UNIT 9 CONCEPTS OF TESTING OF HYPOTHESIS

Structure

9.1 INTRODUCTION

In previous block of this course, we have discussed one part of statistical inference, that is, estimation and we have learnt how we estimate the unknown population parameter(s) by using **point estimation** and **interval estimation**. In this block, we will focus on the second part of statistical inference which is known as **testing of hypothesis**.

In our day-to-day life, we see different commercials advertisements in television, newspapers, magazines, etc. such as

(i) The refrigerator of certain brand saves up to 20% electric bill,

(ii) The motorcycle of certain brand gives 60 km/liter mileage,

(iii) A detergent of certain brand produces the cleanest wash,

(iv) Ninety nine out of hundred dentists recommend brand A toothpaste for their patients to save the teeth against cavity, etc.

Now, the question may arise in our mind "can such types of claims be verified statistically?" Fortunately, in many cases the answer is "yes".

The technique of testing such type of claims or statements or assumptions is known as testing of hypothesis. The truth or falsity of a claim or statement is never known unless we examine the entire population. But practically it is not possible in mostly situations so we take a random sample from the population under study and use the information contained in this sample to take the decision whether a claim is true or false.

This unit is divided into 11 sections. Section 9.1 is introductory in nature. In Section 9.2, we defined the hypothesis. The concept and role of critical region in testing of hypothesis is described in Section 9.3. In Section 9.4, we explored the types of errors in testing of hypothesis whereas level of significance is explored in Section 9.5. In Section 9.6, we explored the types of tests in testing of hypothesis. The general procedure of testing a hypothesis is discussed in Section 9.7. In Section 9.8, the concept of p-value in decision making about the null hypothesis is discussed whereas the relation between confidence interval and testing of hypothesis is discussed in Section 9.9. Unit ends by providing summary of what we have discussed in this unit in Section 9.10 and solution of exercises in Section 9.11.

Objectives

After reading this unit, you should be able to:

- define a hypothesis;
- formulate the null and alternative hypotheses;
- explain what we mean by type-I and type-II errors;
- explore the concept of critical region and level of significance;
- define one-tailed and two-tailed tests;
- describe the general procedure of testing a hypothesis;
- concept of p-value; and
- test a hypothesis by using confidence interval.

Before coming to the procedure of testing of hypothesis, we will discuss the basis terms used in this procedure one by one in subsequent sections.

9.2 HYPOTHESIS

As we have discussed in previous section that in our day-to-day life, we see different commercials advertisements in television, newspapers, magazines, etc. and if someone may be interested to test such type of claims or statement then we come across the problem of testing of hypothesis. For example,

(i) a customer of motorcycle wants to test whether the claim of motorcycle of certain brand gives the average mileage 60 km/liter is true or false,

(ii) the businessman of banana wants to test whether the average weight of a banana of Kerala is more than 200 gm,

(iii) a doctor wants to test whether new medicine is really more effective for controlling blood pressure than old medicine,

(iv) an economist wants to test whether the variability in incomes differ in two populations,

(v) a psychologist wants to test whether the proportion of literates between two groups of people is same, etc.

In all the cases discussed above, the decision maker is interested in making inference about the population parameter(s). However, he/she is not interested in estimating the value of parameter(s) but he/she is interested in testing a claim or statement or assumption about the value of population parameter(s). Such claim or statement is postulated in terms of hypothesis.

In statistics, a hypothesis is a statement or a claim or an assumption about the value of a population parameter (e.g., mean, median, variance, proportion, etc.).

Similarly, in case of two or more populations a hypothesis is comparative statement or a claim or an assumption about the values of population parameters. (e.g., means of two populations are equal, variance of one population is greater than other, etc.). The plural of hypothesis is hypotheses.

In hypothesis testing problems first of all we should being identifying the claim or statement or assumption or hypothesis to be tested and write it in the words. Once the claim has been identified then we write it in symbolical form if possible. As in the above examples,

(i) Customer of motorcycle may write the claim or postulate the hypothesis "the motorcycle of certain brand gives the average mileage 60 km/liter." Here, we are concerning the **average** mileage of the motorcycle so let μ represents the average mileage then our hypothesis becomes $\mu = 60$ km / liter.

(ii) Similarly, the businessman of banana may write the statement or postulate the hypothesis "the average weight of a banana of Kerala is greater than 200 gm." So our hypothesis becomes $\mu > 200$ gm.

(iii) Doctor may write the claim or postulate the hypothesis " the new medicine is really more effective for controlling blood pressure than old medicine." Here, we are concerning the **average** effect of the medicines so let μ_1 and μ_2 represent the average effect of new and old medicines respectively on controlling blood pressure then our hypothesis becomes $\mu_1 > \mu_2$.

(iv) Economist may write the statement or postulate the hypothesis " the variability in incomes differ in two populations." Here, we are concerning the **variability** in income so let σ_1^2 and σ_2^2 represent the variability in incomes in two populations respectively then our hypothesis becomes $\sigma_1^2 \neq \sigma_2^2$.

(v) Psychologist may write the statement or postulate the hypothesis "the proportion of literates between two groups of people is same." Here, we are concerning the **proportion** of literates so let P_1 and P_2 represent the proportions of literates of two groups of people respectively then our hypothesis becomes $P_1 = P_2$ or $P_1 - P_2 = 0$.

The hypothesis is classified according to its nature and usage as we will discuss in subsequent subsections.

9.2.1 Simple and Composite Hypotheses

In general sense, if a hypothesis specifies only one value or exact value of the population parameter then it is known as simple hypothesis. And 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 hypothesis postulated in (i) $\mu = 60$ km/liter is simple hypothesis because it gives a single value of parameter ($\mu = 60$) whereas the hypothesis postulated in (ii) $\mu > 200$ gm is composite hypothesis because it does not specify the exact average value of weight of a banana. It may be 260, 350, 400 gm or any other.

