, n<sub>2</sub>, ...n<sub>p</sub> – are labels of other statements in a program. The labels n<sub>1</sub>, n<sub>2</sub>, ...n<sub>p</sub> must be separated from each other by a comma.

This statement functions as follows

When  $v = 1$ , control goes to statement labeled  $n1$ ,

When  $v = 2$ , control goes to statement labeled  $n2$ ,

.....

.....

When  $v = p$ , control goes to statement labeled  $np$ .

**Program:**

```
10 LET X=3
20 ON X GOTO 80, 60, 30
30 PRINT X
40 X=X-1
50 GOTO 20
60 PRINT X
70 GOTO 40
80 PRINT X
90 STOP
100 END
```

**Output**

```
3
2
1
```

**The IF Statement**

This statement helps in transferring the control of program execution from one part of the program to another depending on some test-condition.

There are two types

- 1) IF-THEN statement
- 2) IF-THEN-ELSE statement

### 1) IF-THEN statement

Its format is

```
1 IF test-condition THEN n
```

Where

n – is the statement number

test-condition – may be relational or logical expression.

When the IF-THEN statement is executed in a program, the test-condition is tested. If the test-condition is true, control of program execution is transferred to statement labeled n. If the test-condition is False, the statement immediately following the IF-THEN statement is executed.

Eg:

```
5 IF A>10 THEN 125
```

### Program

Given a set of 6 values as -5, 2, 6, -1.1, 0, 4

Develop a program that computes the sum of the values greater than zero.

```
10 DATA -5, 2, 6, -1.1, 0, 4
20 LET N=0
25 LET S=0
30 READ A
40 LET N = N+1
50 IF A<=0 THEN 70
60 LET S=S+A
70 IF N<6 THEN 30
80 PRINT "SUM=";S
90 STOP
100 END
```

### Output

SUM=12

### 2) IF-THEN-ELSE statement

This is more general form of IF statement.



Its syntax is

---

1 IF test-condition THEN n1 ELSE n2

---

Where

n1 and n2 are labels of other statements.

When the test-condition is true, statement labeled n1 is executed. However when the test-condition is false, statement n2 is executed.

Eg:

60 IF A <> B THEN 100 ELSE 200

## LOOP STATEMENTS

Loop statements are used to execute one or more statements multiple times. These are

**The FOR & NEXT statements.**

The form of the FOR - NEXT statement is

---

1 FOR K= m1 TO m2 STEP m3

---

Where

K—must be a variable.

m1 – initial value of K

m2– final value which K can assume.

m3– size of the step by which the value of K will be changed.

m1, m2, m3 may be constants, variables or expressions. Their values may be negative, positive or zero. K is called the counter variable or the index of the FOR- NEXT statement.

Eg:

14 FOR L= 1 TO 10 STEP 2

Here K=L, m1=1, m2=10, m3=2. The counter variable L can assume the values 1, 3, 5, 7,9.

If the step is 1, the specification STEP m3 may be omitted. The form of the FOR-TO statement becomes

---

1 FOR K= m1 TO m2

---

eg:

```
30 FOR K=1 TO 50
```

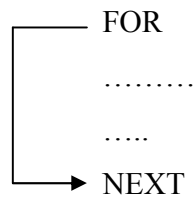
Whenever a FOR-TO statement is used, there must be a corresponding NEXT statement. Its function is to increase the value of the counter variable by the step size. The format of this statement is

```
    I NEXT K
```

Where

k -- must be the variable used in the FOR-TO statement as the counter variable.

The NEXT statement cannot be used independently and must follow the FOR-TO statement. We denote the loop as



### **Subscripted Variable**

Subscripted variables are used to store the values of the same type in an array. It also helps us to store more values in a single variable name. Subscripted variables are declared by giving the variable name along with the subscript.

Example: a(10)

Here a is the variable name and 10 is known as the subscripted value through which we can store 10 values in the variable name a. The 10 values are stored in a (1), a (2),..., a(10). As only one subscript is given for the above variable, the array is called as the single dimensional array.

### **Program**

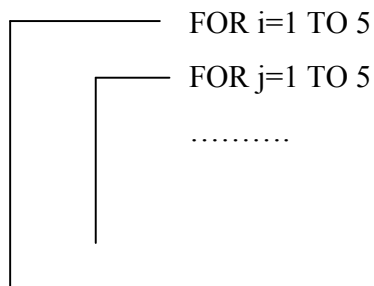
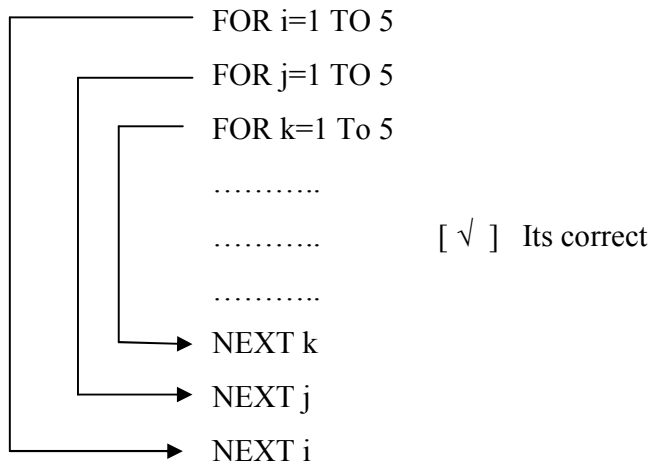
Given the number 20,13,-14,11,50,16,-17,18,19,2. Store them in the computer memory. Read a number from the terminal. Verify if this number is present in the above list

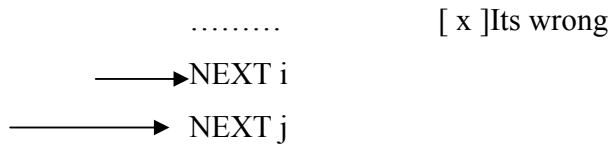
```

5 REM NUMBERS ARE SPECIFIED IN A
7 REM DATA STATEMENT
10 DATA 20, 13,-14, 11, 50, 16,-17, 18,19,2
20 INPUT X
30 FOR K=1 TO 10
40 READ Y
50 IF X=Y THEN 80
60 NEXT K
70 GOTO 100
80 PRINT " THE NUMBER : " ;x, "IS PRESENT"
90 GOTO 130
100 PRINT " THE NUMBER : " ;x, "IS ABSENT"
110 RESTORE
120 GOTO 20
130 STOP
140 END

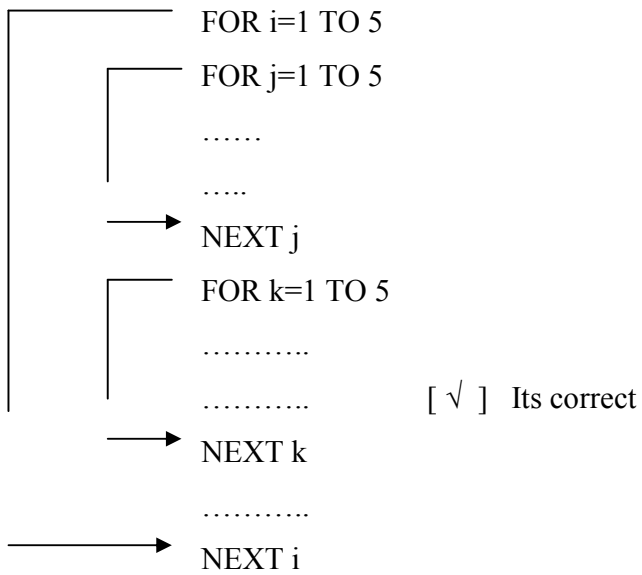
```

When the range of a loop lies within the range of another loop, the first loop is said to be nester in the latter. The loops may be placed as





The loops may also be specified as



### Program

Develop a program to calculate the sum

$$\sum_{j=1}^5 \left( \sum_{k=1}^5 k^j + \sum_{k=1}^5 j^k \right)$$

```

10 REM TO ILLUSTRATE NESTING OF LOOPS
20 PRINT "J", "SUM"
30 PRINT
40 REM FIRST LOOP STARTS
50 FOR J = 1 TO 5
60 S=0
70 REM SECOND LOOP STARTS
80 FOR K=1 TO 5
90 S = S+K^J
100 NEXT K

```

```
110 T=0
120 REM THIRD LOOP STARTS
130 FOR K=1 TO 5
140 T = T+K*J
150 NEXT K
160 Q=S+T
170 PRINT J,Q
180 NEXT J
190 END
```

### **Output**

J	SUM
1	30
2	85
3	270
4	1039
5	4500

### **The WHILE-DO statement**

The format of this statement is

---

```
WHILE test-condition
S1
S2
S3
.
.
.
DO
```

---

Where

test-condition --is relational or Boolean expression whose value may be true or false

S1,S2,... -- are BASIC statements

Each WHILE must have its corresponding DO statement. The action of WHILE –DO statement is as follows.

Execute statement S1, S2, ... up to DO ,as long as the test-condition is true. When it becomes false, execute statement following DO. Thus the execution of WHILE-DO statement is completed only when the test-condition becomes false.

### **Program**

```
10 REM ACTION OF WHILE –DO STATEMENT
20 LET SUM = 0
30 LET I= 1
40 WHILE I<=10
50 SUM =SUM + I
60 I = I+1
70 DO
80 PRINT “SUM=”;SUM
90 END
```

### **Questions**

1. \_\_\_\_\_ statement is used to insert remarks or identification comments in a program.
2. \_\_\_\_\_ is a transfer and control statement in BASIC.
3. \_\_\_\_\_
4. LET or READ statement is used to assign values to variables in BASIC
5. The DATA statement gets the data values from the input device and stores them in the computer memory.
6. What is a variable?
7. What is a constant?

8. Define a compiler.
9. Define an interpreter.
10. Develop a program that computes the sum of n values.
11. Explain loop statement.

## **Lecture.31**

**Visual Basic - Introduction - integrated development environment (IDE) - properties, method and events - variables - data types - working with forms - working with controls - dragging and dropping - message box**

### **NETWORKS**

#### **Definition**

- It is an interconnected collection of autonomous computers.
- Two computers are said to be interconnected if they are able to exchange information.
- The connection need not be via copper wire. fiber optics, microwaves and communication satellites can also be used.

#### **Distributed Systems**

- The existence of autonomous computers is transparent (i.e., not visible) to the user.
- The user can type a command to run a program, and it runs.
- The operating system selects the best processor, finds and transports all input files to that processor, and puts the result in the appropriate place.
- The user of a distributed system is not aware that there are multiple processors, it looks like a virtual uniprocessor.
- The software is built on top of a network.

#### **Computer Network**

- In this the user should explicitly log on to the machine, explicitly submit jobs remotely, explicitly move files around and generally handle all the network management personally.
- The difference between the distributed system and computer Network lies in who invokes the movement, the system or the user.

#### **Uses of Computer Networks**

- Computer Networks are used for



- 1) Companies
- 2) People
- 3) Social Issues

### **Networks for companies**

- The first goal is to provide resource sharing
- It makes all the programs, equipment and especially data available to anyone on the network without regard to the physical location of the resource and the user.
- The second goal is to provide high reliability
- It provides alternative sources for supply.
- For example, all files could be replicated on two or three machines, so if one of them is unavailable, the other copies could be used.
- The process of multiple CPUs means if one goes down, the others are available to take over the work.
- For military, banking, air traffic control and many other applications, the ability to continue operating in the face of hardware problems is of utmost importance.
- The third goal is saving money.
- Small computers have a better price/performance ratio than large ones.
- Many system designers build systems consisting of personal computers, one per user with data kept on one or more shared file server machines.
- The user is called the client and the whole arrangement is called the client-server model.
- The communication generally takes the form of a request message from the client to the server asking for some work to be done.
- The server then does the work and sends back the reply.
- The fourth goal is scalability.
- The ability to increase system performance gradually as the workload grows just by adding more processors.

- The fifth goal is to provide powerful communication medium among widely separated employee.
- When one worker makes a change to an on-line document, the others can see the change immediately, instead of waiting for several days for a letter.

### **Networks for people**

- It includes
  - 1) Access to remote information
  - 2) Person-to-person communication
  - 3) Interactive entertainment
- The access to remote information is in many forms. Many people pay their bills, manage their bank accounts and handle their investments electronically.
- Home shopping has become popular, with ability to inspect the on-line catalogues of thousands of companies.
- Newspapers will go on-line and be personalized
- Access to the World Wide Web, which contains information about the arts, business, government and many other topics.
- E-mail is widely used by millions of people.
- Virtual meeting called video conference for getting medial opinions from distant specialist and numerous other applications are used.
- Another application is entertainment which is huge and growing industry.

### **Social Issues**

- A popular feature of networks is newsgroups, where people can exchange message with like-minded individuals.
- The network like the printing press, allow ordinary citizens to distribute their views in different ways.

### **Advantages and disadvantages of network**

#### **Speed**

Sharing and transferring files within Networks are very rapid.

Thus saving time, while maintaining the integrity of the file.

### **Cost**

- Individually licensed copies of many popular software programs can be costly.
- Networkable versions are available at considerable savings.
- Shared programs, on a network allows for easier upgrading of the program on one single file server, instead of upgrading individual workstations.

### **Security**

- Sensitive files and programs on a network are passwords protected (established for specific directories to restrict access to authorized users).

### **Centralized Software Management**

- Software can be loaded on one computer (the file server) eliminating that need to spend time and energy installing updates and tracking files on independent computers throughout the building.

### **Resource Sharing**

- Resources such as, printers, fax machines and modems can be shared.

### **Electronic Mail**

- E-mail aids in personal and professional communication.
- Electronic mail on a LAN can enable staff to communicate within the building having not to leave their desk.

### **Flexible Access**

- Access their files from computers throughout the firm.

### **Workgroup Computing**

- Workgroup software (such as Microsoft BackOffice) allows many users to work on a document or project concurrently.

### **Disadvantages of Network**

- Server faults stop applications being available
- Network faults can cause loss of data.

