UNIT 6 TESTS OF SIGNIFICANCE

Structure **Page No.**

6.1 INTRODUCTION

Statistical data are collected in many diverse fields. The main purpose of data collection is to arrive at certain decisions. For example, a company may want to decide whether or not to accept a consignment of ball bearings. A student may want to know whether or not to believe the claim of a coaching class owner that students in his class get 90% marks on the average in Maths. A doctor may need to test whether or not to prefer the new pain medicine to the old.

In all these situations one can identify a random variable whose distribution is not known to the decision maker, but will be useful for decision making. The company may identify a quality characteristic associated with each ball bearing and the company may want to know if the quality characteristic associated with a great majority of the ball bearings in the consignment satisfy the upper and lower specifications on it. The student wants to know if the expected value of the random variable, namely the score in Mathematics of a student of the coaching centre is 90% or not. The doctor wants to know whether the expected value of the random variable, which is the 'duration of relief' for the new pain medicine is more than that for the old medicine or not. In each of the above cases the decision has to be based on the information obtained from a random sample. In the previous unit you have seen how we can get a point and an interval estimates of parameters of a distribution by observing a finite number of realizations of the random variable. In this unit we shall see how a hunch or a claim or a hypothesis about the actual value of a parameter can be tested and a decision taken to reject or accept the hypothesis.

Let us consider again the random variables that were of interest in the above mentioned situations. For the company that has received a consignment of ball bearings, the quality characteristic of interest may be its thickness. The thickness of a ball bearing can be measured on a continuous scale and if $X_1, X_2, \ldots X_n$ denote n realizations (the thickness of n randomly chosen ball bearings, from a very large consignment),it is easy to see that they are a set of independent and identically distributed random variables. It is reasonable to assume that the common distribution is the normal distribution with some unknown mean μ and variance σ^2. The student can reasonably assume that the scores in Mathematics of a random sample of size n from among the alumni of the coaching centre are independent and identically distributed as normal with mean μ and variance σ^2. Similarly, the time durations of relief reported by n_1 randomly chosen test patients using one dose of the new pain medicine can be modelled as n independent and identically distributed random variables distributed as normal(μ_1, σ_1) and those of another set of randomly chosen n_2 patients using one dose of the old medicine as i.i.d. normal(μ_2, σ_2^2). We shall deal with the problem of testing hypotheses about the mean and variance of normally distributed random variables in sections 6.3 and 6.4. In Sec. 6.5 we shall illustrate the problem of testing hypothesis about the proportion.

Objectives

After reading this unit, you should be able to

- formulate null and alternative hypothesis for a given problem,
- describe Type 1 and Type 2 errors,
- differentiate between one sided alternatives and 2-sided alternatives,
- test whether μ has the specified value μ_0 against either a one sided alternative or against a two sided alternative assuming that we have n independent observations on a random variable that is distributed as normal(μ, σ) where σ is either assumed known or is unknown,
- test whether the two normal populations have the same mean or not,
- test whether the population variance from a normally distributed population, has a specified value against either a one sided or two sided alternative,
- test if the population proportion is significantly different from the hypothetical population proportion,
- test if the difference between two population proportions is zero or not based on the observed sample proportions.

6.2 SOME BASIC CONCEPTS

Let us begin with a problem.

A company manufacturing Aluminium rods wants to find the average mean diameter of the length of rods manufactured in the company. Based on previous experience, the firm's statistician knew that, on the average, the rods should be 2 cm. in diameter. He also knew that the variance of the diameters of the rods produced by the process is $\sigma_0^2 = 0.25$ A test was needed for detecting any change in this mean diameter every day so that whatever factors were responsible for such a change could be corrected.

Note that in this situation the statistician is interested in making inference about the population parameter mean μ. However, she is not interested in estimating the value of μ, rather she is interested in testing a hypothesis about the value of μ. The hypothesis is that the mean diameter of the rods produced on a particular day is 2cm. (That is there has been no change in the mean diameter (μ) of all the rods produced in a day' from the established standard length) Let us denote this hypothesis by H_0, and call it the null (no change) hypothesis. The statistician wishes to test this null hypothesis against the alternative that it is not true.

In statistics, the **null hypothesis is the hypothesis that is being tested**. A statistician has to be careful while formulating the null and alternative hypotheses since it is presumed that the null hypothesis holds unless there is a strong statistical evidence obtained from the sample of observations against it and in favour of the alternative hypothesis. The situation is somewhat similar to the modern system of justice. A person accused of a crime is presumed to be not guilty, unless the prosecution can produce strong evidence to the contrary. ~~To make the concept clear, let me pose a question. A~~

strong evidence ~~to the contrary~~. To make the concept clear, let me pose a question. A manufacturer of paints wants to use a new drying process. The present process requires a drying time of 15 minutes and it is claimed that the new process requires only 12 minutes of drying. The new process is to be tested on n sample pieces. What should be the null and alternative hypothesis if the manufacturer wants to continue to use the old process unless there is strong evidence for the claim regarding the new process? In this case you can see that the null hypothesis is $H_0 : \mu = \mu_0 = 15$ minutes and the alternative hypothesis is $H_1 : \mu = \mu_1 = 12$ minutes, where μ is the mean drying time of the new process. The new process is not accepted unless the null hypothesis is rejected. This is the case of a conservative manufacturer. A more risk taking manufacturer may decide that the new process may be accepted unless there is statistical evidence against the claim. In this case the null hypothesis is formulated as $H_0 : \mu = \mu_1 = 12$ minutes against the alternative hypothesis $H_1 : \mu = \mu_0 = 15$ minutes. Unless the null hypothesis in this formulation is rejected, the manufacturer goes for the new process.

We now continue with the case of the company manufacturing aluminium rods. The statistician is investigating if the diameter of a rod produced today is distributed as normal with a mean that is different from 2.0 cm. Unless there is strong statistical evidence for such a change, the statistician will not like to reject the hypothesis that $\mu = 2$ as that might entail stopping the process. If such a change is absent, the mean diameter is 2 cm. Thus the null hypothesis H_0 in this case is

$$H_0 : \text{The mean diameter is 2 cm.}$$

Now to make any decision, the company have to choose between this hypothesis and the alternative hypothesis, which we denote H_1, and this is stated as

$$H_1 : \text{The mean diameter of the given lot of rods is not 2 cm.}$$

As the name suggests this hypothesis is alternate to H_0.

Thus the statistician has made a claim/hypothesis and she wants to test this claim against a suitable alternative hypothesis on the basis of a sample of observations.

Suppose $X_1, X_2 \ldots X_n$ are i.i.d. (independently and identically distributed) random variables having the common distribution function $F(\theta)$ where θ is a real valued parameter. The set of all possible values of θ is known as the parameter space and is denoted as Θ. We wish to test the hypothesis that $\theta \in \Theta_1$ against the alternative that $\theta \in \Theta_2$, where Θ_1 and Θ_2 are non-intersecting subsets of Θ. In the above example, the statistician will observe from the day's production a random sample of n rods. Let $X_1, \ldots X_n$ denote their diameters. Clearly, $X_1, X_2 \ldots X_n$ are i.i.d random variables. The statistician also assumes that their common distribution is normal with mean μ and variance σ_0^2, where σ_0^2 is assumed to be 0.25 The parameter space Θ may be taken as $(-\infty, \infty)$ although the mean diameter can never be negative. Also we take $\Theta_1 = \{2.0\}$ and $\Theta_2 = \Theta \setminus \{2.0\}$.

We call the null hypothesis *simple* if Θ_1 is a singleton set. Otherwise we call it a *composite hypothesis*. Similarly, the alternative hypothesis is called simple if the set Θ_2 is a singleton set; else it is called a composite hypothesis. The testing problem of the company manufacturing aluminium rods has a simple null hypothesis against a two sided (i.e., both $\mu > \mu_0$ and $\mu < \mu_0$) composite alternative hypothesis.

In any hypothesis testing problem such as the above you may easily recognise four possibilities, two for the true value of the hypothesis H_0 and two possibilities for the outcome of any test procedure. It is therefore clear that there can be two kinds of errors of judgement. First, one can reject the null hypothesis when it is true. Second, one can fail to reject the null hypothesis when it is false. These two kinds of error are called Type I and Type II errors and are defined as follows:

when it is true. A Type II error occurs if the null hypothesis is not rejected (or accepted) when it is false.

The four possibilities are described in the table below:

Table I

Two possible outcomes for H_0	Decision taken	
	Do Not reject H_0	Reject H_0
H_0 is true	Correct decision	Type I error
H_0 is not true	Type II error	Correct decision

To familiarise you more with the notions of Type I and Type II error, let us look at the problem of 'Milk packets' we discussed in the Unit 5. Suppose we state the hypothesis H_0 and H_1 as

H_0 : The machine is set for 1 litre. (i.e. the average quantity of milk is 1 litre)
H_1 : The average is not 1 litre.

Let us see what are the four possible situations.

Case-1: Suppose that the hypothesis H_0 is actually (really) true. Then if we decide to accept the hypothesis 'H_0 is true', then we have made the right decision. That means if the machine is really set for 1 litre, and if we accept this following our test procedure, then we have made a correct decision.

Case 2: Suppose that, H_0 is really true and on the basis of our procedure we reject it, then we have made a mistake. That means if the machine is really set for 1 litre and our decision is to reject this and confront the milk man, then we are making a **mistake/error.**

Case 3: Suppose that the hypothesis H_0 is actually false. If we accept such a hypothesis, then we have made a mistake. That means if the machine is not properly set for 1 litre, and based on the procedure we conclude that it is set for 1 litre, then we commit a **mistake/error**.

