UNIT 16 HYPOTHESIS TESTS

Structure

16.1 INTRODUCTION

Many television commercials often make various performance statements. For example, consider the following statements :

(i) A particular brand of tyre will last an average of 40,000 miles before replacement is needed.

(ii) A certain detergent produces the cleanest wash.

(iii) Brand X of disposable nappies are stronger and are more absorbent.

How much confidence can one have in such statements? Can they be verified statistically? Fortunately, in many cases the answer is yes. In Unit 15, you have seen that samples are taken and partial bits of informations are gathered from the sample about such statements. In this unit we will study how to make a decision on whether to accept or to reject a statement on the basis of sample information. This process is called **hypothesis testing.**

In Unit 15, we have studied that population parameter can be estimated by using samples. These estimates are in turn used in arriving at a decision to either accept or reject the hypothesis. By a **hypothesis** we mean an assumption about one or more of the population parameters that will either be accepted or rejected on the basis of the information obtained from a sample.

In this unit we will discuss three types of tests : chi-square test, t-test and analysis of variance for testing a hypothesis. These tests are widely used in making decisions concerning problems in biology and in other fields. We shall not talk about the theories of these tests, but shall consider only their applications under simple situations. To apply these tests you should know what is meant by null hypothesis, level of significance and degrees of freedom. We shall begin with discussing these concepts in brief.

Objectives

After you have completed this unit, you should be able to

- select an appropriate test to analyse a given problem ;
- apply the chi-square test, t-test and analysis of variance ;
- substitute numerical data into selected formulas and solve the corresponding

Appendix

Random Numbers

03	47	43	73	86	36	96	47	36	61	46	98	63	71	62
97	74	24	67	62	42	81	14	57	20	42	53	32	37	32
16	76	62	27	60	56	50	26	71	07	32	90	79	78	53
12	56	85	99	26	96	96	68	27	31	05	03	72	93	15
55	59	56	35	64	38	54	82	46	22	31	62	43	09	90
16	22	77	94	39	49	54	43	54	82	17	37	93	23	78
84	42	17	53	31	57	24	55	06	88	77	04	74	47	67
63	01	63	78	59	16	95	55	67	19	98	10	50	71	75
33	21	12	34	29	78	64	56	07	82	52	42	07	44	38
57	60	86	32	44	09	47	27	96	54	49	17	46	09	62
18	18	07	92	46	44	17	16	58	09	79	83	86	19	62
26	62	38	97	75	84	16	07	44	99	83	11	46	32	24
23	42	40	64	74	82	97	77	77	81	07	45	32	14	08
52	36	28	19	95	50	92	26	11	97	00	56	76	31	38
37	85	94	35	12	83	39	50	08	30	42	34	07	96	88
70	29	17	12	13	40	33	20	38	26	13	89	51	03	74
56	62	18	37	35	96	83	50	87	75	97	12	25	93	47
99	49	57	22	77	88	42	95	45	72	16	64	36	16	00
16	08	15	04	72	33	27	14	34	09	45	59	34	68	49
31	16	93	32	43	50	27	89	87	19	20	15	37	00	49
68	34	30	13	70	55	74	30	77	40	44	22	78	84	26
74	57	25	65	76	59	29	97	68	60	71	91	38	67	54
27	42	37	86	53	48	55	90	65	72	96	57	69	36	10
00	39	68	29	61	66	37	32	20	30	77	84	57	03	29
29	94	98	94	24	68	49	69	10	82	53	75	91	93	30

16.2 STATISTICAL HYPOTHESIS

The testing of a statistical hypothesis is perhaps the most important area of statistical inference. In Sec. 16.1, we have already stated the meaning of a statistical hypothesis. We now give its formal definition.

Definition : A **statistical hypothesis** is an assertion or conjecture concerning one or more populations.

The truth or falsity of a statistical hypothesis is never known with certainty unless we examine the entire population. This, of course, is not practical in most situations. Instead, we take a random sample from the population of interest and use the information contained in this sample to decide whether the hypothesis is likely to be true or false. Evidence from the sample that is against the stated hypothesis leads to a rejection of the hypothesis, whereas evidence supporting the hypothesis leads to its acceptance. We should make it clear at this point that the acceptance of a statistical hypothesis is a result of insufficient evidence to reject it and does not necessarily imply that it is true.

Let us consider a very popular example of tossing a coin. When you toss a coin, only two events can occur—either you get a 'head' or a 'tail'. If you toss the coin 100 times, what will be the expected frequency for 'head' and the expected frequency for 'tail'? In other words, how many times out of 100 observations would you get a 'head'? Suppose, therefore, that you formulate the following hypothesis : "head will occur 50 times". What is the reason behind this prediction? The reason is : ideally, tossing of a coin is a random process, and therefore, the probability of getting a 'head' should be equal to the probability of getting a 'tail'. Now, when you do the actual tossing for 100 times, you know from experience that you may not get exactly 50 heads and 50 tails. There may be a small discrepancy from the ideal ratio of 50:50. This discrepancy, which is called chance variation, is an inherent character of any true random process (ref. Unit 11).

Let us now suppose that in a given situation the tossing of the coin is not a true random process. It may happen that the coin is defective and there is a bias for one of the two possibilities, say, 'head'. It may also happen that the person who is tossing the coin is not doing it correctly, but is applying a trick so that there is a bias for 'head'. In this non-random situation you will get a discrepancy from the ideal ratio of 50:50 when the coin is tossed 100 times. From common sense we can say that the discrepancy must be larger when the situation is non-random. If the bias is for 'head', then 'head' will come much more frequently than 'tail'. Therefore, the question is, how to decide, from the observed discrepancy, whether the discrepancy is large enough to reject the hypothesis or small enough to accept the hypothesis. We use the terms "accept" and "reject" frequently. It is important to understand that the rejection of a hypothesis is to conclude that it is false, while the acceptance of a hypothesis merely implies that we have no evidence to believe otherwise. Because of this terminology we usually state a hypothesis that we hope to reject. For example, to prove that one teaching technique is superior to another, we test the hypothesis that there is no difference in the two techniques.

A hypothesis formulated with the hope that it is most likely to be rejected is called the **null hypothesis.** Normally this term is applied to any hypothesis that we wish to test and is denoted by H_o. The rejection of H_o leads to the acceptance of an **alternative hypothesis,** denoted by H_1.

