UNIT 15

BASIC CONCEPTS OF TESTING OF HYPOTHESIS

Structure

15.1 INTRODUCTION

In our day-to-day life, we see different commercial advertisements on YouTube, Facebook, television, newspapers, magazines, etc. such as

(i) An AC of a certain brand saves up to 50% on electric bills,

(ii) A car of a certain brand gives an average mileage of 25 km/litre,

(iii) A new medicine is more effective for controlling systolic blood pressure than old medicine,

(iv) A detergent of a certain brand produces the cleanest wash, etc.

Also, sometimes, we have research questions such as

(i) The mean starting salary earned by data scientists is greater than that of software engineers,

(ii) The counselling sessions improve the performance of the students of the MSCAST programme,

This course began our study with the basics of statistical inference. Till that, we described how we could prepare the sampling distribution of a statistic and use the concepts of that for estimating the value of an unknown population parameter by using point estimation and interval estimation. Frequently, however, the objective of an investigation is not to estimate a parameter but also to test whether a statement/claim made about the true value of a population parameter is true or false. This approach of statistical inferential is called **testing of hypothesis**.

Unit Writer-Dr. Prabhat Kumar Sangal, School of Sciences, IGNOU, New Delhi

(iii) An online teaching method is more effective than face-to-face, etc.

Now, the question may arise in our mind **“Can such types of claims be verified statistically?**” Fortunately, in most cases, the answer is **“yes”**.

For testing such statements, we may examine every unit of the population and estimate the parameter. If it is equal to the population parameter, then we accept it otherwise reject it. But in most cases, a population is too large or the units or items of the population are destructive in nature or there are limited resources such as manpower, money, etc. therefore, it is not feasible to study all items, objects, or persons of the population. For example, it would not be possible to check each car produced by the company to determine whether it will give a mileage of 25 km/litre. In such cases, an alternative is to take a sample from the population under study and use the information contained in this sample to make the decision whether a claim is true or false.

The technique of testing claims or statements or assumptions about the population parameter(s) on the basis of a sample is known as testing of hypothesis.

In this unit, we will discuss some basic concepts and terminology of hypothesis testing. This unit is divided into 12 sections. Section 15.1 is introductory in nature and discusses the need of hypothesis testing. In Section 15.2, we defined the hypothesis, null and alternative, simple and composite hypotheses. The concept and role of rejection/critical region in testing of hypothesis is described in Section 15.3. In Section 15.4, we explored the types of errors in testing of hypothesis and power of a test whereas the level of significance is explained in Section 15.5. In Section 15.6, we discuss the types of tests in testing of hypothesis. The general procedure for testing a hypothesis is discussed in Section 15.7. In Sections 15.8 and 15.9, we discussed p-value and confidence interval approaches for testing of hypothesis, respectively. The unit ends by providing a summary of what we have discussed in this unit in Section 15.10. The terminal questions and the solution of the SAQs/TQs are given in Sections 15.11 and 15.12, respectively. In the next unit, we shall discuss testing of hypothesis for means.

Expected Learning Outcomes

After studying this unit, you should be able to:

- understand the philosophy and scientific principles underlying hypothesis testing;
- explain why hypothesis testing is important;
- describe the role of sampling in hypothesis testing;
- formulate the null and alternative hypotheses;
- explain the concept of critical region;
- identify Type-I and Type-II errors and discuss how they conflict with each other;
- interpret the level of significance and power of a test; and
- p-value and confidence interval approaches for testing of hypothesis.

Before discussing the procedure of testing of hypothesis, you should learn the basic terms used in hypothesis testing. Let us discuss these one by one in subsequent sections.

15.2 HYPOTHESIS

The terms **hypothesis testing** and **testing a hypothesis** are used interchangeably. The procedure of hypothesis testing starts with a statement, or assumption about a population parameter(s) such as an AC of a certain brand saves up to 50% on electric bills, the average starting salaries earned by data scientists is greater than the software engineers. Such types of statements are referred to as hypothesis. It is derived from a management or research question, or a claim made. If someone may be interested in testing such types of claims or statements as discussed in the previous section, then we come across the problem of hypothesis testing. For example:

(i) An automobile engineer wants to test whether the claim that the average mileage of the car is 25 km/litre is true or false,

(ii) A cigarette manufacturer wants to test whether the variance of the nicotine content of its cigarettes is less than 0.60 mg,

(iii) A research scholar is interested in testing whether the passing ratio of the students of the MSCAST programme is greater than 50%,

(iv) A medicine researcher wants to test whether a new medicine is really more effective for controlling systolic blood pressure than old medicine,

(v) A factory manager wants to test whether the day shift workers are more consistent/efficient than the night shift workers,

(vi) A psychologist wants to test whether the proportion of literate in City A is less than in City B, etc.

In all cases discussed above, the decision maker is interested in testing the claim about the population characteristics such as population mean, proportion, variance, etc. Such claims or statements are postulated in terms of hypothesis. We now formally define a hypothesis as:

In statistics, a hypothesis or statistical hypothesis is a statement or a claim or an assumption about population characteristics/ population parameters such as mean, median, variance, proportion, etc.

For example, a car of a certain brand gives an average mileage of 25 km/litre, the variance of the nicotine content in cigarettes is less than 0.60 mg, etc.

Similarly, in the case of two or more populations, we can define it as

A hypothesis is a comparative statement or a claim or an assumption about the population characteristics/parameters of two or more populations.

For example, the mean starting salary earned by data scientists is greater than that of software engineers, new medicine is more effective for controlling systolic blood pressure than old medicine, etc.

In hypothesis testing problems, first of all, we identify the claim or statement or assumption to be tested which is called a hypothesis. Once the claim has been identified then we try to write it in words and then in symbolical form if possible. As in the above examples:

(i) The automobile engineer may write the claim or postulate the hypothesis as *"the car of the certain brand gives the average mileage of 25 km/litre."* Here, we are concerning the **average** mileage of the car, therefore, if μ represents the average mileage of the cars then our hypothesis becomes $\mu = 25$ km/litre.

(ii) The cigarette manufacturer can present the claim in words such as *"the variation of the nicotine content in the cigarettes is less than 0.60 mg."* Since we are concerned about the **variation** which can be measured in terms of standard deviation (σ), therefore, we can write the hypothesis as $\sigma < 0.60$ mg.

(iii) The research scholar may formulate the hypothesis as *"the passing ratio of the students of the MSCAST programme is greater than 50%."* Here, we are concerning the **proportion**, therefore, if P represents the proportion of the students of the MSCAST programme who have passed, then we can write the hypothesis as $P = 0.50$.

(iv) The medicine researcher may postulate the hypothesis *"the new medicine is more effective for controlling systolic blood pressure than old medicine."* Here, we can say that the medicine is more effective in comparison to others if it reduces the systolic blood pressure more in comparison to the old. So if μ_N and μ_O represent the average blood pressure after taking the new and old medicines, respectively, then the hypothesis becomes $\mu_N < \mu_O$.

A dataset which has less variation in comparison of other is called more **consistent**.

(v) The factory manager postulates the hypothesis *"the day shift workers are more consistent than the night shift workers."* We measure the consistency of the workers using the variation in the number of units produced by the worker, therefore, if σ_1^2 and σ_2^2 represent the variability in the production of the units produced by the day and night shift workers, respectively, then the hypothesis becomes $\sigma_1^2 < \sigma_2^2$.

(vi) The psychologist may postulate the hypothesis *"the proportion of literates in City A is less than in City B."* Here, we are concerning the **proportion** of literates so if P_1 and P_2 represent the proportions of literate people in City A and City B, respectively, then our hypothesis becomes $P_1 < P_2$.

The hypothesis is classified as null and alternative according to its nature. We will discuss these in a subsequent subsection.

15.2.1 Null and Alternative Hypotheses

The concept of the null and alternative hypotheses lies at the heart of the testing of hypothesis. The following example will help you to clarify what is meant by the null and alternate hypotheses and how to decide which statement is considered null and which is an alternative hypothesis.

When a study starts, the researchers must have some idea of what they think the outcome may be. For example, a study of researching a new medication to lower systolic blood pressure. After the proper study, the researchers believe that the new medication lowers systolic blood pressure than the old one. The researchers want to check their beliefs and form a research question and formulate the hypothesis that **“taking the new medicine will lower systolic blood pressure than the old medicine."** The researchers can never prove any statement as there are infinite alternatives as any outcome may have occurred. So, the researchers prepare a compliment/inverse of the hypothesis as they want to test and write both statements/hypotheses (plural of hypothesis) in such a way that they cover all possibilities so that if one is rejected then the second one will not be rejected. Here, the researchers take an alternative as **"taking the new medicine will not lower systolic blood pressure than the old medicine."** Out of these statements, we have to decide which one is taken as the null or alternative hypothesis. Therefore, first, we learn about these two:

<u>Null Hypothesis</u>

The null hypothesis is denoted by H_0. The capital letter H stands for hypothesis, and the subscript zero implies **"no difference."** Therefore,

A null hypothesis is a hypothesis that states that there is no change, no difference, or no relationship, that is, there is nothing new happening, the old theory is still true, or the old standard is correct.

According to Prof. Ronald A. Fisher **“A hypothesis which is tested for possible rejection under the assumption that it is true”** is called null

<u>Alternative Hypothesis</u>

The alternative hypothesis is denoted by H_A or H_1. The capital letter H stands for hypothesis, and the subscript A or 1 implies **"alternate."** A statement that provides an alternative to the null hypothesis is called an alternative hypothesis. In other words,

An alternative hypothesis is a hypothesis that states that there is a change, a difference, or a relationship, that is, a new theory is true or there are new standards.

After understanding the definition of the null and alternative hypotheses, we continue our example and try to understand the null and alternative hypotheses in this case. According to the definition of the null hypothesis, the null hypothesis represents that there is no change, or the old standard is still true, therefore, the second statement (taking the new medicine will not lower systolic blood pressure than the old medicine) shows that there is no change so take it as the null hypothesis and the first statement as the alternative hypothesis, therefore, we can write the null and alternative hypotheses as

H_0: Taking the new medicine will not lower the systolic blood pressure than the old medicine.

H_1: Taking the new medicine will lower the systolic blood pressure than the old medicine.

