UNIT 1 PARAMETRIC AND NON-PARAMETRIC STATISTICS

Structure

1.0 INTRODUCTION

In this unit you will be able to know the various aspects of parametric and non-parametric statistics. A parametric statistical test specifies certain conditions such as the data should be normally distributed etc. The non-parametric statistics does not require the conditions of parametric stats. In fact non-parametric tests are known as distribution free tests.

In this unit we will study the nature of quantitative data and various descriptive statistical measures which are used in the analysis of such data. These include measures of central tendency, variability, relative position and relationships of normal probability curve etc. will be explained.

The computed values of various statistics are used to describe the properties of particular samples. In this unit we shall discuss inferential or sampling statistics, which are useful to a researcher in making generalisations of inferences about the populations from the observations of the characteristics of samples.

For making inferences about various population values (parameters), we generally make use of parametric and non-parametric tests. The concept and assumptions of parametric tests will be explained to you in this section along with the inference regarding the means and correlations of large and small samples, and significance of the difference between the means and correlations in large and small independent samples.

The assumptions and applications of analysis of variance and co-variance for testing the significance of the difference between the means of three or more samples will also be discussed.

In the use of parametric tests for making statistical inferences, we need to take into account certain assumptions about the nature of the population distribution, and also the type of the measurement scale used to quantify the data. In this unit you will learn about another category of tests which do not make stringent assumptions about the nature of the population distribution. This category of test is called distribution free or non-parametric tests. The use and application of several non-parametric tests involving unrelated and related samples will be explained in this unit. These would include chi-square test, median test, Man-Whitney U test, sign test and Wilcoxon-matched pairs signed-ranks test.

1.1 OBJECTIVES

After reading this unit, you will be able to:

- define the terms parametric and non-parametric statistics;
- differentiate between parametric and non-parametric statistics;
- describe the nature and meaning of parametric and non-parametric statistics;
- delineate the assumptions of parametric and non-parametric statistics; and
- list the advantages and disadvantages of parametric and non-parametric statistics.

1.2 DEFINITION OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

Statistics is an Independent branch and its use is highly prevalent in all the fields of knowledge. Many methods and techniques are used in statistics. These have been grouped under parametric and and non-parametric statistics. Statistical tests which are not based on a *normal distribution* of data or on any other assumption are also known as distribution-free tests and the data are generally ranked or grouped. Examples include the *chi-square test* and *Spearman's rank correlation coefficient*.

The first meaning of *non-parametric* covers techniques that do not rely on data belonging to any particular distribution. These include, among others:

1) Distribution free methods: This means that there are no assumptions that the data have been drawn from a normally distributed population. This consists of non-parametric *statistical models*, *inference* and *statistical tests*.

2) Non-parametric statistics: In this the statistics is based on the *ranks* of observations and do not depend on any distribution of the population.

3) No assumption of a structure of a model: In non-parametric statistics, the techniques do not assume that the *structure* of a model is fixed. In this, the

individual variables *are* typically assumed to belong to parametric distributions, and assumptions about the types of connections among variables are also made. These techniques include, among others:

a) Non-parametric regression

b) Non-parametric hierarchical Bayesian models.

In non-parametric regression, the structure of the relationship is treated non-parametrically.

In regard to the Bayesian models, these are based on the *Dirichlet process*, which allows the number of *latent variables* to grow as necessary to fit the data. In this the individual variables however follow parametric distributions and even the process controlling the rate of growth of latent variables follows a parametric distribution.

4) The assumptions of a *classical* or *standard tests* are not applied to non-parametric tests.

Parametric tests

Parametric tests normally involve data expressed in absolute numbers or values rather than ranks; an example is the *Student's t-test*.

The parametric statistical test operates under certain conditions. Since these conditions are not ordinarily tested, they are assumed to hold valid. The meaningfulness of the results of a parametric test depends on the validity of the assumption. Proper interpretation of parametric test based on normal distribution also assumes that the scene being analysed results from measurement in at least an interval scale.

Let us try to understand the term population. Population refers to the entire group of people which a researcher intends to understand in regard to a phenomenon. The study is generally conducted on a sample of the said population and the obtained results are then applied to the larger population from which the sample was selected.

Tests like t, z, and F are called parametrical statistical tests.

T-tests: A T-test is used to determine if the scores of two groups differ on a single variable.

A t-test is designed to test for the differences in mean scores. For instance, you could use t-test to determine whether writing ability differs among students in two classrooms.

It may be mentioned here that the parametric tests, namely, t-test and F-test, are considered to be quite robust and are appropriate even when some assumptions are not met.

Parametric tests are useful as these tests are most powerful for testing the significance or trustworthiness of the computed sample statistics. However, their use is based upon certain assumptions. These assumptions are based on the nature of the population distribution and on the way the type of scale is used to quantify the data measures.

Let us try to understand what is a scale and its types. There are four types of scales used in measurement viz., nominal scale, ordinal scale, interval scale, and ratio scale.

1) Nominal scale deals with nominal data or classified data such as for example the population divided into males and females. There is no ordering of the data in that it has no meaning when we say male > female. These data are also given

arbitrary labels such as m / f and 1 //0 . These are also called as categorical scale , that is these are scales with values that are in terms of categories (i.e. they are names rather than numbers).

2) Ordinal scale deals with interval data. These are in certain order but the differences between values are not important. For example, degree of satisfaction ranging in a 5 point scale of 1 to 5, with 1 indicating least satisfaction and 5 indicating high satisfaction.

3) Interval scale deals with ordered data with interval. This is a constant scale but has no natural zero. Differences do make sense . Example of this kind of data includes for instance temperature in Centigrade or Fahrenheit. The dates in a calendar. Interval scale possesses two out of three important requirements of a good measurement scale, that is, magnitude and equal intervals but lacks the real or absolute zero point.

4) Ratio scale deals with ordered, constant scale with a natural zero. Example of this type of data include for instance, height, weight, age, length etc.

The sample with small number of items are treated with non-parametric statistics because of the absence of normal distribution, e.g. if our sample size is 30 or less; ($N \leq 30$). It can be used even for nominal data along with the ordinal data.

A non-parametric statistical test is based on model that specifies only very general conditions and none regarding the specific form of the distribution from which the sample was drawn.

Certain assumptions are associated with most non-parametric statistical tests, namely that the observation are independent, perhaps that variable under study had underlying continuity, but these assumptions are fewer and weaker than those associated with parametric tests.

More over as we shall see, non-parametric procedures often test different hypotheses about population than do parametric procedures.

Finally, unlike parametric tests, there are non-parametric procedures that may be applied appropriately to data measured in an ordinal scale, or in a nominal scale or categorical scale.

Non-parametric statistics deals with small sample sizes.

Non-parametric statistics are assumption free meaning these are not bound by any assumptions.

Non-parametric statistics are user friendly compared with parametric statistics and economical in time.

We have learnt that parametric tests are generally quite robust and are useful even when some of their mathematical assumptions are violated. However, these tests are used only with the data based upon ratio or interval measurements.

In case of counted or ranked data, we make use of non-parametric tests. It is argued that non-parametric tests have greater merit because their validity is not based upon assumptions about the nature of the population distribution, assumptions that are so frequently ignored or violated by researchers using parametric tests. It may be noted that non-parametric tests are less precise and have less power than the parametric tests.

1.3 ASSUMPTIONS OF PARAMETRIC AND NON-PARAMETRIC STATISTICS

1.3.1 Assumptions of Parametric Statistics

Parametric tests like, ‘t and f’ tests may be used for analysing the data which satisfy the following conditions :

The population from which the sample have been drawn should be normally distributed.

Normal Distributions refer to Frequency distribution following a normal curve, which is infinite at both the ends.

The variables involved must have been measured interval or ratio scale.

Variable and its types: characteristic that can have different values.

Types of Variables

Dependent Variable: Variable considered to be an effect; usually a measured variable. Independent Variable: Variable considered being a cause.

The observation must be independent. The inclusion or exclusion of any case in the sample should not unduly affect the results of study.

These populations must have the same variance or, in special cases, must have a known ratio of variance. This we call homosedasticity.

The samples have equal or nearly equal variances. This condition is known as equality or homogeneity of variances and is particularly important to determine when the samples are small.

The observations are independent. The selection of one case in the sample is not dependent upon the selection of any other case.

1.3.2 Assumptions of Non-parametric Statistics

We face many situations where we can not meet the assumptions and conditions and thus cannot use parametric statistical procedures. In such situation we are bound to apply non-parametric statistics.

If our sample is in the form of nominal or ordinal scale and the distribution of sample is not normally distributed, and also the sample size is very small, it is always advisable to make use of the non-parametric tests for comparing samples and to make inferences or test the significance or trust worthiness of the computed statistics. In other words, the use of non-parametric tests is recommended in the following situations:

Where sample size is quite small. If the size of the sample is as small as N=5 or N=6, the only alternative is to make use of non-parametric tests.

When assumption like normality of the distribution of scores in the population are doubtful, we use non-parametric tests.

When the measurement of data is available either in the form of ordinal or nominal scales or when the data can be expressed in the form of ranks or in the shape of + signs or – signs and classification like “good-bad”, etc., we use non-parametric statistics.

The nature of the population from which samples are drawn is not known to be normal.

The variables are expressed in nominal form.

The data are measures which are ranked or expressed in numerical scores which have the strength of ranks.

1.4 ADVANTAGES OF NON-PARAMETRIC STATISTICS

If the sample size is very small, there may be no alternative except to use a non-parametric statistical test.

Non-parametric tests typically make fewer assumptions about the data and may be relevant to a particular situation.

The hypothesis tested by the non-parametric test may be more appropriate for research investigation.

Non-parametric statistical tests are available to analyse data which are inherently in ranks as well as data whose seemingly numerical scores have the strength of ranks.

For example, in studying a variable such as anxiety, we may be able to state that subject A is more anxious than subject B without knowing at all exactly how much more anxious A is. Thus if the data are inherently in ranks, or even if they can be categorised only as plus or minus (more or less, better or worse), they can be treated by non-parametric methods.

Non-parametric methods are available to treat data which are simply classificatory and categorical, i.e., are measured in nominal scale.

Samples made up of observations from several different populations at times cannot be handled by Parametric tests.

Non-parametric statistical tests typically are much easier to learn and to apply than are parametric tests. In addition, their interpretation often is more direct than the interpretation of parametric tests.

1.5 DISADVANTAGES OF NON-PARAMETRIC STATISTICAL TESTS

If all the assumptions of a parametric statistical model are in fact met in the data and the research hypothesis could be tested with a parametric test, then non-parametric statistical tests are wasteful.

The degree of wastefulness is expressed by the power-efficiency of the non-parametric test. It will be remembered that, if a non-parametric statistical test has power-efficiency of say, 90 percent, this means that when all conditions of parametric statistical test are satisfied the appropriate parametric test would be just as effective with a sample which is 10 percent smaller than that used in non-parametric analysis.

Another objection to non-parametric statistical test has to do with convenience. Tables necessary to implement non-parametric tests are scattered widely and appear in different formats (The same is true of many parametric tests too).

1.6 PARAMETRIC STATISTICAL TESTS FOR DIFFERENT SAMPLES

Suppose we wish to measure teaching aptitude of M.A. Psychology Students(LARGE SAMPLE) by using a verbal aptitude teaching test.

It is not possible and convenient to measure the teaching aptitude of all the enrolled M.A. Psychology Students trainees and hence we must usually be satisfied with a sample drawn from this population.

However, this sample should be as large and as randomly drawn as possible so as to represent adequately all the M.A. Psychology Students of IGNOU.

If we select a large number of random samples of 100 trainees each from the population of all trainees, the mean values of teaching aptitude scores for all samples would not be identical.

A few would be relatively high, a few relatively low, but most of them would tend to cluster around the population mean.

The sample means due to 'sampling error' will not vary from sample to sample but will also usually deviate from the population mean. Each of these sample means can be treated as a single observation and these means can be put in a frequency distribution which is known as sampling distribution of the means.

An important principle, known as the 'Central Limit Theorem', describes the characteristics of sample means. According to this theorem, if a large number of equal-sized samples, greater than 30 in size, are selected at random from an infinite population:

The means of the samples will be normally distributed.

The average value of the sample means will be the same as the mean of the population.

The distribution of sample means will have its own standard deviation.

This standard deviation is known as the 'standard error of the mean' which is denoted as SE_M or $ó_M$.

It gives us a clue as to how far such sample means may be expected to deviate from the population mean.

The standard error of a mean tells us how large the errors are in any particular sampling situation.

The formula for the standard error of the mean in a large sample is:

$$SE_M \text{ or } ó_M = ó \,/\, \sqrt{N}$$

Where

ó = the standard deviation of the population

N = the size of the sample

In case of **small samples**, the sampling distribution of means is not normal. It was in about 1815 when William Seely Gosset developed the concept of small sample size. He found that the distribution curves of small sample means were some what different from the normal curve. This distribution was named as t-distribution. When the size of the sample is small, the t-distribution lies under the normal curve.

1.7 PARAMETRIC STATISTICAL MEASURES FOR CALCULATING DIFFERENCE BETWEEN MEANS

In some research situations we require the use of a statistical technique to determine whether a true difference exists between the population parameters of two samples. The parameters may be means, standard deviations, correlations etc. For example, suppose we wish to determine whether the population of male M.A. Psychology Students enrolled with IGNOU differs from their female counterparts in their attitude towards teaching… In this case we would first draw samples of male and female M.A. Psychology Students. Next, we would administer an attitude scale measuring attitude towards teaching on the selected samples, compute the means of the two samples, and find the difference between them. Let the mean of the male sample be 55 and that of the females 59. Then it has to be ascertained if the difference of 4 between the sample means is large enough to be taken as real and not due only to sampling error or chance.

In order to test the significance of the obtained difference of 4, we need to first find out the standard error of the difference of the two means because it is reasonable to expect that the difference between two means will be subject to sampling errors. Then from the difference between the sample means and its standard error we can determine whether a difference probably exists between the population means.

In the following sections we will discuss the procedure of testing the significance of the difference between the means and correlations of the samples.

1.7.1 Significance of the Difference between the Means of Two Independent Large and Small Samples

Means are said to be independent or uncorrelated when computed from samples drawn at random from totally different and unrelated groups.

