Chapter - 8

HYPOTHESES TEST CONCERNING MEAN

Study Strategy and Learning Objectives

Please Remember The Following Study Strategy and Learning Objectives:

Study Strategy:

1. First, read this section with the limited objective of simply trying to understand the following important key terms and concepts: *Null hypothesis, Sample size determination for hypothesis test, One-and two- sided alternative hypotheses tests, acceptance and Critical regions for a test statistic, Test statistic, Connection between hypothesis tests and condense intervals, Significance level of a test, Type I and type II errors. Z-test, t-test, large sample test, small sample test, difference between Z-test and t-test, Test for homogeneity, Power of the test, ANOVA.*
2. Second, try to understand what they accomplish, and why they are needed; and develop the ability to calculate or select them.
3. Third, learn how to interpret them.
4. Fourth, read the section once again and try to understand the underlying theory.

You will always enjoy much greater success if you understand what you are doing, instead of blindly applying mechanical steps in order to obtain an answer that may or may not make any sense.

Learning Objectives:

After careful study of this chapter you should be able to do the following:

1. Structure engineering decision-making problems as hypothesis tests on mean.
2. Structure comparative experiments involving two samples as hypothesis tests
3. Test hypotheses and construct confidence intervals on single mean of a normal distribution and on the difference in two population means using either a *Z-test* or a *t-test* procedure.
4. Use the *P*-value approach for making decisions in hypotheses tests.
5. Compute power, type II error probability, and make sample size selection decisions for one sample and two sample tests on means.
6. Explain and use the relationship between confidence intervals and hypothesis tests.
7. Understand how the analysis of variance is used to analyze the data from the experiments

8.1 Introduction

Estimation theory and hypothesis testing are integral parts of statistical inference. A parameter can be estimated from sample data either by a single number (a point estimation) or an entire interval (a confidence interval). However, the objective of an investigator is not to estimate a parameter but to decide which of two contradictory claims about the parameter is correct. Methods for accomplishing this consist of the part of statistical inference called hypotheses testing.

Therefore, in estimate theory, we learned how to estimate the values of population parameter. In this chapter, we will learn how to test the values of population parameters.

Definitions: Hypothesis and Test of hypothesis

Hypothesis is a quantative statement (or claim) about a property of a population parameter. In other words, it is an *assumption* or *presumption* or *claim* about population parameter. It may or may not be true.

A ***hypothesis test*** (or ***test of significance***) is a standard procedure for testing a claim about a property of a population

That is, a statistical procedure that involves formulating hypotheses and testing of *validity* (or *reliability*) of the hypothesis on the basis of sample information is called ***test of hypothesis*** or ***test of significance***. By test of hypothesis we can find out whether it deserves acceptance or rejection.

There can be various types of hypotheses (claims).

For example:

1. The defective items produced by a particular machine is 3% of the total production.
2. A particular drug cures 80% of the patients suffering from malaria.
3. A reporter claims that the majority of Nepalese derivers run red lights.
4. Medical researchers claim that the mean body temperature of healthy adults is not equal to 98.6°F.
5. When new equipment is used two manufacture aircraft altimeter, the new altimeters are better because the variation in the errors is reduced so that the reading are more consistent.

A hypothesis is concerned with the behavior of an observable random variable and is tested on the basis of limited results obtained from samples. An assumption is made about the parameter. To test the assumption, a sample is selected from the population, a sample statistic observed, the difference between the sample mean and the hypothesized value calculated, and the difference is tested for significance. Smaller the difference, closer is the sample mean to the hypothesized value and vice versa.

Before beginning to study this chapter, you should recall—and understand clearly—this basic rule.

Rare Event Rule for Inferential Statistics

If, under a given assumption, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct.

Following this rule, we test a claim by analyzing sample data in an attempt to distinguish between results that *can easily occur by chance* and results that are *highly unlikely to occur by chance*. We can explain the occurrence of highly unlikely results by saying that either a rare event has indeed occurred or that the underlying assumption is not true.

8.2 Basics of Hypothesis Testing

In this section, we describe the formal components used in hypothesis testing: null hypothesis, alternative hypothesis, test statistic, critical region, significance level, critical value, *P*-value, type I error, and type II error. The focus in this section is on the individual components of the hypothesis test, whereas the following sections will bring those components together in comprehensive procedures. Here are the objectives for this section.

Objectives for this Section

1. Given a claim, identify the null hypothesis and the alternative hypothesis, and express them both in symbolic form.
2. Given a claim and sample data, calculate the value of the test statistic.
3. Given a significance level, identify the critical value(s).
4. Given a value of the test statistic, identify the *P*-value.
5. State the conclusion of a hypothesis test in simple, nontechnical terms.
6. Identify the type I and type II errors that could be made when testing a given claim.

8.2.1 Types of Hypotheses

There are two types of Hypotheses:

(1) Null hypothesis and; (2) Alternative hypothesis

8.2.2 Null Hypothesis (or Neutral hypothesis)

The ***null hypothesis*** (denoted by H_0) is the claim that is initially assumed to be true (the '*prior belief*' claim). This hypothesis is under test or verification. It is assumption or presumption about the population parameter. Generally Null hypothesis is a hypothesis of no difference which means that there is no significant difference between the sample statistic and the population parameter, in case difference is seen, that is merely due to fluctuation of sampling.

Therefore, the ***null hypothesis*** is a statement that the value of a population parameter (such as proportion, mean, or standard deviation) is a ***equal to*** some claimed value. Here are some typical null hypotheses of the type considered in this chapter:

$H_0: \mu = 98.6 \quad H_0: p = 0.5 \quad H_0: \sigma = 15$

We test the null hypothesis directly in the sense that we assume it is true and reach a conclusion to either reject H_0 or fail to reject H_0.

If the population has specified value, say μ_0, the null hypothesis can be set up as:

$H_0: \mu = \mu_0$, i.e., the population has specified mean value μ_0.

For example, a manufacturer of dairy milk claims that, on an average, its pocket contains 1000 ml of milk. In reality, this claim may or may not be true. However, we will initially assume that the manufacturer's claim is true. To test this claim, we set up null hypothesis as:

$H_0: \mu = 1000$ *ml*, i.e. manufacturer's claim is true. Here, the manufacturer's claim will be true if all the packets, on an average, contain 1000 *ml* of milk.

8.2.3 Alternative hypothesis

The ***alternative hypothesis*** (denoted by H_1 or H_a) is the statement that the parameter has a value that somehow differs from the null hypothesis. Alternative hypothesis is likely to be accepted if the null hypothesis is rejected. It is set up against null hypothesis. So it is complementary hypothesis to the null hypothesis. It is also called ***hypothesis of difference***.

For the methods of this chapter, the symbolic form of the alternatives hypothesis must use one of these symbols: < or > or ≠ . Here are nine different examples of alternative hypotheses involving proportions, means, and standard deviations:

Means:	$H_1: \mu > 98.6$	$H_1: \mu < 98.6$	$H_1: \mu \neq 98.6$
Proportions:	$H_1: p > 0.5$	$H_1: p < 0.5$	$H_1: p \neq 0.5$
Standard Deviations:	$H_1: \sigma > 15$	$H_1: \sigma < 15$	$H_1: \sigma \neq 15$

Suppose, we want to test the null hypothesis that the population mean μ has a specified value μ_0. Now, we set up alternative hypothesis H_1 follows:

$H_0: \mu = \mu_0$.

We will consider three possible alternative hypotheses:

(i) $H_1: \mu \neq \mu_0$. That is, there is significant difference between sample statistic and population parameter.

(ii) $H_1: \mu > \mu_0$. i.e., population mean is greater than μ_0.

(iii) $H_1: \mu < \mu_0$, i.e. population mean is less than μ_0.

Here (i) is called ***two tailed*** or ***two sided*** alternative hypothesis and (ii), (iii) are called ***right*** and ***left tailed*** or ***one sided*** alternative hypotheses respectively.

General guideline for selecting the null hypothesis H_0:

When the goal of an experiment is to establish an assertion, the negation of the assertion should be taken as the ***null hypothesis***. The assertion becomes the ***alternative hypothesis***.

Notation for the hypotheses:

H_1: The ***alternative hypothesis*** is the claim we wish to establish.

H_0: The ***null hypothesis*** is the negation of the claim.

For example, suppose a consumer protection group wishes to test the claim of the manufacturer. In order to test the claim of the manufacturer for the above example, H_1 will be set up as follows:

$H_1: \mu < 1000$ *ml*, i.e., manufacturer's claim is false. The manufacturer's claim will be false if its milk packets contain, on average, less than 1000 *ml* of milk.

Note:

1. In above example, we do not set up alternative hypothesis as $H_1: \mu > 1000$ *ml*, because if all packets contain more than 1000 *ml* milk, then the manufacturer's claim is also valid. So, we formulate alternative hypothesis from consumer's side in the sense that if each packet contains more than 1000 *ml*, then consumer gets more quantity at the same cost. Hence, we wish to test whether each packet contains less than 1000 *ml*. We will reject the manufacturer's claim if each packet, in an average; contains less than 1000 *ml*.
2. If the manufacturer wishes to test his own claim, then manufacturer sets up alternative hypothesis as: $H_1: \mu \neq 1000$ *ml*.

Here, he wishes to test to determine whether each packet contains significantly less or more than 1000 *ml*. He always wants to fill the amount of milk in each pocket exactly or approximately 1000 *ml* because if he fills more than 1000 *ml*, he will be in loss and if he fills less than 1000 *ml* he will lose the customers.

The following rules will help while setting up H_0 and H_1

1. Null hypothesis are of the form

   $H_0: \theta = \theta_0$ or $H_0: \theta \geq \theta_0$ or $H_0: \theta \leq \theta_0$

   but while conducting the test, in general, hull hypothesis is expressed as an equality, i.e., $H_0: \theta = \theta_0$.

   and the alternative hypothesis are expressed in the form of strict inequality like $H_1: \theta > \theta_0$ or $H_1: \theta < \theta_0$ or $H_1: \theta \neq \theta_0$

   [Here parameter θ could be mean μ, proportion p, variance σ^2 and so on.]

   If we want to test, say,

   $H_0: \theta \leq \theta_0$ against $H_1: \theta > \theta_0$, it suffices to test

   $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$.
2. The conclusion expected as a result of the test should be placed in the alternative hypothesis.
3. Null hypothesis is tested or verified.
4. Null and alternative hypothesis are complementary.

Note about forming your own Claims (Hypotheses)

If you are conducting a study and want to use a hypothesis test to support your claim, the claim must be worded so that it becomes the alternative hypothesis. This means that your claim must be expressed using only these symbols: < or > or ≠. You cannot use a hypothesis test to support a claim that some parameter is equal to some specified value.

For example, suppose you have developed a magic potion that raises *IQ* scores so that the mean becomes greater than 100. If you want to provide evidence of the potion's effectiveness, you must state the claim as $\mu > 100$. (In this context of trying to support the goal of the research, the alternative hypothesis is sometimes referred to as the *research hypothesis*. Also in this context, the null hypothesis of $\mu = 100$ is assumed to be true for the purpose of conducting the hypothesis test, but it is hoped that the conclusion will be rejection of the null hypothesis so that the claim of $\mu > 100$ is supported)

Note about Identifying H_0 and H_1: The following Figure summarizes the procedures for identifying the null and alternative hypotheses. Note that the original statement could become the null hypothesis. It could become the alternative hypothesis, or it might not correspond exactly to either the null hypothesis or the alternative hypothesis.

Figure: Identifying H_0 and H_1

Start

↓

Identify the specific claim or hypothesis to be tested and express it in symbolic form.

↓

Give the symbolic form that must be true when the original claim is false.

↓

Of the two symbolic expressions obtained so far, let the alternative hypothesis H_1 be the one not containing equality, so that H_1 uses the symbol < or > or ≠. Let the null hypothesis H_0 be the symbolic expression that the parameter equals the fixed value being considered.

For example, we sometimes test the validity of someone else's claim, such as the claim of the Coca Cola Bottling Company that "the mean amount of Cock in cans is at least 12 *oz*." That claim can be expressed in symbols as $\mu \geq 12$. We see that if that original claim is false, then $\mu < 13$, but the null hypothesis is $\mu = 12$. We will be able to address the original claim after determining whether there is sufficient evidence to reject the null hypothesis of $\mu = 12$.

Example 1: *(Identifying the Null and Alternative Hypotheses)*

Use the given claims to express the corresponding null and alternative hypotheses in symbolic form,

(a) The proportion of drivers who admit to running red lights is greater than 0.5

(b) The mean height of professional basketball players is at most 7 ft.

(c) The standard deviation of *IQ* scores of actors in equal to 15.

Solution: See above figure, which shows the three-step procedure.

(a) In steps 1 we express the given claim as $p > 0.5$. In step 2 we see that if $p > 0.5$ is false, then $p \leq 0.5$ must be true. In Step 3, we see that the expression $p > 0.5$ does not contain equality, so we let the alternative hypothesis H_1 be $p > 0.5$, and we let H_0 be $p = 0.5$.

(b) In Step 1 we express "a mean of at most 7 ft" in symbols as $\mu \leq 7$. In Step 2 we see that if $\mu \leq 7$ is false, then $\mu > 7$ must be true. In Step 3, we see that the expression $\mu > 7$ does not contain equality, so we let alternative hypothesis H_1 be $\mu > 7$ and we let H_0 be $\mu = 7$.

(c) In Step 1 we express the given claim as $\sigma = 15$. In Step 2 we see that if $\sigma = 15$ is false, then $\sigma \neq 15$ be must be true. In Step 3, we let alternative hypothesis H_1 be is $\sigma \neq 15$,and we let H_0 be $\sigma = 15$.

Example 2: *(Set up null and alternative hypotheses for the following claims)*

(a) Suppose a company has implemented a new advertising program in the hope of increasing sales from last year's annual average of Rs. 10 million. Test whether the new advertising program was successful. Here we set up hypotheses as:

Null Hypothesis H_0: $\mu = 10$ millions, i.e., the new advertising program was not successful.

Alternative hypothesis H_1: $\mu > 10$ millions. i.e. the new advertising program was successful. The program is successful if sales of current year is greater than that of last year, otherwise, not successful.

(b) Set up *alternative hypothesis* against the *null hypothesis* that the per capita income of Nepalese income is (i) different from Rs 10,000; (ii) Less than Rs. 10,000, (iii) more than Rs. 10,000.

Solution:

(i) *Null hypothesis* H_0: μ = Rs. 10,000

Alternative hypothesis H_1 : $\mu \neq$ Rs. 10,000.

(ii) *Null hypothesis* H_0: $\mu \geq$ Rs. 10,000 (Note that this contains equality sign also)

Alternative hypothesis H_1 : $\mu <$ Rs. 10,000.

(iii) *Null hypothesis* H_0: $\mu \leq$ Rs. 10,000.

Alternative hypothesis H_1: $\mu >$ Rs. 10,000.

(c) If we think about the judicial system in terms of a hypothesis test we set up null and alternative hypotheses as :

Null hypothesis: H_0: the person is innocent

Alternative hypothesis: H_1: the person is not innocent.

8.3 Types of errors in testing of hypothesis

Since the decision to accept or reject the null hypothesis H_0 is made on the basis of informations obtained from the sample data only, there is always a possibility of error. There are four possible decisions:

(i) Accepting H_0 when H_0 is true

(ii) Rejecting H_0 when H_0 is true

(iii) Accepting H_0 when H_0 is false

(iv) Rejecting H_0 when H_0 is false.

Here decisions (i) and (iv) are correct decisions but (ii) and (iii) are wrong decisions.

Thus, in testing of hypothesis, we may commit two types of errors:

(1) Type I error , and (2) Type II error.

8.3.1 Type I error

The error committed in rejecting null hypothesis H_0 when it is true, is called ***Type I error*** or the ***error of the first kind***. The probability of making Type I error is denoted by α. That is,

P(Reject H_0 when H_0 is true) = P(Type I error) = α.

Also α is also called size of type I error.

Here, α is referred to as ***level of significance***. The significance level is always specified by the decision maker himself / herself. In most of the hypothesis testing problems, level of significance is fixed at 5%. Hence, the risk of committing Type I error is always under the control of decision maker. The complement of α (i.e., $1 - \alpha$) is called the ***confidence coefficient***. The confidence coefficient $(1-\alpha)$ is the probability of accepting H_0 when H_0 is true. That is,

P(Accept H_0 when H_0 is true) = $1 - \alpha$.

8.3.2 Type II error

The error committed in accepting null hypothesis H_0 when it is false, is called ***Type II error***. The probability of making Type *II* error is denoted by β. That is,

$$P(\text{Accept } H_0 \text{ when it is false}) = P(\text{Type II error}) = \beta.$$

Here β is also called size of type II error. The probability β depends upon the gap between sample statistic and the population parameter. If the gap is too low, β is high and vice-versa. The complement of β, $(1 - \beta)$ is called power of the test which is the probability of rejecting the null hypothesis when it is false. That is,

$$P(\text{Reject } H_0 \text{ when } H_0 \text{ is false}) = 1 - \beta.$$

These two types of errors can be presented in table as:

Statistical decision	*Actual Situation*	
	H_0 is true	H_0 is false
Accept H_0	correct decision, confidence coefficient = $1 - \alpha$	Type *II* error, P(Type *II* error) = β
Reject H_0	Type I error, P(type I error) = α	Correct decision, Power of test = $1 - \beta$

Notation:

α (alpha) = probability of a type I error (the probability of rejecting the null hypothesis when it is true)

β (beta) = probability of a type II error (the probability of failing to reject a null hypothesis when it is false)

In terminology of Industrial Quality control while inspecting the quality of manufactured lot.

$$\alpha = P(\text{type I error}) = P[\text{rejecting a good lot}]$$

$$\text{and} \quad \beta = P(\text{type II error}) = P[\text{accepting a bad lot}]$$

where α and β are known a ***producer's risk*** and ***consumer's risk*** respectively. If we consider consequences of both types of errors, we find that type II error more serious than type I error.