We state the null and alternative hypotheses in such a way that they cover all possibility of the value of population parameter.

We now try to write the statements in symbolic form. If μ_N and μ_O represent the average systolic blood pressure after taking the new and old medicines, respectively, then we can write the above two statements as

H_0: $\mu_N \geq \mu_O$

H_1: $\mu_N < \mu_O$

Similarly, in the cigarette nicotine example (case (ii)), after the proper quality control management, the cigarette manufacturer believes that the variation of the nicotine content of its cigarettes is less than 0.60 mg then its complement is that the variation of the nicotine content of its cigarettes is still 0.60 mg or perhaps it has greater than 0.60 mg. So, we can write the two statements in symbolic form as

$\sigma < 0.60$ mg and $\sigma \geq 0.60$ mg

Since the null hypothesis represents that there is no change, therefore, we take the null hypothesis as

$H_0: \sigma \geq 0.60$

And the alternative hypothesis represents that there is a change, therefore, we take the alternative hypothesis as

$H_1: \sigma < 0.60$.

Sometimes it becomes confusing to decide which one is a null or alternative hypothesis using the definition of these. In such situations, we can also use the following technique to decide the same:

The statement containing an equality sign is always taken as a null hypothesis. That is, if a hypothesis contains sign = or ≤ or ≥, then we take it as a null hypothesis and if a hypothesis contains only sign > or < then we take it as an alternative hypothesis. It is important to remember that no matter how the problem is stated, the null hypothesis will always contain the equality sign. The equality sign (=) will never appear in the alternate hypothesis. Why? Because the null hypothesis is the statement being tested, and we need a specific value to include in our calculations.

In our example of new medicine, the second statement $\mu_N \geq \mu_O$ has the equality sign so we take $\mu_N \geq \mu_O$ as the null hypothesis and $\mu_N < \mu_O$ as an alternative hypothesis. Therefore,

$H_0: \mu_N \geq \mu_O$ and

$H_1: \mu_N < \mu_O$

Similarly in case (ii), the claim is $\sigma < 0.60$ mg and its complement is $\sigma \geq 0.60$ mg. Since the complement $\sigma \geq 0.60$ mg contains an equality sign so we take the complement as a null hypothesis and claim $\sigma < 0.60$ mg as an alternative hypothesis, that is,

$H_0: \sigma \geq 0.60$ mg and

$H_1: \sigma < 0.60$ mg

In our example of the average mileage of the car, the claim is $\mu = 25$ km/litre and its complement is $\mu \neq 25$ km/litre. Since the claim contains the equality sign, so we take it as a null hypothesis and the complement $\mu \neq 25$ km/litre as an alternative hypothesis, that is,

$H_0: \mu = 25$ km/litre and

$H_1: \mu \neq 25$ km/litre

The following points should be kept in mind to test a hypothesis:

- The null and alternative hypotheses are two mutually exclusive statements about a population or population parameter(s) and cover all possibilities.

- In the testing of hypothesis procedure, we assume that the null hypothesis is true until the sample provides sufficient evidence that it is far more likely that the alternative hypothesis is true. If there is enough sample evidence to suggest that the null hypothesis is false, then we reject the null hypothesis and support the alternative hypothesis. If the sample fails to provide us with sufficient evidence against the null hypothesis, we are not saying that the null hypothesis is true because here, we take the decision on the basis of a random sample which is a small part of the population. To say that the null hypothesis is true we must study all observations of the population under study. For example, if someone wants to test that the monkey of India has two hands then to prove that this is true, we must check all monkeys of India whereas to prove that it is false, we require a single monkey which has one hand or no hand.
- Some authors use the equality sign (=) in the null hypothesis instead of $\geq$ and $\leq$ signs.

After understanding the null and alternative hypotheses, we study the types of the alternative hypothesis. The alternative hypothesis has two types:

Two-sided (tailed) alternative hypothesis

If the alternative hypothesis gives the alternative of the null hypothesis in both directions (less than and greater than), that is, it contains $\neq$ sign then it is known as a two-sided alternative hypothesis.

One-sided (tailed) alternative hypothesis

If the alternative hypothesis gives the alternative of the null hypothesis only in one direction (less than or greater than), that is, it contains > or < sign then it is known as a one-sided alternative hypothesis. If the alternative hypothesis contains **>** sign, then it is known as a **right-tailed hypothesis** and if it contains **<** sign then it is known as a **left-tailed hypothesis**.

For example, in our example of the average mileage of the car, the null and alternative hypotheses are

H_0: μ = 25 km/litre and H_1: $\mu \neq$ 25 km/litre

The alternative hypothesis H_1: $\mu \neq$ 25 km/litre is a two-sided alternative hypothesis because $\mu \neq 25$ means that the average mileage may be greater than or less than 25 km/litre. Similarly, if H_1: μ > 25 km/litre then it is a right-sided alternative hypothesis because it means that the average milage is greater than 25 km/litre and if H_1: μ < 25 km/litre then it is a left-sided alternative hypothesis because it means that the average milage is less than 25 km/litre.

After understanding the concept of null and alternative hypotheses, let us discuss simple and composite hypotheses in the next subsection.

15.2.2 Simple and Composite Hypotheses

In a general sense, if a hypothesis specifies only one value or exact value of the population parameter then it is known as a **simple hypothesis**. If a hypothesis specifies not just one value but a range of values that the population parameter may assume is called a **composite hypothesis**.

A hypothesis which completely specifies parameter(s) of a theoretical population (probability distribution) is called a simple hypothesis otherwise called composite hypothesis.

As in the above examples, the average mileage of the car H_0: μ = 25 km/litre is a simple hypothesis because it gives a single value of the parameter (μ = 25) whereas the alternative hypothesis H_0: μ ≠ 25 km/litre is a composite hypothesis because it does not specify the exact average mileage it may be greater than 25 such as 26,28,30, etc. or less than 25 such as 20,22,24, etc.

You should try the following exercises to check your understanding before moving to the next term of the testing of hypothesis.

SAQ 1

Write the null and alternative hypotheses in cases (iii), (v) and (vi) discussed in Section 15.2.

After describing the hypothesis and its types, our next point in the testing of hypothesis is to find/locate rejection and non-rejection regions which will be described in the next section.

15.3 REJECTION REGION AND NON-REJECTION REGION

As you studied in the previous section, the null hypothesis is tested with the help of sample data. The sample data either provide support for the null hypothesis or refuse the null hypothesis.

A test statistic is a function/statistic that identifies how far or how many standard deviations a sample outcome is from the value stated in a null hypothesis.

In particular, if there is a big discrepancy between the data and the null hypothesis, then we conclude that the null hypothesis is wrong. In the example of the average mileage of the car, the null hypothesis stated that the average mileage of the car is 25 km/litre. To check it, the automobile engineer selected, say, 20 cars of the same company and same model and noted the mileage of each car. After that, he calculated the sample mean. Suppose it comes out as 20 km/litre which is less than the claim mean (25 km/litre). Are we rejecting the claim because it is less than the claim's mean? The situation is not as straightforward as you might think because we are analysing a sample instead of all cars (population). It may be possible a second sample has a sample mean near 25 km/litre. Therefore, it is impossible to assess this possibility by looking at only the sample mean. We need to use a hypothesis test to determine the likelihood of obtaining our sample mean if the null hypothesis is true i.e. the population mean is 25. Therefore, we examine all of the possible values of a test statistic (sample mean) that could be obtained if the null hypothesis is true and prepare the sampling distribution of the statistic as we have studied in Units 2 and 3. The entire area under the sampling distribution probability curve equals 1. Suppose in our case it is shown in Fig. 15.1.

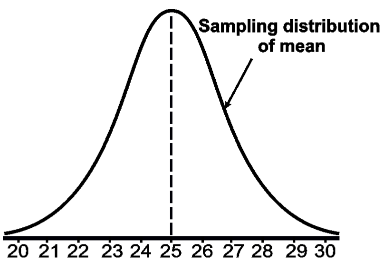


Fig. 15.1

If you look at the sampling distribution curve, then you will notice that there is no particular location on the curve where you can definitively say which values are "near" and which values are "very far" from the population mean 25 km/litre. There is only a consistent decrease in the likelihood of observing sample means that are farther from the null hypothesis value. How do we decide a sample mean is far away enough?

To answer this question, we will need a hypothesis testing tool! The concept of rejection region and non-rejection region help us.

We can divide the sampling distribution of the test statistic into two regions as follows:

1. A region, in the sampling distribution of a test statistic, in which the calculated value of the test statistic is likely to lie if the null hypothesis H_0 is true is called **the region of non-rejection**. This region lies around the centre value of the sampling distribution of the test statistic. If the calculated value of the test statistic lies in this region, we do not reject the null hypothesis.
2. A region, in the sampling distribution of a test statistic, in which the calculated value of the test statistic is very unlikely to lie if the null hypothesis H_0 is true is called **the rejection region** (sometimes it is also called the **critical region**). This region locates all those values that are so large or so small and the probability of their occurrence under a true null hypothesis is very small when the null hypothesis is true. This region lies very far from the centre and lies at the tail of the sampling distribution of the test statistic. If the test statistic lies in this region, we reject the null hypothesis.

For better understanding, we explain the concept of the rejection region and non-rejection region by taking a very simple example.

Suppose 100 students of class 10 of a school appeared in the annual exam. Suppose there are 5 subjects (Hindi, Mathematics, English, Science and Social Science) in class 10^{th} and the scores in these papers are denoted by $X_1, X_2, \ldots, X_5$. The maximum mark for each paper is 100. To obtain the grade A1 (outstanding) in all subjects the students need to have a total score equal to or more than 450 which is a rule.

Suppose we select one student randomly out of 100 students, and we want to test that the selected student is an outstanding awarded holder. So we can take the null and alternative hypotheses as

H_0: Selected student is not an outstanding awarded holder

H_1: Selected student is an outstanding awarded holder

For taking the decision whether the selected student is an outstanding awarded holder or not, on the basis of his/her total marks in all 5 subjects, therefore, we define a statistic $T_5 = \sum_{i=1}^{5} X_i$ as the sum of the scores in all 5 papers of the student. The range of T_5 is $0 \leq T_5 \leq 500$. Now, we divide the range of the test statistic (0-500) into two regions: non-outstanding awarded region (the region in which the null hypothesis is not rejected, i.e. less than 450) and outstanding awarded region (the region in which the null hypothesis is rejected, i.e. greater than or equal to 450) as shown in Fig. 15.2.