Large Samples

You have leant that the frequency distribution of large sample means, drawn from the same population, fall into a normal distribution around the population mean (M_{pop}) as their measure of central tendency. It is reasonable to expect that the frequency distribution of the difference between the means computed from the samples drawn from two different populations will also tend to be normal with a mean of zero and standard deviation which is called the standard error of the difference of means.

The standard error is denoted by ó$_{dm}$ which is estimated from the standard errors of the two sample means, ó$_{m1}$ and ó$_{m2}$. The formula is:

ó$_{dM}$ = (ó$_{M1}$2 + ó$_{M1}$2)$^{\text{Under root}}$

in which

ó$_{M1}$ = SE of the mean of the first sample

ó$_{Mw}$ = SE of the mean of the second sample

N_1 = Number of cases in first sample

N_2 = Number of cases in second sample

1.7.2 Significance of the Differences between the Means of Two Dependent Samples

Means are said to be dependent or correlated when obtained from the scores of the same test administered to the same sample upon two occasions, or when the same test is administered to equivalent samples in which the members of the group have been matched person for person, by one or more attributes.

$T = M_1 + M_2 \,/\, óM^2_1 + - 2\ r_{12}\ óM_1\ óM_2$

in which

M_1 and M_2 = Means of the scores of the initial and final testing.

$óM_1$ = Standard error of the initial test mean.

$óM_2$ = Standard error of the final test mean.

r_{12} = Correlation between the scores on initial and final testing.

1.7.3 Significance of the Difference between the Means of Three or More Samples

We compute CR and t-values to determine whether there is any significant difference between the means of two random samples. Suppose we have N(N>2) random samples and we want to determine whether there are any significant differences among their means. For this we have to compute F value that is Analysis of Variance.

Analysis of variance has the following basic assumptions underlying it which should be fulfilled in the use of this technique.

The population distribution should be normal. This assumption, however, is not especially important.

Eden and Yates showed that even with a population departing considerably from normality, the effectiveness of the normal distribution still held.

All the groups of certain criterion or of the combination of more than one criterion should be randomly chosen from the sub-population having the same criterion or having the same combination of more than one criterion.

For instance, if we wish to select two groups in a population of M.A. Psychology Student trainees enrolled with IGNOU, one of males and the other of females, we must choose randomly from the respective sub populations. The assumption of randomness is the key stone of the analysis of variance technique. There is no substitute for randomization.

The sub-groups under investigation should have the same variability. This assumption is tested by applying F_{max} test.

F_{max} =Largest Variance / Smallest Variance

In analysis of variance, we have usually three or more groups i.e. there will be three or more variances.

Unless the computed value of F_{max} equals or exceeds the appropriate F critical value at .05 level in the Table N of the Appendix, (Statistics book) it is assumed that the variances are homogeneous and the difference is not significant.

1.8 PARAMETRIC STATISTICS MEASURES RELATED TO PEARSON'S 'r'

The mathematical basis for standard error of a Pearson's co-efficient of correlation 'r' is rather complicated because of the difficulty in its nature of sampling distribution.

The sampling distribution of r is not normal except when population r is near zero and size of the sample is large (N=30 or greater).

When r is high (0.80 or more) and N is small, the sampling distribution of r is skewed. It is also true when r is low (0.20 or less).

In view of this, a sound method for making the inference regarding Pearson's r, especially when its magnitude is very high or very low, is to convert r into Fisher's Z coefficient using conversion table provided in the Appendix (Statistics book) and find the standard error (SE) of Z.

The sampling distribution of Z co-efficient is normal regardless of the size of sample N and the size of the population r. Furthermore, the SE of Z depends only upon the size of sample N.

The formula for standard error of Z ($\acute{o}_z$) is:

$$SE_z = 1/\sqrt{N-3}$$

The method of determining the standard error of the difference between Pearson's co-efficient of correlation of two samples is first to convert the r's into Fisher's Z co-efficient and then to determine the significance of the difference between the two Z's.

When we have two correlations between the same two variables, X and Y, computed from two totally different and unmatched samples, the standard error of a difference between two corresponding Z's is computed by the formula:

$$SE_{dz} = \acute{o}_{z1-z2} = \sqrt{(1/N_{1-}\ 3 + 1/\ N_{2-}\ 3)}$$

in which

N_1 and N_2 = sizes of the two samples

The significance of the difference between the two Z's is tested with the following formula:

$$CR = Z_1 - Z_2 / SE_{DZ}$$

1.8.1 Non-parametric Tests Used for Inference

The most frequently used non-parametric tests for drawing statistical inferences in case of unrelated or independent samples are:

1) Chi square test;

2) Median test; and

3) Mann-Whitney 'U' test.

The use and application of these tests are discussed below:

The Chi Square (X^2) Test

The chi square test is applied only to discrete data. The data that are counted rather than measured. It is a test of independence and is used to estimate the likelihood that some factor other than chance accounts for the observed relationship.

The Chi square (X^2) is not a measure of the degree of relationship between the variables under study.

The Chi square test merely evaluates the probability that the observed relationship results from chance. The basic assumption, as in case of other statistical significance, is that the sample observations have been randomly selected.

The formula for chi-square (X^2) is:

$$(X^2) = \sum[(fo-fe)2 / fe]$$

In which

Fo = frequency of occurrence of observed or experimentally determined facts.

Fe = expected frequency of occurrence.

The Median Test

The median test is used for testing whether two independent samples differ in central tendencies. It gives information as to whether it is likely that two independent samples have been drawn from populations with the same median. It is particularly useful when even the measurements for the two samples are expressed in an ordinal scale. In using the median test, we first calculate the combined median for all measures (scores) in both samples. Then both sets of scores at the combined median are dichotomized and the data are set in a 2 x 2 table with two rows one containing below median and the other row containing above median. On the column side we have two columns, one containing the sample 1 and the other column containing sample 2.

The Mann-Whitney U Test

The Mann-Whitney U test is more useful than the Median test. It is one of the most useful alternative to the parametric t test when the parametric assumptions cannot be met and when the measurements are expressed in ordinal scale values.

1.9 SOME NON-PARAMETRIC TESTS FOR RELATED SAMPLES

Various tests are used in drawing statistical inferences in case of related samples. In this section we shall confine our discussion to the use of Sign Test and Wilcoxon Matched-Pairs Signed-Ranks Test Only

The Sign Test

The sign test is the simplest test of significance in the category of non-parametric tests. It makes use of plus and minus signs rather than quantitative measures as its data. It is particularly useful in situations in which quantitative measurement is impossible or inconvenient, but on the basis of superior or inferior performance it is possible to rank with respect to each other, the two members of each pair.

The sign test is used either in the case of single sample from which observations are obtained under two experimental conditions. The researcher wants to establish that the two conditions are different.

The use of this test does not make any assumption about the form of the distribution of differences. The only assumption underlying this test is that the variable under investigation has a continuous distribution.

The Wilcoxon Matched Pairs Signed Ranks Test

The Wilcoxon matched pairs signed ranks test is more powerful than the sign test because it tests not only direction but also the magnitude of differences within pairs of matched groups.

This test, like the sign test, deals with dependent groups made up of matched pairs of individuals and is not applicable to independent groups. The null hypothesis would assume that the direction and magnitude of pair difference would be about the same.

1.10 LET US SUM UP

Parametric and non-parametric tests are important for students especially researchers working in any field. Parametric tests include all methods of statistics when the sample size is large whereas in non-parametric test the sample size is small. There are some advantages and disadvantages of both the tests. In this unit we discussed the statistical inference based on parametric tests. It included the assumptions on which the use of parametric tests are based; inferences regarding means of large and small samples; significance of the difference between the means of two large and small independent samples; significance of the difference between means of the two dependent samples; significance of the difference between means of three or more samples; significance of Pearson's coefficients of correlation; and significance of the difference between Pearson's coefficients of correlation of two independent samples. F test is used for testing the significance between the means of three or more samples. It involves the use of analysis of variance or analysis of co-variance. For testing the significance of Pearson's r, we make use of Fisher's Z transformation or t-test.

1.11 UNIT END QUESTIONS

1) Define parametric statistics.
2) Discuss non-parametric statistics?
3) Write various assumptions of parametric statistics?
4) What are the advantages of non-parametric statistics?
5) Differentiate between parametric and non-parametric statistics?
6) List the assumptions on which the use of Parametric Tests is base.
7) Describe the characteristics of Central Limit Theorem.
8) Define the standard error of mean.

1.12 GLOSSARY

Statistics : Measurement which are associated with sample

Parameters : Measurements which are associated with population

Assumptions	:	Prerequisite conditions
Population	:	Larger group of people to which inferences are made.
Sample	:	Small proportion of the population which we assert representing population.
Normal Curve	:	Bell shaped frequency distribution that is symmetrical and unimodel.
Distribution free tests	:	Hypothesis – testing procedure making non assumptions about population parameters.
Categorical Scale	:	Variable with values that are categories that is, they are name rather than numbers.
Test	:	Test is a tool to measure observable behaviour
Homosedasity	:	Populations must have some variance or in special cases must have a known ratio of variance.

1.13 SUGGESTED READINGS

Asthana H.S, and Bhushan. B. (2007) *Statistics for Social Sciences* (with SPSS Applications).

B.L. Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Guilford, J.P. (1965); *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill Book Company.

Siegel, S. (1956): *Non-parametric Statistics for Behavioural Sciences*. Tokyo: McGraw Hill Hoga Kunsa Ltd.

Sidney Siegel, & N. John Castetellan, Jr. (1958) *Non-parametric Statistics for the Behavioural Science*. McGraw Hill Books company, New Delhi

UNIT 2 DESCRIPTIVE AND INFERENTIAL STATISTICS

Structure

2.0 INTRODUCTION

In this unit we will be dealing with descriptive and inferential statistics. First we start with defining descriptive statistics and indicate how to organise the data , classify, tabulate etc. This unit also presents as to how the data should be presented graphically. Once the data is collected the same has to be made meaningful which can be done through averaging the data or working out the variances in the data etc. Then we deal with the advantages and disadvantages of descriptive statistics. This is followed by defining what is inferential statistics and delineating its meaning. In this unit the student will also gain knowledge regarding point and interval estimation so as to validate the results. We also learn in this unit about hypothesis testing, how it is done and the methods thereof. We also deal with different types of errors in hypothesis testing including sampling error etc.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- define the nature and meaning of descriptive statistics;
- describe the methods of organising and condensing raw data;
- explain concept and meaning of different measures of central tendency;
- analyse the meaning of different measures of dispersion;
- define inferential statistics;
- explain the concept of estimation;
- distinguish between point estimation and interval estimation; and
- explain the different concepts involved in hypothesis testing.

2.2 MEANING OF DESCRIPTIVE STATISTICS

The word statistics has different meaning to different persons. For some, it is a one-number description of a set of data. Some consider statistics in terms of numbers used as measurements or counts. Mathematicians use statistics to describe data in one word. It is a summary of an event for them. Number , n, is the statistic describing how big the set of numbers is, how many pieces of data are in the set.

Also, knowledge of statistics is applicable in day to day life in different ways. Statistics is used by people to take decision about the problems on the basis of different types of information available to them. However, in behavioural sciences the word 'statistics' means something different, that is its prime function is to draw statistical inference about population on the basis of available quantitative and qualitative information.

The word statistics can be defined in two different ways. In singular sense 'Statistics' refers to what is called statistical methods. When 'Statistics' is used in plural sense it refers to 'data'.

In this unit we will use the term 'statistics' in singular sense. In this context, it is described as a branch of science which deals with the collection of data, their classification, analysis and interpretations of statistical data.

The science of statistics may be broadly studied under two headings:

i) Descriptive Statistics, and (ii) Inferential Statistics

i) Descriptive Statistics: Most of the observations in this universe are subject to variability, especially observations related to human behaviour. It is a well known fact that attitude, intelligence and personality differ from individual to individual. In order to make a sensible definition of the group or to identify the group with reference to their observations/ scores, it is necessary to express them in a precise manner. For this purpose observations need to be expressed as a single estimate which summarises the observations.

Descriptive statistics is a branch of statistics, which deals with descriptions of obtained data. On the basis of these descriptions a particular group of population is defined for corresponding characteristics. The descriptive statistics include classification, tabulation, diagrammatic and graphical presentation of data, measures of central tendency and variability. These measures enable the researchers to know about the tendency of data or the scores, which further enhance the ease in description of the phenomena. Such single estimate of the series of data which summarises the distribution are known as parameters of the distribution. These parameters define the distribution completely.

Bascially descriptive statistics involves two operations:

(i) Organisation of data, and (ii) Summarination of data

2.3 ORGANISATION OF DATA

There are four major statistical techniques for organising the data. These are:

i) Classification

ii) Tabulation

iii) Graphical Presentation, and

iv) Diagrammatical Presentation

2.3.1 Classification

The arrangement of data in groups according to similarities is known as classification. A classification is a summary of the frequency of individual scores or ranges of scores for a variable. In the simplest form of a distribution, we will have such value of variable as well as the number of persons who have had each value.

Once data are collected, it should be arranged in a format from which they would be able to draw some conclusions. Thus by classifying data, the investigators move a step ahead in regard to making a decision.

A much clear picture of the information of score emerges when the raw data are organised as a frequency distribution. Frequency distribution shows the number of cases following within a given class interval or range of scores. A frequency distribution is a table that shows each score as obtained by a group of individuals and how frequently each score occurred.

2.3.1.1 Frequency Distribution can be with Ungrouped Data and Grouped Data

i) An ungrouped frequency distribution may be constructed by listing all score values either from highest to lowest or lowest to highest and placing a tally mark (/) besides each scores every times it occurs. The frequency of occurrence of each score is denoted by ‘f’.

ii) Grouped frequency distribution: If there is a wide range of score value in the data, then it is difficult to get a clear picture of such series of data. In this case grouped frequency distribution should be constructed to have a clear picture of the data. A group frequency distribution is a table that organises data into classes.

It shows the number of observations from the data set that fall into each of the class.

Construction of frequency distribution

To prepare a frequency distribution it is essential to determine the following:

1) The range of the given data =, the difference between the highest and lowest scores.

2) The number of class intervals = There is no hard and fast rules regarding the number of classes into which data should be grouped. If there are very few scores it is useless to have a large number of class-intervals. Ordinarily, the number of classes should be between 5 to 30.

3) Limits of each class interval = Another factor used in determining the number of classes is the size/ width or range of the class which is known as 'class interval' and is denoted by 'i'.