For example, suppose a drug is administered to a few patients to cure the particular disease and the drug is curing the disease. But if it is discontinued by claiming that the drug has adverse effect, it is type I error. In contrary to this, if the drug has, in fact, adverse effect and continued to administer to patients claiming that the drug has good effect, then it is type II error.

How to remember type I and type II errors?

Please remember the words "ROUTINE FOR FUN." Using only the consonants from those words (RouTiNe FoR FuN), we can easily remember that a type I error is RTN: reject true null (hypothesis), whereas a type II error is FRFN: failure to reject a false null (hypothesis).

Example 3: (*Identifying Type I and Type II Errors*)

Assume that we are conducting a hypothesis test of the claim that $p > 0.5$. Here are the null and alternative hypotheses: $H_0: p = 0.5$ versus $H_1: p > 0.5$

Give statements identifying **(a)** a type I error. **(b)** a type II error.

Solution:

(a) A type I error is the mistake of rejecting a true null hypothesis, so this is a type I error: Conclude that there is sufficient evidence to support $p > 0.5$, when in reality $p = 0.5$.

(b) A type II error is the mistake of failing to reject the null hypothesis when it is false, so this is a type II error: Fail to reject $p = 0.5$ (and therefore fail to support $p > 0.5$) when in reality $p > 0.5$.

Controlling Type I and Type II Errors

One-step in our standard procedure for testing hypotheses involves the selection of the significance level α, which is the probability of a type I error. However, we don't select β [P(type II error)]. It would be great if we could always have $\alpha = 0$ and $\beta = 0$, but in reality that is not possible, so we must attempt to manage the α and β error probabilities. Mathematically, it can be shown that α, β, and the sample size n are all related, so when you choose or determine any two of them, the third is automatically determined. The usual practice in research and industry is to select the values of α and n, so the value of β is determined. Depending on the seriousness of a type I error, try to use the largest α that you can tolerate. For type I errors with more serious consequences, select smaller values of α. Then choose a sample size n as large as is reasonable, based on considerations of time, cost, and other relevant factors. (Sample size determinations were discussed in previous Chapters.) The following practical considerations may be relevant:

1. For any fixed α, an increase in the sample size n will cause a decrease in β. That is, a larger sample will lessen the chance that you make the error of not rejecting the null hypothesis when it's actually false.
2. For any fixed sample size n, a decrease in α will cause an increase in β. Conversely, an increase in α will cause a decrease in β.
3. To decrease both α and β, increase the sample size n.

These two types of errors depend on each other and cannot be minimized simultaneously for a test of hypothesis for a fixed sample size. Lowering the value of α will raise the value of β and lowering the value of β will raise the value of α. However, we can decrease both α and β simultaneously by increasing the sample size but due to the limit of resources, it is not possible. Therefore, for given sample we minimize more serious error after fixing up less serious error. Thus, we fix α, the size of ***type* I *error*** and then try to obtain a criterion which minimizes β, the size of ***type* II *error***. A test with both α and β small is desirable.

Power of a Test

We use β to denote the probability of failing to reject a false null hypothesis (type II error). It follows that $1 - \beta$ is the probability of rejecting a false null hypothesis. Statisticians refer to this probability as the power of a test, and they often use it to gauge the test's effectiveness in recognizing that a null hypothesis is false.

Definition: The ***power*** of the hypothesis test is the probability $(1 - \beta)$ of rejecting a false null hypothesis, which is computed by using a particular significant level α and a particular value of the population parameter that is an alternative to the value assumed true in the null hypothesis. That is, the power of hypothesis test is the probability of supporting an alternative hypothesis that is true.

8.4 Test statistic, Critical Region and Acceptance region, Significance level, Critical Value, and P–Value

Test statistic

Definition: The ***test statistic*** is a value calculated from the sample data, and it is used in making the decision whether the null hypothesis should be accepted or rejected in our hypothesis test.

The test statistic is found by converting the sample statistic (such as the sample proportion $\hat{p}$, or the sample mean $\bar{X}$, or the sample standard deviation S) to a score (such as z, t, or χ^2) with the assumption that the null hypothesis is true. The test statistic can therefore be used for determining whether there is significant evidence against the null hypothesis. In this chapter we consider hypothesis tests involving means and. Based on results from preceding chapters about the sampling distributions of proportions, means and standard deviation, we use the following test statistics:

The test statistic fits the common format of:

$$\textit{Test statistic} = \frac{(\textit{Sample statistic}) - (\textit{Claimed value of population parameter})}{(\textit{Standard error of sampal statistic})}$$

$$= \frac{\hat{\theta} - E(\hat{\theta})}{S.E.(\hat{\theta})}$$

Test statistic for mean μ: $\quad Z = \frac{\bar{X} - E(\bar{X})}{S.E.(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ if σ^2 is known

or, $t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ if σ^2 is not known.

Test Statistic for proportion p: $\quad Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$

Test statistic for standard deviation σ: $\quad \chi^2 = \frac{(n-1)s^2}{\sigma^2}$

The above test statistic for a proportion does not include the continuity correction that we usually use when approximating a binomial distribution by a normal distribution. When working with proportions in this chapter, we will work with large samples so the continuity correction can be ignored because its effect is small. Also, the test statistic for a mean can be based on the *normal* or *Student t distribution*, depending on the conditions that are satisfied. When choosing between the normal or student t distributions, this chapter will use the same criteria described as before.

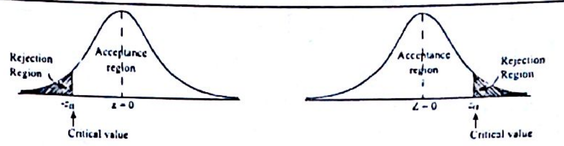
Critical Region and Acceptance region

The sampling distribution of test statistic is divided into two regions: ***Rejection region*** and ***Acceptance region***. The region where the true null hypothesis H_0 is rejected is called ***critical region*** or ***rejection region***. The critical region is the set of all values of the test statistic that cause us to reject the null hypothesis. In other words, the area covered by ***Type I error*** in the probability density curve is known as critical region.

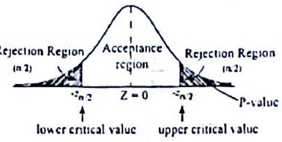
The region where the true null hypothesis H_0 is accepted is called ***acceptance region***. Thus acceptance region consists of all the values of test statistic for which H_0 is accepted.

Critical value:

A ***critical value*** is any value that separates the critical regions (where we reject the null hypothesis) from the values of the test statistic that do not lead to rejection of the null hypothesis. That is, the value of the test statistic that separates the critical and acceptance regions is called the ***critical value*** or ***significant value***. The critical values depend on the nature of the null hypothesis, the sampling distribution that applies, and the significance level α.



(i) Left tailed test (ii) Right tailed test



(iii) Two tailed test

Significance level

The significance level (denoted by α) is the probability that the test statistic will fall in the critical region when the null hypothesis is actually true. If the text statistic falls in the critical region, we will reject the null hypothesis, so α is the probability of making the mistake of rejecting the null by prosthesis when it is true.

So the maximum size of ***type I error***, which we are prepared to risk is known as the significance level. So, $\alpha = P$(type I error). We define the confidence level for a confidence interval to be the probability $1 - \alpha$.

Commonly used levels of significance are 1% and 5%. Significant at 5% is stated as ***significant*** (for moderate precision) and 1% level as ***highly significant*** (for high precision). ***Not significant*** is related to the acceptance of the null hypothesis. If we adopt α = 5% level of significance, it shows that in 5 true samples out of 100, we are likely to reject a correct H_0. In other words, we are ready to take a 5% risk of rejecting true null hypothesis H_0, i.e. the probability of rejecting the true H_0 is 5% = 0.05.

So it is a value indicating the percentage of sample values that is outside certain limits assuming the hypothesis is correct; that is the probability of rejecting the null hypothesis when it is true.

8.5 One–Tailed Test and Two–Tailed Test

The *tails* in a distribution are the extreme region bounded by critical values. Some hypothesis tests are two tailed. Some are right tailed, and some are left tailed

One–Tailed: A hypothesis test in which there is only one rejection is called ***one–tailed test***. It can be left or right tailed.

Left–tailed test: The one tailed test is called ***left–tailed test*** or ***lower tailed test*** if the rejection region is in the extreme left region (tail) under the distribution curve.

Right–tailed test: The one tailed test is called ***right–tailed test*** or ***upper tailed test*** if the rejection region is in the extreme right region (tail) under the distribution curve.

Two–tailed test A hypothesis test in which the critical region is the two extreme regions (tails) under the curve (i.e., there are two rejection region), is called a ***two–tailed test***.

Important Properties of a Test of Hypothesis

1. The type I error and type II error and related. A decrease in the probability of one generally results in an increase in the probability of the other
2. The size of the critical region, and therefore the probability of committing a type I error, can always be reduced by adjusting the critical values(s).
3. An increase in the sample size n will reduce α and β simultaneously.
4. If the null hypothesis is false, β is a maximum when the true value of a parameter approaches the hypothesized value. The greater the distance between the true value and the hypothesized value, the smaller β will be.

Techniques of identifying one and two tests for single mean

A test which seeks to establish whether two values are significantly different or whether the sample has been drawn from parent population but the direction of difference (i.e., word *more* or *less*) is not specified will required a ***two–tailed test.***

A test which seeks to compare two values and the direction of the difference is specified, will require ***one tailed test.*** i.e. for comparative study of identifying the population parameter, we use one tailed test.

1. **How to detect left–tailed test?**

   The test will be left tailed if:

   (i) Problem statement has keywords less than, decreased, reduced, inferior, minority, below, smaller, shorter, at least etc.

   (ii) Left–tailed test is used if population parameter has been shifted to a number less than a specified number.

   (iii) If the alternative hypothesis has a less than (<) sign, we use left tailed test.

2. **How to detect right–tailed test?**

   The test will be right–tailed test if

   (i) Problem statement has keywords greater than, increase more than, above, superior, enhance, gained, improvement, at most etc.

   (ii) Population parameter has been shifted to a number more than a specified number.

   (iii) The alternative hypothesis has a greater than (>) sign.

3. **How to detect two–tailed test?**

   The test will be two tailed test if

   (i) The problem statement has the keywords changed, different from, no longer than, same, equal, exactly, unbiased etc. (i.e., direction of difference is not given).

   (ii) The population parameter has been shifted away from a specified number in either direction, increased or decreased.

   (iii) The alternative hypothesis has a not equal to (≠) sign.

For example:

1. The test for testing the mean of a population

   $H_0: \mu = \mu_0$ against the alternative hypothesis

   $H_1: \mu > \mu_0$ (Right tailed)

   or, $H_1: \mu < \mu_0$ (left tailed) is one tailed test.

2. If the hypothesis is set up as follows:

   $H_0: \mu = \mu_0$, against the alternative hypothesis

   $H_1: \mu \neq \mu_0$ ($\mu > \mu_0$ or $\mu < \mu_0$) then the test is called two tailed test.

8.6 Simple and composite hypothesis

If a statistical hypothesis completely determines the population, it is called a ***simple hypothesis.*** If a statistical hypothesis partially specifies the population, it is called ***composite hypothesis.***

For example: In a normal population if a hypothesis determines both the parameter μ and σ^2, then it is called *simple hypothesis*. But if it determines only one of the two parameters μ and σ^2, then it is called *composite hypothesis.*

Simple hypothesis $H: \mu = \mu_0,\ \sigma^2 = \sigma_0^2$

Composite hypothesis (1) $H: \mu = \mu_0$; (2) $H: \mu = \mu_0$; (3) $H: \sigma^2 = \sigma_0^2$ etc.

8.7 Hypothesis Test or Test of Significance

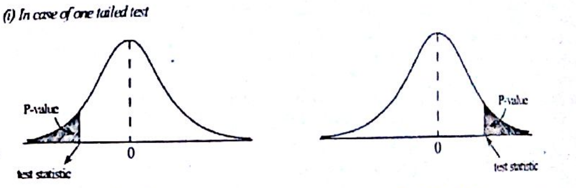
Focus on the use of the rare event rule of inferential statistics: ***If, under a given assumption, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct.*** However, if the probability of a particular observed sample result is not very small, then we do not have sufficient evidence to reject the assumption. In Section we will describe the specific steps used in hypothesis testing.

Definition: The ***P*-Value** (or ***probability value***): The probability that the value of test statistic is *at least as extreme* as the computed value of the test statistic on the basis of the sample data under H_0 is called its ***P–value*** or ***tail probability***. The null hypothesis is rejected if the *P*-value is very small, such as 0.05 or less.

P-values can be found by using the procedures summarized in the following figures:

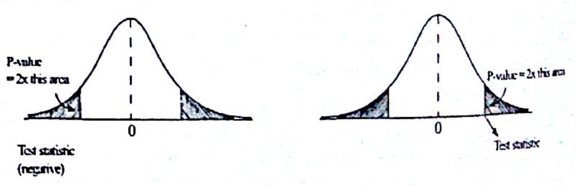
Finding *P*–value

(i) In case of one tailed test



Left Tailed Test   Right Tailed Test

(ii) In case of two tailed tests



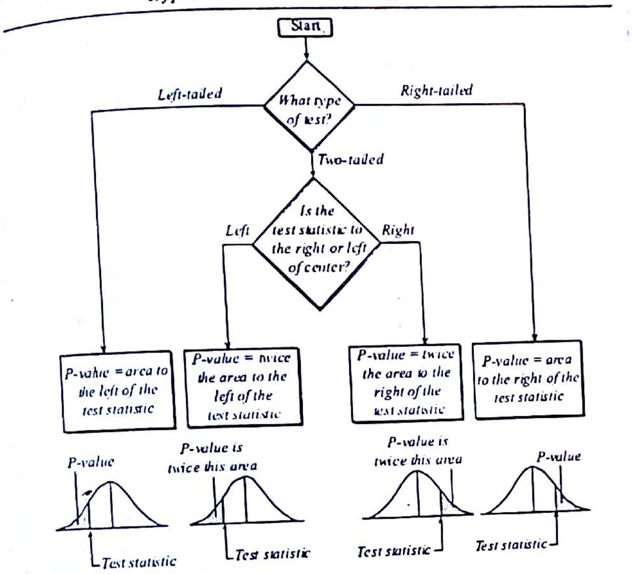


Fig. Procedure for finding P-values

Example 4: (*Finding p-values*): First determine whether the given conditions result in a right-tailed test, a left-tailed test, or a two-tailed test, then use above Figure to find the *P*-value, then state a conclusion about the null hypothesis.

(a) A significance level of $\alpha = 0.05$ is used in testing the claim that $p > 0.25$, and the sample data result in a test statistic of $z = 1.18$.

(b) A significance level of $\alpha = 0.05$ is used in testing the claim that $p \neq 0.25$, and the sample data result in a test statistic of $z = 2.34$.

Solution:

a) With a claim of $p > 0.25$, the test is right-tailed. Because the test is right-tailed, the *P*-value is the area to the right of the test statistic $z = 1.18$. Using the Table A-3 the area to the right of $z = 1.18$ is 0.1190. The *P*-value of 0.1190 is greater than the significance level $\alpha = 0.05$, so we fail to reject the null hypothesis. The *P*-value of 0.1190 is relatively large, indicating that the sample results could easily occur by chance.

b) With a claim of $p \neq 0.25$, the test is two-tailed. Because the test is two-tailed, and because the test statistic of $z = 2.34$ is to the right of the center, the *P*-value is twice the area to the right of $z = 2.34$. Using the Table A-3 the area to right of $z = 2.34$ is 0.0096, so *P*-value $= 2 \times 0.0096 = 0.0192$. The *P*-value of 0.0192 is less than or equal to the significance level, so we reject the null hypothesis. The small *P*-value of 0.0192 shows that the sample results are not likely to occur by chance.

Wording the Final Conclusion

The conclusion of rejecting the null hypothesis or failing to reject it is fine for those of us with the wisdom to take a statistics course, but we should use simple, nontechnical terms in stating what the conclusion really means. The following Figure summarizes a procedure for wording of the final conclusion. Note that only one case leads to wording indicating that the sample data actually support the conclusion. If you want to support some claim, state it in such a way that it becomes the alternative hypothesis, and then hope that the null hypothesis gets rejected. For example, to support the claim that the mean body temperature is different from 98.6°, make the claim that $\mu \neq 298.6°$. This claim will be an alternative hypothesis that will be supported if you reject the null hypothesis, H_0: $\mu = 98.6°$. If, on the other hand, you claim that $\mu = 98.6°$, you will either reject or fail to reject the claim; in either case, you will never support the claim that $\mu = 98.6°$.

Accept/Fail to Reject

Some texts say "accept the null hypothesis" instead of "fail to reject the null hypothesis." Whether we use the term accept or fail to reject, we should recognize that we are not proving the null hypothesis; we are merely saying that the sample evidence is not strong enough to warrant rejection of the null hypothesis. ***It's like a jury's saying that there is not enough evidence to convict a suspect***. The term accept is somewhat misleading, because it seems to imply incorrectly that the null hypothesis has been proved. (It is misleading to state "there is sufficient evidence to accept the null hypothesis.") The phrase fail to reject says more correctly that the available evidence is not strong enough to warrant rejection of the null hypothesis. In this text we will use both the terminology ***fail to reject*** the null hypothesis, or ***accept*** the null hypothesis.