Before coming to different tests for testing a hypothesis we shall learn about two concepts: **level of significance** and **degrees of freedom.** These two concepts will be used in the discussion of all the tests that we will take up in this unit.

16.2.1 Level of Significance

We always start with the hypothesis that the process is random. Therefore, the observed discrepancy is due to chance variation. If the hypothesis is valid, then the discrepancy of the experimental value from the expected value should be small. If the hypothesis is not valid, then the discrepancy is large. How will we decide which value of discrepancy is large and which is small? In other words, how to take any decision to either accept or reject a null hypothesis?

Since any decision to either accept or reject a null hypothesis is to be made on the basis of information obtained from sample data, there is a chance that we will be making an error. There are two possible errors that we can make. We may reject a null hypothesis when we really should accept it. Alternately, we may accept a null hypothesis when we should reject it. These two errors are referred to as a **Type I** and a **Type II error,** respectively. In either case we make a wrong decision. We define these formally as follows :

Definition : A **Type-I error** is made when a true null hypothesis is rejected, that is, we reject a null hypothesis when we should accept it.

Definition : A **Type-II error** is made when a false null hypothesis is accepted, that is, we accept a null hypothesis when we should reject it.

In the following box we now show how these two errors are made :

		We claim that	
		H_o is true	H_o is false
If	H_o is true	correct decision (no error)	Type-I error
	H_o is false	Type-II error	correct decision (no error)

When deciding whether to accept or reject a null hypothesis, we always wish to minimise the probability of making a Type I error or a Type II error. The relationship between the probabilities of the two types of error is of such a nature that if we reduce the probability of making one type of error, we usually increase the probability of making the other type of error. In most practical problems one type of error is more serious than the other depending upon the nature of the problem. In such situations careful attention is given to the more serious error.

How much of a risk should we take in rejecting a true hypothesis, that is, in making a Type I error? In practice the limit of 0.05 or 0.01 is customary, although other values are also used. Each of these limits is called a **level of significance** or **significance level.** We have the following definition.

Definition : The **significance level** of a test is the probability that the test statistic falls within the rejection region when the null hypothesis is true.

In other words, we can say that the probability of committing a Type I error is called the **level of significance.** The **0.05 level of significance** is used when we want that the risk of rejecting a true null hypothesis should not exceed 0.05. The **0.01 level of significance** is used when we want that the risk of rejecting a true null hypothesis should not exceed 0.01.

If, for example, a 0.05 or 5% level of significance is chosen in designing a test of a hypothesis, then there are about 5 chances in 100 that we would reject the hypothesis when it should be accepted. i.e. we are about 95% confident that we have made the right decision. In such cases we say that the hypothesis has been rejected at a 0.05 level of significance, which means that we could be wrong with a probability of 0.05. Thus, we use the level of significance as a guide in decision making.

Remark : If the test statistic falls within the acceptance region, we do not reject the null hypothesis. When a null hypothesis is not rejected, this does not mean that the null hypothesis statement is guaranteed to be true. It simply means that on the basis of the information obtained from the sample data there is not enough evidence to reject the null hypothesis.

In actual situation the discrepancy value is calculated from the observed (experimental) and the expected (theoretical) values using the appropriate test formula. Then, the probability of getting this discrepancy is obtained from the available standard tables using the number of degrees of freedom involved. We shall be illustrating this procedure through examples. But, before that let us now discuss degrees of freedom.

16.2.2 Degrees of Freedom

To explain what degrees of freedom mean, let us consider the tossing of a coin again. In the tossing of a coin, there are two events that can occur; getting a 'head' or getting a 'tail'. These two events or possibilities are also called variables. Suppose, we have tossed the coin for a given number of times, say 50 times. Once this total number of observations is fixed, let us see how we can distribute the numbers between the two variables, 'head' and 'tail'. If we say, the number for head is 20, the number of 'tail' is automatically determined, and you know that the number is 30 for 'tail'. This means that between two variables, you can **freely** choose the number for one variable only. So we say that there is only one degree of freedom when there are two variables.

Let us take another simple example: suppose you have 30 pencils in all. How can you divide these 30 pencils among three children? You can, of course, do it in many ways. For example, you can give 10 to the first child, 5 to the second and then the third will get 15. But each time you divide these 30 pencils, you can freely choose the numbers for two children; once these numbers are chosen freely, the number of pencils for the third is automatically fixed. So you have two degrees of freedom when there are three possibilities (or variables). Notice that the number of degrees of freedom is always 1 less than the number of sample proportions that we are testing. As we proceed in this unit we will examine several examples of how to determine the degrees of freedom.

You may wonder why we have to consider the degrees of freedom. The reason is that in statistical analysis the degrees of freedom influence the extent to which the experimental results will deviate from the expected (theoretical) results. In other words, the degrees of freedom form one of the factors that influence the value of discrepancy between the observed and the expected results. So, for a given experiment, we need to determine the degrees of freedom.

Now let us look at some tests of hypothesis. We start with the chi-square test.

16.3 CHI-SQUARE TEST

We have been talking about measuring the discrepancy between the observed and the expected results. How to measure this discrepancy? One way of measuring this is the chi-square method. Chi-square (χ^2) is commonly used statistical procedure that is quite simple to apply. The formula for computing χ^2 is

χ is a Greek letter pronounced as 'kai'

$$\chi^2 = \sum \frac{(O - E)^2}{E},$$

where O = observed or experimental value,
E = expected value from the hypothesis.

Let us now go through a few examples which will help you in understanding how the χ^2-test is applied.

Example 1 : A coin is tossed 100 times. The head appeared 60 times and the tail appeared 40 times. Will you accept or reject the hypothesis that the tossing was normal (truly random)?

Solution : Since it is a question of accepting or rejecting a hypothesis, let us apply the chi-square test. We are given the observed frequencies for head and tail. We have to first determine the expected frequencies for the head and tail. The hypothesis is that the tossing was normal, that is, there was no bias for head or tail. In that case, both head and tail will have the same probability. So the expected frequency for head or for tail is 50. Therefore

		Head	Tail	Total
Observed value (O)	=	60	40	100
Expected value (E)	=	50	50	100
O − E	=	10	− 10	
$(O-E)^2$	=	100	100	
$\frac{(O-E)^2}{E}$	=	$\frac{100}{50} = 2.0$	$\frac{100}{50} = 2.0$	

There are two values of $\frac{(O-E)^2}{E}$ in this case.