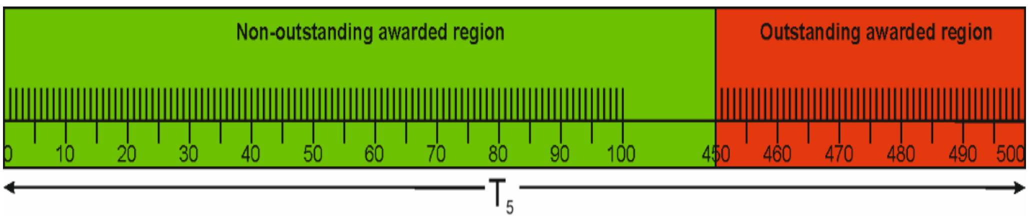


Fig. 15.2: Rejection and non-rejection regions for outstanding award.

Critical value is a value which separates the rejection region and non-rejection region.

Here, the value 450 is the **critical value** which separates the non-outstanding and outstanding awarded regions. On the basis of scores in all papers of the selected student, we calculate the value of the test statistic $T_5 = \sum_{i=1}^{5} X_i$. If the calculated value of test statistic T_5 lies in the non-outstanding awarded region (non-rejection region), that is, $T_5 < 450$ then we do not reject the null hypothesis H_0 and if the calculated value of the test statistic T_5 lies in the outstanding awarded region (rejection/critical region), that is, $T_5 \geq 450$ then we reject the null hypothesis H_0. It is a basic structure of the procedure of testing of hypothesis which needs two regions:

(i) Region of rejection of null hypothesis H_0

(ii) Region of non-rejection of null hypothesis H_0

I think that you understand the concept of rejection and non-rejection regions. This is a very simple example and by rule, we take 450 as the cut-off/critical value but in cases such as the average mileage of the car what will be the cut-off value? How do we decide it? To answer this, the statisticians find the values of the test statistics that are very unlikely to be obtained if the null hypothesis is true with the help of the sampling distribution of the test statistic and prepare the tables of the same. Thus, we obtain critical values for the size of the critical region (α) by using the tables of the sampling distributions of the test statistic as we have discussed in Unit 5.

The rejection (critical) region lies in one-tail or two-tails on the probability curve of the sampling distribution of the test statistic depending upon the alternative hypothesis. Therefore, three cases arise:

Case I: If the alternative hypothesis is right-sided such as $H_1: \theta > \theta_0$ or $H_1: \theta_1 > \theta_2$, then the entire critical/rejection region of size α lies on the right-tail of the probability curve of the sampling distribution of the test statistic as shown in Fig. 15.3.

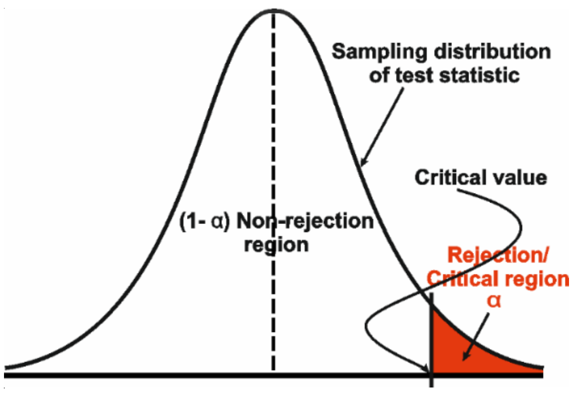


Fig. 15.3: Rejection and non-rejection regions for right-sided alternative hypothesis.

Case II: If the alternative hypothesis is left-sided such as $H_1: \theta < \theta_0$ or $H_1: \theta_1 < \theta_2$, then the entire critical/rejection region of size α lies on the left-tail of the probability curve of the sampling distribution of the test statistic as shown in Fig. 15.4.

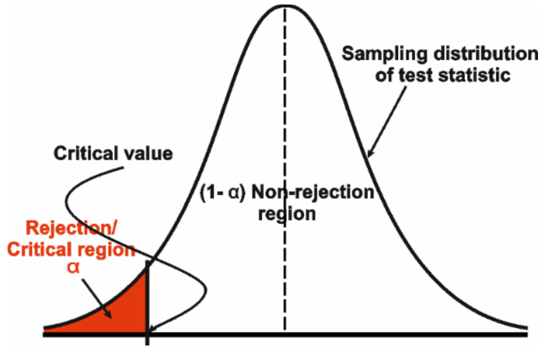


Fig. 15.4: Rejection and non-rejection regions for left-sided alternative hypothesis.

Case III: If the alternative hypothesis is two-sided such as H_1: $\theta \neq \theta_0$ or H_1: $\theta_1 \neq \theta_2$, then the total critical/rejection region of size α is equally divided into two regions and half-half critical region of size α/2 lies on both tails of the probability curve of the sampling distribution of the test statistic as shown in Fig. 15.5.

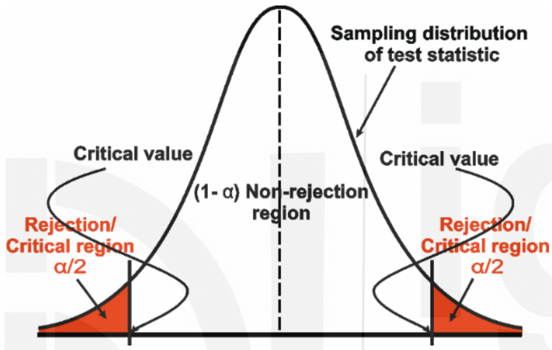


Fig. 15.5: Rejection and non-rejection regions for two-sided alternative hypothesis.

After understanding the concept of the critical region and where it lies, you can try the following exercise.

SAQ 2

Consider our example of the average mileage of the car and we formulated the null and alternative hypothesis as H_0: $\mu = 25$ k m/litre and H_1: $\mu \neq 25$ km/litre. Write where the critical region lies in this case.

15.4 TYPE-I AND TYPE-II ERRORS

In Section 15.3, we have discussed a rule that if the value of the test statistic falls in the rejection (critical) region, then we reject the null hypothesis and if it falls in the non-rejection region, then we do not reject the null hypothesis. A test statistic is calculated based on observed sample observations. But a sample is a small part of the population and based on this, we take the decision about the null hypothesis. A random sample may or may not be a good representative of the population.

A faulty sample misleads the inference (or conclusion) relating to the null hypothesis. For example, an engineer infers that a packet of screws is sub-standard when actually it is not. It is an error caused due to poor or inappropriate (faulty) sample. Similarly, a packet of screws may infer good when actually it is sub-standard. So we can commit two kinds of errors while testing a hypothesis which are summarised in Table 15.1:

Table 15.1: Type of Errors

Decision	H_0 True	H_1 True
Reject H_0	Type-I Error	Correct Decision
Do not reject H_0	Correct Decision	Type-II Error

To discuss these errors, let us take a situation where a patient suffering from high fever reaches a doctor. Suppose the doctor formulates the null and alternative hypotheses as

H_0: The patient is a malaria patient

H_1: The patient is not a malaria patient

Then the following cases arise:

Case I: Suppose that the null hypothesis H_0 is really true, that is, the patient is actually a malaria patient and after observation (pathological and clinical examination) the doctor rejects the null hypothesis H_0, that is, he/she declares himself /herself a non-malaria patient. It is not a correct decision, and he /she commits an error in the decision. This error is known as a Type-I error.

Case II: Suppose that the null hypothesis H_0 is actually false, that is, the patient is actually a non-malaria patient and after the observations, the doctor rejects H_0, that is, he/she declares himself /herself a non-malaria-patient. It is a correct decision.

Case III: Suppose that the null hypothesis H_0 is really true, that is, the patient is actually a malaria patient and after the observations, the doctor does not reject H_0, that is, he/she declares himself /herself a malaria patient. It is a correct decision.

Case IV: Suppose that the null hypothesis H_0 is actually false, that is, the patient is actually a non-malaria patient and after the observations, the doctor does not reject H_0, that is, he/she declares himself /herself a malaria patient. It is not a correct decision, and he/she commits an error in the decision. This error is known as a **Type-II error**.

Thus, we formally define Type-I and Type-II errors as follows:

Type-I Error

The rejection of a null hypothesis H_0 when it is true is called a Type-I error. The probability of committing the Type-I error is denoted by α and is given by

$$\alpha = P\,[\text{Reject } H_0 \text{ when } H_0 \text{ is true}] = P\,[\text{Reject } H_0 \mid H_0 \text{ is true}]$$

We reject the null hypothesis if a random sample/test statistic falls in the rejection region, therefore if $X = (X_1, X_2, \ldots, X_n)$ is a random sample and ω is the rejection region, then

$$\alpha = P\left[X \in \omega \mid H_0\right]$$

and

$$1 - \alpha = 1 - P[\text{Reject } H_0 \mid H_0 \text{ is true}]$$

$$= P[\text{Do not reject } H_0 \mid H_0 \text{ is true}] = P[\text{Correct decision}]$$

The $(1 - \alpha)$ is the probability of a correct decision and it correlates to the concept of $(1 - \alpha)100\%$ confidence interval used in estimation.

Type-II Error

The non-rejection of a null hypothesis H_0 when it is false (i.e. H_1 is true) is called a Type-II error. The probability of committing a Type-II error is generally denoted by β and is given by

$$\beta = P[\text{Do not reject } H_0 \text{ when } H_0 \text{ is false}] = P[\text{Do not reject } H_0 \text{ when } H_1 \text{ is true}]$$

$$= P[\text{Do not reject } H_0 \,|\, H_1 \text{ is true}]$$

$$= P\left[X \in \bar{\omega} \,|\, H_1\right] \quad \text{where, } \bar{\omega} \text{ is the non-rejection region.}$$

Also,

$$1 - \beta = 1 - P[\text{Do not reject } H_0 \,|\, H_1 \text{ is true}]$$

$$= P[\text{Reject } H_0 \,|\, H_1 \text{ is true}] = P[\text{Correct decision}]$$

The $(1 - \beta)$ is the probability of a correct decision and is also known as the **"power of the test"**. Since it indicates the ability or power of the test to recognize correctly that the null hypothesis is false, therefore, we wish for a test that yields a larger power.