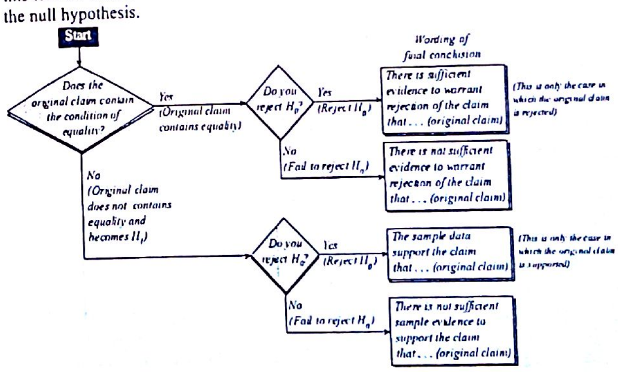


Fig. Wording of final Conclusion

8.7.1 Procedure for Testing of Hypothesis (Test of Significance)

From the above figure, we see that the following steps should be considered while testing a hypothesis.

Step 1. Set up the *null hypothesis* H_0

Step 2. Set up the *alternative hypothesis* H_1. It may be either one tailed or two tailed alternative.

Step 3. *Level of significance* (α): The level of significance is fixed in advance. Usually it is taken as 5%.

Step 4. *Test statistic*: Under null hypothesis, define and compute test statistic as

$$\text{Test statistic} = \frac{(\text{Sample statistic}) - (\text{Claimed value of population parameter})}{(\text{Standard error of sampal statistic})}$$

$$= \frac{\hat{\theta} - E(\hat{\theta})}{S.E.(\hat{\theta})}$$

Step 5. *Critical value*: Obtain critical value(s) or tabulated value and critical region from the appropriate table according to predetermined level of significance.

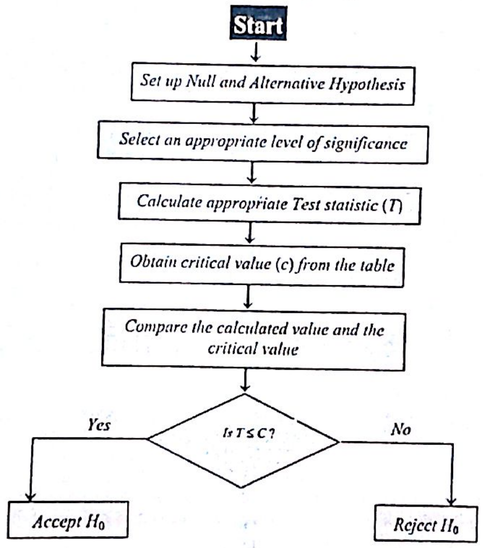
Step 6. *Decision*:

(i) If the calculated value of test statistic is less than or equal to tabulated value, then H_0 accepted and reject H_1 rejected which means that there is no significant difference between the sample statistic and the population parameter and the difference is due to sampling fluctuations.

(ii) If the calculated value of test statistic is greater the tabulated value, then H_0 is rejected and H_1 is accepted which means that there is significant difference between the sample statistic and population parameter and difference observed is not due to sample fluctuations.

Step 7. *Conclusion*: Write the conclusion of the test in simple language.

Flow Chart for Testing of Hypothesis



This method of testing statistical hypothesis is known as *Rejection Region Method or traditional method.*

Test of Significance: The terms *test of significance* and *testing of hypothesis* both refer more or less the same statistical process, of whether or not there is a difference between two values or two sets of results.

Note: We can also use another method called '*P–value Estimation Method*' for testing of a statistical hypothesis.

P–value Estimation Method

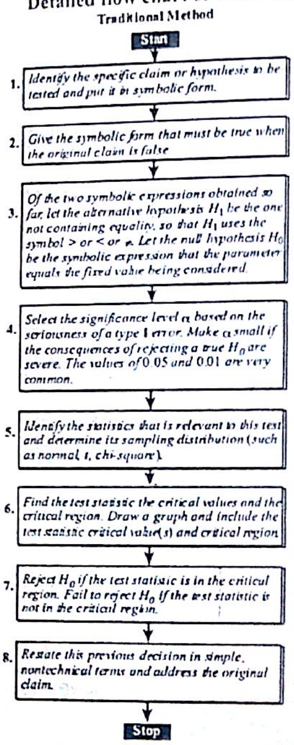
Step. 1, 2, 3, 4 are same as in Rejection Region Method.

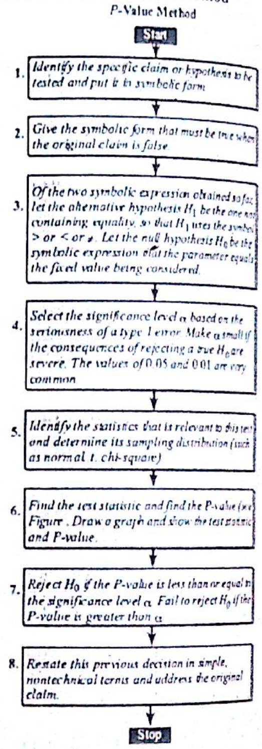
Step 5. Find the *P–value* of the test statistic under H_0 in step 4.

Step 7. If $P\text{–value} \le \alpha$ we reject H_0 at α level of significance. If $P\text{–value} > \alpha$ we fail to reject H_0 at α level of significance.

Step 8. *Decision*: Write the conclusion of the test in simple language.

Detailed flow chart of Traditional Method and *P*-Value Method





8.8 Relation between Confidence Interval and Testing of hypothesis

If a sample statistic, say $\hat{\theta}$, is computed from the sample drawn from an unknown population and assumed as an estimate of θ, falls within the confidence limits, say $\hat{\theta}_1$ and $\hat{\theta}_2$, then we conclude that it would have come from the population whose parameter is θ. Otherwise, if the sample statistic falls outside the confidence limits, then we conclude that the sample is not from the population whose parameter is θ. Therefore, we have

$$P(\hat{\theta}_1 \le \hat{\theta} \le \hat{\theta}_2) = 1 - \alpha,\ P = \frac{\alpha}{2},\ P(\hat{\theta} > \hat{\theta}_2) = \frac{\alpha}{2}$$

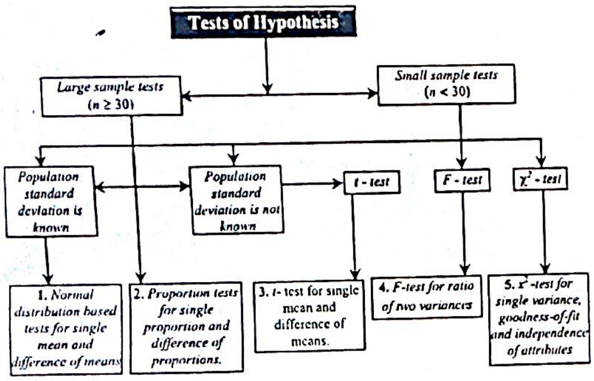
Then $P(\hat{\theta}_1 \le \hat{\theta} \le \hat{\theta}_2) = 1 - \alpha$ is termed as the acceptance region and

$$P(\hat{\theta} < \hat{\theta}_1) + P(\hat{\theta} > \hat{\theta}_2) = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$$

is termed as the rejection region.

8.9 Classification of Hypothesis Tests

Depending upon whether the ***population parameters are known or not and whether the sample size is large or small***, various tests are classified accordingly as shown in the following figure in the form of a flow diagram.



8.10 Test of Significance for large samples (i.e., Z–test)

In this section we discuss about the test of significance when the samples are large. For practical purpose, the sample is considered large if $n \ge 30$. The large sample is generally desirable when the units in the population under study are not homogeneous or uniform. To get more reliable results about the population parameter, small sample is not sufficient for heterogeneous population, so large sample test should be carried out. Large sample test is generally used when sample size n is greater than or equal to 30 (i.e., $n \ge 30$). ***The test which is applied in the case of large samples or σ known case is called*** Z-test.

Assumptions of Z-test

The Z-test is used under the following assumptions:

1. The sample is a simple random sample.(All samples of same size have an equal chance of being selected)
2. The value of the population standard deviation σ is known.
3. The samples are independent
4. Either or both of these conditions are satisfied: The population from which the samples are drawn is normally distributed or, $n \ge 30$.

8.10.1 Applications of Z-test

The Z-test has a large number of applications such as

1. Test of significance of sampling of variables
   (i) Test of significance of a single mean (testing the significance of difference between sample mean and population mean)
   (ii) Test of significance of difference between two means (testing the significance of difference between two independent sample means)
2. Test of significance of sampling attributes.
   (i) Test of significance of a single sample proportion (testing the significance of difference between sample proportion and population proportion)
   (ii) Test of significance of difference between two proportions (testing the significance of difference between two sample proportions).

8.10.2 Test of significance of a single Mean

(σ^2 Known or large sample case)

Suppose a random sample of size n is drawn from a normal population with mean μ and variance σ^2. The sample mean $\bar{X}$ is also a normal variate with mean $E(\bar{X}) = \mu$ and variance $V(\bar{X}) = \sigma^2/n$. Then Z-statistic

$$Z = \frac{\bar{X} - E(\bar{X})}{S.E.(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$ So, the steps in test of significance of a single mean for large samples ($n \ge 30$) are as follows:

Step 1. ***Set up Null Hypothesis:*** H_0: $\mu = \mu_0$ [i.e., the population mean has a specified value μ_0. In other words, there is no significant difference between sample mean $\bar{X}$ and population mean μ, or the sample has been drawn from given large normal population with mean μ_0 and standard deviation σ.]

Step 2. ***Set up Alternative Hypothesis:*** H_1: $\mu \ne \mu_0$ [*Two tailed test*] [i.e., population mean is not equal to μ_0. In other words, there is significant difference between sample mean $\bar{X}$ and population mean μ, or the sample has not been drawn from normal population with mean μ_0 and standard deviation σ or, H_1: $\mu > \mu_0$ (*Right tailed test*) [i.e., population mean is greater than specified mean μ_0]
or, H_1: $\mu < \mu_0$ (*Left tailed test*) [i.e., population mean is less than specified mean μ_0]

Step 3. ***Level of significance*** (α): Choose the appropriate level of significance. The most commonly used level of significance is 5% unless otherwise stated.

Step 4. ***Test Statistic:*** Under H_0, test statistic is given by

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \text{ if } \sigma^2 \text{ is known.}$$

If σ^2 is not known, then ***for large sample*** we use S^2, an unbiased estimator of σ^2 in place of σ^2, so $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0, 1)$ [$\because \hat{\sigma} = S$ *for large samples*]

where $S^2 = \frac{1}{n-1} \Sigma(X_i - \bar{X})^2$

[We can use $S^2 = \frac{1}{n}\Sigma(X_i - \bar{X})^2$ because for large sample

$$\frac{1}{n}\Sigma(X_i - \bar{X})^2 = \frac{1}{n-1}\Sigma(X_i - \bar{X})^2]$$

Step 5. ***Critical value***: Obtain critical value or tabulated value of Z (i.e., z_α or $z_{\alpha/2}$) at pre-specified level of significance.

Step 6. ***Decision:***

(i) If $|Z| > z_{\alpha/2}$ (for two tailed Z-test) or, $|Z| > z_\alpha$ (for one tailed Z-test) it is significant and reject H_0. Hence accept H_1 [i.e. the population mean has not a specified value μ_0. In other words, there is significant difference between sample mean $\bar{X}$ and population mean μ, or the sample has not been drawn from a normal population with mean μ_0 and standard deviation σ.]

(ii) If $|Z| \le z_{\alpha/2}$ (for two tailed) or, $|Z| \le z_\alpha$ (for one tailed) it is not significant and accept H_0. Hence reject H_1. [i.e., the population mean has specified value μ_0. In other words, there is no significant difference between sample mean $\bar{X}$ and population mean μ, or sample has been drawn from a normal population with mean μ_0 and standard deviation σ.]

Note:

1. A small risk $\alpha = 1\%$ is used if it is a matter of life or death.
2. Test always involves '*risks of making false decision*".
3. If the difference between two means is zero and if our test suggests rejection of the null hypothesis we commit ***Type I error***. If the difference between two means is not zero but our test suggests acceptance of null hypothesis we commit type II error.
4. **Summary of Decision Rule for Z-test**

Alternative hypothesis	*Reject H_0 if*
$\mu < \mu_0$	$Z < -z_\alpha$ i.e., $\|Z\| > z_\alpha$
$\mu > \mu_0$	$Z > z_\alpha$ i.e., $\|Z\| > z_\alpha$
$\mu \ne \mu_0$	$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$ i.e., $\|Z\| > z_{\alpha/2}$

5. **Choosing Z and t distribution**

Method	*Conditions*
Use normal (Z) distribution	σ known and normally distributed population or σ known and $n \ge 30$
Use t distribution	σ not known and normally distributed population or σ not known and $n \ge 30$
Use a nonparametric method or bootstrapping	Population is not normally distributed and $n < 30$
Notes: 1. For large sample i.e., $n \ge 30$, t distribution and Z distribution give same result.	

2. **Criteria for deciding whether the population is normally distributed:**
   Population need not be exactly normal, but it should appear to be somewhat symmetric with one mode and no outliers.
3. **Sample size $n \ge 30$:** This is a commonly used guideline, but sample sizes of 15 to 30 are adequate if the population appears to have a distribution that is not far from being normal and there are no outliers.

Example 5: The mean income of the random sample of 100 employees of an industrial concern was found to be Rs. 3000. If the standard deviation of the population was 25, find the standard error of the mean and also test whether the sample mean differs from the population mean of 2850. [*TU, MBS/MPA 2069*]

Solution: Given: Mean income of the sample $(\bar{x})$ = Rs. 3000

Size of sample (n) = 100, population *s.d.* (σ) = Rs. 25

population mean (μ) = Rs. 2850

Standard error of mean $(\bar{x}) = S.E.(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{100}} = 2.5$

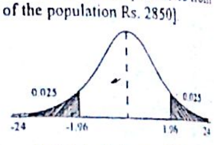
Let μ be the true population mean income.

Step 1. ***Null Hypothesis***: H_0: $\mu = 2850$ [i.e. the mean income of the sample does not differ from mean income of population Rs. 2850]

Step 2. ***Hypothesis***: H_1: $\mu \ne 2850$[i.e. the mean income of the sample differs from the mean income of the population Rs. 2850]

Step 3. ***Test statistic***: Under null hypo-thesis H_0, the test statistic

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{3000-2850}{2.5} = 60.$$



Two tailed critical region, |Z| > 1.96

Step 4. ***Level of significance*** (α): Take $\alpha = 5\% = 0.05$ as not mentioned.

Step 5. ***Critical value***: At $\alpha = 5\% = 0.05$, $\alpha/2 = 0.025$, $z_{\alpha/2} = z_{0.025} = 1.96$

Step 6. ***Decision*** : Here $z_{\alpha/2} < |z|$. So, H_0 is rejected i.e., the mean income of the sample differs from the mean income of the population.

Determination of α and β

Example 6: A process for making steel pipe is under control if the diameter of the pipe has a mean of 3 inches with a standard deviation of 0.0250 inches. To check whether the process is under control, a random sample of size $n = 30$ is taken each day and the null hypothesis $\mu = 3$ is rejected if $\bar{X}$ is less than 2.9960 or greater than 3.0040. Find **(a)** The probability of a ***Type I error***;

(b) The probability of a ***Type II error*** when $\mu = 3.0050$ inches.

Solution: (a) $\alpha = P$(type I error) = P(reject H_0 when H_0 is true

$= P(\bar{X} < 2.9960$ or $\bar{X} > 3.0040$ when $\mu_0 = 3)$

$$= P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < \frac{\frac{2.9960-3}{0.0250}}{\sqrt{3}} \text{ or } \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \frac{\frac{3.0040-3}{0.0250}}{\sqrt{30}}\right)$$

$= P(Z < -0.876$ or $Z > 0.876)$

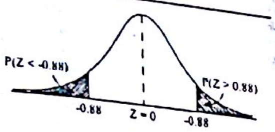
$= P(Z < -0.876) + P(Z > 0.876)$ [*being disjoint event*]

$= 2P(Z > 0.876)$ [*By symmetry*]

$= 2[1 - P(Z < 0.8106)]$

$= 2[1 - 0.8106]$

$= 0.3788$



(b) Probability of type *II* error (β) =

$P(\text{Accept } H_0 \text{ when } H_0 \text{ is false})$

$= P(\text{Accept } H_0 \text{ when } H_1 \text{ is true})$

$= P(\bar{X} \geq 2.9960 \text{ and } \bar{X} \leq 3.0040 \text{ when } \mu_1 = 3.005)$

$= P(2.9960 \leq \bar{X} \leq 3.0040 \text{ when } \mu_1 = 3.005)$

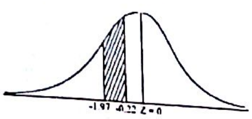
$$= P\left[\left(\frac{2.9960 - 3.005}{0.025}\right)\sqrt{30} \leq \frac{\bar{X} - \mu_1}{\sigma\sqrt{n}} \leq \left(\frac{3.0040 - 3.005}{0.025}\right)\sqrt{30}\right]$$

$= P(-1.97 \leq Z \leq -0.22)$

$= F(-0.22) - F(-1.97)$

$= 0.4129 - 0.0244$ [*From Normal table*]

$= 0.3885$



Example 7: Suppose that for a given population with $\sigma = 8.4$ square inches we want to test the null hypothesis $\mu = 80$ square inches against the alternative hypothesis $\mu < 80$ square inches on the basis of a random sample of size $n = 100$.

(a) If the null hypothesis is rejected for $\bar{X} < 78$ square inches and otherwise it is accepted, what is the probability of *type I errors*?

(b) What is the answer to part (a) if the null hypothesis is $\mu \geq 80$ square inches instead of $\mu = 80$ square inches?

Solution:

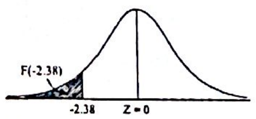
(a) α = probability of type I error

$= P(\text{type I error})$

$= P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$

$= P(\bar{X} < 78 \text{ when } \mu_0 = 80)$

$$= P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{78 - 80}{8.4}\sqrt{100}\right)$$



$= P(Z < -2.38) = 0.0087$ [*From Normal table*]

(b) If $H_1: \mu \geq 80$, then $\alpha \leq 0.0087$.

