UNIT - 3

PART - 1

NON-PARAMETRIC TESTS

Points to be covered in this topic

1. INTRODUCTION
2. WILCOXON RANK SUM TEST
3. MANN-WHITNEY U TEST
4. KRUSKAL-WALLIS TEST
5. FRIEDMAN TEST

❑ NON-PARAMETRIC TESTS

- A parameter is a summary value or numerical index for the **population such as mean, median, SD or the variance of a variable**.
- Non-parametric tests are **mathematical procedures concerned with the treatment of standard statistical problems** when the assumptions of normality are replaced with general assumption for the distribution function.
- In order to conclude **data distribution, comparing the means of two populations is very important**. In some experiments the data is not normal, and sometimes the sample size is also so small Thus, for such non-normal data distribution, which is independent of **any assumption about the distribution, an approach used is known as non-parametric tests.**
- These tests are very robust regardless of the **distribution of the data**. Therefore, non-parametric models are sometimes referred to as **distribution-free methods.**
- As non-parametric test **does not match normal data distribution** and thus assumes that it does not depend on arithmetic properties Consequently, all tests involving the ranking of data are non-parametric and also no statement about the distribution of data is made.

(a) Characteristics:

1. It involves only **fewer assumptions.**
2. It works without any **pre-computed statistic** as an estimate of parameter.
3. It can be used when **sample size is too small.**
4. It can be used for **non-normal distribution of the variables.**
5. The methods of computation are **very simple and easy.**

(b) Advantages:

1. If the sample size is very small, **these tests are the only alternative.**
2. These are useful when observations are **nominal, ordinal (ranked), subject to outliers.**
3. These tests are relatively much **easier to learn and to apply.**

4. They involve **comparatively short calculations.**
5. **No assumption** is required to be made about distribution.
6. Suitable for treating samples made up of **observations from different populations.**
7. Suitable to treat data which are **simply classificatory or categorical.**
8. Suitable to **analyze data** which are inherently in ranks as well as data whose apparent numerical scores have the strength of ranks.

(c) Disadvantages:

1. Non-parametric tests are usually **less powerful and less efficient** when the sample size is small or when the normality assumption holds.
2. Non-parametric tests often **require us to modify the hypotheses rather than estimation**. For example, most non-parametric tests about the population center are tests about the median instead of the mean.
3. The results may or may not provide an **accurate answer** because they are distribution free.
4. If all of the assumptions of a parametric statistical method are met in the data and the **research hypothesis could be tested with a parametric test,** then non-parametric statistical tests are wasteful.
5. These tests are **not systematic.**
6. These tests are not convenient to use as Tables required to implement these tests are **scattered widely and appear in different formats.**

(d) Assumptions:

A non-parametric statistical test is based on a model that **specifies or specifies only general conditions and nothing about the specific form** of the distribution from which the sample is drawn. Certain assumptions that are associated with most non-parametric statistical tests are:

1. The observations are independent.
2. The variable under study has underlying continuity.
3. Non-parametric procedures makes different hypothesis about population.
4. It may be applied to data measured in ordinal, nominal or categorical scales.

(e) Test procedure:

The following are general steps to carry out non-parametric tests:

1. **Stating hypotheses:** The null hypothesis Ho, and the alternative hypothesis H_1, are stated.
2. **Setting significance level:** The significance level a related with the null hypothesis is set. It is normally set at 5% and therefore the confidence level is 95%.
3. **Selecting test:** The suitable statistic test is chosen by considering, number of samples, whether the samples are dependent or independent, and the type of data.
4. **Calculating statistic:** The test statistic is then calculated. For small sample, a method particular to a specific statistical test is used. For large samples, the data is approximated to a normal distribution and the z-score is evaluated.
5. **Comparing values:** The value required to reject the null hypothesis is determined using the suitable Table of critical values for the specific statistic. This value is compared with the critical value which enables us to find the difference based on a specific significance level. Then, we can state whether the null hypothesis should be rejected or not. For example, for a two-tailed hypothesis with alpha = 0.05 the null hypothesis is not rejected if - 1.96 <= z <= 1.96.
6. **Making decision:** The results are explained and a conclusion is drawn.

❖ Wilcoxon Rank Sum Test:

- The Wilcoxon rank sum test, also be known as **Mann Whitney Wilcoxon test,** is a non- parametric dependent samples **t-test that can be performed on ordinal (ranked) data.**
- This test is used to test the **null hypothesis**, that is, **the median of a distribution is equal to some value.** It may also be used to assess whether the distributions of observations obtained **between two separate groups on a dependent variable are systematically different** from one another.
- Additionally, it can be used for **ordered categorical data where a numerical scale is inappropriate** but possible to rank the observations.

Steps to Perform Test:

(a) The general way:

Step 1: State the null hypothesis, Ho and the alternate hypothesis, H1.

Step 2: Define a level.

Step 3: Define decision rule

Step 4: Calculate z-statistics

Step 5: Calculate results.

Step 6: Make conclusion.

(b) For paired data:

1. **State the null hypothesis:** It is that the median difference, M, which is equal to zero.
2. **Calculate each paired difference:** Calculate difference (d_i) using equation, $d_i = x_i = y_i$ where, x_i, y_i are the pairs of observations.
3. Rank d_i's ignoring the signs (i.e. assign rank 1 to the smallest $|d_i|$ rank 2 to the next.
4. Designate each rank along with its sign, based on the sign of d_i
5. Calculate W+, the sum of the ranks of the positive d_i and w−, the sum of the ranks of the negative d_i

$(W+) + (W-) = n\ (n+1)^2$ **Where, n is the number of pairs of observations in the sample.**

(c) For single set of data:

1. State the null hypothesis: The median value is equal to some value M.
2. Calculate the difference between each observation and the hypothesized median, $d_i = x_i - M$
3. Rank d_i' s, ignoring the signs (assign rank 1 to the smallest $|d_i|$ rank 2 to the next, etc.)
4. Designate each rank with its sign, based on the sign of d_i
5. Calculate W+, the sum of the ranks of the positive d_i's and w - , the sum of the ranks of the negative d's

$$(W+) + (W-) = n\,(n+1)^2$$

Under the null hypothesis, we would expect the distribution of the differences to be **approximately symmetric around zero** and **the distribution of positives and negatives to be distributed** at random among the ranks. Under this assumption, it is possible to **work out the exact probability of every possible outcome for W**. To continue with the test, procedure given below is followed:

1. Select W equal to minimum of W -,W+

2. Using Tables of critical values for the Wilcoxon signed rank sum test find the probability of observing a value of W or more extreme. Most Tables give both one- sided and two-sided p-values. If not, double the one-sided p-value to obtain the two- sided p-value. This is an exact test.

❖ Mann-Whitney U-Test:

- The Mann-Whitney U-test is a non-parametric statistical method that **compares two groups that are independent of sample data.**
- It is used to test the null hypothesis that the two samples have similar median or, conversely, whether observations in one sample are likely to have larger values than those in the other sample. **The parametric equivalent of Mann-Whitney U-test is t-test of unrelated samples.**

Assumptions:

1. The two samples are random.
2. The two samples are independent of each other, thus observations are within each sample.
3. The measurement scale is of ordinal type; thus, observations can be arranged in ranks.

➢ Calculations:

Consider a **sample of n_a** observations from one population, where, **$a=[a_1, a_2,.....,a_n]$ are from one group** and **sample of n_b** observations from another population, **where, $b=[b_1, b_2,.....,b_n]$** are from another group.

This test is centered on the difference of each observation a_i in the first sample with each observation b_i in the other sample.

The total number of **pairwise difference that can be executed is $n_a n_b$** If the median is similar for both samples, then each a_i has an even possibility of being smaller or larger than each b_j assuming a probability of 1/2.

Hence, Null hypothesis	$H_0 = P(a_i > b_j) = 1/2$
Alternative hypothesis	$H_1 = P(a_i > b_j) \neq 1/2$

- ✓ The number of times an a_i, from sample 1 is greater than b_j, from sample 2 is calculated and represented by R_a.
- ✓ Similarly, the number of times an a_i, from sample 1 is smaller than b from sample 2 is counted and represented by R_b.
- ✓ Below the null hypothesis, R_a and R_b are likely to be almost the same.

Steps to Perform the Mann-Whitney Test:

1. The null hypothesis H_o, and the alternative hypothesis H_1, are identified.
2. The significance level (α), related with the null hypothesis is stated. Usually α is set at 5% and therefore, the confidence level is 95%.
3. All of the observations are arranged in terms of magnitude.
4. Under every observation in table, note down A or B or any other suitable symbol to demonstrate from which sample they come.
5. Below each a, the number of b's, that is lesser than it, is recorded. This implies $a_i > b_j$ Similarly, below each b, note down the number of a's that is lesser than it. This means $b_j > a_i$.
6. The total number of times $a_i > b_j$ is calculated and represented by R_a. Equally, the number of times $b_j > a_i$ is counted and denoted by R_b. The R_a denote the sum of the ranks in group a, and R_b, denote the sum of the ranks in group b. The Mann-Whitney U-test statistic for each sample is determined by:

$$U_a = n_a n_b + \frac{n_a(n_a+1)}{2} - R_a$$

verify that $U_a + U_b = n_a n_b$

7. Evaluate U= min (U_a, U_b). The obtained value is the smaller of the two U statistics.
8. Using Tables of critical values for the Mann-Whitney U-test evaluate the possibility of obtaining value of U or lower. If the test is one-sided, the p-value is the probability been calculated itself and if it is two-sided the probability is doubled to get the p-value.
9. The critical value determined is compared with the obtained value.
10. The results are then interpreted to draw conclusion.

❖ Kruskal-Wallis H-Test:

- The Kruskal-Wallis H-test, also **called H-Test,** is a non-parametric statistical procedure used for comparing **more than two independent samples.**
- The parametric equivalent to this test is the **one-way ANOVA**. The H-test is used for non-normally distributed data.
- This test is a generalization of the Mann-Whitney test which is a test for determining whether **the two samples selected are taken from the same population.**
- The **p-values in both the Kruskal Wallis and the Mann-Whitney tests are equal**. However, Kruskal Wallis test is used for samples to evaluate their **degree of association.**

Description of Samples: This test is performed when different samples of data comply with the following criteria:

1. There must be similar data distributions.
2. Data in each sample should be more than 5.
3. Both distribution and the population have same shape.
4. Data can and must be ranked.
5. Samples must be independent.

Characteristics:

1. **Test statistic is applied when data is not normally distributed or nearly x^2 distributed**
2. **Test equality is more than 2 of the population median.**
3. **Test uses k samples of data.**
4. **Test can be used for one nominal and one ranked variable.**
5. **Significance level is denoted with a.**
6. **Data is ranked and df = n - 1**
7. **The rank of each sample is calculated and average rank is applied in case of tie**

Steps and Calculations:

The hypothesis is stated in terms of populations. While performing H-test it is important to note the observation of the median of the population. The first step is to combine all the samples and rank the order of all the data values and then calculate test statistic using following formula:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

Where, N is number of values obtained from every grouped sample, is summation of ranks taken from a particular sample and n_i is number of values from the equivalent sum of rank. The degree of freedom is determined by using following formula:

$$df = k - 1$$

Where, df is degrees of freedom and k is number of groups or samples.