Class interval should be of uniform width resulting in the same-size classes of frequency distribution. The width of the class should be a whole number and conveniently divisible by 2, 3, 5, 10, or 20.

There are three methods for describing the class limits for distribution:

(i) Exclusive method, (ii) Inclusive method and (iii) True or actual class method.

i) **Exclusive method**

In this method of class formation, the classes are so formed that the upper limit of one class become the lower limit of the next class. In this classification, it is presumed that score equal to the upper limit of the class is exclusive, i.e., a score of 40 will be included in the class of 40 to 50 and not in a class of 30 to 40 (30-40, 40-50, 50-60)

ii) **Inclusive method**

In this method the classes are so formed that the upper limit of one class does not become the lower limit of the next class. This classification includes scores, which are equal to the upper limit of the class. Inclusive method is preferred when measurements are given in whole numbers. (30-39, 40-49, 50-59)

iii) **True or Actual class method**

Mathematically, a score is an internal when it extends from 0.5 units below to 0.5 units above the face value of the score on a continuum. These class limits are known as true or actual class limits. (29.5 to 39.5, 39.5 to 49.5) etc.

2.3.1.2 Types of Frequency Distribution

There are various ways to arrange frequencies of a data array based on the requirement of the statistical analysis or the study. A couple of them are discussed below:

i) Relative frequency distribution: A relative frequency distribution is a distribution that indicates the proportion of the total number of cases observed at each score value or internal of score values.

ii) Cumulative frequency distribution: Sometimes investigator may be interested to know the number of observations less than a particular value. This is possible by computing the cumulative frequency. A cumulative frequency corresponding to a class-interval is the sum of frequencies for that class and of all classes prior to that class.

iii) Cumulative relative frequency distribution: A cumulative relative frequency distribution is one in which the entry of any score of class interval expresses that score's cumulative frequency as a proportion of the total number of cases.

Self Assessment Questions

1) Complete the following statements

   i) Statistics in plural means ....

   ii) Statistics in singular means ....

   iii) Data collection is .... step in statistics.

   iv) The last step in statistics is ....

2) Define following concepts

   1) Descriptive statistics

   ....

   ....

   2) Inferential statistics

   ....

   ....

   3) Exclusive method of classification

   ....

   ....

   4) Actual method of classification

   ....

   ....

   5) Frequency distribution

   ....

   ....

2.3.2 Tabulation

Frequency distribution can be either in the form of a table or it can be in the form of graph. Tabulation is the process of presenting the classified data in the form of a table. A tabular presentation of data becomes more intelligible and fit for further statistical analysis. A table is a systematic arrangement of classified data in row and columns with appropriate headings and sub-headings. The main components of a table are:

i) Table number: When there is more than one table in a particular analysis a table should be marked with a number for their reference and identification. The number should be written in the center at the top of the table.

ii) Title of the table: Every table should have an appropriate title, which describes the content of the table. The title should be clear, brief, and self-explanatory. Title of the table should be placed either centrally on the top of the table or just below or after the table number.

iii) Caption: Captions are brief and self-explanatory headings for columns. Captions may involve headings and sub-headings. The captions should be placed in the middle of the columns. For example, we can divide students of a class into males and females, rural and urban, high SES and Low SES etc.

iv) Stub: Stubs stand for brief and self-explanatory headings for rows.

v) Body of the table: This is the real table and contains numerical information or data in different cells. This arrangement of data remains according to the description of captions and stubs.

vi) Head note: This is written at the extreme right hand below the title and explains the unit of the measurements used in the body of the tables.

vii) Footnote: This is a qualifying statement which is to be written below the table explaining certain points related to the data which have not been covered in title, caption, and stubs.

viii) Source of data: The source from which data have been taken is to be mentioned at the end of the table.

TITLE

<table>
<tr><th>Stub Head</th><th colspan="4">Caption</th></tr>
<tr><td rowspan="2">Stub Entries</td><td colspan="2">Column Head I</td><td colspan="2">Column Head II</td></tr>
<tr><td>Sub Head</td><td>Sub Head</td><td>Sub Head</td><td>Sub Head</td></tr>
<tr><td></td><td>MAIN BODY</td><td>OF</td><td colspan="2">THE TABLE</td></tr>
<tr><td>Total</td><td></td><td></td><td colspan="2"></td></tr>
</table>

Footnote(s):

Source :

2.3.3 Graphical Presentation of Data

The purpose of preparing a frequency distribution is to provide a systematic way of " looking at" and understanding data. To extend this understanding, the information contained in a frequency distribution often is displayed in graphic and/or diagrammatic forms. In graphical presentation of frequency distribution, frequencies are plotted on a pictorial platform formed of horizontal and vertical lines known as graph.

A graph is created on two mutually perpendicular lines called the X and Y–axes on which appropriate scales are indicated. The horizontal line is called the abscissa and vertical the ordinate. Like different kinds of frequency distributions there are many kinds of graph too, which enhance the scientific understanding of the reader. The commonly used graphs are Histogram, Frequency polygon, Frequency curve, Cumulative frequency curve. Here we will discuss some of the important types of graphical patterns used in statistics.

i) Histogram: It is one of the most popular methods for presenting continuous frequency distribution in a form of graph. In this type of distribution the upper limit of a class is the lower limit of the following class. The histogram consists of series of rectangles, with its width equal to the class interval of the variable on horizontal axis and the corresponding frequency on the vertical axis as its heights.

ii) Frequency polygon: Prepare an abscissa originating from 'O' and ending to 'X'. Again construct the ordinate starting from 'O' and ending at 'Y'. Now label the class-intervals on abscissa stating the exact limits or midpoints of the class-intervals. You can also add one extra limit keeping zero frequency on both side of the class-interval range.

The size of measurement of small squares on graph paper depends upon the number of classes to be plotted. Next step is to plot the frequencies on ordinate using the most comfortable measurement of small squares depending on the range of whole distribution.

To plot a frequency polygon you have to mark each frequency against its concerned class on the height of its respective ordinate. After putting all frequency marks a draw a line joining the points. This is the polygon.

iii) Frequency curve: A frequency curve is a smooth free hand curve drawn through frequency polygon. The objective of smoothing of the frequency polygon is to eliminate as far as possible the random or erratic fluctuations that are present in the data.

2.3.3.1 Cumulative Frequency Curve or Ogive

The graph of a cumulative frequency distribution is known as cumulative frequency curve or ogive. Since there are two types of cumulative frequency distribution e.g., " less than" and " more than" cumulative frequencies. We can have two types of ogives.

i) 'Less than' Ogive: In 'less than' ogive , the less than cumulative frequencies are plotted against the upper class boundaries of the respective classes. It is an increasing curve having slopes upwards from left to right.

ii) 'More than' Ogive : In more than ogive , the more than cumulative frequencies are plotted against the lower class boundaries of the respective classes. It is decreasing curve and slopes downwards from left to right.

2.3.4 Diagrammatic Presentation of Data

A diagram is a visual form for the presentation of statistical data. They present the data in simple , readily comprehensible form. Diagrammatic presentation is used only for presentation of the data in visual form, whereas graphic presentation of the data can be used for further analysis. There are different forms of diagram e.g., Bar diagram, Sub-divided bar diagram, Multiple bar diagram, Pie diagram and Pictogram.

i) Bar diagram: Bar diagram is most useful for categorical data. A bar is defined as a thick line. Bar diagram is drawn from the frequency distribution table representing the variable on the horizontal axis and the frequency on the vertical axis. The height of each bar will be corresponding to the frequency or value of the variable.

ii) Sub- divided bar diagram: Study of sub classification of a phenomenon can be done by using sub-divided bar diagram. Corresponding to each sub-category of the data the bar is divided and shaded. There will be as many shades as there will sub portion in a group of data. The portion of the bar occupied by each sub-class reflects its proportion in the total.

iii) Multiple Bar diagram: This diagram is used when comparisons are to be shown between two or more sets of interrelated phenomena or variables. A set of bars for person, place or related phenomena are drawn side by side without any gap. To distinguish between the different bars in a set , different colours , shades are used.

iv) Pie diagram: It is also known as angular diagram. A pie chart or diagram is a circle divided into component sectors corresponding to the frequencies of the variables in the distribution. Each sector will be proportional to the frequency of the variable in the group. A circle represents 360^{O}. So 360^{O} angles is divided in proportion to percentages. The degrees represented by the various component parts of given magnitude can be obtained by using this formula.

After the calculation of the angles for each component, segments are drawn in the circle in succession, corresponding to the angles at the center for each segment. Different segments are shaded with different colours, shades or numbers.

Self Assessment Questions

1) In ' less than' cumulative frequency distribution, which class limit is omitted

   i) upper

   ii) lower

   iii) last

   iv) none of these

2) Differentiate between following components of a statistical table that is "Caption" and "Stub head" "Head note" and "Foot note".

   ....

   ....

3) Explain the following terms

   i) Histogram,

   ....

   ii) Bar diagram,

   ....

   iii) Frequency polygon, and

   ....

   iv) Pie diagram.

   ....

2.4 SUMMARISATION OF DATA

In the previous section we have discussed about tabulation of the data and its representation in the form of graphical presentation. In research, comparison between two or more series of the same type is needed to find out the trends of variables. For such comparison, tabulation of data is not sufficient and it is further required to investigate the characteristics of data. The frequency distribution of obtained data may differ in two ways, first in measures of central tendency and second, in the extent to which scores are spread over the central value. Both types of differences are the components of summary statistics.

2.4.1 Measures of Central Tendency

It is the middle point of a distribution. Tabulated data provides the data in a systematic order and enhances their understanding. Generally, in any distribution values of the variables tend to cluster around a central value of the distribution. This tendency of the distribution is known as central tendency and measures devised to consider this tendency is know as measures of central tendency. A measure of central tendency is useful if it represents accurately the distribution of scores on which it is based. A good measure of central tendency must possess the following characteristics:

It should be clearly defined- The definition of a measure of central tendency should be clear and unambiguous so that it leads to one and only one information.

It should be readily comprehensible and easy to compute.

It should be based on all observations- A good measure of central tendency should be based on all the values of the distribution of scores.

It should be amenable for further mathematical treatment.

It should be least affected by the fluctuation of sampling.

In Statistics there are three most commonly used measures of central tendency. These are:

1) Arithmetic Mean 2) Median, and 3) Mode

1) Arithmetic Mean: The arithmetic mean is most popular and widely used measure of central tendency. Whenever we refer to the average of data, it means we are talking about its arithmetic mean. This is obtained by dividing the sum of the values of the variable by the number of values. It is also a useful measure for further statistics and comparisons among different data sets. One of the major limitations of arithmetic mean is that it cannot be computed for open-ended class-intervals.

2) Median: Median is the middle most value in a data distribution. It divides the distribution into two equal parts so that exactly one half of the observations is below and one half is above that point. Since median clearly denotes the position of an observation in an array, it is also called a position average. Thus more technically, median of an array of numbers arranged in order of their magnitude is either the middle value or the arithmetic mean of the two middle values. It is not affected by extreme values in the distribution.

3) Mode: Mode is the value in a distribution that corresponds to the maximum concentration of frequencies. It may be regarded as the most typical of a series value. In more simple words, mode is the point in the distribution comprising maximum frequencies therein.

2.4.2 Measures of Dispersion

In the previous section we have discussed about measures of central tendency. By knowing only the mean, median or mode, it is not possible to have a complete picture of a set of data. Average does not tell us about how the score or measurements are arranged in relation to the center. It is possible that two sets of data with equal mean or median may differ in terms of their variability. Therefore, it is essential to know how far these observations are scattered from each other or from the mean. Measures of these variations are known as the 'measures of dispersion'. The most commonly used measures of dispersion are range, average deviation, quartile deviation, variance and standard deviation.

i) Range

Range is one of the simplest measures of dispersion. It is designated by 'R'. The range is defined as the difference between the largest score and the smallest score in the distribution. It gives the two extreme values of the variable. A large value of range indicates greater dispersion while a smaller value indicates lesser dispersion among the scores. Range can be a good measure if the distribution is not much skewed.

ii) Average deviation

Average deviation refers to the arithmetic mean of the differences between each score and the mean. It is always better to find the deviation of the individual observations with reference to a certain value in the series of observation and then take an average of these deviations. This deviation is usually measured from mean or median. Mean, however, is more commonly used for this measurement.

Merits: It is less affected by extreme values as compared to standard deviation. It provides better measure for comparison about the formation of different distributions.

iii) Standard deviation

Standard deviation is the most stable index of variability. In standard deviation, instead of the actual values of the deviations we consider the squares of deviations and the outcome is known as variance. Further, the square root of this variance is known as standard deviation and designated as SD. Thus, standard deviation is the square root of the mean of the squared deviations of the individual observations from the mean. The standard deviation of the sample (ó) and population denoted by (ì) respectively. If all the score have an identical value in a sample, the SD will be 0 (zero).

Merits: It is based on all observations. It is amenable to further mathematical treatments.

Of all measures of dispersion, standard deviation is least affected by fluctuation of sampling.

2.4.3 Skewness and Kurtosis

There are two other important characteristics of frequency distribution that provide useful information about its nature. They are known as skewness and kurtosis.

i) Skewness

Skewness is the degree of asymmetry of the distribution. In some frequency distributions scores are more concentrated at one end of the scale. Such a distribution is called a skewed distribution. Thus, skewness refers to the extent to which a distribution of data points is concentrated at one end or the other. Skewness and variability are usually related, the more the skewness the greater the variability.

ii) Kurtosis

The term 'kurtosis' refers to the 'peakedness' or flatness of a frequency distribution curve when compared with normal distribution curve. The kurtosis of a distribution is the curvedness or peakedness of the graph.

If a distribution is more peaked than normal it is said to be leptokurtic. This kind of peakedness implies a thin distribution.

On the other hand, if a distribution is more flat than the normal distribution it is known as Platykurtic distribution.

A normal curve is known as mesokurtic.

2.4.4 Advantages and Disadvantages of Descriptive Statistics

The *Advantages* of Descriptive statistics are given below:

- It is essential for arranging and displaying data.
- It forms the basis of rigorous data analysis.
- It is easier to work with, interpret, and discuss than raw data.
- It helps in examining the tendencies, variability, and normality of a data set.
- It can be rendered both graphically and numerically.
- It forms the basis for more advanced statistical methods.