Therefore, the sum is

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 2.0 + 2.0 = 4.0$$

This is the measure of discrepancy between the observed and the expected values. Having got this measure the next step is to draw conclusions regarding the hypothesis based on this measure. Here we make use of the level of significance and degrees of freedom for making decisions.

First we determine the degrees of freedom. Since there are only two variables, head and tail, the degree of freedom (D.F.) is

D.F. = 2− 1 = 1

Now we consult the χ^2 table (Appendix 1) which gives the probability of a given χ^2 with a given degree of freedom. In our problem χ^2 value is 4.0 and the degree of freedom is 1. So in the table look at the row where D.F. = 1. In that row if you go from left to right, you will see that the value 4.0 falls after 3.84 which corresponds to the probability of 0.05. Therefore, 4.0 has the probability of less than 0.05. Therefore, the discrepancy is large enough to reject the hypothesis. This means, the tossing is not normal. Thus, the difference between the frequencies of head and tail is not due to chance variation alone.

Once you conclude that the tossing is not normal, your exercise with statistics is over, but your detective search begins with the question : why is the tossing not normal? Such detective search (scientific investigation) may reveal interesting facts of life.

So let us see the sequence of steps we have followed in performing the χ^2-test.

Sequence of steps in Chi-square test

1) Determine the expected values from the hypothesis.
2) Make a table with the observed and the expected values.
3) Calculate the value of χ^2 .
4) Determine the degrees of freedom.
5) Determine, from the χ^2 table, the probability value of χ^2 obtained in Step 3 for the degrees of freedom determined in Step 4.
6) If the probability value obtained in Step 5 is less than 0.05 reject the hypothesis.

The χ^2-test has wide application in biology, especially in genetics. We will try to solve some biological problems using this test.

Example 2 : A garden pea plant is heterozygous (genetically mixed) for the gene pair Tt, where the gene T (for tall) is dominant over the gene t (for short). The plant produced 35 tall and 25 short offspring. Find out whether the plant was self-fertilised or fertilised by a short plant.

Solution : Let us begin with the hypothesis that the given plant is self-fertilised. This means that the cross is as follows:

	T	t
T	TT tall	Tt tall
t	tT tall	tt short
Result	1TT + 2Tt = 3 tall	1tt = 1 short

The above diagram shows that the offspring have 3 kinds of genotypes, TT, Tt and tt in the ratio of 1:2:1. But TT and Tt are tall plants; so the phenotypes are of two kinds: tall and short in the ratio of 3:1. So the expected ratio of tall:short is 3:1.

Since the total number of offspring is 60, the expected number of tall plants

$= \frac{60 \times 3}{4} = 45$ and the expected number of short plants $= 60 - 45 = 15$.

Now form a table :

	Tall	Short	Total
Observed value (O)	35	25	60
Expected value (E)	45	15	60
(O—E)	—10	+ 10	
$\frac{(O-E)^2}{E}$	$\frac{100}{45}$	$\frac{100}{15}$	

$$\therefore \chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{100}{45} + \frac{100}{15} = \frac{400}{45} = 8.88.$$

Now, the degrees of freedom = D.F. = 2 − 1 = 1.

Thus, we look at the χ^2 table and find that the probability of getting χ^2 value of 8.88 with 1 degree of freedom is less than 0.05. So the hypothesis is rejected at the 0.05 level of significance. That is, the plant is not self-fertilised.

Now let us assume the hypothesis that the plant is fertilised by a short plant. Since the genotype of a short plant is tt, the cross is as follows.

	T	t
t	tT	tt
t	tT	tt
Result	2 tall	2 short

The diagram shows that the offspring have two phenotypes, tall and short in the ratio of 2:2 = 1:1. This is the expected ratio assuming the hypothesis that the given plant is fertilised by a short plant.

Therefore, the expected number of tall plants $= \frac{60}{2} = 30$ and the expected number of short plant = 60 − 30. Now form a table.

	Tall	Short	Total
Observed value (O)	35	25	60
Expected value (E)	30	30	60
(O—E)	5	—5	
$(O-E)^2$	25	25	
$\frac{(O-E)^2}{E}$	$\frac{25}{30}$	$\frac{25}{30}$	

$$\therefore \quad \chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{25}{30} + \frac{25}{30} = 1.66.$$

The degrees of freedom = 2 — 1 = 1.

From χ^2 table the probability of getting χ^2 value of 1.66 with 1 degree of freedom is greater than 0.05. Therefore, the hypothesis is accepted at the .05 level of significance. That means, the given plant is fertilised by a short plant.

You may now try this exercise.

E 1) In an experiment Ramesh obtained the following results: He fertilised 7324 pea plants having smooth seeds. Among the offspring, 5474 plants produced smooth seeds and the rest 1850 plants produced wrinkled seeds. Perform the statistical analysis to determine whether the parents were heterozygous for this trait or not.

In the examples discussed so far, there were two variables. In the case of tossing a single coin, the variables (also called possibilities or classes) were 'head' and 'tail'. In the case of the offspring from the crosses Tt × Tt or Tt × tt, there were two phenotype cases, 'tall' and 'short'. In each of these cases the degree of freedom was one. But, there are situations where more than two variables are involved. Consequently, in such cases the degree of freedom is more than one. We shall now see how the χ^2-test can be applied in such cases for testing a hypothesis.

Chi-square Test with More Than One Degree of Freedom

We begin by considering some examples where three or more classes are scored in an experiment.

Example 3 : We tossed two identical coins simultaneously for 200 times. We got two heads 55 times, two tails 55 times, and one head and one tail 90 times. Does this result agree with the hypothesis that the tossing is random?

