UNIT 3 OVERVIEW OF HYPOTHESIS TESTING*

Structure

3.0 OBJECTIVES

After going through this unit, you will be able to

- explain the concept and significance of hypothesis testing;
- describe the applications of a test statistic;
- explain the procedure of testing of hypothesis of population parameters;
- distinguish between the Type I and Type II errors; and
- apply the tests for comparing parameters from two different samples.

3.1 INTRODUCTION

The purpose behind statistical inference is to use the sample to make judgement about the population parameters. The concept of hypothesis testing is crucial for predicting the value of population parameters using the sample. Various test statistics are used to test hypotheses related to population mean and variance. The variance of two different samples can also be compared using hypothesis testing. There are two approaches to testing of hypothesis: (i) test of significance

* Dr. Pooja Sharma, Assistant Professor, Daulat Ram College, University of Delhi

approach, and (ii) confidence interval approach. While testing a hypothesis, there is a likelihood of committing two types of errors: (i) type I error, and (ii) type II error. In this unit we will elaborate on the process of hypothesis testing, and explain the method of rejecting the null hypothesis on the basis of appropriate test statistic.

3.2 PROCEDURE OF HYPOTHESIS TESTING

We formulate a hypothesis on the basis of economic theory or logic. A hypothesis is a tentative statement about certain characteristic of a population. As you know, a population is described by its parameters (such as mean, standard deviation, etc.). Thus a hypothesis is an assumption about a population parameter. A hypothesis may or may not be true. For finding out that we test a hypothesis by certain econometric method.

Formulation of a hypothesis involves a prior judgement or expectation about what value a particular parameter may assume. For example, prior knowledge or an expert opinion tells us that the true average price to earnings (P/E) ratio in the local stock exchange is 20. Thus our hypothesis is that the P/E ratio is equal to 20.

In order to test this hypothesis, suppose we collect a random sample of stocks and find that the average P/E ratio is 23. Is the figure 23 statistically different from 20? Because of sampling variation there is likely to be a difference between a sample estimate and its population value. It is possible that statistically the number 23 may not be very different from the number 20. If this is the case, then we should not reject the hypothesis that the average P/E ratio is 20.

In hypothesis testing there are four important components: i) null hypothesis, ii) alternative hypothesis, iii) test statistic, and iv) interpretation of results. We elaborate on these components below.

(i) Formulation of null and alternative hypotheses: There are two types of hypothesis, viz., null hypothesis and alternative hypothesis. A 'null hypothesis' is the statement that we consider to be true about the population. It is called 'null' thereby meaning empty or void. For example, a null hypothesis could be: there is no relationship between employment and education. Therefore, if we carry out a regression of employment on education, the regression coefficient should be zero. Usually we denote null hypothesis by H_0. The alternative hypothesis is the opposite of the null hypothesis. Alternative hypothesis is usually denoted by H_1. You should note that H_0 and H_1 are 'mutually exclusive'; they cannot occur simultaneously.

(i) Identification of the test statistic: The null hypothesis is put to test by a test statistic. There are several test statistics (such as t, F, chi-square, etc.) available in econometrics. We have to identify the appropriate test statistic.

(ii) Interpretation of the results based on the value of the test statistic: After carrying out the test, we interpret the results. When we apply the test statistic to the sample data that we have, we obtain certain value of the test statistic (for example, t-ratio of 2.535). Interpretation of results involves comparison of two values: tabulated value of the test statistics and the computed value. If the computed value exceeds the tabulated value we reject the null hypothesis.

The sampling distribution of a test statistic under the null hypothesis is called the 'null distribution'. When the data depicts strong evidence against the null hypothesis, the value of test statistic becomes very large. By observing the computed value of the test statistic we draw inferences. Apart from the test statistic econometric software provides a *p-value*. The p-value indicates the probability of the null hypothesis being true. Thus, if we obtain a p-value of 0.04, it says the probability of the null hypothesis being true is 0.04 or 4 per cent. Therefore, if we take 5 per cent level of significance, we reject the null hypothesis.

3.3 ESTIMATION METHODS

In Unit 2 we described about two concepts; point estimation and interval estimation. We also discussed about certain probability distribution functions such as normal, *t*, *F* and chi-square.

There are basically three estimation methods: (i) least squares, (ii) maximum likelihood, and (iii) method of moments. We will use the least squares estimation method extensively in this course. In Unit 7 of this course we have introduced the maximum likelihood method. You are not introduced to 'Method of Moments' in this course.

In Unit 5 of the course BECC 107 we discussed with the concept of regression. In Section 5.9 that Unit we mentioned that the error variable in the regression should be minimised. For that purpose, we minimised the sum of squares of the error terms ($\sum u_i^2$). Now you can guess why it is called the least squares method. In this course we confine to ordinary least squares (OLS) method. We deal with OLS method first with the two-variable case. Subsequently, we extend it to more than two variables. This leads us the multiple regression model.