Then, p-value is calculated. When p-value is less than 5% or greater than 10%, null hypothesis is rejected. While, if p-value falls between 5% and 10% null hypothesis is accepted.

❖ Friedman Test:

- Friedman test (FM) is used for finding differences in treatments across multiple attempts by comparing 3 or more dependent samples.
- This testing procedure consist ranking of each row together followed by considering the values of ranks by columns. No assumptions are made when implementing this test.
- This test may be used as an **alternative to ANOVA, when the assumption of normality is not met.** The test statistic using this test is calculated using ranks of data instead of unprocessed data.
- Since no assumptions are made in this test, it is not as powerful as the ANOVA. If it is equivalent to the Wilcoxon signed-rank test, then this implies that there are only two measures for this test.

- FM test is used to test for differences between groups when the dependent variable being measured is ordinal. It can also be used for continuous data that has marked as deviations from normality with repeated measures.

Descriptions and Requirements:

1. Dependent variable should be measured at the ordinal or continuous level.
2. Data comes from a single group, measured on at least three different occasions.
3. The sample must be drawn with a random sampling method.
4. Blocks are mutually independent (i.e. all of the pairs are independent).
5. Observations are ranked within blocks with no ties.
6. Samples do not need to be normally distributed.

UNIT - 3

PART - 2

INTRODUCTION TO RESEARCH

Points to be covered in this topic

1. INTRODUCTION
2. NEED FOR RESEARCH
3. NEED FOR DESIGN OF EXPERIMENTS
4. EXPERIENTIAL DESIGN TECHNIQUE
5. PLAGIARISM

❑ INTRODUCTION TO RESEARCH

- Research is a process of discovering new knowledge. It can be defined as, **"the creation of new knowledge and/or the use of existing knowledge in a new and creative way to generate new concepts, methodologies and understandings".**
- Research could include **gathering, synthesis and analysis of previous research to the extent that it leads to new and creative outcomes.**
- The goal of research is to extend human's logic, mathematics, science, history, humanities, philosophy, and perseverance of knowledge beyond what is already known.

❖ Need for Research

In the domain of healthcare, **there are gaps in the knowledge and theories about how something might work better** and ideas for its improvement. We **cannot afford to take risks and thus research is needed.** The main objective of research is to inform accomplishments, prove theories, and make contributions to develop knowledge in a field or study. The different needs of research are as follows:

1. **Developing efficient learning process:** It is used to build basic **knowledge, provide information, correct wrong notions** and develop critical thinking habits.
2. **Discover issues:** Through research, issues that one did not know to be in **existence can be uncovered; questions** that were not apparent may arise.
3. **Finding opportunities:** It enables researchers to **assess their potentials, achieve their goals** through available opportunities in the form of employment, funding, training. grants, etc.
4. **Sharing of valuable information:** Reading and writing stimulates computation and understanding. Reading opens the mind to knowledge. **Writing enables individual transform ideas to more concrete perspective.**

5. **Nourishing the mind:** Curiosity stimulates the **mind to search for solutions, allowing creativity and logical reasoning**. Research develops systems that will lead the frontiers of development. **The research manager identifies skilled and qualified professionals** and co-ordinate with them to conduct research.
6. **Understanding issues:** Research is needed to **establish facts, theories or hypotheses** It is needed to disprove what was thought to be truth and to **obtain in-depth information that will make for better understanding of issues**. It can help reveal what we did not existed.
7. **Business success:** Research is needed to **add value to our businesses when research findings lead to new ways** of executing the business or development of new products.
8. **New opportunities:** Research enables us to seek opportunities; we can develop ourselves further and key into new opportunities based on our interests.
9. **Innovation:** Knowledge is **developed and improvement of practical approaches is achieved through research**. Students learn most when actively involved in developing their knowledge. Research is used to initiate innovation and invention.
10. **Legal requirement:** In clinical trials, research is a **legal requirement wherein producers can obtain marketing authorization by proving to the relevant authorities** that the drug is safe and effective. They do this by complying with the regulations and guidelines.
11. **Knowledge building:** Research enables us to **improve our analytical, listening and speaking skills**. These functions improve our understanding and mathematical skills.

❖ Need for Design of Experiments

- A research design is a framework for conducting the research experiments or projects. It describes the steps for **acquiring the information needed to solve research problems.**

- A research design is **used to plan the research**, to show how all of the major components of the research project **contribute to attempt the main research problem.**
- It provides the steps and the **plan for successfully carrying out the study**. It is the strength of the any research protocol, **a document that outlines the planning of study.**
- It increases the chances of gathering all of the **information required to answer a research problem**. If design is sound and strictly follows the protocol, then and then only the **information gathered during research is useful.**
- Strict execution of the procedures and techniques specified in the research protocol increases the chance of results that **the research will be accurate and significant to the other researchers.**
- Research design is significant in **terms of reliability of the results outcomes**. Therefore, it becomes a **firm foundation for the entire research**. It is important because by **employing design the results can be reproduced by the other researchers.**
- If the results are reproduced; **there is more chance that other researchers and the public accept these findings as true**. In addition, the research design makes the **procedures clearer that are used to ensure the protection of research matter**.
- It helps to maintain **the integrity of the information gathered in the study**. Research design facilitates the **smooth completion of the various research activities**. It makes research as efficient as possible by establishing maximum information with minimal expenditure on labor, material, and time and money.
- In fact, research design is an advance planning of the methods, those can be employed in **gathering most relevant data, and the techniques to be used in their analysis.**
- The advantages for research design are:

1. Research design minimizes inaccuracy.
2. It helps to achieve maximum efficiency and reliability.
3. It eliminates bias and marginal errors.
4. It minimizes wastage of resources such as labour, materials and, time and money.
5. It is helpful for identifying and collecting research materials.
6. It is helpful for testing of hypothesis.
7. It provides an overview to other expert researchers.
8. It guides the research activities in the right direction.

A research design is thought to be **based on some methodology**. It should be finalized, **once the research problem is formulated, objectives are properly outlined,** concepts are properly defined and the hypothesis is properly stated.

❖ Experiential Design Techniques

- The **design of fundamental relationship under controlled situation** is called experimental design.
- An experiment is an observation under controlled conditions or it is a design in which **some of the variables being studied are manipulated or searched to control the conditions.**
- Controlling of conditions is a process that **does not allow the conditions to change when the experimental activities** are going on. During experiment, various types of facts are controlled so that the alternative hypothesis can be tested, and fundamental relationship may be obtained.
- In brief, controlling is nothing but **holding one variable constant while other variables are free to vary within the experiment.** Independent variables are manipulated in order to measure their effects on dependent variable, and other variables which may confound.
- The three main types of experimental designs are **pre-experimental or experimental designs, quasi-experimental designs** and observational or true experimental designs.

(1) Pre-experimental designs:

- It is simplest design in which the basic experimental steps are followed without assigning any control group.
- Using this design, the researcher studies a single group with no any comparisons within this group as there is no equivalent non-treatment group.
- They follow some basic steps used in experiments, but unfortunately these designs either fail to include a pretest, a control or comparison group, or both.
- In addition, no randomization procedures are used to control for extraneous variables. They are considered as "pre-", because they are preparatory to true experimental designs.

(2) Quasi-experimental designs:

- In this experimental design, the study subjects are not randomly assigned to the groups. This research design is similar to experimental research. In that, there is manipulation of an independent variable.
- It differs from experimental research because either there is no control group, no random selection, no random assignment, and/or no active manipulation.
- This type of research is often performed when a control group cannot be created or random selection cannot be performed. This is often used in certain medical and pharmaceutical studies.

(3) True experimental designs:

- True experimental design employs statistical analysis to support or reject a hypothesis. This design has control group and can assess cause and effect relationships between the variables.
- A true experiment design is thought to be the most accurate type of experimental research. This is because it supports or disproves a hypothesis using statistical analysis. It is also thought to be the only experimental design that can establish cause and effect relationships.

A true experimental design is supposed to satisfy following conditions:

1. There must be a control group which won't change, and an experimental group which will experience the changed variables.
2. There must be variables those can be manipulated by the researcher.
3. There must be random distribution.

Amongst the three designs described above, true experimental design is the most accurate research design. These true experimental study designs are further categorized into:

(a) **After-only designs:** In this design, the experimental group and the control group are similar. **The uncontrolled extreme factors may affect both the group causally.** The experimental group is exposed to the assumed causal variable (X) and **the control group remains unexposed.** After the experimentation, **both groups are compared, and there is some effect (Y) produced in the experimental group.** It is possible that response (Y) produced might not be by (X), but by the joint interaction of X and other external factors.

(b) **Before-after designs:** In this design, the **dependent variable (effect) is measured both before and after the exposure of groups to experiment.** The experiment in this design may use one or several groups and may have one or more than one control group. **The major weakness of this design is that during an experiment,** a group may be influenced by the external factors in a different ways than the other group, **with no any assurance of equal effects.** Though, **this design is more reliable than after-only design.**

(c) **Ex-post facto designs:** This is the type of experimental research design **wherein researcher depends on historical background** because sometimes it is not possible to **divide the population into two clear and similar groups.** Thus, it may be necessary to study the **entire historical background and the past is studied through the present.**

(d) **Panel study designs:** Panel study is a **method of study of a particular subjects over time by using different kinds of data**. The researcher may secure direct evidence of time dependent relationship among variables. It involves **repeated observations on the same subjects at different period of time** and thus also known as a time-series study. In this design, result variations may be **attributed to a real change in phenomena**. It is a continuous, deep and reliable design, but it has limitations of loss of panel members, absenteeism and rigid attitude of the members, etc.

(e) **Post-test only designs:** This type of design has **two randomly assigned groups, namely an experimental group and a control group**. None of these groups are pretested before the execution of the treatment. **The treatment is applied only to the experimental group whereas post-test is conducted on both groups** to measure the effects of the treatment. When pretesting of the subjects is not possible or not required, this design is commonly used.

(f) **Pretest, post-test only designs:** In this design, the experimental and the control groups are formed, and then **both groups are pretested for the independent variable**. During experiment, only experimental group receives the treatment. Finally, **both groups are post-tested to investigate the effects of the independent variable** on the dependent one.

(g) **Solomon four-group designs :** In this design, 4 groups are randomly formed including **2 experimental groups and 2 control groups**. **Only 2 groups are pretested.** One pretested group and **one un-pretested group receives the treatment**. All 4 groups receive the post-test treatment and, finally, the post-test results demonstrate the effects of the dependent variable comparing to the effects of **the independent variable on the dependent variable**. This design is a combination of pretest, post-test only and Solomon four-group designs. It eliminates sources of error.

(h) **Factorial designs:** In this design, **two or more independent variables are simultaneously manipulated to observe their effects on the dependent variable.** This design is useful to the researcher to test two or more hypotheses in a single experiment/project. **The independent variables in this design are called as factors, and sub-division of factors is called a levels.** This design helps researcher to investigate the individual effect (main effects) of each treatment on the dependent variables as well as their joint effect (interaction effects).

(i) **Randomized block designs:** In this design, subject population is grouped into **relatively homogeneous subgroups** (also called blocks) within which the experiment is replicated. **This design is used when there are inherent differences between subjects and the possible differences in experimental conditions.** If there are a large number of experimental groups, this design is used because it makes the groups homogenous. **It is useful when it needs to reduce the noise or variance in data.** The noise may be attributed to the differences between the blocks.