A test is said to be ideal if it minimises the probability of both types of errors and maximises the probability of a correct decision. But for fixed sample size, α and β are so interrelated that the decrement in one result in an increment in the other. Therefore, for a fixed sample size, the minimisation of both probabilities of Type-I and Type-II errors is not possible simultaneously. Also, the probabilities of both types of errors will be at zero level (i.e. no error in decision) if the size of the sample is equal to the population size. But it involves a huge cost if the population size is large and it is not possible in all situations such as testing of blood. Depending on the problem at hand, we have to choose the type of error which has to be minimised.

So instead of demanding error-free procedures, we must seek procedures for which type of error is unlikely to occur. The choice of a particular rejection region, the critical/cutoff value fixes the probabilities of Type-I and Type-II errors. Because the null hypothesis specifies a unique value of the parameter, therefore, there is a single value of α. However, there is a different value of β for each value of the parameter consistent with the alternative hypothesis β (see Example 1).

Generally, strong control on α is necessary. It should be kept as low as possible. In the test procedure, we prefix it at a very low level like $\alpha = 0.05$ (5%) or 0.01 (1%) and try to minimise β or maximise the power of the test.

After understanding the concepts of the Type-I and Type-II errors, we now learn the power of a test in the next sub-section.

15.4.1 Power of a Test

Whenever we conduct a hypothesis test, we would like to make sure that it is a test of high quality. One way of quantifying the quality of a hypothesis test is to ensure that it is a "**powerful**" test. In this section, we will learn what it means to have a powerful hypothesis test as well as which factors affect the power of a test.

The power of the test is the probability that the test correctly rejects the null hypothesis H_0 when the alternative hypothesis H_1 is true. It is commonly denoted by $1-\beta$, where β is the probability of making a Type-II error. The power of a hypothesis test is a probability, therefore, it lies between 0 and 1. For a good test, the power should be as 0.8 and higher.

There are four main things that primarily affect the power of a test of significance. They are:

- **The significance level α of the test.** If all other things are held constant, then as α decreases, the power of the test also decreases. If we reduce α (from 0.05 to 0.01), then the non-rejection region increases. As a result, we are less likely to reject the null hypothesis. This means we are less likely to reject the null hypothesis when it is false, so we are more likely to make a Type-II error and hence reduce the power of the test.
- **The sample size n.** As n increases, the power of the test increases. This is because a larger sample size narrows the sampling distribution of the test statistic (as you have seen in Unit 2). The hypothesised distribution of the test statistic when the null hypothesis is true and the true distribution of the test statistic (the null hypothesis is false) become more distinct from one another as they become narrower, so it becomes easier to tell whether the observed statistic comes from one distribution or the other.
- **The inherent variability in the measured response variable.** As the variability increases, the power of the test decreases. One way to think of this is that a test of significance is like trying to detect the presence of a "signal," such as the effect of a treatment, and the inherent variability in the response variable is "noise" that will drown out the signal if it is too great. We cannot completely control the variability in the response variable, but they can sometimes reduce it through especially careful data collection and conscientiously uniform handling of experimental units or subjects.
- **The difference between the hypothesised value of a parameter and its true value.** The greater the difference between the "true" value of a parameter and the value specified in the null hypothesis, the greater the power of the test (see Example 1). This is because when the difference is large, the true distribution of the test statistic is far from its hypothesised distribution, so the two distributions are distinct, and it is easy to tell which one an observation came from.

Let us take an example to understand how to calculate the probability of types of errors and the power of a test.

Example 1: Suppose a pizza restaurant claims its average pizza delivery time is up to 30 minutes. But the students of the MSCAST programme believe that the restaurant takes more than 30 minutes. They formulate the hypothesis as

$H_0: \mu \le 30$ minutes and $H_1: \mu > 30$ minutes

To test this claim they noted 10 randomly chosen pizza deliveries of the restaurant. Also, the student rejects the null hypothesis if the sample mean exceeds 31 minutes. If the delivery time of pizza follows a normal distribution with a mean of 30 minutes and a standard deviation of 5 minutes. What would be the probability of committing Type-I and Type-II errors? Also, find the power of the test.

Solution: For calculating the error and power of the test, first of all, we search for the rejection region. Here, the student rejects the null hypothesis if the sample mean exceeds 31 minutes, therefore,

The rejection region is $\omega = P[\bar{X} > 31]$

Thus, we can calculate the Type-I error as

α = P[Reject H_0 when H_0 is true] $= P[\bar{X} > 31 | H_0]$

To calculate the above probability, we require the sampling distribution of the sample mean. Here, it is given that the delivery time follows a normal distribution with a mean of 30 minutes and a standard deviation of 5 minutes. Therefore, the sampling distribution of the sample mean follows a normal distribution with mean μ = 30 and standard deviation $\sigma / \sqrt{n} = 5 / \sqrt{10} = 1.58$, and we calculate the desired probability as we discussed in Units 2 and 3. Thus, we convert the sample mean delivery time to the standard normal Z-score (as discussed in Unit 2) as follows:

$$Z = \frac{\bar{X} - \mu}{SD(\bar{X})}$$

We can transform the above expression of the probability to the Z-scores as follows:

$$P\left[\bar{X} > 31 | H_0\right] = P\left[\frac{\bar{X} - 30}{1.58} > \frac{31 - 30}{1.58}\right] \text{ (since } H_0 \text{ is true so we take } \mu = 30)$$

$$= P\left[Z > 0.63\right]$$

From the standard normal table (Table IV) given in the Appendix of this volume, we get

$$P\left[Z > 0.63\right] = 1 - P\left[Z \le 0.63\right]$$

$$= 1 - 0.7357 = 0.2643 \text{ (from Table III)}$$

Therefore, the probability of the Type-I error is 0.2643. It means that 26.43% of all experiments carried out as described will result in the null hypothesis being rejected when it is actually true. Hence, the Type-I error is very large.

Now we come to calculate the probability of committing the Type-II error. By the definition of it, we have

β = P[Do not reject H_0 when H_1 is true]

$$= P\left[\bar{X} \le 31 | H_1\right] = P\left[\frac{\bar{X} - \mu}{SD(\bar{X})} \le \frac{31 - \mu}{SD(\bar{X})} | H_1 \text{ is true}\right]$$

But when H_1 is true, we do not know the exact value of μ, therefore, there is not a single β. Instead, there is a different β for each different μ that exceeds 30. Thus, there is a value of β for μ = 31, another value of β for μ = 32, and so on. If we take H_1: μ = 31.5 then $\bar{X} \sim N(31.5, 1.58)$, therefore,

$$\beta = P\left[\bar{X} \le 31 | H_1\right] = P\left[\frac{\bar{X}-\mu}{SD\left(\bar{X}\right)} \le \frac{31-\mu}{SD\left(\bar{X}\right)} | H_1 \text{ is true}\right] = P\left[\frac{\bar{X}-31.5}{1.58} \le \frac{31-31.5}{1.58}\right]$$

$$\beta = P\left[Z \le -0.32\right] = 0.3745 \text{ (from Table IV)}$$

If we take H_1: μ = 35 then $\bar{X} \sim N(35, 1.58)$, therefore,

$$\beta = P\left[\bar{X} \le 31 | H_1\right] = P\left[\frac{\bar{X}-\mu}{SD\left(\bar{X}\right)} \le \frac{31-\mu}{SD\left(\bar{X}\right)} | H_1 \text{ is true}\right] = P\left[\frac{\bar{X}-35}{1.58} \le \frac{31-35}{1.58}\right]$$

$$\beta = P\left[Z \le -2.53\right] = 0.0057 \quad \text{(from Table III)}$$

Since the power of the test (1 – β) also depends on the Type-II error, therefore, it is different for each different μ that exceeds 30. If we take H_1: μ = 31.5 then

Power of the test $= 1-\beta = 1-0.3745 = 0.6255$

If we take H_1: μ = 35 then

Power of the test $= 1-\beta = 1-0.0057 = 0.9943$

Hence, as the difference between the "true" value of a parameter and the value specified in the null hypothesis increases, the power of the test also increases.

Now, you can try the following exercise.

SAQ 3

A person gets a coin on the road and wants to test whether it is unbiased or not. Therefore, the person formulates the hypotheses as

$$H_0: p = p_0 = \frac{1}{2} \text{ and } H_1: p = p_1 = \frac{1}{4}$$

where p represents the probability of getting a head.

The person decides that he will toss the coin one time and reject the null hypothesis if a head appears. Find the value of α, β and power of the test.

The next section will address the idea of **"level of significance"**, which is quite valuable in making decisions while testing a hypothesis.

15.5 LEVEL OF SIGNIFICANCE

For testing a hypothesis, we take a sample from the population under study and use the information contained in this sample to make the decision whether a null hypothesis is true or false. Since a sample is a small part of the population and we take the decision based on it, therefore, we cannot be 100% sure of the decision. Therefore, we set some level of confidence, or level of significance, at which we trust that our conclusion is accurate.
The level of significance (or α level) is specified as the probability that the decision is incorrect. In other words:

The probability of rejecting a null hypothesis when it is true is called the level of significance.

Since it is the risk of rejecting the null hypothesis when it is really true, therefore, it is sometimes called the **level of risk**. The level of significance is denoted by the Greek letter alpha (α) and it is decided at the initial stages of the procedure of hypothesis testing. Generally, we set the level of significance at 5% (or 0.05). Although other levels may also be used depending on the study. Traditionally, the 5% (0.05) level is used for common purposes, 1% (0.01) for quality assurance and medical, and 10% (0.10) for political polling. If we set it at 5% then it means our decision is willing to be incorrect (rejecting H_0 when it is true) 5% of the time however we will be confident that the "concluding decision about H_0" is true with 95% assurance.

For example, consider our example of the average mileage of the car. If 100 automobile engineers want to test the claim of the company μ = 25 km/litre and each draws a random sample of 20 cars. Suppose they set a 5% level of significance and use the same test statistic to test the same null hypothesis H_0 conducting the same experiment, then 95 of them will reach the same conclusion about H_0. But still, 5 of them may differ (i.e. against the earlier conclusion).