The name ordinary least squares (OLS) suggests that it is the simplest of the least squares methods. It implies that further complexities can be brought into the OLS method. Correctly so; there are generalised least squares (GLS), two-stage least squares (2SLS), three-stage least squares (3SLS), etc. Therefore, be careful when you read about the least squares method – notice which method the text is referring to. When you come across the term GLS in some context do not confuse it with OLS – both methods are different. In both OLS and GLS the sum of squares of the error terms is minimised (that is why both are referred to as least squares method) but there is some transformation of the regression model in the case of GLS. The advanced methods of least squares are not dealt with in this course. Remember that for carrying out the least squares method you do not need to assume any probability distribution function about the variables.

The maximum likelihood (ML) method assumes a probability distribution about the variables. Normal distribution is the most commonly used probability distribution function in maximum likelihood estimation. In ML method we form a likelihood function, which is derived from the probability distribution function. Note that in econometrics we are given the data – the data is obtained from a sample survey. We estimate the parameters of the regression model, under that the assumption that the data follows certain probability distribution function (for example, normal distribution). The likelihood function can follow any of the probability distribution functions; not just normal distribution. Recall from your statistics course that in probability distribution function we are given the parameters and we find out the probability of occurrence of particular dataset. In ML method, we do the opposite – we are provided with the data, and we are estimating the parameters.

The method of moments (MOM) makes use of the moment generating function (MGF) properties. You have been introduced to the concept of 'moments' in Unit 4 of BECC 107. The moment generating function of certain probability distributions are used for estimation of the parameters. The method of moments is quite advanced and beyond the scope of this course.

3.4 REJECTION REGION AND TYPES OF ERRORS

In the previous Unit we discussed about point estimation and interval estimation. The underlying idea behind hypothesis testing and interval estimation is the same. Recall that a confidence interval is built around sample mean with certain confidence level. A confidence level of 95 per cent implies that in 95 per cent cases the population mean would remain in the confidence interval estimated from the sample mean. It is implicit that in 5 per cent cases the population mean will not remain within the confidence interval. Note that when the population mean does not remain within the confidence interval our test statistic should reject the null hypothesis.

3.4.1 Rejection Region for Large Samples

Let us explain the concept of critical region. Sampling distribution of sample mean ($\bar{x}$) follows normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

The standard deviation of a sampling distribution is known as 'standard error'.

Thus, $\bar{x}$ can be transformed into a standard normal variable, z, so that it follows normal distribution with mean 0 and standard deviation 1.

In notations, $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ and $z \sim N(0,1)$.

Recall that area under the standard normal curve gives the probability for different range of values assumed by *z*. These probabilities are presented as the area under standard normal curve.

Let us explain the concept of critical region or rejection region through the standard normal curve given in Fig. 14.1 below. When we have a confidence coefficient of 95 percent, the area covered under the standard normal curve is 95 per cent. Thus 95 per cent area under the curve is bounded by $-1.96 \le z \le 1.96$. The remaining 5 per cent area is covered by $z \le -1.96$ and $z \ge 1.96$. Thus 2.5 per cent of area on both sides of the standard normal curve constitute the rejection region. This area is shown in Fig. 3.1. If the sample mean falls in the rejection region we reject the null hypothesis.

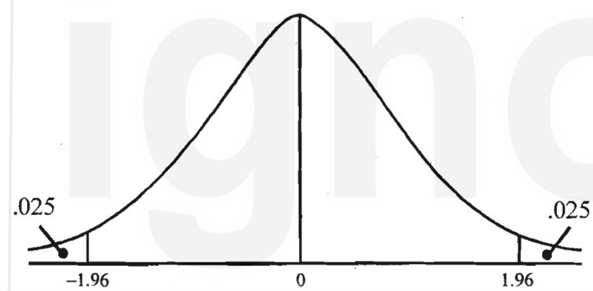

Fig. 3.1: Critical Regions

3.4.2 One-tail and Two-tail Tests

In Fig. 3.1 we have shown the rejection region on both sides of the standard normal curve. However, in many cases we may place the rejection region on one side (either left or right) of the standard normal curve. Remember that if α is the level of significance, then for a two-tail test $\frac{\alpha}{2}$ area is placed on both sides of the standard normal curve. But if it is a one-tail test, then α area is placed on one-side of the standard normal curve. Thus the critical value for one-tail and two tail test differ.