Similarly, (iii) $\mu_1 > \mu_2$ or $\mu_1 - \mu_2 > 0$ and (iv) $\sigma_1^2 \neq \sigma_2^2$ or $\sigma_1^2 - \sigma_2^2 \neq 0$ are not simple hypotheses because they specify more than one value as $\mu_1 - \mu_2 = 4$, $\mu_1 - \mu_2 = 7$, $\sigma_1^2 - \sigma_2^2 = 2, \sigma_1^2 - \sigma_2^2 = -5$, etc. and (v) $P_1 = P_2$ or $P_1 - P_2 = 0$ is simple hypothesis because it gives a single value of parameter as $P_1 - P_2 = 0$.

9.2.2 Null and Alternative Hypotheses

As we have discussed in last page that in hypothesis testing problems first of all we identify the claim or statement to be tested and write it in symbolical form. After that we write the complement or opposite of the claim or statement in symbolical form. In our example of motorcycle, the claim is $\mu = 60$ km/liter then its complement is $\mu \neq 60$ km/liter. In (ii) the claim is $\mu > 200$ gm then its complement is $\mu \leq 200$ gm. If the claim is $\mu < 200$ gm then its complement is $\mu \geq 200$ gm. The claim and its complement are formed in such a way that they cover all possibility of the value of population parameter.

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

Once the claim and its compliment have been established then we decide of these two which is the null hypothesis and which is the alternative hypothesis. The thump rule is that the statement containing equality is the null hypothesis. That is, the hypothesis which contains symbols $=$ or $\leq$ or $\geq$ is taken as null hypothesis and the hypothesis which does not contain equality i.e. contains $\neq$ or $<$ or $>$ is taken as alternative hypothesis. The null hypothesis is denoted by H_0 and alternative hypothesis is denoted by H_1 or H_A.

In our example of motorcycle, the claim is $\mu = 60$ km/liter and its complement is $\mu \neq 60$ km/liter. Since claim $\mu = 60$ km/liter contains equality sign so we take it as a null hypothesis and complement $\mu \neq 60$ km/liter as an alternative hypothesis, that is,

$$H_0: \mu = 60 \text{ km/liter and } H_1: \mu \neq 60 \text{ km/liter}$$

In our second example of banana, the claim is $\mu > 200$ gm and its complement is $\mu \leq 200$ gm. Since complement $\mu \leq 200$ gm contains equality sign so we take complement as a null hypothesis and claim $\mu > 200$ gm as an alternative hypothesis, that is,

$$H_0: \mu \leq 200 \text{ gm and } H_1: \mu > 200 \text{ gm}$$

Formally these hypotheses are defined as

The hypothesis which we wish to test is called as the null hypothesis.

According to Prof. R.A. Fisher,

"A null hypothesis is a hypothesis which is tested for possible rejection under the assumption that it is true."

The hypothesis which complements to the null hypothesis is called alternative hypothesis.

Note 1: Some authors use equality sign in null hypothesis instead of $\geq$ and $\leq$ signs.

The alternative hypothesis has two types:

(i) Two-sided (tailed) alternative hypothesis

(ii) One-sided (tailed) alternative hypothesis

If the alternative hypothesis gives the alternate of null hypothesis in both directions (less than and greater than) of the value of parameter specified in null hypothesis then it is known as two-sided alternative hypothesis and if it gives an alternate only in one direction(less than or greater than) only then it is known as one-sided alternative hypothesis. For example, if our alternative hypothesis is $H_1: \theta \neq 60$ then it is a two-sided alternative hypothesis because its means that the value of parameter θ is greater than or less than 60. Similarly, if $H_1: \theta > 60$ then it is a right-sided alternative hypothesis because its means that the value of parameter θ is greater than 60 and if $H_1: \theta < 60$ then it is a left-sided alternative hypothesis because its means that the value of parameter θ is less than 60.

In testing procedure, we assume that the null hypothesis is true until there is sufficient evidence to prove that it is false. Generally, the hypothesis is tested with the help of a sample so evidence in testing of hypothesis comes from a sample. 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 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 null hypothesis is true we must study all observations of the population under study. For example, if someone wants to test that the person of India has two hands then to prove that this is true we must check all the persons of India whereas to prove that it is false we require a person he / she has one hand or no hand. So we can only say that there is not enough evidence against the null hypothesis.

Note 2: When we assume that null hypothesis is true then we are actually assuming that the population parameter is equal to the value in the claim. In our example of motorcycle, we assume that $\mu = 60$ km/liter whether the null hypothesis is $\mu = 60$ km/liter or $\mu \leq 60$ km/liter or $\mu \geq 60$ km/liter.

Now, you can try the following exercises.

E1) A company manufactures car tyres. Company claims that the average life of its tyres is 50000 miles. To test the claim of the company, formulate the null and alternative hypotheses.

E2) Write the null and alternative hypotheses in case (iii), (iv) and (v) of our example given in Section 9.2.

E3) A businessman of orange formulates different hypotheses about the average weight of the orange which are given below:

(i) $H_0: \mu = 100$ (ii) $H_1: \mu > 100$ (iii) $H_0: \mu \leq 100$ (iv) $H_1: \mu \neq 100$

(v) $H_1: \mu > 150$ (vi) $H_0: \mu = 130$ (vii) $H_1: \mu \neq \mu_0$

Categorize the above cases into simple and composite hypotheses.

After describing the hypothesis and its types our next point in the testing of hypothesis is critical region which will be described in next section.