- Network fault could lead to loss of resources
- System open to hackers
- Decisions tend to become centralised
- Could degrade in performance

### **Types of networking**

#### **Network hardware**

- There are two types of transmission technology:
  1. Broadcast networks
  2. Point-to-point networks

#### **Broadcast networks**

- They have a single communication channel that is shared by all machines on the network.
- Short messages called packets are sent by any machine and received by all others.
- An address field within the packet specifies for whom it is intended.
- Upon receiving a packet, a machine checks the address field. If the packet is intended for itself, it processes the packet, else ignores it.
- This mode of operation is called broadcasting.
- Some broadcast system also support transmission to a subset of the machines, known as multicasting.

#### **Point-to-point networks**

- This network consists of many connections between individual pairs of machines.
- To go from the source to the destination, a packet on this type of network may have to first visit one or more intermediate machines.

Inter processor Distance	Processors located in same	Example
0.1 m	Circuit Board	→ Data Flow Machine
1 m	System	→ Multi Computer
10 m	Room	↕ Local Area Network
100 m	Building	
1 km	Campus	↕ Metropolitan Area Network
10 km	City	
100 km	Country	↕ Wide Area Network
1,000 km	Continent	
10,000 km	Planet	→ The Internet

Classification of interconnected processors by scale

- In the data flow machines, highly parallel computers with many functional units all working on the same program.
- In multicomputer systems communicate by sending messages over very short, very fast buses
- Computers communicate by exchanging messages over longer cables.
- These can be divided into local , metropolitan and wide area networks.
- Connection of two or more networks is called internetwork. World wide internet is an example of an internetwork.

### Local area networks

Local Area Networks are generally called LANs, are privately owned networks within a single building or campus a few kilometers in size.

- They are widely used to connect personal computers and workstations in company offices and factories to share resources and exchange information.
- LANs distinguished from other kinds of networks by three characteristics:
  - 1) Their size
  - 2) transmission technology
  - 3) their topology
- LANs are restricted in size.
- They use a transmission technology consisting of a single cable to which all the machines are attached.

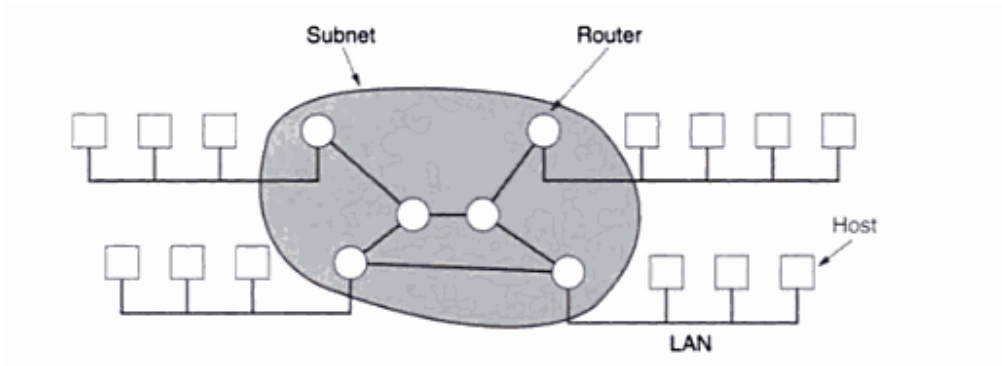
- Traditional LANs run at speeds of 10 to 100 mbps have low delay and make very few errors.
- Newer LANs, may operate at higher speeds, up to hundreds of megabits/sec(1,000,000 bits)
- Broadcast networks can be further divided into static and dynamic, depending on how the channel is allocated.
- Static allocation would divide time into discrete intervals and run a round robin algorithm, allowing each machine to broadcast only when its time slot comes up.
- They waste the channel capacity when a machine has nothing to say during its allocated slot, so dynamic method is preferred.
- Dynamic allocation methods for a common channel are either centralized or decentralized.

#### **Metrololitan area network**

- A MAN can support both data and voice, and can even be related to local cable television network.
- It has one or two cables and does not have switching elements. It has a standard called DQDB (Distributed Queue Dual Bus).
- It consists of two unidirectional buses to which all the computers are connected.
- Each bus has a head-end, a device that initiates transmission activity.
- Traffic that is destined for a computer to the right of a sender uses the upper bus.
- Traffic to the left uses the lower bus.
- There is a broadcast medium to which all the computers are attached.

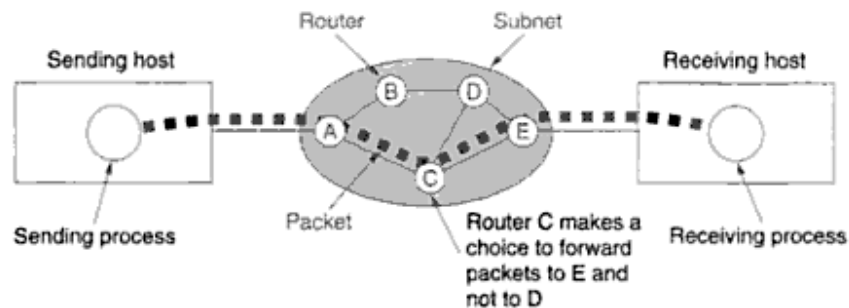
#### **Wide area networks**

- Wide Area Network or WAN spans a large geographical area.It contains a collection of machines for running the user programs called hosts.
- The host are connected by a communication subnet.



### Relation between the host and the subnet

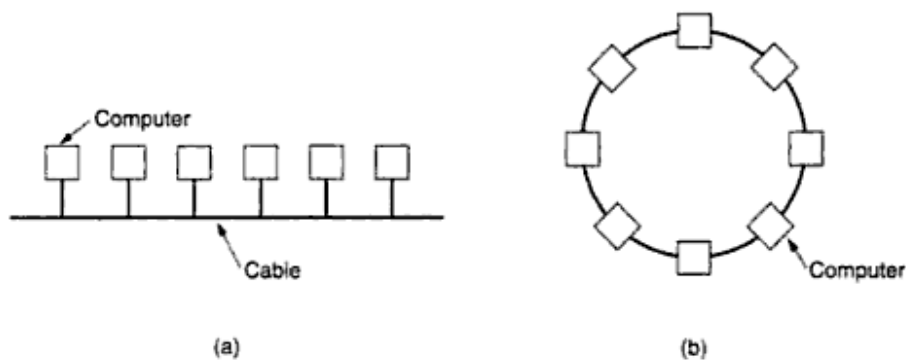
- The subnet is used to carry messages from host to host.
- Subnet consists of two different components: transmission lines and switching elements.
- Transmission lines (called circuits, channels or trunks) move bits between machines.



- Switching elements are specialized computers used to connect two or more transmission lines. When data arrive on an incoming line, the switching element must choose an outgoing line to forward them on.
- They are called packet switching nodes, intermediate systems and data switching exchanges.
- A generic term for switching computers is called the router.
- When a packet is sent from one router to another via one or more intermediate routers, the packet is received at its each intermediate router, stored there until the required output line is free, and then forwarded.

- A subnet using this principle is called a point-to-point, store and forward or packet-switched subnet.
- When packets are small and all the same size, they are often called cells.
- Some topologies of a point-to-point subnet (a)star (b)Ring (c)Tree (d)Complete  
(e)Intersecting Rings (f) Irregular

Two broadcast networks are (a) Bus (b) Ring



- Various topologies are possible for broadcast LANs (Bus and Ring).
- In **bus network**(a linear cable), at any instant one machine is the master and is allowed to transmit.
- IEEE 802.3, popularly called Ethernet, for example is a bus-based broadcast network with decentralized control operating at 10 or 100 mbps.
- In **ring network**, each bit propagates on its own, not waiting for the rest of the packet to which it belongs.
- IEEE 802.5(the IBM token ring),is a popular ring based LAN operating at 4 and 16 Mbps.

### Wireless networks

- Having a wired connection is impossible in cars and airplanes, there is a lot of interest in wireless networks.
- Wireless networks have many uses. A common one is portable office.



- Portable computers are sometime wired. Wireless LANs are easy to install.
- The computers communicate I wireless LAN in digital form.
- There is a possibility of using a cellular telephone with a traditional analog modem.
- Direct digital cellular service called CDPD (Cellular Digital Packet Data) is available in many cities.

### **Internetworks**

- Internet is a collection of LANs connected by WAN.
- The real distinction between a subnet and a WAN is whether or not host is present.
- If the system within the closed curve contains only routers, it is a subnet.
- If it contains routers and host with their own users, it is a WAN. In case of a LAN, the cable and the host form the network. There is no subnet.
- Internetwork is formed when distinct networks are connected together. Connecting a LAN and a WAN or connecting two LANs forms an internetwork.

### **Router**

- Device forwards data packets -networks.
- Router connected -at least two networks- two LANs or WANs or a LAN and its ISP's network.
- Located at gateways-where two or more networks connect.
- Use headers and forwarding tables –determine best path for forwarding the packets
- use protocols- communicate with each other and configure the best route between any two hosts.
- Tasks -routing and forwarding, generally containing a specialized operating system.

- Routers operate in two different planes:

1. Control Plane - learns the outgoing interface- appropriate for forwarding specific packets to specific destinations.

2. Forwarding Plane-responsible for actual process of sending a packet received on a logical interface to an outbound logical interface.

**Routers** are completely different devices.

- Where a hub or switch is concerned with transmitting frames.
- Router's job, as its name implies, is to route packets to other networks.
- One of the key features of a packet is that it not only contains data, but the destination address of where it's going.
- Router use protocols such as ICMP to communicate with each other and configure the best route between any two hosts.

## **Hub**

- Common connection point for devices in a network.
- Hubs used to connect segments of a LAN.
- Contains multiple ports.
- Packet arrives at one port, it is copied to the other ports so that all segments of the LAN can see all packets.
- Do not manage any of the traffic that comes through them.
- Packet collisions result.
- Frames carry data.
- When a frame is received, it is amplified and then transmitted on to the port of the destination PC.
- Difference between these two devices is in the method in which frames are being delivered.  
In a hub, a frame is passed along or broadcast" to every one of its ports.
- The hub has no way of distinguishing which port a frame should be sent to.

- Passing it along to every port ensures that it will reach its intended destination.
- This places a lot of traffic on the network and can lead to poor network response times.
- One PC is broadcasting-access to the maximum available bandwidth.
- Multiple PCs are broadcasting- bandwidth will need to be divided among all of those systems, which will degrade performance.
- Hub -collisions -on individual ports-*partition* the port, disconnecting from the shared medium.
- Status lights –hub-easily detected-coaxial cable.
- A compromise between a hub and a switch appeared known as a "**dual speed hub**".
- Two hubs (one of each speed) and a two port bridge between them.
- Devices were connected to the appropriate hub automatically based on their speed and the bridge handled inter-speed traffic.

### Switches

- A switch, however, keeps a record of the MAC addresses of all the devices connected to it.
- In networks, a device that filters and forwards packets between LAN segments.
- Switches operate at the data link layer (layer 2) and sometimes the network layer (layer 3) of the OSI Reference Model and therefore support any packet protocol.
- Network switches are capable of inspecting data packets as they are received, determining the source and destination device of that packet, and forwarding it appropriately.
- By delivering each message only to the connected device it was intended for, a network switch conserves network bandwidth and offers generally better performance than a hub.

### Questions

1. Network is an interconnected collection of \_\_\_\_\_ computers.

**Ans: autonomous**

2. LAN stands for \_\_\_\_\_

**Ans: Local Area Network**

3. A network operating within a city is called as Metropolitan Area network.

**Ans: True**

4. Connection of two or more networks is called internetwork.

**Ans: True**

5. Name the different types of networks.

6. Mention the two types of transmission technology.

7. Define a router.

8. What are the two broadcast networks?

9. Explain Wide Area Network.

10. Write about the uses of computer networks.

## Lecture.32

### Multimedia – Introduction – multimedia hardware and storage devices – communication devices – multimedia software – presentation tools

#### Introduction to Multimedia

Multimedia = Multi + media

- Multi = many
- Media = medium or means by which information is stored, transmitted, presented or perceived.
- By simple definition: **Multimedia** can be any combination of text, graphics, sound, animation and video, to effectively communicate ideas to users

#### Types of Multimedia Presentation

##### 1. Linear Multimedia

- Users have little control over the presentation, just sit back and watches the presentation
- The presentation normally plays from the start to end or even loops continually to present the information.
- **A movie** is a common type of linear multimedia

##### 2. Interactive Multimedia

- Users dictate the flow of delivery, to control the *what* and *when*.
- Users have the ability to move around or follow different path through the information presentation.
- Advantage: complex domain of information can be presented.
- Disadvantage: users might lost in the massive “information highway”.
- Useful for: **information archive (encyclopedia), education, training and entertainment.**

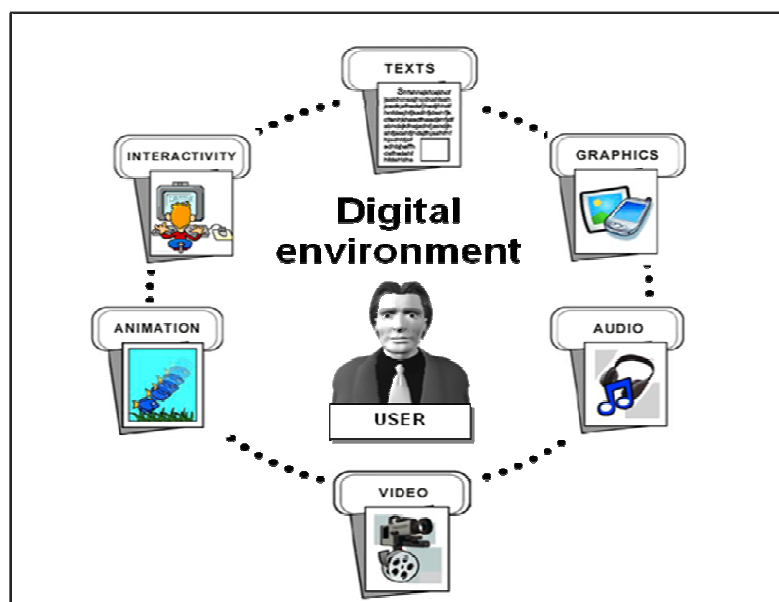
## Multimedia System Characteristics

1. Multimedia systems must be **computer controlled**.
2. All multimedia **components are integrated**.
3. The interface to the final user may **permit interactivity**.
4. **Information** must be represented **digitally**.