Solution : First we have to find out the expected values (frequencies) of the various classes following the hypothesis. Let us name the coins A and B. The tossing of coins A and B shows four possibilities : (1) two heads, (2) two tails, (3) head of A and tail of B, and (4) head of B and tail of A. Since the tossing is assumed to be random, the probability for each of these possibilities is one-fourth, that is $\frac{1}{4}$. However, since the coins are identical, the possibilities (3) and (4) are not distinguishable. Therefore, these two possibilities should be considered as one possibility, having the probability of $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Therefore,

expected frequency for two heads $= \frac{1}{4} \times 200 = 50$

expected frequency for two tails $= \frac{1}{4} \times 200 = 50$

expected frequency for head + tail $= \frac{1}{2} \times 200 = 100$

Now let us form a table to calculate the value of χ^2.

	Head	Tail	Head & Tail
Observed value (O)	55	55	90
Expected value (E)	50	50	100
(O − E)	5	5	−10
$(O - E)^2$	25	25	100
$\frac{(O-E)^2}{E}$	$\frac{25}{50} = \frac{1}{2}$	$\frac{25}{50} = \frac{1}{2}$	$\frac{100}{100} = 1$

Thus, $\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{1}{2} + \frac{1}{2} + 1 = 2.0.$

Now, the degrees of freedom = 3 − 1 = 2, because there are three variables.

The χ^2 table shows that the probability of getting a χ^2 value of 2.0 with 2 degrees of freedom is greater than 0.05. So the hypothesis is accepted at the .05 level of significance. In other words, the tossing is random, that is, the observed discrepancy from the expected values is due to chance variation only.

Let us consider another example.

Example 4 : Genetic theory states that children having one parent of blood type M and the other of blood type N will always be of one of the three types M, MN, N and that proportions of these types will, on an average be 1:2:1. A report states that out of 300 children having one M parent and one N parent, 30% were found to be of type M, 45% of type MN and the remainder of type N. Test the hypothesis by χ^2-test.

Solution : If the genetic theory be true, the expected distribution of 300 children should be as follows:

Children of type M = $\frac{300 \times 1}{1+2+1} = 75$

Children of type MN = $\frac{300 \times 2}{4} = 150$

Children of type N = $\frac{300 \times 1}{4} = 75$

The observed frequencies are

Type M = $\frac{300 \times 30}{100} = 90$,

type MN = $\frac{300 \times 45}{100} = 135$,

type N = $\frac{25 \times 300}{100} = 75$

Therefore, forming a table we have

	M	MN	N
Observed values (O)	90	135	75
Expected values (E)	75	150	75
(O−E)	15	− 15	0
$\frac{(O-E)^2}{E}$	$\frac{225}{75}$	$\frac{225}{150}$	0

Therefore, $\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{225}{75} + \frac{225}{150} + 0 = 4.5$

Since there are three different types, the degrees of freedom = 3 − 1 = 2. The value of $\chi^2 = 4.5$ for 2 degrees of freedom is greater than 0.05. So the hypothesis is accepted at the 0.05 level of significance.

How about doing an exercise now?

E 2) Radha made a cross between two pea plants, one having smooth-yellow seeds and the other having wrinkled-green seeds. She observed 208 seeds produced from this cross with the following results:

56 seeds were smooth-yellow (SY)

51 seeds were smooth-green (Sy)

49 seeds were wrinkled-yellow (sY)

52 seeds were wrinkled-green (sy)

Show with statistical analysis that these results fit the hypothesis that the genotype of the smooth-yellow parent was SsYy and the genotype of the wrinkled-green parent was ssYY.

There may be cases when you like to compare the means of two samples. For example, you want to know whether the average height of the adult male in country A is significantly different from that in country B. You may wish to determine whether the average weight of eggs laid by hens of type A is significantly different from the average weight of eggs laid down by hens of type B. A farmer may want to compare the rates of growth of a plant species grown in the presence of two different kinds of fertiliser. To solve these kinds of problems, we perform a test called a t-test. Let us now discuss it in detail.

16.4 t-TEST

This test was invented by a famous statistician, W.S. Gosset (1876-1937), who was a statistician for Guinness, an Irish brewing company. Gosset was the first to develop methods for interpreting information obtained from small samples. But, his company did not allow any of its employees to publish his/her articles. So, Gosset secretly published his findings in 1907 under the name "student". Because of this the t-test is also referred to as the student's t-test to this day. The t-test is used for comparing the means of two samples. Let us now see what the various steps involved in this t-test are:

Sequence of steps in t-Test

Let us consider two samples, whose means we want to compare. Let us denote the data in these samples by x_1 and x_2, the corresponding sample sizes by n_1, n_2 and the means of two populations by $\bar{x}_1$ and $\bar{x}_2$ respectively. We take the following steps to apply the t-test.

1) Calculate the sample means $\bar{x}_1$ and $\bar{x}_2$.

2) Calculate the standard deviations s_1 and s_2 of the two samples, given by

$$s_1 = \sqrt{\frac{\Sigma (x_1 - \bar{x}_1)^2}{n_1 - 1}}, \quad s_2 = \sqrt{\frac{\Sigma (x_2 - \bar{x}_2)^2}{n_2 - 1}}$$

3) Calculate the t-value using the equation

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

4) Calculate the number of degrees of freedom ν, by using the formula: $\nu = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

5) Using the table given in Appendix 2, find the probability value corresponding to the t-value obtained for the given degrees of freedom.

6) If the probability value obtained from the table is less than 0.05, it is concluded that the difference between the means $\bar{x}_1$ and $\bar{x}_2$ is statistically significant. That is, $\bar{x}_1$ is significantly different from $\bar{x}_2$ at the 0.05 level of significance.

   If the probability value obtained from the table is more 0.05, it is concluded that the difference between $\bar{x}_1$ and $\bar{x}_2$ is not statistically significant. That is, the observed difference between the two means is just due to chance variation.

Let us now do some examples which illustrate this procedure.

Example 5: A sample of 125 male birds of a given species had a mean weight of 93.3 gms. with a variance of 56.2 gms. A sample of 85 female birds of the same species had a mean weight of 88.8 gms. with a variance of 65.4 gms. Is there any real difference between the average weights of the male and the female birds?