9.3 CRITICAL REGION

As we have discussed in Section 9.1 that generally, null hypothesis is tested by the sample data. Suppose $X_1, X_2, \ldots, X_n$ be a random sample drawn from a population having unknown population parameter θ. The collection of all possible values of $X_1, X_2, \ldots, X_n$ is a set called sample space(S) and a particular value of $X_1, X_2, \ldots, X_n$ represents a point in that space.

In order to test a hypothesis, the entire sample space is partitioned into two disjoint sub-spaces, say, ω and $S - \omega = \bar{\omega}$. If calculated value of the test statistic lies in ω, then we reject the null hypothesis and if it lies in $\bar{\omega}$, then we do not reject the null hypothesis. The region ω is called a **"rejection region or critical region"** and the region $\bar{\omega}$ is called a **"non-rejection region"**. Therefore, we can say that

A statistic is a function of sample observations (not including parameter). Basically, a test statistic is a statistic which is used in decision making about the null hypothesis.

"A region in the sample space in which if the calculated value of the test statistic lies, we reject the null hypothesis then it is called critical region or rejection region."

This can be better understood with the help of an example. Suppose, 100 students appear in total 10 papers two of each in English, Physics, Chemistry, Mathematics and Computer Science of a Programme. Suppose the scores in these papers are denoted by $X_1, X_2, \ldots, X_{10}$ and maximum marks =100 for each paper. For obtaining the distinction award in this Programme, a student needs to have total score equal to or more than 750 which is a rule.

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

H_0: Selected student is a distinction award holder

H_1: Selected student is not a distinction award holder

For taking the decision about the selected student, we define a statistic $T_{10} = \sum_{i=1}^{10} X_i$ as the sum of the scores in all the 10 papers of the student. The range of T_{10} is $0 \le T_{10} \le 1000$. Now, we divide the whole space (0-1000) into two regions as no-distinction awarded region (less than 750) and distinction awarded region (greater than or equal to 750) as shown in Fig. 9.1. Here, 750 is the critical value which separates the no-distinction and distinction awarded regions.

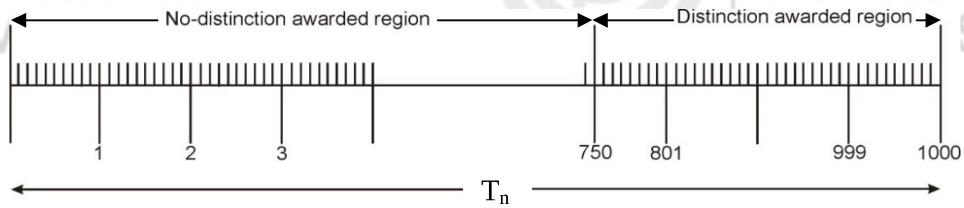

Fig. 9.1: Non-rejection and critical regions for distinction award

On the basis of scores in all the papers of the selected student, we calculate the value of the statistic $T_{10} = \sum_{i=1}^{10} X_i$. And calculated value may fall in distinction award region or not, depending upon the observed value of test statistic.

For making a decision to reject or do not reject H_0, we use test statistic $T_{10} = \sum_{i=1}^{10} X_i$ (sum of scores of 10 papers). If calculated value of test statistic T_{10} lies in no-distinction awarded region (critical region), that is, $T_{10} < 750$ then we reject H_0 and if calculated value of test statistic T_{10} lies in distinction awarded region (non-rejection region), that is, $T_{10} \ge 750$ then we do not reject H_0. It is a basic structure of the procedure of testing of hypothesis which needs two regions like:

(i) Region of rejection of null hypothesis H_0

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

The point of discussion in this test procedure is “**how to fix the cut off value 750”**? What is the justification for this value? The distinction award region may be like $T_{10} \ge 800$ or at $T_{10} \ge 850$ or at $T_{10} \ge 900$. So, there must be a scientific justification for the cut-off value 750. In a statistical test procedure it is obtained by using the probability distribution of the test statistic.

The region of rejection is called critical region. It has a pre-fixed area generally denoted by α, corresponding to a cut-off value in a probability distribution of test statistic.

The rejection (critical) region lies in one-tail or two-tails on the probability curve of sampling distribution of the test statistic its depends 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 or rejection region of size α lies on right tail of the probability curve of sampling distribution of the test statistic as shown in Fig. 9.2.

Critical value is a value or values that separate the region of rejection from the non-rejection region.

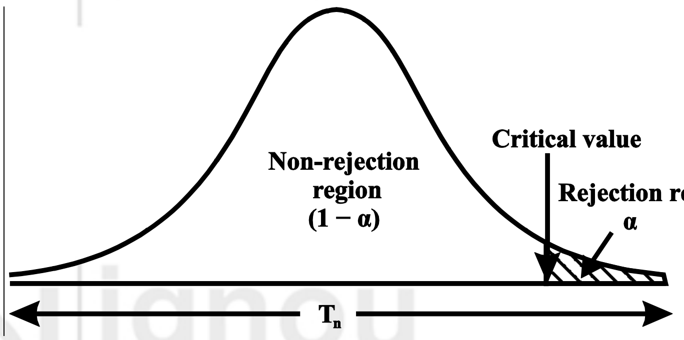

Fig. 9.2

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 or rejection region of size α lies on left tail of the probability curve of sampling distribution of the test statistic as shown in Fig. 9.3.

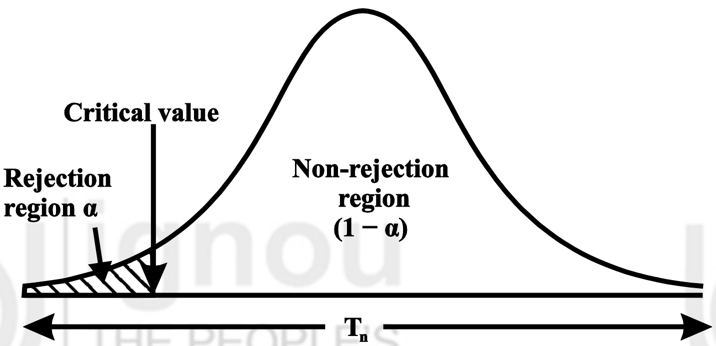

Fig. 9.3

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 critical or rejection regions of size α/2 lies on both tails of the probability curve of sampling distribution of the test statistic as shown in Fig. 9.4.