Case 4: Suppose that the H_0 is actually false. If we reject H_0 then our conclusion is correct. That means if the machine is not properly set, and our procedure also concludes so, then we would be fully justified in demanding an explanation from the milk man. Thus there are two situations in which an error occurs.

1) **Hypothesis is true, but we reject it. This is called Type 1 error.**
2) **Hypothesis is false, but we accept it. This is called Type 2 error.**

Now you can try some exercises to see how much you have followed.

E1) Suppose you have a cough. You open the medicine cupboard and find an unmarked bottle. You have a hunch that it is not a cough medicine, rather, it is some poison. If you are using hypothesis test to arrive at a decision, how will you state your null hypothesis and alternative hypothesis? What are the two situations that lead to Type 1 and Type II errors?

E2) If you test a hypothesis and reject the null hypothesis in favour of the alternative hypothesis, does your test prove that the alternative hypothesis is correct? Justify your answer.

Now that we are aware of the types of errors, our aim is to reduce the probability of their occurrence. It is, of course, not possible to eliminate both the errors at the same time.

Depending on the problem in hand, we'll have to choose the type of error which we may prefer to have. For this, we have to look at the consequences: We may have a

Depending on the problem in hand, we'll have to choose the type of error which we may prefer to have. For this, we have to look at the consequences: We may have a situation where, if we make a Type 1 error, we lose Rs.100, and if we make a Type 2 error, then we lose Rs.10,000. In this case we'll try to eliminate the Type 2 error, since it's more expensive.

Similarly let us look at the situation in E1. You open the medicine cupboard and find an unmarked bottle. You have a hunch that it is not a cough medicine, rather it's some poison! You are not sure though. If you reject this hunch and take it to be the cough medicine, when actually it is a poison, then you have made a Type 1 error. Of course, being dead, you would be beyond caring about errors by then! On the other hand, if it is really the cough medicine, but you accept your hunch and refuse to drink it, the only consequence will be that your cough would last a little longer. In this case you would surely opt for the Type 2 error.

Often we have to properly balance the two types of error.

We denote by α, the probability of committing an error of Type 1, and by β, the probability of an error of Type 2.

Let us look at another situation suppose our supplier has sent us 5000 pens. We want to decide whether to accept or reject this lot. We would like all the pens to be in working order, but also realise that there would be some defective ones. We are ready to allow 5% defective. It is not possible to test all the pens. So we decide to test a sample of 20. We now have to decide the criterion of acceptance or rejection of the lot. For example, we may decide to accept the lot if we find at most two defectives in the sample. So we reject the lot (equivalently, decide that the lot has more than 5 % defectives) if we find 2 or more defectives in the sample. If the hypothesis that there are 5 % defectives were true, we can find the probability of rejecting the lot, that is, P (2 or more pens are defective) by using a binomial probability distribution with $n = 20, p = 0.05$. This probability is α, the probability of committing an error of Type 1. The computation of β is not always possible. Because as in this example if $p \neq 0.05$, there are infinitely many alternatives for p. Now, in this example, as soon as we fix the criterion for rejection, α is fixed.

Thus, before start testing a hypothesis, we fix or specify the value of α, or equivalently the critical region. α is called the **level of significance** of the test.

So, it is in our hands to keep the probability of Type 1 error low.

So far we have discussed that, to carry out a test for making decision, we have to choose between the null hypothesis and the alternative. We also note that we have to be careful in making decision because it can lead to errors like Type 1 and Type 2 discussed earlier. So, the tests should be such that it should be possible to measure these errors and to some extent, at least be reduced.

Let us now go back to the problem of the company manufacturing aluminium rods. Let us see how the statistician conducts a test. Note that in this case the statistician actually wants to test whether there is any change in the mean diameter of the rods. Essentially he is testing the change in the parameter mean.

Note that the change can occur in both directions. Either the mean diameter can be very large or it can very small, and neither is desirable. This is why we formulated a two sided composite alternative.

Let us now see how the statistician proceeds. The first step is to state H_0 and H_1 which has been done earlier. We restate it as

$$H_0 : \mu = 2, \quad H_1 : \mu \neq 2$$

where μ is the population mean.

Since the random sample of observations $X_1, X_2 \ldots X_n$ are assumed to be i.i.d, with the common distribution function as Normal(μ, σ), the distribution of $\bar{X}$ is normal and if the null hypothesis is true (That is, if the population mean is 2cm.), the sampling distribution of the sample mean is as shown in Fig.1. We know from Unit 4 that the sampling distribution of mean is normally distributed and has a mean of 2 cm. and a standard deviation $\frac{\sigma}{\sqrt{n}}$ ~~(where σ is the population standard deviation and n is the sample~~ size.) How does this information help the company to decide on whether its hypothesis is acceptable or not? First, they have to choose a level of significance which is "reasonable" Let us say this is $\alpha = 0.05$ i.e. 5%. This α defines a region, which is the total shaded region in the figure below (See Fig1 below). The unshaded portion in the figure below is called the critical region. Let us see how this region is calculated. Take $\frac{1-\alpha}{2} = \frac{0.95}{2} = 0.4750$. Look at the normal distribution table, and see where the value 0.4750 is given. You will find this listed in the row 1.9 and column 0.06. Thus it corresponds to z = 1.96. So, we consider the region below the curve, and bounded by $\mu_0 \pm 1.96\sigma/\sqrt{n}$ where μ_0 is the hypothetical mean (Here $\mu_0 = 2$). This region is the critical region shown in Fig. 1. The values z = 1.96 and z = −1.96 are called the **critical values.**

Now we calculate the sample mean and check whether this mean lies within the critical region or outside.

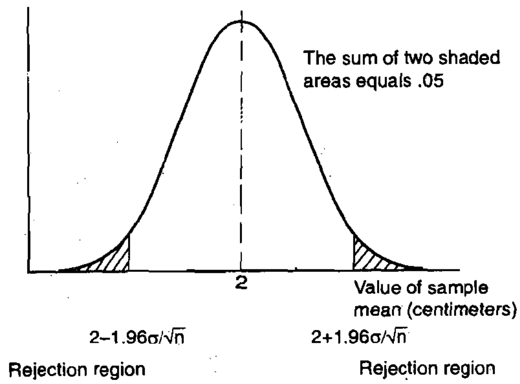

Fig.1

Critical region is calculated using the confidence level discussed in Unit 5. If it lies outside the critical region, then the decision is to reject H_0 in favour of H_1. That means the assumption that the mean diameter is 2 cm. is rejected. If it lies inside the critical region, then the conclusion is that we cannot reject H_0. That is, there is insufficient evidence to conclude that the population μ is not 2cm. Even if our sample statistic Fig.1 does fall in the unshaded region (the region that makes up 95% of the area under the curve), this does not prove that our null hypothesis ~~(H_0) is true, it simply does not~~ provide statistical evidence to reject it, why? Because the only way in which the hypothesis can be accepted with certainty is for us to know the population parameter, unfortunately this is not possible. Therefore, whenever we say that we accept null hypothesis, ~~we actually mean~~ that there is not sufficient statistical evidence to reject it.

Let us now summarise the important steps that we have discussed in the testing procedure of any hypothesis.

1) Formulate the null and alternate hypothesis .
2) Specify the significance level of the test (i.e. fix α).
3) Choose a test statistic
4) Find the critical region
5) Collect the sample and calculate the numerical value of the test statistic based on the sample of observations.

6) Conclusion:
   i) If the numerical value of the test statistic falls in the rejection region, we reject the null hypothesis and conclude that the alternative hypothesis is true. We know that the hypothesis-testing process will lead to this conclusion incorrectly (Type I error) only $100\alpha\%$ of the time when H_0 is true.
   ii) If the test statistic does not fall in the rejection region, we do not reject H_0. Thus, we reserve judgement about which hypothesis is true. We do not conclude that the null hypothesis is true, because we do not (in general) know the probability β that our test procedure will lead to an incorrect acceptance of H_0 (Type II error).

In the next section we shall illustrate these steps with many examples.

Let us look at another example.

Suppose a light bulb manufacturer believes and advertises that her bulbs have a mean life-time of 1500 hours, and she wants to test her belief. So She formulates the null hypothesis as $H_0 : \mu = 1500$ hours (h)and the alternative as $H_1 : \mu \neq 1500$h. However, she realises that her customers won't complain if the life of the bulbs they buy from her have a life span of more than 1500 hrs. So she reformulates her problem as

$$\text{Test } H_0 : \mu = 1500\text{h against } H_1 : \mu < 1500\text{h}.$$

So in this case we reject H_0 only if the mean life of the sampled bulbs is significantly below 1,000 hours. This situation is illustrated in Fig. 2.