Solution: Here, we are given $n_1 = 125$, $n_2 = 85$, $\bar{x}_1 = 93.3$, $\bar{x}_2 = 88.8$ and the corresponding variances are $s^2_1 = 56.2$ and $s^2_2 = 65.4$, respectively. Calculating the t-value we get,

$$t = \frac{93.3 - 88.8}{\sqrt{\frac{56.2}{125} + \frac{65.4}{85}}} = \frac{4.4}{1.1} = 4.0$$

Number of degrees of freedom $= (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$
$= 125 + 85 - 2 = 208.$

The table (ref. Appendix 2) shows that the probability P is less than 0.01 for $t = 4.0$ having degrees of freedom greater than or equal to 30. This means the probability that there is no difference is even less than 0.01. Thus, there is a real difference between the two mean weights. In other words, the difference is highly significant.

Let us now consider another example.

Example 6: Two sets of 50 plants were grown on two different fertilisers. The average height of one group of plants was 138 cm., with a standard deviation of 7 cm. For the other group, the average height was 145 cm. with a standard deviation of 10 cm.

(a) What is the probability that the observed difference is due to chance variation and not because of superiority of one fertiliser over the other? In other words, what is the probability that there is no significant difference between the two averages?

(b) Would you accept the hypothesis that one fertiliser is better than the other in terms of the plant growth?

Solution: We assume that the observed difference is due to chance variation only. It is not because of superiority of one fertiliser over the other. In this case we have $\bar{x}_1 = 145$, $\bar{x}_2 = 138$, $s_1 = 10$, $s_2 = 7$ and $n_1 = n_2 = 50$. Then the t-value is given by

$$t = \frac{145 - 138}{\sqrt{\frac{10^2}{50} + \frac{7^2}{50}}} = \frac{7}{\sqrt{2.98}} = \frac{7}{1.72} = 4.06$$

The number of degrees of freedom $= (50 - 1) + (50 - 1) = 98.$

(a) The table (ref. Appendix 2) shows that the probability of getting a t-value of 4.06 with 98 degrees of freedom is less than 0.01.

(b) Since the above mentioned probability of 0.01 is less than 0.05 the null hypothesis is rejected. That is, the observed difference in the mean values is not due to chance variation. So we accept that one fertiliser is better than the other.

You may now try the following exercise:

E 3) A biscuit manufacturer wishes to compare the performance of two biscuit production lines, A and B. The lines produce packets of biscuits with a nominal weight of 300 grams. Two random samples of 15 packets each from the two lines are weighed (in grams). The sample data is summarised as follows:
Line A mean $\bar{x}_1 = 309.8$; standard deviation $s_1 = 3.5$
Line B mean $\bar{x}_2 = 305.2$; standard deviation $s_2 = 7.0$

Carry out a t-test on this sample data to test whether the two production lines produce packets of different weights.

So far we have considered two sets of data arising from two different populations, as in the case of a population of male birds and a population of female birds, or growth of plants with different fertilisers. Sometimes it may happen that we have to compare two sets of data arising from the same population under two different conditions. In such conditions, measurements are made on each subject twice—once under control conditions and once under experimental conditions. Such measurements are known as **paired measurements.** For example, temperature of a patient measured before and after taking a particular drug. We shall now see how t-test is performed for such paired data.

t-test for paired data

Pairing of samples is always desirable because it helps in eliminating variations between samples arising due to factors other than the factor in question. To find the t-value in case of paired data the following formula is used

$$t = \frac{\bar{d}/s}{\sqrt{N}}$$

where, d = difference between each pair of values
s = standard deviation of the difference
N = number of pairs measured
$\bar{d} = \sum d/N$ = mean of the difference between paired values
The sign, positive or negative, of the difference between each pair should be considered.

For this formula to be meaningful N should be large.

To have a clear picture about it let us consider an example.

Example 7: Suppose we want to see the effect of a drug on blood pressure. Ten subjects are chosen and the blood pressure is measured for each subject before and after the administration of the drug. The result is shown in the second and the third columns of Table 1. Does the drug have any significant effect on blood pressure?

Table 1

Subject	Blood Pressure		Difference	
	before	after	d	d^2
1st	118	127	9	81
2nd	113	121	8	64
3rd	128	136	8	64
4th	124	131	7	49
5th	136	138	2	4
6th	130	132	2	4
7th	140	141	1	1
8th	130	131	1	1
9th	140	132	− 8	64
10th	128	120	− 8	64
			22	396

Solution : We assume that the drug has no significant effect.

Thus, the mean difference $= \bar{d} = \frac{22}{10} = 2.2$

The standard deviation of the differences (i.e. of the d values)

$$s = \sqrt{\frac{N\Sigma\, d^2 - (\Sigma\, d)^2}{N\,(N-1)}} = \sqrt{\frac{(10 \times 396) - (22)^2}{10\,(10-1)}} = 5.89$$

So,

$$t = \frac{\frac{\bar{d}}{s}}{\sqrt{N}} = \frac{\frac{2.20}{5.89}}{\sqrt{10}}$$

$$= 1.179$$

The degrees of freedom = 10 − 1 = 9.
With t = 1.179 and 9 degrees of freedom, the probability P is greater than .2 (from Appendix 2). Considering a level of .05 significance, this value of P suggests that our hypothesis is true. That is, the drug has no significant effect.

You may now try this exercise.

E 4) A study was conducted to test the effects of growth hormone on the rate of growth of 10 children. Growth rates were measured before and after the subjects were given growth hormone three times a week for an year. Based on the data given below, does the sample show a significant difference in the growth rate?

Subject	Before treatment	After treatment
1	3.4	4.5
2	3.6	5.2
3	4.8	6.5
4	3.4	5.2
5	4.8	7.4
6	5.8	8.9
7	4.2	8.4
8	5.7	8.5
9	4.1	7.5
10	4.3	8.2

We shall now discuss a generalised form of the t-test, namely the analysis of variance.

16.5 THE ANALYSIS OF VARIANCE

The analysis of variance is a statistical technique which can be used to make comparisons among more than two groups. For instance, suppose we want to see the effects of three different fertilisers on the yield of a crop. We select one hundred identical plots of which 25 are control plots and 75 are experimental plots. Plants are grown in all the 100 plots in the same manner except that no fertiliser is applied in the control plots, while the three fertilisers are applied in the 75 experimental plots. The 75 experimental plots are divided into 3 groups of 25 plots each. Each of these groups is treated with different kinds of fertiliser. When the crop is harvested, the yield of each individual plot is recorded in four groups. We calculate the mean yield of all the 100 plots. Then we apply the technique (analysis of variance) to find out if there is any significant difference between the overall mean yield and the mean yields of the individual groups.