Other Solved Examples

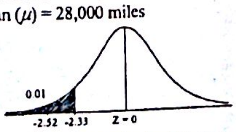
Example 8: (*Large sample test of mean tire life, σ unknown*): A trucking firm is suspicious of the claim that the average lifetime of a certain tires is at least 28,000 miles. To check the claim, the firm puts 40 of these tires on its trucks and gets a mean lifetime 27,436 miles with a standard deviation of 1,348 miles. What can it conclude if the probability of type *I* error (α) is to be at most 0.01 ?

[*TU, BE 2062 Jestha / 2066 Magh / 2067 Shrawan (BIE) / 2067 Mangsir*]

Solution: Given: Sample size (n) = 40, population mean (μ) = 28,000 miles

sample mean ($\bar{x}$) = 27,463 miles, sample *s.d.* (s) = 1348 miles

Standard error of mean ($\bar{x}$) $= \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$



Left tailed critical region, Z < -2.33

$$= \frac{27436 - 28000}{1248/\sqrt{40}} = -2.$$

[$\because \hat{\sigma} = s$ *for large sample*]

Now let μ be true mean lifetime.

Step 1. ***Set up hypotheses***

Null Hypothesis H_0: $\mu \geq 28{,}000$ miles

Alternative hypothesis H_1: $\mu < 28{,}000$ miles [*Left tailed test*]

Step 2. ***Level of significance:*** $\alpha \leq 0.01$

Step 3. ***Test statistic:*** Under null hypothesis H_0, since $n > 30$, the

Z–statistic is $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = -2.52$

Step 4. ***Critical value:*** At $\alpha \leq 0.1$, $z_\alpha = -2.33$

Step 5. ***Decision:*** $z < -z_\alpha$ i.e., $|z| > z_\alpha$ at $\alpha \leq 0.01$. So it is significant and null hypothesis H_0 is rejected at $\alpha \leq 0.01$. In other words, the trucking firm's suspicion that $\mu < 28{,}000$ miles is confirmed.

P–Value Approach

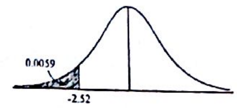
Steps 1,2,3 are same as above.

Step 4. Determination of *P*–value

P–value = $F(-2.52) = 0.0059$

[$\because$ *test is one tailed*]

Here *P*–value < α. So we reject H_0.



99% confidence interval for the mean lifetime μ is given by

C.I. for $\mu = \bar{x} \pm z_\alpha S.E.(\bar{x}) = 27463 \pm (-2.33)\dfrac{s}{\sqrt{n}}$

$$= 27463 \pm (-2.33)\left(\frac{1348}{\sqrt{40}}\right) = (27463 \pm 496.61) = (26966.39, 27959.61)$$

Example 9: The breaking strength of cables produced by a manufacturer have mean 815 kg and standard deviation 45 kg. By a new technique in the manufacturing process it is claimed that the breaking strength can be increased. To test this claim a sample of 50 cables is tested and it is found that the mean breaking strength is 840 kg. Can we support the claim at 0.01 level of significance? [*TU, BE 2058 Shrawan/ Purbanchal Uni. BE 2001*]

Solution: Given: sample size (n) = 50, sample mean ($\bar{x}$) = 840 kg

population mean (μ) = 815 kg (=μ_0), population s.d (σ) = 45 kg

Now, let μ be true mean breaking strength of cables

Step 1. ***Set up Hypotheses***

Null hypothesis $H_0: \mu = 815$ kg (i.e., $\mu_0 = 815$)

Alternative hypothesis $H_1: \mu > 815$ kg (*Right tailed test*)

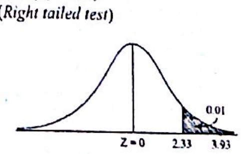
Step 2. ***Test statistic:*** Under null hypothesis H_0, test statistic is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{840 - 815}{45/\sqrt{50}} = 3.93$$

Step 3. ***Level of significance:*** $\alpha = 0.01$

Step 4. ***Critical value***

At $\alpha = 0.01$, $z_\alpha = 2.33$ [*from table*]



Right tailed critical region, Z > 2.33

Step 5. ***Decision:*** $z > z_\alpha$ i.e., $|z| > z_\alpha$ at $\alpha = 0.01$. Therefore, we reject null hypothesis. This means that the breaking strength of the cable has been increased by a new technique.

Example 10: ***(Large sample, σ unknown)***: A random sample of size 60 from a large population gave the following distribution:

Age	5–10	10–15	15–20	20–25	25–30
No. of children	5	19	24	8	4

Test the hypothesis whether this sample comes from a population with mean 20. Also calculate the 95% confidence limits for the population mean μ.

Solution: Using calculator

Sample mean ($\bar{x}$) = 16.42 , Sample *s.d.* (s) = 5.05, Sample size(n)= 60

Let μ be the true population mean.

Step 1. ***Null hypothesis:*** $H_0 : \mu = 20$ (i.e., population mean is 20)

Alternative hypothesis $H_1 : \mu \neq 20$ (*Two tailed test*)

Step 2. ***Test statistic:*** Under null hypothesis, test statistic is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad [\because \hat{\sigma} = s \textit{ for large sample}]$$

$$= \frac{16.42 - 20}{5.05/\sqrt{60}} = -5.491. \qquad \text{So, } |z| = 5.491$$

Step 3. ***Level of significance:*** $\alpha = 5\%$ (as not mentioned)

Step 4. ***Critical value:*** The tabulated value at $\alpha = 5\% = 0.05$ is

$z_{\alpha/2} = z_{0.025} = 1.96$ [$\because$ *two tailed test*]

Step 5. ***Decision :*** Since $z < z_{\alpha/2}$ i.e. $|z| > z_{\alpha/2}$, it is significant and H_0 is rejected. That is the population mean is not 20.

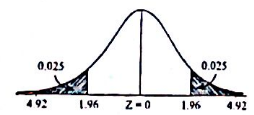
For 95% confidence limits for μ

95% large sample confidence interval for μ

$$= \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$= 16.2 \pm 1.96 \times \frac{5.46}{\sqrt{50}} = 16.2 \pm 1.51$$

$$= (14.69, 17.71)$$



Two tailed critical region, |Z| > 1.96

The required limits that will include the true population mean are

L= Lower confidence limit = 14.69 and U = Upper confidence = 17.71

Example 11: ***(Large sample, σ unknown)***: A machine shop is interested in determining a measure of the current year's sales revenue in order to compare it with known results from last year. From the 9682 sales invoices (bills) for the current year to data, the management randomly selected invoices and from each recorded X, the sales revenue per invoice. Using the following data summary, test the hypothesis that the mean revenue per invoice is Rs. 6.35, the same as last year, versus the alternative hypothesis that the mean revenue per invoice is different from Rs. 6.35, with $n = 400$, $\alpha = 0.05$. *[TU, BE, 2064 Poush]*

Data Summary: $\sum_{i=1}^{400} X_i$ = Rs. 2464.40, $\sum_{i=1}^{400} X_i^2 = 16{,}156.728$.

Solution: Given: $n = 400$

Population mean revenue per invoice (μ) = Rs. 6.35

Sample variance $S^2 = \frac{1}{n-1}\Sigma(X_i - \bar{X})^2 = \frac{1}{399}\left[16156.728 - \frac{(2464.40)^2}{399}\right] = 2.440$

$\Rightarrow s = 1.56$

$\bar{X} = \frac{\Sigma X_i}{n} = \frac{2464.40}{400} = 6.16 = \bar{x}$

[**Note:** $S^2 = \frac{1}{n-1}\Sigma(X_i - \bar{X})^2 \approx \frac{1}{n}\Sigma(X_i - \bar{X})^2$ as $n - 1 \approx n$ *for large n.*

$S^2 = \frac{1}{n}\Sigma(X_i - \bar{X})^2 = \frac{1}{n}\left[\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right] = \frac{1}{400}\left[16156.728 - \frac{(2464.40)^2}{400}\right] = 2.43$]

Now,

Step 1. ***Null hypothesis:*** $H_0 : \mu = 6.35$

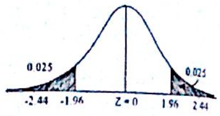
Alternative hypothesis: $H_1 : \mu = 6.35$

Step 2. ***Test statistic:*** Under null hypothesis H_0,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad [\because \hat{\sigma} = s \textit{ for large n}]$$

$$= \frac{6.16 - 6.35}{1.56/\sqrt{400}} = -2.44$$

So, $|z| = 2.44$



Step 3. ***Critical value:***

Since test is two tailed the tabulated value of z at $\alpha = 0.05$ is $z_{\alpha/2} = z_{0.025} = 1.96$

Step 5. ***Decision:*** Since $|z| > z_{\alpha/2}$, z lies in rejection region. So, it is significant. Hence H_1 is accepted by rejecting H_0. That is, the mean revenue per invoice is different from Rs. 6.35.

95% large sample confidence interval for estimating μ

$$C.I. \text{ for } \mu = \bar{x} \pm z_{\alpha/2}\, S.E(\bar{x}) = \bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$= 6.16 \pm (1.96)\frac{1.56}{\sqrt{400}} = 6.16 \pm 0.15 = (6.01, 6.31)$$

$\therefore$ This is the required interval that includes the true population mean.

Example 12: ***(P–value approach to hypothesis is testing).***

The target thickness for silicon wafers used in a certain type of integrated circuit of 245 μm. A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of 246.18 μm and a sample standard deviation of 3.60 μm. Does this data suggest that true average wafer thickness is something other than the target value?

Solution: Let μ be the true average wafer thickness

Step 1. ***Null hypothesis:*** $H_0 : \mu = 245\ \mu m$

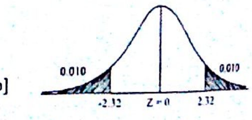
Alternative hypothesis: $H_1 : \mu \neq 245$

Step 2 . ***Test statistic*** : Under null hypothesis $H_0 : \mu = 245$ the test statistic is

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$$

Step 3. ***Level of significance:***

Take $\alpha = 0.01$ [We can take $\alpha = 5\%$ also]



Step 4. ***Determination of P–value:***

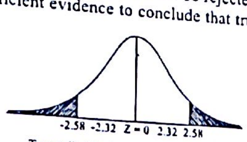
P–value = 2 $F(-2.32)$=2×0.0102
= 0.0204 [Using table A-3]

Step 5. ***Decision***: P–value > α i.e., 0.0204 > 0.01. Hence H_0 would not be rejected. At this significance level, there is insufficient evidence to conclude that true average thickness differs from the target value.

By Rejection Region method

Steps 1,2,3,4 are same as above.

Step 5. at $\alpha = 0.01$, for two tailed test,
$z_{\alpha/2} = z_{0.05} = 2.58$



Two tailed Rejection region, $|Z| < Z_{\alpha/2}$

Step 6. ***Decision***: Since $|z| < z_{\alpha/2}$
z lies in acceptance region, so H_0 is accepted.

Example 13: A quality control inspector for a microwave transmitter manufacture is accessing a shipment of 400 crystal control. The actual broadcast frequency will depend on the resonant frequency, which will vary slightly from crystal to crystal, but the mean level should achieve the related target of 0.55 *MHz*. The exact standard deviation for the individual crystal frequency is unknown. A random sample of 50 crystals from the shipment will be tested and resonant frequency determined for each. The entire shipment will be rejected if the observed mean is significantly above or below the rated level and accepted otherwise. The inspector wants just a 1% chance of rejecting a shipment in which the mean frequency exactly matches the rated level. (i) Formulate the inspector's hypothesis. **[TU, BE, 2064 Shrawan]**

(ii) Sample result provide $\bar{x} = 0.5503$ *MHz* and $s = 485$ *Hz*.
At this situation what decision should the inspector take?

Solution: Given data : $n = 50$, $s = 485$ *Hz*, [1 *MHz* = 10^6 *Hz*]

Step 1. ***Null hypothesis***: $H_0 : \mu = 0.55$ *MHz*
Alternative hypothesis: $H_1 : \mu \neq 0.55$

Step 2. ***Level of significance***: $\alpha = 1\% = 0.01$

Step 3. ***Test statistic***: Under $H_0 : \mu_1 = 0.55$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{0.5503 - 0.55}{s/\sqrt{50}} = \frac{3 \times 10^{-4}}{485 \times 10^{-6}/\sqrt{50}} = 4.3739 \quad [\because \textit{For large sample } \hat{\sigma} = s]$$

Step 4. ***Critical value***: Tabulated value of z for two tailed test at $\alpha = 0.01$ level of significance is $z_{\alpha/2} = z_{0.05} = 2.58$

Step 5. ***Decision***: Since $|z| > z_{\alpha/2}$, it is significant and H_0 is rejected. So H_1 is accepted which means that observed mean is significantly above or below the rated level. At this situation, inspector must reject entire shipment.

8.10.3 Test significance of Difference of two means (Large samples case)

Assumptions for large samples

This Z-statistic requires that:

(i) $X_1, X_2, \ldots, X_{n_1}$ is a random sample of size $n_1 \geq 30$ from population 1 which has mean $=\mu_1$ and variance $= \sigma_1^2$.

(ii) $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample of size $n_2 \geq 30$ from population 2 which has mean $=\mu_2$ and variance $= \sigma_2^2$.

(iii) Two samples $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are independent.

Suppose that the two independent random samples of sized n_1 and n_2 be drawn from two different populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2

respectively. Let $\bar{X}$ and $\bar{Y}$ be their corresponding sample means. Then, for large samples (i.e., as $n_1 \to \infty, n_2 \to \infty$) $\bar{X} \sim N(\mu, \sigma_1^2/n_1)$ and $\bar{Y} \sim N(\mu_2, \sigma_2^2/n_2)$.
Hence $(\bar{X}-\bar{Y})$, being the difference of two independent normal variates, is also a normal variate with mean.

$E(\bar{X}-\bar{Y}) = \mu_1 - \mu_2$ and variance $V(\bar{X}-\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

The standardized variate Z corresponding to the test statistic $(\bar{X}-\bar{Y})$ is

$$Z = \frac{(\bar{X}-\bar{Y}) - E(\bar{X}-\bar{Y})}{S.E.(\bar{X}-\bar{Y})} = \frac{(\bar{X}-\bar{Y}) - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Large sample test for the difference of two population means

Step 1. ***Set up hypothesis***

Null hypothesis: $H_0: \mu_1 - \mu_2 = \delta$
Alternative hypothesis: $H_1: \mu_1 - \mu_2 \neq \delta$ (*Two tailed test*)
or, $H_1: \mu_1 - \mu_2 > \delta$ (*Right tailed test*)
or, $H_1: \mu_1 - \mu_2 < \delta$ (*Left tailed test*)

Step 2. ***Level of significance*** (α): Take the most commonly used α = 5% unless otherwise stated.

Step 3. ***Test statistic***: Under the null hypothesis, $H_0: \mu_1 - \mu_2 = \delta$, Z–statistic is

(i) If σ_1^2 and σ_2^2 are known and unequal

$$Z = \frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

(ii) If σ_1^2 and σ_2^2 are known and equal [i.e.,$\sigma_1^2 = \sigma_2^2 = \sigma^2$]

$$Z = \frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(iii) If σ_1^2 and σ_2^2 are unknown and unequal

$$Z = \frac{(\bar{X}-\bar{Y}) - (\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_1^2}{n_2}}}$$

Here $\hat{\sigma}_1 = s_1$, $\hat{\sigma}_2 = s_2$ for large samples.

(iv) If σ_1^2 and σ_2^2 are unknown but equal [i.e.,$\sigma_1^2 = \sigma_2^2 = \sigma^2$]

$$Z = \frac{(\bar{X}-\bar{Y}) - \delta}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where } S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

[Here we can use $S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$ because *for large samples*

$$\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} \approx \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}$$]

Step 4. ***Critical value*:** Obtain critical or tabulated of statistic z at the pre-specified level significance according as whether the test is two tailed or one tailed.

Step 5. ***Decision*:** If $|Z| > z_{\alpha/2}$ for two tailed test

or $|Z| > z_\alpha$ for one tailed test, reject H_0.

If $|Z| \le z_{\alpha/2}$ for two tailed test

or $|Z| \le z_\alpha$ for one tailed test, accept H_0.

Large sample test for the equality of two populations means

Step 1. ***Set up Hypotheses*:**

Null hypothesis: H_0: $\mu_1 = \mu_2$ (i.e., two independent population means are equal. In other words, there is no significant difference between the sample means.)

Alternative hypothesis: H_1: $\mu_1 \neq \mu_2$ (*Two tailed test*) (i.e., two independent population means are not equal. In other words, there is significant difference between the sample means.)

or, H_1: $\mu_1 > \mu_2$ (*Right tailed test*) [i.e., the mean of the first population is greater than the mean of second population]

or, H_1: $\mu_1 < \mu_2$ (*Left tailed test*) [That is, mean of the first population is less than the mean of second population]

Step 2. ***Level of significance*** **(α):** Choose the appropriate level of significance. Take the most commonly used $\alpha = 5\%$ unless otherwise stated.

Step 3. ***Test statistic*:** Under the null hypothesis H_0: $\mu_1 = \mu_2$, the test statistic is,

$$Z = \frac{(\bar{X}-\bar{Y}) - E(\bar{X}-\bar{Y})}{S.E.(\bar{X}-\bar{Y})} = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Here σ_1^2 and σ_2^2 are known. If they are unknown, then we use their estimates provided by sample variances S_1^2 and S_2^2 respectively. That is, $\hat{\sigma}_1^2 = S_1^2$ and $\hat{\sigma}_2^2 = S_2^2$ for large samples. So,

$$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

If we want to test if two independent samples have come from same population or if two independent population have same variance i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$ then

$$Z = \frac{(\bar{X}-\bar{Y})}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is unknown, we use combined variance

$$\hat{\sigma}^2 = S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2} \quad \textit{[for large samples]}$$

Step 4. ***Critical value*** **(z_α or $z_{\alpha/2}$):** Obtain critical or tabulated of statistic Z at the pre-specified level of significance according as whether the test is two tailed or one tailed.