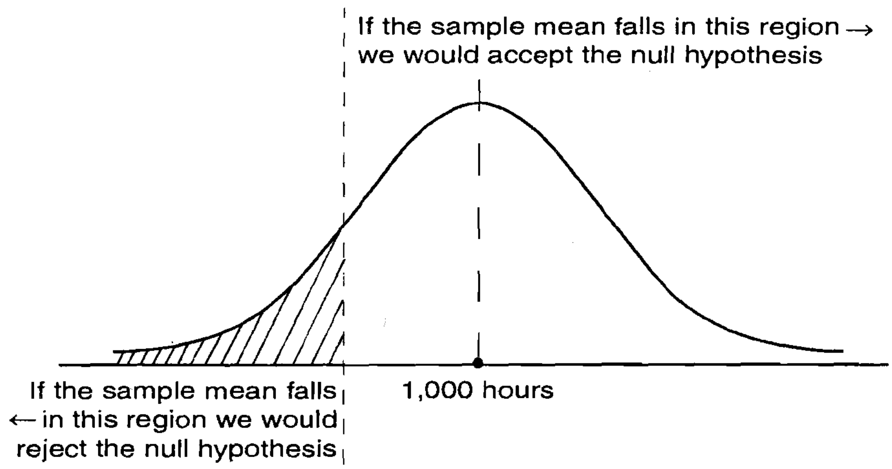

Fig. 2

Note that in this situation, the rejection region is in the left tail of the distribution of the sample mean, and so we call this test a left-tailed (or one-tailed) test.

A left-tailed test is one of the two kinds of one-tailed test. As you have probably guessed, the other kind of one-tailed test is right-tailed test which is used in the situations as given below:

A sales manager has asked her sales people to observe a limit on travelling expenses. The manager hopes to keep expenses to an average of Rs. 100 per salesperson per day. One month after the limit is imposed, a sample of submitted daily expenses is taken to see whether the limit is being observed. Here the null hypothesis is $H_0 = \mu =$ Rs.100, but the manager is concerned only with excessively high expenses. Thus, the appropriate alternative hypothesis in this case is $H_1 : \mu >$ Rs.100. So, in this case the null hypothesis is rejected (and corrective measures taken) only if the sample mean is significantly higher than Rs. 100. The situation is illustrated in Fig. 3.(See next page.) Note that in this situation, the rejection region is in the right-tail of the distribution of the sample means, and so we call this test a right-tailed test.

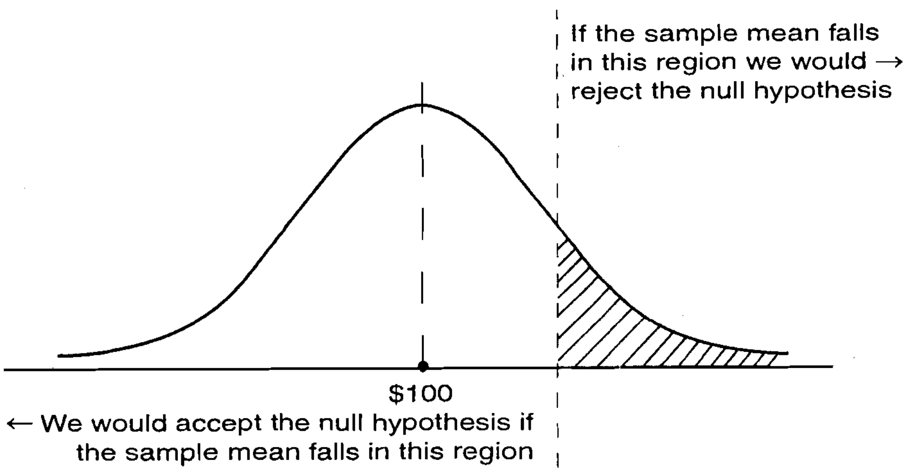

Fig. 3

Let us now summarise our discussion on the different types of tests.

The first case, where we test $H_0 : \mu = 1500h$ against $H_1 : \mu \neq 1500h$, we need to use a **two tailed test**; in the second case where we test $H_0 : \mu = 1500h$ against $H_1 : \mu < 1500h$, we use a **one-tailed test** (i.e. left-tailed test), and in the third case where we test $\mu =$Rs. 100.00 against $H_1 : \mu >$ Rs. 100.00, we use a one-tailed test (i.e. a right-tailed test.).

So, in a 2-tailed test we are concerned about any difference from the hypothetical value of the parameter, whereas in a 1-tail test we are concerned with values which are either lower or higher than the hypothetical value. The difference would be clear through the examples. Try this exercise now.

E3) Radhika, a highway safety engineer, decides to test the load-bearing capacity of a bridge that is 20 years old. Considerable data are available from similar tests on the same type of bridge. Which is appropriate, a one-tailed or two-tailed test? If the minimum load-bearing capacity of the bridge must be 10 tons, what are the null and alternate hypothesis?

In the next section we shall now show how hypotheses about parameters are tested through some examples. In this unit we shall deal with only those cases, where the population is taken to be normally distributed.

We start with tests involving the parameter mean of the population.

6.3 TESTS ABOUT THE MEAN (z-test and t-test)

In this section we show how to test whether a sample is drawn from a population with a given mean. We'll also show how to test whether two samples belong to the same population or not.

6.3.1 Comparing the Mean of a Population and a Sample

We first illustrate the technique when the variance of the population is known.

Let us look at a problem.

Problem 1: The mean marks obtained by the students of a mathematics course in IGNOU is 54.5 with a standard deviation 8.0. At one of the study centres, where 100 students took the examination, the mean marks were 55.9. Are the students of this study centre significantly 1) different from 2) better than, the rest of the students of that course in IGNOU at 0.01 level?

Solution: We shall first solve (1). Let $X_1, X_2, \ldots X_{100}$ denote the random variables which are the marks obtained by the 100 students. We assume that these random variables are i.i.d as normal with mean μ and standard deviation $\sigma_0 = 8.0$ We want to test if $\mu = 54.5$ against the alternative that $\mu \neq 54.5$

Let us apply the six steps given in the earlier section.

1) Here $H_0 : \mu = 54.5, H_1 : \mu \neq 54.5$ and $\alpha = 0.01$. It is a two-tailed test. The test statistic is $\overline{x}$. Now we have to fix the critical region as discussed in the previous section. You know that if you take samples of size 100, then the sample means $\overline{x}$ are normally distributed with mean $\mu = 54.5$ and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{100}} = 0.8$. This means that $\frac{\overline{x} - 54.5}{0.8}$ follows a standard normal distribution. Now, since $\alpha = 0.01$ (1 % level), $\frac{1-\alpha}{2} = 0.4950$. So we find the z-value corresponding to 0.4950, which we denote by $z_{0.4950}$, such that
$$P\left[-z_{0.4950} < \frac{\overline{X} - 54.5}{0.8} < z_{0.4950}\right] = 1 - \alpha = 0.99.$$
Now we look at the normal distribution table given at the end of the block and see where the value 0.4950 is given. This is listed in the row for 2.5 and column for 0.08. That is $z_{0.4950} = 2.58$.
Therefore, we have $P\left[-2.58 < Z = \frac{\overline{X} - 54.5}{0.8} < 2.58\right] = 0.99$. Also 2.58 and -2.58 are the critical values and the accepted region is as shown in Fig. 4.

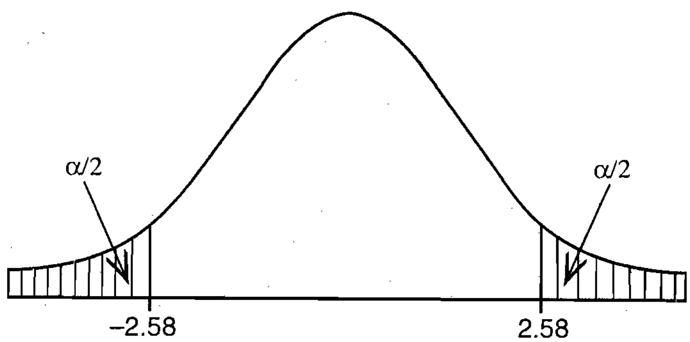

Fig.4: The total area of rejection (the shaded part) should be α

It is already given in the problem that the value of the sample statistic $\overline{x}$ is 55.9. The corresponding z-value is
$$\begin{aligned} z &= \frac{55.9 - 54.5}{0.8} \\ &= 1.75 \end{aligned}$$
Since $z = 1.75$ is in the acceptance region, we accept the hypothesis, or more precisely, we fail to reject it.
This means the difference in the two mean marks is not significant enough to suggest that the students at that study centre are different from the rest of the IGNOU students.

2) Here we apply a one-tailed test.
$H_0 : \mu = 54.5, H_1 : \mu > 54.5$
Since we are interested in knowing whether the group of students is **better**, we need to look at only those values of $\overline{x}$ which are higher than 54.5. So we find the appropriate z-value such that the rejection area is the shaded portion α as shown Fig. 5 and the acceptance region is the unshaded portion. From the normal distribution table, we determine that the value of z for 40% of the area under the

curve is z = 2.33. Here we have only one critical value 2.33. Also we have

$$P\left[Z = \frac{\overline{X} - 54.5}{0.8} < 2.33\right] = 0.99.$$

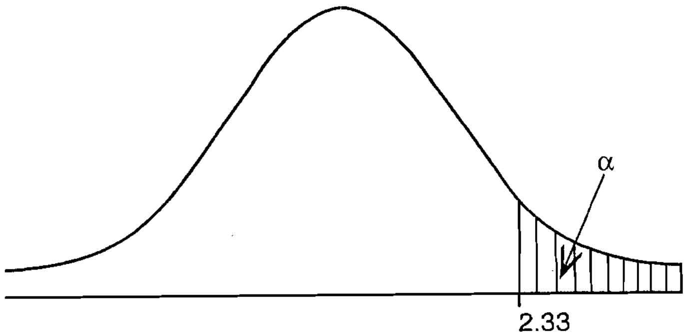

Fig.5

The shaded region is given in Fig.5. Again, the test statistic here is 1.75, which is less than 2.33 and therefore falls in the acceptance region and so we fail to reject H_0. That is, we conclude that the students from that study centre are not better than other students at 1 % level of significance.