As another example, suppose we want to test the effectiveness of various doses of a drug on blood pressure. We take a large number of individuals and divide them into several groups depending on the number of doses. For example, if we want to test four different doses, we make five groups (one control group and four experimental groups). Each experimental group is meant for a given dose of the drug.

There are various ways of applying the technique of the analysis of variance. These are known as one-way, two-way, or three-way, ANOVA (ANOVA is the abbreviated form of ANalysis Of VAriance). One-way ANOVA is used to see the effects of a **single factor** (independent variable) on a given dependent variable. In case of the fertiliser experiment described above, we only wanted to test the effects of fertilisers on the **yield** (dependent variable). Two-way ANOVA is used to see the effects of two independent variables on a given dependent variable. For example, we would use this test if we were interested in examining the effects of three fertilisers at three different levels of irrigation on the yield of a given crop. Here the two independent factors would be **fertilisers** and **irrigation.** We will only discuss about one-way ANOVA, which is the simplest one.

In the analysis of variance, as the name suggests, we essentially compare the variations existing in the observed values of the dependent variable in the classified groups. We already know that one measure of variation is **variance.** Variance is calculated as the average of the squares of the deviations of each of the observed values from their mean value. It gives us the total variation of the group containing the observed values. In the analysis of variance, which we shall be explaining very soon, we calculate the variance of the **total** population, the variance **between** the groups, and the variance **within** the groups. Then, we examine whether there is any significant difference between the calculated variances. As in most statistical techniques, we assume the null hypothesis that the means of all the groups are not different from the mean of the total population. If we find, by ANOVA, that the variance between the groups is not significantly different from the variance within the groups, we conclude that there is no significant difference between the means of the groups, and hence the independent variable (factor) has no effect on the dependent variable. Before we go into the actual analysis of variance, you may try the following exercises.

E 5) Suppose you want to see the effects of the hormone glucagon on the blood sugar level.

(a) How will you design the experiment?

(b) Which is the independent variable and which is the dependent variable?

E 6) Suppose you want to see the effects of glucagon on the blood sugar level of people maintained on two different diets.

(a) How will you design the experiment?

(b) Point out the independent and dependent variables.

Let us now discuss the analysis of one-way ANOVA.

One-way ANOVA

In actual analysis we will calculate the variance between the groups and the variance within the groups. Then we will calculate the so-called **F-value** by taking the ratio of these two variances. Then the value of F will be interpreted with reference to a table (Appendix 3) which gives the value of F for different degrees of freedom and at various levels of significance. If our calculated value of F is higher than the tabular value of F for the appropriate degrees of freedom at the selected level of significance, we will conclude that the difference between the variances within groups and between groups is significant. That is, the independent variable (factor) is effective. Let us now consider an example.

Example 8 : In an experiment with three groups of animals, the Group 1 animals were injected with distilled water (control group), the Group 2 animals with 50 units of a drug, and the Group 3 animals with 100 units of the same drug. The blood sugar levels (in mg/100 ml of blood) were estimated one hour after the injection, and the values obtained are given by the three columns of Table 2. Apply ANOVA to find out whether the mean values of the three groups differ significantly.

Table 2

Group 1	Group 2	Group 3
76	100	108
83	82	104
85	95	109
75	90	110
74	98	112
80	89	106
82	92	108
77	90	112

Solution : Step 1 : We first calculate the mean of each group and mean of the total population.

$$\text{Mean of Group 1} = \bar{x}_1 = \frac{76 + 83 + 85 + 75 + 74 + 80 + 82 + 77}{8}$$

$$= \frac{632}{8} = 79.0$$

In this operation we have actually used a formula $\bar{x}_1 = \frac{\Sigma x_1}{n}$

where Σx_1 is the sum of the x_1 values, that is, the values of Group 1; n_1 is the number of observation in Group 1.

Similarly, the mean of Group 2 is

$$\bar{x}_2 = \frac{\Sigma x_2}{n_2} = \frac{100 + 82 + 95 + 90 + 98 + 89 + 92 + 90}{8} = \frac{736}{8} = 92.0$$

The mean of Group 3 is

$$\bar{x}_3 = \frac{\Sigma x_3}{n_3} = \frac{108 + 104 + 109 + 110 + 112 + 106 + 108 + 112}{8}$$

$$= \frac{872}{8} = 109.0$$

So, the mean of the total population is

$$\mu = \frac{\Sigma x_1 + \Sigma x_2 + \Sigma x_3}{n_1 + n_2 + n_3} = \frac{632 + 736 + 872}{24} = 93.33.$$

Step II : We then find the variance within the groups. For this we calculate the squares of the deviations of each observation from the mean of the respective groups. Then we take the sum of these squared deviations. In the following table we have done it only for Group 1.

x_1	Deviation from $\bar{x}_1$ $(= \bar{x}_1 - x_1)$	$(\bar{x}_1 - x_1)^2$
76	+ 3	9
83	− 4	16
85	− 6	36
75	+ 4	16
74	+ 5	25
80	− 1	1
82	− 3	9
77	+ 2	4
		$\Sigma (\bar{x}_1 - x_1)^2 = 116$

We follow the same procedure for Group 2 and Group 3.

E 7) Obtain $\Sigma (\bar{x}_2 - x_2)^2$ and $\Sigma (\bar{x}_3 - x_3)^2$.

If you have done E 7, you will find that $\Sigma (\bar{x}_2 - x_2)^2 = 226$ and $\Sigma (\bar{x}_3 - x_3)^2 = 70$.

Therefore, the total of squares $= \sum_{i=1}^{3} (\bar{x}_i - x_i)^2$

$$= \Sigma (\bar{x}_1 - x_1)^2 + \Sigma (\bar{x}_2 - x_2)^2 + \Sigma (\bar{x}_3 - x_3)^2$$
$$= 116 + 226 + 70 = 412.$$

Now we find the degrees of freedom in this case. Since there are a total of 24 observations and 3 groups, the number of degrees of freedom in this case D.F. = 24 − 3 = 21.