The *disadvantages* of descriptive statistics can be listed as given below:

- It can be misused, misinterpreted, and incomplete.
- It can be of limited use when samples and populations are small.
- It offers little information about causes and effects.
- It can be dangerous if not analysed completely.
- There is a risk of distorting the original data or losing important detail.

Self Assessment Questions

1) Which one of the alternative is appropriate for descriptive statistics?

   i) In a sample of school children, the investigator found an average IQ was 110.

   ii) A class teacher calculates the class average on their final exam. Was 64%.

2) State whether the following statements are *True* (T) or *False* (F).

   i) Mean is affected by extreme values ( )

   ii) Mode is affected by extreme values ( )

iii)	Mode is useful in studying qualitative facts such as intelligence	( )
iv)	Median is not affected by extreme values	( )
v)	Range is most stable measures of variability	( )
vi)	Standard deviation is most suitable measures of dispersion	( )
vii)	Skewness is always positive	( )

2.5 MEANING OF INFERENTIAL STATISTICS

In the previous section we discussed about descriptive statistics, which basically describes some characteristics of data. But the description or definition of the distribution or observations is not the prime objective of any scientific investigation.

Organising and summarising data is only one step in the process of analysing the data. In any scientific investigation either the entire population or a sample is considered for the study.

In most of the scientific investigations a sample, a small portion of the population under investigation, is used for the study. On the basis of the information contained in the sample we try to draw conclusions about the population. This process is known as statistical inference.

Statistical inference is widely applicable in behavioural sciences, especially in psychology. For example, before the Lok Sabha or Vidhan Sabha election process starts or just before the declaration of election results print media and electronic media conduct exit poll to predict the election result. In this process all voters are not included in the survey, only a portion of voters i.e. sample is included to infer about the population. This is called inferential statistics.

Inferential statistics deals with drawing of conclusions about large group of individuals (population) on the basis of observation of a few participants from among them or about the events which are yet to occur on the basis of past events. It provides tools to compute the probabilities of future behaviour of the subjects.

Inferential statistics is the mathematics and logic of how this generalisation from sample to population can be made.

There are two types of inferential procedures: (1) Estimation, (2) Hypothesis testing.

2.5.1 Estimation

In estimation, inference is made about the population characteristics on the basis of what is discovered about the sample. There may be sampling variations because of chance fluctuations, variations in sampling techniques, and other sampling errors. Estimation about population characteristics may be influenced by such factors. Therefore, in estimation the important point is that to what extent our estimate is close to the true value.

Characteristics of Good Estimator: A good statistical estimator should have the following characteristics, (i) Unbiased (ii) Consistent (iii) Accuracy . These are being dealt with in detail below.

i) Unbiased

An unbiased estimator is one in which, if we were to obtain an infinite number of random samples of a certain size, the mean of the statistic would be equal to the parameter. The sample mean, ($\overline{x}$) is an unbiased estimate of population mean (μ) because if we look at possible random samples of size N from a population, then mean of the sample would be equal to μ.

ii) Consistent

A consistent estimator is one that as the sample size increased, the probability that estimate has a value close to the parameter also increased. Because it is a consistent estimator, a sample mean based on 20 scores has a greater probability of being closer to (μ) than does a sample mean based upon only 5 scores

ii) Accuracy

The sample mean is an unbiased and consistent estimator of population mean (μ). But we should not over look the fact that an estimate is just a rough or approximate calculation. It is unlikely in any estimate that ($\overline{x}$) will be exactly equal to population mean (μ). Whether or not x is a good estimate of (μ) depends upon the representative ness of sample, the sample size, and the variability of scores in the population.

2.5.2 Point Estimation

We have indicated that $\overline{x}$ obtained from a sample is an unbiased and consistent estimator of the population mean (μ). Thus, if an investigator obtains Adjustment Score from 100 students and wanted to estimate the value of (μ) for the population from which these scores were selected, researcher would use the value of $\overline{x}$ as an estimate of population mean (μ). If the obtained value of x were 45.0 then this value would be used as estimate of population mean (μ).

This form of estimate of population parameters from sample statistic is called point estimation. Point estimation is estimating the value of a parameter as a single point, for example, population mean (μ) = 45.0 from the value of the statistic $\overline{x}$ = 45.0

2.5.3 Interval Estimation

A point estimate of the population mean (μ) almost is assured of being in error, the estimate from the sample will not equal to the exact value of the parameter. To gain confidence about the accuracy of this estimate we may also construct an interval of scores that is expected to include the value of the population mean. Such intervals are called confidence interval. A confidence interval is a range of scores that is expected to contain the value of (μ). The lower and upper scores that determine the interval are called confidence limits. A level of confidence can be attached to this estimate so that, the researcher can be 95% or 99% confidence level that encompasses the population mean.

Self Assessment Questions

1) What is statistical inference?

....

....

2) Explain with illustrations the concept of

   i) Estimation,

   ....

ii) Point estimation,

....

iii) Interval estimation

....

3) What are the procedures involved in statistical inference?

....

....

2.6 HYPOTHESIS TESTING

Inferential statistics is closely tied to the logic of hypothesis testing. In hypothesis testing we have a particular value in mind. We hypothesize that this value characterise the population of observations. The question is whether that hypothesis is reasonable in the light of the evidence from the sample. In estimation no particular population value need be stated. Rather, the question is , what is the population value. For example, Hypothesis testing is one of the important areas of statistical analyses. Sometimes hypothesis testing is referred to as statistical decision-making process. In day-to-day situations we are required to take decisions about the population on the basis of sample information.

2.6.1 Statement of Hypothesis

A statistical hypothesis is defined as a statement, which may or may not be true about the population parameter or about the probability distribution of the parameter that we wish to validate on the basis of sample information.

Most times, experiments are performed with random samples instead of the entire population and inferences drawn from the observed results are then generalised over to the entire population.

But before drawing inferences about the population it should be always kept in mind that the observed results might have come due to chance factor. In order to have an accurate or more precise inference, the chance factor should be ruled out.

The probability of chance occurrence of the observed results is examined by the null hypothesis (H_0). Null hypothesis is a statement of no differences. The other way to state null hypothesis is that the two samples came from the same population. Here, we assume that population is normally distributed and both the groups have equal means and standard deviations.

Since the null hypothesis is a testable proposition, there is counter proposition to it known as alternative hypothesis and denoted by H_1. In contrast to null hypothesis, the alternative hypothesis ($H_{1)}$ proposes that

i) the two samples belong to two different populations,

ii) their means are estimates of two different parametric means of the respective population, and

iii) there is a significant difference between their sample means.

The alternative hypothesis (H_1) is not directly tested statistically; rather its acceptance or rejection is determined by the rejection or retention of the null hypothesis. The probability ‘p’ of the null hypothesis being correct is assessed by a statistical test. If probability ‘p’ is too low, H_0 is rejected and H_1 is accepted.

It is inferred that the observed difference is significant. If probability ‘p’ is high, H_0 is accepted and it is inferred that the difference is due to the chance factor and not due to the variable factor.

2.6.2 Level of Significance

The level of significance ($p<.05$) is that probability of chance occurrence of observed results up to and below which the probability ‘p’ of the null hypothesis being correct is considered too low and the results of the experiment are considered significant ($p\leq$).

On the other hand, if p exceeds , the null hypothesis (H_0) cannot be rejected because the probability of it being correct is considered quite high and in such case, observed results are not considered significant ($p >$).

The selection of level of significance depends on the choice of the researcher. Generally level of significance is taken to be 5% or 1%, i.e., $= .05$ or $= .01$). If null hypothesis is rejected at .05 level, it means that the results are considered significant so long as the probability ‘p’ of getting it by mere chance of random sampling works out to be 0.05 or less ($p< .05$). In other words, the results are considered significant if out of 100 such trials only 5 or less number of the times the observed results may arise from the accidental choice in the particular sample by random sampling.

2.6.3 One-tail and Two-tail Test

Depending upon the statement in alternative hypothesis (H_1), either a one-tail or two-tail test is chosen for knowing the statistical significance. A one-tail test is a directional test. It is formulated to find the significance of both the magnitude and the direction (algebraic sign) of the observed difference between two statistics. Thus, in two-tailed tests researcher is interested in testing whether one sample mean is significantly higher (alternatively lower) than the other sample mean.

2.7 ERRORS IN HYPOTHESIS TESTING

In hypothesis testing, there would be no errors in decision making as long as a null hypothesis is rejected when it is false and also a null hypothesis is accepted when it is true. But the decision to accept or reject the null hypothesis is based on sample data. There is no testing procedure that will ensure absolutely correct decision on the basis of sampled data. There are two types of errors regarding decision to accept or to reject a null hypothesis.

2.7.1 Type I Error

When the null hypothesis is true, a decision to reject it is an error and this kind of error is known as type I error in statistics. The probability of making a type I error is denoted as ‘á’ (read as alpha). The null hypothesis is rejected if the probability ‘p’ of its being correct does not exceed the p. The higher the chosen level of p for considering the null hypothesis, the greater is the probability of type I error.

2.7.2 Type II Error

When null hypothesis is false, a decision to accept it is known as type II error. The probability of making a type II error is denoted as ‘â’ (read as beta). The lower the chosen level of significance p for rejecting the null hypothesis, the higher is the probability of the type II error. With a lowering of p, the rejection region as well as the probability of the type I error declines and the acceptance region (1-p) widens correspondingly.

The goodness of a statistical test is measured by the probability of making a type I or type II error. For a fixed sample size n, á and â are so related that reduction in one causes increase in the other. Therefore, simultaneous reductions in á and â are not possible. If n is increased, it is possible to decrease both á and â.

2.7.3 Power of a Test

The probability of committing type II error is designated by â. Therefore, 1-â is the probability of rejecting null hypothesis when it is false. This probability is known as the power of a statistical test. It measures how well the test is working. The probability of type II error depends upon the true value of the population parameter and sample size n.

Self Assessment Questions

1) Fill in the blanks

   i) Alternative hypothesis is a statement of .... difference.

   ii) Null hypothesis is denoted by ....

   iii) Alternative hypothesis is .... directly tested statistically.

   iv) .... is that probability of chance of occurrence of observed results.

   v) One tail test is a .... statistical test.

   vi) When the null hypothesis is true, a decision to reject is known as.........

   vii) When a null hypothesis is false, a decision to accept is known as.........

2.8 GENERAL PROCEDURE FOR TESTING A HYPOTHESIS

Step 1. Set up a null hypothesis suitable to the problem.

Step 2. Define the alternative hypothesis.

Step 3. Calculate the suitable test statistics.

Step 4. Define the degrees of freedom for the test situation.

Step 5. Find the probability level ‘p’ corresponding to the calculated value of the test statistics and its degree of freedom. This can be obtained from the relevant tables.

Step 6. Reject or accept null hypothesis on the basis of tabulated value and calculated value at practical probability level.

There are some situations in which inferential statistics is carried out to test the hypothesis and draw conclusion about the population , for example (i) Test of hypothesis about a population mean (Z test), (ii) Testing hypothesis about a population mean (small sample ‘ t’ test).

2.9 LET US SUM UP

Descriptive statistics are used to describe the basic features of the data in investigation. Such statistics provide summaries about the sample and measures. Data description comprises two operations: organising data and describing data. Organising data includes: classification, tabulation, graphical and diagrammatic presentation of raw scores. Whereas, measures of central tendency and measures of dispersion are used in describing the raw scores.

In the above section, the basic concepts and general procedure involved in inferential statistics are also discussed. Inferential statistics is about inferring or drawing conclusions from the sample to population. This process is known as statistical inference. There are two types of inferential procedures: estimation and hypothesis testing. An estimate of unknown parameter could be either point or interval. Hypothesis is a statement about a parameter. There are two types of hypotheses: null and alternative hypotheses. Important concepts involved in the process of hypothesis testing example, level of significance, one tail test, two tail test, type I error, type II error, power of a test are explained. General procedure for hypothesis testing is also given.

2.10 UNIT END QUESTIONS

1) What is descriptive statistics? Discuss its advantages and disadvantages.
2) What do you mean by organisation of data? State different methods of organising raw data.
3) Define measures of dispersion. Why is it that standard deviation is considered as the best measures of variability?
4) Explain the importance of inferential statistics.
5) Describe the important properties of good estimators.
6) Discuss the different types of hypothesis formulated in hypothesis testing.
7) Discuss the errors involved in hypothesis testing.
8) Explain the various steps involved in hypothesis testing.

2.11 GLOSSARY

Classification : A systematic grouping of data

Cumulative frequency distribution : A classification, which shows the cumulative frequency below, the upper real limit of the corresponding class interval.

Data : Any sort of information that can be analysed.

Discrete data : When data are counted in a classification.

Exclusive classification : The classification system in which the upper limit of the class becomes the lower limit of next class

Frequency distribution : Arrangement of data values according to their magnitude.

Inclusive classification	:	When the lower limit of a class differs the upper limit of its successive class.
Mean	:	The ratio between total and numbers of scores.
Median	:	The mid point of a score distribution.
Mode	:	The maximum occurring score in a score distribution.
Central Tendency	:	The tendency of scores to bend towards center of distribution.
Dispersion	:	The extent to which scores tend to scatter from their mean and from each other.
Standard Deviation	:	The square root of the sum of squared deviations of scores from their mean.
Skewness	:	Tendency of scores to polarize on either side of abscissa.
Kurtosis	:	Curvedness of a frequency distribution graph.
Range	:	Difference between the two extremes of a score distribution.
Confidence level	:	It gives the percentage (probability) of samples where the population mean would remain within the confidence interval around the sample mean.
Estimation	:	It is a method of prediction about parameter value on the basis Statistic.
Hypothesis testing	:	The statistical procedures for testing hypotheses..
Level of significance	:	The probability value that forms the boundary between rejecting and not rejecting the null hypothesis.
Null hypothesis	:	The hypothesis that is tentatively held to be true (symbolised by Ho)
One-tail test	:	A statistical test in which the alternative hypothesis specifies direction of the departure from what is expected under the null hypothesis.
Parameter	:	It is a measure of some characteristic of the population.
Population	:	The entire number of units of research interest.
Power of a test	:	An index that reflects the probability that a statistical test will correctly reject the null hypothesis relative to the size of the sample involved.
Sample	:	A sub set of the population under study.

Statistical inference : It is the process of concluding about an unknown population from known sample drawn from it.

Statistical hypothesis : The hypothesis which may or may not be true about the population parameter.

t-test : It is a parametric test for the significance of differences between means.