The selection of one-tail or two-tail test depends upon the formulation of the alternative hypothesis. When the alternative hypothesis is of the type $H_A : \bar{x} \ne \mu$ we have a two-tail test, because $\bar{x}$ could be either greater than or less than μ. On the other hand, if alternative hypothesis is of the type $H_A : \bar{x} < \mu$, then entire rejection is on the left hand side of the standard normal curve. Similarly, if the alternative hypothesis is of the type $H_A : \bar{x} > \mu$, then the entire rejection is on the right hand side of the standard normal curve.

The critical values for z depend upon the level of significance. In the appendix tables at the end of this book Table 14.1 these critical values for certain specified levels of significance (α) are given.

3.4.3 Rejection Region for Small Samples

In the case of small samples ($n \leq 30$), if population standard deviation is known we apply z-statistic for hypothesis testing. On the other hand, if population standard deviation is not known we apply t-statistic. The same criteria apply to hypothesis testing also.

In the case of small samples if population standard deviation is known the test statistic is

$$z = \frac{|\bar{x} - \mu|}{\sigma/\sqrt{n}} \qquad \ldots(3.1)$$

On the other hand, if population standard deviation is not known the test statistic is

$$t = \frac{|\bar{x} - \mu|}{s/\sqrt{n}} \qquad \ldots(3.2)$$

In the case of t-distribution, however, the area under the curve (which implies probability) changes according to degrees of freedom. Thus while finding the critical value of t we should take into account the degrees of freedom. You should remember two things while finding critical value of t. These are: i) level of significance, and ii) degrees of freedom.

3.5 TYPES OF ERRORS

In hypothesis testing we reject or do not reject a hypothesis with certain degree of confidence. As you know, a confidence coefficient of 0.95 implies that in 95 out of 100 samples the parameter remains within the acceptance region and in 5 per cent cases the parameter remains in the rejection region. Thus in 5 per cent cases the sample is drawn from the population but sample mean is too far away from the population mean. In such cases the sample belongs to the population but our test procedure rejects it. Obviously we commit an error such that H_0 is true but gets rejected. This is called 'Type I error'. Similarly there could be situations when the H_0 is not true, but on the basis of sample information we do not reject it. Such an error in decision making is termed 'Type II error' (see Table 3.1).

Note that Type I error specifies how much error we are in a position to tolerate. Type I error is equal to the level of significance, and is denoted by α. Remember that confidence coefficient is equal to $1-\alpha$.

The probability of committing a type I error is designated as α and is called the level of significance. The probability of committing type II error is called β. Thus, Type I error = α = prob (rejecting H_0 | H_0 is true)

Type II error = β = prob (accepting H_0 | H_0 is false)

Table 3.1: Type of Errors

	H_0 **true**	H_0 **not true**
Reject H_0	Type I Error	Correct decision
Do not reject H_0	Correct decision	Type II Error

Check Your Progress 1

1) Distinguish between one-tail and two-tail tests.

....

....

....

....

....

....

2) Distinguish between Type I and Type II errors.

....

....

....

....

....

....

3) Suppose the cholesterol level of an individual is normally distributed with mean of 180 and standard deviation of 20. Cholesterol level of over 225 is diagnosed as not healthy.

a) What is the probability of making type I error?

b) What level should people be diagnosed as not healthy if we want the probability of type I error to be 2%?

....

....

....

....

....

3.6 POWER OF TEST

As pointed out above, there are types I and type II errors in hypothesis testing. Thus, there are two types of risks: (i) α represents the probability that the null hypothesis is rejected when it is true and should not be rejected. (ii) β represents the probability that null hypothesis is not rejected when in reality it is false. The power of test is referred to as $(1-\beta)$, that is the complement of β. It is basically the probability of not committing a type II error.

A 95% confidence coefficient means that we are prepared to accept at most 5% probability of committing type I error. We do not want to reject a true hypothesis by more than 5 out of 100 times. This is called 5% level of significance.

The power of test depends on the extent of difference between the actual population mean and the hypothesized mean. If the difference is large then the power of test will be much greater than if the difference is small. Therefore, selection of level of significance α is very crucial. Selecting large value of α makes it easier to reject the null hypothesis thereby increasing the power of the test $(1-\beta)$.

At the same time increasing the sample size increases the precision in the estimates and increases the ability to detect the difference between the population parameter and sample, increasing the power of the test.

3.7 APPROACHES TO PARAMETER ESTIMATION

In statistical hypothesis testing, estimation theory deals with estimating the values of parameters based on measurement of empirical data that has a random component. The method of estimation requires setting up of a null hypothesis and a corresponding alternative hypothesis, which are further rejected or not rejected based on the two approaches used to make decision regarding the null hypothesis. The two methods have been described in the following section.