———— × ————

You can see that the basic technique is to come up with an acceptance region by using the given level of significance (α), after noting whether it is a 1-tailed or a 2-tailed test. We then accept or reject the null hypothesis, depending on whether the given value of $\bar{x}$ falls in the region or not. You must have noticed here the similarity with the computation of confidence intervals discussed in Unit 5. There we find an interval around $\bar{x}$. Here we find an interval around μ_1, the hypothetical mean.

In the above example, α was given to be 0.01. If $\alpha = 0.05$, the only difference is in the value of z that we read from the tables. This value for a 2-tailed test is 1.96 and for a 1-tailed test is 1.64. You have been using these same values of z for calculating 95% and 99% confidence intervals in Unit 5.

In the following table, we give the critical values for both one-tailed and two-tailed tests at three significance levels, $\alpha = 0.1$, 0.01 and 0.05.

Table 2

level of significance α	0.1 (90%)	0.05 (95%)	0.01 (99%)
critical value for left-tailed test	−1.28	−1.64	−2.33
critical value for right-tailed test	1.28	1.64	2.33
critical values for two-tailed tests	−1.64 and 1.64	−1.96 and 1.96	−2.58 and 2.58

Note: - The test procedure applied in problems 1 and 2 is called **z-test.** Note that z-test is applicable when we assume that the random variables representing the sample of observations, $X_1, X_2, \ldots X_n$ are i.i.d normal(μ, σ^2) where σ^2 is assumed to be known. It is also valid if the sample size n is large i.e. $n > 30$ (by the central limit theorem) even if the common distribution is not normal or if σ^2 is unknown and has to be estimated from the sample.

Let us consider the following problem.

Problem 2: The manufacturer of an antacid claims that it relieves discomfort in 5 minutes (with a standard deviation of 2 minutes). Ten people volunteer to take it to test the claim. The average time to get relief was 7.5 minutes. Do you accept the claim at a 10% level?

Solution: We shall assume that the time to get relief is distributed normally with unknown mean μ and known standard deviation 2. Here $H_0 : \mu = 5$mins, $H_1 : \mu \neq 5$ mins.

Here the test is two-tailed and from Table 1 we get that the critical values are 1.64 and -1.64. Also we have

$$P\left[-1.64 < \frac{\overline{X} - 5}{2/\sqrt{10}} < 1.64\right] = 0.9$$

Since $\dfrac{\overline{X} - 5}{2/\sqrt{10}} = \dfrac{7.5 - 5}{2/\sqrt{10}} = 3.95$ is greater than 1.64, does not lie in the acceptance region, so we reject H_0, and conclude that the claim is not justified at 10% level.

———— × ————

In the next problem we consider the situation where the sample size n is large i.e. $n \geq 30$ and σ is unknown whereas the sample standard deviation s is given. Note that in this situation we apply z-test by replacing σ by s.

Let us see an example.

Problem 3: : A consumer magazine, when comparing various brands of paints, stated that the drying time of one particular brand was found to be four hours. The manufacturer was not particularly pleased with this and consequently modified the paint to try to reduce the drying time. The paint was then tested by a random sample of 40 customers all of whom were decorating their living rooms. For this sample the mean drying time in hours was found to be 3.85 and the sample standard deviation was 0.55.

a) Analyse the sample data using the one-sided z-test.

b) Find a 95% confidence interval for the population mean of the drying times for the modified paint.

c) What can you conclude about the drying time of the modified paint?

Solution:

a) We want to test the hypothesis that the population mean (μ) of the drying times (in hours) of the modified paint is equal to 4. The appropriate null hypothesis is therefore

$$H_0 : \mu = 4$$

Since the manufacturer is looking for a reduction in the drying time (at least he does not expect that it should have increased!), the alternative hypothesis should be one-sided, giving

$$H_1 : \mu < 4.$$

The test statistic is

$$z = \frac{\overline{x} - 4}{\sigma}$$

Here we do not know the population variance. But we know that the sample standard deviation is 0.55. Therefore using the estimation, we can find an estimate of the population S.D as $\sigma = \dfrac{s}{\sqrt{n}} = \dfrac{0.55}{\sqrt{40}} = 0.0869626$.

Note that here the test is left-sided and $\alpha = 0.05$. From Table 1, we get the critical value as -1.64.

Since the sample mean $\bar{x}$ is 3.85, we get that $z = \frac{3.85 - 4}{0.869626} \sim -1.72$ (cutting to two decimal places).
Since the value of the test statistic is -1.72 is less than the critical value -1.64, we reject H_0 in favour of H_1 at the 5% significance level and conclude that it looks as though the population mean of the drying times for the modified paint is less than four hours as the manufacturer hoped. Of course, this result only applies to paint drying in living rooms.

b) A 95% confidence interval for μ is given by $(\bar{x} - 1.96\ s/\sqrt{n}, \bar{x} + 1.96\ s/\sqrt{n})$ which is

$$[3.85 - (1.96 \times 0.0869626), 3.85 + (1.96 \times 0.0869626)]$$

giving

$$[3.67,\ 4.02]$$

c)

Fig.6

———— × ————

As in the case of the earlier problem, it is always a good idea to follow up a hypothesis by finding a confidence interval since such an estimate provides useful extra information. Of course, with some hypothesis tests you may not be able to do this with one reason or other.

You can try some exercises now.

E4) The breaking strengths of cables made by a company had a mean of 1800 N. The company then adopted a new technique which is believed to increase the breaking strengths. 50 cables made by the new technique were tested to see if the belief is justified. or not. The mean breaking strength of these 50 is found to be 1850N with a standard deviation of 100N. Is the belief justified at a) 5% level b) 1% level.

E5) As part of a survey on drivers' reaction times for a driving magazine, 300 drivers were subjected to the following test : each driver was asked to press a lever with his/her foot in response to a flashing light. The reaction times (in seconds) were recorded and the sample mean was found to be 0.83. The sample standard deviation was 0.31. What can you conclude about drivers' reaction times?

So far we have seen that z-test can be used when σ is known whether the sample size is large or small. But when the sample size is small, z-test can not be applied when σ is not known and in the case we use a test based on t-distribution instead of normal distribution. In the following problem we illustrate the use of t-distribution.

Problem 4: A machine manufactures standard weights to be used in weighing scales. To check if the machine is working properly, a random sample of five 2-kg. weights was taken. Each 2-kg. weight was weighed on a special scale and the actual weights were found to have a mean of 1.962 kg. and a standard deviation of 0.038 kg. If $\alpha = 0.05$, can you say that the machine is in proper working order?

Solution: Assume that the random observations $X_1, X_2, \ldots X_5$ are distributed as i.i.d normal (μ, σ^2) Our null hypothesis is that the machine is in proper working order. That

is $H_0 : \mu = 2kg, H_1 : \mu \neq 2kg$.
Note that here we do not know the population standard deviation (This ~~means that we are testing a composite hypothesis against a two sided composite alternative.~~)

Now we form the test statistic $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{1.962 - 2}{0.038/\sqrt{5}} = -2.24$

This follows a t distribution with 4 d.f. Now this is a 2-tailed test. So we need to get $t_{.975}$ for 4 d.f. from the t- table. We see that $t_{.975} = 2.776$. So the acceptance region is $(-2.776, 2.776)$ and -2.24 falls in this region.

So, we conclude that at 5% level of significance, the mean of the weights manufactured by the machine is 2Kg. That is, we say that the machine is in proper working order. In other words, we say that we have not found sufficient evidence to suggest that the machine is not working properly.

———— × ————

The situation in the above example called for a 2-tailed test because we were interested in testing against a two sided alternative. whether the weights were any **different** from 2 kg. The next example presents the case for a 1-tailed test.

Problem 5: A management school claims that the starting salaries for its graduates average Rs.10,000 or more per month. A random sample of 7 students who had recently graduated, showed an average salary of Rs.9700 with a standard deviation of Rs.306. At a 5% level of significance would you accept the claim?

Solution: Here you would be interested in checking if the average salaries are **less than** claimed. (If the average is more than Rs.10,000, the claim is justified of course). So

$$H_0 : \mu = \text{Rs.}10,000, \ H_1 : \mu < \text{Rs.}10,000$$

Here $t = \frac{9700 - 10,000}{306/\sqrt{7}} = -2.59.$

Since we have to apply a 1-tail test to check the lower end, we have to get a lower limit of t corresponding to 5% level with 6 d.f. Now, to get $t_{0.05}$, we find $t_{.95}$ with 6 d.f. and take its negative. So, $t_{.05} = -1.943$. Since $t = -2.59$ falls in the rejection region or the critical region, we reject H_c. This means that the claim is not justified at 5% level.

———— × ————

See if you can do this exercise now.

E6) The specifications for the production of a certain alloy call for 23.2% copper. In 10 analyses, the mean copper content was found to be 23.5 n of 0.24%. Can we conclude that the product meets the specifications if $\alpha = 0.05$?

E7) The diameters of bolts manufactured by a machine are known to have a ~~standard deviation of 0.0002 cm~~. A random sample of 10 bolts has an average diameter of 0.5046 cm and standard deviation. Test the hypothesis that the true mean diameter of bolts is 0.51 cm, using $\alpha = 0.01$.