(j) **Crossover designs:** In crossover design, also known as **repeat measures design; different orders of treatment are randomly assigned to the subjects.** The treatments assigned may be more than one. **The groups for comparison need to have an equal distribution of features as well as there should be a high level of similarity among the subjects.** The subjects in this design may serve as their own control groups. This design has major limitation that the subjects' exposed to first treatment may affect their responses to the second treatment.

- ✓ Experimental research is aimed to draw conclusions by controlling the external factors in order to provide definite conclusions about the effect of the independent variable on the dependent variables.
- ✓ The purpose of experimental research design is to test the hypothesis of a fundamental relationship between variables.

❖ Plagiarism

- Plagiarism is nothing but **stealing and passing off the ideas or words of others** as our own with or without their consent, **using other's production without crediting the source**; committing literary theft, presenting **as new and original idea or product derived from an existing source without full acknowledgement.**
- According to American Association of University Professors' plagiarism is **"Taking over the ideas, methods, or written words of another, without acknowledgment and with the intention that they be taken as the work of the deceiver".**
- It is an unethical practice **in scientific writing** and is an act of fraud. It is a common problem which is **often the result of a lack of knowledge and skills.**
- The expression of original ideas is considered as an **intellectual property** and is protected by **copyright laws through intellectual property rights** (IPR).
- Almost all **forms of expressions fall under copyright protection** if they are recorded in as a book or a computer file. **Copyright, means enhancing knowledge and useful arts by providing limited time security** for authors and inventors through exclusive rights regarding their writings and inventions.
- Although **no degree of plagiarism is acceptable**, it can range from complete plagiarism to **accidental plagiarism.**
- Commonly, plagiarism can be divided into **intentional plagiarism** where the author is fully aware of the plagiarism and is willing to do it and the second is **unintentional plagiarism** where a person plagiarizes due to his or her unawareness and lack of skill in writing.

➢ Classification of Plagiarism

1. **Complete plagiarism:** Complete plagiarism, also known as **global plagiarism**, is the severest form of plagiarism, where a researcher presents an entire text of others as his/her own.

For example, someone submits script as his or her own work, which has been written by someone else. **It also includes plagiarism of unpublished work written by someone else.**

2. **Source-based plagiarism:** This plagiarism falls under serious category and **occurs due to use of different types of sources**, for example, **citing a source that is not accurate or does not even exist.** This plagiarism usually occurs when an **author uses a secondary source of information**, and cite only primary source. **This type of plagiarism unnecessarily increases the number of reference sources** that creates problems for readers to opt for multiple cross- referencing.
3. **Direct plagiarism:** Direct plagiarism occurs when an **author directly copy a passage. of text without citation.** It is also called verbatim plagiarism. It is similar to complete plagiarism, but it refers to part of another script. This type of plagiarism is serious, considered as dishonest and it calls for academic disciplinary actions.
4. **Paraphrasing plagiarism:** Paraphrasing plagiarism is **common and serious type of plagiarism**. It involves **rephrasing of someone else's texts and ideas without citation**. Even if the words or part of sentence differ, the original idea and meanings remains the same.
5. **Mosaic plagiarism:** Mosaic plagiarism is also known as **patchwork plagiarism.** It involves combining text and ideas from different sources without citation. **This is a serious kind of plagiarism which is difficult to detect because it involves inclusion of someone else's phrases** or text within its own research. It is an intentional and dishonest act.
6. **Accidental plagiarism:** It occurs when someone **neglect to cite their sources, misquotes their source, or unintentionally rephrases a source by using similar words, groups of words**, and/or sentence structure without giving due credit. While this plagiarism is serious and real, situations where a plagiarist acts so negligently as to plagiarize are not accidents at all; rather they are the natural outcome of gross negligence.

7. **Self plagiarism:** Self-plagiarism, also called **auto-plagiarism or duplication, is when an author reuses significant portions of his or her own previously published works** without any acknowledgements. For example, publishing a paper which overlaps another paper without due acknowledgement, breaking a large paper into a few smaller papers and publishing them separately and republishing the same full work at different publishers. This is a moderate plagiarism and its severity is under debate.
8. **Inaccurate authorship: Inaccurate authorship is a moderate kind of plagiarism that can happen in two different ways.** First, when someone contributes to a manuscript but does not get credit for it and second, when someone gets credit without contributing to the work.
9. **Casual plagiarism:** It occurs due to **lack of awareness about plagiarism, or insufficient understanding and knowledge of citing sources** or writing references. This type of plagiarism is net result of poor scholarship.

➢ Reasons for Plagiarism:

1. **Lack of understanding:** Some authors plagiarize unintentionally. This usually happens when authors are not familiar with proper ways of quoting, paraphrasing, citing and referencing sources.
2. **Competency gain:** Some authors plagiarize text or ideas to get better opportunities and positions amongst the competitors and to save the time.
3. **Time management:** The authors have to manage their time for many reasons. Thus, to save time they may plagiarize.
4. **Personal values**: Some authors find no reason for why they should not plagiarize or do it because of social pressure, because it makes them feel good.
5. **Defiance:** Some authors' prefer plagiarism as a tangible way of opposing and expressing a lack of respect for their authorities.

6. **Attitude:** Some authors cheat because they have negative attitudes towards assignments and tasks that authorities think have meaning but they do not. Some authors believe that authority does not read their papers or review their work.
7. **Denial:** Some authors deny to themselves that they are cheating or find ways of legitimizing it by passing the blame on to others.
8. **Temptation:** Some authors plagiarize as information becomes more accessible on the Internet and web search tools which make it easier and quicker to find and copy.
9. **Lack of deterrence:** To some authors the benefits of plagiarizing outweigh the risks. Some authors prefer plagiarism when they think of little or no chance of getting caught and there is little or no punishment, if they are caught.

➢ Detecting Plagiarism:

Detecting plagiarism is hard task and this makes plagiarism a threat to the health of scientific literature. **Often plagiarism is recognized by learned reviewers who possess up-to- date knowledge in their own specialist field.** The following include some of the methods that can be used by researchers to detect plagiarism.

1. General sight overview: The academic staff should assess the sentence structure, grammar and idioms used in the author's manuscripts or assignments. They should examine the work which is lower or higher than the author's abilities can afford.

2. Search of online bookstores: Online bookstores help the authorities to decide whether the authors mentioned the right dates for publications or whether the sources used were appropriate to the subject in hand.

3. Search of keywords: Searching keywords in search engines is another tool available to authorities to find instances of plagiarism.

4. Use of plagiarism services: There are many software applications, tools and web sites that can help to detect plagiarized texts. In plagiarism software one can access the plagiarism checker tool, upload the text.

➢ Strategies to Avoid Plagiarism:

Scientific writing requires a great deal of perception that demands exceptional degrees of clarity. In addition, accuracy and transparency are two fundamental aspects and are also critical components of scientific writing. In most cases plagiarism can be avoided by citing sources. Simply acknowledging that certain material has been borrowed and providing readers with the information necessary to find that source is usually enough to prevent plagiarism. Additionally, following considerations would help to avoid plagiarism.

1. Authors must carefully read 'Instructions to Authors' provided by the publisher.
2. Always acknowledge the contributions of others and the source of ideas and words.
3. Use of verbatim text/material must be enclosed in quotation marks.
4. While paraphrasing, understand the material completely and use own words without changing its significance. 5. When in doubt about whether or not the concept or fact is a common knowledge, put it into reference list and provide citation.
5. Always prefer to present results of a single study as a whole.
6. When submitting a manuscript for publication that has already been published, make editors and readers aware about the same through letter.
7. If any doubts about duplication of manuscripts, the authors should alert the editors

UNIT - 3

PART - 3

GRAPHS

Points to be covered in this topic

1. Histogram
2. Pie Chart
3. Cubic Graph
4. Response surface plot
5. Counter Plot graph

❑ GRAPHS

- Graphs and charts are used for easier **interpretation of numerical information**. There are different types of graphs in statistics which can be used to **represent data in a pictorial form.**
- A statistical graph is defined as, **"the pictorial representation of statistical data in graphical form"**. A chart can be used to represent numeric data, functions or some kinds of quality structure in tabular format that provides different information.
- Most commonly used graphs are **line graphs, bar graphs and histograms, pie charts, cubic graphs, response surface plots and counter plots.**

Histogram

- Histogram is a specific type of **bar chart,** where categories are ranges of numbers.
- A histogram is **an area diagram**. It can be defined as, "a **set of rectangles with bases** along with the intervals between **class boundaries and with areas proportional to frequencies** in the corresponding classes".
- In histogram, all the rectangles are **adjacent since the base covers the intervals between class boundaries**. The heights of rectangles are proportional to corresponding frequencies of similar classes and for different classes; **the heights will be proportional to corresponding frequency densities.**
- Histogram illustrates the distribution of **numeric data across categories and therefore show combined continuous data**. For example, histograms can be used to present percentage of patients who received number of drugs in hospital.

❖ Steps to Construct a Histogram:

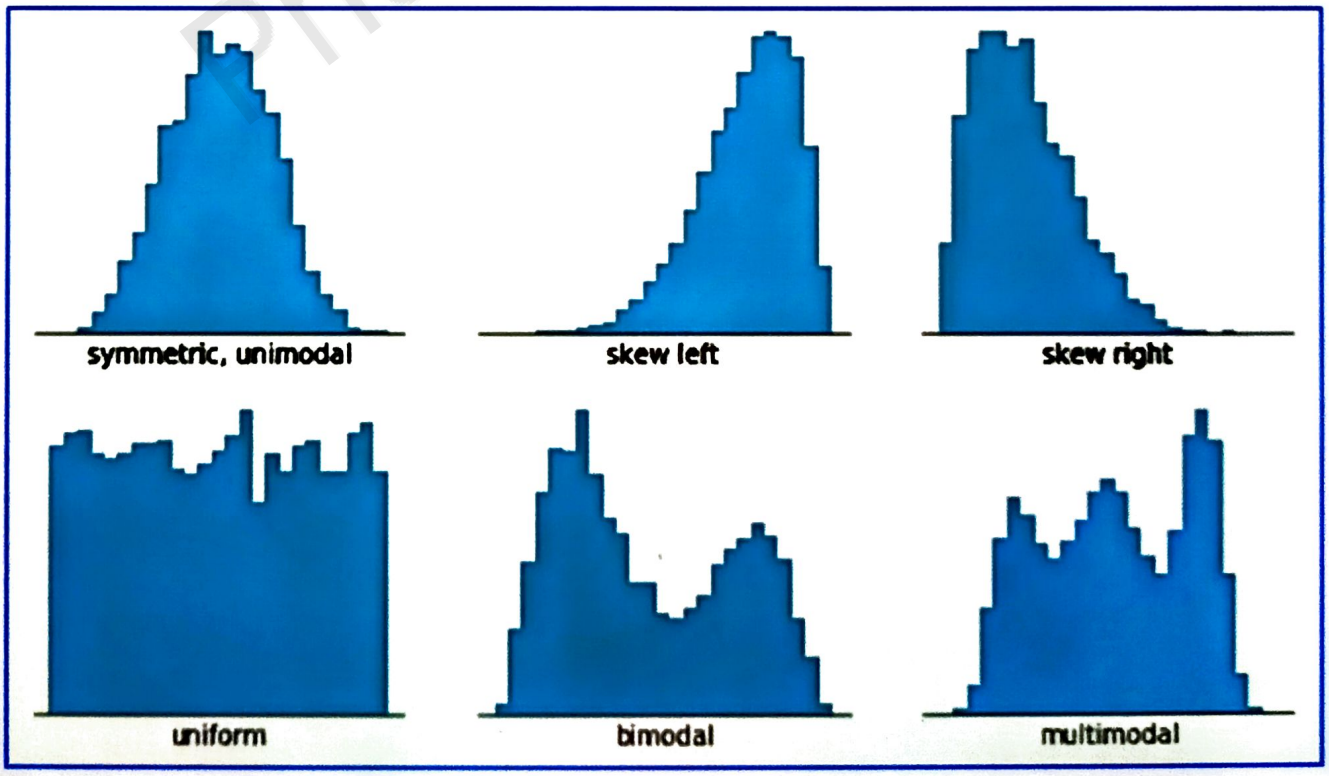
1. Begin by marking class intervals on x-axis and frequencies on y-axis.
2. The scales for both the axes have to be same.
3. Class intervals need to be exclusive