## Why Represent Information in Digital Form?

- **Storage**
  - Digital representation permits the storage of different information types on the same devices.
- **Transmission**
  - Information may also be transmitted over a single digital network.
- **Processing**
  - When digitized, all form of information may be treated by computer programs, for editing, quality improvement, or recognition of the meaning of the information

## Multimedia Building Block



## **Multimedia Takes Many Forms**

### **Multimedia takes many forms**

- **Courseware**
- **Movies**
- **Photo albums**
- **Image catalogs**

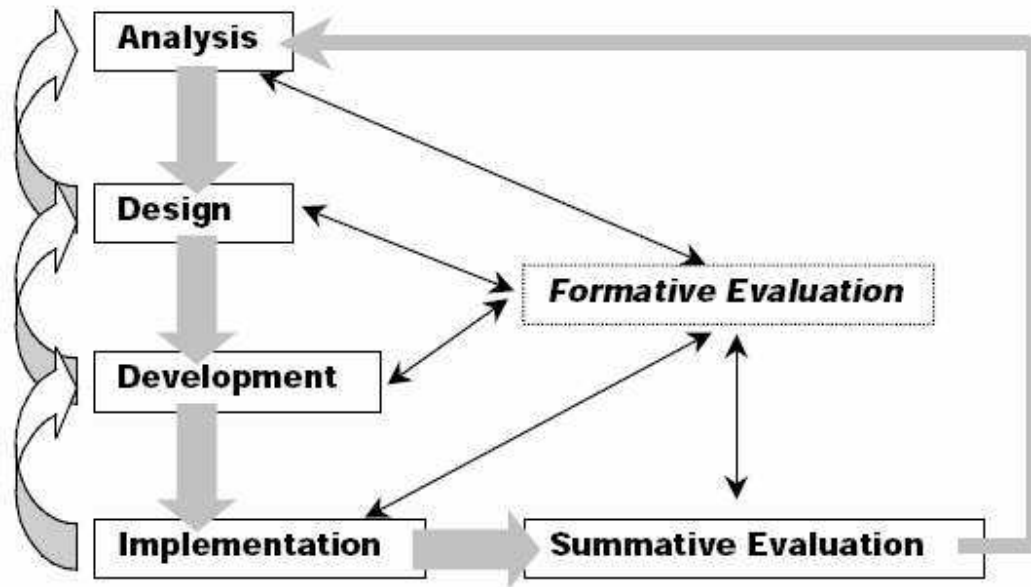
## **Delivering and Using Multimedia**

- Multimedia Can Be Delivered Online
- Online uses include:
  - Books and magazines
  - Movies
  - News and weather
  - Education
  - Maps
  - Entertainment

## **Creating & Developing Multimedia**

- Requirements
  - Creative skills
  - Technology & tools
    - Hardware
    - Software
  - Organization and business talent

## Multimedia Development Process



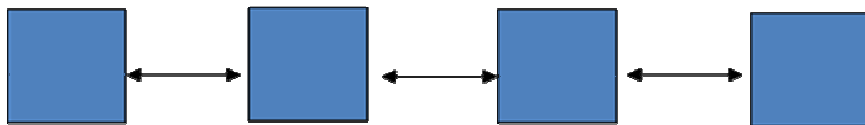
## Multimedia Design Process

- **Approaching the design**
  - Designers work closely with producers or clients throughout the process.
  - The conceptual model between the designers and the producers/clients must be closely matched.
  - The more works put in design help to eliminate tedious and costly alteration at the end of the project.
- **Design Aspects**
  - Interactivity – how does the process work? Allow users to control the application in a way that works with the content.
  - Structure – the flow ( navigation) and content structure of the multimedia application
  - Appearance – determine how the screen will look like.
- **Designing the interactivity**
  - Implemented in different ways depending on the product, content and development tools

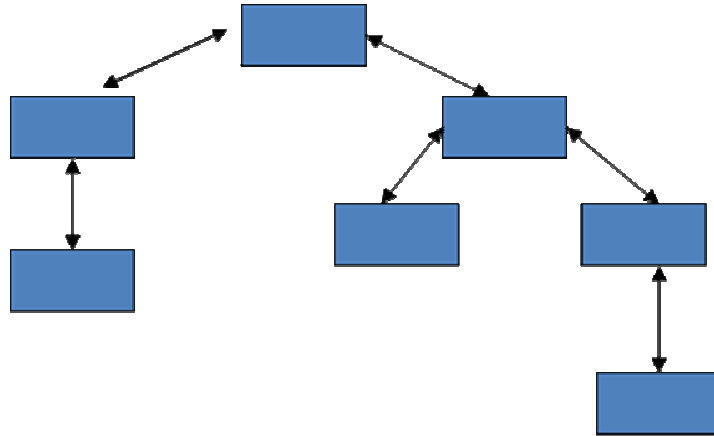


- Three approaches
  - Linear
  - Programmed branching
  - Hypermedia
- **Designing the structure**
  - Consideration on Navigation
    - The term navigation means finding your way to certain information.
    - Maps the structures and navigation paths as early as possible.
  - Common Navigation structures
    - Linear
    - Hierarchical
    - Nonlinear
    - Composite

- Common Navigation structures
  - **Linear:** Users navigate sequentially, from one frame of information to another



- **Hierarchical:** Users navigate along the branches of a tree structure that is shaped by the natural logic of the content



- **Common Navigation structures**

- **Nonlinear:** Users navigate freely through the content of the project, unbound by predetermined routes
- **Composite:** Users may navigate freely, but are occasionally constrained to linear presentations of movies or date.

- **Designing the structure**

- Consideration on Hot Spot, button and icon
  - Hot spot: Areas of the screen that are ‘clickable’.
  - Button: Provide button feedback whenever possible
    - Have a unique ‘pressed’ or ‘clicked’ state
    - In HTML, feedback is provided by colour changing
  - Icon: Avoid forcing users to learn special icons in your project. Try to follow the design convention set by other leading software.

- **Designing the Appearance**

- Usually called the *user interface design* or the *look and feel design*.
- The target is to facilitate communication, provide entertainment, elicit emotion, etc

- The graphic design of your multimedia product is the users first impression. A well design interfaces provide a good match between the user's task needs, skill level and learning ability and will lead to satisfied and productive users.
- **Storyboarding**
  - A storyboard is a series of screens, organized sequentially, screen by screen, and each screen is sketched out with design notes and specification before final rendering.
  - A storyboard can be developed with simple pencil and paper as well as using software applications.

## **Lecture.33**

### **Tools for object generation, video and sound – audio file formats – Video file format**

#### **Multimedia Building Blocks**

#### **Multimedia Building Block I Text**

##### **Text in History**

- Written text came into use about 6000 years ago
  - Mesopotamia, Egypt, Sumeria and Babylonia
- Initially text was written in symbols
  - Pictographics signs and cuneiforms

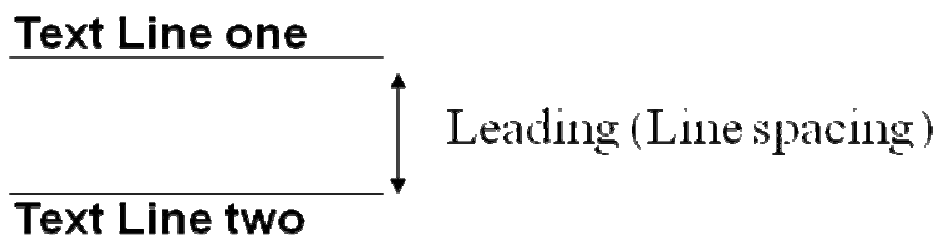
##### **Text in Multimedia**

- A key component of any multimedia product
- Multimedia products depend on text for various reasons
  - Page title
  - Content of the information
  - Label and Instruction
- Text/Word must be chosen carefully
  - Precise and accurate meaning to describe what you mean
- Words appears in titles, menus and navigational aids
- Test the words that you plan to use on several users and observe their reaction.

##### **Text Types Terminology**

- A 'typeface' is a family of graphic characters that usually includes many type sizes and styles.
  - ARIAL
  - Courier

- Times
- A ‘font’ is a collection of characters of a single size and style belonging to a particular typeface family.
  - This is an Arial font size 14, bold and underlined
  - *This is a Times New Roman font size 20, bold, italic and underlined*
- Point: the font size is usually measured in point.
  - A point is 1/72 of an inch (0.0138)
- Leading: the vertical space between lines of text



- Kerning: the space between two characters
- Adjusting the space between the characters is also called *tracking*

Tighter Track                      A V                      A V  
 Looser Track                      Kerned                      Unkerned

#### Serif versus Sans Serif

- The serif is the little decoration at the end of a letter stroke
- Serif
  - Old-fashioned, suitable for body text easy to read (printed)
  - Serif fonts are considered to be more readable on printed pages.
  - Times , New Century Schoolbook , Bookman
- Sans serif (without serif)
  - Clean, simple and modern, suitable for headlines and bold statements

- Sans serif fonts are usually more readable on computer screens.
- **Arial , Century Gothic, Verdana**

### **Hypertext**

- Hypertext is defined as the organized cross-linking of words, images, and other Web elements.
- A system in which words are keyed or indexed to other words is referred to as a hypertext system.
- A hypertext system enables the user to navigate through text in a non-linear way.
- HTML → Hypertext Markup Language.
  - Standard document format used for Web pages.

### **Hypertext Document**

Data is stored in a network of “nodes” connected by “links”.

Nodes can contain text, graphics, audio, video, and other forms of data.

### **Multimedia Building Block II –Images**

#### **Making Still Images**

- Still images are drawn in one of two ways:
  - Bitmapped images
    - Clip art
    - Bitmap software
    - Capture
    - Scanning
  - Vector-drawn images
- Images are usually compressed to save space
  - Formats like GIF, JPEG and PNG incorporate compression

#### **Bitmapped Image**

- A simple matrix or grid of dots with color information.

i.e. an array of color dots that when looked at from distance forms an image.

- The smallest element of a bitmap is a pixel (Picture element)

### **Vector Image**

- Image data are stored in the form of
  - Data points that describe the collection of lines, curves, circle, ellipses, text, polygon and other shape
  - The characteristic of each shape such as line type and fill/shading specification
- The information of the images can be stored as coordinates
- The computer recreates the image based on the information describing the image.

### **Bitmapped vs Vector Images**

- Vector images are easily scaled without quality loss.
- Bitmapped images get grainy and pixelated when zoomed in

### **Bitmapped vs Vector Images**

- Vector image require plug-ins
- Vector image can easily be edited
- Vector image files are usually smaller compared to bitmap
  - Contain only information how to recreate the image
- Vector graphics are web friendly
- However, to display complex vector images might require more system resources
- Bitmaps are more suitable for large images with many different colors (photograph)
- Special effect can easily be applied on bitmapped image (distortion, blurring).
- To apply the same effect, vector image need to be transformed to bitmapped first.

### **Image File Formats**

- There are many file formats used to store bitmaps and drawings :

- Windows device-independent bitmap (DIB)
  - .bmp, .dib, .rle
- CompuServe GIF (.gif)
- JPEG (.jpg)
- Apple Macintosh PICT (.pic or .pct)
- Portable Network Graphic (.png)
- Windows Metafile (.wmf)

### **Multimedia Building Block III -Sound**

#### **Understanding Sound**

- Sound as a form of energy
- Sound travels as waves at 750mph (at sea level)
- Sound pressure is measured in decibels (dB)
- Human ear can detect frequencies in the range between 20Hz to 20kHz
  - For example, a turbojet engine might be as loud as 165 dB. A car driving on the highway, about 100 dB. And, a whisper, averaging around 35 dB.
  - We cannot hear very high frequencies outside our hearing range and neither can we hear very low ones.

#### **Types of Sound**

- **Music**
  - Background - To set the mood
  - Attention grabber - To catch the interest of the audience
  - Special effect
    - To reinforce message or theme
    - To alert the audience
- **Speech**
  - effective for training and educational application.



- Narration
- Instruction

### **Sound Consideration**

- While sound (music or narration) can be very useful, most of the time, they can be irritating as well.
  - Use appropriate music to reflect the mood or theme.
  - Give the users choice of turning off/on the sound.
  - In cases where the content of a page depends on sound (speech), consider providing transcript as the alternative. (for deaf people, non multimedia pc)

### **Digital Audio Overview**

- Digital audio is a representation of the original sound.
- Sound digitization is obtained through
  - Sampling
  - Quantization
  - Code-word generation process
- Digitized sound is called sampled sound.
- Digital audio usually uses large storage space
- Digital audio is less affected by hardware, thus providing consistent playback quality.
  - Audio track for multimedia project will sound as good in the end as it did in the beginning when it is created.
- Require less musical talent to record.

### **MIDI Audio Overview**

- **Musical Instruments Digital Interface**
- 1980s → a standard communications protocol between electronic musical instruments and computer

- MIDI file is not a digitized sound, it is the shorthand representation of music stored in numeric form (binary)
- MIDI file contains instructional data (notes, sequences of notes, which instruments to play etc.) to other devices on how to generate an appropriate sound,
- It does not have the actual audio data, just the instruction to the instrument or sound card on how to reproduce it.
- Offer opportunity for developers to compose their own music.
- MIDI files are much more smaller than digital audio files, thus,
  - It requires less storage space.
  - Suitable to be used with web page.
- Inconsistency in playback quality
  - Playback will only be accurate if the MIDI playback device is identical to the device used for production.
- Cannot be used for spoken dialogue.