In all the examples and exercises till now, we have been using a function of the sample mean to test hypothesis about the population mean. Suppose now we want to compare the effect of two medicines in providing relief from pain. The duration for which a medicine provides relief from pain can be observed and we may assume that the duration for medicine 1 is a random variable X that is distributed as normal with mean μ_X and standard deviation σ_X. Similarly the duration for which medicine 2 provides relief is a random variable Y which is assumed to be distributed as normal with mean

μ_Y and standard deviation σ_Y. We want to test the null hypothesis (the hypothesis of no difference) $H_0 : \mu_X = \mu_Y$ against the alternative that H_0 is not true. We discuss similar problems in the next sub-section.

6.3.2 Difference in the Means of Two Populations

Here we shall be dealing with two populations or two variables, X_1, X_2 assumed to be normally distributed with means μ_1, μ_2 and variances σ_1^2 and σ_1^2, respectively. We shall have to divide our discussion in two parts: 1) when σ_1, σ_2 are known, and 2) when σ_1, σ_2 are unknown. In Sec.6.3.1 you have seen that the test statistic for the problems discussed there follows a standard normal distribution when the standard deviation is known. You have also noted that it follows the t-distribution in case the standard deviation is not known.

We start with the first case now.

σ_1, σ_2 **known**:

We shall illustrate the method through an example.

Example 1: Suppose we want to investigate the following:

"Was there a difference in the performance of male and female students of IGNOU in the examination for Mathematics Elective courses in a particular year, say 2000? If so, what was the difference?

Let us see how we can use z-test to find an answer to this question.

We shall first set up the null and alternate hypothesis in terms of the population means as follows:

H_0: The mean examination mark for the population of all male students is the same as the mean examination mark for the population of all female students.

H_1: The mean examination mark for the population of all male students is not the same as the mean examination mark for the population of all female students.

We can now introduce some symbols that enable us to express these hypotheses more concisely. We let

μ_m denote the population mean of the marks of all male students

and

μ_f denote the population mean of the marks of all female students.

Then H_0 and H_1 become

$$H_0 : \mu_m = \mu_f$$
$$H_1 : \mu_m \neq \mu_f.$$

We also introduce now some other symbols that will be useful in our analysis of the sample data. Let

σ_m and σ_f denote the population standard deviations of the marks of all male and female students respectively;

$\overline{X}_m$ and $\overline{X}_f$ denote the two sample means;
s_m and s_f denote the two sample standard deviations;
n_m and n_f denote the two sample sizes.

Now suppose we take a sample of 150 male students and 100 female students. Then the table in the next page gives the sample means and sample s.d.'s for the samples taken. We may use $\frac{\overline{X}_m - \overline{X}_f}{SSD((\overline{X})_m - \overline{X}_f)}$ as the test statistic, where SSD denotes the estimate $\sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}$.

	Sample size	Sample mean	Sample standard deviation
Male	$n_m = 150$	$\overline{x}_m = 57.88$	$s_m = 20.00$
Female	$n_f = 100$	$\overline{x}_f = 63.73$	$s_f = 22.10$

Here, we need to find the distribution of $\dfrac{\overline{X}_m - \overline{X}_f}{SSD((\overline{X}_m) - \overline{X}_f)}$. In our problem we have assumed that the marks obtained by the male or female students follow a normal distribution. Therefore the distributions of the sample means are also normal with appropriate means and standard deviations. Also $\overline{X}_m - \overline{X}_f$ will have a normal distribution with an appropriate mean and standard deviation. However the distribution of $\dfrac{\overline{X}_m - \overline{X}_f}{SSD((\overline{X}_m) - \overline{X}_f)}$ is unknown but can be taken approximately as standard normal by using the central limit theorem, provided the sample sizes n_m and n_f are large, say more than 30. From Unit 4 you know that the sampling distribution $\overline{X}_m$ of the mean $\overline{X}_m$ is normal with mean μ_m and standard deviation $SE = \sigma_m/\sqrt{n_m}$. The standard deviation is given by

$$SE = \frac{\sigma_m}{\sqrt{n_m}} = \frac{\sigma_m}{\sqrt{150}}.$$

(See Fig.7)

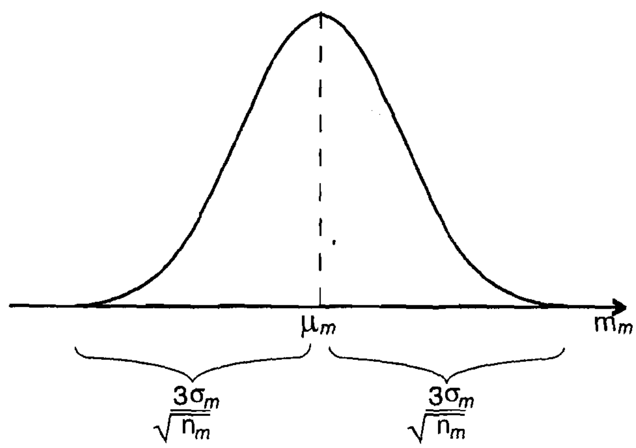

Fig. 7 Sampling distribution of $\overline{X}_m$

Similarly the sampling distribution $\overline{X}_f$ is normal with mean μ_f and standard deviation

$$SE = \frac{\sigma_f}{\sqrt{n_f}} = \frac{\sigma_f}{\sqrt{100}}.$$

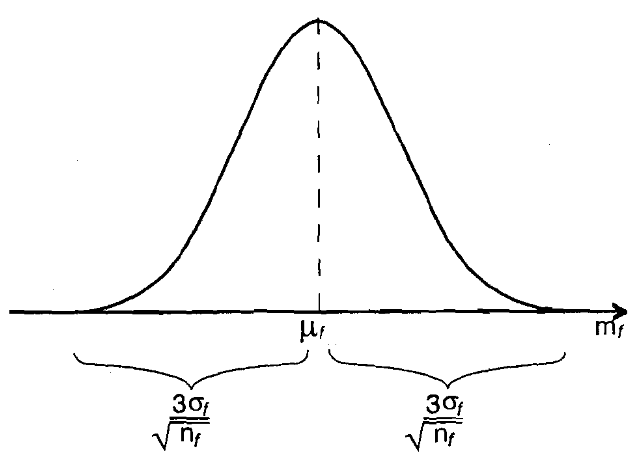

Fig. 8: Sampling distribution of $\overline{X}_f$

Now, consider all the possible pairs of samples of sizes 150 and 100 respectively that we could select from the two populations of male and female students. For each of these pairs of samples (and there is a huge number of possible combinations) we can calculate the **difference between the sample means** $(m_m - m_f)$ and by considering all the combinations of such samples we obtain the **sampling distribution of** $\frac{\overline{X}_m - \overline{X}_f}{SSD((\overline{X}_m) - \overline{X}_f)}$. It turns out that the distribution of the statistic is also approximately Normal.

Moreover, the mean of this statistic is

$$\mu_m - \mu_f$$

and its standard deviation is given by

$$SE = \sqrt{\frac{\sigma_m^2}{n_m} + \frac{\sigma_f^2}{n_f}}$$

The sample estimate of the standard deviation is

$$SSD = \sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}$$

and it is this estimate that we shall use here. The sampling distribution of $(\overline{X}_m - \overline{X}_f)$ is given in Fig.9.

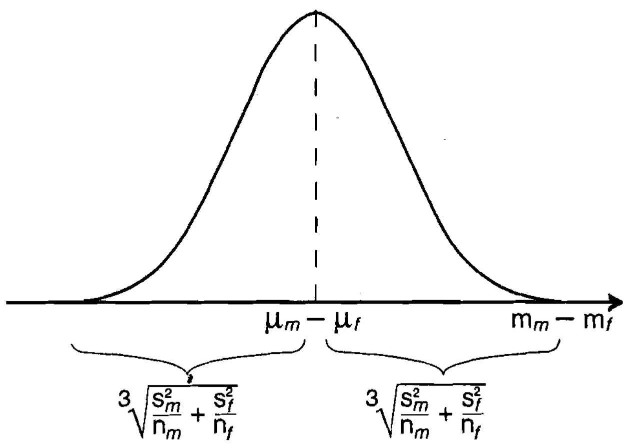

Fig.9 Sampling distribution of $\overline{X}_m - \overline{X}_f$

Before we proceed further we make a note.

Note: If we knew the standard deviations σ_m and σ_f then we didn't have to replace them by the sample estimates. In that case our test statistic will be given by

$$\frac{(\overline{X}_m - \overline{X}_f) - (\mu_m - \mu_f)}{\sqrt{\frac{\sigma_m^2}{n_m} + \frac{\sigma_f^2}{n_f}}}$$

Now we apply the z-test based on the statistic $\overline{X}_m - \overline{X}_f$ to test the given null hypothesis. The hypotheses under consideration are

$$H_0 : \mu_m = \mu_f \text{ and } H_1 : \mu_m \neq \mu_f$$

and, because

$$\mu_m = \mu_f$$

can be rewritten as

$$\mu_m - \mu_f = 0,$$

we can express H_0 and H_1 in terms of the difference between μ_m and μ_f as follows

$$H_0 : \mu_m - \mu_f = 0$$
$$H_1 : \mu_m - \mu_f \neq 0.$$

Now we just apply the z-test discussed in the earlier section replacing μ by $\mu_m - \mu_f$ and $\overline{X}$ by $\overline{X}_m - \overline{X}_f$.
Hence the test statistic is

$$z = \frac{(\overline{x}_m - \overline{x}_f) - (\mu_m - \mu_f)}{SE}$$

and, since the null hypothesis is $\mu_m - \mu_f = 0$, this simplifies to

$$z = \frac{\overline{x}_m - \overline{x}_f}{SE}$$

where, $SE = \sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}$.