4. Draw rectangles with bases as class intervals and frequencies as heights.
5. A rectangle is built on each class interval because the class limits are marked on the horizontal axis, and the frequencies are on vertical axis.
6. If the intervals are equal, the height of each rectangle is proportional to the corresponding class frequency.
7. The area of every individual rectangle is proportional to the corresponding class
8. frequency if the intervals are unequal. The histogram does not involve any gaps between the two successive bars

❖ Types of Histograms:

1. Uniform histogram: A uniform distribution in histogram reveals that, **the number of classes is too small,** and each class has the same number of elements. It may involve distribution that has several peaks.

2. Symmetric histogram: When we draw vertical line down the centre of the histogram, and the two sides are identical in size and shape, the histogram is said to be symmetric. The diagram is perfectly symmetric if the right half portion of the image is exactly identical to the left half, The histograms that are not symmetric are known as skewed histograms.



3. Probability histogram: A probability histogram shows a pictorial representation of a distinct probability distribution. It consists of a rectangle centered on every value of x, and the area of each rectangle is proportional to the probability of the corresponding value. The probability histogram diagram is begun by selecting the classes. The probabilities of each outcome are the heights of the bars of the histogram.

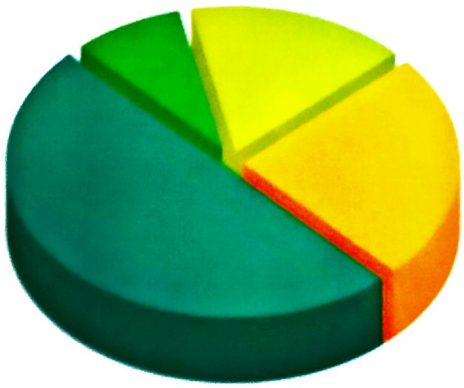
4. Bimodal (side-by-side) histogram: A histogram is unimodal if there is one peak, bimodal if there are two peaks and multimodal if there are many peaks. The bimodality feature results when the data set has observations on two different kinds of individuals or combined groups and if the centers of the two separate histograms are far enough to the variability in both the data sets.

Pie Chart

- A pie chart is a pictorial diagram that shows **how total amount is divided between levels of a categorical variable**. It looks like a circle (or a pie) divided into radial slices or sectors.
- The slices are of **different sizes based on how much of the whole they represent.** Each slice is labeled to represent its value to the whole. It is used to show dependent data that how the whole population breaks down into parts.
- Usually pie chart is not appropriate to use for samples with **more than 5 or 6 different categories.**
- Pie charts are used in **research and business presentations to demonstrate population segments**, outcome responses and budget allocations.
- The pie chart can also be used to **illustrate the numerical problems and also to find out the composition of something.**Pie charts, are preferred choice as contains different segments and sectors and **each segment or sector of a pie chart forms a certain portion of the total percentage.**

- The total of **all the data is equal to 360° and total value of the pie is 100%** (i.e. 360° 100%). The steps performed to compute the percentage of sector a pie chart include categorization of the data, calculating the total, dividing the categories and converting them into percentages and finally, calculating the degrees. The pie chart formula for segment is given by:

Pie Chart



$$\text{Segment (\%) in a pie chart} = \frac{\text{given data}}{\text{total value of data}} \times 3600$$

❖ Uses of Pie Chart:

Pie chart can be used for displaying data as an alternative to table formats. These are useful for displaying data that are classified into nominal or ordinal categories. It helps to categorize data into different ranks as very poor, poor, fair, good, very good. Pie charts are used to show percentage or proportions of data. These charts are good for displaying data for around 5-7 categories or fewer which becomes easier to interpret.

❖ Advantages:

1. The pie picture is simple and easy-to-understand.
2. Data can be represented visually as a fractional part of a whole.
3. It helps in providing an effective communication tool for uninformed audience.
4. Provides a data comparison for the readers at a glance for immediate analysis.
5. No need to examine underlying numbers which can be removed in this chart.
6. Easy to manipulate pieces of data in the pie chart.

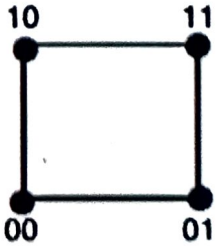
❖ Disadvantages:

1. Pie chart becomes less effective, if there are too many categories of data to use.
2. If there are too many pieces of data, they may become crowded and difficult to read.
3. As this chart only represents one data set, we need series to compare multiple sets.
4. Pie charts with many categories are difficult to analyze and assimilate information

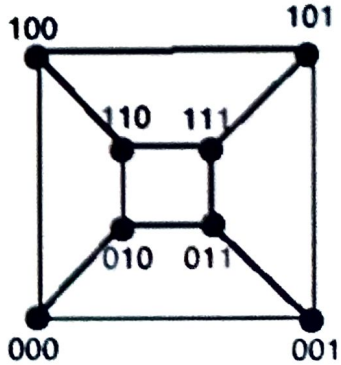
Cubic Graphs

- A cubic graph is one that is obtained by taking all vertices denoted as binary words and joining the vertices with the edge whenever the binary words differ by 1.
- It is a graph in which all vertices have degree 3. Therefore, a cubic graph is a 3-regular graph and is also called as trivalent graphs.
- These graphs are symmetric or semi-symmetric. A graph is said to be regular if all its vertices are of same degree and k-regular if all its vertices are of degree k.
- A 3-regular graph is also called a cubic graph. A vertex with degree zero is an isolated vertex, a vertex with degree one is a pendant vertex and the unique edge incident to a pendant vertex is a pendant edge.
- A vertex of odd degree is an odd vertex and a vertex of even degree is an even vertex.

k = 1



k = 2



k = 3

❖ Properties of Cubic Graph:

1. Cubic graph can be colored with at most three colors. Contour plot a
2. Every connected cubic graph has an independent set of at least n/3 vertices, where n is the number of vertices in the graph.
3. Cubic graphs arise naturally in topology as simple polyhedra and polyhedra.
4. The path width of any n-vertex cubic graph is at most n/6.
5. Every cubic graph has an even number of vertices.
6. Every cubic bridgeless graph has a perfect matching.
7. Every cubic bridgeless graph has an exponential number of perfect matching.
8. Every cubic bridgeless graph with n vertices has at least 2/3656 perfect matching.
9. Cubic graphs can be used to model many types of relations and process dynamics in computer science, physical and biological and social systems

Response Surface Plot

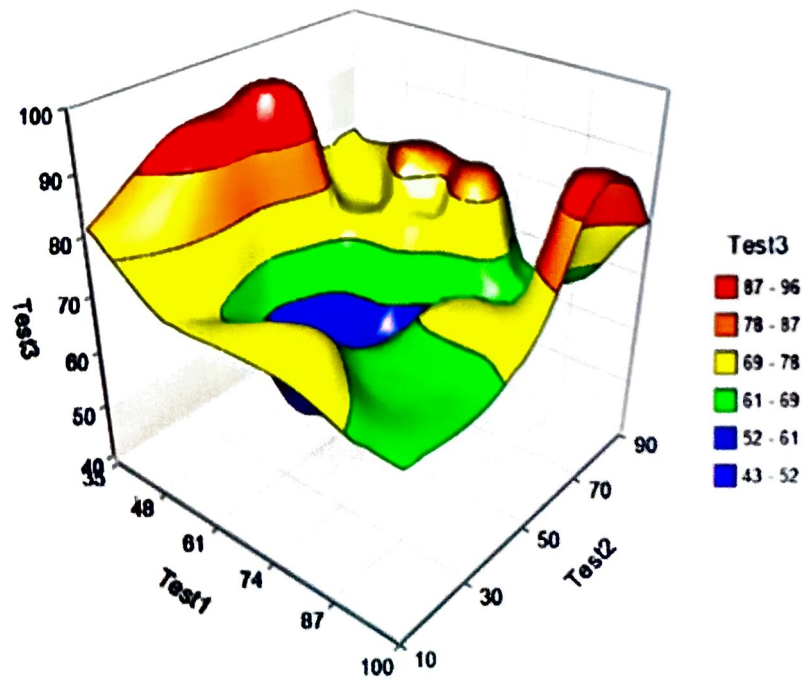
- Response surface plots are 3-D diagrams that show a **functional relationship between a dependent variable (y), and two independent variables (x and z)**. In order to construct surface plot, first 2-D grid of x and z variables is constructed.
- The range of this grid is equal to the range of the data. It is followed by **calculating y value for each grid point**. The y value is a weighted average of all data values that are near this grid point. **The number of points to be averaged needs to be specified.**
- Then the 3-D surface is constructed using these averaged values. Therefore, surface plot does not show any variation at each grid point.
- Surface plots are useful in regression analysis for viewing the relationship among a dependent and two independent variables.

- The **multiple regressions assume that surface is a perfectly flat surface** and hence it helps to visually determine if multiple regression is appropriate. Generally, a surface plot is constructed from three variables using Windows MS Excel.
- The condition to construct the surface plot is that all three variables must be numeric. The x variable is displayed along the horizontal axis, the y variable is displayed along the vertical axis, and the z variable is displayed on the depth axis.
- Surface charts are useful to find **the optimum combinations between two sets of data**. The general stepwise procedure to create a surface plot using MS Excel is given below

1. Arrange the data in columns or rows and select the data.
2. On the INSERT tab, in the Charts group, click the Stock, Surface or Radar Chart icon to select the Surface chart. Different sub-types of surface charts are 3-D Surface, Wireframe 3-D Surface, Contour and Wireframe Contour.
3. When mouse point on each of the icons, a preview of that chart type appears on the worksheet.
4. Select the suitable chart type from the list.