### **Multimedia Software:**

#### **Software that enables Multimedia:**

- ❖ System Software
  - OS
  - Utilities
  - Networking
- ❖ Development Software
  - Graphics
  - Sound
  - Text
  - Web Developing
  - Multimedia Authoring

❖ Delivery Software

- Stand-Alone Programs
- Players

**System Software**

- Written specifically to control its CPU
- Unique to the CPU brand and model
- Written in or translated into the machine language of the CPU
- Used to run computer
- Utilities
  - To play sound
  - Print the screen image
  - Eventually became operational extensions of OS
  - Important consideration when choosing OS

**Painting and Drawing Tools**

Most multimedia authoring programs, at least for now, import bitmaps.

- Painting Software (bitmap)
  - PhotoShop, PicturePublisher, and Fractal Designer
- Drawing Software (vector)
  - CorelDraw, FreeHand, Illustrator, Designer, and Canvas
  - Macromedia's Flash vector
  - Fireworks

**Image-editing Tools**

Are specialized and powerful tools for enhancing and retouching, existing bitmapped images

- PaintShop Pro & PhotoShop

- Plug-ins
  - Vertigo's HotTEXT lets you turn text into 3-D objects
  - EyeCandy
  - Kai's power tools
  - Xaos tool's paint alchemy

### **3-D Modeling**

- Extreme 3-D, 3d/fx, AutoDesk's 3D Studio Max, StrataVision's 3D, Infini-D, etc.
- Bryce
- Poser

### **Sound Applications**

- Sound Editing
  - Sound Recorder
  - Sound Forge
  - SoundEdit
  - MIDI
- Sound Playback
  - Windows Media Player
  - QuickTime
  - Real Player
  - WinAmp

### **Animation, Video and Digital Movie Tools**

Animation and the perception of motion

QuickTime, Microsoft's video for windows, Adobe Premiere,

Morphing

Video formats avi, mov, mpeg

Movie editors – decoders – Targa boards 30+, Adobe after effects, Pinnacle boards

Compression, Image quality, Compress/decompress speed

Lossless or lossy

### **Linking Multimedia Objects**

DDE & ole

Office suites

Presentation tools

Astound

Persuasion

PowerPoint

DeltaGraph professional

Harvard graphics

### **Networking Software**

#### ■ Network Hardware Protocols

- Ethernet
- Token Ring
- ATM

#### ■ Message Passing Protocols

- HTTP
- SMTP
- AppleTalk
- TCP/IP
- IPX/SPX
- SNA

## **Web Development**

### ❖ **Authoring Tools**

- Microsoft FrontPage
- Macromedia's Dreamweaver
- Adobe's GoLive

### ❖ **Languages**

- JavaScript
- Visual Basic

## **Helpful Accessories**

### ❖ **Screen Grabber**

- PaintShop Pro
- Capture
- SnapPro
- Hijaak 95

## **Video can be stored in many different formats**

### **The AVI Format**

The AVI (Audio Video Interleave) format was developed by Microsoft.

The AVI format is supported by all computers running Windows, and by all the most popular web browsers. It is a very common format on the Internet, but not always possible to play on non-Windows computers.

Videos stored in the AVI format have the extension .avi.

### **The Windows Media Format**

The Windows Media format is developed by Microsoft.

Windows Media is a common format on the Internet, but Windows Media movies cannot be played on non-Windows computer without an extra (free) component installed. Some later Windows Media movies cannot play at all on non-Windows computers because no player is available.

Videos stored in the Windows Media format have the extension .wmv.

### **The MPEG Format**

The MPEG (Moving Pictures Expert Group) format is the most popular format on the Internet. It is cross-platform, and supported by all the most popular web browsers.

Videos stored in the MPEG format have the extension .mpg or .mpeg.

### **The QuickTime Format**

The QuickTime format is developed by Apple.

QuickTime is a common format on the Internet, but QuickTime movies cannot be played on a Windows computer without an extra (free) component installed.

Videos stored in the QuickTime format have the extension .mov.

### **The RealVideo Format**

The RealVideo format was developed for the Internet by Real Media.

The format allows streaming of video (on-line video, Internet TV) with low bandwidths. Because of the low bandwidth priority, quality is often reduced.

Videos stored in the RealVideo format have the extension .rm or .ram.

### **The Shockwave (Flash) Format**

The Shockwave format was developed by Macromedia.

The Shockwave format requires an extra component to play. This component comes preinstalled with the latest versions of Netscape and Internet Explorer.

Videos stored in the Shockwave format have the extension .swf.

## **Audio File Formats**

- Common file types used for digitized sounds
  - AIFF audio (.aifc, .aiff, .aif)
  - MIDI (.midi, .mid)
  - MP3 Audio (.mp3)

- Ogg-Vorbis (.ogg)
- Real Audio (.ra, .ram, .rm)
- Windows Media (.wma)
- WAV audio (.wav)

## **Tools for object Generation**

### **Animation**

#### **The Power of Animation**

- Animation is achieved by adding motion to still image / object.
- May also be defined as the creation of moving pictures one frame at a time.
- Animation grabs attention.
- Few types of animation
  - Layout transition (i.e. in PowerPoint)
  - Process/ information transition
  - Object movement

#### **Layout Transition**

- The simplest form of animation is transition
- Transitions specify how the display changes (such as fading to black) as a user moves from one item (such as slide or Web page) to another)
- Examples :
  - Spiral
  - Stretch
  - Zoom
  - Checkerboard

#### **Process / Information Transition**

Animation can be used to describe complex information / process in an easier way



- Perform visual cues (e.g. how things work)

### **How Animation Works**

It is believed that animation is possible because of

- a biological phenomenon known as persistence of vision
  - An object seen by human eye remains chemically mapped on the eye's retina for a brief time after viewing
- a psychological phenomenon called **phi**.
  - Human's mind need to conceptually complete the perceived action i.e. translating the action
- Combination of these two (**persistence of vision + phi**) make it possible for a series of images that are changed very slightly and very rapidly, one after another, to seemingly blend together into a visual illusion of movement.
- E.g. a few cells or frames of rotating logo, when continuously and rapidly changed, the arrow of the compass is perceived to be spinning.
- The speed of the image changes is called the frame rate.
- Movie/Film is typically delivered at 24 – 30 frames per second (fps)
- Computer animations can be effective at 12 to 15 frames per second

### **Animations File Formats**

- Some file formats are designed specifically to contain animations :
  - Macromedia Director (.dir and .dcr)
  - Macromedia Flash (.fla and .swf)
  - AnimatorPro (.fli and .flc)
  - 3D Studio max (.max)
  - SuperCard (.pics)
  - CompuServe GIF89a (.gif)

## Video

- The term "video" commonly refers to related types of carrier formats —which can either be **digital** (DVD, Quicktime, Ogg) or **analog** videotape (VHS, Betamax).
- Television broadcasting and home movies have long been the traditional application of video technology
- The Internet has made possible the rise of compressed video file formats to syndicate video files to a global audience.
- Video is also used in scientific engineering,

manufacturing, and security applications.

## Digital Video

Advantages	Disadvantages
Easy to <u>edit sequences</u> of video (frame by frame)	high capacity storage
Easy to add special effects	Easily duplicated (piracy)
<u>Quality</u> of the original video are retained / maintained in the new copy. Digital mastering means that quality will never be an issue	Require high-end pc to process and deliver (fast processor, plenty of RAM, fast and big hard disk
Better image and audio quality (digital quality). Digital video <u>reduces generational losses</u> suffered by analog video	User need to have good knowledge and background in video compression and technology
Long lasting, offer superior quality at a given <u>cost</u>	

## Video Consideration

- To get the highest video performance, we should:
  - Use **video compression hardware** to allow you to work with full-screen, full-motion video.
  - Use a sophisticated **audio board** to allow you to use CD-quality sounds.
  - Install a Super fast **RAID** (Redundant Array of Independent Disks) system that will support high-speed data transfer rates.

## Obtaining Video Clip

- If using analog video, we need to convert it to digital format first (in other words, need to digitize the analog video first).
- Source for analog video can come from:
  - **Existing video content / clips**
    - beware of licensing and copyright issues
  - **Take a new footage (i.e. shoot your own video)**
    - Ask permission from all the persons who appear or speak, as well as the permission for the audio or music used.

## Video basics

- Light passes through the camera lens and is converted to an electronic signal by a **Charge Coupled Device (CCD)**
- Most consumer-grade cameras have a single **CCD**.
- Professional-grade cameras have three **CCDs**, one for each **Red, Green** and **Blue** color information (**RGB**)
- The output of the **CCD** is processed by the camera into a signal containing three channels of color information and synchronization pulse (sync).

## Video Compression

- Because of the large sizes associated with video files, video compression/decompression programs (Codec) have been developed.
  - Lossless compression → preserve image throughout the compress/decompress process.
  - Lossy compression → eliminates some of the data in the image (greater compression ratios) → usually used for video as some drop in quality is not noticeable in moving images
  - Trade-off is file size versus image quality.
  - Common compression standards
    - **MPEG** (Motion Picture Experts Group)

- **JPEG** (Joint Photographic Experts Groups)

## **Multimedia Authoring Tools**

### **Introduction to Multimedia Authoring Tools**

- ▶ Provide the framework for organizing and editing the elements of a multimedia project.
- ▶ Provides an integrated environment for combining the content and functions of a project.
- ▶ Enables the developer to create, edit, and import data

### **Word**

- ▶ Including various image formats, movies, and
- ▶ digitized sounds (including voice annotations).

### **Spreadsheets**

- Spreadsheets can include embedded objects made with other applications.

### **FileMaker**

- A FileMaker Pro employee database can include image and sound resources.

### **PowerPoint**

- Microsoft PowerPoint provides multimedia linking and embedding features.

### **Types of Authoring Tools:**

- ▶ Card- or page-based tools.
- ▶ Icon-based, event-driven tools.
- ▶ Time-based tools.

### **Card-based or page-based tools**

- The elements are organized as pages of a book or a stack of cards.
- Card-or page-based authoring systems

- best used when the bulk of your content consists of elements that can be viewed individually the pages of a book or cards in a card file.
- link these pages or cards into organized sequences.
- jump, on command, to any page play sound elements and launch animations and digital video.

### **Icon-or object-based, event-driven tools**

- multimedia elements and interaction cues ( events) are organized as objects in a structural framework or process.
- simplify the organization of your project

display flow diagrams of activities along branching paths.

- In complicated navigational structures, this charting is particularly useful during development.

### **Time-based tools**

- Elements and events are organized along a timeline with high resolutions.

### **Time-based tools**

- best to use when you have a message with a beginning and an end.
- played back at a speed that you can set
- Other elements (such as audio events) are triggered at a given time or location in the sequence of events.
- Jumps to any location in a sequence navigation and interactive control.

### **Introduction to Multimedia Authoring Tools**

- ▶ Authoring system in multimedia.
- ▶ Features of authoring tools.

### **Authoring System in Multimedia**

- ▶ Multimedia elements and events are often regarded as objects.
- ▶ Objects exist in a hierarchical order of parent and child relationships.
- ▶ Each object is assigned properties and modifiers.

- ▶ On receiving messages, objects perform tasks depending on the properties and modifiers.

### **The Right Tool for the Job**

- ▶ Flash and PowerPoint a presentation over the Web.
- video, slide shows, and diagrams with sound and graphics.

### **Different Stages of Authoring**

#### **Analysis**

- What do you need to do and what do you use to do it?

#### **Design**

- Create storyboards to tell the story of the project.

#### **Development**

- Incorporate data and set it up as a prototype or model.

#### **Evaluation**

- When the prototype application works the way you want it to, test it again, fine-tune it, and then review your work.

#### **Distribution**

- When it is ready to go (after the evaluation phase), make it real Package and distribute it.

### **Features of Authoring Tools**

- ▶ Editing and organizing features.
- ▶ Programming features.
- ▶ Interactivity features.
- ▶ Performance tuning and playback features.
- ▶ Delivery, cross-platform, and Internet playability features.

## **Editing and Organizing Features**

### **Editing tools**

- ▶ to create, edit, and convert multimedia elements such as animation and video clips.
- ▶ The organization, design, and production process for multimedia involves storyboarding and flowcharting.
- ▶ Visual flowcharting or overview facility illustrates project structure at a macro level.

### **Programming Features**

- ▶ Visual programming with icons or objects
  - the simplest and easiest authoring process. Authorware and IconAuthor suitable for slide shows and presentations.
- ▶ Authoring tools offer ‘very high level language’ (VHLL) or interpreted scripting environment.

## **Interactivity Features**

### **Interactivity**

- the end user control over the content and flow of information
  - Simple branching
    - go to
  - Conditional branching
    - IF-THEN decisions or events.

### **Structured language**

- complex programming logic, subroutines, event tracking, and message passing among objects and elements.

## **Performance Tuning and Playback Features**

- ▶ Synchronization is difficult
- ▶ Authoring system should facilitate precise timing of events.

- ▶ It should enable developers to build a part of a project and then test it immediately.

### **Delivery, Cross-Platform, and Internet Playability Features**

- ▶ Delivering the project may require building a run-time version of the project, using the multimedia authoring software.
- ▶ Run-time version or standalone
  - allows a project to play back without the complete authoring software and all its tools and editors.
- ▶ Across platforms
- ▶ Authoring systems provide a means for converting their output to be delivered within the context of HTML or DHTML.

### **Card- and Page-Based Tools**

- ▶ Card- and page-based authoring systems
  - provide a simple and easily understood metaphor for organizing multimedia elements.
  - contains media objects such as buttons, text fields, and graphic objects.
  - provides a facility for linking objects to pages or cards.
- ▶ **Card-and page-based systems typically provided two separate layers on each card**
  - a background layer that could be shared among many cards
  - a foreground layer that was specific to a single card.

### **Icon-Based, Event-Driven Tools**

- ▶ Icon-based, event-driven tools provide a visual programming approach to organize and present multimedia.
  - Multimedia elements and interaction cues are organized as objects in a flowchart.
  - Flowchart can be built by dragging appropriate icons from a library, and then adding the content.



## **Time-Based Tools**

- ▶ Time-based tools are best suited for messages with a beginning and an end.
  - Some time-based tools facilitate navigation and interactive control.
  - Macromedia's Director and Flash are time-based development environments.