Since in hypothesis testing, we assume the null hypothesis is true, therefore, we control for Type-I error by stating a level of significance. The level we set is the largest probability of committing a Type-I error that we will allow and still decide to reject the null hypothesis.

The level of significance is also called the size of the critical region and is used to find the critical value which separates the rejection region from the non-rejection region. The level of significance is the area (probability) in the 'tails' of a sampling distribution of the test statistic where the rejection region is located. Once we set the level of significance (α) (usually at 5%, 1% or 10%), then we know the size of the rejection region and use the table of the sampling distribution of the test statistic (Z, t, chi-square, F, etc.) to find the cut-off value(s) that provides us rejection region (critical area) equal to 5% (or 1%).

Let us move to the next term which is widely used in hypothesis testing.

> Researchers use the term statistical significance to ascertain whether the observed difference or effect between groups is not mere chance but true. The researchers, in this way, can understand whether the study results are meaningful, or whether they could have happened just by chance. Level of significance means how sure a researcher is that the results found are not accidental (not by chance). A level of significance of α = 0.05 means that there is a 95% probability that the results found in the study are the result of a true relationship/difference between groups being compared. It also means that there is a 5% chance that the results were found by chance alone.

15.6 ONE-TAILED AND TWO-TAILED TESTS

When we use hypothesis testing to determine whether a claim is true or not. We have to judge whether it is a one-tailed or a two-tailed test (also known as "directional" and "non-directional" tests respectively) so that we can find the critical values from the tables and we can take the decision whether a claim is true or not. Now, a question may arise, "How do we decide whether it is a one-tailed or a two-tailed test?" As the critical region depends on whether the alternative hypothesis is one-tailed or two-tailed, the test also depends on the alternative hypothesis in the question. Let us learn how to decide whether it is a one-tailed or a two-tailed test.

One-tailed Test

In hypothesis testing, when we test (check) the claim whether the parameter is higher or lower (< or >) than the current population parameter, it is known as the one-tailed test. It means that if the words such as **"increased, greater, larger, improved and so on", or "decreased, less, smaller and so on"** in the original claim of a question (or >, < are used in the alternative hypothesis), a one-tail test is applied. One-tailed hypothesis tests are also known as **directional and one-sided tests** because we can test for effects in only one direction. It can be a left-tailed test or a right-tailed test.

- **Left-tailed test:** If the claim has words such as **"decreased, less, smaller and so on"** (or < is used in the alternative hypothesis), that is,

  $H_0 : \theta \geq \theta_0$ and $H_1 : \theta < \theta_0$

  then a test is used to test such a hypothesis is called a left-tailed test. The critical region for the left-tailed test lies in the extreme left region (tail) under the probability curve of the sampling distribution of the test statistic as shown in Fig. 15.6.

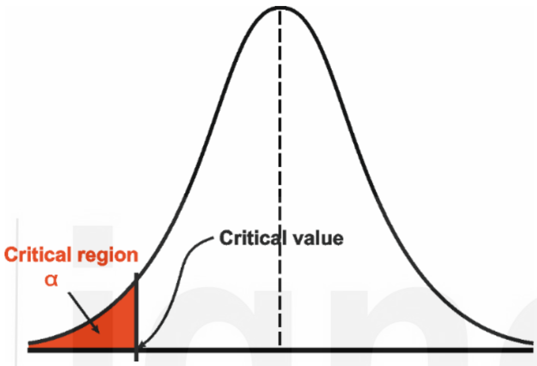


Fig. 15.6: Left-tailed test

- **Right-tailed test:** If the claim has words such as **"increased, greater, larger, improved and so on"**(or > is used in the alternative hypothesis), that is,

  $H_0 : \theta \leq \theta_0$ and $H_1 : \theta > \theta_0$

  then a test is used to test such a hypothesis is called a right-tailed test. The critical region for the right-tailed test lies in the extreme right region (tail) under the probability curve of the sampling distribution of the test statistic as shown in Fig. 15.7.

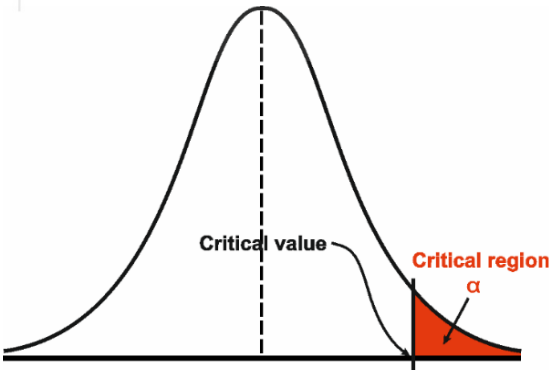


Fig. 15.7: Right-tailed test

<u>Two-tailed Test</u>

In hypothesis testing, when we test the claim whether the parameter changes (could be increased or decreased), then it is known as the two-tailed test. It means that if words such as **"change, different/difference"** and so on are used in the question (or ≠ is used in the alternative hypothesis), that is,

$H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$

then a test is used to test such a hypothesis is called a two-tailed test. It is also known as **a non-directional or two-sided test**. When we perform a two-tailed test, the entire critical region lies equally at the end of both tails of the probability curve of the sampling distribution of the test statistic as shown in Fig. 15.8.

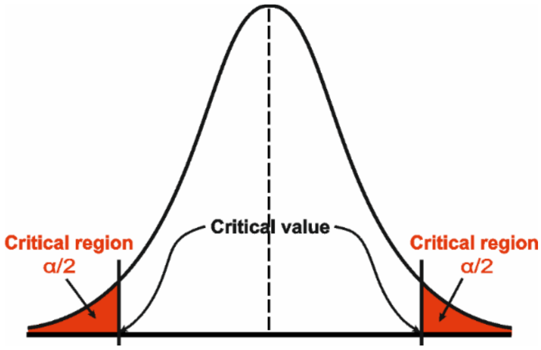


Fig. 15.8: Two-tailed test

The above discussion can be summarised in Table 15.2.

Table 15.2: Null and Alternative Hypotheses and Corresponding One-tailed and Two-tailed Tests

Null Hypothesis	Alternative Hypothesis	Types of Critical Region / Test
$H_0 : \theta = \theta_0$	$H_1 : \theta \neq \theta_0$	Two-tailed test having critical regions under both tails.
$H_0 : \theta \leq \theta_0$	$H_1 : \theta > \theta_0$	Right-tailed test having critical region under right tail only.
$H_0 : \theta \geq \theta_0$	$H_1 : \theta < \theta_0$	Left-tailed test having critical region under left tail only.

Let us take an example for illustration purposes.

Example 2: The manufacturer of electronic components has stated that the percentage of defective components is less than 12% in its shipments. Set up null and alternate hypotheses to check the claim and suggest whether we will use a one-tailed or two-tailed test.

Solution: The manufacturer of electronic components claimed that the percentage of defective electronic components in its shipment is less than 12%. We may write the claim in the form of a hypothesis as **"the proportion of defective electronic components in its shipment is less than 0.12."** Here, we are concerning the proportion of defective components. If the proportion of defective electronic components in its shipment is denoted by P, then the claim is $P < 0.12$ and its complement is $P \geq 0.12$. Since the complement contains the equality sign so we take it as a null hypothesis. Hence, we can formulate the null and alternative hypotheses as

H_0: $P \geq 0.12$

and

H_1: $P < 0.12$

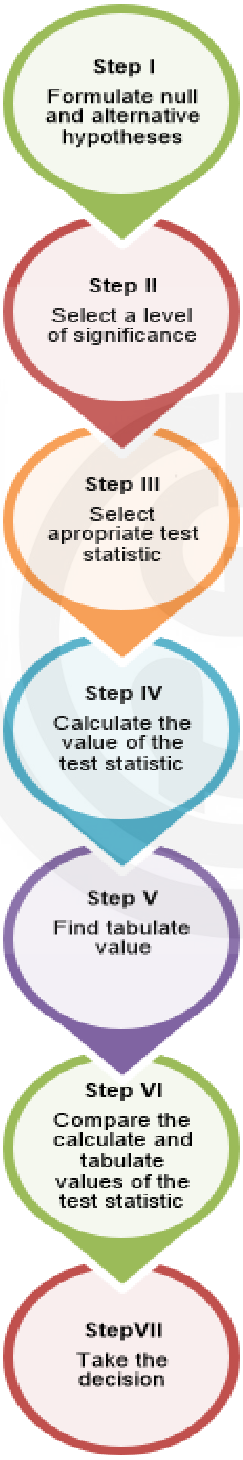
Since in the question, less than word is used and the alternative hypothesis contains less than sign (<) so we will use a left-tailed test.

Now, you can try the following exercises.

SAQ 4

A company has replaced its original technology of producing LED electric bulbs with the new technology. The company manager claims that the average life of LED bulbs produced by the new technology is greater than the original technology. Write appropriate null and alternative hypotheses. Also, suggest whether a one-tailed or a two-tailed test will be used to test the claim.

After understanding the basic terms of testing of hypothesis, you are now in a position to understand the procedure of hypothesis testing. The same is discussed in the next section.



15.7 GENERAL PROCEDURE OF TESTING A HYPOTHESIS

Testing of hypothesis is a huge demanded statistical tool by many disciplines and professionals. It is a step-by-step procedure as you will see in the next three units through a large number of examples. The aim of this section is just to give you the flavour of hypothesis testing which involves the following steps:

Step I: First of all, we have to set up null and alternative hypotheses. Suppose we want to test the hypothetical/claimed/assumed value θ_0 of the parameter θ. So we can take the null and alternative hypotheses as

$$H_0 : \theta = \theta_0 \text{ and } H_1 : \theta \neq \theta_0 \qquad [\text{two-tailed test}]$$

or

$$\left.\begin{array}{ll} H_0 : \theta \leq \theta_0 \text{ and } H_1 : \theta > \theta_0 & [\text{right-tailed test}] \\ H_0 : \theta \geq \theta_0 \text{ and } H_1 : \theta < \theta_0 & [\text{left-tailed test}] \end{array}\right\}$$

In case of comparing the same parameter in two populations of interest, say, θ_1 and θ_2, then our null and alternative hypotheses would be

$$H_0 : \theta_1 = \theta_2 \text{ and } H_1 : \theta_1 \neq \theta_2 \qquad [\text{two-tailed test}]$$

or

$$\left.\begin{array}{ll} H_0 : \theta_1 \leq \theta_2 \text{ and } H_1 : \theta_1 > \theta_2 & [\text{right-tailed test}] \\ H_0 : \theta_1 \geq \theta_2 \text{ and } H_1 : \theta_1 < \theta_2 & [\text{left-tailed test}] \end{array}\right\}$$

Step II: After setting the null and alternative hypotheses, we establish a criterion for rejection or non-rejection of the null hypothesis, that is, we decide the level of significance (α) at which we want to test our hypothesis. Generally, it is taken as 5% or 1% (α = 0.05 or 0.01).

