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:

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

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

(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:

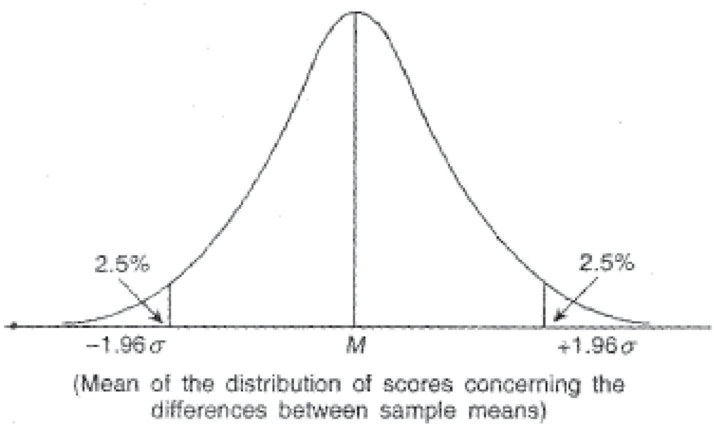

Fig.: Two-tailed test at the 5% level

(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-Il 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-Il 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.