## **Flash**

### **Adobe Flash**

- ▶ delivering rich multimedia content to the Web.
- ▶ the creation of simple static HTML pages with the Flash Player plug-in.
- ▶ ActionScript
  - based upon the international ECMAScript standard (<http://www.ecma-international.org>) derived from Netscape's original JavaScript.

## **Macintosh Versus Windows Platform**

### **The Macintosh platform**

- ▶ Apple in 1984.
- ▶ a good built-in audio and high-quality graphics capability
- ▶ Includes hardware and software for digitizing and editing
- ▶ Makes multimedia project development easier and smoother.

### **The Windows platform**

- ▶ different vendor-neutral components that are tied together by the requirements of the Windows operating system.
- ▶ Initially focused on business computing
- ▶ was not suitable for multimedia
- ▶ It is now easier to find multimedia hardware and software for Windows as compared to the Macintosh.
- ▶ Windows-95%
- ▶ Mac-3.71%

- ▶ Others-1.23%

## **Networking Macintosh and Windows Computers**

### **Networking**

- ▶ for direct communication
- ▶ sharing of resources across platforms.
- ▶ Local area network (LAN)
- ▶ Wide area network (WAN)
- ▶ Internet

### **LAN**

- ▶ workstations are located within a short distance

### **WAN**

- ▶ communication systems span great distances
- ▶ set up and managed by large corporations
- ▶ expensive to install and maintain
- ▶ A dial-up connection

The Internet through an Internet Service Provider (ISP)

Communication between a Macintosh and Windows PC:

- Ethernet system and client-server software.
- ▶ Macintosh computers: built-in Ethernet networking
- ▶ Windows PCs: require an additional Ethernet card.

### **Connections**

The various connection methodologies

- ▶ Small Computer System Interface (SCSI).
- ▶ Integrated Drive Electronics (IDE).
- ▶ Universal Serial Bus (USB).

### **Small Computer System Interface (SCSI)**

- connect internal and external peripheral equipments
  - SCSI cards
  - Preferred for real-time video editing, network servers and situations that require mirroring.
- ▶ SCSI ID conflicts should be avoided by providing unique IDs to devices.

### **Questions**

1. HTML stands for \_\_\_\_\_

**Ans: Hyper Text Markup Language\**

2. MIDI stands for \_\_\_\_\_

**Ans: Musical Instrument Digital Interface**

3. MPEG is Moving Photographic Experts Group.

**Ans: False**

4. The AVI format was defined by Microsoft

**Ans: True**

5. Define animation?

6. Name two image file formats.

7. What is a vector image?

8. List four audio file formats.

9. What are multimedia authoring tools?

10. What is MIDI?

## Lecture.34

**Internet – introduction – network and types of network – LAN, WAN and MAN – advantages and disadvantages of net works – devices – hub, router and switch**

### **VISUAL BASIC**

#### **Origins through BASIC**

- Beginners All-purpose Symbolic Instruction Code
- A simple programming language
- Developed mid 1960's

#### **Graphical User Interface**

- Visual Basic allows you to create GUIs
- Graphic User Interface (GUI) comprises
  - Forms
  - Controls
  - Event-driven programming

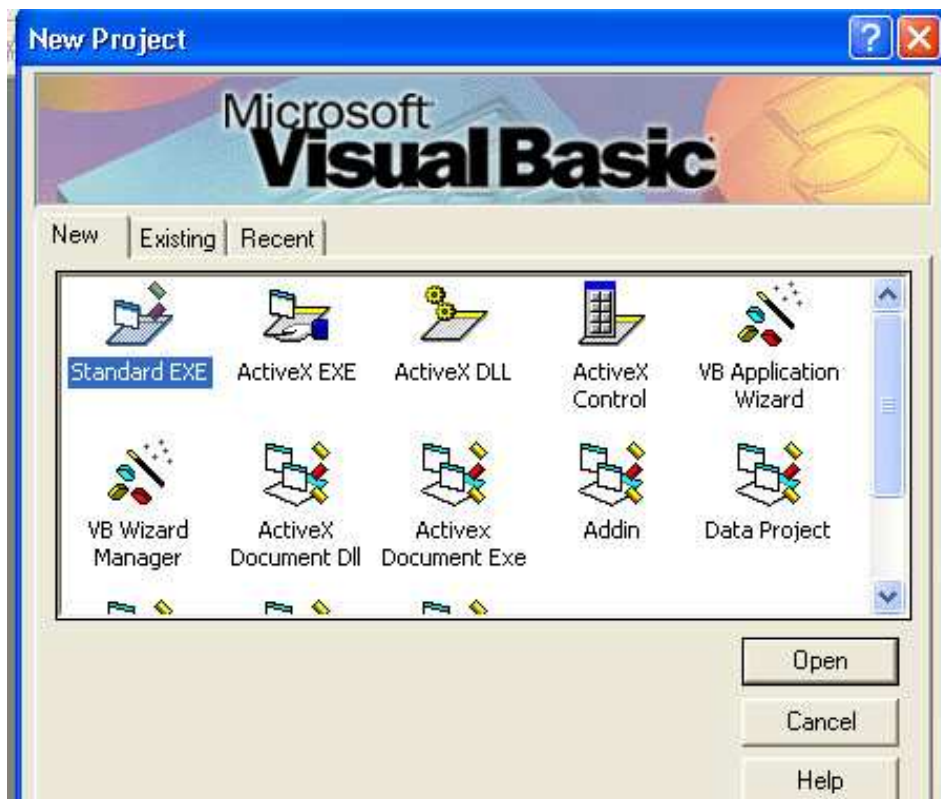
#### **Starting Visual Basic**

- o Click the Start button.
- o Choose Programs -> Microsoft Visual

Studio 6.0 ->Microsoft Visual Basic 6.0.

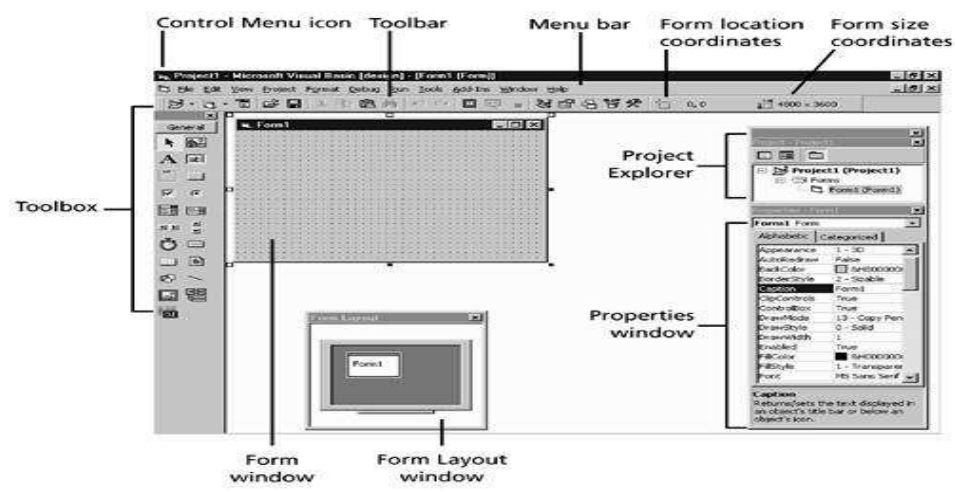
(Or)

Double click the shortcut icon, if it is available.



## INTEGRATED DEVELOPMENT ENVIRONMENT

The windows that are displayed when we start VB are collectively known as the Visual Basic Integrated Development Environment (IDE). It is an event-driven programming.

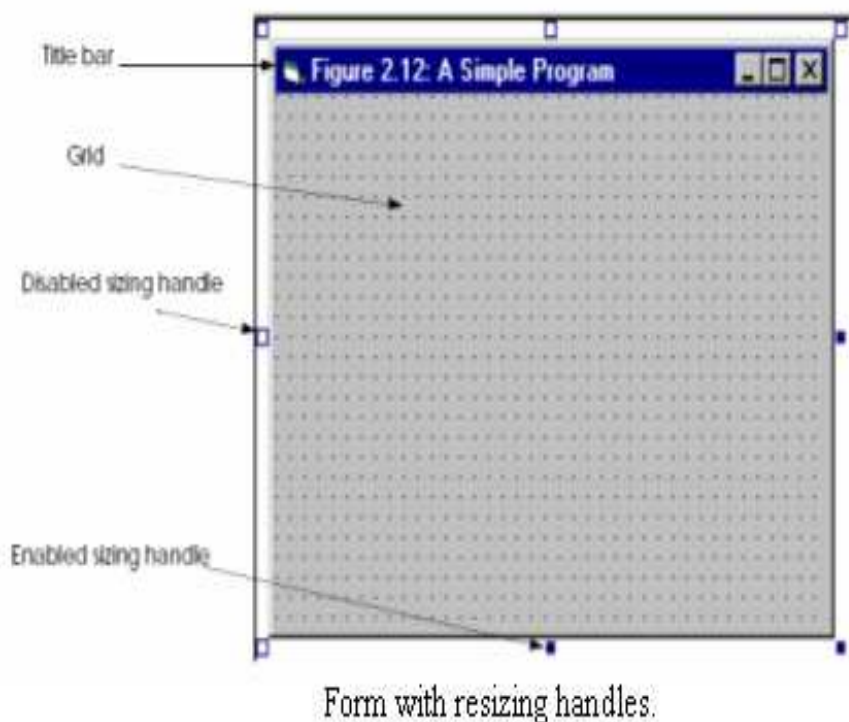


## IDE

- Form Window.
- Toolbars
- Menu bar
- Visual Basic toolbox
- Properties window
- Project Explorer
- Immediate window
- Form Layout window

## Form Window

- A standard grid (a window consisting of regularly spaced dots).
- Form window grid - **create the user interface.**



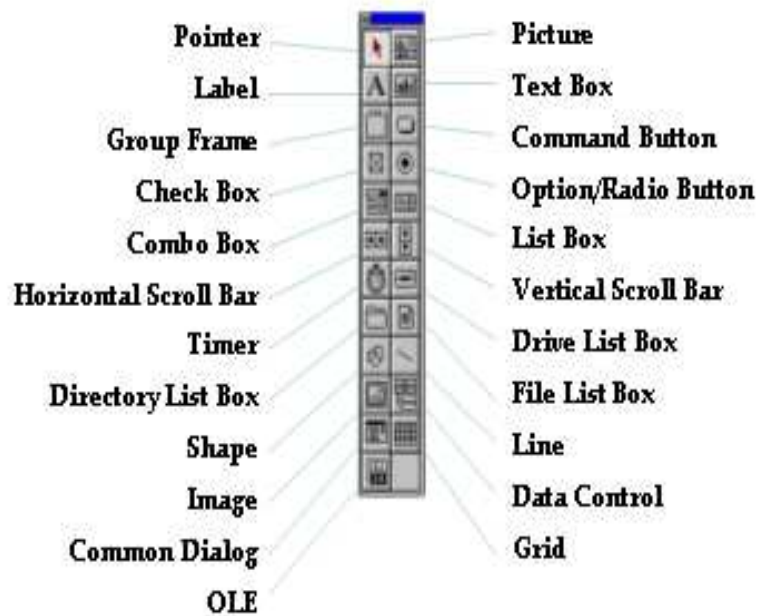
### Building Interface Elements:

- Adjusting Form Size
- Controlling Form Placement
  - Form Layout window.

### VB Environment: Menu Bar

- The Menu Bar consists of 3 elements the
  - Title Bar
  - Menu Bar
  - Toolbar

### Toolbox Controls



---

### Controls in the toolbox

**Pointer** - use to move or change the size of a control

**Picture box** - use to display graphics

**Image** - use to display a graphic, uses fewer resources than a picture box

**Label** - use for text that the user cannot change

**Text box** - use for text that the user can enter or change

**Frame** - use to group controls together

**Command button** - use to create a clickable button

**Check box** - use for choices where more than one choice can be selected

**Option button** - use for choices where choices are mutually exclusive

**Combo box** - user can either select an item from the list or enter a value

**List box** - use to display a list of items where only one choice is allowed

**Horizontal or Vertical Scroll bars** - use to quickly navigated through a long list of items or a large amount of information

**Timer** - use like a stopwatch to trigger events

**Drive list box** - displays list of disk drives

**Directory list box** - displays list of folders

**File List box** - displays a list of files

**Shape** - use to draw a rectangle, square, oval or circle

**Line** - use to draw a variety of lines

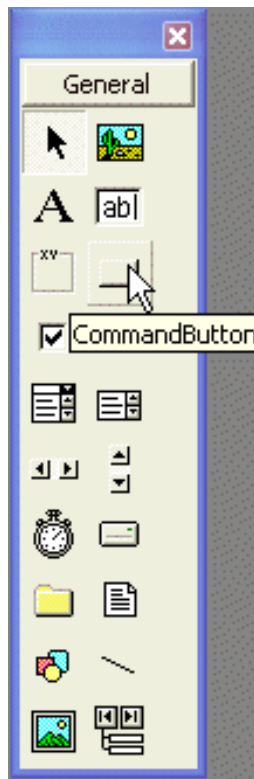
**OLE Container** - use to link or embed OLE objects

**Data** - use to create applications from many types of databases

### **Control identification**

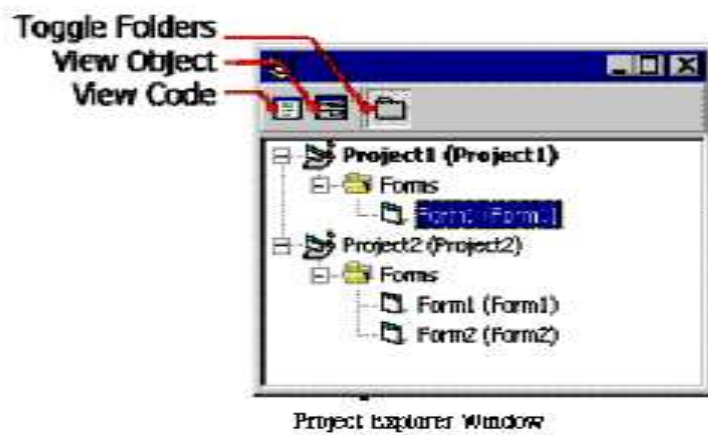
- We can position the mouse on it. A tool tip would appear.