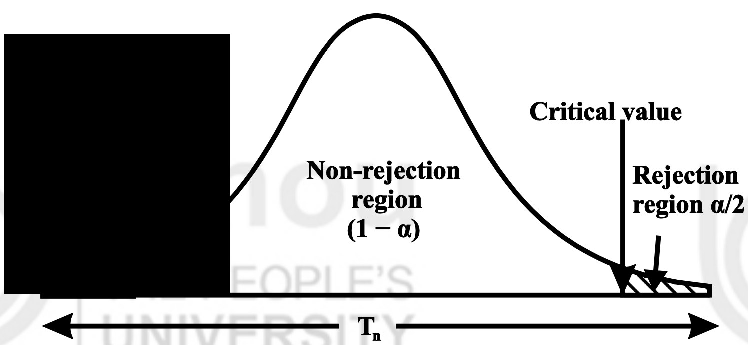

Fig. 9.4

Now, you can try the following exercise.

E4) If $H_0: \theta = 60$ and $H_1: \theta \neq 60$ then critical region lies in one-tail or two-tails.

9.4 TYPE-I AND TYPE-II ERRORS

In Section 9.3, we have discussed a rule that if the value of test statistic falls in 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 on the basis of observed sample observations. But a sample is a small part of the population about which decision is to be taken. 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 9.1 which is given below:

Table 9.1: 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

Let us take a situation where a patient suffering from high fever reaches to a doctor. And 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 following cases arise:

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

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

Case III: Suppose that the hypothesis H_0 is really true, that is, patient actually a malaria patient and after observation, the doctor does not reject H_0, that is, he / she declares him / her a malaria-patient. It is a correct decision.

Case IV: Suppose that the hypothesis H_0 is actually false, that is, patient actually a non-malaria patient and after observation, the doctor does not reject H_0, that is, he / she declares him / her a malaria-patient. It is not a correct decision and he / she commits an error in decision known as type-II error.

Thus, we formally define type-I and type-II errors as below:

Type-I Error:

The decision relating to rejection of null hypothesis H_0 when it is true is called type-I error. The probability of committing the type-I error is called size of test, denoted by α and is given by

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

We reject the null hypothesis if random sample / test statistic falls in rejection region, therefore,

$$\alpha = P[X \in \omega / H_0]$$

where $X = (X_1, X_2, \ldots, X_n)$ is a random sample and ω is the rejection region and

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

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

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

Type-II Error:

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

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

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

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

$$= P[X \in \overline{\omega} / H_1] \text{ where, } \overline{\omega} \text{ is the non-rejection region.}$$

and

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

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

The $(1-\beta)$ is the probability of correct decision and also known as **"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 a test that yields a large power.

We say that a statistical test is ideal if it minimizes the probability of both types of errors and maximizes the probability of correct decision. But for fix sample size, α and β are so interrelated that the decrement in one results into the increment in other. So minimization of both probabilities of type-I and type-II errors simultaneously for fixed sample size is not possible without increasing sample size. Also both types of errors will be at zero level (i.e. no error in decision) if size of the sample is equal to the population size. But it involves huge cost if population size is large. And it is not possible in all situations such as testing of blood.

Depending on the problem in hand, we have to choose the type of error which has to minimize. For this, we have to look at a situation, suppose there is a decision making problem and there is a rule that if we make type-I error, we lose10 rupees and if we make type-II error we lose 1000 rupees. In this case, we try to eliminate the type-II error, since it is more expensive.

In another situation, suppose the Delhi police arrests a person whom they suspect is a murderer. Now, policemen have to test hypothesis:

H_0: Arrested person is innocent (not murderer)

H_1: Arrested person is a murderer

The type-I error is

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

That is, suspected person who is actually an innocent will be sent to jail when H_0 rejects, although H_0 being a true.

The type-II error is

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

That is, when arrested person truly a murderer but released by the police. Now, we see that in this case type-I error is more serious than type-II error because a murderer may be arrested / punished later on but sending jail to an innocent person is serious.

Consider another situation, suppose we want to test the null hypothesis $H_0 : p = 0.5$ against $H_1 : p \neq 0.5$ on the basis of tossing a coin once, where p is the probability of getting a head in a single toss (trial). And we reject the null hypothesis if a head appears and do not reject otherwise. The type-I error, that is, the probability of Reject H_0 when it is true can be calculated easily(as shown in Example 1) but the computation of type-II error is not possible because there are infinitely many alternatives for p such as p = 0.6, p = 0.1, etc.

Generally, strong control on α is necessary. It should be kept as low as possible. In test procedure, we prefix it at very low level like $\alpha = 0.05$ (5%) or 0.01 (1%) .

Now, it is time to do some examples relating to α and β.

Example 1: It is desired to test a hypothesis $H_0 : p = p_0 = 1/2$ against the alternative hypothesis $H_1 : p = p_1 = 1/4$ on the basis of tossing a coin once, where p is the probability of "getting a head" in a single toss (trial) and agreeing to reject H_0 if a head appears and accept H_0 otherwise. Find the value of α and β.

Solution: In such type of problems, first of all we search for critical region.