Step III: The third step is to choose an appropriate test statistic for testing the null hypothesis assuming that the null hypothesis is true. It is too important step and you have to choose intelligently the appropriate test for a given situation. The given information in the question helps you to select the same. The general formula of the test statistic is given as follows:

$$\text{Test statistic} = \frac{\text{Statistic} - \text{Value of the parameter under } H_0}{\text{Standard error of the statistic}}$$

After that, we specify the sampling distribution of the test statistic preferably in the standard form like Z (standard normal), chi-square, t, F or any other well-known in the literature.

Step IV: After selecting the appropriate test statistic, we calculate the value of the test statistic on the basis of the observed sample.

Step V: We obtain the critical (or cut-off) value(s) in the sampling distribution of the test statistic and construct the rejection (critical) region of size α. Generally, the critical values for various levels of significance are put in the form of a table for various standard sampling distributions of the test statistic such as Z-table, chi-square table, t-table, etc.

Step VI: The test statistic (calculated in Step IV) is now compared to the critical value (obtained in Step VI). The test statistic either lies within (inside) the non-rejection region or it lies within the rejection region of H_0.

Step VII: In testing of hypothesis ultimately, we have to reach a conclusion. It is done as explained below:

(i) If the calculated value of the test statistic falls within the rejection region, then we reject the null hypothesis at the given level of significance. It means that the sample observations provide us with sufficient evidence against the null hypothesis and there is a significant difference between the hypothesised value and observed value of the parameter.

(ii) If the calculated value of the test statistic falls within the non-rejection region, then we do not reject the null hypothesis at the given level of significance. It means that the sample data fails to provide us with sufficient evidence against the null hypothesis and the difference between the hypothesised value and the observed value of the parameter is due to chance.

Step VIII: After that, we draw the final conclusion about the claim as

- If the null hypothesis is rejected and the claim exists in the null hypothesis, then the claim is probably false, so we reject it.
- If the null hypothesis is rejected and the claim exists in the alternative hypothesis, then the claim is probably true, so we do not reject it and support the claim.
- If the null hypothesis is not rejected and the claim exists in the null hypothesis, then the claim is probably true, and we may not reject the claim.
- If the null hypothesis is not rejected and the claim exists in the alternative hypothesis, then the claim is probably false, so we reject it.

To apply the above procedure, let us take an example.

Example 3: Suppose it is found that the average battery backup of smart mobile phones is 10 hours (screen-on-time) with a standard deviation of 3.6 hours nearly 2 years ago. Battery Cell Module (BCM) engineers believe that the average battery backup has increased due to advancements in technology. To test their belief, they randomly selected 26 smart mobile phones and calculated the average battery backup as 12 hours. If the battery backup of the smart mobile phones follows normal distribution, then describe the procedure to carry out this test at a 1% level of significance.

Solution: Here, we are given that

Here the average battery backup and standard deviation are given. We take these as the values of population because these are the general statements before taking the sample.

Specified value of population mean (μ_0) = 10 hours

Population standard deviation (σ) = 3.6 hours

Sample size (n) = 26

Sample mean $(\bar{X})$ = 12 hours

To carry out the above test, we have to follow up the following steps:

Step I: First of all, we set up null and alternative hypotheses. Here, we want to test that the average battery backup has been increased. So our claim is "average battery backup has been increased". If we denote the average battery backup by μ then the claim is $\mu > 10$ and its complement is $\mu \le 10$. Since the null hypothesis represents that there is no change or the old standard is still true, therefore, we take the null hypothesis as

$H_0: \mu \le 10$

And the alternative hypothesis represents that there is a change, or new standards are true, therefore, we take the alternative hypothesis as

$H_1: \mu > 10$.

In another way, the complement ($\mu \le 10$) contains an equality sign so we can take the complement as the null hypothesis and claim as the alternative hypothesis.

Step II: After setting the null and alternative hypotheses, we decide the level of significance. Here, it is given as 0.01 (= 1%).

Step III: To test the null hypothesis, we have to define a test statistic. Here, we are interested in hypothesis testing related to the mean, therefore, the test statistic depends on the sampling distribution of the mean. Since the population standard deviation is given and the battery backup of the smart mobile phones follows normal distribution, therefore, we consider the following test statistic as

$$Z_{cal} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

The sampling distribution of test statistic approximately follows standard normal distribution (as explained in Unit 2 of this course), i.e. $Z \sim N(0,1)$

Step IV: We now calculate the value of the test statistic based on sample observations as

$$Z_{cal} = \frac{12 - 10}{3.6 / \sqrt{26}} = \frac{2 \times 5.1}{3.6} = 2.83$$

Step V: Now, we find the critical value. Since the sampling distribution follows the standard normal distribution, therefore, we use the standard normal table

(**Table I**) to find out the critical value or cut-off value. Since the alternative hypothesis is right-tailed so the critical region will lie at the right-tailed of the distribution as shown in Fig. 15.9. Thus, from the Z-table, the critical value for the right-tailed test at $\alpha = 0.01$ is $z_{\alpha} = z_{0.01} = 2.33$.

Step IV: To take the decision about the null hypothesis, we compare the calculated value of the test statistic with the critical value.

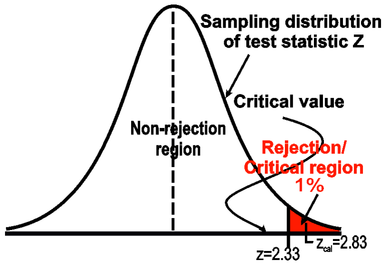


Fig. 15.9

Since the calculated value of the test statistic (= 2.83) is greater than the critical value (= 2.33), it means that it lies in the rejection region at the 1% level of significance as shown in Fig. 15.9. So, we reject the null hypothesis and support the alternative hypothesis. Since the alternative hypothesis is our claim, so we support the claim.

Thus, we conclude that the sample does not provide us with sufficient evidence against the claim so we may assume that the average battery backup of the smart mobile phones has been increased due to advancements in technology.

You can try the following Self Assessment Question.

SAQ 5

The income tax department of a country introduced new Income Tax Return (ITR) forms and believes that it takes, typical salary-earning taxpayers, less than 15 minutes, on average, to complete their tax return using eFiling (the online tax return submission system). To test this claim, a chartered accountant (CA) randomly selected 12 salary-earning taxpayers who had registered for eFiling and recorded their time to complete the eFiling process. The CA observed that the sample average completion time for the 12 taxpayers is 12.5 minutes, with a sample standard deviation of 5.2 minutes. If the time to complete the eFiling process follows normal distribution, then test whether the department's claim is likely to be true at a 5% level of significance.

Another method of performing hypothesis testing is to determine the probability of the null hypothesis being true or false. This is known as the p-value approach. In this next section, we will discuss it in detail.

15.8 p-VALUE APPROACH FOR TESTING OF HYPOTHESIS

In the previous section, you studied the procedure of testing a hypothesis. According to that, we compare the test statistic to a critical value. If the test statistic lies in the rejection region, then we reject the null hypothesis and if it lies in the non-rejection region then we do not reject the null hypothesis. This is often called **fixed significance level testing**. If we use the critical value approach (fixed significance level), then we may face the following difficulties:

- We only know that the results are significant or not significant at the stated level of significance. This statement of conclusion is often inadequate because it does not give whether the computed value of the test statistic was just barely in the rejection region or whether it was very far into this region.

The use of the p-value is becoming more and more popular because of the following reasons:

- The p-value provides more information compared to critical value as far as rejection or do not rejection of the null hypothesis.
- It allows us to check on how strong the statistical evidence is in favour of or against a null hypothesis.

- Furthermore, stating the results this way imposes a predefined level of significance on other users. This approach may be unsatisfactory because some decision-makers might be uncomfortable with the risks implied by $\alpha = 0.05$.
- When analysing a problem of hypothesis testing, various researchers employ varying levels of significance, therefore, it can occasionally be challenging for a reader to compare the outcomes of two distinct tests.

To avoid such difficulties, the p-value approach has been adopted widely in practice.

Moving in this direction, we note that in scientific applications someone is not only interested simply in rejecting or not rejecting the null hypothesis, but he/she is also interested in assessing how strong the data has the evidence to reject the null hypothesis. For example, as we have seen in Example 3 where we tested the null hypothesis

$H_0: \mu \le 10$ against $H_1: \mu > 10$

To test the null hypothesis, we calculated the value of the test statistic as 2.83 and the critical value (z_α) at $\alpha = 0.01$ was $z_\alpha = 2.33$.

Since the calculated value of the test statistic (= 2.83) lies in the critical region (2.33 to ∞), therefore, we rejected the null hypothesis at the 1% level of significance.

Now, if we reject the null hypothesis at this level (1%) surely, we will reject it at a higher level because at $\alpha = 0.05$, $z_\alpha = 1.645$ and at $\alpha = 0.10$, $z_\alpha = 1.28$, that is, the size of the critical region will increase. The question arises "could the null hypothesis also be rejected at values of α smaller than 0.01?" The answer is "yes" and we can compute the smallest level of significance (α) at which a null hypothesis can be rejected. This smallest level of significance (α) is known as **"p-value"**. The p-value is the smallest value of the level of significance(α) at which a null hypothesis can be rejected. We can also define the p-value as:

A p-value is the probability of getting sample evidence that is equally or more unfavourable to the null hypothesis while the null hypothesis is actually true.