Type I error : A decision error in which the statistical decision is to reject the null hypothesis when it is actually true.

Type II error : A decision error in which the statistical decision is not to reject the null hypothesis when it is actually false.

Two-tail test : A statistical test in which the alternative hypothesis does not specify the direction of departure from what is expected under the null hypothesis.

2.12 SUGGESTED READINGS

Asthana, H. S. and Bhushan, B. (2007). *Statistics for Social Sciences* (with SPSS Application). Prentice Hall of India, New Delhi.

Garret, H. E. (2005). *Statistics in Psychology and Education*. Jain publishing, India.

Elhance, D. N., and Elhance, V. (1988). *Fundamentals of Statistics*. Kitab Mahal, Allahabad

Nagar, A. L., and Das, R. K. (1983). *Basic Statistics*. Oxford University Press, Delhi.

Sani, F., and Todman, J. (2006). *Experimental Design and Statistics for Psychology.* Blackwell Publishing U.K.

Yale, G. U., and M.G. Kendall (1991). *An Introduction to the Theory of Statistics*. Universal Books, Delhi.

UNIT 3 TYPE I AND TYPE II ERRORS

Structure

3.0 INTRODUCTION

Each and every discipline needs Statistics and thus is the importance of statistics. One finds that statistics is of great importance to government organisations, non government organisation, experts of all the fields and also to students. Statistics is used for a wide variety of purposes. It is also true that all the time they are not accurate and correct. Sometimes the results are known and sometimes unknown. In the words of Statistics these are known as errors. To achieve accuracy in the concerned field it is important to understand these concepts in detail. It is also important to understand and discuss the related concepts which would be helpful to understand type I and type II errors. In this unit we would be dealing with the definition and concept of errors in statistics and focus on type I and type II errors wich are essential to understand when we deal with statistics and make interpretation of the results using statistics.

3.1 OBJECTIVES

After completing this unit, you will be able to:

- define and differentiate between Type I and Type II errors;
- describe probability concept and the level of significance;
- define and differentiate between one tailed and two tailed tests;
- explain the significance of Normal probability curve;
- define the Cut off sample scores; and
- describe what is z-scores.

3.2 DEFINITION AND CONCEPTS

Before moving onwards we should know the related concepts of Type I and Type II Errors. The concepts that need to be understood include the following:

1) Hypothesis testing
2) The hypothesis – testing process
3) Null Hypothesis
4) Population
5) Sample
6) Rejecting and accepting null hypothesis
7) One-tailed and two-tailed hypothesis
8) Decision errors

3.2.1 Hypothesis Testing

Hypothesis testing has a vital role in psychological measurements. By hypothesis we mean the tentative answer to any questions. Hypothesis testing is a systematic procedure for deciding whether the results of a research study, which examines a sample, support a particular theory or practical innovation, which applies to a population. Hypothesis testing is the central theme in most psychology research.

Hypothesis testing involves grasping ideas that make little sense. Real life psychology research involves samples of many individuals. At the same time there are studies which involve a single individual.

3.2.2 The Core Logic of Hypothesis Testing

There is a standard kind of reasoning researchers use for any hypothesis/testing problem. For this example, it works as follows. Ordinarily, among the population of babies that are not given the specially purified vitamin, the chance of a baby's starting to walk at age 8 months or earlier would be less than 2% Thus, walking at 8 months or earlier is highly unlikely among such babies. But what if the randomly selected sample of one baby in our study does start walking by 8 months? If the specially purified vitamin had no effect on this particular baby's walking age (which means that the baby's walking age should be similar to that of babies that were not given the vitamin), it is highly unlikely (less than a 2% chance) that the particular baby we selected at random would start walking by 8 months. So, if the baby in our study does in fact start walking by 8 months, that allows us to *reject* the idea that the specially purified vitamin has no effect. And if we reject the idea that the specially purified vitamin has no effect, then we must also accept the idea that the specially purified vitamin does have an effect. Using the same reasoning, if the baby starts walking by 8 months, we can reject the idea that this baby comes from a population of babies with a mean walking age of 14 months. We therefore conclude that babies given the specially purified vitamin will start to walk before 14 months. Our explanation for the baby's early walking age in the study is that the specially purified vitamin speeded up the baby's development.

The researchers first spelled out what would have to happen for them to conclude that the special purification procedure makes a differences. Having laid this out in advance the researchers could then go on to carry out their study. In this example, carrying out the study means giving the specially purified vitamin to a randomly selected baby and watching to see how early that baby walks. Suppose the result of the study is that the baby starts walking before 8 months. The researchers would then conclude that it is unlikely the specially purified vitamin makes no difference and thus also conclude that it does make a difference.

This kind of testing the opposite-of-what-you-predict, roundabout reasoning is at the heart of inferential statistics in psychology. It is something like a double negative. One reason for this approach is that we have the information to figure the probability of getting a particular experimental result if the situation of there being no difference is true. In the purified vitamin example, the researchers know what the probabilities are of babies walking at different ages if the specially purified vitamin does not have any effect. It is the probability of babies walking at various ages that is already known from studies of babies in general – that is, babies who have not received the specially purified vitamin. (Suppose the specially purified vitamin has no effect. In that situation, the age at which babies start walking is the same whether or not they receive the specially purified vitamin.)

Without such a tortuous way of going at the problem, in most cases you could just not do hypothesis testing at all. In almost all psychology research, we base our conclusions on this question: What is the probability of getting our research results. If the opposite of what we are predicting were true? That is, we are usually predicting an effect of some kind. However, we decide on whether there is such an effect by seeing if it is unlikely that there is not such an effect. If it is highly unlikely that we would get our research results if the opposite of what we are predicting were true, that allows us to reject that opposite prediction. If we reject that opposite prediction, we are able to accept our prediction. However, if it is likely that we would get our research results if the opposite of what we are predicting were true, we are not able to reject that opposite prediction. If we are not able to reject that opposite prediction, we are not able to accept our prediction.

3.2.3 The Hypothesis – Testing Process

Let's look at example in this time going over each step in some detail. Along the way, we cover the special terminology of hypothesis-testing. Most important, we introduce five steps of hypothesis testing you use for the rest of the course.

Step 1: Restate the question as a research Hypothesis and Null Hypothesis about the populations

Our researchers are interested in the effects on babies in general (not just this particular baby). That is, the purpose of studying samples is to know about populations thus, it is useful to restate the research question in terms of populations.

In our example, we can think of two populations of babies.

Population 1: Babies who take the specially purified vitamin.

Population 2: Babies who do not take the specially purified vitamin

Population 1 comprises those babies who receive the experimental treatment. In our example, we use a sample of one baby to draw a conclusion about the age that babies in Population 1 start to walk. Population 2 is a kind of comparison baseline of what is already known.

The prediction of our research team is that Population 1 babies (those who take the specially purified vitamin) will on the average walk earlier than population 2 babies (those who do not take the specially purified vitamin) $\mu_1 < \mu_2$

The opposite of the research hypothesis is that the populations are not different in the way predicted. Under this scenario, population 1 babies (those who take the specially purified vitamin) will on the average not walk earlier than Population 2 babies (those who do not take the specially purified vitamin). That is, this prediction is that there is no difference in when population 1 and Population 2 babies start walking. They start at the same time. A statement like this, about a lack of difference between populations, is the crucial opposite of the research hypothesis. It is called a null hypothesis. It has this name because it states the situation in which there is no difference (the difference is "null") between the between the populations. In symbols, the null hypothesis is $\mu_1 < \mu_2^1$.

The research hypothesis and the null hypothesis are complete opposites: if one is true, the other cannot be. In fact, the research hypothesis is sometimes called the alternative hypothesis – that is, it is the alternative to the null hypothesis. This is a bit ironic. As researchers, we care most about the research hypothesis. But when doing the steps of hypothesis so that we can decide about its alternative (the research hypothesis).

Step 2: Determine the Characteristics of the comparison Distribution

Recall that the overall logic of hypothesis testing involves figuring out the probability of getting a particular result if the null hypothesis is true. Thus, you need to know what the situation would be if the null hypothesis were true. Population 2 we know $\mu = 14$, $\sigma = 3$, and it is normally distributed. If the null hypothesis is true,

Population 1 and Population 2 are the same – in our example, this would mean Populations 1 and 2 both follow a normal curve, $\mu = 14$, $\sigma = 3$.

In the hypothesis-testing process, you want to find out the probability that you could have gotten a sample score as extreme as what you got (say, a baby walking very early) if your sample were from a population with a distribution of the sort you would have if the null hypothesis were true. Thus, in this book we call this distribution a comparison distribution. (The comparison distribution is sometimes called a statistical model or a sampling distribution – an idea we discuss in Chapter 5.) That is, in the hypothesis-testing process, you compare the actual sample's score to this comparison distribution.

In our vitamin example, the null hypothesis is that there is no difference in walking age between babies that take the specially purified vitamin (Population 1) and babies that do not take the specially purified vitamin (Population 2). The comparison distribution is the distribution for Population 2, since this population represents the walking age of babies if the null hypothesis is true. In later chapters, you will learn about different types of comparison distributions, but the same principle applies in all cases: The comparison distribution is the distribution that represents the population situation if the null hypothesis is true.

Step 3-Determine the Cutoff Sample Score on the comparison Distribution at Which the null hypothesis sould be rejected.

Ideally, before conducting a study, researchers set a target against which they will compare their result – how extreme a sample score they would need to decide against the null hypothesis: that is, how extreme the sample score would have to be for it to be too unlikely that they could get such an extreme score if the null hypothesis were true. This is called the cutoff sample score. (The cutoff sample score is also known as the critical value.)

Step 4: Determine your sample's Score on the Comparison Distribution

The next step is to carry out the study and get the actual result for your sample. Once you have the results for your sample, you figure the Z score for the sample's raw score based on the population mean and standard deviation of the comparison distribution.

Assume that the researchers did the study and the baby who was given the specially purified vitamin started walking at 6 months. The mean of the comparison distribution to which we are comparing these results is 14 months and the standard deviation is 3 months. That is $\mu = 14, \sigma = 3$. Thus, a baby who walks at 6 months is 8 months below the population mean. This puts this baby $2\frac{2}{3}$ standard deviations below the population mean. The Z score for this sample baby on the comparison distribution is thus – 2.67 (Z = [6 - 14]/3=-2.67).

Step 5: Decide Whether to reject the null hypothesis

To decide whether to reject the null hypothesis, you compare your actual sample's Z score (from Step 4) to the cutoff Z score (from Step 3). In our example, the actual result was – 2.67. Let's suppose the researchers had decided in advance that they would reject the null hypothesis if the sample's Z score was below – 2. Since – 2.67 is below – 2, the researchers would reject he null hypothesis.

Or, suppose the researchers had used the more conservative 1% significance level. The needed Z score to reject the null hypothesis would then have been – 2.33 or lower. But, again, the actual Z for the randomly selected baby was – 2.67 (a more extreme score than – 2.33). Thus, even with this more conservative cutoff, they would still reject the null hypothesis.

3.2.4 Implications of Rejecting or Failing to Reject the Null Hypothesis

It is important to emphasise two points about the conclusions you can make from the hypothesis-testing process. First, suppose you reject the null hypothesis. Therefore, your result supports the research hypothesis (as in our example). You would still not say that the results prove the research hypothesis or that the results show that the research hypothesis is true. This would be too strong because the results of research studies are based on probabilities. Specifically, they are based on the probability being low of getting your result if the null hypothesis were true. Proven and true are okay in logic and mathematics, but to use these worlds in conclusions from scientific research is quite unprofessional. (It is okay to use true when speaking hypothetically) – for example, "if this hypothesis were true, then…" – but not when speaking of conclusions about an actual result.) what you do say when you reject the null hypothesis is that the results are statistically significant.

Second, when a result is not extreme enough to reject the null hypothesis, you do not say that the result supports the null hypothesis. You simply say the result is not statistically significant.

A result that is not strong enough to reject the null hypothesis means the study was inconclusive. The results may not be extreme enough to reject the null hypothesis, but the null hypothesis might still be false (and the research hypothesis true). Suppose in our example that the specially purified vitamin had only a slight but still real effect. In that case, we would not expect to find a baby given the purified vitamin to be walking a lot earlier than babies in general. Thus, we would not be able to reject the null hypothesis, even though it is false. (You will learn more about such situations in the Decision Errors section later in this chapter).

Showing the null hypothesis to be true would mean showing that there is absolutely no difference between the populations it is always possible that there is a difference between the populations, but that the difference is much smaller than what the particular study was able to detect. Therefore, when a result is not extreme enough to reject the null hypothesis, the results are inconclusive. Sometimes, however, if studies have been done using large samples and accurate measuring procedures, evidence may build up in support of something close to the null hypothesis – that there is at most very little difference between the populations.

3.2.5 One-Tailed and Two-Tailed Hypothesis Tests

In our examples so far, the researchers were interested in only one direction of result. In our first example, researchers tested whether babies given the specially purified vitamin would walk earlier than babies in general. In the happiness example, the personality psychologists predicted the person who received $10 million would be happier than other people. The researchers in these studies were not interested in the possibility that giving the specially purified vitamin would cause babies to start walking later or that people getting $10 million might become less happy.

Directional hypotheses and One-Tailed tests

The purified vitamin and happiness studies are examples of testing directional hypotheses. Both studies focused on a specific direction of effect. When a researcher makes a directional hypothesis, the null hypothesis is also, in a sense, directional. Suppose the research hypothesis is that getting $10 million will make a person happier. The null hypothesis, then, is that the money will either have no effect or make the person less happy (in symbols, if the research hypothesis is $\mu > \mu_{2,}$ then the null hypothesis is $\mu_1 \leq \mu_2$ $\leq$ is the symbol for less than or equal to.) thus to reject the null hypothesis, the sample had to have a score in one particular tail of the comparison distribution – the upper extreme or tail (in this example, the top 5%) of the comparison distribution. (When it comes to rejecting the null hypothesis with a directional hypothesis, a score at the other tail would be the same as a score in the middle – that is, it would not allow you to reject the null hypothesis). For this reason, the test of a directional hypothesis is called a one-tailed test. A one-tailed test can be one-tailed in either direction. In the happiness study example, the tail for the predicted effect was at the high end. In the baby study example, the tail for the predicted effect was at the low end (that is, the prediction tested was that babies given the specially purified vitamin would start walking unusually early).