Step 5. ***Decision*:**

If $|Z| > z_{\alpha/2}$ (*For two tailed test*), or $|Z| > z_\alpha$ (*For one tailed test*), then Z lies in rejection region. Reject H_0 and accept H_1. That is population means are not equal. In other words, there is significant difference between the sample means.

If $|Z| > z_{\alpha/2}$ (*For two tailed test*), or $|Z| > z_\alpha$ (*For one tailed test*), then Z lies in acceptance region. Accept H_0 and reject H_1. That is the two population means are equal. In other words, there is no significance difference between the sample means.

Step 6. ***Conclusion*:** Write the conclusion in simple language.

Note: Identification of left tailed and right tailed tests:

(i) If first sample statistic < second sample statistic, then we use left tailed test.

(ii) If first sample statistic > second sample statistic, then we use right tailed test.

Example 14: A company claims that its light bulbs are superior to those of its competitor. If a study showed that a sample of $n_1 = 40$ of its bulbs has a mean lifetime 1647 hours of continuous use with a standard deviation of 27 hours, while a sample of $n_2 = 40$ bulbs made by its competitor had a mean lifetime of 1638 hours of continuous use with a standard deviation of 31 hours, does this substantiate the claim at the 0.05 level of significance? *[TU, BE, 2061 Ashadh/ 062 Jestha]*

Solution: With the usual notations, we have

$n_1 = 40, \bar{x} = 1647$ hours, $\sigma_1 = 27$ hours

$n_2 = 40, \bar{y} = 1638$ hours, $\sigma_2 = 31$ hours.

Let μ_1 and μ_2 be the lifetimes of bulbs of these companies.

Step 1. ***Set up hypotheses*:**

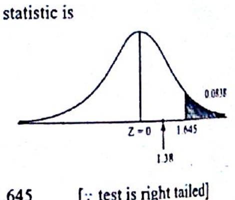
Null hypothesis: $H_0 : \mu_1 = \mu_2$ (i.e. *do not differ in quality*)

Alternate hypothesis: $H_1 : \mu_1 > \mu_2$ (i.e. *differ in quality*)[*Right tailed test*]

Step 2. ***Level of significance*:** $\alpha = 0.05$

Step 3. ***Test statistic*:** Under $H_0 : \mu_1 = \mu_2$, the test statistic is

$$z = \frac{(\bar{x}-\bar{y})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{(647-638)}{\sqrt{\frac{729}{40}+\frac{961}{40}}} = 1.38$$



Step 4. ***Critical value*:** At $\alpha = 0.05$, $z_\alpha = z_{0.05} = 1.645$ [$\because$ test is right tailed]

Step 5. ***Decision*:** Here, $|z| < z_\alpha$, so test statistic lies in acceptance Region. So, H_0 is accepted. That is, the observed difference between the two sample means is not significant, or the mean lifetime of the light bulbs do not differ significantly. Hence, this does not substantiate the claim at $\alpha = 0.05$.

Example 15: If the mean height of 60 Engineering students of Tribhuvan University is found to be 68.6 inches and the mean height of 50 Medicine students of the same University is found to be 69.51 inches, would you conclude that the Medicine students are taller than Engineering students? Assume that standard deviation of height of the students of Tribhuvan university to be 2.48 inches. *[TU, BIE, 2067 Mangsir]*

Solution: With the usual notations, we have,

Engineering students	*Medicine students*
$n_1 = 60$	$n_2 = 50$
$\bar{x} = 68.60$ inches	$\bar{y} = 69.51$ inches

σ = common population standard deviation = 2.48 inches.

Let μ_1 and μ_2 be the true average heights of Engineering and Medicine students respectively.

Step 1. *Null hypothesis:* $H_0: \mu_1 = \mu_2$. That is Medicine students are not taller than Engineering students.

Alternative hypothesis: $H_1: \mu_1 < \mu_2$ (Left tailed test).

That is, Medicine students are taller than Engineering students.

Step 2. *Test statistic:* Under null hypothesis $H_0: \mu_1 = \mu_2$, test statistic is

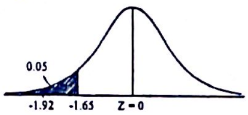
$$z = \frac{(\bar{x} - \bar{y})}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(68.6 - 69.51)}{\sqrt{\sigma^2\left(\frac{1}{60} + \frac{1}{50}\right)}} = -1.92$$

Hence, $|z| = -1.92$

Step 3. *Level of significance* (α): Take $\alpha = 5\% = 0.05$ as not mentioned.

Step 4. *Critical value:* The tabulated or critical value of z at 5% level of significance for left tailed test is

$z_\alpha = z_{0.05} = -1.645$



Step 5. *Decision* : Since $|z| > z_\alpha$, it is significant. i.e. z lies in rejection region. So, we reject H_0 and accept H_1 which means that Medicine students are taller than the engineering students.

Example 16: An examination was given to 50 students at college A and 60 students at college B. At A mean grade was 75 with standard deviation of 9. At B mean grade was 79 with standard deviation of 7. Is there significant difference between the performance of students at A and those at B, given at $\alpha = 0.05$ and 0.01? Also find confidence intervals. **[TU, BE, 2056 Bhadra]**

Solution: Given data:

Class A	$\bar{x} = 75$	$n_1 = 50$	$S_1 = 9$
Class B	$\bar{y} = 79$	$n_2 = 60$	$S_2 = 7$

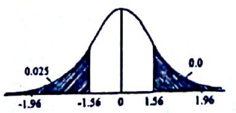
Suppose the two sample classes come from two independent populations with true mean μ_1 and μ_2 respectively.

Step 1. *Null hypothesis:* $H_0: \mu_1 = \mu_2$

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2$ (*Two tailed test*)

Step 2. *Test statistic:* Under null hypothesis $H_0 : \mu_1 = \mu_2$ is

$$z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(75 - 79)}{\sqrt{\frac{9^2}{50} + \frac{7^2}{60}}} = -1.561$$



(a) Two tailed critical region, |Z| > 1.56

Step 3. *Level of significance:*

(a) $\alpha = 5\% = 0.05$

(b) $\alpha = 1\% = 0.01$

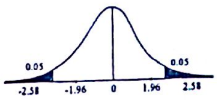
Step 4. Critical value:

(a) at $\alpha = 0.05$,

$z_{\alpha/2} = z_{0.025} = 1.96$

(b) at $\alpha = 0.01$,

$z_{\alpha/2} = z_{0.05} = 2.58$



(b) Two tailed critical region, |Z| < 2.58

Step 5. *Decision:* (a) At $\alpha = 0.05$, $|z| > z_\alpha$. So, H_0 is rejected and H_1 is accepted. This means, there is a significance difference between the performances of two classes.

(b) At $\alpha = 0.01$, $|z| < z_\alpha$. i.e., z lies in acceptance region. So, H_0 is accepted and H_1 is rejected. This means, there is no significant difference between the performances of two classes.

Again,

(a) 95% confidence interval for the difference of population means ($\mu_1 - \mu_2$) is

$$C.I. \text{ for } (\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$= (75 - 79) \pm 1.96\sqrt{\frac{9^2}{50} + \frac{7^2}{60}} = (-4 \pm 3.06) = (-7.06, -0.94)$$

(b) 99% confidence interval for the difference ($\mu_1 - \mu_2$) of population means is

$$C.I. \text{ for } (\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}}$$

$$= (75 - 79) \pm 2.58\sqrt{\frac{9^2}{50} + \frac{7^2}{60}} = (-4 \pm 4.03) = (-8.03, 0.03).$$

Example 17: (*Testing a non zero difference in means with two large samples*): To test the claim that the resistance of electric wire can be reduced by more than 0.050 ohm by alloying, 32 values obtained for alloyed wire yielded $\bar{x} = 0.136$ ohm and $s_1 = 0.004$ ohm, and 32 values obtained for alloyed wire yielded $\bar{y} = 0.083$ ohm and $s_2 = 0.005$ ohm. At the 0.05 level of significance, does this support the claim? **[TU, BIE 2066 Magh]**

Solution: Give data: $\bar{x} = 0.136$ ohm, $n_1 = 32$, $s_1 = 0.004$ ohm

$\bar{y} = 0.083$ ohm, $n_2 = 32$, $s_2 = 0.005$ ohm.

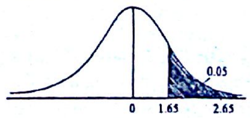
Step 1. *Null hypothesis:* $H_0: \mu_1 - \mu_2 = 0.050$ ohm $[\mu_1 - \mu_2 = \delta = 0]$

Alternative hypothesis: $H_1: \mu_1 - \mu_2 > 0.050$ (*Right tailed*)

Step 2. *Level of significance:* $\alpha = 0.05$

Step 3. *Test statistic:* Under $H_0 : \mu_1 - \mu_2 = 0.050$, test statistic is

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{S.E.(\bar{x} - \bar{y})} = \frac{0.136 - 0.083 - 0.050}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$



Scanned by CamScanner

$$= \frac{0.003}{\sqrt{\frac{(0.004)^2}{32} + \frac{(.005)^2}{32}}} = 2.65$$

Step 4. ***Critical value:*** $\alpha = 0.05$. Since test is right tailed, at $\alpha = 0.05$

$z_\alpha = z_{0.05} = 1.645 = 1.65$

Step 5. ***Decision:*** Since $|z| > z_\alpha$, z lies in rejection region, H_0 is rejected and H_1 is accepted. That is, the data substantiate the claim.

Example 18: A college conducts survey to test whether there is any difference of score in assessment in the subject 'Research Methodology and Statistical Methods' for morning–program students and day program students. A sample of 50 students in the morning yield average score in assessment as 80% with *s.d.* 14%. Similarly, a sample of 50 students in day–program yield average assessment score of 92% with *s.d.* of 25%. Based on the data what conclusion do you make? [TU 2065]

Solution: Given data :

Morning-Program : $n_1 = 50, \bar{x} = 80\%, s_1 = 14\%$

Day- Program : $n_2 = 50, \bar{y} = 92\%, s_2 = 25\%$

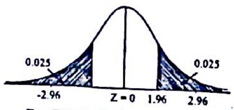
Let μ_1 and μ_2 be the true averages of scores of morning and day program students.

Step 1. ***Null hypothesis*** H_0: $\mu_1 = \mu_2$ i.e., there is no significance difference between scores in assessment.

Alternative hypothesis: H_1: $\mu_1 \neq \mu_2$ i.e. there is significant difference between scores in assessment.

Step 2. ***Test statistic:*** Under $H_0 : \mu_1 = \mu_2$

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{S.E.(\bar{x} - \bar{y})} = \frac{\bar{x} - \bar{y} - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} =$$



Two Tailed critical region, $|Z| > Z_{\alpha/2}$

$$\frac{80 - 92}{\sqrt{\frac{(14)^2}{50} + \frac{(25)^2}{50}}} = -2.96$$

Step 3. ***Level of significance:*** Take $\alpha = 5\% = 0.05$ as not mentioned

Step 4. ***Critical value:*** The tabulated or critical value of z at $\alpha = 0.05$ for two tailed test is $z_{\alpha/2} = z_{0.025} = 1.96$

Step 5. ***Decision:*** Since $|z| > z_{\alpha/2}$, z lies in rejection region. So H_0 is rejected. Hence we conclude that there is significant difference between the scores in assessment.

8.11 Testing of Hypotheses for small sample cases ($n < 30$)

In large sample cases while testing hypothesis we use Z–formula because sampling distribution of sample mean $\bar{X}$ is also normal and

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \text{ if } \sigma \text{ is known.}$$

If σ is unknown, then in large sample cases this population *s.d.* σ is replaced by sample standard deviation S because S is a good approximation of population *s.d.* σ for normal population. But *Z-formula* are not applicable if the population *s.d.* σ is not known and sample size is small (i.e., $n < 30$).

This problem was solved by *W.S.* Gosset by developing *t–test*. Gosset was a student of Mathematics and published the work on *t-test* under his pen name 'student' (as he was not allowed to publish the work with his own name). Therefore, the distribution developed by Gosset is called ***student's t–distribution*** or simply *t-distribution*.

If $X_1, X_2, \ldots, X_n$ is a random sample of size n drawn from a normal population with mean μ and variance σ^2 (*not known*) then the students *t–statistic* is defined

$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$, which follows *t–distribution* with $(n - 1)$ degrees of freedom, where

$\bar{X}$ = sample mean, n = sample size, S = unbiased estimator of population *s.d.* σ

8.11.1 Properties of *t*–distribution

1. Like *Z–distribution*, *t–distribution* is also a continuous distribution.
2. *t–distribution* curve is symmetrical about the line $t = 0$ but flatter than the standard normal distribution curve.
3. The value of t ranges from $-\infty$ to $+\infty$ (i.e., x–axis is asymptote to the *t–distribution* curve)
4. *t–distribution* is lower at the mean and higher at the tails than a normal distribution. However as the sample size increase (i.e., $n \to \infty$) the *t*–distribution tends to a normal distribution curve.
5. The *t–distribution* can be used in case of large sample but large sample theory cannot be used for small sample.
6. Total area under a *t–distribution* curve is 1 or 100%.
7. The shape of the curve of *t–distribution* varies with the degrees of freedom.

Small Sample Test (*t*–test)

If small sample is sufficient to get information about population parameter, it is not necessary to take a large sample to make decision about parameter because it saves our time, available resources and budget. Therefore, in many times we take small samples because of limited resources and nature of the experiment.

8.11.2 Assumptions for *t*–test

1. The sample size is less than 30 (i.e., $n < 30$).
2. The population standard deviation σ is not known.
3. The parent population(s) from which samples are drawn is(are) normally distributed.
4. The samples are random and independent of each other.

8.11.3 Application of *t* –test

The *t*–distribution has the following applications:

1. *t*–test for significance of single mean, population variance σ^2 being unknown.
2. *t*–test for the significance of the difference between two sample means, the population variance being equal but unknown.
3. Paired *t*–test for difference of two means.
4. *t*–test for significance of an observed sample correlation coefficient.

8.11.4 Test of Significance of a Single Mean (Small sample)

The procedures of testing hypothesis in small sample case ($n < 30$) and large sample case ($n \geq 30$) are similar except in test statistics.

The procedure of testing hypothesis of a single mean in small case is given below: Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n < 30$ drawn from a normal population with mean μ and unknown variance σ^2.

Step 1. ***Set up hypotheses***

Null hypothesis H_0: $\mu = \mu_0$. [That is, the population mean has some specified value μ_0. In other words, there is no significant difference between the sample mean $\bar{X}$ and population mean μ or, the difference between $\bar{X}$ and μ is due to fluctuations of sampling or, the sample has been drawn from normal population]

Alternative hypothesis: H_1: $\mu \neq \mu_0$ (two tailed test)

[That is population mean has not specified value μ_0. In other words, there is significant difference between sample mean and population mean or, sample has not been drawn from normal population.]

or, H_1: $\mu > \mu_0$ (*Right tailed test*) [That is population mean is greater than μ_0]

or, H_1: $\mu < \mu_0$ (*Left tailed test*) [That is population mean is less than μ_0]

Step 2. ***Level of significance*** **(α)**: Take $\alpha = 5\%$ unless otherwise stated and specify whether the alternative hypothesis is one tailed or two tailed.

Step 3. ***Test statistic***: Under null hypothesis, the test statistic is

$$t = \frac{\bar{X} - E(\bar{X})}{S.E.\ (\bar{X})} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \text{ where } S^2 = \frac{1}{n-1}\Sigma(X_i - \bar{X})^2$$

Step 4. ***Degree of freedom*** **(*d.f.*) = $n - 1$.**

Step 5. ***Critical value***: Obtain critical or tabulated value of t at pre-specified level of significance for (n – 1) degree of freedom according to one tailed or two tailed test.

Step 6. ***Decision***:

(i) If calculated value of t ≤ tabulated value of t

i.e., $|t| \le t_{\alpha/2,\, n-1}$ (*for two tailed test*), or $|t| \le t_{\alpha,\, n-1}$ (*for one tailed test*) then accept H_0. That is, the population mean has specified value μ_0. In other words, there is no significant difference between $\bar{X}$ and μ or sample is drawn from the normal population with mean μ_0.

(ii) If calculated value of t > tabulated value of t

i.e. $|t| > t_{\alpha/2,\, n-1}$ (*for two tailed test*)

or, $|t| > t_{\alpha,\, n-1}$ (*for one tailed test*), then reject H_0. That is, the population mean has not specified value μ_0. In other words, there is significant difference between $\bar{X}$ and μ.

Confidence limits in Estimating population mean μ for small samples

$(1 - \alpha)$ 100% confidential or Fiducial limits for population mean μ are

C.I. for $\mu = \bar{X} \pm t_{\alpha/2,\, n-1}\ S.E.\ (\bar{X})$

$$= \bar{X} \pm t_{\alpha/2,\, n-1} \frac{S}{\sqrt{n}} \text{ if } S \text{ is unbiased estimate of population } s.d.$$

$$\text{i.e., } S^2 = \frac{1}{n-1}\ \Sigma(X - \bar{X})^2$$

Summary decision rule for *t*-test

Alternative Hypothesis	*Reject null hypothesis* H_0 *if*
$\mu < \mu_0$	$t < -t_\alpha$ i.e., $\lvert t\rvert > t_\alpha$
$\mu > \mu_0$	$t > t_\alpha$ i.e., $\lvert t\rvert > t_\alpha$
$\mu \neq \mu_0$	$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$ i.e., $\lvert t\rvert > t_{\alpha/2}$.