Here, we have critical region ω = {head}

Therefore, the probability of type-I error can be obtained as

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

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

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

Also,

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

$$= P[X \notin \omega / H_1] = P\,[\text{Tail appears} / H_1]$$

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

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

Example 2: For testing H_0: $\theta = 1$ against H_1: $\theta = 2$, the pdf of the variable is given by

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}; & 0 \le x \le \theta \\ 0; & \text{elsewhere} \end{cases}$$

Obtain type-I and type-II errors when critical region is $X \geq 0.4$. Also obtain power function of the test.

Solution: Here, we have critical (rejection) and non-rejection regions as

$$\omega = \{X : X \geq 0.4\} \text{ and } \overline{\omega} = \{X : X < 0.4\}$$

We have to test the null hypothesis

$$H_0 : \theta = 1 \text{ against } H_1 : \theta = 2$$

The size of type-I error is given by

$$\alpha = P[X \in \omega / H_0] = P[X \geq 0.4 / \theta = 1]$$

$$= \left[\int_{0.4}^{\theta} f(x,\theta)\,dx\right]_{\theta=1} \qquad \left[\because P[X \geq a] = \int_{a}^{\infty} f(x,\theta)\,dx\right] \qquad \dots (1)$$

Now, by using $f(x,\theta) = \frac{1}{\theta}; \quad 0 \leq x \leq \theta$, we get from equation (1)

$$\alpha = \left[\int_{0.4}^{\theta} \frac{1}{\theta}\,dx\right]_{\theta=1} = \int_{0.4}^{1} dx = (x)_{0.4}^{1} = 1 - 0.4 = 0.6$$

Similarly, the size of type-II error is given by

$$\beta = P[X \in \overline{\omega} / H_1] = P[X < 0.4 / \theta = 2]$$

$$\beta = \left[\int_{0}^{0.4} \frac{1}{\theta}\,dx\right]_{\theta=2} = \int_{0}^{0.4} \frac{1}{2}\,dx = \frac{1}{2}(x)_{0}^{0.4} = \frac{1}{2}(0.4 - 0) = 0.2$$

The power function of the test = 1 - β = 1 - 0.2 = 0.8.

Now, you can try the following exercise.

E5) An urn contains either 4 white and 2 black balls or 2 white and 4 black balls. Two balls are to be drawn from the urn. If less than two white balls are obtained, it will be decided that this urn contains 2 white and 4 black balls. Calculate the values of α and β.

9.5 LEVEL OF SIGNIFICANCE

So far in this unit, we have discussed the hypothesis, types of hypothesis, critical region and types of errors. In this section, we shall discuss very useful concept **"level of significance"**, which play an important role in decision making while testing a hypothesis.

The probability of type-I error is known as level of significance of a test. It is also called the size of the test or size of critical region, denoted by α. Generally, it is pre-fixed as 5% or 1% level (α = 0.05 or 0.01). As we have discussed in Section 9.3 that if calculated value of the test statistic lies in rejection(critical) region, then we reject the null hypothesis and if it lies in non-rejection region, then we do not reject the null hypothesis. Also we note that when H_0 is rejected then automatically the alternative hypothesis H_1 is accepted. Now, one point of our discussion is that how to decide critical value(s) or cut-off value(s) for a known test statistic.

If distribution of test statistic could be expressed into some well-known distributions like Z, χ^2, t, F etc. then our problem will be solved and using the probability distribution of test statistic, we can find the cut-off value(s) that provides us critical area equal to 5% (or 1%).

Another viewpoint about the level of significance relates to the trueness of the conclusion. If H_0 do not reject at level, say, $\alpha = 0.05$ (5% level) then a person will be confident that "concluding statement about H_0" is true with 95% assurance. But even then it may false with 5% chance. There is no cent-percent assurance about the trueness of statement made for H_0.

As an example, if among 100 scientists, each draws a random sample and use the same test statistic to test the same hypothesis H_0 conducting same experiment, then 95 of them will reach to the same conclusion about H_0. But still 5 of them may differ (i.e. against the earlier conclusion).

Similar argument can be made for, say, $\alpha = 0.01$ (=1%). It is like when H_0 is rejected at $\alpha = 0.01$by a scientist , then out of 100 similar researchers who work together at the same time for the same problem, but with different random samples, 99 of them would reach to the same conclusion however, one may differ.

Now, you can try the following exercise.

E6) If probability of type-I error is 0.05 then what is the level of significance?

9.6 ONE-TAILED AND TWO-TAILED TESTS

We have seen in Section 9.3 that rejection (critical) region lies at one-tail or two-tails on the probability curve of sampling distribution of the test statistic its depend upon the form of alternative hypothesis. Similarly, the test of testing the null hypothesis also depends on the alternative hypothesis.

A test of testing the null hypothesis is said to be two-tailed test if the alternative hypothesis is two-tailed whereas if the alternative hypothesis is one-tailed then a test of testing the null hypothesis is said to be one-tailed test.

For example, if our null and alternative hypothesis are

$$H_0 : \theta = \theta_0 \text{ and } H_1 : \theta \neq \theta_0$$

then the test for testing the null hypothesis is two-tailed test because the alternative hypothesis is two-tailed that means, the parameter θ can take value greater than θ_0 or less than θ_0.

If the null and alternative hypotheses are

$$H_0 : \theta \leq \theta_0 \text{ and } H_1 : \theta > \theta_0$$

then the test for testing the null hypothesis is right-tailed test because the alternative hypothesis is right-tailed.

Similarly, if the null and alternative hypotheses are

$$H_0 : \theta \geq \theta_0 \text{ and } H_1 : \theta < \theta_0$$

then the test for testing the null hypothesis is left-tailed test because the alternative hypothesis is left-tailed.

The above discussion can be summarised in Table 9.2.

Table 9.2: Null and Alternative Hypotheses and Corresponding One-tailed and Two-tailed Tests

Null Hypothesis	Alterative 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 do one example based on type of tests.