Non-directional hypotheses and two-tailed tests

Sometimes, a research hypothesis states that an experimental procedure will have an effect, without saying whether it will produce a very high score or a very low score. Suppose an organisational psychologist is interested in how a new social skills program will affect productivity. The program could improve productivity by making the working environment more pleasant. Or, the program could hurt productivity by encouraging people to socialise instead of work. The research hypothesis is that the social skills program changes the level of productivity; the null hypothesis is that the program does not change productivity one way or the other. In symbols, the research hypothesis is $\mu_1 \neq \mu_2$ the null hypothesis is $\mu_1 = \mu_2$

When a research hypothesis predicts an effect but does not predict a particular direction for the effect, it is called a non-directional hypothesis. To test the significance of a non-directional hypothesis, you have to take into account the possibility chance of non-directional hypothesis, you have to take into account the possibility that the sample could be extreme at either tail of the comparison distribution. Thus this is called a two-tailed test.

3.2.6 Decision Errors

Another crucial topic for making sense of statistical significance is the kind of errors that are possible in the hypothesis-testing process. The kind of errors we consider here are about how, in spite of doing all your figuring correctly, your consider here are about how, in spite of doing all your figuring correctly, your conclusions from hypothesis-testing can still be incorrect. It is not about making mistakes in calculations or even about using the wrong procedures. That is, mistakes in calculations or even about using the wrong procedures. That is, decision errors are situations in which the right procedures lead to the wrong decisions.

Decision errors are possible in hypothesis testing because you are making decisions about populations based on information in samples. The whole hypothesis testing process is based on probabilities. The hypothesis-testing process is set up to make the probability of decision errors as small as possible. For example, we only decide to reject the null hypothesis if a sample's mean is so extreme that there is a very small probability (say, less than 5%) that we could have gotten such an extreme sample if the null hypothesis is true. But a very small probability is not the same as a zero probability! Thus, in spite of your best intentions, decision errors are always II errors.

3.3 TYPE I ERROR

You make a Type I error if you reject the null hypothesis when in fact the null hypothesis is true. Or, to put it in terms of the research hypothesis, you make a Type I error when you conclude that the study supports the research hypothesis when in reality the research hypothesis is false.

Suppose you carried out a study in which you had set the significance level cut off at a very lenient probability level, such as 20%. This would mean that it would not take a very extreme result to reject the null hypothesis. If you did many studies like this, you would often (about 20% of the time) be deciding to consider the research hypothesis supported when you should not. That is, you would have a 20% chance of making a Type I error.

Even when you set the probability at the conventional .05 or .01 levels, you will still make a Type I error sometimes (5% or 1% of the time). Consider again the example of giving the new therapy to a depressed patient. Suppose the new therapy is not more effective than the usual therapy. However, in randomly picking a sample of one depressed patient to study, the clinical psychologists might just happen to pick a patient whose depression would respond equally well to the new therapy and the usual therapy. Randomly selecting a sample patient like this is unlikely, but such extreme samples are possible, and should this happen, the clinical psychologists would reject the null hypothesis and conclude that the new therapy is different than the usual therapy. Their decision to reject the null hypothesis would be wrong – a Type I error. Of course, the researchers could not know they had made a decision error of this kind. What reassures researchers is that they know from the logic of hypothesis testing that the probability of making such a decision error is kept low (less than 5% if you use the .05 significance level).

Still, the fact that Type I errors can happen at all is of serious concern to psychologists, who might construct entire theories and research programs, not to mention practical applications, based on a conclusion from hypothesis testing that is in fact mistaken. It is because these errors are of such serious concern that they are called Type I.

As we have noted, researchers cannot tell when they have made a Type I error. However, they can try to carry out studies so that the chance of making a Type I error is as small as possible.

What is the chance of making a Type I error? It is the same as the significance level you set. If you set the significance level at $p < .05$, you are saying you will reject the null hypothesis if there is less than a 5% (.05) chance that you could have gotten your result if the null hypothesis were true. When rejecting the null hypothesis in this way, you are allowing up to a 5% chance that you got your results even though the null hypothesis was actually true. That is, you are allowing a 5% chance of a Type I error.

The significance level, which is the chance of making a Type I error, is called alpha (the Greek letter Ü). The lower the alpha, the smaller the chance of a Type I error. Researchers who do not want to take a lot of risk set alpha lower than .05 such a $p<.001$ in this way the result of a study has to be very extreme in order for the hypothesis testing process to reject the null hypothesis.

Using a .001 significance level is like buying insurance against making a Type I error. However, when buying insurance, the better the protection, the higher the cost. There is a cost in setting the significance level at too extreme a level. We turn to that cost next.

3.4 TYPE II ERROR

If you set a very stringent significance level, such as .001, you run a different kind of risk. With a very stringent significance level, you may carry out a study in which in reality the research hypothesis is true, but the result does not come out extreme enough to reject the null hypothesis. Thus, the decision error you would make is in not rejecting the null hypothesis when in reality the null hypothesis is false to put this in terms of the research hypothesis, you make this kind of decision error when the hypothesis-testing procedure leads you to decide that the results of the study are inconclusive when in reality the research hypothesis is true. This is called a Type II error. The probability of making a Type II error is called beta (the Greek letter β).

3.5 RELATIONSHIP BETWEEN TYPE I AND TYPE II ERRORS

When it comes to setting significance levels, protecting against one kind of decision error increases the chance of making the other. The insurance policy against Type I error (setting a significance level of, say, .001) has the cost of increasing the chance of making a Type II error. (This is because with a stringent significance level like

.001, even if the research hypothesis is true, the results have to be quite strong to be extreme enough to reject the null hypothesis.) The insurance policy against Type II error (setting a significance level of say .20) has the cost of increasing the chance of making a Type I error. (This is because with a level of significance like .20, even if the null hypothesis is true, it is fairly easy to get a significant result just by accidentally getting a sample that is higher or lower than the general population before doing the study.)

3.6 LET US SUM UP

Hypothesis testing considers the probability that the result of a study could have come about even if the experimental procedure had no effect. If this probability is low, the scenario of no effect is rejected and the theory behind the experimental procedure is supported.

The expectation of an effect is the research hypothesis, and the hypothetical situation of no effect is the null hypothesis.

When a result (that is, a sample score) is so extreme that the result would be very unlikely if the null hypothesis were true, the null hypothesis is reject and the research hypothesis supported. If the result is not that extreme, the null hypothesis is not rejected and the study is inconclusive.

Psychologists usually consider a result too extreme if it is less likely than 5% (that is, a significance level of .05) to have come about, if the null hypothesis were true. Psychologists sometimes use a more stringent 1% (.01 significance level), or even .01% (.001) significance level), cutoff.

The cutoff percentage is the probability of the result being extreme in a predicted direction in a directional or one-tailed test. The cutoff percentages are the probability of the result being extreme in either direction in a non-directional or two-tailed test.

There are two kinds of decision errors one can make in hypothesis testing. A Type I error is when a researcher rejects the null hypothesis, but the null hypothesis is actually true. A Type II error is when a researcher does not reject the null hypothesis, but the null hypothesis is actually false.

There has been much controversy about significance tests, including critiques of the basic logic and, especially, that they are often misused. One major way significance tests are misused is when researchers interpret not rejecting the null hypothesis as demonstrating that the null hypothesis is true.

Research articles typically report the results of hypothesis testing by saying a result was or was not significant and giving the probability level cutoff (usually 5% or 1%) the decision was based on. Research articles rarely mention decision errors.

3.7 UNIT END QUESTIONS

1) Fill in the blanks with appropriate terms:

   i) The research hypothesis and_____________are completely opposite.

   ii) Cutoff sample score are also known as________________

   iii) In Type I Error we __________ hypothesis when it is true.

   iv) The hypothesis in which we Accept Null hypothesis is called___________

2) Mark (T) for True statement and (F) for False Statement:

   i) The probability of making Type II error is called. ( )

   ii) The significance level, which is chance of making Type II Error, is called. ( )

   iii) The significance level, which is chance of making Type I error is called. ( )

   iv) One directional hypothesis is termed as two tailed test. ( )

   v) A hypothesis which predicts an effect but does not predict a particular direction for the effect is called non-directional hypothesis. ( )

   vi) One tailed test has either direction only. ( )

3) Give brief answers of the following questions:

   i) What is Comparison distribution?

   ii) What is research hypothesis?

   iii) What is directional hypothesis?

   iv) What is Type I error?

   v) What do you mean by two tailed test?

   vi) Describe briefly the relationship between Type I & Type II errors.

3.8 GLOSSARY

Hypothesis : Tentative statement which can be tested.

Research hypothesis : Statement about the predicted relation between populations.

Null hypothesis : A Statement opposite to the research hypothesis.

Alternate hypothesis : A statement which is opposite to the null hypothesis

Level of Significance : Probability of getting statistical significance of null hypothesis is accurately true.

Comparison distributions : Distribution used in hypothesis testing.

One tailed test : Hypothesis testing procedure for a directional hypothesis

Two tailed test : Hypothesis testing procedure for a non-directional hypothesis

Sample : Scores of particular group of people studied.

Type I Error : When we reject a null hypothesis when it is true

Type II : When we accept a null hypothesis when it is false

α(alpha) : Probability of making type – I Error

β(Beta) : Sampling distribution Probability of making type I Error

Normal curve : Bell shaped frequency distribution that is Symmetrical and unimodel.

3.9 SUGGESTED READINGS

Asthana H.S, and Bhushan. B. (2007) *Statistics for Social Sciences* (with SPSS Applications).

B.L. Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Gupta, S.C. (1990) *Fundamentals of Statistics*

Sidney Siegel, & N. John Castetellan, Jr. *Non-parametric Statistics for the Behavioural Science*. McGraw Hill Books company

Y.P. Aggarwal. *Statistical Methods Concepts*, Application & Computation Sterling homosedasity publishers Pvt. Ltd.

UNIT 4 SETTING UP THE LEVELS OF SIGNIFICANCE

Structure

4.0 INTRODUCTION

In behavioural sciences nothing is absolute. Therefore, while obtaining the findings through statistical analyses, behavioural scientists usually ignore the error to the maximum of 5%. In statistics, a result is called statistically significant if it is unlikely to have occurred by chance. The phrase, "test of significance" was coined by Ronald Fisher. As used in statistics, significance does not mean importance or meaning fitness as it does in everyday speech. In this unit we will be dealing with the definition and concept of level of significance and how the level of significance is decided. Since level of significance is related to hypothesis testing we will be dealing with null hypothesis and alternative hypothesis and how these are to be tested in different types of experiments. While dealing with hypothesis testing we will also be dealing with experimental designs , errors in hypothesis testing etc. We will also learn what is meant by confidence limits and how these are established.

4.1 OBJECTIVES

After completing this unit, you will be able to:

- define and put forward the concept of null hypothesis;

- describe the process of hypothesis testing;
- explain the confidence limits;
- elucidate the errors in hypothesis testing and its relationship to levels of significance;
- explain level of significance;
- describe the setting up of level of significance; and
- analyse the experimental designs in relation to levels of significance.

4.2 HYPOTHESIS TESTING

Many a time, we strongly believe some results to be true. But after taking a sample, we notice that one sample data does not wholly support the result. The difference is due to

i) the original belief being wrong, or

ii) the sample being slightly one sided.

Tests are, therefore, needed to distinguish between two possibilities. These tests tell about the likely possibilities and reveal whether or not the difference can be due to only chance elements. If the difference is not due to chance elements, it is significant and, therefore, these tests are called tests of significance. The whole procedure is known as *Testing of Hypothesis*.

Setting up and testing hypotheses are essential part of statistical inference. In order to formulate such a test, usually some theory is put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. For example, the hypothesis may be the claim that a new drug is better than the current drug for treatment of a disease, diagnosed through a set of symptoms.

In each problem considered, the question of interest is simplified into two competing claims that is the hypotheses between which we have a choice; the null hypothesis, denoted by Ho, against the alternative hypothesis, denoted by H_1. These two competing claims or hypotheses are not however treated on an equal basis. Special consideration is given to the null hypothesis which states that there is no difference between the two drugs. Null hypothesis is also called as "No Difference" hypothesis. We have two common situations: Let us say we have formulated the null hypothesis stating that "There will be no difference between the two drugs in regard to treating a disorder" We carry out an experiment to prove that the null hypothesis is true or reject that null hypothesis as the experimental results show that there is a difference in the treatment by the two drugs. Thus we have two situations as given below:

i) The experiment has been carried out in an attempt to prove or reject a particular hypothesis, the null hypothesis. We give priority to the null hypothesis and say that we will reject it only if the evidence against it is sufficiently strong.

ii) If one of the two hypotheses is 'simpler', we give it priority so that a more 'complicated' theory is not adopted unless there is sufficient evidence against the simpler one. For example, it is 'simpler' to claim that there is no difference in the treatment between the two drugs than to state the alternate hypothesis that drug A will be better than dug B, that is stating that there is a difference.

The hypotheses are often statements about population parameters like expected value and variance. Take for example another null hypothesis, H_0, the statement that the expected value of the height of ten year old boys in the Indian population is not different from that of ten year old girls.

A hypothesis can also be a statement about the distributional form of a characteristic of interest. For instance, the statement that the height of ten year old boys is normally distributed within the Indian population. This is a hypothesis in statement form regarding the distribution of 10 year old boys height in the population.

4.3 NULL HYPOTHESIS

The null hypothesis, H_0, represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved. For example, in respect of a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug. We would write H_0: there is no difference between the two drugs on average.

We give special consideration to the null hypothesis. This is due to the fact that the null hypothesis relates to the statement being tested, whereas the alternative hypothesis relates to the statement to be accepted if when the null hypothesis is rejected.

The alternative hypothesis, H_1, is a statement of what a statistical hypothesis test is set up to establish. For example, in a clinical trial of a new drug, the alternative hypothesis might be that the new drug has a different effect, on average, compared to that of the current drug.

We would write H_1: the two drugs have different effects, on average.

The alternative hypothesis might also be that the new drug is better, on average, than the current drug. In this case, we would write H_1: the new drug is better than the current drug, on average.

The final conclusion once the test has been carried out is always given in terms of the null hypothesis. We either 'reject H_0 in favour of H_1' or 'do not reject H_o'; we never conclude 'reject H_1', or even 'accept H_1'.

If we conclude 'do not reject H_0', this does not necessarily mean that the null hypothesis is true, It only suggests that there is not sufficient evidence against H0 in favour of H_1 that we reject the null hypothesis. It only suggests that the alternative hypothesis may be true.