Example 19: A manufacture of gun power has developed a new power which is designed to produce a muzzle velocity equal to 3000 ft/sec. Seven shells are loaded with the charge and the muzzle velocities measured. The resulting velocities are as

follows: 3005, 2935, 2995, 2965, 3905, 2395 and 2905. Do these data present sufficient evidence to indicate that the average velocity differs from 3,000 ft/sec (use $\alpha = 5\%$) [*Pokhara Uni 2001*

Solution: Specified mean of muzzle velocity (μ_0) = 3000 ft/sec. Random sample size (n) = 7.

Step 1. ***Set up Hypothesis***

Null hypothesis: H_0: $\mu = 3000$

Alternative hypothesis: H_1: $\mu \neq 3000$ (*Two tailed*)

Step 2. ***Level of significance***: $\alpha = 5\% = 0.05$

Step 3. ***Test Statistic***: Under H_0, test statistic is

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}.$$

For $\bar{X}$ and S^2: $\bar{X} = \frac{\Sigma X_i}{n} = \frac{21645}{7} = 3092.14$

$$S^2 = \frac{1}{n-1}\Sigma(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\Sigma X_i^2 - \frac{(\Sigma X)^2}{n}\right]$$

$$= \frac{1}{n-1}[\Sigma X^2 - n\bar{X}^2] = \frac{1}{6}\ [6{,}77{,}07{,}775 - 7 \times (3092.14)^2] = 129744.42$$

Therefore, $s = 360.20$

$$\therefore t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{3092.14 - 3000}{360.20\sqrt{7}} = 0.677$$

Step 4. ***degree of freedom***: $\nu = n - 1 = 7 - 1 = 6$

Step 5. ***Critical value***: Since the test is two tailed test, the tabulated or critical value of t is $t_{\alpha/2,\, n-1} = t_{0.025,6} = 2.365$

Step 6. ***Decision***: Since $|t| < t_{\alpha/2,\, n-1}$ i.e. calculated value < tabulated value, we accept H_0. Hence, we may conclude that data don't differ sufficiently from the velocity 3000 ft/sec.

95% confidence interval for true population mean μ

$(1 - \alpha)$ 100% = 95% $\Rightarrow \alpha = 5\% = 0.05$

So, 95% confidence interval for μ is given by:

$$\text{C.I. for } \mu = \bar{X} \pm t_{\alpha/2,\, n-1}\frac{s}{\sqrt{n}} = 3092.14 \pm (2.365)\left(\frac{360.20}{\sqrt{7}}\right)$$

$$= 3092.14 \pm 321.98 = (2770.16,\ 3414.12)$$

Hence, 95% confidence that include true mean μ is (2770.16, 3414.12).

Example 20: A random sample of 16 values from a normal population showed mean of 41.5 inches and the sum of square of deviations from this mean equal to 135 square inches. Show that the assumption of a mean of 43.5 inches for the population is not reasonable. Obtain 95% fiducial limits for the same.

Solution: Given data: $n = 16$, $\bar{X} = 41.5$ inches, $\Sigma(X - \bar{X})^2 = 135$, $\mu = 43.5$.

$$\therefore \quad s = \sqrt{\frac{1}{n-1}\Sigma(X - \bar{X})^2} = \sqrt{\frac{1}{15} \times 135} = 3.$$

Step 1. ***Set up hypothesis***

Null hypothesis: H_0: $\mu = 43.5$

Alternative hypothesis: H_1: $\mu \neq 43.5$ (*Two tailed*)

Step 2. ***Level of significance***: $\alpha = 5\% = 0.05$.

Step 3. *Test Statistic*: Under H_0: $\mu = 43.5$, test statistic is

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{41.5 - 43.5}{3/\sqrt{16}} = -2.67. \text{ So } |t| = 2.67.$$

Step 4. *Degree of freedom* $(d.f.) = n - 1 = 16 - 1 = 15$

Step 5. *Critical value*: Since test is two tailed, the tabulated or critical value of t for 15 degree of freedom is $t_{\alpha/2, n-1} = t_{0.025, 15} = 2.131$.

Step 6. *Decision*: Since $|t| > t_{\alpha/2, n-1}$, i.e. calculated value of t is greater than tabulated value of t, we reject H_0, and accept H_1. Hence we conclude that the mean of the population is not 43.5 inches.

For 95% confidence limits are calculated as follows

$1 - \alpha = 95\% \Rightarrow 5\%$, $d.f. = 16 - 1 = 15$, $t_{\alpha/2, n-1} = t_{0.025, 15} = 2.131$.

Thus, 95% confidence limits for the population mean are given by

$$\text{C.I. for } \mu = \bar{X} \pm t_{\alpha/2, n-1}\, S.E.(\bar{X}) = 41.5 \pm (2.131)\frac{s}{\sqrt{n}}$$

$$= 41.5 \pm (2.131)\frac{3}{\sqrt{16}} = 41.5 \pm 1.6 = (39.9, 43.1)$$

$\therefore$ Lower limit = 39.9 and upper limit = 43.1

Example 21: The specimen of copper wires drawn from a large lot has the following breaking strength (in *kg. wt*) 578, 572, 568, 570, 572, 578, 570, 572, 596 and 544. Test whether the mean breaking strength of the lot may be taken at least 578 kg wt at 5% level of significance.
[TU, BE, 2063 Kartik]

Solution: Given data: $n = 10$, $\mu = 578$ *kg.wt.*

Calculation of $\bar{X}$ and s^2:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{5720}{10} = 572$$

$$S^2 = \frac{1}{n-1}\Sigma(X - \bar{X})^2 = \frac{1}{9} \times 1456 = 161.78 \quad \therefore s = 12.72$$

Step 1. *Null hypothesis*: H_0: $\mu \geq 578$ *kgs*.

[To conduct the hypothesis we may state H_0: $\mu = 578$ *kgs* also:]

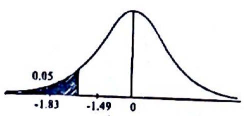
***Alternative hypothesis*:** H_1: $\mu < 578$ *kgs*. (*Left tailed test*)

Step 2. *Level of significance*: $\alpha = 5\% = 0.05$

Step 3. *Test statistic*: Under H_0, the test satistic is

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{572 - 578}{12.72/\sqrt{10}} = -1.49.$$

So, $|t| = 1.49$



Step 4. *Degree of freedom*: $(d.f.) = n - 1 = 10 - 1 = 9$

Step 5. *Critical value*: Since test is one tailed, the tabulated or critical value of t for 9 degrees of freedom is $t_{\alpha, n-1} = t_{0.05, 9} = 1.833$ *[from t-table A-4]*

Step 6. *Decision*: Since $|t| < t_{\alpha, n-1}$ i.e. calculated value of t is less than tabulated value of t, we accept H_0. Hence the population mean of the breaking strength is at least 578 *kgs*.

Example 22: The specifications for a certain kind of ribbon call for a mean breaking strength of 180 pounds. If five pieces of the ribbon (randomly selected from different

rolls) have a mean of 169.5 pounds with a standard deviation of 5.7 pounds, test the null hypothesis $\mu = 180$ pounds against the alternative hypothesis $\mu < 180$ pounds at the 0.01 level of significance. Assume that the population distribution is normal.
[TU, BE, 2061 Ashwin/ 2062 Baisakh/ 2065 Chaitra]

Solution: Given data : $\mu = 180$ pounds, $\bar{x} = 169.5$ pounds, $n = 5 < 30$ $s = 5.7$

Step 1. *Null hypothesis*: H_0: $\mu = 180$ pounds

***Alternative hypothesis*:** H_1: $\mu < 180$ pounds [*Left tailed test*]

Step 2. *Level of significance* : $\alpha = 0.01$

Step 3. *Test statistic* : Under H_0 test statistic is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{169.5 - 180}{5.7/\sqrt{5}} = -4.12. \text{ So, } |t| = 4.12$$

Step 4. *Degree of freedom*: $(d.f.) = n - 1 = 5 - 1 = 4$

Step 5. *Critical (significant) value*: Tabulated or critical value of t for 4 degree of freedom is $t_{\alpha, n-1} = t_{0.01, 4} = 3.747$

Step 6. *Decision*: Since $|t| > t_{\alpha, n-1}$, the null hypothesis H_0 must be rejected at $\alpha = 0.01$. So we accept H_1, i.e. mean breaking strength is less than 180 pounds.

Example 23: The Edison Electric Institute has published figures on the annual value of *kilowatt* hours consumed by various home appliances. It is claimed that vacuum cleaner consumed an average of 46 *kilowatt* hours per year. If a random sample of 12 homes included in a planned study indicates that vacuum cleaner consumes an average of 42 *kilowatt* hours per year with a standard deviation of 11.9 *kilowatt*, does this suggest at the 0.05 level of significance that vacuum cleaners consumes, on the average, less than 46 *kilowatt* hours annually? Assume that population of kilowatt hours to be normal.
[TU, BE, 2067 Mangsir]

Solution: Hint: H_0: $\mu = 46$ versus H_1: $\mu < 46$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{42 - 46}{11.9/\sqrt{12}} = -1.164$$

$t_{\alpha, n-1} = t_{0.05, 11} = 1.796$

$|t| < t_{\alpha, n-1}$. Accept H_0 and reject H_1.

Example 24: The following data were obtained for the amount of time (second) taken by a proposed computer system to compile a sample of short FORTRAN program:

2.3	6.7	3.8	5.0	4.9	6.1	4.4	5.2	3.9	4.8
4.6	5.7	5.3	4.7	4.2	4.7	5.7	4.8		

At 5% level of significance, test the null hypothesis that the mean compilation time for all short of FORTRAN program run on the system is at least 4 seconds. Should it be accepted or rejected.
[TU, BE, 2064 Shrawan]

Solution: Set H_0: $\mu \geq 4$ seconds versus H_1: $\mu < 4$ seconds

Calculate $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ and solve as before.

Example 25: The following samples failure data (thousands of miles) were obtained for a type of catalytic converter:

62.3, 44.4, 49.2, 49.2, 63.3, 47.6, 60.1, 37.4, 55.8, 57.5, 58.3, 56.2, 54.3.

At 5% level of significance, must you accept or reject the null hypothesis that the catalytic converter will last, on average, 50 thousand miles?
[TU, BE, 2064 Shrawan]

Solution: Set H_0: $\mu \leq 50$ thousand miles versus H_1: $\mu > 50$ thousand miles and solve as before.

Example 26: The life in hours of a battery is known to be approximately normally distributed. A random sample of 10 batteries has a mean life of 40.5 hours and standard deviation of 1.25 hours. Is there evidence to support the claim that battery life exceeds 40 hours? Use $\alpha = 0.05$. [TU, BE, 2066 Magh]

Solution: Data given: $n = 18$, $\bar{x} = 40.5$, $s = 1.25$

Set hypothesis $H_0: \mu \geq 40$ versus $H_1: \mu < 40$.

Calculate $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ and proceed as above.

8.11.5 The Two Sample t–test and Confidence Interval (Inference concerning two means: Independent samples)

Assumptions for testing the difference of means:

This *t-statistic requires* that:

1. $X_1, X_2, \ldots, X_{n_1}$ is a random samples of sizes n_1 from population 1 which has mean $= \mu_1$ and variance $= \sigma_1^2$.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random samples of sizes n_2 from population 2 which has mean $= \mu_2$ and variance $= \sigma_2^2$.
3. The populations from which the samples are drawn are normally distributed.
4. Population variances unknown.
5. Two samples $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are random samples and independent.

8.11.6 Small sample test of significance for the difference between two means

Small sample test of significance for the difference between means is as follows:

Step 1. ***Set up hypothesis***

Null hypothesis: $H_0: \mu_1 - \mu_2 = \delta$

Alternative hypothesis: $H_1: \mu_1 - \mu_2 \neq \delta$ (*Two tailed test*)

or, $H_1: \mu_1 - \mu_2 > \delta$ (*Right tailed test*)

or, $H_1: \mu_1 - \mu_2 < \delta$ (*Left tailed test*)

Step 2. ***Level of significance*** (α): Take most commonly used $\alpha = 5\%$ unless otherwise stated.

Step 3. ***Test Statistic***: Under the null hypothesis $H_0: \mu_1 - \mu_2 = \delta$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ the test statistic is

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \delta}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2}$$

where *pooled* (i.e., *combined*) variance S_p^2 is unbiased estimate of the common population variance σ^2.

Calculation of S_p^2:

(i) If the unbiased estimates of population variances (i.e., S_1^2 and S_2^2) are known, then $S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

(ii) If S_1^2 and S_2^2 are not given, we calculate S_p^2 as follows:

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma X^2 - \frac{(\Sigma X)^2}{n_1} + \Sigma Y^2 - \frac{(\Sigma Y)^2}{n_2}\right]$$

Step 4. ***Degree of freedom*** $(d.f.) = \nu = n_1 + n_2 - 2$

Step 5. ***Critical value***: Obtain the tabulated or critical value of t at α level of significance for $(n_1 + n_2 - 2)$ degree of freedom in one/two tailed te

Step 6. ***Decision***: If $|t| > t_{\alpha, n_1 + n_2 - 2}$ for one tailed test

or, $|t| > t_{\alpha/2, n_1 + n_2 - 2}$ for two tailed test, then reject H_0 and accept H_1.

If $|t| \leq t_{\alpha, n_1 + n_2 - 2}$ for one tailed test

or, $|t| \leq t_{\alpha/2, n_1 + n_2 - 2}$ for two tailed test, then accept H_0 and reject H_1.

Summary of decision rule for t- test

Alternative hypothesis	*Reject null hypothesis if*
$\mu_1 - \mu_2 < \delta$	$t < -t_\alpha$ i.e., $\lvert t\rvert > t_\alpha$
$\mu_1 - \mu_2 > \delta$	$t > t_\alpha$ i.e., $\lvert t\rvert > t_\alpha$
$\mu_1 - \mu_2 \neq \delta$	$t < -t_{\alpha/2}$, i.e., $\lvert t\rvert > t_{\alpha/2}$ or, $t > t_{\alpha/2}$, i.e., $\lvert t\rvert > t_{\alpha/2}$

Case I Unknown and Equal Variances

Small sample test of significance for the equality of two means $\mu_1 = \mu_2$

Assumptions

This *t-statistic requires* that:

1. $X_1, X_2, \ldots, X_{n_1}$ is a random sample of size $n_1 < 30$ from population 1 which has mean $= \mu_1$ and variance $= \sigma_1^2$.
2. $Y_1, Y_2, \ldots, Y_{n_2}$ is a random sample of size $n_2 < 30$ from population 2 which has mean $= \mu_2$ and variance $= \sigma_2^2$.
3. The populations from which the samples are drawn are normally distributed.
4. Population variances are *equal* but *unknown*. It follows that the standard deviations of two populations are also equal i.e., $\sigma_1 = \sigma_2 = \sigma$.
5. Two samples $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are random samples and independent.

Small sample test of significance for the equality of two means

Step 1. ***Set up hypotheses***

Null hypothesis: $H_0: \mu_1 = \mu_2$. That is sample have been drawn from the normal populations with the same means. In other words, there is no significant difference between two sample means $\bar{X}$ and $\bar{Y}$.

Alternative hypothesis:

That is samples have not been drawn from the normal populations with the same means. In other words, there is significant difference between two sample means $\bar{X}$ and $\bar{Y}$.

or, $H_1: \mu_1 > \mu_2$ (*Right tailed test*). That is first population mean is significantly greater than the second population mean.

or, $H_1: \mu_1 < \mu_2$ (*Left tailed test*). That is, first population mean is significantly less than the second population mean.

Step 2. ***Level of significance*** (α): Take $\alpha = 5\%$ unless otherwise stated.

Step 3. ***Test statistic***: Under null hypothesis $H_0: \mu_1 = \mu_2$, t–statistic is

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S.E.(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Step 4,5,6 are same as in previous test of significance.

The two sample t confidence interval for $\mu_1 - \mu_2$ with confidence level $(1-\alpha)$ 100% is

$$\left[(\bar{X}-\bar{Y}) \pm t_{\alpha/2,\, n_1+n_2-2}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\,\right]$$

Note: For $d.f. > 30$, the significant values of t are same as those of Z for the normal test. So, $t_\alpha = z_\alpha$ and $t_{\alpha/2} = z_{\alpha/2}$ for $d.f. > 30$.

Example 27: The following random samples are measurements of the heat–producing capacity (in millions of calories per ton) of specimens of coal from two mines:

Mine 1	8260	8130	8350	8070	8340	
Mine 2	7950	7890	7900	8140	7920	7840

Use the 0.01 level of significance to test whether the difference between means of the these two samples is significant? *TU, BE, 2062 Baishakh/TU, BE, 2068 Magh (Back)]*

Solution: Let μ_1 and μ_2 be the population means of mine 1 and mine 2 respectively.

Step 1. ***Null hypothesis:*** $H_0: \mu_1 = \mu_2$

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2$ (two tailed test)

Step 2. ***Level of significance:*** $\alpha = 0.01$

Step 3. ***Test statistic:*** Under $H_0: \mu_1 = \mu_2$, test statistic is

$$t = \frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Calculation of sample means and s_p^2

Have $n_1 = 5$, $n_2 = 6$

Now, $\bar{x} = A + \frac{\Sigma d_1}{n_1} = 8260 + \frac{-150}{5} = 8260 + (-30) = 8230$

$\bar{y} = B + \frac{\Sigma d_2}{n_2} = 7900 + \frac{240}{6} = 7940$

$$s_p^2 = \frac{1}{n_1+n_2-2}\left[\Sigma d_1^2 - \frac{(\Sigma d_1^2)}{n_1} + \Sigma d_2^2 - \frac{(\Sigma d_2^2)}{n_2}\right]$$

$$= \frac{1}{9}\left[67500 - \frac{(-150)}{5} + 64200 - \frac{(240)}{6}\right] = 13066.67$$

$\therefore\ s_p = 114.31$

So, test statistic is: $t = \frac{\bar{x}-\bar{y}}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{8230-7940}{114.31\sqrt{\left(\frac{1}{5}+\frac{1}{6}\right)}} = 4.19$

Step 4. ***Degree of freedom*** $(d.f.) = n_1+n_2-2 = 9$

Step 5. ***Critical value*** **:** Tabulated value of the test statistic at $\alpha = 0.01$ for 9 degree of freedom in two tailed test is $t_{\alpha/2, n_1+n_2-2} = t_{0.05,9} = 1.833$

Step 6. ***Decision:***

Since $|t| > 1.833$, H_0 is rejected and H_1 is accepted. Hence, the difference between means of these two samples is significant.