Therefore, it provides a measure of how much evidence there is to reject or not reject the null hypothesis. Since a p-value is a probability, therefore, its value lies between 0 and 1. A p-value close to 0 suggests that the test statistic calculated from the data gets further away from the range of the test statistic when the null hypothesis is true. It means that the observed difference between the statistic and parameter is less likely due to random chance, it suggests that there is a significant difference between them and the strong evidence against the null hypothesis, so the null hypothesis is rejected. On the other hand, a p-value close to 1 suggests that there is no significant difference between them other than chance and there is not enough evidence against the null hypothesis, therefore, the null hypothesis is not rejected. The p-values are expressed as decimals and can be converted into percentages. For example, a p-value of 0.0237 is 2.37%, which means there is a 2.37% chance that the results are random or happened by chance.

After understanding the concept of the p-value, let us learn how to calculate it.

Computation of p-Value

The calculation for a p-value depends on the type of test performed. As you know, three types of tests are used in hypothesis testing which depend on the alternative hypothesis.

- Right-tailed test (One-sided upper-tailed test)
- Left-tailed test (One-sided lower-tailed test)
- Two-tailed test

The concept of calculating the p-value is the same in all three cases but the procedure is different. By the definition of the p-value, the p-value is the probability of getting evidence that is equally or more unfavourable to the null hypothesis. It means that to calculate the p-value, we have to find the probability that the test statistic takes values which are not in favour/possible when the null hypothesis is true.

Thus, we can calculate the p-value in the following cases:

For right-tailed test:

In this case, the null and alternative hypotheses are

$H_0 : \theta \le \theta_0$

$H_1 : \theta > \theta_0$

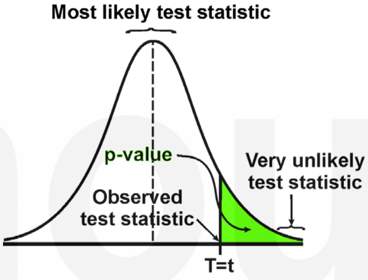


Fig. 15.10

In this case, the p-value is the area to the right of the calculated value of the test statistic as shown in Fig. 15.10.

Thus, we can calculate the p-value as

p-value = P[Test statistic (T) ≥ observed value of the test statistic (T_{cal})]

For left-tailed test:

In this case, the null and alternative hypotheses are

$H_0 : \theta \ge \theta_0$ and $H_1 : \theta < \theta_0$

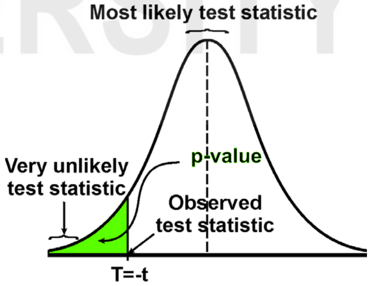


Fig. 15.11

In this case, the p-value is the area to the left of the calculated value of the test statistic as shown in Fig. 15.11.

Thus, we can calculate the p-value as

p-value = P[Test statistic (T) ≤ observed value of the test statistic (T_{cal})]

If the sampling distribution of the test statistic is symmetrical such as Z, t, etc., then the left-tail probability is equal to the right-tailed probability, therefore, we can use either the right-tailed or left-tailed formula to calculate the p-value.

For two-tailed test:

In this case, the null and alternative hypotheses are

$H_0 : \theta = \theta_0$ and $H_1 : \theta \ne \theta_0$

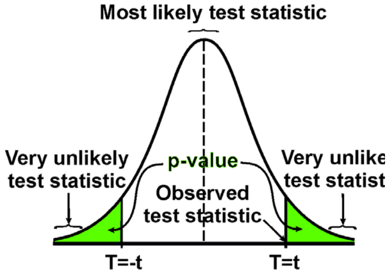


Fig. 15.12

In this case, the p-value is double the area to the right of the absolute value of the calculated test statistic as shown in Fig. 15.12.

Thus, we can calculate the p-value as

p-value = 2P[Test statistic (T) ≥ |observed value of the test statistic (T_{cal})|]

We can calculate the p-value by using the sampling distribution of the test statistic as we have calculated in Units 1 to 3. That is, we use the tables of the sampling distribution of the test statistics which are given at the end of this volume. But unless we are dealing with the standard normal distribution, the exact p-value is not obtained with the tables such as t, chi-square and F tables. However, with the help of computer packages/ software such as R, SPSS, SAS, MINITAB, STATA, EXCEL, etc., we can easily calculate the exact p-value. Also, such types of computer packages/ software present the p-value as part of the output for each hypothesis testing procedure.

Making a Decision on the Basis of a p-value

A p-value is typically used in hypothesis testing to determine whether the results obtained from a statistical analysis are significant or not. On the basis of the p-value, we can take the decision of rejection or non-rejection of the null hypothesis. The following table helps us to take the decision:

p-value	Decision
p-value > 0.05	The result is not statistically significant and hence, we do not reject the null hypothesis.
0.01 < p-value < 0.05	The result is statistically significant. We reject the null hypothesis in favour of the alternative hypothesis.
p-value < 0.01	The result is highly statistically significant. Hence, we reject the null hypothesis in favour of the alternative hypothesis.

For simplicity, we can also compare the p-value to a predetermined level of significance (α), usually set at 0.05, to determine whether to reject or not reject the null hypothesis. Let us try to understand the same.

- If the p-value is less than or equal to the level of significance, that is, **p-value ≤ α**, then the results are considered statistically significant, and **the null hypothesis is rejected**.
- If the p-value is greater than the level of significance, that is, **p-value > α**, then the results are not statistically significant, and **the null hypothesis is not rejected**.

Let us see how to compute the p-value with the help of an example.

Example 4: Consider Example 3 of the average battery backup of the smart mobile phones. Find the p-value and test whether the claim is rejected or not rejected using the p-value approach.

Solution: In Example 3, we have calculated the test statistic as

$$Z_{cal} = \frac{12-10}{3.6/\sqrt{26}} = \frac{2 \times 5.1}{3.6} = 2.83$$

Since the test is right-tailed, therefore, we can calculate the p-value as

$$\text{p-value} = P[\text{Test Statistic (T)} \geq \text{observed value of the test statistic}]$$
$$= P\left[Z \geq Z_{cal}\right] = P\left[Z \geq 2.83\right]$$

Here, the sampling distribution of the test statistic follows a standard normal distribution, therefore, we can use the standard normal table (Table IV) given at the end of this volume to calculate this probability (p-value) as we have calculated in Unit 2.

$$\text{p-value} = P[Z \ge 2.83] = 1 - P[Z < 2.83] = 1 - 0.9977 = 0.0023$$

Since the p-value for this hypothesis test is less than 5% (i.e. the p-value = 0.0023 < 0.05), therefore, there is strong sample evidence to reject H_0 in favour of H_1. In other words, we can say that there is only a 0.23% chance that the null hypothesis is true. Since this is a very low probability (less than 5%), therefore, we reject H_0 in favour of H_1.

Thus, we conclude that the sample does not provide us with sufficient evidence against the claim so we may assume that the average battery backup of the smart mobile phones has been increased.

You noticed that we get the same conclusion when we have used the critical values approach.

Now, it is time for you to try the following Self Assessment Question to make sure that you have learnt how to calculate the p-value.

SAQ 6

Consider SAQ 5 of the income tax department. Find the p-value and test whether the claim that the average eFiling completion time is less than 15 minutes is rejected or not rejected using the p-value approach.

15.9 CONFIDENCE INTERVAL APPROACH FOR TESTING OF HYPOTHESIS

In Units 12 and 13, you have learned what is confidence interval and how to construct a confidence interval estimate of a population parameter with certain confidence. When we construct a (1 − α)100% confidence interval for an unknown parameter then this interval contains the true value of the parameter with (1 − α) 100% confidence and the chance that it does not lie in the interval is 100α%. Therefore, we can use this concept to take the decision about the null hypothesis. For that, we construct a (1 − α)100% confidence interval and reject the null hypothesis at α level of significance if the value of the parameter specified in the null hypothesis does not lie in that interval. We may summarise how to use a confidence interval approach for hypothesis testing as follows:

Case I: If we want to test the null hypothesis H_0: $\theta = \theta_0$ against the alternative hypothesis H_1: $\theta \neq \theta_0$ (two-tailed) at 5% or 1% level of significance then we construct a two-sided (1 – α)100% = 95% or 99% confidence interval for the parameter θ. We have 95% or 99% (as may be the case) confidence that this interval will include the value of the parameter specified under the null hypothesis, that is, $\theta = \theta_0$. If the value of the parameter specified by the null hypothesis i.e. θ_0 does not lie in this confidence interval, then we reject the null hypothesis otherwise we may not reject the null hypothesis.

Case II: If we want to test the null hypothesis H_0: $\theta \le \theta_0$ against the alternative hypothesis H_1: $\theta > \theta_0$ then we construct (1 – α) 100% upper one-sided confidence interval (left-tail) (because the alternative hypothesis is right/upper tail) for the parameter θ. If the confidence interval does not contain the value of the parameter specified by the null hypothesis i.e. $\theta \le \theta_0$ then we reject the null hypothesis otherwise we may not reject the null hypothesis.

Case III: If we want to test the null hypothesis H_0: $\theta \ge \theta_0$ against the alternative hypothesis H_1: $\theta < \theta_0$ then we construct $(1 - \alpha)100\%$ lower one-sided confidence interval (right-tail) (because the alternative hypothesis is a left/lower-tail) for the parameter θ. If the confidence interval does not contain the value of the parameter specified by the null hypothesis i.e. $\theta \ge \theta_0$ then we reject the null hypothesis otherwise we may not reject the null hypothesis.

Let us take an example to illustrate purpose.

Example 5: Consider Example 3 of the average battery backup of smart mobile phones. Test the claim at a 1% level of significance using the confidence interval approach.

Solution: Here, we are given that

$\mu_0 = 10, \sigma = 3.6, n = 50, \bar{X} = 12$

In this example, we formulated the null and alternative hypotheses which are given as follows:

$H_0: \mu \le 10$ and $H_1: \mu > 10$

Since we are interested in testing the hypothesis related to the mean and the alternative hypothesis is $H_1: \mu > 10$, therefore, we construct $(1 - \alpha)$ 100% upper/right one-sided confidence interval for the mean.