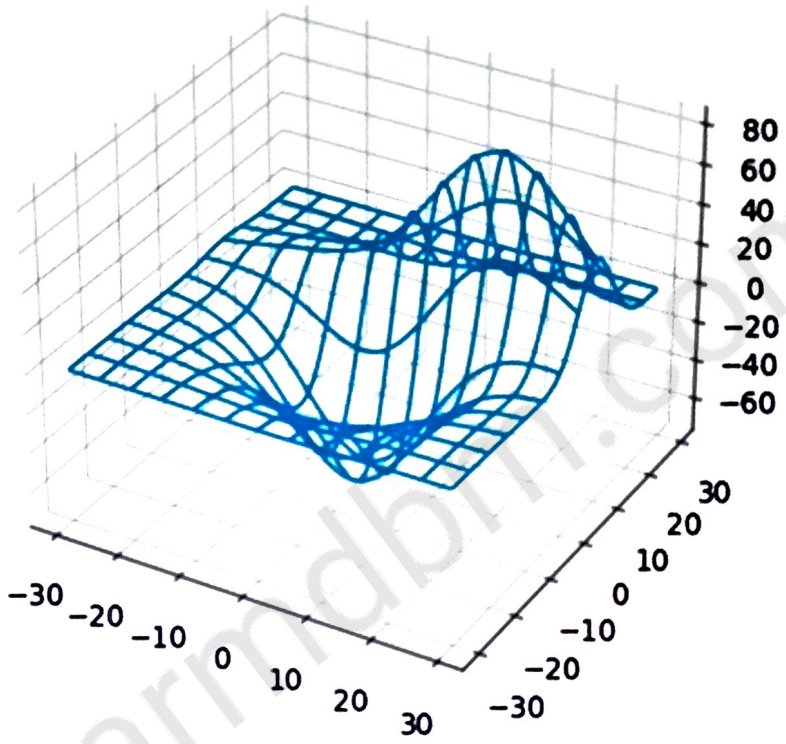
❖ 3-D surface plots:

A 3-D surface chart shows 3-D view of the data, which can be imagined as a rubber sheet stretched over a 3-D column chart. Color bands in a surface chart do not represent the data series but it indicates the difference between the values.



❖ Wireframe 3-D surface plots:

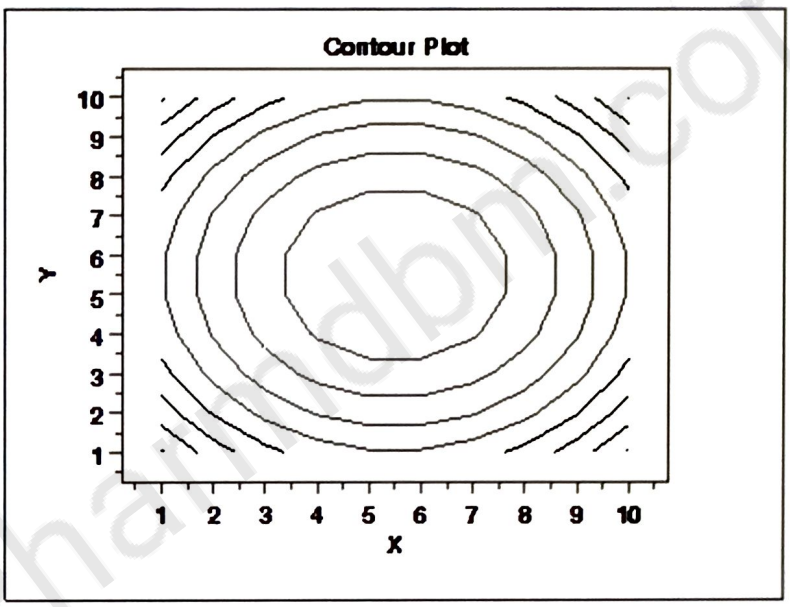
This is a 3-D surface chart shown without color on the surface. This type of surface **chart shows only the lines**. This type of surface chart is difficult to read, but it can plot large data sets much faster as compared to 3-D surface chart. **Wireframe 3-D surface chart is used to show the trends in values across two dimensions in a continuous curve**, when the categories and the series are both numeric values and when the data curves behind itself.



Counter Plot Graph

- Contour plots are similar to the response surface plots. **A contour plot is a graphical technique for representing a 3-D surface by plotting constant z slices, called contours, on a 2-D format.**
- For example, if values for z are given, lines are drawn for connecting the (x, y) co- ordinates where that z value occurs.
- The contour plot is an **alternative plot to a 3-D surface plot**. This plot shows symmetry of the surface and peaks in the center.
- The contour plot is formed by vertical axis (independent variable 2), horizontal axis (independent variable 1) and lines (iso-response values).

- The independent variables are usually placed on x-axis. The determination of correct iso- response values is complex and is always done by using computer software.
- An additional variable may be required to specify the z values for drawing the iso-lines. Some software packages require explicit values while some software packages determine them automatically.
- If the data is not suitable to form a regular grid, in such cases it needs to use 2-D interpolation to form a regular grid. The contour plot is used to obtain information about how does z change as a function of x and y.
- The contour plot is a specialized type of plot used in the design of experiments, especially for full and fractional factorial designs.



❖ 3-D Counter Plot:

- Contour plots are surface charts that are viewed from top and are the form of statistical software programs which vary widely in their capabilities to plot and generate contour plots.
- Many programs provide simply the basic contour plot over a rectangular grid while higher versions permits color filled or shaded contours. Statistical software programs that supports design of experiments also has a DOE contour plot facility.

- The counter plots can also be generated using Windows MS Excel program. In these plots the color bands represent specific ranges of the values and the lines connect the interpolated points of equal value.
- Contour charts are used to show the 2-D top view of a 3-D surface chart, to represent the ranges of the values using color and when both the categories and the series are numeric.

❖ Wireframe Contour Chart:

- Wireframe contour charts are also surface charts viewed from above. These charts shows only the lines without the color bands on the surface but are not easy to read.
- Thus, usually 3-D surface chart is used. Wireframe contour chart is used to show the 2-D top view of a 3-D surface chart only with lines and when both the categories and the series are numeric.
- In order to look at the case where x and y do not form a perfect grid, an example is a surface.
- If we add noise to the x and y values, it will not have a regular grid with which we may draw the wireframe. The surface command will round the x and y values to reduce the number of different values and then attempt to draw the frame.

UNIT - 3

PART - 4

DESIGNING THE METHODOLOGY

Points to be covered in this topic

1. Sample size determination and
2. Power of Study
3. Report writing
4. presentation of data
5. Protocol,
6. Cohorts studies,
7. Observational studies
8. Experimental studies,
9. Designing clinical trial, various phases

❑ DESIGNING THE METHODOLOGY

- Designing the methodology refers to the **development of a system or method or plan for a solving research question at hand**. The key purpose of design methodology is to find the best solution for each experimental condition. **It involves brainstorming to encourage innovative ideas and collaborative thinking to work through each proposed idea** and achieve the best solution. Design methodology can employ basic research methods which involves analysis as well as testing

Sample Size Determination

- In research methodology a **sample is a set of data collected and/or selected from a statistical population by a defined method**. The elements of sample are sample points and sampling units.
- A most important and critical aspect of any research study is **determining the appropriate sample size to answer the research question.**
- Sample size determination means **selecting the number of observations to include in a statistical sample.**
- The sample size is an important feature in **making inferences about a population from a sample.**
- The sample size in a study is determined on the **basis of data collection costs and efforts and the need of sufficient statistical power**. Studies should be designed to include a sufficient number of samples to adequately address the research question.

❖ Considerations for Sample Size:

1. Confidence: Confidence is the idea of being certain that the estimate based on the sample correctly represents the population. This is used in association with an interval, within which the true and unknown population value is expected to reside. Confidence provides the result's ability to be convincing. Low confidence implies the results based on the sample data are not convincing.

2. **Significance:** Significance is the idea that the results are not due to random chance alone. It is the notion that there is convincing evidence based on the sample data that there really is a difference. This is commonly used when accepting the alternative hypothesis with a specified level of statistical significance.
3. **Variance:** The higher the population's variation or spread, the more samples will be needed to determine the same result over a smaller variation population. The chance of selecting samples further from the mean value will require more samples to get an accurate estimate than the population that has a very tight variance. The more precise or smaller the difference that we want to detect, the more samples will be required.
4. **Population size:** The larger is the population; larger will be the sample size.
5. **Appropriate sample size:** Selection of proper sample size is must to make conclusions about a population. Selection should be made on the basis of sampling risk, population's variance and amount of change to be detected. To minimize risk related to the sample selection that represents the population better is to consider large samples. For population with homogenous units' small sample size is enough but for population that has heterogeneous units large sample size is must to obtain reliable results.
6. **Nature of study:** If study is intensive and continuous, small sample size is enough but for studies which are not likely to be repeated may select large sample size.
7. **Other considerations:** The availability of trained expert, use of appropriate sampling technique, categorization of population, time constraints, budgetary provisions etc.

1. One Sample, Continuous Outcome:

The required sample size is estimated to minimize the margin of error. The margin of error (E) in the one sample confidence interval (μ) can be calculated as follows:

$$E = Z\frac{\sigma}{\sqrt{n}}$$

Where, n is sample size, E is margin of error, Z is the value from the Table of probabilities of the standard normal distribution for the desired confidence level (for example, Z = 1.96 for 95% confidence) and a is the SD of the outcome of interest.

2. One Sample, Dichotomous Outcome:

To estimate the proportion of successes in a dichotomous outcome variable in a single population, the formula for determining sample size is:

$$n = P(1-P)\left(\frac{Z}{E}\right)^2$$

- Where, Z is the value from the standard normal distribution reflecting the confidence level that will be used (Z = 1.96 for 95%), E is desired margin of error and P is the proportion of successes in the population.
- When a study is planned to generate a 95% confidence interval for the unknown population proportion (P), the equation to determine the sample size for determining P requires knowledge of P. In fact, we need an approximate value of P or an anticipated value. The range of P is 0 to 1, and therefore the range of P(1 - P) is 0 to 1. The value of P that maximizes P(1 - P) is 0.5. If no information is available to approximate P, then 0.5 can be used to generate the most conservative (as largest possible) sample size.

3. Two Independent Samples, Continuous Outcome:

In order to estimate the difference in means between two independent populations, the formula to determine the sample sizes required in each comparison group is given by equation

$$n_i = 2\left(\frac{Z\sigma}{E}\right)^2$$

Where, n_i is the sample size required in each group (i = 1, 2) , Z is the value from the standard normal distribution reflecting the confidence level that will be used and E is the desired margin of error, and σ is the SD of the outcome variable

4. Matched Samples, Continuous Outcome:

In order to estimate the mean difference of a continuous outcome based on matched data, the formula used for determining sample size is:

$$n = \left(\frac{Z\sigma_d}{E}\right)^2$$

Where, Z is the value from the standard normal distribution reflecting the confidence level that will be used (Z = 1.96 for 95%), E is the desired margin of error, and σ_d is the SD of the difference scores.

5. Two Independent Samples, Dichotomous Outcome:

In order to estimate the difference in proportions between two independent populations, that is to estimate the risk difference, the formula for determining the sample sizes required in each comparison group is:

$$n_i = \{P_1(1-P_1)+P_2(1-P_2)\}\left(\frac{Z}{E}\right)^2$$

Where, n_i is the sample size required in each group (i = 1, 2) Z is the value from the standard normal distribution reflecting the confidence level (Z = 1.96 for 95%), and E is the desired margin of error. The terms P_{1} and P_{2} are the proportions of successes in each comparison group.