99% confidence interval

$(1-\alpha)$ 100% = 99% $\Rightarrow \alpha = 1\% = 0.01$

$t_{\alpha/2,\, n_1+n_2-2} = t_{0.05,\,9} = 1.833$

Hence 99% C.I. for $(\mu_1-\mu_2) = \left[(\bar{x}-\bar{y}) \pm t_{\alpha/2,\, n_1+n_2-2}\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\,\right]$

$= 290 \pm 1.833 \times 69.22 = (290 \pm 126.88) = (163.12, 416.88)$

Example 28: The average number of articles produced by two machines per day are 200 and 250 with standard deviation 20 and 25 respectively on the basis of records of 25 days production. Can you regard both the machines equally efficient at 1% level of significance? *[TU, BE, 2057 Bhadra/2065 Chaitra]*

Solution: Data given: $n_1 = n_2 = 25$; $\bar{x} = 200$; $\bar{y} = 250$; $s_1 = 20$; $s_2 = 25$

$$s_p^2 = \frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1+n_2-2} = 512.5$$

Step 1. ***Null hypothesis:*** $H_0: \mu_1 = \mu_2$ (i.e., both machines are equally efficient)

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2$ (i.e. both are not equally efficient)

Step 2. ***Level of significance:*** $\alpha = 1\% = 0.01$

Step 3. ***Test statistic:*** Under $H_0: \mu_1 = \mu_2$, test statistic is

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{200-250}{533.85\sqrt{\left(\frac{1}{25}+\frac{1}{25}\right)}} = -7.65$$

Step 4. ***Degree of freedom*** $(d.f.)$: $n_1 + n_2 - 2 = 48$

Step 5. ***Critical value:*** For $d.f. > 30$ the significant values of t are same as those of z for the normal test. So, $t_{\alpha/2} = z_{\alpha/2} = z_{0.005} = 2.58$.

Step 6. Decision: Since $|t| > t_{\alpha/2}$ (i.e., $t < t_{\alpha/2}$), it is highly significant. Hence, H_0 is rejected and we conclude that both machines are not equally efficient at 1% level of significance.

Example 29: A college conducts both day and night classes intended to be identical. A sample of 15 students yield examination results as $\bar{x} = 72.4$ and $s_1 = 14.8$. A sample of 10 students yield examination results as $\bar{y} = 73.9$ and $s_2 = 15.9$. Is there any effect on the result by the shift? Test it at 5% level of significance. *[TU, BE, 2063 Kartik]*

Solution: Set $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ and solve as before.

Example 30: The nicotine contain in mg of two samples of tobacco were found to be as follows: *[TU, BE, 2063 Kartik]*

Sample *A*	24	27	26	21	25	
Sample *B*	27	30	28	31	22	36

Can it be said that two sample come from normal population having some mean?

Solution: Set $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ and solve as before.

Example 31: A company claims that their light bulbs are superior to those of its main competitor. If a study shows that a sample of $n_1 = 10$ of its bulbs had a mean lifetime of 647 hours of continuous use with a *s.d.* of 27 hours, while a sample of $n_2 = 15$ bulbs made by its main competitor had mean lifetime of 638 hours of continuous use with *s.d.* of 31 hours, does this substantiate the claim at the 0.05 level of significance? *[TU, BE, 2062 Jestha]*

Solution: Set $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 > \mu_2$ and solve as before.

Example 32: In the process of making a decision of either continue operating or close a civil health centre, a random sample of 25 people who had visited the centre, at least once was chosen and each person asked whether he or she felt the centre should be closed. In addition, the distance between each person's place of residence and health center was computed and recorded. Of the 25 people responding, 16 were in favour of continued operation. For these 16 people, the average distance from the centre was 5.2

miles with a standard deviation of 2.5 miles. The remaining 9 people who were in favour of closing the center lived at an average of 8.7 miles from the center with a *s.d.* of 5.3 miles. Do these data indicate that there is a significant difference in mean distance to the health centre for these two groups? **[TU, BE, 2064 Poush]**

Solution: Data given : $n_1 = 16, \bar{x} = 5.2$ miles, $s_1 = 2.5$ miles

$n_2 = 9, \bar{y} = 8.7$ miles, $s_2 = 5.3$ miles.

Set $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ and solve as before.

Example 33: Two types of drugs were used on 5 and 7 patients for reducing their weight. Drug *A* was imported and drug *B* was indigenous. The decrease in the weight after using the drugs for six months was as follows:

Drug A	10	12	13	11	14		
Drug B	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of two drugs? If not, which drug should you buy? **[TU, 2055]**

Solution:

Step 1. ***Null hypothesis***: $H_0: \mu_1 = \mu_2$ .That is, there is no significant difference in the efficacy of two drugs *A* and *B*.

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2$ (*Two tailed test*).That is, there is significant difference in the efficacy of two drugs *A* and *B*.

Step 2. ***Level of significance***: Since the level of significance is not given, we take $\alpha = 5\% = 0.05$.

Step 3. ***Test statistic***: Under $H_0: \mu_1 = \mu_2$, the test statistic is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Calculation of sample means and s_p^2

Here, $n_1 = 5, n_2 = 7$. So, $\bar{x} = \frac{\Sigma x}{n_1} = \frac{60}{5} = 12, \bar{y} = \frac{\Sigma y}{n_2} = \frac{77}{7} = 11$

$$s_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2\right] = \frac{1}{5 + 7 - 2}[10 + 44] = 5.4$$

$$\text{So, } t = \frac{12 - 11}{\sqrt{5.4\left(\frac{1}{5} + \frac{1}{7}\right)}} = 0.735$$

Step 4. ***Degree of freedom*** $(d.f.) = n_1 + n_2 - 2 = 10$

Step 5. ***Critical value***: The tabulated of the test statistic t at 5% level of significance for 10 degree of freedom in two tailed test is $t_{\alpha/2,\, n_1+n_2-2} = t_{0.025,10} = 2.228$

Step 6. ***Decision***: Since the calculated value of $t = 0.735$ is less than the tabulated value of $t_{0.025,\,10} = 2.228$. That is, $|t| < t_{\alpha/2}$, H_0 is accepted. Hence, we conclude that there is no significant difference in the efficacy of two drugs *A* and *B*. Therefore, any drug *A* or *B* can be bought.

Example 34: A group of six months of old chickens reared on a high protein diet weigh 1.2, 1.5, 1.1, 1.6, 1.4, 1.4 and 1.6 *kgs*. A second group of five chickens similarly treated except that they receive a low protein diet weigh 0.8, 1.0, 1.4, 1.0 and 1.3 kg. Test whether there is significant evidence that high protein diet has increased the weight of chickens. **[TU, MBA, 2045]**

Solution:

Step 1. ***Null hypothesis***: $H_0: \mu_1 = \mu_2$; i.e., the high protein diet has not increased the weight of the chickens.

Alternative hypothesis: $H_1: \mu_1 > \mu_2$ (*Right tailed test*). That is, high protein has increased the weight of the chicken.

Step 2. ***Level of significance***: Since the level of significance is not given, we take $\alpha = 0.05$.

Step 3. ***Test statistic***: Under $H_0: \mu_1 = \mu_2$, test statistic is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Calculation of sample means and s_p^2

Here $\bar{x} = \frac{\Sigma x}{n_1} = \frac{9.8}{7} = 1.4, \bar{y} = \frac{\Sigma y}{n_2} = \frac{5.5}{5} = 1.1$

$$s_p^2 = \frac{1}{n_1 + n_2 - 2}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n_1} + \Sigma y^2 - \frac{(\Sigma y)^2}{n_2}\right]$$

$$= \frac{1}{10}[13.94 - 13.82 + 6.29 - 6.05] = 0.046$$

$$\therefore \quad t = \frac{1.4 - 1.1}{\sqrt{0.046\left(\frac{1}{7} + \frac{1}{5}\right)}} = 2.38$$

Step 4. ***Degree of freedom*** $(d.f.) = n_1 + n_2 - 2 = 10$

Step 5. ***Critical value***: Tabulated or critical value of t for 10 *d.f.* at 5% level of significance for right tailed test is $t_{\alpha,\, n_1+n_2-2} = t_{0.05,\,10} = 1.812$

Step 6. ***Decision***: Since $|t| > t_\alpha$ it is significant and H_0 is rejected and hence H_1 is accepted which means there is significant evidence that high protein diet has increased the weight of the chicken.

Example 35: The means of two random samples of size 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviations from mean are 26.94 and 18.73 respectively. Can the samples be considered to have drawn from the same normal population? (Use $\alpha = 5\%$) **[TU, BE, 2065 Chaitra]**

Solution: Given data: $n_1 = 9, \bar{x} = 196.42, \Sigma(x - \bar{x})^2 = 26.94$

$n_2 = 7, \bar{y} = 198.82, \Sigma(y - \bar{y})^2 = 18.73$

Set $H_0: \mu_1 = \mu_2$ versus $\mu_1 \neq \mu_2$ and solve as above.

Example 36: In investigating which of two presentations of subject matter to use in a computer programmed course, an experimenter randomly choose two groups of 18 students each, and assigned one group to receive presentation I and the second to receive presentation II. A short quiz on the presentation was given to each group and their grades recorded. Do the following data indicate that a difference in the mean quiz scores (hence a difference in effectiveness of presentation) exists for the two methods. **[TU, BE, 2066 Magh]**

	Mean	*Variance*
Presentation I	81.7	23.2
Presentation II	77.2	19.8

Solution: Given data: $n_1 = 18, \bar{x} = 81.7, s_1^2 = 23.2$

$n_2 = 18, \bar{y} = 77.2, s_2^2 = 19.8$, compute s_p^2

Using $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

Set hypothesis $H_0: \mu_1 = \mu_2$ versus $H_1: \mu \neq \mu$ and solve as above.

Example 37: To find out whether a new serum will arrest leukemia, 9 mice, all with an advanced stage of the disease, are selected. Five mice receive the treatment and 4 do not. Survival times, in years, from the time experiment commenced are as follows:

Treatment	2.1	5.3	1.4	4.6	0.9
No Treatment	1.9	0.5	2.8	3.1	

At the 0.05 level of significance can the serum be said to be effective. Assume the two distributions to be normally distributed with equal variances.

[TU, BE, (II/II) 2067 Mangsir]

Solution: Given data: $n_1 = 5, n_2 = 4$. Set up hypothesis

$H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ and proceed as above.

Please Remember: If the degrees of freedom is 30 or more than 30 (i.e., $d.f. \geq 30$) then Z-values and t-values are same. See the following example:

Example 38: According to Chemical Engineering an important property of fiber is its water absorbency. The average percent absorbency of 25 randomly selected pieces of cotton fiber was found to be 20 with a standard deviation of 1.25. A random sample of 25 pieces of acetate yield on average percent of 2 with a standard deviation of 1.25. Is there strong evidence that the population mean present absorbency for cotton fiber is significantly higher than the mean for acetate? Assume that the percent absorbency is approximately normally distributed and that the population variances in present absorbency for the two fibers are the same. Use a significance level of 0.05. *[TU, BE (II/II) 2067 Mangsir]*

Solution: Given data: $n_1 = 25, \bar{x} = 20, s_1 = 1.25, n_2 = 25, \bar{y} = 12, s_2 = 1.25$.

So, $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \dfrac{24 \times (1.25)^2 + 24 \times (1.25)^2}{25+25-2} = 1.563$

Step 1. ***Null hypothesis:*** $H_0: \mu_1 = \mu_2$. That is population mean percent absorbency for cotton and acetate are equal.

Alternative hypothesis: $H_1: \mu_1 > \mu_2$ (*Right tailed test*). That is, mean percent absorbency for cotton fiber is significantly higher than for acetate.

Step 2. ***Level of significance:*** $\alpha = 0.5$

Step 3. ***Test statistic:*** Under $H_0: \mu_1 = \mu_2$,

$$t = \frac{\bar{x}-\bar{y}}{s_p\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{20-12}{1.63\sqrt{\frac{1}{25}+\frac{1}{25}}} = 17.35$$

Step 4. ***Critical value:*** Tabulated or critical value of t for $d.f. = 48$ at $\alpha = 0.5$ for right tailed test is $t_{\alpha,\, n_1+n_2-2} = t_{0.05,\,48} \approx z_\alpha = z_{0.05} = 2.58$

Step 5. ***Decision:*** Since $|t| > t_\alpha$, it is significant. Therefore, we reject H_0 and accept H_1. Hence, there is strong evidence that the population mean percent absorbency for cotton fiber is significantly higher than the mean for acetate.

Example 39: An experiment as performed to compare the abrasive wear of two different laminated material. Twelve pieces of material 1 were tested by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 unites with a sample standard deviation of 4 while the samples of material 2 gave an average of 81 and sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the abrasive wear of arterial 1

exceeds that of material 2 by more than 2 units? Assume the populations to be approximately normal with equal variances.

Solution: Let μ_1 and μ_2 represent the population means of the abrasive wear for material 1 and material 2, respectively.

Step 1. ***Set up hypotheses***

Null hypothesis: $H_0: \mu_1 - \mu_2 = 2$.

Alternative hypothesis: $H_1: \mu_1 - \mu_2 > 2$.

Step 3. ***Level of significance:*** $\alpha = 0.05$

Step 4. ***Critical region:*** $t > 1.725$, where $t = \dfrac{(\bar{x}-\bar{y}) - d_0}{S_p\sqrt{1/n_1 + 1/n_2}}$ with $\nu = 20$ degrees of freedom.

Step 5. ***Computations:*** $\bar{x} = 85, s_1 = 4, n_1 = 12, \quad \bar{y} = 81, s_2 = 5, n_2 = 10$.

Hence $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

$$S_p = \sqrt{\frac{(11)(16)+(9)(25)}{12+10-2}} = 4.478,$$

$$t = \frac{(85-81)-2}{4.478\sqrt{1/12+1/10}} = 1.04$$

$p = P(T > 1.04) \approx 0.16$.[From t-table A-4]

Step 6. ***Decision:*** Do not reject H_0. We are unable to conclude that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units.

Case II Unknown but Unequal Variances

There the situations where the analyst is not able to assume that $\sigma_1 = \sigma_2$. When we deal with independent random samples from normal populations whose variances seem to be unequal, we should not pool.

Statistic for small samples inference, $\sigma_1 \neq \sigma_2$, normal populations

For normal populations, when the sample sizes n_1 and n_2 are not large and $\sigma_1 \neq \sigma_2$,

$$t' = \frac{(\bar{X}-\bar{Y}) - \delta}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

is approximately distributed as a t with estimated degrees freedom.

The estimated degrees of freedom for t' are calculated from the observed values of the sample variances s_1^2 and s_2^2

$$\text{Estimated degrees of freedom} = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$$

The estimated degrees of freedom are often rounded down to an integer so a t table can be consulted.

The test based on t' is called the **Smith-Satterthwaite test.**

Example40: (***Testing equality of mean product volume***)

One process of making green gasoline takes sucrose, which can be derived from biomass, and converts it into gasoline using catalytic reactions. This is not a process for making a gasoline additive but fuel itself, so research is still at the pilot plant stage. At one step in a pilot plant process, the product consists of carbon chains of length 3. Nine runs were made with each of two catalysts and the product volumes (gal) are

Catalyst 1	0.63	2.64	1.85	1.68	1.09	1.67	0.73	1.04	0.68
Catalyst 2	3.71	4.09	4.11	3.75	3.49	3.27	3.72	3.49	4.26

Step 3. *Test Statistic*: Under H_0: $\mu_X = \mu_Y$, the test statistic is $t = \frac{\bar{D}}{s_D/\sqrt{n}} \sim t_{n-1}$

Calculation of $\bar{D}$ and s_D

Set A (X)	Set B (X)	$D = X - Y$	D^2
10	12	−2	4
8	8	0	0
7	8	−1	1
9	10	−1	1
8	8	0	0
10	11	−1	1
9	9	0	0
6	8	−2	4
7	9	−2	4
8	9	−1	1
		$\Sigma D = -10$	$\Sigma D^2 = 16$

Here, $n = 10$, $\bar{D} = \frac{\Sigma D}{n} = \frac{-10}{10} = -1$

$$S_D^2 = \frac{1}{n-1}\left[\Sigma D^2 - \frac{(\Sigma D)^2}{n}\right] = \frac{1}{9}\left[16 - \frac{(-10)^2}{10}\right] = 0.667. \text{ So, } s_D = 0.82$$

Hence $t = \frac{\bar{D}}{s_D/\sqrt{n}} = \frac{-1}{0.82/\sqrt{10}} = -3.86$

Step 4. *Degree of freedom* (*d.f.*): $n - 1 = 10 - 1 = 9$

Step 5. *Critical value*: The tabulated or critical value of t at $\alpha = 0.05$ for 9 degree of freedom in one tailed test is $t_{0.05,\,9} = 1.833$

Step 6. *Decision*: Since $|t| > t_{\alpha,\,n-1}$ (i.e. $t < t_{\alpha,\,n-1}$), we reject H_0 and accept H_1. Hence we conclude that training has benefited the set B students.

Example 44: A certain stimulus administered to each 12 patients resulted in the following increase (change) of blood pressure:

5, 2, 8, −1, 3, 0, −2, 1, 5, 0, 4, 6

Can it be concluded that the stimulus will be in general, accompanied by increase in blood pressure? **[TU, BA 2059/ BSc 2057, 2060]**

Solution: Here, we are given the increments $D = Y - X$ in the blood pressure readings of 12 patients.