Example 3: A company has replaced its original technology of producing electric bulbs by CFL technology. The company manager wants to compare the average life of bulbs manufactured by original technology and new technology CFL. Write appropriate null and alternate hypotheses. Also say about one tailed and two tailed tests.

Solution: Suppose the average lives of original and CFL technology bulbs are denoted by μ_1 and μ_2 respectively.

If company manager is interested just to know whether any significant difference exists in average-life time of two types of bulbs then null and alternative hypotheses will be:

$H_0: \mu_1 = \mu_2$ [average lives of two types of bulbs are same]

$H_1: \mu_1 \neq \mu_2$ [average lives of two types of bulbs are different]

Since alternative hypothesis is two-tailed therefore corresponding test will be two-tailed.

If company manager is interested just to know whether average life of CFL is greater than original technology bulbs then our null and alternative hypotheses will be

$H_0: \mu_1 \geq \mu_2$

$$H_1: \mu_1 < \mu_2 \left[\begin{array}{l}\text{average life of CFL technology bulbs} \\ \text{is greater than average life of original technology}\end{array}\right]$$

Since alternative hypothesis is left-tailed therefore corresponding test will be left-tailed test.

Now, you can try the following exercises.

E7) If we have null and alternative hypotheses as

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

then corresponding test will be

(i) left-tailed test (ii) right-tailed test (iii) both-tailed test

Write the correct option.

E8) The test whether one-tailed or two-tailed depends on

(i) Null hypothesis (H_0) (ii) Alternative hypothesis (H_1)

(iii) Neither H_0 nor H_1 (iv) Both H_0 and H_1

9.7 GENERAL PROCEDURE OF TESTING A HYPOTHESIS

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

Step I: First of all, we have to setup null hypothesis H_0 and alternative hypothesis H_1. Suppose, we want to test the hypothetical / claimed /

assumed value θ_0 of parameter θ. So we can take the null and alternative hypotheses as

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

$$\text{or} \quad \left.\begin{array}{l} H_0: \theta \leq \theta_0 \text{ and } H_1: \theta > \theta_0 \\ H_0: \theta \geq \theta_0 \text{ and } H_1: \theta < \theta_0 \end{array}\right\} \qquad [\text{for one-tailed test}]$$

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

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

$$\text{or} \quad \left.\begin{array}{l} H_0: \theta_1 \leq \theta_2 \text{ and } H_1: \theta_1 > \theta_2 \\ H_0: \theta_1 \geq \theta_2 \text{ and } H_1: \theta_1 < \theta_2 \end{array}\right\} \qquad [\text{for one-tailed test}]$$

Step II: After setting the null and alternative hypotheses, we establish a criteria for rejection or non-rejection of null hypothesis, that is, 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 under H_0 for testing the null hypothesis as given below:

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

After that, specify the sampling distribution of the test statistic preferably in the standard form like Z (standard normal), χ^2, t, F or any other well-known in literature.

Step IV: Calculate the value of the test statistic described in Step III on the basis of observed sample observations.

Step V: Obtain the critical (or cut-off) value(s) in the sampling distribution of the test statistic and construct rejection (critical) region of size α. Generally, critical values for various levels of significance are putted in the form of a table for various standard sampling distributions of test statistic such as Z-table, χ^2-table, t-table, etc.

Step VI: After that, compare the calculated value of test statistic obtained from Step IV, with the critical value(s) obtained in Step V and locates the position of the calculated test statistic, that is, it lies in rejection region or non-rejection region.

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

(i) If calculated value of test statistic lies in rejection region at α level of significance then we reject null hypothesis. It means that the sample data provide us sufficient evidence against the null hypothesis and there is a significant difference between hypothesized value and observed value of the parameter.

(ii) If calculated value of test statistic lies in non-rejection region at α level of significance then we do not reject null hypothesis. Its means that the sample data fails to provide us sufficient evidence against the null hypothesis and the difference between hypothesized value and observed value of the parameter due to fluctuation of sample.

Note 3: Nowadays the decision about the null hypothesis is taken with the help of p-value. The concept of p-value is very important, because computer packages and statistical software such as SPSS, SAS, STATA, MINITAB, EXCEL, etc. all provide p-value. So, Section 9.8 is devoted to explain the concept of p-value.

Now, with the help of an example we explain the above procedure.

Example 4: Suppose, it is found that average weight of a potato was 50 gm and standard deviation was 5.1 gm nearly 5 years ago. We want to test that due to advancement in agricultural technology, the average weight of a potato has been increased. To test this, a random sample of 50 potatoes is taken and calculate the sample mean $(\bar{X})$ as 52gm. Describe the procedure to carry out this test.

Solution: Here, we are given that

Specified value of population mean = μ_0 = 50 gm,

Population standard deviation = σ = 5.1 gm,

Sample size = n = 50,

Sample mean = $\bar{X}$ = 52 gm

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

Step I: First of all, we setup null and alternative hypotheses. Here, we want to test that the average weight of potato is increased. So our claim is "average weight of potato has increased" i.e. $\mu > 50$ and its complement is $\mu \le 50$. Since complement contains equality sign so we can take the complement as the null hypothesis and claim as the alternative hypothesis, that is,

$$H_0: \mu \le 50 \text{ gm and } H_1: \mu > 50 \text{ gm [Here, } \theta = \mu]$$

Since the alternative hypothesis is right-tailed, so our test is right-tailed.

Step II: After setting the null and alternative hypotheses, we fix level of significance α. Suppose, $\alpha = 0.01$ (= 1 % level).