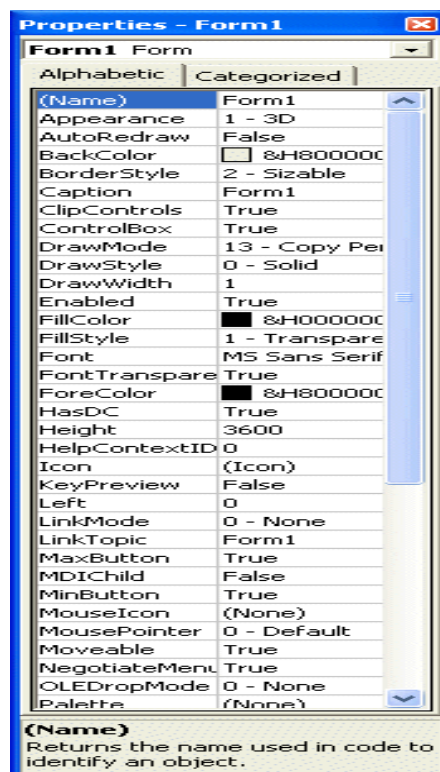
## PROJECT EXPLORER WINDOW

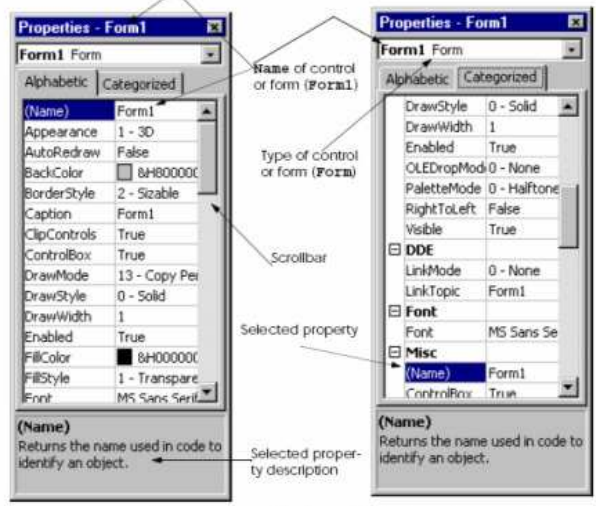
- Switch back and forth between these components as we work on a project.
- Buttons: **View Code** , **View Object** ,toggle folders



## Properties window

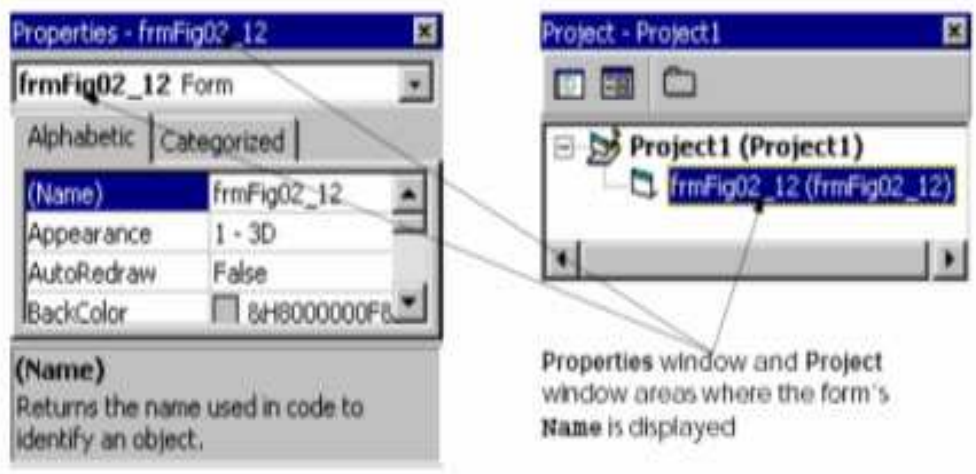
- We can change the characteristics (property settings) of the user interface elements on a form.
- To display the Properties window, click the **Properties Window** button on the toolbar.





Properties window with the Alphabetic and Categorized tab.

**Changing the property settings**



Properties window and project window after the name prosperity is set.

**Code window**

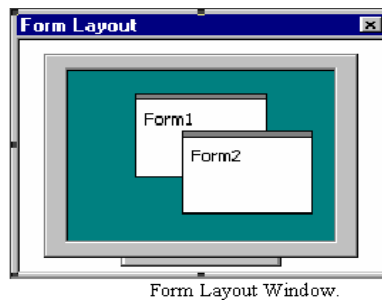
We can display the Code window in either of two ways:

- By clicking View Code in the Project window
- (or)

- By clicking the View menu Code command.

### **Form Layout Window**

- The Form Layout window is a visual design tool.
- We can arrange the forms -dragging the miniature form to the desired location.



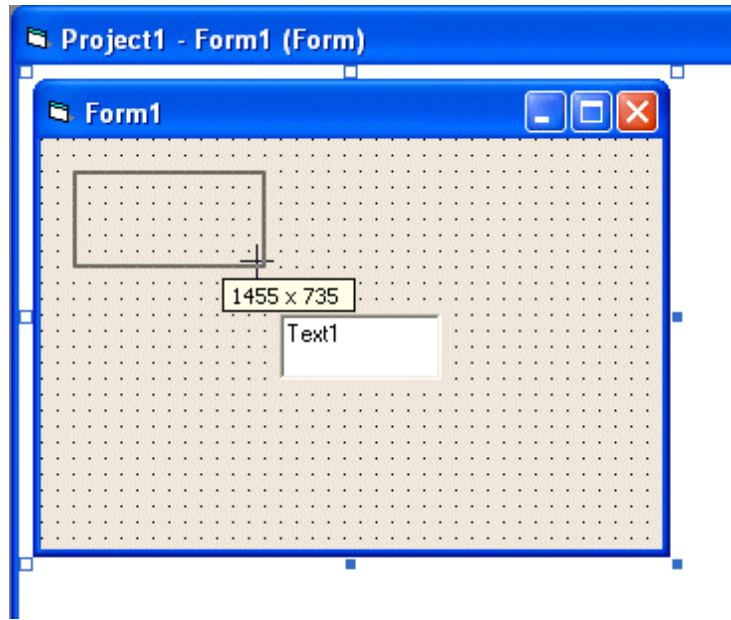
### **Immediate window**

- The **Immediate** window is used at design time to debug and evaluate expressions, execute statements, print variable values, and so forth.
- It allows you to enter expressions to be evaluated or executed by the development language during debugging.
- To display the **Immediate** window, open a project for editing, then choose **Windows** from the **Debug** menu and select **Immediate**.
- You can use this window to issue individual Visual Studio commands.
- The available commands include EvaluateStatement, which can be used to assign values to variables

### **Adding controls to a form**

- To add a control to a form, on the toolbox, you can double-click it.
- Alternatively, you can click the control on the toolbox and then "draw" it on the form.

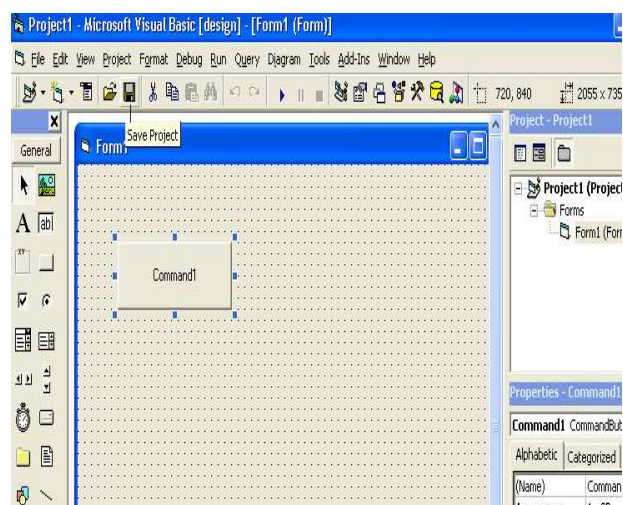
## Placing controls on a form

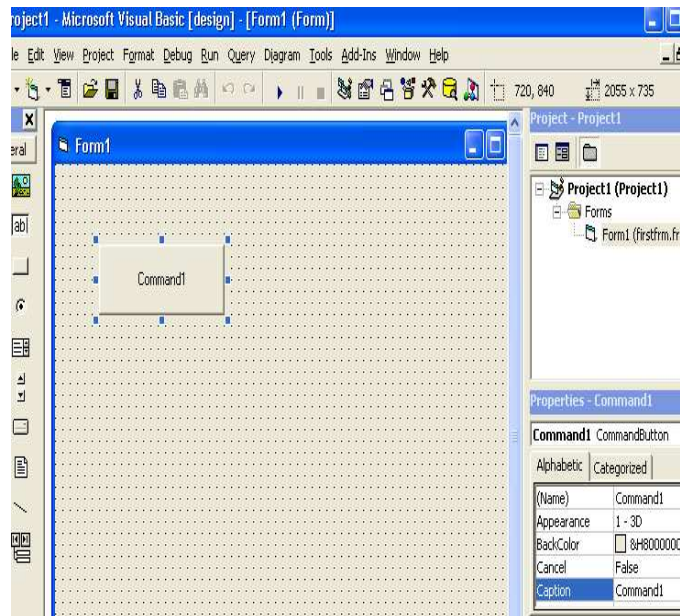
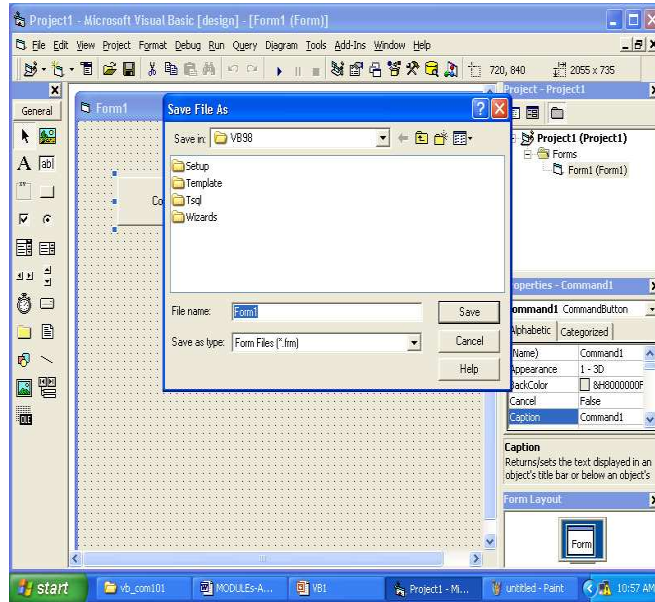


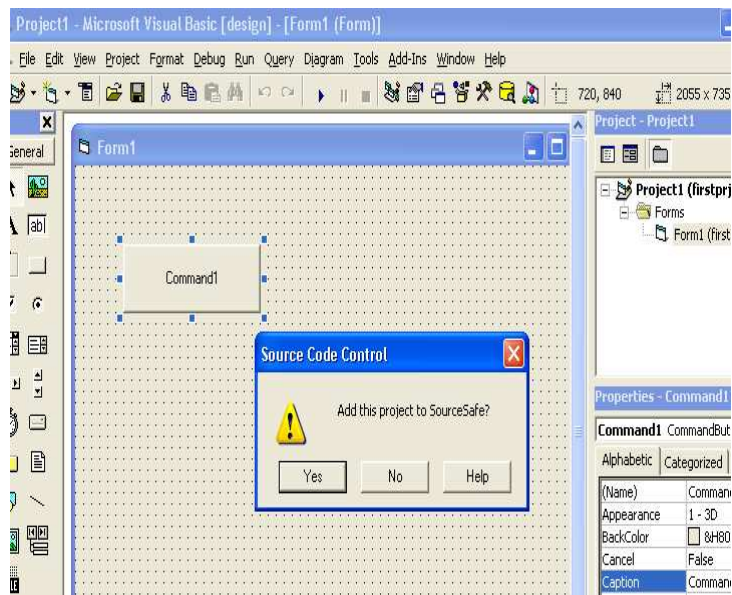
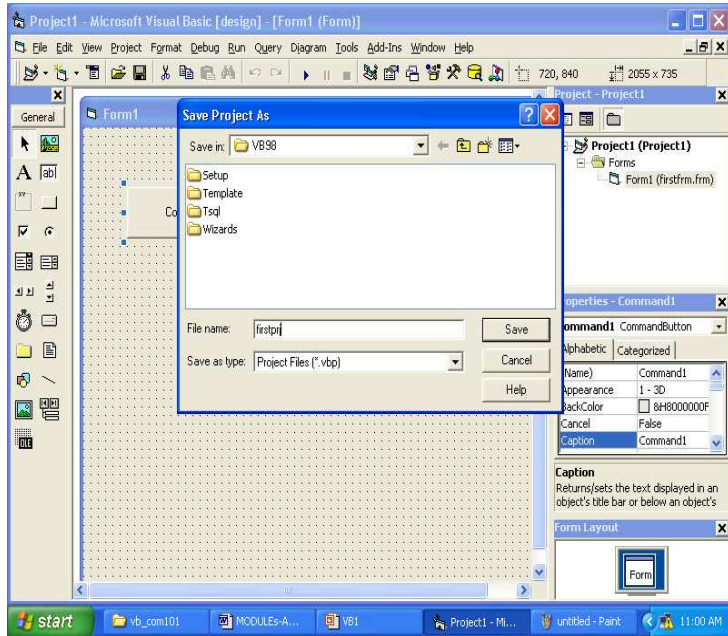
## Deleting Controls in a form

- Click the TextBox on the form to select it
- Press Delete to remove it
- Right-click the CommandButton on the form and click Cut

## Saving the Project

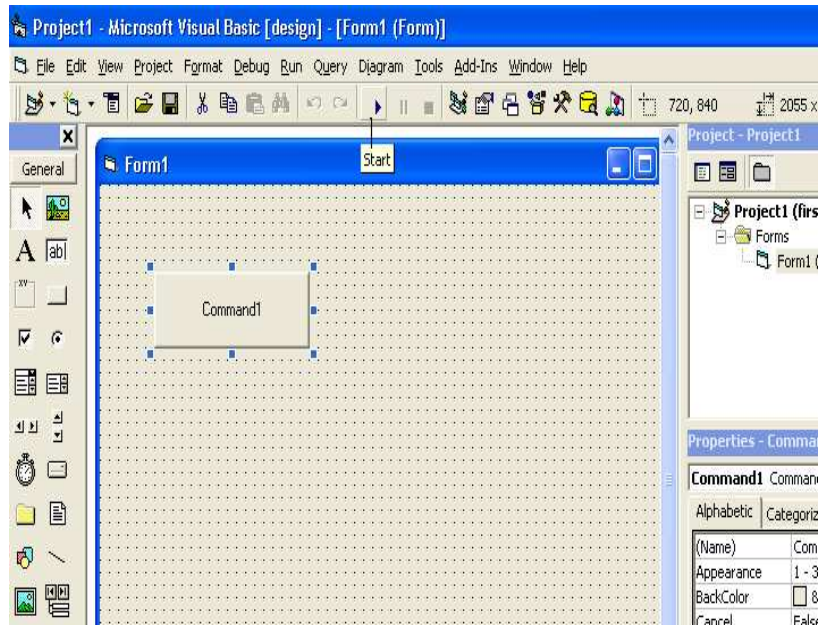








## Run an application



## Objects

- An application is a collection of objects that work together to accomplish something useful. In VB the application is called a **Project**.
- A window is an **object**.