Now from the data given in Fig. we can calculate the z-value. We have

$$\begin{aligned}
SE &= \sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}} \\
&= \sqrt{\frac{(20.00)^2}{150} + \frac{(22.10)^2}{100}} \\
&= \sqrt{2.6666667 + 4.8841} \\
&= \sqrt{7.5507667} \\
&= 2.7478658 \\
\text{and } \overline{x}_m - \overline{x}_f &= 57.88 - 6373 = -5.85 \\
\text{so } z &= \frac{-5.85}{2.7478659} \sim -2.212
\end{aligned}$$

Hence the test statistic is $z = -2.12$ and now the procedure is exactly the same as it was in previous Section.

Since the test statistic -2.12 is less than the critical value -1.96, we reject H_0 in favour of H_1 at the 5% significance level and conclude that $\mu_m - \mu_f$ does not seem to be equal to zero. Indeed, because -2.12 is less than -1.96, it looks as though $\mu_m - \mu_f$ is less than zero and this means that μ_m seems to be less than μ_f. This suggests that there is a difference in examination performance between the sexes; it seems that the females perform better than the males.

* * *

We shall now summarise the steps used in the example above, for conducting the test.

Our aim here is to test whether two given normally distributed random variables have the same mean or not. For this, we first state H_0 and H_1 as

$$H_0 : \mu_1 = \mu_2 \text{ i.e. } (\mu_1 - \mu_2) = 0; \quad H_1 = \mu_1 \neq \mu_2.$$

Then take a random sample of size n_1 from the first population and a random sample of size n_2 from the second. We find the means of these samples: $\overline{X}_1$ and $\overline{X}_2$. Here The difference $\overline{X}_1 - \overline{X}_2$ is normally distributed with mean $\mu_1 - \mu_2$ and variance $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.

So, if H_0 is true, then the test statistic $Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ is distributed as standard

normal.
So we find $Z_{1-\alpha/2}$ where α is the level of significance. We then reject H_0 if $Z > Z_{1-\alpha/2}$, or $Z < -Z_{1-\alpha/2}$.
If we have a situation where a one-tailed test is to be applied, say,

$$H_0 : \mu_1 = \mu_2, H_1 : \mu_1 < \mu_2,$$

then we find $Z_{1-\alpha}$ and reject H_0 (that is, accept H_1) if $Z < -Z_{1-\alpha}$.

Study the following examples carefully now, so that you can solve the exercises which come later.

Problem 6: Two machines are used to fill cans with 200 ml. of a drink. The filling processes are assumed to be normal, with standard deviations $\sigma = 0.2$ and $\sigma = 0.25$. The quality control department wants to check if the two machines fill the same volume. A random sample is taken from the output of each machine:

M_1	M_2
200.3 200.1 200.4 198.9 200.5	200.2 2003 199.7 200.4 199.4
199.2 200.5 200.2 200.2 199.5	200.2 200.1 200.1 199.8 200

what is your conclusion? Use $\alpha = 0.05$.

Solution: Note that the standard deviations are assumed to be known. We state the hypothesis as

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_A : \mu_1 &\neq \mu_2 \end{aligned}$$

The test statistics is

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

From the data given in the two samples we have
$\overline{x}_1 = 199.98$ and $\overline{x_2} = 200.02$

$$z = \frac{199.98 - 200.02}{\sqrt{\frac{0.2}{10} + \frac{0.25}{10}}} = 0.395$$

Now, for $\alpha = 0.5, z_{1-\alpha/2} = 1.96$. Since $z < z_{1-\alpha/2}$, we accept H_0.

That is, we have not found any evidence to suggest that the two machines fill different volumes of the drink.

———————×———————

Problem 7: Two different formulations of petrol are tested to study their road octane numbers. The variance for Formulation 1 is $\sigma_1^2 = 1.8$ and for Formulation 2 is $\sigma_2^2 = 1.2$. Two random samples of size $n_1 = 15$ and $n_2 = 20$ are taken. The mean road octane numbers are $\overline{x_1} = 89.6$ and $\overline{x_2} = 92.5$. Can you say that Formulation 2 produces a higher road octane number than Formulation 1 if 1) $\alpha = 0.05$, 2) $\alpha = 0.01$?
Solution: Here $H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$.

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_1}}} = \frac{89.6 - 92.5}{\sqrt{\frac{1.8}{15} + \frac{1.2}{20}}} = -6.84$$

1) For $\alpha = 0.5, z < -z_{.95} = -1.64$. Therefore we reject H_0, that is accept H_1.

2) For $\alpha = 0.1, z < -z_{.99} = -2.33$. Therefore we again reject H_0, and accept H_1.

So, at both the levels we say that Formulation 2 produces a higher road octane number than Formulation 1.

———— × ————

If you have followed these examples, you would certainly be able to these exercises.

E8) A new teaching technique is to be tested. A group of 22 students were taught in the traditional way. Another group of 18 students was taught with the help of the new technique. The two groups were then given a standardised test which is known to have a standard deviation of 25. The mean score of the traditional group was 127 and that of the experimental group was 136. If $\alpha = 0.1$, do you think that the new technique is significantly better?

E9) A psychologist gave a test to decide if male students are as smart as female students. The sample of 40 female students had a mean score of 131 and the sample of 36 males had a mean score of 126. The test has a standard deviation of 16. Is there a difference at 0.01 level of significance?

We have been considering cases where σ_1 and σ_2 are known. If they are not known, they have to be estimated from the sample. If the samples are large, then these estimates are quite close to the real values and so we can use them in forming the test statistic Z. In the next exercise you see one such situation.

E10) A sample of 100 electric light bulbs produced by manufacturer A showed a mean life-time of 1190h and a standard of 90h. A sample of 75 bulbs produced by manufacturer B showed a mean life-time of 1230h and a standard deviation of 120h. a) Is there a difference between the two brands of bulbs at a significance level of 0.05? b) Are the bulbs of manufacturer B superior to those of manufacturer A at the same level?

Now we come to the second case. Here we shall see how to test for the difference between means when the population variance is not known and the sample sizes are small.

σ_1, σ_2 unknown

Before we go any further, we must make an additional assumption. The population variances are unknown here, but whatever they are, we are going to assume that they are equal. This is because, the situation becomes very complicated if the population variances are not equal, and in this course we are not yet ready to tackle it. The common but unknown population variance has to be then estimated from the samples. So, obviously, we have to compute s_1^2 and s_2^2 which are unbiased estimates of σ^2. Now, though we have assumed the populations variances to be equal. s_1^2 and s_2^2 may not be equal. Therefore, we use them to form a pooled variance, which we then take as a single estimate of σ^2.

We form $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_2 + n_2 - 2}$, and then compute the test statistic

$t = \dfrac{\overline{x}_1 - \overline{x}_2}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, which is distributed as student's t with $n_1 + n_2 - 2$ degrees of freedom. Here is an example to illustrate this procedure.

Problem 8: Suppose we have to choose between two types of paying surfaces to be used on a highway. One of the considerations is the stopping distance of cars. The distance taken by cars travelling at 80 km. per hour to come to a complete stop was measured. The results (in meters) were as follows:

Surface A : $n_1 = 8, \bar{x}_1 = 42.3, s_1^2 = 38.8$

Surface B : $n_2 = 8, \bar{x}_2 = 43.2, s_2^2 = 51$.

Is there a difference in the stopping distance between the two surfaces at 0.05 level of significance?

Solution: $H_0 : \mu_1 = \mu_2, H_1 : \mu_1 \neq \mu_2$

$$\text{Now, } s_p^2 = \frac{7(38.8 + 51)}{14} = 44.9$$

$$\text{Then } t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{42.3 - 43.2}{\sqrt{44.9}\sqrt{\frac{1}{8} + \frac{1}{8}}} = -0.25$$

The degrees of freedom are $8 + 8 - 2 = 14$. This is a two-tailed test. So, $t_{.975}$ for 14 d.f. is 2.14. Since t is within the limits, -2.14 and 2.14, we fail to reject H_0. Therefore, we conclude that at the given level, there is no difference between the two types of surfaces.

———— × ————

We have been testing for the equality of the means of two normally distributed populations by using the sample means. The samples that we had considered were independent samples. That is, the individuals in one sample were not matched or paired with those in the other in any way.

We now look at some situations where the individuals in the samples are paired or the samples are dependent.

For example, suppose the students of a particular class are given a crash course in speed reading. To test whether their reading speed has really increased or not, we need to take a sample and note the reading speeds of the students in this sample both before and after the course. In this case, the two data sets (comprising the before and after reading speeds) are not independent since each observation in the before set is matched with an observation in the after set. Here we compare the score of each student before he/she took the course to his/her score after the course, and try to find out if there is a pattern. If 60% of the students show an increased speed, is it reasonable to call the course a success? We can answer such questions with the help of t-distribution.

We start by calculating the difference for each of the pairs. In this particular in example? It means that we take the difference between the reading speeds of each student. We then assume that these values (differences) are normally distributed. Next, we calculate the mean and standard deviation from the sample and apply the t-test. See the following situation.

Problem 9: The following table gives the data on the speeds of 10 students:

Table 3

Before	9.4	10.3	8.4	6.8	7.8	9.8	9.2	11.2	9.4	9.0
After	9.3	10.6	8.8	7.0	7.7	10.0	9.8	11.7	9.7	9.0
Difference	-0.1	0.3	0.4	0.2	-0.1	0.2	0.6	0.5	0.3	0.0

Can you say that the reading course is a success at 0.05 level of significance?