Step III: Define a test statistic to test the null hypothesis as

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

$$T = \frac{\bar{X} - 50}{\sigma/\sqrt{n}}$$

Since sample size is large (n = 50 > 30) so by central limit theorem the sampling distribution of test statistic approximately follows standard normal distribution (as explained in Unit 1 of this course), i.e. $T \sim N(0,1)$

Step IV: Calculate the value of test statistic on the basis of sample observations as

$$T = \frac{52 - 50}{5.1/\sqrt{50}} = \frac{2}{0.72} = 2.78$$

Step V: Now, we find the critical value. The critical value or cut-off value for standard normal distribution is given in **Table I (Z-table)** in the Appendix at the end of Block 1 of this course. So from this table, the critical value for right-tailed test at $\alpha = 0.01$ is $z_\alpha = 2.33$.

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

Since calculated value of test statistic (= 2.78) is greater than critical value (= 2.33), that means calculated value of test statistic lies in rejection region at 1% level of significance as shown in Fig. 9.5. So we reject null hypothesis and support the alternative hypothesis. Since alternative hypothesis is our claim, so we support the claim.

Thus, we conclude that sample does not provide us sufficient evidence against the claim so we may assume that the average weight of potato has increased.

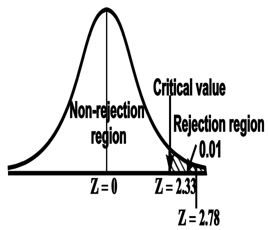

Fig. 9.5

Now, you can try the following exercise.

E9) What is the first step in testing of hypothesis?

9.8 CONCEPT OF p-VALUE

In Note 3 of Section 9.6, we promised that p-value will be discussed in Section 9.8. So, it is the time to keep our promise. Nowadays use of p-value is becoming more and more popular because of the following two reasons:

- most of the statistical software provides p-value rather than critical value.
- p-value provides more information compared to critical value as far as rejection or do not rejection of H_0.

The first point listed above needs no explanation. But second point lies in the heart of p-value and needs to explain more clearly. Moving in this direction, we note that in scientific applications one is not only interested simply in rejecting or not rejecting the null hypothesis but he/she is also interested to assess how strong the data has the evidence to reject H_0. For example, as we have seen in Example 4 based on general procedure of testing a hypothesis where we tested the null hypothesis

$$H_0: \theta \leq 50 \text{ gm against } H_1: \theta > 50 \text{ gm}$$

To test the null hypothesis, we calculated the value of test statistic as 2.78 and the critical value (z_α) at $\alpha = 0.01$ was $z_\alpha = 2.33$.

Since calculated value of test statistic (= 2.78) is greater than critical (tabulated) value (= 2.33), therefore, we reject the null hypothesis at 1% level of significance.

Now, if we reject the null hypothesis at this level (1%) surely we have to reject it at higher level because at $\alpha = 0.05$, $z_\alpha = 1.645$ and at $\alpha = 0.10$, $z_\alpha = 1.28$. However, the calculated value of test statistic is much higher than 1.645 and 1.28, therefore, 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 level of significance(α) at which a null hypothesis can be rejected using the obtained value of the test statistic and can be defined as:

The p-value is the probability of obtaining a test statistic equal to or more extreme (in the direction of sporting H_1) than the actual value obtained when null hypothesis is true.

The p-value also depends on the type of the test. If test is one-tailed then the p-value is defined as:

For right-tailed test:

$$\text{p-value} = P[\text{Test Statistic } (T) \geq \text{observed value of the test statistic}]$$

For left-tailed test:

$$\text{p-value} = P[\text{Test Statistic } (T) \leq \text{observed value of the test statistic}]$$

If test is two-tailed then the p-value is defined as:

For two-tailed test:

$$\text{p-value} = 2P\left[T \geq \left|\text{observed value of the test statistic}\right|\right]$$

Procedure of taking the decision about the null hypothesis on the basis of p-value:

To take the decision about the null hypothesis based on p-value, the p-value is compared with level of significance (α) and if p-value is equal or less than α then we reject the null hypothesis and if the p-value is greater than α we do not reject the null hypothesis.

The p-value for various tests can be obtained with the help of the tables given in the Appendix of the Block 1 of this course. But unless we are dealing with the standard normal distribution, the exact p-value is not obtained with the tables as mentioned above. But if we test our hypothesis with the help of computer packages or softwares such as SPSS, SAS, MINITAB, STATA, EXCEL, etc. then these types of computer packages or softwares present the p-value as part of the output for each hypothesis testing procedure. Therefore, in this block we will also describe the procedure to take the decision about the null hypothesis on the basis of critical value as well as p-value concepts.

9.9 RELATION BETWEEN CONFIDENCE INTERVAL AND TESTING OF HYPOTHESIS

In Units 7 and 8 of this course, we have learned about confidence intervals which were used to estimate the unknown population parameters with certain confidence. In Section 9.7, we have discussed the general procedure of testing a hypothesis which has been used in making decision about the specified/ assumed/ hypothetical values of population parameters. Both confidence interval and hypothesis testing have been used for different purposes but have been based on the same set of concepts. Therefore, there is an extremely close relationship between confidence interval and hypothesis testing.

In confidence interval, if we construct $(1-\alpha)$ 100% confidence interval for an unknown parameter then this interval contained all probable values for the parameter being estimated and relatively improbable values are not contained by this interval.

So this concept can also be used in hypothesis testing. For this, we contract an appropriate $(1-\alpha)$ 100% confidence interval for the parameter specified by the null hypothesis as we have discussed in Units 7 and 8 of this course and if the value of the parameter specified by the null hypothesis lies in this confident interval then we do not reject the null hypothesis and if this specified value does not lie in this confidence interval then we reject the null hypothesis.

Therefore, three cases may arise:

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

Case II: If we want to test the null hypothesis H_0: $\theta \leq \theta_0$ against the alternative hypothesis H_1: $\theta > \theta_0$ then we construct the lower one-sided confidence bound for parameter θ. If the value of the parameter specified by the null hypothesis i.e. θ_0 is greater than or equal to this lower bound then we do not reject the null hypothesis otherwise we reject the null hypothesis.