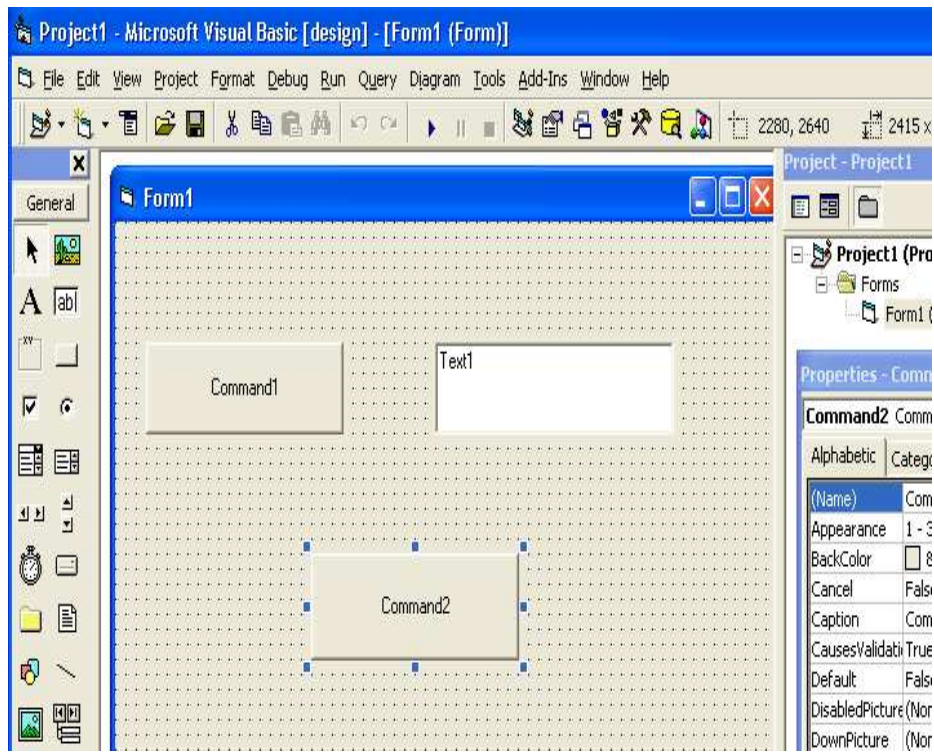
## Event

- In a VB project, the processes that occur have to be associated with events.
- An event is something that happens - the user clicks on a button, a form is opened, the result of a calculation is too large.

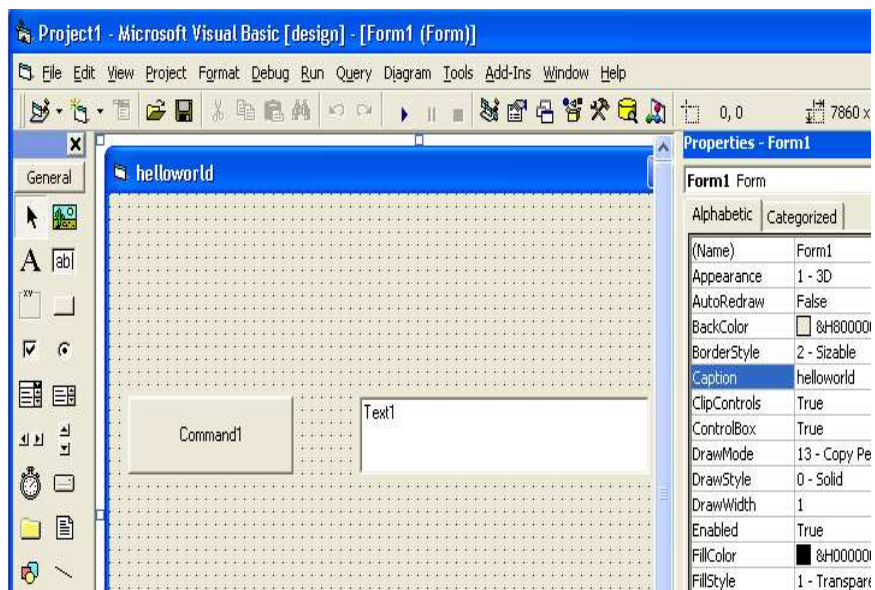
## Hello World Program

- A simple program using
  - 2 command buttons and
  - a text box

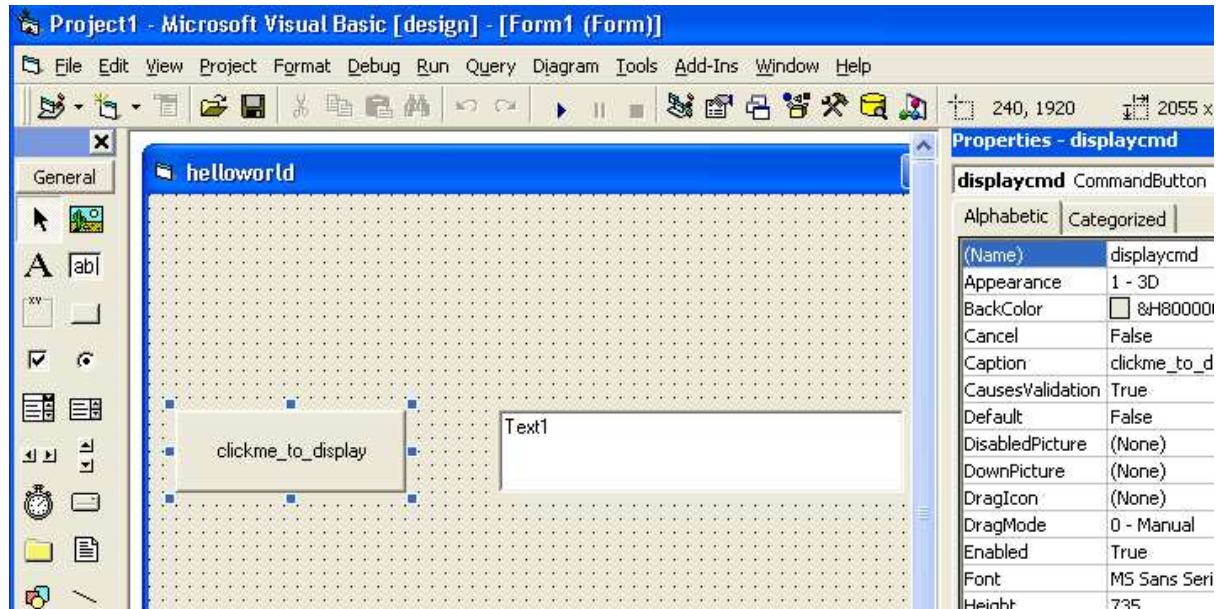




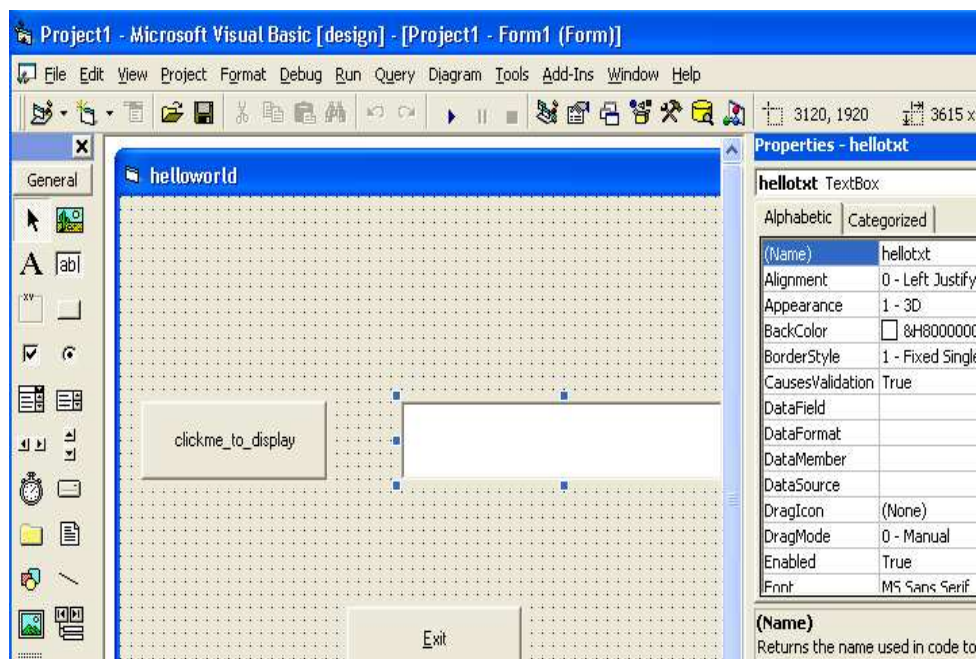
### Setting the property of the form



## Setting the Property of the cmdbutton

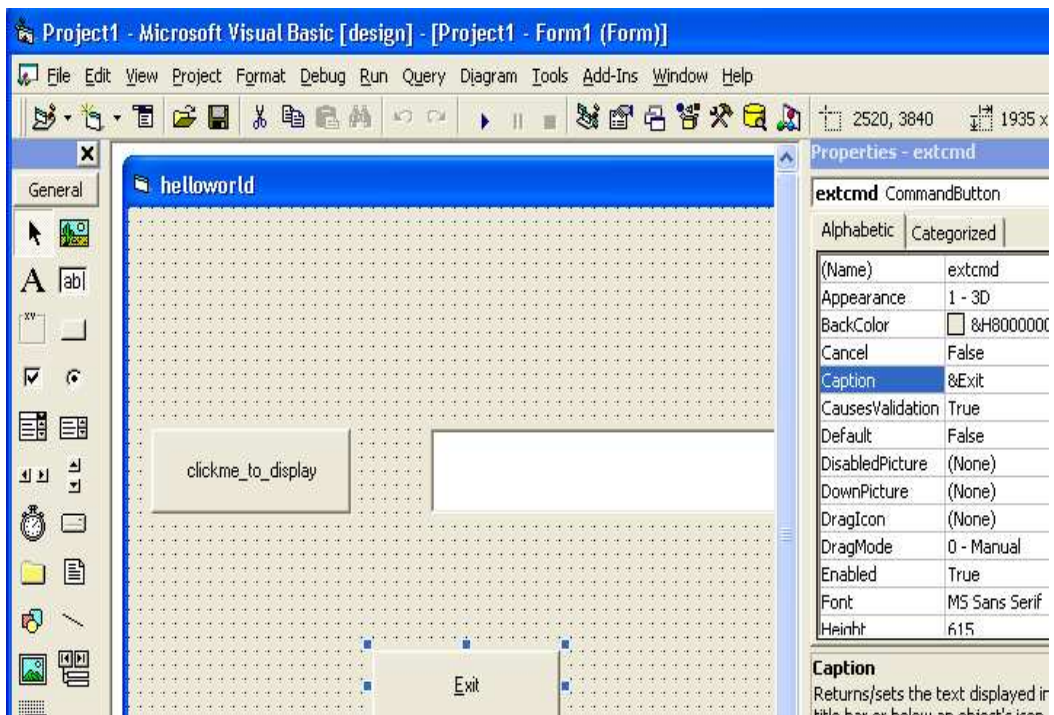
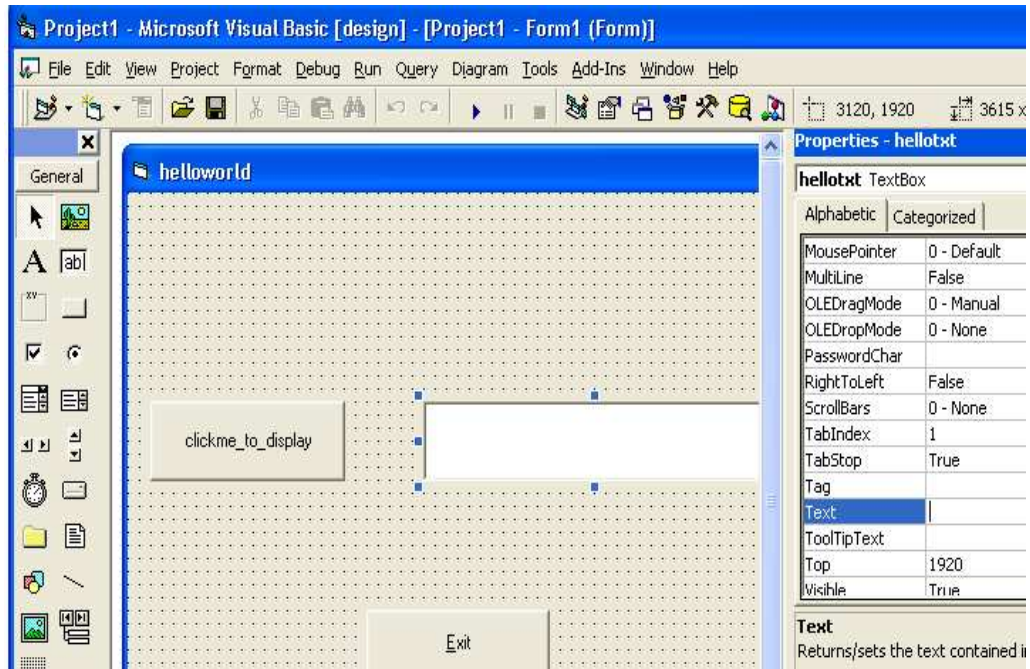