Now, the formula for the variance within groups is

$$\sigma_\omega^2 = \frac{\text{Total squares}}{\text{Degrees of freedom}} = \frac{412}{21} = 19.62.$$

Step III: Now we calculate the variance between groups. For this we calculate the deviation of each group mean from the population mean μ, that is, the values of $(\mu - \bar{x}_1)$, $(\mu - \bar{x}_2)$, and $(\mu - \bar{x}_3)$. Then, we square these deviations and multiply these squared deviations by the number of observations in the respective groups. That is, we find the values $n_1 (\mu - \bar{x}_1)^2$, $n_2 (\mu - \bar{x}_2)^2$, $n_3 (\mu - \bar{x}_3)^2$. Let us perform these calculations for Group 1.

For Group 1, $\bar{x}_1 = 79.0$ and $n_1 = 8$.
Therefore, $\mu - \bar{x}_1 = 93.33 - 79.00 = 14.33$ and $(\mu - \bar{x}_1)^2 = 205.35$
Thus, $n_1 (\mu - \bar{x}_1)^2 = 8 \times 205.35 = 1642.80$

Similar calculations you can perform for Group 2 and Group 3.

E 8) Find out the values of $n_2 (\mu - \bar{x}_2)^2$ and $n_3 (\mu - \bar{x}_3)^2$ for Groups 2 and 3.

On solving E 8) you would have found that $n_2 (\mu - \bar{x}_2)^2 = 14.16$ and $n_3 (\mu - \bar{x}_3)^2 = 1964.40$. We add these values and get the sum of squares as

$$\sum_{i=1}^{3} n_i (\mu - \bar{x}_i)^2 = n_1 (\mu - \bar{x}_1)^2 + n_2 (\mu - \bar{x}_2)^2 + n_3 (\mu - \bar{x}_3)^2$$

$$= 1642.80 + 14.16 + 1964.40$$
$$= 3621.36$$

The degrees of freedom in this case is
D.F. = (No of groups) − 1 = 3 − 1 = 2.

We now calculate the variance between groups

The variance between groups is σ_b^2

$$\sigma_b^2 = \frac{n_1 (\mu - \bar{x}_1)^2 + n_2(\mu - \bar{x}_2)^2 + n_3 (\mu - \bar{x}_3)^2}{\text{Number of degrees of freedom}}$$

$$= \frac{3621.36}{2} = 1810.68.$$

Let us summarise our results in Table 3.

Table-3 of ANOVA

Sources of variation	Sum of squares	D.F.	Variance = $\frac{\text{sum of squares}}{\text{D.F.}}$
Between groups	3621.36	2	1810.68
Within groups	412.00	21	19.62
Total	4033.36	23	1830.30

Step IV: We now obtain the value of F.

The value of F is obtained by using the formula

$$F = \frac{\text{variance between groups}}{\text{variance within groups}} = \frac{\sigma_b^2}{\sigma_\omega^2}$$

$$= \frac{1810.68}{19.62} = 92.29.$$

So, we have obtained the value of F in a situation where the D.F. between groups is 2 and the D.F. within groups is 21. To associate the two degrees of freedom with the value of F, we write

$F_{(2,21)} = 92.29$

where the first number (that is 2) in the subscript of F denotes the degrees of freedom of the variance between groups, and the second number (that is 21) denotes the degrees of freedom of the variance within groups. With the value of F in the hand, we now give you a method of looking at ANOVA table and interpret the result.

ANOVA table and interpretation of the result

Now we have to choose the level of significance. Let us fix it this time at .01. Now, we are in a position to consult the table in Appendix 3 to find the critical values of F for a given set of degrees of freedom and at a selected level of significance.

The numbers in the first vertical column of the table are the values of degrees of freedom for the **lesser mean square**. Almost always the lesser mean square comes from the variation **within** groups, and the greater mean square comes from variation **between** groups, as it has happened in our example (see Table 3). Therefore, you go down the first column of the table and stop when you reach the number 21, because 21 indicates the degrees of freedom for the lesser mean square. From here you go across the table, that is, go towards your right to a place for degrees of freedom 2. There you find two values: one is 3.47 and below that is 5.78 which is printed in boldface. And 3.47 is the critical value of F at the .05 level of significance, so we ignore it. On the other hand 5.78 is the critical value of F at .01 level of significance, the value we want.

Now what do you do with this critical value of F? We find that the observed F-value is greater than the critical F value, that is, we can conclude that there is a significant difference among the means of the three groups. Thus, the observed differences among these means do not arise from chance variation. In other words, the drug has a significant effect on blood sugar level.

And now exercises for you.

E 9) From the data given below calculate the following:

a) The mean of each group and the mean of the total population.

b) The variance within groups.

c) The variance between groups.

d) The F value.

Group 1	Group 2	Group 3
32	42	29
33	37	32
35	36	30
34	41	31
37	40	25
36	39	27

E 10) From the data given below calculate the following:

a) The mean of each group and the mean of the total population.

b) The variance within groups.

c) The variance between groups.

c) The F-value.

e) The critical value of F at 0.01 level of significance.

f) On the basis of values obtained in (d) and (e) what will you conclude about the difference between the two variances in (b) and (c).

Group 1	Group 2	Group 3
80	56	97
73	72	90
79	61	75
88	64	87
68	80	88
75	74	83

16.6 SUMMARY

We conclude this unit by giving a summary of what we have covered in it.

1) The meaning of a statistical hypothesis.
2) The definition of a level of significance of a hypothesis.
3) What we mean by the degrees of freedom of a given sample.
4) The χ^2-test is performed to measure the deviation of the observed values from the expected values.
5) How to apply the χ^2-test.
6) The t-test is performed to decide whether the difference between the means of two samples is significant or just due to chance variations.
7) How to apply the t-test.

8) The analysis of variance (ANOVA) is performed to find out whether the means of the groups are significantly different from the population mean.

9) In ANOVA the deviation is measured in terms of F-value.

10) Various steps involved in 1-way ANOVA are:

    i) if the data is not grouped, divide it into appropriate groups.

    ii) calculate the mean of each group and the mean of the total population.

    iii) calculate variance within groups.

    iv) calculate variance between groups.

    v) summarise the results in ANOVA table.

    vi) calculate the value of F.

    vii) consult the standard table of F to find the critical value of F for the given set of degree of freedom and at the selected level of significance.

    viii) compare the observed F value with the critical F value to conclude whether there is significant difference between the means of the groups.