Power of Study

- The statistical power of a study is its ability to detect a difference if a difference really exists. The power depends on the sample size (number of subjects), and the effect size (difference in outcomes between two groups).
- For common studies that involve comparison between two groups, for example, comparison of blood pressure levels between smokers and non-smokers the t-test is usually used.
- The power of such study is relatively small and easy to compute if the sample size and the difference in blood pressure between the two groups are known.
- Many such small studies are under-powered to detect a true difference because they do not have enough sample size, and research ends up with a large insignificant p-value.
- The major reason for the lack of significance is a sample size. The freely available software package G*Power can be used to compute power of the study.
- It also determines the desired effect size, or the sample size, for a given power. This is very useful software for planning sample size for a study as well as post-hoc analysis of studies to examine if they had enough power.
- Nearly all studies require studying a subject with a particular characteristic rather than the whole population. These subjects are then used to draw inferences about the whole population.
- These power estimations are used to determine number of subjects required to answer the research problem (or null hypothesis). Generally, a power of 0.80 (80%) or higher is considered good for a studies.
- This means, there is an 80% chance of detecting a difference as statistically significant, if a true difference exists. The higher is the power of a study, the more are the subjects and/or the larger is the effect size (or the smaller is the p- value).

- With a very large number of subjects, a study may have good power to detect even a very small effect size.
- To detect a small difference between groups, higher number of subjects is required to contain adequate power and vice-a-versa. Along with types of potential bias, the limitations and types of validity; statistical power and the factors that influence it affects the effect size, sample size and the p-value.

While determining the statistical power of study we need to consider:

1. The null hypothesis that, there is no difference between the treatments in terms of mortality.
2. The alternative hypothesis that, there is a difference between the treatments in terms of mortality.

❖ Reasons to run a power analysis:

We can run a power analysis for many reasons that include:

1. To find the number of trials needed to get an effect of a certain size.
2. To find the power, given an effect size and the number of trials available..
3. To validate research. Conducting power analysis is a good science

Calculating power is complex and is usually performed with a computer. There is a list of links to online power calculators available at https://statpages.info/#Power

❖ Factors affecting a power of study:

(b) There are several factors that can affect the power of a study.

1. Sample size.
2. The precision and variance of measurements within any sample.
3. The true value of the parameter being tested.
4. Magnitude of a significant difference.
5. How certain we want to avoid Type I error.
6. Statistical test applied.
7. Significance level.

Report Writing

A research report is a formal account of **how a scientific research project was conducted:** and what its outcome is. **Writing research reports can be one of the most difficult tasks for the researchers.** It involves long and stressed duration with difficult concepts, trying to produce a cohesive description of how a research project was conducted and what is

❖ Significance of Research Report:

- For the researchers, a research report provides a lasting record and reminder of the work accomplished and its outcomes.
- The feedback other people give on the report can help extend the researchers' knowledge and understanding of a topic. It can also help them to work on research design and management skills for future projects.
- For readers/audience, a research report makes available the project information to read about the project's aims, methods and findings, assess the quality of the project, provide feedback to the project's researchers on what they like or dislike about the project and incorporate aspects of the project's methods or findings into their own work or thinking.

❖ Potential readers of research reports:

- The potential readers of any research project are other researchers, academicians or knowledge workers, managers in organizations who use research to inform decisions, front- line staff delivering programmes or services, clients or consumers of programmes or services, people living in communities and neighborhoods, journalists and other media representatives and the general public.
- Using information provided in research reports, people can increase their knowledge and understanding of a topic. This can lead to changes in people's thinking and behaviour, or to the design and operation of human services and systems.

❖ Report Format:

- Report formats are likely to vary with the researcher or research firm conducting the project, the client for whom the project is being conducted, and the nature of the project itself.
- The usual steps involved in writing report includes logical analysis of the subject- matter, preparation of the final outline, preparation of the rough draft, rewriting and polishing, preparation of the final bibliography and writing the final draft.
- Most research reports are divided into three parts as: formality part, main report and appendix. The detailed description for the elements of the report is described below.

(A) First Part (Formality Part):

1. **Cover page:** A cover page of research report enables the reader to immediately read the title of the report and realize what the report is about. A good cover page gives all the essential information about a report on the first page such as the title, year of publication and the name(s) of author(s) and their affiliations.
2. **Submission letter:** A formal report generally contains a letter of submission that delivers the report to the client and summarizes the researcher's overall experience with the project, without mentioning the findings. A
3. **Title page:** The title page include the title of the report, information (name, address and telephone number, email address) about the researcher or organization conducting the research, the name of the client.
4. **Certificate:** The certification page of a project report is where researcher confirms that the research was carried out by him or his team. In a certificate, researcher states that the research work is original and was conducted by him.
5. **Acknowledgement:** In this section the writer acknowledges and show appreciation to everyone who has helped in the project.

6. **Table of contents:** The table of contents is an organized listing of the chapters and major sections of research report. This includes main sections, list of tables, list of graphs, list of appendices and list of exhibits.
7. **Preface:** A preface introduces the subject matter to a reader. It should be brief, and it comes before anything else that researcher write.
8. **Abstract:** An abstract is a research summary report which is a brief overview of what the whole research is about. It comprises of summary of prime objectives, major findings and conclusions and recommendations. It briefly but perfectly illustrate the core meaning of the research altogether.

(B) Main Report (Central Part of Report):

1. **Introduction:** An introduction should state the research topic; provide background. and a rationale for the work, before stating research questions and hypothesis. It lays the foundation for the report, catch the reader's interest, and communicate the hypothesis statement.
2. **Problem definition:** This section of the report gives the background to the problem. This part also summarize statement of the problem and statement of the research objectives.
3. **Methodology and research design:** This section is focused to the theoretical foundations that guided the research, any analytical models formulated, research questions, hypotheses, and the factors that influenced the research design.
4. **Data analysis:** This section presents quantitative or qualitative data analysis and also describes the plan of data analysis that justifies analytical strategy and techniques used.
5. **Results:** In this section, researcher reports the findings of his study based upon the information gathered as a result of the methodology or methodologies applied.
6. **Limitations:** As no study is completely flawless or inclusive of all possible aspects, limitations of the study are incorporated as point of significance.

7. **Summary, Conclusions and Recommendations:** A research summary provides a brief overview of the whole study. It is essential to identify the important information in a study, and condense it to make it a quality contents for the readers. It includes interpretations, conclusions and the findings or results of an investigation. Recommendations are usually followed by conclusions, and writers opinions are supported by the report's findings.

(C) Appendix (Additional Details):

- At the end of the report, various technical documents are compiled that may be used by different readers to help them to understand characteristics of the research project in more detail.
- These documents include the letter of authorization to conduct the research; this authorization could include the agreed research proposal including statement of expenses.
- Details that relate to individual techniques should be included relating to questionnaires, interview guides, sampling and fieldwork activities. The final part of the appendix should include lists of contacts, references used and further sources of reference.
- Any other relevant information may include errata, publication or any funding generated out of said research work.

Presentation of Data

- Research report writing and its presentation to audience is the final step in the any research project. This process begins with interpretation of data analysis results and leads to conclusions and recommendations.
- Finally, the formal report is written and an oral presentation is made. After audience has read the report, the researcher should conduct a follow-up, assisting audience and undertaking a thorough evaluation of the research project.

- An oral presentation provides a chance for researchers to present their research by reading a paper and/or showing PowerPoint slides to a group of relevant audience (viz. faculty, students, judges, managers/other researchers/client firm/any related authority).
- Presentations, allow researchers to experience what it is like to present their research at conferences. Presentation helps the audience to understand and accept the written report.
- Any preliminary questions that the audience may have can be addressed in the presentation. Because many executives may form their first and last impressions about the project based on the oral presentation, and thus its importance cannot be over emphasized.
- The key to any effective presentation is its preparation. A detailed outline should be prepared following the format of the written research report. The presentation must be focused to the audience.
- For this purpose, the researcher should determine the backgrounds, interests and involvement of all those in the project, as well as the extent to which they are likely to be affected by it.
- The presentation should be rehearsed several times before it is made to the audience. Visual aids including figures, tables and graphs should be displayed using variety of media.
- Flip charts of large pads of blank paper mounted on an easel (support stand) enable the researcher to manipulate numbers. They are particularly useful in communicating answers to technical questions.
- Visual supplements can be prepared on the pages in advance, and the presenter flips through the pages during the presentation.
- Although not as flexible, over head projector and felt boards allow for instant presentation of previously prepared information. Overhead projectors are suitable to present simple charts as well as complex overlays produced by the successive additions of new images to the screen. Currently, use of computer packages such as Microsoft's PowerPoint is of immense help.

Research Protocol

- Once the research study is properly and completely planned, the plan should be written down. **Protocol is defined as a document that describes the background, rationale, objectives, design, methodology, statistical considerations, and organization of a research project.**
- The written protocol compels the investigators to simplify their thoughts and to think about all aspects of the study. **Protocol is a necessary guide if researchers are working as a team on the project.**
- It is essential, if the study involves **research on human subjects or is on experimental animals in order to get the institution's ethical approval.** In addition, it is an essential component of a research proposal submitted for funding.
- The first step before beginning and developing a detailed protocol is **selection of a research topic.** The topic should be able to stand alone as an explanation of the study.
- In doing so, **researchers need to understand prevailing operational realities and how to work within their limits.** During development of the protocol, investigators can and should obtain advice of colleagues and experts in refining their plans.
- Once a protocol for the research study is **developed and approved, and the study is started and progressing,** it should adhere strictly to the protocol and should not change. Any violations of the protocol can discredit the whole study. **If the changes made are minor, that part of the study should be excluded from the analysis.**
- After writing the protocol, particularly in large studies, **key is to develop the operations manual for the study.** This includes detailed instruction to the investigators in order to assure a uniform and standardized approach for performing study with good quality control.
- A well-thought out and well-written protocol can be judged on the basis of following criteria:

1. **Adequacy to answer the research question/s and achieve the study objective.**
2. **Feasibility in the particular set-up for the study.**
3. **Provision of enough details that can allow another investigator to do the study and arrive at comparable conclusions.**

The protocol should outline the **rationale for the study, its objective, the methodology used and how the data will be managed and analyzed**. In addition, it should indicate how ethical issues have been considered and handled, how gender issues are being addressed. While formulating a **scientifically sound research protocol, its various elements should be taken into consideration.** The commonly followed research protocol along with its elements is written according to format described below.

1) **Project title**: The title **should be descriptive and concise**. It can be revised once the protocol writing is completed to reflect closer logic of the study.
2) **Protocol summary:** The summary should be **concise, and should summarize all the elements of the protocol**. It is a sketch plan of the study which should stand on its own, and not refer the reader to points in the project description. **Summary of the protocol is usually written after completion of the protocol writing.** Commonly, it is located at the beginning in the protocol.
3) **Introduction:**

(a) Study question: Protocol should start with a **clear and precise formulation of the research question**. It may be a good practice to write this in the form of a question and not as a statement. **While formulating research question, its relevance, feasibility of conducting the study to answer the question and likelihood of implementation of the findings** must be kept in mind. The more precise the question, the more likely it is that research will provide new knowledge.