We denote the mean difference in the speeds of the population by $\bar{D}$ and that of the sample by $\bar{d}$.

Solution: Suppose that $H_0 : \bar{D} = 0$ (i.e. there is not difference in the reading speeds before and after). $H_1 : \overline{D} > 0$.

Here $\overline{d} = 0.23$, and the standard deviation for the differences (S_α) is 0.241. $n = 10$. The test statistic is:

$$t = \frac{\overline{d}}{S_d/\sqrt{n}} = \frac{0.23}{0.241/\sqrt{10}} = 3.02.$$

This is a one-tailed test. Therefore $t_{0.05}$ for 9 degrees of freedom is 1.83. Since t=3.02 is more than this upper limit, we reject the null hypothesis and accept the alternative one. So, we can conclude that the speed reading course is helpful in increasing the reading speeds of students.

———— × ————

We are sure you can do these exercises now.

E11) We want to test the effect of a new fertiliser on wheat production. For this, 24 plots of land of equal area were chosen. Half of these were treated with the new fertiliser and the other half were treated with old one. With the new fertiliser, the mean yield was 48 kg. with a standard deviation of 4 kg. With the old fertiliser, the mean yield was 51 kg, with a standard deviation of 3.6 kg. Can we say at 5 % level of significance that there is an improvement in the yield because of the new fertiliser? What will be your conclusion at 1 % level?

E12) A botanist was interested in knowing if there was a difference in the time fruits matured on different parts of a plant, and recorded the day of the first fruit on the top and on the bottom for 15 plants. All the fruits came out during the same month.

Top	3	6	7	5	8	9	10	10	7	8	6	9	10	12	4
Bottom	7	9	5	8	8	10	11	12	6	9	7	13	8	13	8

Is there a significant difference in the time to mature at the 1% significance level?

E13) The pulse rates of 12 people were recorded before and after taking a new drug.

Before	68	71	84	93	67	74	82	77	71	83	62	66
After	71	70	81	97	73	80	90	76	80	79	80	67

Using 10% level, can you say that there is a significant increase in the pulse rate?

When you have solved E11, you might have found that you reject the hypothesis at 5 % level, but accept it at 1 % level. In such a case we say that the results are probably significant, that is, the new fertiliser is probably better than the old one. But we need to get more evidence to make a decision. New para After these numerous tests about the mean we now turn our attention to the variance of a population.

6.4 TEST ABOUT THE VARIANCE

You must have seen in Unit 4 that if the population is normally distributed with mean μ and variance σ^2, then the ratio $\frac{(n-1)s^2}{\sigma^2}$ is a χ^2 variable with (n-1) d.f. Now here we are again going to use the χ^2 distribution to test if a given sample could have come from a population with a given variance, σ^2. Of course, before applying this test, we should make sure that the population is normal. Actually, even when we discussed the tests about the mean of a population, we had assumed the population to be normal. But we have to be especially careful in this case, because this test is particularly sensitive to the shape of the distribution. If we apply the test to the variance of a non-normal population, then chances are that we may be committing mistakes much more frequently than the α value indicates.

Suppose we have a sample of size n from a normal population. For this test our null hypothesis is that it comes from a population with a given σ^2. We form the test statistic, $\frac{(n-1)s^2}{\sigma^2}$. Now we compare the value of the test statistic with χ^2_α for a significance level of α. If the value of the test statistic is greater than or equal to χ^2_α, then we reject H_0. Otherwise we accept it.

Problem 10: : A machine is used to fill 2-kg packages of rice. It is known to have a standard deviation of 12.5 g. To check if the machine has become more erratic, a sample of 20 packages was taken. This showed a standard deviation of 16 g. Is the increase in variability significant at 1) 0.05 and 2) 0.01 levels?

Solution: $H_0 : \sigma = 12.5$g., $H_1 : \sigma > 12.5$g.

Now, $\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{19(16^2)}{(12.5)^2} = 31.13$

Here we have to use a one-tailed test.

1) $\chi^2_{.05}$ for 19 d.f. is 30.1. So, we reject H_0 at 0.05 level.

2) $\chi^2_{.01}$ for 19 d.f. is 36.2. So, we fail to reject H_0 at 0.01 level.

Because of the nature of conclusions in 1) and 2), we say that the variability of the machine has probability increased, and recommend that the machine should be examined. (Also see E14)

———— × ————

We now ask you to do a few exercises.

E14) It has been found that the variability of driving speeds among drivers (and not speeding) is the main cause of accidents. It has also been found that the optimum standard deviation of speeds on a highway is 3 km. per hour. A sample of speeds of 16 cars was taken and its s was found to be 14. Is this variability greater than the optimum at 5 % level of significance?

E15) The quality of printing paper depends on the variation in thickness. According to the specifications of a printer, the optimum variance of thickness is 0.0022cm. A sample of 13 readings of the thickness of the paper supplied showed a S^2 of 0.0031 cm. Should the printer reject the paper at 1 % level of significance?

We shall take up the last category of tests that we are going to discuss here: tests about proportions.

6.5 TESTS ABOUT THE POPULATION PROPORTION

In many applications we come across a binomial variable. These are the situations in which the population is divided in exactly two categories: male female, good bad, acceptable defective, and so on. Here we are interested in the proportion of one of the categories. Again, to estimate the proportion in the population we take the help of a random sample. You have already seen this in Unit 5. Even though the variable is binomial, p is still found to follow a normal distribution. This is of course true, when the sample size is large and p is not too close to either 0 or 1. We use this fact now to test the hypotheses about the population proportion. You must have realised by now that tests of hypotheses run exactly parallel to the computation of confidence intervals. Before we illustrate the procedure through some examples, here are a few points to remember.

1) If π is the proportion of a certain category in the population, and p is that calculated from the sample, then under the above restrictions, $z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$ is a standard normal variable.

2) If two samples of sizes n_1 and n_2 are drawn from the same population, and if p_1 and p_2 are the estimates of π obtained from them, then $z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ is a standard normal variable, where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$.

Let us consider this problem.

Problem 11: : A private gallery purchased a rare painting, expecting that it would attract 75 % of its visitors. To verify this, a sample of 60 people was taken. It was found that out of these, 35 had looked at the painting. Do you think that the expectation was justified at 5% level?

Solution: $H_0 : \pi = 0.75, H_1 : \pi \neq 0.75$

Here $n = 60$ and $p = \frac{35}{60} = 0.58$

$$z = \frac{0.58 - 0.75}{\sqrt{\frac{(0.75)(0.25)}{60}}} = -3.036$$

Since this value is outside the two-tailed 5% limits, ± 1.96, we reject the hypothesis and conclude that the expectation was not correctly estimated.

———— × ————

Problem 12: Two types of computer systems are being tested for use in a new gun. The first system gave 250 hits out of 300 rounds, and the second one gave 182 hits out of 240 rounds. At 1% level can you say that the two systems differ?

Solution: $H_0 : \pi_1 = \pi_2, H_1 : \pi_1 \neq \pi_2$

Here $n_1 = 300, n_2 = 240, p_1 = \frac{250}{300}, 0.833, p_2 = \frac{182}{240} = 0.758$

Now, $p = \frac{250 + 182}{300 + 240} = 0.8$

The test statistic $z = \frac{0.833 - 0.758}{\sqrt{(0.8 \times 0.2)(\frac{1}{300} + \frac{1}{240})}} = 2.17$

The two-tailed limits at 1% level are ± 2.58. Since z lies within these limits, we cannot reject H_0. Therefore, we conclude that the two systems do not differ.

———— × ————

In the two examples above, we have shown how to test the hypotheses about proportions in two-tailed situations. We are sure you will be able to modify the procedure in case you have to deal with a one-tailed situation. You can check it out by doing these exercises now. We have included two-tailed as well as one-tailed situations here. In case you have a problem, you know you can always find the solution at the end of the unit.

E16) Two different machines are used to make plastic bottles. Two random samples, each of size 500 are selected from the bottles produced by the machined. The sample from the first machine was found to contain 250 defective bottles, while that from the second machine contained 40 defective bottles. Is it reasonable to say that both the machines produce the same proportion of defectives if we use $\alpha = 0.05$?

E17) A random sample of size 1000 from machine 1 contained 20 defectives, and a random sample of size 1500 from machine 2 contained 40 defectives. If $\alpha = 0.05$, can you say that machine 1 is better than machine 2?

E18) A flue vaccine was given to 125 of a total of 200 employees of a firm. Thirty employees who had received the vaccine were down with flue, while 25 of those who did not, also were stricken. At 1% level of significance would you say that the vaccine was effective?

That brings us to the end of this unit. We now summarise our discussion.

6.6 SUMMARY

In this unit we have discussed

- some test procedures that are called significance test or hypothesis test for making inference about the population parameter using sample results. These tests are applied when we have a claim/hypothesis about the population parameter.
- null and alternate hypothesis, significance level, test statistic, rejection region as the basic elements of these tests.
- two types of errors - Type 1 and Type 2 errors
- when to use one-tailed or two-tailed tests.
- how to apply z-test and t-test.

6.7 SOLUTIONS/ANSWERS

E1) H_0: It is not a cough medicine.
H_1 : It is a cough medicine.