16.7 SOLUTIONS/ANSWERS

E 1) When the parents are heterozygous, the genotypes of the offspring plants are as follows:

	S	s
S	SS	Ss
s	sS	ss

Of these, SS, Ss and sS plants produce smooth seeds and ss plants produce wrinkled seeds. So, the ratio of smooth to wrinkled is expected to be 3 : 1.

So, out of 7324 offspring plants, $7324 \times \frac{3}{4} = 5493$ are expected to produce smooth seeds and $7324 - 5493 = 1831$ are expected to produce wrinkled seeds.

	Smooth	Wrinkled	Total
Observed value (O)	5474	1850	7324
Expected value (E)	5493	1831	7324
(O−E)	− 19	+ 19	
$(O-E)^2$	361	361	
$\frac{(O-E)^2}{E}$	0.065	0.197	

$\chi^2 = 0.065 + 0.197 = 0.262$

Degrees of freedom $= 2 - 1 = 1$.

From the χ^2 table (Appendix 2) the probability of getting χ^2 value of 0.262 with one degree of freedom is greater than 0.05. Therefore, the hypothesis is accepted at the level of .05 significance. That is, the parents are heterozygous.

E 2) From a cross SsYy × ssyy, the genotypes of the offspring will be

SY	Sy	sY	sy
SsYy smooth yellow	Ssyy smooth green	ssYy wrinkled yellow	ssyy wrinkled green

Thus, four phenotypic combinations of offspring are expected in equal proportion, that is, $\frac{208}{4} = 52$ for each type.

	Smooth yellow	Smooth green	Wrinkled yellow	Wrinkled green
Observed value (O)	56	51	49	52
Expected value (E)	52	52	52	52
O-E	4	−1	−3	0
$(O-E)^2$	16	1	9	0
$\frac{(O-E)^2}{E}$	0.307	0.019	0.173	0

$\chi^2 = 0.307 + 0.019 + 0.173 + 0 = 0.499$

Degrees of freedom = 4 − 1 = 3.

The χ^2 table (Appendix 1) shows that the probability of getting a χ^2 value of 0.499 with one degree of freedom is greater than 0.05. So, the hypothesis is accepted at 0.05 level of significance.

E 3) The value of the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{309.8 - 305.2}{\sqrt{\frac{(3.5)^2}{15} + \frac{(7.0)^2}{15}}}$$

$$= \frac{4.6}{\sqrt{4.08}} = \frac{4.6}{2.02} \approx 2.27.$$

The table (Appendix 2) shows that the probability of getting a t-value of 2.27 with 28 degrees of freedom is less than 0.05. Thus, the hypothesis is rejected. That is, the two production lines do appear to produce packets of different weights.

E 4) $\bar{d} = 2.62$, $\Sigma d = 26.20$, $\Sigma\, d^2 = 78.52$
$s = 1.05$, $t = 7.91$
degrees of freedom = 9.
Reject the hypothesis with $P < 0.05$.

E 5) a) The experiment can be designed in many ways. One simple way is to form four groups of subjects, one control group to which no glucagon is administered, and three experimental groups to which glucagon is administered at three different doses. Then the level of blood sugar is measured in each individual in each of the four groups. The data can be analysed using ANOVA.

b) The dose of glucagon is the independent variable and the blood sugar level is the dependent variable.

E 6) a) The experiment can be designed in many ways. One simple way is to form eight groups — four groups for one diet and four groups for the other diet. Out of four groups for a given diet, one group is not administered with glucagon (control group) and the three groups are administered with three different doses of glucagon. Similar design is made for the other four groups but with the other diet. Blood sugar level is measured in each individual in all the eight groups, and the data is analysed using ANOVA.

b) The independent variables are diet and glucagon. The dependent variable is blood sugar level.

E 7) $\Sigma\,(\bar{x}_2 - x_2)^2 = 226$, $\Sigma\,(\bar{x}_3 - x_3)^2 = 70$

E 8) $n_2\,(\mu - \bar{x}_2)^2 = 14.16$, $n_3\,(\mu - \bar{x}_3)^2 = 1964.40$

E 9) a) $\bar{x}_1 = 34.5$, $\bar{x}_2 = 39.1$, $\bar{x}_3 = 29$

b) Total of the sum of the squares of deviation =
$\Sigma\,(\bar{x}_1 - x_1)^2 + \Sigma\,(\bar{x}_2 - x_2)^2 + \Sigma\,(\bar{x}_3 - x_3)^2$
$= 17.50 + 26.86 + 34.00 = 78.36$

Degrees of freedom (D.F) = 18 − 3 = 15 (because there is a total of 18 observations and 3 groups).

$$\therefore \text{Variance within groups} = \sigma_\omega^2 = \frac{\text{Total squares}}{\text{DF}}$$

$$= \frac{78.36}{15} = 5.22$$

c) Mean of the total population $\mu = 34.2$

$\therefore n_1 (\mu - \bar{x}_1)^2 + n_2 (\mu - \bar{x}_2)^2 + n_3 (\mu - \bar{x}_3)^3$

$= 6\,(34.2 - 34.5)^2 + 6\,(34.2 - 39.1)^2 + 6\,(34.2 - 29)^2$

$= 0.54 + 144.06 + 162.24 = 306.84$

Degrees of freedom = (no. of groups) − 1 = 3 − 1 = 2

$$\therefore \text{Variance between groups} = \frac{306.84}{2}$$

$\sigma_b^2 = 153.42.$

d) $$\text{F value} = \frac{\text{Variance between groups}}{\text{Variance within groups}} = \frac{\sigma_b^2}{\sigma_\omega^2} = \frac{153.42}{5.22} = 29.39.$$

E 10) a) $\bar{x}_1 = 77.17$, $\bar{x}_2 = 67.83$, $\bar{x}_3 = 86.67$

$\mu = 77.22$

b) $\text{Variance within groups} = \dfrac{909}{15} = 60.60$

c) $\text{Variance between groups} = \dfrac{1064.11}{2} = 532.06$

d) F-value = 8.78

e) Critical value of F at 0.01 level of significance = 6.36

f) Difference is significant