(b) Rationale: Rationale should answer the questions such as **why the research needs to be done and what will be its relevance**. A brief description of the most relevant studies published on the subject should be provided to support the rationale for the study. **The rationale states how the research question arose from current knowledge about the subject.** The progression of ideas needs to be set out in a logical sequence. **Researcher must be very concise to include key references, but not a complete review of the literature**. The following points must be taken into consideration while writing this section:

1. Discuss the importance of the topic.
2. Review relevant literature and current knowledge.
3. Describe any results that have been already gathered in the area of proposed study.
4. Indicate how the research question has emerged and fits logically.
5. Outline the approach to address the research question.
6. Explain how proposed study will benefit the community.

4) Objectives: Specific objective is a statement of the research question. It should be. **simple and specific, and must be stated in advance**. Next to the statement of **the primary objective, secondary objectives, if any, may be described in brief**. It is advised to resist the temptation to put too many objectives or over-ambitious objectives that cannot be adequately achieved upon implementation of the protocol.

5) Design and methods

(a) Study design: This section should state the selected design of the study. The study designs can be **Cross-sectional study, Cohort study, Case-control study or Experimental or Intervention study**. Protocol needs to explain why a particular study design has been chosen in preference to others.

(b) Study population: This section defines **the group in which the study will be conducted and to whom will the results refer**.

(c) Sample size: The protocol should provide information and justification for the selected sample size. **If study involves human subjects it is unethical to expose human subjects unnecessarily to any potential unnecessary risk without additional benefits.** The basis for selected sample size should be explained in the methodology section of the protocol.

(d) Study subject: This section should explain **how many subjects will be selected, where and why**. It includes defining **eligibility, inclusion and exclusion criteria and mechanism of selection**. It must give information about method employed to estimate numbers of potentially eligible subjects.

(e) Data collection methods: It is essential to state how the **data will be gathered to obtain the outcome of the study**. Quality control procedures should also be specified.

(f) Data management and statistical analysis: The protocol should provide **information on how the data will be managed, including data coding for computer analysis, monitoring and verification**. In addition to statistical methods planned for the analysis of data, detail information about the available computer facility should also be provided.

6) Project management

(a) Personnel: This section identifies the manpower requirement to conduct the study. and define their tasks. **It must justify proposed manpower in terms of their expertise, tasks and the amount of time required.**

(b) Action-plan: Action-plan (time schedule) is the anticipated time required for each phase of the study, including **pilot testing, recruiting of subjects, preparation of study questionnaires or formats, logistics including personnel and training, data collection, follow- up procedures**.

7) **Strengths and limitations**: It is important to include the possible criticisms of proposed design and methods. It must provide reasons why investigator thinks the limitations imposed by his choices are not the serious ones. Similarly, it is useful to identify important aspects of the protocol that are particularly strong and worthy of financial support.

8) **Ethical considerations:** Ethical considerations are applicable to all types of health research. The ethical issues involved in the study like benefits of the study procedures, hazards, responsibility of injury; voluntary consent, etc. must be specified.

9) **Expected outcome:** This section restates the **justification for the study in terms of the anticipated results.** It will specify the implications of the potential results and, how the results of the study may be useful to the policy makers, community at large and for future research.

10) **Budget summary:** In this section, a brief outline of the budget requirement showing head-wise expenditure such as manpower, logistics, transportation, etc.

11) **References:** A list of the various references quoted while formulating protocol may be listed in a sequential manner. **The references should follow uniform format.**

12) **Annexures:** These include study formats and questionnaires that are designed in such a manner that it will contain all the required information in a systematic manner with a brief introduction on the study.

13) **Curriculum-vitae of investigators:** This section describes role of investigators in the study, and clearly states the responsibilities for each component of the study. The curriculum-vitae should provide a clear description of the qualification and experience of the investigators, including training, academic degree (certificates), research experience and scientific publications. The protocol should end with this section.

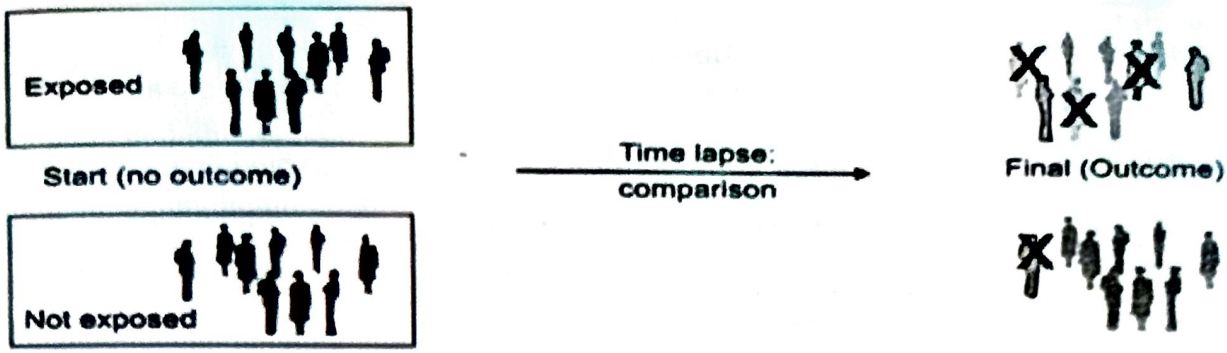
Cohorts Studies

- The word cohort means a **group of people**. Cohort studies are a **type of medical research studies used to determine the causes of disease and to establish relationships between risk factors and health outcomes**.
- The cohort study design identifies a people exposed to a particular factor and a comparison group that was not exposed to that factor and measures and compares the incidence of disease in the two groups.
- Cohort studies require **long periods of follow-up since disease may occur a long time after exposure**. When risks are computed in a study, the risk ratio is the measure that compares the $Risk_{exposed}$ to the $Risk_{non\text{-}exposed}$.
- The risk ratio is defined as, **"the risk in the exposed cohort (the index group) divided by the risk in the unexposed cohort (the reference group)".** Cohort's studies can be forward-looking (prospective) or backward- looking (retrospective)

❖ Types of Cohort Studies:

(i) Prospective cohort studies: These studies are **planned in advance and carried out over a future period of time**. In these studies the investigators visualize and **design the study, recruit subjects, and collect baseline exposure data from all subjects**, before any of the subjects develops any of the outcomes of interest. The subjects are then observed into the future to record the development of any of the outcomes of interest. **The follow up is usually conducted by mail questionnaires, by phone interviews, via the internet, in person, etc.**

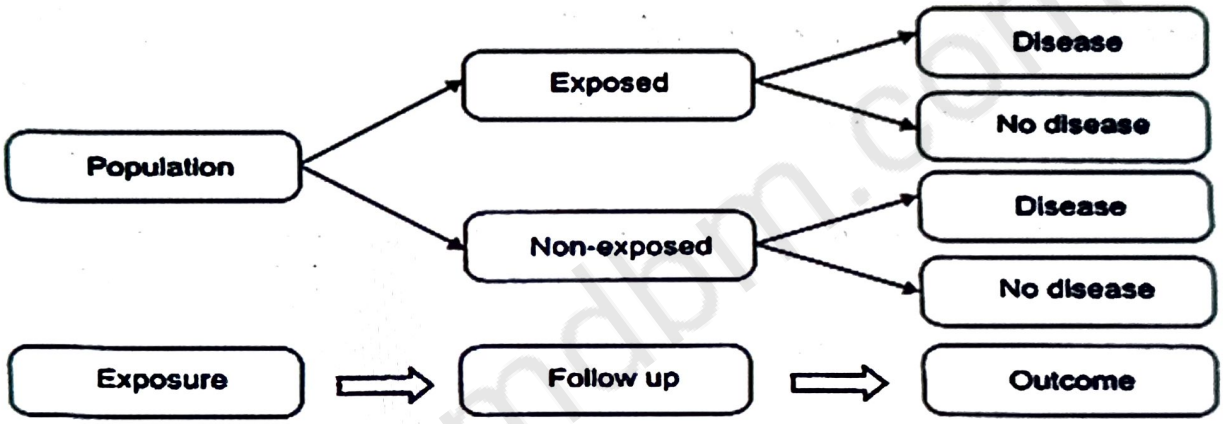
(ii) Retrospective cohort studies: In these studies, subjects are grouped on the basis of **their exposure status and are compared for their incidences of disease**. Both, exposure status and outcome in these studies are determined retrospectively. These **studies examine data that already exist and try to identify risk factors for particular conditions.**



A cohort study concept

❖ Cohort Study Design:

These studies consist of two groups namely; **first group which is exposed to some risk factor** and **second group that is free from exposure.**



Schematic Design of Cohort Study

- Cohort studies typically observe **large groups of individuals and record their exposure to certain risk factors** to find reasons about the possible causes of disease.
- These long-term studies are sometimes called **longitudinal studies.** The cohort study design is the best available **scientific method to measure the effects of a suspected risk factor.**
- In cohort study, researchers raise a question and **form a hypothesis about probable cause of a disease.** This is followed by observing cohort over a period of time which may take several years.
- They collect data that may be relevant to the disease. **Through this procedure, researchers aim to detect any changes in health linked to the probable risk factors** those have been identified.

- Cohort studies are **best to find and establish relationships between health and environmental factors** such as chemicals in the air, water and food.
- Pooling data from different studies is helpful to **increase the sample size. The large sample size makes the results more reliable**, especially for rare conditions such as some types of cancer.
- Measures of relative effect express the outcome in one group relative to that in the other. Non-cases may be **enrolled from a well-defined population, current exposure status** (at time to) determined, and the onset of disease observed in the subjects over time.
- Disease status at time t_1 can be compared to exposure status at time t_o. The data collected may be presented in tabular format as shown in Table.