Since the population standard deviation is given and the battery backup of the smart mobile phones follows a normal distribution, therefore, we use $(1 - \alpha)$ 100% upper one-sided confidence interval for the population mean which is given by

Upper one-sided confidence interval = $\left[\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right]$

For α = 0.01, we have, $z_\alpha = z_{0.01} = 1.645$

Thus, we can calculate the 95% upper one-sided confidence interval as

$$\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}} = 12 - 1.645 \frac{3.6}{\sqrt{26}} = 12 - 1.16 = 10.84$$

Therefore, the required confidence interval is given by

$$\left[\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right] = [10.84, \infty]$$

Since the 95% upper one-sided confidence interval does not contain 10 and lower values, therefore, we reject the null hypothesis. Therefore, we draw the same conclusion as we have obtained using critical value and p-value approaches.

Thus, we can use three approaches (critical value, p-value and confidence interval) for taking the decision about the null hypothesis.

With this, we end this unit. Let us summarise what we have discussed in this unit.

15.10 SUMMARY

In this unit, we have covered the following points:

- A hypothesis or statistical hypothesis is a statement or a claim or an assumption about the value of a population parameter such as mean, median, variance, proportion, etc.
- A null hypothesis is a hypothesis that states that there is no change, no difference, or no relationship, that is, there is nothing new happening, the old theory is still true, old standard is correct.
- An alternative hypothesis is a hypothesis that states that there is a change, a difference, or a relationship, that is, the new theory is true, there are new standards.
- The rejection of a null hypothesis H_0 when it is true is called a Type-I error.
- The non-rejection of a null hypothesis H_0 when it is false (i.e. H_1 is true) is called a Type-II error.
- The power of the test is the probability that the test correctly rejects the null hypothesis H_0 when the alternative hypothesis H_1 is true.
- The probability of rejecting a null hypothesis when it is true is called the level of significance.
- A p-value is the probability of getting sample evidence that is equally or more unfavourable to the null hypothesis while the null hypothesis is actually true.

15.11 TERMINAL QUESTIONS

1. Describe the null and alternative hypotheses.

2. If the calculated value of a test statistic is greater than the tabulated value for the right-tailed test, then what will be the decision about the null hypothesis?

3. How to take the decision of rejection and non-rejection of the null hypothesis using the p-value and level of significance?

15.12 SOLUTIONS / ANSWERS

Self Assessment Questions(SAQs)

1. In part (iii), the claim is P = 0.50 then its complement is P ≠ 0.50. Since the claim contains the equality sign so we take the claim as a null hypothesis and the complement as the alternative hypothesis. Therefore,

H_0: P = 0.50 and H_1: P ≠ 0.50

In part (v), the claim is $\sigma_1^2 < \sigma_2^2$ then its complement will be $\sigma_1^2 \geq \sigma_2^2$. Since the complement contains the equality sign so we take it as a null hypothesis and the claim as the alternative hypothesis. Therefore,

H_0: $\sigma_1^2 \geq \sigma_2^2$ and H_1: $\sigma_1^2 < \sigma_2^2$

In part (v), the claim is $P_1 < P_2$ then its complement will be $P_1 \geq P_2$. Since the complement contains the equality sign, we take it as a null hypothesis and the claim as the alternative hypothesis. Therefore,

$H_0: P_1 \geq P_2$ and $H_1: P_1 < P_2$

2. Since alternative hypothesis $H_1: \mu \neq 25$ km/litre is two-tailed so critical region lies in both tails.

3. First of all, we search for the critical region. Here, the person rejects the null hypothesis if a head appears, therefore, the critical region is given as follows:

ω = {getting head}

Thus, we can compute the probability of the Type-I error as

α = P[Reject H_0 when H_0 is true]

$= P[X \in \omega | H_0] = P[\text{Head appears} | H_0]$

$$= P\left[\text{Head appears}\right]_{p=\frac{1}{2}} = \frac{1}{2} \quad \left[\begin{array}{l}\because H_0 \text{ is true so we take value} \\ \text{of the parameter p given in } H_0\end{array}\right]$$

Also,

β = P[Do not reject H_0 when H_1 is true]

$= P\left[X \notin \omega | H_1\right] = P\ [\text{Tail appears} | H_1]$

$$= P\left[\text{Tail appears}\right]_{p=\frac{1}{4}} \quad \left[\begin{array}{l}\because H_1 \text{ is true so we take value} \\ \text{of the parameter p given in } H_1\end{array}\right]$$

$$= 1 - P\left[\text{Head appears}\right]_{p=\frac{1}{4}} = 1 - \frac{1}{4} = \frac{3}{4}$$

Therefore, the power of the test $= 1 - \beta = 1 - \frac{3}{4} = \frac{1}{4} = 0.25$

4. We may write the claim as "the average life of LED bulbs produced by the new technology is greater than original technology." Here, we are concerning the average of the two groups. If we represent the average life of LED bulbs produced by the new and old technologies by μ_N and μ_O, respectively, then the claim is $\mu_N > \mu_O$ and its complement is $\mu_N \leq \mu_O$. Since the complement contains the equality sign so we take it as a null hypothesis. Hence, we can formulate the null and alternative hypotheses as

$H_0: \mu_N \leq \mu_O$ and $H_1: \mu_N > \mu_O$

Since the claim has the word **greater than** and the alternative hypothesis is right-tailed, therefore, the corresponding test will be the right-tailed test.

5. Here, we are given that

Specified value of population mean = μ_0 = 15 minutes

Sample size = n = 12

Sample mean = $\bar{X}$ = 12.5 hours

Sample standard deviation = S = 5.2 minutes

To carry out the above test, we have to follow up the following steps:

Step I: First of all, we set up the null and alternative hypotheses. Here, we want to test the claim that the average time to complete the eFiling process by salary-earning taxpayers is less than 15 minutes. So our claim is "average time to complete the eFiling process by salary-earning taxpayers is less than 15 minutes" If we denote the average time by μ, then the claim is $\mu < 15$ and its complement will be $\mu \geq 15$. Since the null hypothesis represents that there is no change or the new form takes time as old (larger time), therefore, we take the null hypothesis as

$H_0: \mu \geq 15$

And the alternative hypothesis represents that there is a change, or new form reduces the time, therefore, we take the alternative hypothesis as

$H_1: \mu < 15.$

In another way, since the complement ($\mu \geq 15$) contains an equality sign so we can take it as the null hypothesis and claim as the alternative hypothesis.

Step II: In the question, the level of significance (α) is given as $\alpha = 0.05$ (= 5 % level).

Step III: To test the null hypothesis, we have to define a test statistic. Here, we are interested in testing the hypothesis related to the mean then the test statistic depends on the sampling distribution of the mean. Since the population standard deviation is not given and the time to complete the eFiling process follows normal distribution, therefore, we consider the following test statistic:

$$t = \frac{\bar{X} - \mu}{S / \sqrt{n}}$$

As we know, the sampling distribution of the test statistic follows t distribution (as explained in Unit 2 of this course).

Step IV: We now calculate the value of the test statistic based on sample observations as

$$t_{cal} = \frac{12.5 - 15}{5.2 / \sqrt{12}} = \frac{-2.5 \times 3.46}{5.2} = -1.66$$

Step V: Now, we find the critical value. Since the sampling distribution follows the t-distribution, therefore, to find out the critical value or cut-off value, we use the t-table. From the t-table (Table V), the critical value for the left-tailed test corresponding (n –1) = 12 –1 =11 df at the 5% level of significance is $-t_{(n-1),\alpha} = -t_{(11),0.05} = -1.796.$

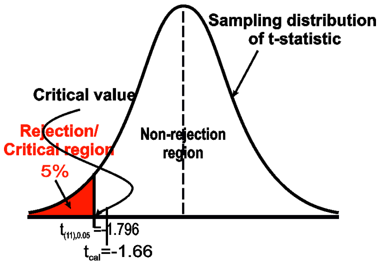


Fig. 15.13

Step IV: Since the calculated value of the test statistic (= – 1.66) lies in the non-rejection region as shown in Fig. 15.13. So, we do not reject the null hypothesis and support it. Since the claim is under the alternative hypothesis so reject the claim at the 5% level of significance.

Thus, we conclude that the sample does not provide us sufficient evidence against the claim. It can be concluded, with 95% confidence, that the average eFiling completion time is less than 15 minutes. Thus, the income tax department's claim may be supported.

6. In SAQ 5, we have calculated the test statistic as

$$t_{cal} = \frac{12.5-15}{5.2/\sqrt{12}} = \frac{-2.5\times 3.46}{5.2} = -1.66$$

Since the test is left-tailed, therefore, we can calculate the p-value as

p-value = P[Test Statistic (T) ≤ observed value of the test statistic]

$$= P\left[t \le -1.66\right]$$

Here, the sampling distribution of the test statistic follows a t-distribution with n – 1 = 11 degrees of freedom, therefore, we can use the t-table (given at the end of this volume) to calculate this probability (p-value) as we have calculated in Unit 2.

$$\text{p-value} = P\left[t_{(11)} \le -1.66\right] = P\left[t_{(11)} \ge 1.66\right] \begin{bmatrix}\text{Since the t-distribution} \\ \text{is symmetrical.}\end{bmatrix}$$

From the t-table (Table V), we get

$$P\left[t_{(11)} > 1.363\right] = 0.10 \text{ and } P\left[t_{(11)} > 1.796\right] = 0.05$$

Therefore, we can find the required probability as:

$$0.05 < P\left[t_{(11)} \le -1.66\right] < 0.10$$

P[t ≤ –1.66]
R code
= pt(–1.66,11)= 0.063

Note: With the help of the t-table, we cannot calculate the exact probability. But computer packages or software such as R, SPSS, SAS, MINITAB, STATA, EXCEL, etc. help us to calculate it exactly. From R, we find it as

$$P\left[t_{(11)} \le -1.66\right] = 0.063$$

Since the p-value for this hypothesis test is greater than 5% (i.e. the p-value = 0.063 > 0.05), therefore, we do not reject the null hypothesis.

Thus, we conclude that the sample does not provide sufficient evidence against the claim. It can be concluded, with 95% confidence, that the average eFiling completion time is less than 15 minutes. Thus, the income tax department's claim may be supported.

Terminal Questions (TQs)

1. Referrer to Section 15.2.

2. In this case, the calculated value of the test statistic lies in the rejection region so we will reject the null hypothesis.

3. Referrer to Section 15.8.