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 ì$_1 \neq$ ì$_2$ the null hypothesis is ì$_1 =$ ì$_2$

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.