Cohort 2 × 2 study design table

	Event	No Event	Total exposure
Exposed	Success (S_E)	Fail (F_E)	$Total_{Exposed}$ (N_E)
Non-exposed	Success (S_C)	Fail (F_C)	$Total_{Non\text{-}exposed}$ (N_C)
	$Total_{cases}$	$Total_{non\text{-}cases}$	**Total**

❖ Calculation of statistic:

The measurement of disease frequency and effect/association is calculated as:

(1) Incidence density (Incidence Rate):

Among exposed: Incidence Rate $= \frac{S_E}{N_E}$

Among non-exposed: Incidence Rate $= \frac{S_C}{N_C}$

(2) Incidence density ratio (Risk ratios or Relative risk, RR) $= \frac{\frac{S_E}{N_E}}{\frac{S_C}{N_C}}$

(3) Attributable risk $= \frac{S_E}{N_E} - \frac{S_C}{N_C}$

Risk ratio is, "the ratio of the risk of an event in the two groups". The odd ratio is, "the ratio of the odds of an event"

(4) Odd ratio (OR) = $\frac{\text{Odds of event in experimental group}}{\text{Odd of event in control group}} = \frac{\frac{S_E}{N_E}}{\frac{S_C}{N_C}}$

(5) Risk difference = (Risk of events in experimental group)- (Risk of events in control - group)

$$= \frac{S_E}{N_E} - \frac{S_C}{N_C}$$

❖ Advantages of Cohort Studies:

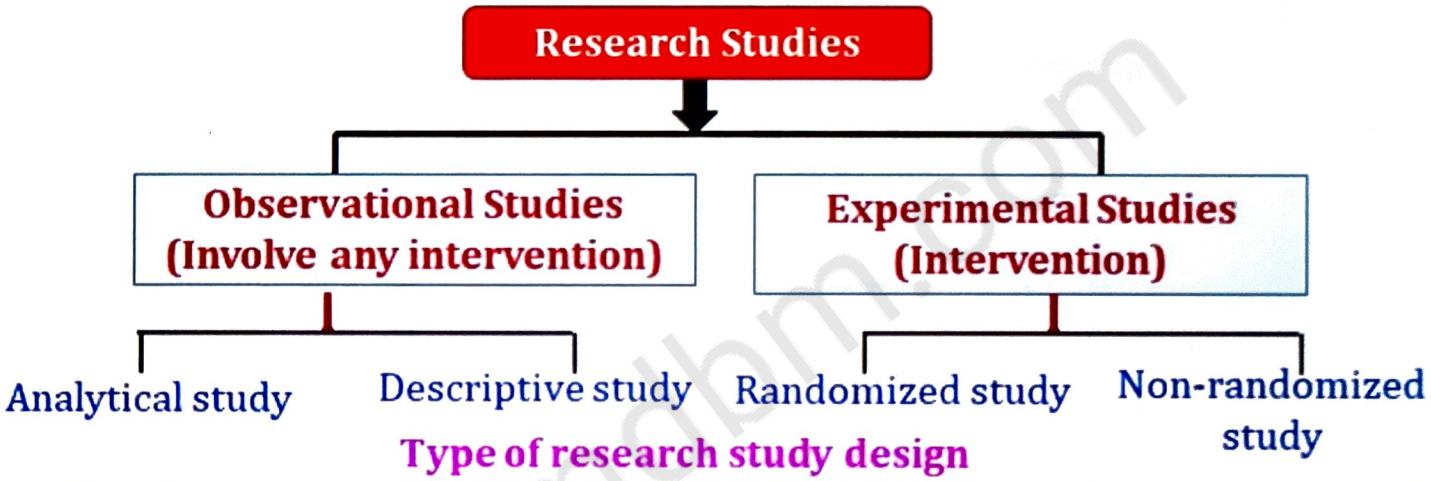
1. Cohort studies precisely indicate **the chronological sequence between exposure and outcome.**
2. Cohort studies allow **calculating the incidence of disease in exposure groups.**
3. Cohort design can be used to **investigate common exposures, for example, exposure to toxic chemicals**, adverse effects of drugs or treatments, etc.
4. Allow examination of **multiple effects of a single exposure.**
5. Cohort studies **reduce the possibility** that the results will be biased

❖ Limitations Cohort Studies:

1. As cohort studies **follow exposure data and watch for any emerging cases of disease** they are less suited to finding clues about rare diseases.
2. Unsuitable for identifying the causes of a **sudden outbreak of disease.**
3. They are **expensive to run and take many years**, often decades, to produce results.
4. They can **only offer clues about the causes of disease** rather than definitive proof.
5. Participants may **leave the cohort for various reasons** that can bias the results.

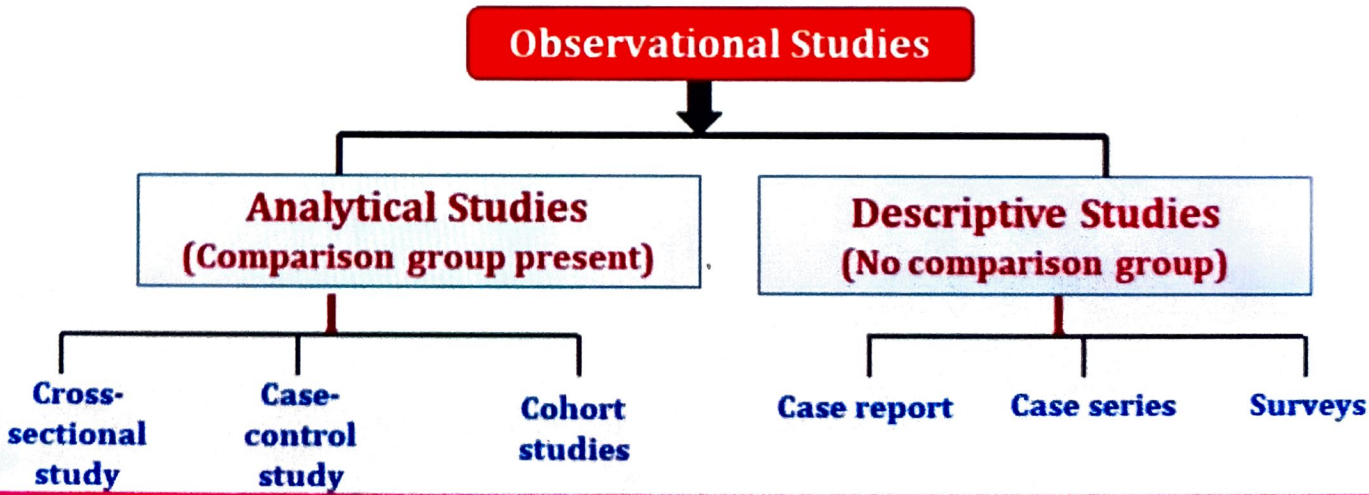
Observational Studies

- Research study designs are classified into two types namely; **observational study and experimental study**. In observational studies, individuals are observed or certain outcomes are measured.
- Researchers observe the effect of a **risk factor, diagnostic test, treatment or other intervention** without trying to change who is or is not exposed to it. Observational studies do not control any variables; thus **the results can only establish relationships and does not allow us to claim association**, and not the causation.



Type of research study design

- Based on the time span, observational study designs are further categorized into analytical study and descriptive study.
- Main objectives of the various study designs include descriptive study to generate hypothesis, analytical study to test hypothesis and experimental study to prove hypothesis.
- There are three major types of observational studies namely; cross-sectional studies, case-control studies, and cohort studies.



❖ Cross-sectional Studies:

- This study involves collecting data about **individuals at a certain point in time. Cross- sectional studies are cheap and easy to conduct**, but they do not give very strong results.
- For example, **a researcher concerned about the effect of working with asbestos might compare the cancer rate** of those who work with asbestos versus those who do not.
- We never be sure that **those working with asbestos and do not report cancer would not eventually develop it.** Researchers tend to use a cross-sectional study to first determine if there might be any link, and then later conduct another study to further investigate.

❖ Case-control Studies:

- In this study researchers **identify people with an existing health problem and a similar group without the problem** and then compare them with respect to an exposure or exposures.
- Researchers in this study attempt to select homogeneous groups, so that on average, **all other characteristics of the individuals will be similar,** with only the characteristic in question differing.
- For example, research on the **link between smoking and lung cancer in the UK.** In the 1950's, **almost 80% of adults in the UK were smokers,** and the connection between smoking and lung cancer was not well established.
- Researchers interviewed about **700 lung cancer patients and tried to determine a possible cause.** This type of study is retrospective, because it asks the individuals to look back and describe their habits.
- The weakness of this study is that **individuals may not only have a smoking accurate memory, but also to respond honestly**. The methods adopted to conduct cross-sectional studies include surveys and database analysis or medication chart reviews.

Experimental studies

- Experimental studies are studies in which the **investigator artificially manipulates study factors or subjects, such as therapeutic regimen, or some other parameter.**
- Experimental studies involve the random assignment of -participants into different groups (e.g. experimental, control) in order to determine the causal effect of a certain condition (independent variable) on a certain outcome (dependent variable).

Experimental study: An example

- On the current topic would be to randomly assign some participants to spend 10 minutes on their face book page (experimental treatment), and the other participants to spend 10 minutes on other websites (control group). Experimental studies are considered the gold standard in social psychological studies because researchers are able to determine causal effects with more confidence than when using any other research method.
- The researcher has no control over the variables in an observational study. An experiment is a method of applying treatments to a group and recording the effects. Remember, a good group experiment will have two basic elements: a control and a treatment.

❖ Principles of experimental design:

There are three important principle of experimental design which given as under:

- **Replication:** to provide an estimate of experimental error; randomization, to ensure that this estimate is statistically valid; and. local control, to reduce experimental error by making the experiment more efficient.
- **Experimental Method:** An experiment is an investigation in which a hypothesis is scientifically tested. In an experiment, an independent variable (the cause) is manipulated and the dependent variable (the effect) is measured; any extraneous variables are controlled. An advantage is that experiments should be objective.

Designing Clinical Trial - Various Phases

- The development of INDs involves performing clinical trials to assess the **safety and efficacy of these drugs in humans.**
- Clinical trials are usually classified into **4 phases** of development as **Phase-I to Phase-IV,** with each potentially lasting for many months to several years.
- Upon successful completion of each phase of trial, necessary approvals are obtained from the appropriate regulatory authority or authorities, for example the European Medicines Agency in the European Union, Food and Drug Administration in the United States of America, Health Canada in Canada, the Ministry of Health, Labour and Welfare in Japan, Drugs Controller General of India in India etc.
- for the progression to the next phase. Satisfactory completion and approval of Phase-I to Phase-III is required for a IND to be approved for marketing.
- Phase-IV studies are conducted after a drug that has been approved by regulatory agency, for the primary purpose of post marketing surveillance. Recently, for speeding up the drug development process and to quickly identify safety issues, Phase 0 studies (human microdosing studies) have been introduced.

➢ **Phase-0:** The official name of a Phase-0 study is an **'Exploratory IND Study'**. Its goal is to quickly establish whether candidate drug will work as desired in humans. This study is based on **in-vivo safety, pharmacology and toxicological preclinical studies**. At this phase studies, a **single sub-therapeutic dose of the IND is administered to a small number of healthy subjects (10 to 15),** over a short duration of 7 days. The dose administered at this phase is too low to result in a therapeutic effect to ensure the **absence of any toxic effects, preliminary pharmacokinetic (PK) and, pharmacodynamic (PD)** data are collected for evaluation to establish suitability.

- **Phase-I:** The Phase-I studies are designed to assess the safety of an IND, to understand its **PK and PD properties**, and to ideally **identify a potential therapeutic dose.** These studies are usually conducted in a small number of healthy **volunteers/subjects (15 to 30).**
- **Phase-II:** These studies are typically conducted to test the IND in a larger group of patients **who have the disease or illness for which the IND is being developed,** to determine whether it is efficacious, at least in the short term. Phase-II studies are larger than those conducted earlier in the drug development, **typically comprising up to 300 patients.**
- **Phase-III:** The Phase-III studies are designed and performed to assess **the efficacy and effectiveness of an IND in a larger cohort of patients**, all of whom have the disease that the treatment is intended to treat. Such studies are typically conducted in several hundred - patients, and are usually conducted at multiple sites in multiple countries. **Phase-III studies compare the new treatment against the current 'gold standard'** treatment for the condition for which the new treatment is being developed.
- **Phase-IV:** **Post marketing surveillance involves monitoring (Pharmacovigilance) for safety concerns once a treatment has been approved by the regulatory authority/authorities.** Such surveillance is aimed to detect any uncommon adverse effects that have not been observed previously or have only been observed occasionally, and to observe the effects of long term administration in a wider population.