Case III: If we want to test the null hypothesis H_0: $\theta \geq \theta_0$ against the alternative hypothesis H_1: $\theta < \theta_0$ then we construct the upper one-sided confidence bound for parameter θ. If the value of the parameter specified by the null hypothesis i.e. θ_0 is less than or equal to this upper bound then we do not reject the null hypothesis otherwise we reject the null hypothesis.

For example, referring back to Example 4 of this unit, here we want to test the null hypothesis

$$H_0 : \mu \leq 50 \text{ gm against } H_1 : \mu > 50 \text{ gm}$$

This was tested with the help of test statistic

$$T = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \qquad [\text{Here}, \theta = \mu]$$

and we reject the null hypothesis at $\alpha = 0.01$.

This problem could also have been solved by obtaining confidence interval estimate of population mean which is described in Section 7.4 of Unit 7.

Here, we are given that

$$n = 50,\ \bar{X} = 50 \text{ and } \sigma = 5.1$$

Since alternative hypothesis is right-tailed, therefore, we construct lower one-sided confidence bound for population mean.

Since population variance is known so lower one-sided $(1-\alpha)$ 100 % confidence bound for population mean when population variance is known is given by

$$\bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

Since we test our null hypothesis at $\alpha = 0.01$ therefore, we contract 99% lower confidence bound and for $\alpha = 0.01$, we have $z_\alpha = z_{0.01} = 2.33$.

Thus, lower one-sided 99% confidence bound for average weight of potatoes is

$$52 - 2.33 \frac{5.1}{\sqrt{50}} = 52 - 1.68 = 50.32$$

Since the value of the parameter specified by the null hypothesis i.e. $\mu = 50$ is less than lower bound for average weight of potato so we reject the null hypothesis.

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

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

9.10 SUMMARY

In this unit, we have covered the following points:

1. Statistical hypothesis, null hypothesis, alternative hypothesis, simple & composite hypotheses.
2. Type-I and Type-II errors.
3. Critical region.
4. One-tailed and two-tailed tests.
5. General procedure of testing a hypothesis.
6. Level of significance.
7. Concept of p-value.
8. Relation between confidence interval and testing of hypothesis.

9.11 SOLUTIONS / ANSWERS

E1) Here, we wish to test the claim of the company that the average life of its car tyres is 50000 miles so

Claim: $\mu = 50000$ miles and complement: $\mu \neq 50000$ miles

Since claim contains equality sign so we take claim as the null hypothesis and complement as the alternative hypothesis i.e.

$H_0: \mu = 50000$ miles

$H_1: \mu \neq 50000$ miles

E2) (iii) Here, doctor wants to test whether new medicine is really more effective for controlling blood pressure than old medicine so

Claim: $\mu_1 > \mu_2$ and complement: $\mu_1 \leq \mu_2$

Since complement contains equality sign so we take complement as the null hypothesis and claim as the alternative hypothesis i.e.

$H_0: \mu_1 \leq \mu_2$

$H_1: \mu_1 > \mu_2$

(iv) Here, economist wants to test whether the variability in incomes differ in two populations so

Claim: $\sigma_1^2 \neq \sigma_2^2$ and complement: $\sigma_1^2 = \sigma_2^2$

Since complement contains equality sign so we take complement as the null hypothesis and claim as the alternative hypothesis i.e.

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

(v) Here, psychologist wants to test whether the proportion of literates between two groups of people is same so

Claim: $P_1 = P_2$ and complement: $P_1 \neq P_2$

Since claim contains equality sign so we take claim as the null hypothesis and complement as the alternative hypothesis i.e.

$H_0: P_1 = P_2$

$H_1: P_1 \neq P_2$

E3) Here, (i) and (vi) represent the simple hypotheses because these hypotheses tell us the exact values of parameter average weight of orange μ as $\mu = 100$ and $\mu = 130$.

The rest (ii), (iii), (iv), (v) and (vii) represent the composite hypotheses because these hypotheses do not tell us the exact values of parameter μ.

E4) Since alternative hypothesis $H_1: \theta \neq 60$ is two tailed so critical region lies in two-tails.

E5) Let A and B denote the number of white balls and black balls in the urn respectively. Further, let X be the number of white balls drawn among the two balls from the urn then we can take the null and alternative hypotheses as

$H_0: A = 4$ & $B = 2$ and $H_1: A = 2$ & $B = 4$

The critical region is given by

$$w = \{X : X < 2\} = \{X : X = 0,1\}$$

Thus,

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

$$= P[X \in w / H_0] = P[X = 0/H_0] + P[X = 1/H_0]$$

$$= \frac{{}^4C_0\ {}^2C_0}{{}^6C_2} + \frac{{}^4C_1\ {}^2C_1}{{}^6C_2} = \frac{1\times 1}{15} + \frac{4\times 2}{15} = \frac{1}{15} + \frac{8}{15}$$

$$\alpha = \frac{9}{15} = \frac{3}{5}$$

Similarly,

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

$$= P[X \notin w / H_1] = P[X = 2/H_1] = \frac{{}^2C_2\ {}^4C_0}{{}^6C_2} = \frac{1\times 1}{15} = \frac{1}{15}$$

E6) Since level of significance is the probability of type-I error so in this case level of significance is 0.05 or 5%.

E7) Here, the alternative hypothesis is two-tailed therefore, the test will be two-tailed test.

E8) Whether the test of testing a hypothesis is one-tailed or two-tailed depends on the alternative hypothesis. So correct option is (ii).

E9) First step in testing of hypothesis is to setup null and alternative hypotheses.