Thus one may state that Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution.

First, a tentative assumption is made about the parameter or distribution. This assumption is called the null hypothesis and is denoted by H_0. An alternative hypothesis (denoted by H_1), which is the opposite of what is stated in the null hypothesis, is then defined. The hypothesis-testing procedure involves using sample data to determine whether or not H_0 can be rejected. If H_0 is rejected, the statistical conclusion is that the alternative hypothesis H_1 is true.

A hypothesis is a statement supposed to be true till it is proved false. It may be based on previous experience or may be derived theoretically. First a statistician or the investigator forms a research hypothesis that an exception is to be tested. Then she/ he derives a statement which is opposite to the research hypothesis (noting as H_0). The approach here is to set up an assumption that there is no contradiction between the believed result and the sample result and that the difference, therefore, can be ascribed solely to chance. Such a hypothesis is called a null hypothesis (H_0). It is the null hypothesis that is actually tested, not the research hypothesis. The object of the test is to see whether the null hypothesis should be rejected or accepted.

If the null hypothesis is rejected, that is taken as evidence in favour of the research hypothesis which is called as the alternative hypothesis (denoted by H_1). In usual practice we do not say that the research hypothesis has been "proved" only that it has been supported.

For example, assume that a radio station selects the music it plays based on the assumption that the average age of its listening audience is 30 years. To determine whether this assumption is valid, a hypothesis test could be conducted with the null hypothesis as H_0: = 30 and the alternative hypothesis as H_1: 30. Based on a sample of individuals from the listening audience, the sample mean age, can be computed and used to determine whether there is sufficient statistical evidence to reject H_0. Conceptually, a value of the sample mean that is "close" to 30 is consistent with the null hypothesis, while a value of the sample mean that is "not close" to 30 provides support for the alternative hypothesis.

Self Assessment Questions

Fill in blanks with appropriate terms.

1) Generally the .05 and the _______________ levels of significance are mostly used.
2) Standard Error of mean is calculated by the formula __________________.
3) In case of two-tailed test (+1.96) with fall on ____________________ of the normal curve.
4) In case of .05 level of significance amount of confidence will be ___________.

4.4 ERRORS IN HYPOTHESIS TESTING

Ideally, the hypothesis-testing procedure leads to the acceptance of H_0 when H_0 is true and the rejection of H_0 when H_0 is false. Unfortunately, since hypothesis tests are based on sample information, the possibility of errors must be considered. A Type-I error corresponds to rejecting H_0 when H_0 is actually true, and a Type-II error corresponds to accepting H_0 when H_0 is false.

In testing any hypothesis, we get only two results: either we accept or we reject it. We do not know whether it is true or false. Hence four possibilities may arise.

i) The hypothesis is true but test rejects it (Type-I error).

ii) The hypothesis is false but test accepts it (Type-II error).

iii) The hypothesis is true and test accepts it (correct decision).

iv) The hypothesis is false and test rejects it (correct decision)

Type-I Error

In a hypothesis test, a Type-I error occurs when the null hypothesis is rejected when it is in fact true. That is, H0 is wrongly rejected. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug. That is, there is no difference between the two drugs on average. A Type-I error would occur if we concluded that the two drugs produced different effects when in fact there was no difference between them.

A Type-I error is often considered to be more serious, and therefore more important to avoid, than a Type-TI error.

The hypothesis test procedure is therefore adjusted so that there is a guaranteed 'low' probability of rejecting the null hypothesis wrongly;

This probability is never 0. This probability of a Type-I error can be precisely computed as, P (Type-I error) significance level =

The exact probability of a Type-I error is generally unknown.

If we do not reject the null hypothesis, it may still be false (a Type-I error) as the sample may not be big enough to identify the falseness of the null hypothesis (especially if the truth is very close to hypothesis).

For any given set of data, Type-I and Type-Il errors are inversely related; the smaller the risk of one, the higher the risk of the other.

A Type-I error can also be referred to as an error of the first kind.

Type-II Error

In a hypothesis test, a Type-II error occurs when the null hypothesis, Ho, is not rejected when it is in fact false. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is Ho: there is no difference between the two drugs on average.

A Type-II error would occur if it was concluded that the two drugs produced the same effect, that is, there is no difference between the two drugs on average, when in fact they produced different effects.

A Type-II error is frequently due to sample sizes being too small.

The probability of a Type-II error is symbolised by â and written:

P (Type-II error) = â (but is generally unknown).

A Type-II error can also be referred to as an error of the second kind.

Hypothesis testing refers to the process of using statistical analysis to determine if the observed differences between two or more samples are due to random chance factor (as stated in the null hypothesis) or is it due to true differences in the samples (as stated in the alternate hypothesis).

A null hypothesis (H_0) is a stated assumption that there is no difference in parameters (mean, variance) for two or more populations. The alternate hypothesis (H_1) is a statement that the observed difference or relationship between two populations is real and not the result of chance or an error in sampling.

Hypothesis testing is the process of using a variety of statistical tools to analyse data and, ultimately, to fail to reject or reject the null hypothesis. From a practical point of view, finding statistical evidence that the null hypothesis is false allows you to reject the null hypothesis and accept the alternate hypothesis.

Because of the difficulty involved in observing every individual in a population for research purposes, researchers normally collect data from a sample and then use the sample data to help answer questions about the population.

A hypothesis test is a statistical method that uses sample data to evaluate a hypothesis about a population parameter.

The hypothesis testing is standard and it follows a specific order as given below.

i) first state a hypothesis about a population (a population parameter, e.g. mean

ii) obtain a random sample from the population and also find its mean , and

iii) compare the sample data with the hypothesis on the scale (standard z or normal distribution).

Self Assessment Questions

1) What is Type I error? Give suitable examples.

....

....

2) What is Type II error? Give example.

....

....

3) What is hypothesis testing? What are the steps for the same?

....

....

4) What is Null hypothesis and alternate hypothesis?

....

....

4.4.1 Basic Experimental Situations for Hypothesis Testing

i) It is assumed that the mean, t, is known before treatment. The purpose of the experiment is to determine whether or not the treatment has an effect on the population mean, e.g. a researcher will like to find out whether increased stimulation of infants has an effect on their weight. It is known from national statistics that the mean weight, t, of 2-year old children is 13 kg. The distribution is normal with the standard deviation, á =2 kg.

ii) To test the truth of the claim a researcher may take 16 new born infants and give their parents detailed instructions for giving these infants increased handling and stimulations. At age 2, each of the 16 children will be weighed and the mean weight for the sample will be computed.

iii) The researcher may conclude that the increased handling and stimulation had an effect on the weight of the children if there is a substantial difference in the weights from the population mean.

A hypothesis test is typically used in the context of a research study, i.e. a researcher completes one round of a field investigation and then uses a hypothesis test to evaluate the results. Depending on the type of research and the type of data, the details will differ from one research situation to another.

The probability of making a Type-I error is denoted by á, and the probability of making a Type-II error is denoted by â.

In using the hypothesis-testing procedure to determine if the null hypothesis should be rejected, the person conducting the hypothesis test specifies the maximum allowable probability of making a Type-I error, called the level of significance for the test.

Common choices for the level of significance are á = 0.05 and á = 0.01. Although most applications of hypothesis testing control the probability of making a Type I error, they do not always control the probability of making a Type-Il error.

A concept known as the p-value provides a convenient basis for drawing conclusions in hypothesis-testing applications. The p-value is a measure of how likely the sample results are, assuming that the null hypothesis is true. The smaller the p-value, the lesser likely are the sample results reliable. If the p-value is less than á, the null hypothesis can be rejected, otherwise, the null hypothesis cannot be rejected. The p-value is often called the observed level of significance for the test.

A hypothesis test can be performed on parameters of one or more populations as well as in a variety of other situations. In each instance, the process begins with the formulation of null and alternative hypotheses about the population. In addition to the population mean, hypothesis-testing procedures are available for population parameters such as proportions, variances, standard deviations, and medians.

Hypothesis tests are also conducted in regression and correlation analysis to determine if the regression relationship and the correlation coefficient are statistically significant.

A goodness-of-fit test refers to a hypothesis test in which the null hypothesis is that the population has a specific probability distribution, such as a normal probability distribution. Non-parametric statistical methods also involve a variety of hypothesis testing procedures.

For example, if it is assumed that the mean of the weights of the population of a college is 55 kg, then the null hypothesis will be: the mean of the population is 55 kg, i.e. Ho: = 55 kg (Null hypothesis). In terms of alternative hypothesis (i) H_1: > 55 kg, (ii) H_1: <55 kg.

Now fixing the limits totally depends upon the accuracy desired. Generally the limits are fixed such that the probability that the difference will exceed the limits is 0.05 or 0.01. These levels are known as the ‘levels of significance’ and are expressed as 5% or 1% levels of significance.

What does this actually mean? When we say the limits not to exceed .05 level, that means what ever result we get we can say with 95% confidence that these are genuine results and not because of any chance factor. If we say .01 level, then it means that we can say with 99% confidence that the obtained results are genuine and not due to any chance factor.

Rejection of null hypothesis does not mean that the hypothesis is disproved. It simply means that the sample value does not support the hypothesis. Also, acceptance does not mean that the hypothesis is proved. It means simply it is being supported.

Self Assessment Questions

1) What is a goodness of fit test?

....

....

2) How do we fix the limits for significance in hyopothesis testing?

....

....

3) What are the basis experimental situations for hypothesis testing?

....

....

4.5 CONFIDENCE LIMITS

The limits (or range) within which the hypothesis should lie with specified probabilities are called the confidence limits or fiduciary limits. It is customary to take these limits as 5% or 1% levels of significance. If sample value lies between the confidence limits, the hypothesis is accepted; if it does not, the hypothesis is rejected at the specified level of significance.

4.5.1 Meaning and Concept of Levels of Significance

Experimenters and researchers have selected some arbitrary standards—called levels of significance to serve as the cut-off points or critical points along the probability scale, so as to separate the significant difference from the non significant difference between the two statistics, like means or SD's.

Generally, the .05 and the .01 levels of significance are the most popular in social sciences research. The confidence with which an experimenter rejects—or retains—a null hypothesis depends upon the level of significance adopted. These may, hence, sometime be termed as levels of confidence. Their meanings may be clear from the following:

Meaning of Levels of Confidence

Level	Amount of confidence	Interpretation
.05	95%	If the experiment is repeated a 100 times, only on five occasions the obtained mean will fall outside the limited $\mu \pm 1.96$ SE
.01	99%	If the experiment is repeated a 100 times, only on one occasions the obtained mean will fall outside the limited $\mu \pm 2.58$ SE

The values 1.96 and 2.58 have been taken from the t tables keeping large samples in view. The .01 level is more rigorous and higher a standard as compared to the .05 level and would require a larger value of the critical ratio for the rejection of the

Ho. Hence if an obtained value of t is significant at 01 level, it is automatically significant at .05 level but the reverse is not always true.

4.5.2 Application and Interpretation of Standard Error of the Mean (SEM) in Small Samples

The procedure of calculation and interpretation of Standard Error of Mean in small samples differs from that for large samples, in two respects.

1) The denominator N-1 instead of N is used in the formula for calculation of the SD of the sample.

2) The appropriate distribution to be used for small samples is t distribution instead of normal distribution.

The rest of the line of reasoning used in determining and interpreting SE in small samples is similar to that for the large samples.

4.5.3 The Standard Error of a Median, σ_{Mdn}

It has been established that the variability of the sample medians is about 25 per cent greater than the variability of means in a normally distributed population. Hence the standard error of a median can be estimated by using the formulas:

$$SE_{Mdn,}\ \sigma_{Mdn} = \frac{1.253\sigma}{\sqrt{N}}$$

$$\sigma_{Mdn} = \frac{1.253\sigma}{\sqrt{N}}$$

(Standard Error of the Median in terms of ó and Q)

Self Assessment Questions

1) Describe confidence limits.

....

....

2) Elucidate the concept of significance level.

....

....

3) What is standard error of the mean? How is it useful in hypothesis testing?

....

....

4) What is standard error of median? How is it calculated ? What is its significance?

....

....

4.6 SETTING UP THE LEVEL OF CONFIDENCE OR SIGNIFICANCE

The experimenter has to take a decision about the level of confidence or significance at which the hypothesis is going to be tested. At times the researcher may decide to use 0:05 or 5% level of significance for rejecting a null hypothesis (when a hypothesis is rejected at the 5% level it is said that the chances are 95 out of 100, that the hypothesis is not true and only 5 chances out of 100 that it is true). At other times, the researcher may prefer to make it more rigid and therefore, use the 0.01 or 1% level of significance. If a hypothesis is rejected at this level, the chances are 99 out of 100, that the hypothesis is not true and that only 1 chance out of 100 is true. This level on which we reject the null hypothesis, is established before doing the actual experiment (before collecting data). Later we have to adhere to it.

4.6.1 Size of the Sample

The sampling distribution of the differences between means may look like a normal curve or *t* distribution curve depending upon the size of the samples drawn from the population. The t distribution is a theoretical probability distribution. It is symmetrical, bell-shaped, and similar to the standard normal curve. It differs from the standard normal curve, however, in that it has an additional parameter, called degrees of freedom, which changes its shape.

If the samples are large (N = 30 or greater than 30), then the distribution of differences between means will be a normal one. If it is small (N is less than 30), then the distribution will take the form of a I distribution and the shape of the t-curve will vary with the number of degrees of freedom.

In this way, for large samples, statistics advocating normal distribution of the characteristics in the given population will be employed, while for small samples, the small sample statistics will be used.

Hence in the case of large samples possessing a normal distribution of the differences of means, the value of standard error used to determine the significance of the difference between means will be in terms of standard sigma (z) scores. On the other hand, in the case of small samples possessing a t- distribution of differences between means, we will make use of t values rather than z scores of the normal curve. From the normal curve table we see that 95% and 99% cases lie at the distance of l.96 and 2.58 . Therefore, the sigma or z scores of 1.96 and 2.58 are taken as critical values for rejecting a null hypothesis.

If a computed z value of the standard error of the differences between means approaches or exceeds the values 1.96 and 2.58, then we may safely reject a null hypothesis at the 0.05 and 0.01 levels.