3.7.1 Test of Significance Approach

Any test statistic can be used for the test of significance approach to hypothesis testing. Let us consider the t-statistic.

$$t = \frac{\bar{X}-\mu_X}{\frac{\sigma_X}{\sqrt{n}}} \quad \dots (3.3)$$

If the difference between $\bar{X}$ and μ_X is small, $|t|$ value will also be small, where $|t|$ is the absolute value of t-statistic. You should note that $t = 0$, if $\bar{X} = \mu_X$. In this case we do not reject the null hypothesis. As | t | gets larger, we would be more inclined to reject the null hypothesis.

Example. Suppose for a dataset $\bar{X} = 23.25, S_X = 4.49,$ and $n = 28$. Our nulland alternative hypothesis are

H_0: $\mu_X = 18.5$ and H_A: $\mu_X \neq 18.5$

$$\therefore \quad t = \frac{23.25 - 18.5}{9.49/\sqrt{28}} = 2.6486 \qquad \ldots (3.4)$$

We need to specify α, the probability of rejecting the null hypothesis (probability of commuting type I error). Let us fix α at 5%.

$H_0: \mu_X = 18.5$

$H_A: \mu_X \neq 18.5$ (two-tailed test)

Since the computed *t* value is 2.6486. This value lies in the right-tail critical region of the *t*-distribution. We therefore reject the null hypothesis (H_0) that the true population mean is 18.5.

A test is statistically significant means that we one can reject the null hypothesis. This implies that the probability of observed difference between the sample value and the critical value (also called tabulated value) is not small and is not due to chance.

A test is statistically not significant means that we do not reject the null hypothesis. The difference between the sample value and the critical value could be due to sampling variation or due to chance mechanism.

3.7.2 Confidence Interval Approach

Let us assume that the level of significance or the probability of commuting type I error is fixed at α = 5%. Suppose the alternative hypothesis is two-sided. Assume that we apply *t*-distribution since variance is not known. From the *t* table we find the critical value of *t* at 8 degree of freedom $(n-K)=(10-2)$ at α = 5%. We find out the value to be 2.360. Thus we construct the confidence interval

$$P(-2.360 \leq t \leq 2.306) = 0.95 \qquad \ldots (3.5)$$

The probability that *t* value lies between the limits ($-2.360 \leq t \leq 2.360$) is 0.95 or 95%. The values -2.360 and 2.360 are the critical *t* values.

If we substitute the t from equation (3.2)

$$P\left(-2.306 \leq \frac{b_2 - \beta_2}{\mathrm{SE(b_2)}} \leq 2.306\right) = 0.95 \qquad \ldots (3.6)$$

As we will see in Unit 4, $\mathrm{SE(b_2)}$ is $\frac{\hat{\sigma}}{\sqrt{\Sigma x_i^2}}$

If we substitute the above value in equation (3.6) and re-arrange terms we obtain

$$P\left(b_2 - 2.306\frac{\hat{\sigma}}{\sqrt{\Sigma x_i^2}} \leq \beta_2 \leq b_2 + 2.306\frac{\hat{\sigma}}{\sqrt{\Sigma x_i^2}}\right) = 0.95 \qquad \ldots (3.7)$$

Equation (3.7) provides a 95% confidence interval for the parameter β_2. Such a confidence interval is known as the region of acceptance (H_0). The area outside the confidence interval is known as the rejection region (H_A).

If the confidence interval includes the value of the parameter β_2,we do not reject the hypothesis. But if the parameter lies outside the confidence interval, we reject the null hypothesis.

Check Your Progress 2

1) What is meant by power of a test?

....

....

....

....

....

....

2) Explain how a confidence interval is built.

....

....

....

....

....

....

3.8 LET US SUM UP

This unit elaborated the procedure of statistical inference regarding the population parameters. There are two approaches to hypothesis testing of population parameters: test of significance approach, and confidence interval approach. The unit also pointed out that there are errors involved in testing of hypothesis. While making a decision regarding acceptance or rejection of a hypothesis, two types of error may be committed: type I error, and type II error. Power of a test is the probability of not committing a type II error, i.e., rejecting H_0 when it is false is $(1 - \beta)$.

3.9 ANSWERS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Go through Sub-Section 3.4.2 and answer.

2) We have given the types of errors in table 3.1. You should elaborate on that.

3) a) In order to test this we use z-statistics $z = (X - \mu)/\sigma$, $z = (225 - 180)/20 = 2.25$

b) The area corresponding to the z value of 2.25 is 0.0122, which the probability of making type I error. An area of tail as 2% corresponds to $Z = 2.05$.

$Z = (X - \mu)/\sigma$

$2.05 = (X - \mu)/20$, i.e., $(X - \mu) = 2.05 * 20 = 41$

$X = 41 + 180 = 221$

Check Your Progress 2

1) Go through Section 3.6 and answer.

2) Go through Section 3.7.2 and answer.