Suppose it is actually true that it is not a cough medicine i.e. H_0 is true, then if you are rejecting this hypothesis, i.e. you are making a decision that it is a cough medicine. Then this will lead to Type 1 error.

Suppose it is actually true that it is a cough medicine (i.e. H_0 is false) and you are accepting the hypothesis and decide not cough medicine then this will lead to Type II error.

E2) No.

E3) The engineer would be interested in whether a bridge of this age could withstand minimum load-bearing capacities necessary for safety purposes. She therefore wants its capacity to be above a certain minimum level, so a one-tailed test would be appropriate. The hypotheses are

$H_0 : \mu = 10$ tons, $H_1 = \mu > 10$ tons

E4) $H_0 : \mu = 1800N$
$H_1 : \mu > 1800N$

$|z| = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| = \left|\frac{1850 - 1800}{100/\sqrt{50}}\right| = 3.5355$ The critical value for one-tailed test at

5% level is 1.64 and that at 1% level is 2.33. Since $|z|$ is greater than both these, we reject the hypothesis (H_0) and accept the alternative hypothesis (H_1) at both levels. So the new technique is effective.

E5) $\bar{x} = 0.83$ sec., $s = 0.31$ sec., $n = 300$

The 95% C.I. for μ is

$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}}$$

$$= 0.83 \pm 1.96\frac{0.31}{\sqrt{300}} = 0.83 \pm 0.035$$

$$= (0.795, 0.865)$$

With 95% confidence we can say that the mean reaction time of drivers is between 0.795 and 0.865 seconds.

E6) d.f. =9 $H_0 : \mu = 23.2\%$

$H_1 : \mu \neq 23.2\%$.

The critical value of t with 9 d.f. at $\alpha = 0.05$ is 2.26

$$|t| = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| = 3.9528$$

Therefore we reject H_0. The product does not meet the specifications.

E7) $\sigma = 0.0002$ cm., $n = 10$, $\bar{x} = 0.5046$, d.f. = 9 and $\alpha = 0.01$

$H_0 : \mu = 0.51$ cm.

$H_1 : \mu \neq 0.51$ cm.

This is a 2-tailed test.

$$|t| = \left|\frac{\bar{x} - \mu}{s/\sqrt{n}}\right| = \left|\frac{0.5046 - 0.51}{0.0002/\sqrt{10}}\right| = 85.3815$$

The critical value of t with 9 d.f. at $\alpha = 0.01$ is 2.821. Hence we reject H_0.

$\therefore$ the true mean diameter is not 0.51 cm.

E8) $n_1 = 22$ $n_2 = 18$, $\bar{x}_1 = 127$, $\bar{x}_2 = 136$ $\sigma = 25$ and $\alpha = 0.1$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_2 > \mu_1$

$$|z| = \left|\frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| = \left|\frac{127 - 136}{25\sqrt{\frac{1}{22} + \frac{1}{18}}}\right| = 1.1327.$$

The critical value of z for $\alpha = 0.1$ is 1.64.

$\therefore$ we do not reject H_0.

$\therefore$ the new technique is not better.

E9) $n_1 = 40$ $n_2 = 36$ $\overline{x_1} = 131$, $\overline{x_2} = 126, \sigma = 16$ and $\alpha = 0.01$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

$$|z| = \left|\frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\right| = \left|\frac{131 - 126}{16\sqrt{\frac{1}{40} + \frac{1}{36}}}\right| = 1.36.$$

The critical value for $\alpha = 0.01$ for a 1-tail test is 2.33.

$\therefore$ We accept H_0

So male students are as smart as female students.

E10) $n_1 = 100$, $n_2 = 75$, $\overline{x_1} = 1190h$, $\overline{x_2} = 1230h$, $s_1 = 90h$ $s_2 = 120h$ and $\alpha = 0.05$

a) $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

$$|z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right| = \left| \frac{1190 - 1230}{\sqrt{\frac{8100}{100} + \frac{14400}{75}}} \right| = 2.421.$$

The critical value is 1.96 for a 2-tail test
$\therefore$ we reject H_0
$\therefore$ There is a significant difference.

b) $H_0 : \mu_1 = \mu_2$
$H_1 : \mu_2 > \mu_1$
$|z| = 2.421$. The critical value is 1.64 for a 1-tail test at $\alpha = 0.05$.
$\therefore$ we reject H_0 and accept H_1.
$\therefore$ The bulbs of manufacturer B are superior to those of manufacturer A.

E11) $n_1 = 12$, $n_2 = 12$, $\overline{x_1} = 48kg$, $\overline{x_2} = 51kg$, $s_1 = 4kg$, $s_2 = 3.6kg$, $\alpha = 0.05$ and d.f. $= 22$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

Now,

$$|t| = \left| \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \quad \text{where } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$s_p = 3.8052$ $|t| = 1.9312$

The critical value $t_{0.05} = 1.72$

$\therefore$ we reject H_0 and accept H_1.

$\therefore$ There is improvement.

The critical value $t_{0.01} = 2.51$

$\therefore$ At $\alpha = 0.01$, we fail to reject H_0, and conclude that there is no improvement.

E12)

Top	3	6	7	5	8	9	10	10	7	8	6	9	10	12	4
Bottom	7	9	5	8	8	10	11	12	6	9	7	13	8	13	8
Difference	−4	−3	2	−3	0	−1	−1	−2	1	−1	−1	−4	2	−1	−4

$n = 15$, $\bar{d} = -1.3333$ $s = 1.955$ d.f. $= 14$, $\alpha = 0.01$

$H_0 : \overline{D} = 0$
$H_1 : \overline{D} \neq 0$

$$|t| = \left| \frac{\bar{d}}{s/\sqrt{n}} \right| = \frac{1.3333}{1.955/\sqrt{15}} = 2.6412$$

This is a 2-tailed test. $t_{0.01}$ for 14 d.f. is 2.98 $\therefore$ we accept H_0.

$\therefore$ there is no significant difference.

E13)

Before	68	71	84	93	67	74	82	77	71	83	62	66
After	71	70	81	97	73	80	90	76	80	79	80	67
Difference	-3	1	3	-4	-6	-6	-8	1	-9	4	-18	-1

$\bar{d} = 3.833 \quad s_d = 5.9 \quad n = 12 \quad d.f. = 11$

$H_0 : \bar{d} = 0$

$H_A : \bar{d} < 0$

$$|t| = \left|\frac{\bar{d}}{s/\sqrt{n}}\right| = \frac{3.833}{5.9/\sqrt{12}} = 2.2505$$

This is a 2-tailed test. The critical value of t for $\alpha = 0.1$ for 11 d.f. is 1.36.

Hence we reject H_0 and accept H_1.

therefore there is a significant increase in the pulse rate.

E14) $\sigma = 3$km/hr. $\quad n = 16, \quad s = 14$km/hr. $\quad \alpha = 0.05$

$H_0 : \sigma = 3$km/hr.

$H_A : \sigma > 3$km/hr.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{15(14)^2}{9} = 326.67.$$

This is a 1-tailed test.

$\chi^2_{.05}$ for 15 d.f. is 25

$\therefore$ we reject H_0 and accept H_1.

E15) $\sigma^2 = 0.0022$ cm. $\quad s^2 = 0.0031$ cm. $\quad n = 13 \quad \alpha = 0.01.$

$H_0 : \sigma^2 = 0.0022$ cm.

$H_A : \sigma^2 > 0.0022$ cm.

$$\chi^2 = \frac{12(0.0031)}{0.0022} = 16.91$$

$\chi^2_{.01}$ for 12 d.f. is 26.2.

$\therefore$ we accept H_0. The printer should not reject.

E16) $n_1 = n_2 = 500 \quad p_1 = \frac{250}{500} = 0.5, \quad p_2 = \frac{40}{500} = 0.08.$

$$p = \frac{250 + 40}{1000} = 0.29$$

$H_0 : \pi_1 = \pi_2$

$H_A : \pi_1 \neq \pi_2$

$$|z| = \left|\frac{0.5 - 0.08}{\sqrt{0.29 \times 0.71\left(\frac{1}{500} + \frac{1}{500}\right)}}\right| = 14.635$$

The critical value for $\alpha = 0.05$ in 1.96. Hence we reject H_0.

E17) $n_1 = 1000, \quad n_2 = 1500, \quad \alpha = 0.05$

$p_1 = \frac{20}{1000} = 0.02 \qquad p_2 = \frac{40}{1500} = 0.0267$

$$p = \frac{60}{2500} = 0.024$$

$H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 < \pi_2$

$$|z| = \left|\frac{0.02 - 0.0267}{\sqrt{0.024 \times 0.976\left(\frac{1}{1000} + \frac{1}{1500}\right)}}\right| = 1.0723$$

1-tailed limit for $\alpha = 0.05$ is 1.64.

$\therefore$ we accept H_0.

E18) $n_1 = 125, \quad n_2 = 75, \quad \alpha = 0.01$

$$p_1 = \frac{30}{125} = 0.24 \qquad p_2 = \frac{25}{75} = 0.333$$

$$p = \frac{55}{200} = 0.275$$

$H_0 : \pi_1 = \pi_2$

$H_1 : \pi_1 < \pi_2$

$$|z| = \left| \frac{0.24 - 0.333}{\sqrt{0.275 \times 0.725 \left(\frac{1}{125} + \frac{1}{75} \right)}} \right| = 1.426$$

This is a 1-tailed test. The critical value of z for $\alpha = 0.01$ is 2.33.

$|z| < 2.33.$

$\therefore$ we accept H_0.