To test the null hypothesis in the case of small sample means, we first compute the t ratio in the same manner as z scores in case of large samples. Then we enter the table of t distribution (Table C in the Appendix) with $N_1 + N_2$ — 2 degrees of freedom and read the values of t given against the row of $N_1 + N_2$ — 2 degrees of freedom and columns headed by 0.05 and 0.01 levels of significance. If our computed t ratio approaches or exceeds the values of t read from the table, we will reject the established null hypothesis at the 0.05 and 0.01 levels of significance, respectively.

4.6.2 Two-Tailed and One-Tailed Tests of Significance

Two-tailed test. In making use of the two-tailed test for determining the significance of the difference between two means, we should know whether or not such a difference between two means really exists and how trustworthy and dependable this difference is.

In all such cases, we merely try to find out if there is a significant difference between two sample means; whether the first mean is larger or smaller than the second, is of no concern. We do not care for the direction of such a difference, whether positive or negative. All that we are interested in, is a difference.

Consequently, when an experimenter wishes to test the null hypothesis, Ho: $M_1 = 0$, against its possible rejection and finds that it is rejected, then the researcher may conclude that a difference really exists between the two means, however no assertion is made about the direction of the difference. Such a test is a non-directional test.

It is also named as *two-tailed test*, because it employs both sides, positive and negative, of the distribution (normal or t distribution) in the estimation of probabilities. Let us consider the probability at 5% significance level in a two-tailed test with the help of figure given below:

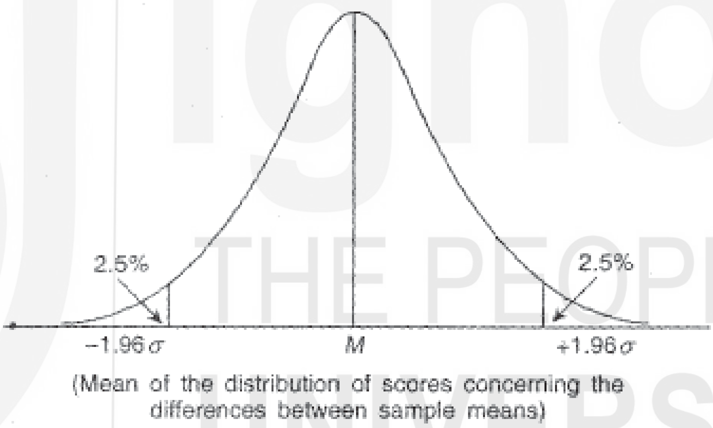

Fig.: Two-tailed test at the 5% level

(Mean of the distribution of scores concerning the differences between sample means)

Two-tailed test at the 5% level.

Therefore, while using both the tails of the distribution we can say that area of the normal curve falls to the right of 1.96 standard deviation units above the mean and 2.5% falls to the left of 1.96 standard deviation units below the mean.

The area outside these limits is 5% of the total area under the curve. In this way, for testing the significance at the 5% level, we may reject a null hypothesis if the computed error of the difference between means reaches or exceeds the yardstick 1.96.

Similarly, we may find that a value of 2.58 is required to test the significance at the 1% level in the case of a two-tailed test.

4.6.3 One-tailed Test

As we have seen, a two-tailed or a non-directional test is appropriate, if we are only concerned with the absolute magnitude of the difference, that is, with the difference regardless of sign.

However, in many experiments, our primary concern may be with the direction of the difference rather than with its existence in absolute terms. For example, if we plan an experiment to study the effect of coaching work on computational skill in mathematics, we take two groups—the experimental group, which is provided an extra one hour coaching work in mathematics, and the control group, which is not provided with such a drill. Here, we have reason to believe that the experimental group will score higher on the mathematical computation ability test which is given at the end of the session.

In our experiment we are interested in finding out the gain in the acquisition of mathematical computation skill (we are not interested in the loss, as it seldom happens that coaching will decrease the level of computation skill).

In cases like these, we make use of the one-tailed or directional test, rather than the two-tailed or non-directional test to test the significance of difference between means.

Consequently, the meaning of null hypothesis, restricted to an hypothesis of no difference with two-tailed test, will be somewhat extended in a one-tailed test to include the direction-positive or negative-of the difference between means.

Self Assessment Questions

1) How is size of sample important in setting up level of confidence?

....

....

2) What is one tailed tests of significance? Explain with examples.

....

....

3) What is a two tailed test? When is it useful? Give suitable examples.

....

....

4.7 STEPS IN SETTING UP THE LEVEL OF SIGNIFICANCE

i) State the null hypothesis and the alternative hypothesis. (Note: The goal of inferential statistics is to make general statements about the population by using sample data. Therefore in testing hypothesis, we make our predictions about the population parameters).

ii) Set the criteria for a decision.

iii) Level of significance or alpha level for the hypothesis test: This is represented by μ which is the probability used to define the very unlikely sample outcomes, if the null hypothesis is true.

In hypothesis testing, the set of potential samples is divided into those that are likely to be obtained and those that are very unlikely if the hypothesis is true.

iv) Critical Region: This is the region which is composed of extreme sample values that are very unlikely outcomes if the null hypothesis is true. The boundaries for the critical region are determined by the alpha level. If sample data fall in the critical region, the null hypothesis is rejected. The cut off level that is set affects the outcome of the research.

v) Collect data and compute sample statistics using the formula given below

$$z = \frac{\overline{x} - \mu}{\sigma_x}$$

where, $\overline{x}$ = sample mean,

μ = hypothesised population mean, and

$ó_x$ = standard error between $\overline{x}$ and μ.

vi) Make a decision and write down the decision rule.

Z-Score is called a test statistics. The purpose of a test statistics is to determine whether the result of a research study (the obtained difference) is more than what would be expected by the chance alone.

$$z = \frac{\text{Obtained difference}}{\text{Difference due to chance}}$$

Now suppose a manufacturer, produces some type of articles of good quality. A purchaser by chance selects a sample randomly. It so happens that the sample contains many defective articles and it leads the purchaser to reject the whole product. Now, the manufacturer suffers a loss even though he has produced a good article of quality. Therefore, this Type-I error is called "*producers risk*".

On the other hand, if we accept the entire lot on the basis of a sample and the lot is not really good, the consumers are put to loss. Therefore, this Type-II error is called the "*consumers risk*".

In practical situations, still other aspects are considered while accepting or rejecting a lot. The risks involved for both producer and consumer are compared. Then Type-I andType-II errors are fixed; and a decision is reached.

In summary, the following procedure is recommended for formulating hypotheses and stating conclusions.

4.7.1 Formulating Hypothesis and Stating Conclusions

i) State the hypothesis as the alternative hypothesis H_1.

ii) The null hypothesis, Ho, will be the opposite of H_1 and will contain an equality sign.

iii) If the sample evidence supports the alternative hypothesis, the null hypothesis will be rejected and the probability of having made an incorrect decision (when in fact Ho is true) is α, a quantity that can be manipulated to be as small as the researcher wishes.

iv) If the sample does not provide sufficient evidence to support the alternative hypothesis, then conclude that the null hypothesis cannot be rejected on the basis of your sample. In this situation, you may wish to collect more information about the phenomenon under study.

An example is given below:

Example

The logic used in hypothesis testing has often been likened to that used in the courtroom in which a defendant is on trial for committing a crime.

i) Formulate appropriate null and alternative hypotheses for judging the guilt or innocence of the defendant.

ii) Interpret the Type-I and Type-II errors in this context.

iii) If you were the defendant, would you want á to be small or large? Explain.

Solution

i) Under a judicial system, a defendant is "innocent until proven guilty". That is, the burden of proof is not on the defendant to prove his or her innocence; rather, the court must collect sufficient evidence to support the claim that the defendant is guilty. Thus, the null and alternative hypotheses would be

H_0 : Defendant is innocent

H_1: Defendant is guilty

ii) The four possible outcomes are shown in the table below. A Type-I error would be to conclude that the defendant is guilty, when in fact he or she is innocent; a Type-II error would be to conclude that the defendant is innocent, when in fact he or she is guilty.

Table : Conclusions and Consequences

		Decision of Court	
		Defendant is Innocent	Defendant is Guilty
True State of Nature	Defendant if Innocent Defendant is Guilty	*Correct decision* Type-I error	*Type-II error* Correct decision

iii) Most people would probably agree that the Type-I error in this situation is by far the more serious. Thus, we would want α, the probability of committing a Type-I error, to be very small indeed.

A convention that is generally observed when formulating the null and alternative hypotheses of any statistical test is to state Ho so that the possible error of incorrectly rejecting Ho (Type-I error) is considered more serious than the possible error of incorrectly failing to reject Ho (Type-II error).

In many cases, the decision as to which type of error is more serious is admittedly not as clear-cut though experience will help to minimize this potential difficulty.

4.7.2 Types of Errors for a Hypothesis Test

The goal of any hypothesis testing is to make a decision. In particular, we will decide whether to reject the null hypothesis, Ho, in favour of the alternative hypothesis, H_1. Although we would like always to be able to make a correct decision, we must remember that the decision will be based on sample information, and thus we are subject to make one of two types of error, as defined in table below.

The null hypothesis can be either true or false. Further, we will make a conclusion either to reject or not to reject the null hypothesis. Thus, there are four possible situations that may arise in testing a hypothesis as shown in table .

Table: Conclusions and Consequences for Testing a Hypothesis

		Decision of Court	
		Defendant is Innocent	Defendant is Guilty
True "State of Nature"	Null Hypothesis	Correct Conclusion	*Type-I error*
	Alternative Hypothesis	*Type-II error*	Correct Conclusion

The kind of error that can be made depends on the actual state of affairs (which, of course, is unknown to the investigator). Note that we risk a Type-I error only if the null hypothesis is rejected, and we risk a Type-II error only if the null hypothesis is not rejected.

Thus, we may make no error, or we may make either a Type-I error (with probability a), or a Type-II error (with probability 3), but not both. We don't know which type of error corresponds to actuality and so would like to keep the probabilities of both types of errors small.

Remember that as α increases, β decreases, similarly, as β increases, α decreases. The only way to reduce α and β simultaneously is to increase the amount of information available in the sample, i.e. to increase the sample size.

You may note that we have carefully avoided stating a decision in terms of "accept the null hypothesis H_0". Instead, if the sample does not provide enough evidence to support the alternative hypothesis H_1 we prefer a decision "not to reject H_0".

This is because, if we were to "accept H_0", the reliability of the conclusion would be measured by α, the probability of Type-II error. However, the value of β is not constant, but depends on the specific alternative value of the parameter and is difficult to compute in most testing situations.

Self Assessment Questions

1) Elucidate the steps in setting up the level of significance.

....

....

2) How do you formulate hypothesis and state the conclusions?

....

....

3) Explain the concept if α increases β decreases. If β increases α decreases.

....

....

4) Why β is more important than β? Explain.

....

....

4.8 LET US SUM UP

In this unit, we pointed out how drawing conclusions about a population on the basis of sample information is called statistical inference. Here we have basically two things to do: statistical estimation and hypothesis testing.

An estimate of an unknown parameter could be either a point or an interval. Sample mean is usually taken as a point estimate of population mean. On the other hand, in interval estimation we construct two limits (upper and lower) around the sample mean. We can say with stipulated level of confidence that the population mean, which we do not know; is likely to remain within the confidence interval.

We learnt about confidence interval and how to set the same. In order to construct confidence interval we need to know the population variance or its estimate. When we know population variance, we apply normal distribution to construct the confidence interval. In cases where population variance is not known, we use t distribution for the above purpose.

Remember that when sample size is large (n>30) t-distribution approximates normal distribution. Thus for large samples, even if population variance is not known, we can use normal distribution for estimation of confidence interval on the basis of sample mean and sample variance.

Subsequently we discussed the methods of testing a hypothesis and drawing conclusions about the population. Hypothesis is a simple statement (assertion or claim) about the va1ue assumed by the parameter. We test a hypothesis on the basis of sample information available to us. In this Unit we considered two situations: i) description of a single sample, and ii) comparison between two samples.

In the case of qualitative data we pointed out how we cannot have parametric values and hypothesis testing on the basis of z statistic or t-statistic cannot be performed.

4.9 UNIT END QUESTIONS

1) What do you mean by a null hypothesis?

2) What is significance of size of sample is hypothesis testing?

3) Write down two levels of significance which are mainly used in hypothesis testing?

4) Write down a short note on level of significance.

4.10 GLOSSARY

Contingency Table	:	A two-way table to present bivariate data. It is called contingency table because we try to find whether one variable is contingent upon the other variable.
Degrees of Freedom	:	It refers to the number of pieces of independent information that are required to compute some characteristic of a given set of observations.
Estimation	:	It is the method of prediction about parameter values on the basis of sample statistics.

Expected Frequency : It is the expected cell frequency under the assumption that both the variables are independent.

Nominal Variable : Such a variable takes qualitative values and do not have any ordering relationships among them. For example, gender is a nominal variable taking only the qualitative values, male and female; there is no ordering in 'male' and 'female' status. A nominal variable is also called an attribute.

Parameter : It is a measure of some characteristic of the population.

Population : It is the entire collection of units of a specified type in a given place and at a particular point of time.

Random Sampling : It is a procedure where every member of the population has a definite chance or probability of being selected in the sample. It is also called probability sampling. Random sampling could be of many types: simple random sampling, systematic random sampling and stratified random sampling.

Sample : It is a sub-set of the population. It can be drawn from the population in a scientific manner by applying the rules of probability so that personal bias is eliminated. Many samples can be drawn from a population and there are many methods of drawing a sample.

Sampling Distribution : It is the relative frequency or probability distribution of the values of a statistic when the number of samples tends to infinity.

Sampling Error : In the sampling method, we try to approximate some feature of a given population from a sample drawn from it. Now, since in the sample all the members of the population are not included, howsoever close the approximation is, it is not identical to the required population feature and some error is committed. This error is called the sampling error.

Significance Level : There may be certain samples where population mean would not remain within the confidence interval around sample mean. The percentage (probability) of such cases is called significance level. It is usually denoted by.

4.11 SUGGESTED READINGS

Asthana H.S, and Bhushan. B. (2007) *Statistics for Social Sciences* (with SPSS Applications).

B.L. Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Guilford, J.P. (1965); *Fundamental Statistics in Psychology and Education*. New York: McGraw Hill Book Company.

Gupta, S.C. (1990) *Fundamentals of Statistics*.

Siegel, S. (1956): *Non-parametric Statistics for Behavioural Sciences*. Tokyo: McGraw Hill Hoga Kunsa Ltd.