## Setting the property of the name for textbox

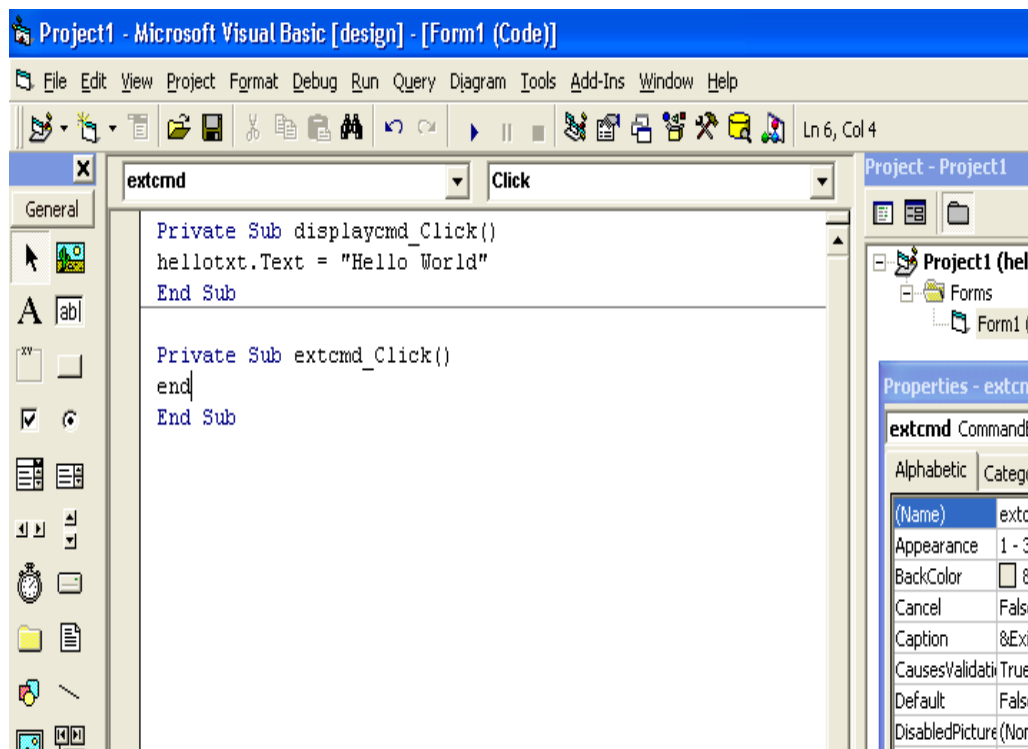
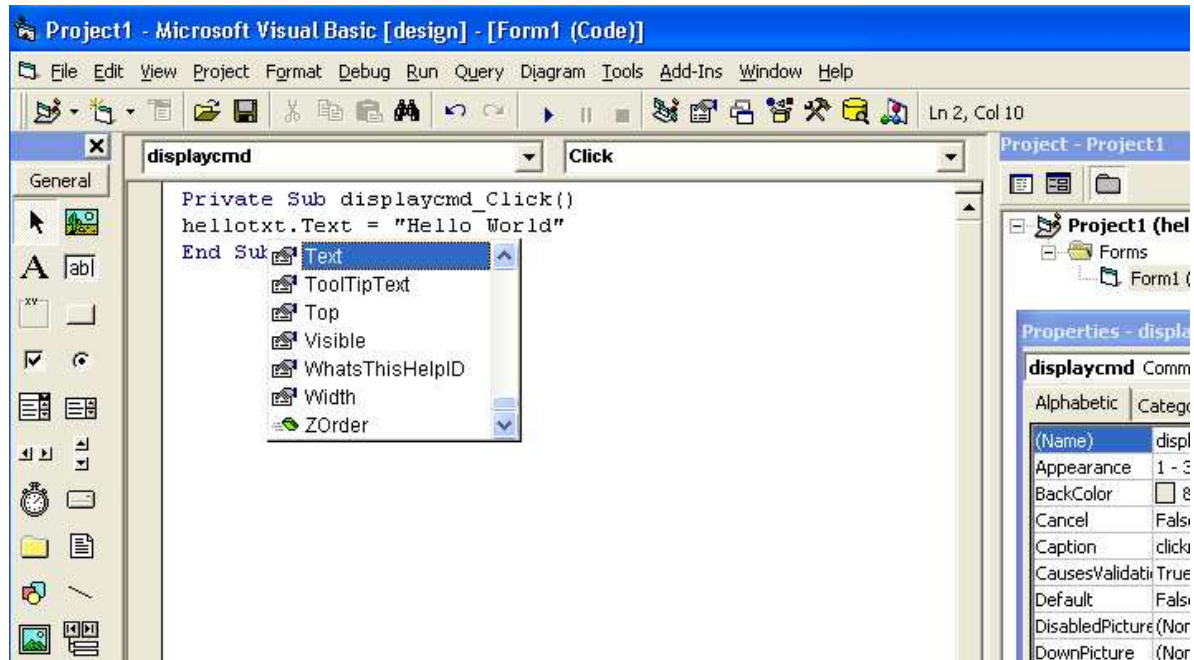




## Setting the Property of the Text for Textbox

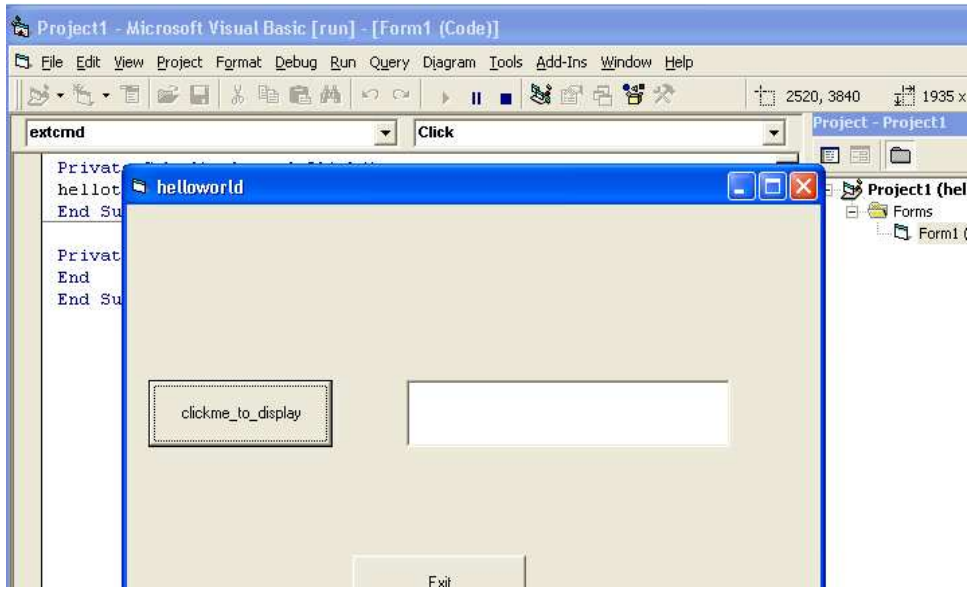


## Code window



## RUNNING THE APPLICATION

- Click the start button



## Variables

- A variable is a value you are ask the computer to store in its memory while the program is running.

### Rules for naming your variable

- The name of a variable must begin with a letter
- Cannot have a period
- Can have up to 255 characters.
- Must be unique inside of the procedure or the module it is used in.
- To declare a variable, type the **Dim** keyword, like this:

Dim

- A variable cannot be used if it has not been primarily declared.

### Option Explicit

This can also be done automatically for each file by checking the Required Variable Declaration in the Options dialog box.

## DATA TYPES

- Before declaring or using a variable, first decide what kind of role that variable will play in your program.
- The kind of variable you want to use is referred to as a data type.
- To specify the kind of variable you want to use, you type the **As** keyword on the right side of the variable's name.
- The formula to declare such a variable is:

**Dim** *VariableName* **As** *DataType*

### String

- A string is an empty text, a letter, a word or a group of words considered.
- To declare a string variable, use the String data type. Here is an example:

```
Private Sub Form_Load ()
```

```
Dim CountryName As String
```

End Sub

```
Private Sub Form_Load ()
```

```
    Dim CountryName As String
```

```
    CountryName = "Great Britain"
```

End Sub

### **Boolean**

- A Boolean variable is one whose value can be only either True or False.
- To declare such a variable, use the Boolean keyword. Example:

```
Private Sub Form_Load ()
```

```
Dim IsMarried As Boolean
```

End Sub

- Boolean variable can be initialized by assigning it either True or False. Example:

```
Private Sub Form_Load ()
```

```
Dim IsMarried As Boolean
```

```
    IsMarried = False
```

End Sub

Like any other variable, after initializing the variable, it keeps its value until you change its value again.

### **Numeric data types**

#### **Byte**

- A byte is a small natural positive number that ranges from 0 to 255. A variable of byte type can be used to hold small values such as a person's age, the number of fingers on an animal, etc.
- To declare a variable for a small number, use the Byte keyword. Here is an example:

```
Private Sub Form_Load ()
```

Dim StudentAge As Byte

End Sub

### **Integer**

- An integer is a natural number larger than the **Byte**. It can hold a value between -32,768 and 32,767. Examples of such ranges are: the number of pages of a book.

### **Long Integer**

- A long integer is a natural number whose value is between -2,147,483,648 and 2,147,483,642.
- Examples are the population of a city, the distance between places of different countries, the number of words of a book.
- To declare a variable that can hold a very large natural number, use the **Long** keyword.

### **Decimal data types**

#### **Single**

- To declare a variable that can hold small decimal numbers with no concern for precision, use the **Single** data type.

#### **Double**

To declare a variable that can store large decimal numbers with a good level of precision, use the **Double** keyword.

In most circumstances, it is preferable to use **Double** instead of **Single** when declaring a variable that would hold a decimal number. Although the **Double** takes more memory spaces (computer memory is not expensive anymore (!)), it provides more precision.

#### **Date**

- To declare a variable that can hold either date values, time values, or both, use the **Date** keyword. After the variable has been declared, you will configure it to the appropriate value. Here are two examples:

Dim DateOfBirth As Date

Dim KickOffTime As Date



## Variant

- A **Variant** can be used to declare any kind of variable. You can use a variant when you can't make up your mind regarding a variable but, as a beginning programmer, you should avoid it.

## FUNCTIONS

### The InputBox ( ) Function

- An InputBox ( ) function will display a message box where the user can enter a value or a message in the form of text. The format is  
**myMessage=InputBox(Prompt, Title, default\_text, x-position, y-position)**
- my Message is a variant data type but typically it is declared as string, which accept the message input by the users.
- The arguments are explained as follows:
  - Prompt - The message displayed normally as a question asked.
  - Title - The title of the Input Box.
  - default-text - The default text that appears in the input field where users can use it as his intended input or he may change to the message he wish to key in.
  - x-position and y-position - the position or the coordinate of the input box.



```

○ Private Sub OK_Click()
  Dim userMsg As String
  userMsg = InputBox("What is your message?", "Message Entry Form", "Enter
  your message here", 500, 700)
  If userMsg <> "" Then
    message.Caption = userMsg
  Else
    message.Caption = "No Message"
  End If

```

End Sub

- When a user click the OK button, the input box as shown in Figure will appear. After user entering the message and click OK, the message will be displayed on the caption, if he click Cancel, "No message" will be displayed.



### Questions

1. Visual Basic form is saved with \_\_\_\_\_ extension

**Ans: [\*frm]**

2. IDE stands for \_\_\_\_\_

**Ans: Integrated Development Environment**

3. In VB the application is called a Project.

**Ans: True**

4. The name of a variable must begin with a letter.

**Ans: True**

5. Name two controls from the toolbox.

6. Name the extension in which the Visual basic form is saved.

7. What are the rules for naming a variable in VB?

8. What is the use of the properties window?

9. Write a hello program using command button and text box.

10. Explain about the data types available in VB.

## References

1. Rangaswamy, R. 2009, A Text book of Agricultural Statistics, Wiley Eastern Limited, New Delhi.
2. Gomez, K.A. and Gomez, A.A., 1984, Statistical Procedures for Agricultural Research, John Wiley and Sons, New York.
3. Panse, V.G. and Sukhatme, P.V. 1961, Statistical methods for Agricultural Workers, ICAR, New Delhi.
4. A First Course in Computers, Sanjay Saxena, Vikas Publishing House Private Limited.
5. Andrew S.Tanenbaum, 1996, Computer Networks, 3<sup>rd</sup> Edition , Prentice Hall of India, New Delhi.
6. Tay Vaughan, 2006. Multimedia: Making IT Work, McGraw-Hill Osborne Media, 7 edition, ISBN-10: 0072264519
7. Byron S. Gotteried, 2002, "Theory and Problems of Programming with Visual Basic", Tata McGraw-Hill Publishing Company Limited, New Delhi
8. Jeffrey R. Shapiro, 2002, "The Complete Reference Visual Basic .Net", Tata McGraw- Hill Publishing Company Limited, New Delhi.