Step 1. *Null hypothesis* H_0: $\mu_X = \mu_Y$. That is there is no significant difference in the blood pressure readings of the patients before and after the drug. In other words, the given increments are not due to the stimulus.

Alternative Hypothesis: H_1: $\mu_Y > \mu_X$ (*right tailed test*) i.e., the stimulus results in an increase in blood pressure.

Step 2. **Level of significance:** Since the level of significance is not given we take, $\alpha = 5\% = 0.05$.

Step 3. *Test statistic*: Under H_0: $\mu_X = \mu_Y$, test statistic is

$$t = \frac{\bar{D}}{s_D/\sqrt{n}} \text{ where, } \bar{D} = \frac{\Sigma D}{n},\ s_D^2 = \frac{1}{n-1}\left[\Sigma D^2 - \frac{(\Sigma D)^2}{n}\right]$$

Calculation of $\bar{D}$ and s_D^2

D	5	2	8	-1	3	0	-2	1	5	0	4	6	$\Sigma D = 31$
D^2	25	4	64	1	9	0	4	1	25	0	16	36	$\Sigma D^2 = 185$

$$\therefore \quad \bar{D} = \frac{\Sigma D}{n} = \frac{31}{12} = 2.58,\ s_D^2 = \frac{1}{n-1}\left[\Sigma D^2 - \frac{(\Sigma D)^2}{n}\right] = 9.5382$$

$$t = \frac{\bar{D}}{\sqrt{\frac{s_D^2}{n}}} = \frac{2.58}{\sqrt{\frac{9.5382}{12}}} = 2.894$$

Step 4. *Degree of freedom* ($d.f.$) $= 12 - 1 = 11$

Step 5. *Critical value*: Critical or tabulated value of t for (11 $d.f.$ at 5% level of significance for right tailed test is $t_{\alpha,\,n-1} = t_{0.05,11} = 1.796$.

Step 6. *Decision*: Since the calculated value of t is greater than the tabulated value of t, *i.e.*, $|t| > t_{\alpha,\,n-1}$, it is significant and H_0 is rejected and hence H_1 is accepted which means that the stimulus is effective in increasing blood pressure.

8.13 Analysis of Variance (ANOVA)

8.13.1 Introduction

As long as we face with the problem of testing the equality of two means of the respective populations, the tests like "normal tests for large samples" and "t - tests for small samples" are used. However, experimenters often prefer to test the equality of more than two means. For example, in order to know the influence of a medicine, it may be tried with different levels of dosage and each level may have its mean effect. One product may be produced by more than two similar machines and each machine may produce the product that has a specific mean value for a dimension (mean width, mean length, etc.) that may be different from the mean of the dimension of the product produced by other machines.

Under these circumstances, there exists a traditional statistical test method called Analysis Of Variance (frequently abbreviated as *ANOVA*) that can be used for comparing means when there are more than two levels of a single factor (or treatment). For example, the effect of different levels of dosage of same medicine (treatment or factor) may be compared. Therefore, *ANOVA* is a statistical method for determining the existence of differences among several population means. In fact, treatments relative to the variance within treatments and hence the name analysis of variance.

To understand the concept of *ANOVA*, let us consider the following example. In a film manufacturing company, photo sensitized solution is coated on a base film. The base film thickness is observed to be a critical factor, as it affects the quality of the film. The company has five machines that produce film base of specified thickness, *i.e.*,(180±7) microns. Since the machines are expected to produce film base of same thickness, they are controlled based on the outcomes of the sample test results on thickness. The means and variances based on ten samples per machine are calculated and are shown in Table. If we assume that the thickness values are from normal populations, then the overall population and the distributions of the samples of five machines can be obtained as shown in the following figure. In figure, the overall population mean is considered as 180 with the standard deviation 2.605 (*pooled from samples*)

Table: Mean and variance of base thickness (*In microns*)

Machine	*Mean thickness*	*Standard deviation*
1	184.9	1.913
2	181.1	2.557
3	177.1	3.108
4	182.4	2.876
5	175.7	2.406

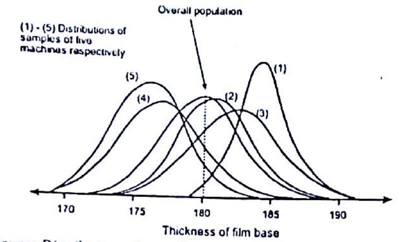


Figure: *Distributions of overall populations and sample observations.*

It may be observed from above figure that while the overall population is the desired result, the distributions, of the samples of thickness values collected from different machines show that their means are different (few are on the left side and few are on the right side) from the overall population mean. Therefore, the experimenter prefers to test whether this difference between the samples are statistically significant or not. If the differences are not significant, then the mean values of thickness can be taken as equal.

Techniques of Analysis of variance (*ANOVA*)

The techniques of analysis of variance for a single variable and two variables is very similar and it can be studied under the following classification

ANOVA

↓

One -way classification | **Two-way classification**

Assumptions of *ANOVA*

1. The populations form which samples are drawn are normally distributed with same mean and variance.
2. The samples drawn from different populations are random and independent.
3. Various treatment and environmental effects are additive in nature

8.13.2 One-way ANOVA (or single -Factor Experiment) and linear statistical model

One-way *ANOVA* involves the statistical model either of data sampled from more than two numerical populations (distributions) or of data from experiments in which more than two treatments have been used. The characteristic that differentiates the treatments or populations from one another is called the ***factor under study*** and the different treatments or populations are called ***levels of the factor.*** Under the one way *ANOVA*, the influence of only one attribute or factor is considered. For example:

1. An experiment to study the effects of five different brands of gasoline on automobile engine operating efficiency (mpg).

2. An experiment to study the effects of the presence of four different sugar solutions (glucose, sucrose, fructose, and a mixture of the three) on bacterial growth.
3. A experiment to decide whether the color density of fabric specimens depends on the amount of dye used.

Now let us generalize the concept of *ANOVA*

Suppose we have k different levels or k independent random samples, from k different populations (that is, data concerning k treatments, k groups, k methods of production, etc) of a single factor that we wish to compare. Each factor level is also called a treatment. The response of each of the k treatment is a random variable. Then we are concerned with testing the hypothesis that the means of these k populations are equal.

Typical data for a single-factor experiment

Treatment	*Observations (response)*	*Means*	*Sum of squares*
Sample 1	$y_{11}, y_{12}, \ldots, y_{1j}, \ldots, y_{1n_1}$	$\bar{y}_1$	$\sum_{j=1}^{n_1}(y_{1j}-\bar{y}_1)^2$
Sample 2	$y_{21}, y_{22}, \ldots, y_{2j}, \ldots, y_{2n_2}$	$\bar{y}_2$	$\sum_{j=1}^{n_2}(y_{2j}-\bar{y}_2)^2$
...	... ...	...	...
Sample i	$y_{i1}, y_{i2}, \ldots, y_{ij}, \ldots, y_{in_i}$	$\bar{y}_i$	$\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$
...	... ...	...	......
Sample k	$y_{k1}, y_{k2}, \ldots, y_{kj}, \ldots, y_{kn_k}$	$\bar{y}_k$	$\sum_{j=1}^{n_k}(y_{kj}-\bar{y}_k)^2$

A cell in Table say y_{ij}, represents the j^{th} observation taken under treatment i. It is assumed that there are an equal number of observations, n, on each treatment. Now, the observations in Table may be described by the linear statistical model

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}\,;\; i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n. \qquad \ldots (1)$$

where Y_{ij} is a random variable denoting the $(ij)^{th}$ observation, μ is a parameter common to all treatments called the *overall population mean*, τ_i, is a parameter associated with the i^{th} treatment called the i^{th} treatment effect and ε_{ij} is *a random error component*. Notice that the model given by Eq. (1) could have been written as

$$Y_{ij} = \mu_i + \varepsilon_{ij}\,;\quad i = 1, 2, \ldots, k \text{ and } j = 1, 2, \ldots, n. \qquad \ldots (2)$$

where $\mu_i = \mu + \tau_i$ is the mean of the i^{th} treatment.

In this form of model given by equation (2), we see that each treatment defines a population that has mean μ_i consisting of the overall mean μ plus and effect τ_i that is due to the particular treatment. It is assumed that the errors ε_{ij} are normally and independently distributed with mean zero and variance σ^2, and τ_i are normal with mean μ_i and variance σ_i^2. Therefore, each treatment can be thought of as a normal population with mean μ_i and variance σ_i^2. Usually, it is expected that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$ (the *homogeneity of variances*). It may be noted that i^{th} treatment mean μ_i is estimated by its unbiased estimator $\bar{y}_i$ and variance σ_i^2 is estimated by its unbiased estimator S_i^2, where

$$\bar{y}_i = \frac{\sum_{j=1}^{n} y_{ij}}{n} = \frac{y_i}{n},\; i = 1, 2, \ldots, k$$

$$\text{and } S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_{ij}-\bar{y}_i)^2,\; i = 1, 2, \ldots, k$$

Similarly, the overall population mean $\bar{\mu}$ and variance σ^2 can be estimated respectively by the unbiased estimators $\bar{y}$ and S^2, where

$$\bar{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{N},\ S^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y})^2$$

S^2 is also known as the ***total mean square*** (***total variance***) of the overall population.

Definition: If T = ***Total sum of all the observations***

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} = \sum_{i=1}^{k} T_i \text{ where } T_i = \sum_{j=1}^{n_i} y_{ij}$$

$$N = \text{Total sample size} = \sum_{i=1}^{k} n_i$$

Then ***overall sample mean*** (or ***grand mean***) $\bar{y}$ is given as

$$\bar{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} n_i \bar{y}_i}{N} = \frac{T}{N}$$

Theorem: **(Identity for one-way analysis of variance)**

$$\boxed{SST = SST_r + SSE}$$

where, $SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2$ = Total sum of squares

$SST_r = \sum_{i=1}^{k} n_i(\bar{y}_i-\bar{y})^2$ = Treatment sum of squares or between-samples sum of squares.

$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$ = error sum of squares or within sample sum of squares.

Decomposition of the degrees of freedom

d.f. total = *d.f. treatment* + *d.f. error*

$$\sum_{i=1}^{k} n_i - 1 = k - 1 + \sum_{i=1}^{k} n_i - k$$

Definition: *Mean square = sum of squares / degrees of freedom*

$$MST_r = \text{treatment mean square} = \frac{SST_r}{K-1}$$

$$MSE = \text{Error mean square} = \frac{SSE}{N-K}$$

Short cut formulas: Sums of squares

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - C;$$

$SST_r = \sum_{i=1}^{k}\frac{T_i^2}{n_i} - C$, where, $C = \frac{T^2}{N}$ is correction term for the mean with

$$N = \sum_{i=1}^{k} n_i,$$

T_i = Total of the n_i observations in the i^{th} sample $= \sum_{j=1}^{n_i} y_{ij}$

T = grand total of all N observations $= \sum_{i=1}^{k} T_i$

Remarks:

When the population means equal, both $MSTr$ and MSE are estimates of σ^2. However when the null hypothesis is false, the treatment or between sample mean square can be expected to exceed the error within sample mean square. If the null hypothesis is true, it can be shown that the two mean squares $MSTr$ and MSE are independent and that their ratio

$$F\text{-ratio of treatment} = \frac{MSTr}{MSE} = \frac{\sum_{i=1}^{k} n_i(\bar{Y}_i-\bar{Y})^2/k-1}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2/N-k}$$

has an F distribution with $k-1$ and $N-k$ degrees of freedom

One-way *ANOVA* Table

Source of Variation	Degrees of freedom	sum of squares	mean square	F ratio
Treatments Error	$k-1$ $N-k$	$SSTr$ SSE	$MSTr = SSTr/k-1$ $MSE = SSE/N-k$	$F = \frac{MSTr}{MSE}$
Total	$N-1$	SST		

Example 45: (*How does ANOVA estimate from a decomposition of individual observations?*)

Suppose 3 drying formulas for curing a glue are studied and the following times observed.

Formula	*Time*					
A	13	10	8	11	8	–
B	13	11	14	14	–	–
C	4	1	3	4	2	4

Construct ANOVA table.

Solution: There are $N = \sum_{i=1}^{k} n_i = n_1 + n_2 + n_3 = 5 + 4 + 6 = 15$ observations

T = Total sum of observations = 120

The grand mean

$$\bar{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{N} = \frac{T}{N} = \frac{120}{15} = 8$$

That is,

$$y_{ij} = \bar{y} + (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)^2$$

Observation   *grand mean*   *deviation due to treatment*   *error*

For example $13 = 8 + (10 - 8) + (13 - 10)$

So, repeating decomposition for each observation we get

Observation y_{ij} *grand mean* $\bar{y}$ *treatment effects* $\bar{y}_i - \bar{y}$ *error* $(y_{ij} - \bar{y}_i)$

$$\begin{bmatrix} 13 & 10 & 8 & 11 & 8 & - \\ 13 & 11 & 14 & 14 & - & - \\ 4 & 1 & 3 & 4 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 8 & 8 & 8 & 8 & 8 & - \\ 8 & 8 & 8 & 8 & - & - \\ 8 & 8 & 8 & 8 & 8 & 8 \end{bmatrix} + \begin{bmatrix} 2 & 2 & 2 & 2 & 2 & - \\ 5 & 5 & 5 & 5 & - & - \\ -5 & -5 & -5 & -5 & -5 & -5 \end{bmatrix} + \begin{bmatrix} 3 & 0 & -2 & 1 & -2 & - \\ 0 & -2 & 1 & 1 & - & - \\ 1 & -2 & 0 & 1 & -1 & 1 \end{bmatrix}$$

Now,

SST_r = treatment sum of squares

$$= \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 = 5(2)^2 + 4(5)^2 + 6(-5)^2 = 270$$

SSE = error sum of squares

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$$

$$= 3^2 + 0^2 + (-2)^2 + \ldots + (-1)^2 + 1^2 = 32$$

SST = Total sum of squares $= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 = 302$

Hence, $SST = SST_r + SSE$

i.e., $302 = 270 + 32$

One-way ANOVA table				
Analysis of variance Table for Cure Times				
Source of variation	*Degrees of freedom*	*Sum of squares (SS)*	*mean square (MS)*	$F = \frac{MSTr}{MSE}$
Treatment *error*	$k - 1 = 2$ $n - k = 12$	$SSTr = 270$ $SSE = 32$	$MSTr = 135$ $MSE = 2.667$	50.6
Total	$N - 1 = 14$	$SST = 302$		

8.13.3 Method of one-way analysis of variance or one-way classification

Set 1. Set up hypotheses

Null hypothesis: $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$; *i.e.*, k independent population means are equal.

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$, *i.e.*, the k independent population means are not equal. In other words, at least two means of the populations are not equal.

Step 2. ***Level of significance*** (α): choose most commonly used $\alpha = 5\%$ unless otherwise stated.

Step 3. ***Test statistic:*** Under null hypothesis H_0 test statistic is given by

$$F = \frac{MSTr}{MSE} = \frac{\text{Variance between samples}}{\text{Variance within samples}}$$

where, $MSTr = SSTr/k-1$; $MSE = SSE/N-k$

$$SSTr = \Sigma_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2,$$

$$SSE = \Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \Sigma(y_1 - \bar{y})^2 + \ldots + \Sigma(y_k - \bar{y}_k)^2$$

One-way *ANOVA* Table

Source of Variation	*Degrees of freedom*	*sum of squares*	*mean square(MS)*	*F ratio*
Treatments *Error*	$k - 1$ $N - k$	*SSTr* *SSE*	$MSTr = SSTr/k-1$ $MSE = SSE/N-k$	$F = \frac{MSTr}{MSE}$
Total	$N-1$	*SST*		

Step 4. ***Degree of freedom (d.f.)*** $= (k - 1, N - k)$

Step 5. ***Critical value:*** Obtain the critical or tabulated value of the test statistic F from F-table at the pre-specified level of significance for *d.f.* $(k - 1, N - k)$

Step 6. ***Decision:***

(i) If the calculated value of F is less than or equal to the tabulated value of F, then we accept H_0. *i.e.*, population means of k independent populations are equal.

(ii) If the calculated value of F is greater than the tabulated value of F, then we reject H_0 i.e. population means of k independent populations are unequal.

Example 46: Suppose 3 drying formulas for curing a glue are studied and the following times observed. **[TU, BE, 2068 Magh (Back)]**

Formula	*Time*					
A	13	10	8	11	8	–
B	13	11	14	14	–	–
C	4	1	3	4	2	4

Construct ANOVA table and test for the equality of the mean curing times.

Solution:

Step 1. Set up hypotheses

Null hypothesis: $H_0: \mu_1 = \mu_2 = \mu_3$; i.e., there is no significant difference in the mean curing times.

Alternative hypothesis: $H_1: \mu_1 \neq \mu_2 \neq \mu_3$; i.e., there is significant difference in the mean curing times.

Step 2. ***Level of significance*** (α): Since α is not given we take $\alpha = 5\% = 0.05$

Step 3. ***Test statistic:*** Under H_0, test statistic is

$$F = \frac{MSTr}{MSE} = \frac{SSTr/k-1}{MSE/N-k}$$

k = number of samples (treatments)

N = Total number of observations.

Calculation of *MSTr* and *MSE*

A y_1	$(y_1 - \bar{y}_1)^2$	*B* y_2	$(y_2 - \bar{y}_2)^2$	C y_3	$(y_3 - \bar{y}_3)^2$
13	9	13	0	4	1
10	0	11	4	1	4
8	4	14	1	3	0
11	1	14	1	4	1
8	4	-	-	2	1
-	-	-	-	4	1
$\Sigma y_1 = 50$	$\Sigma(y_1 - \bar{y}_1)^2 = 18$	$\Sigma y_2 = 52$	$\Sigma(y_2 - \bar{y}_2)^2 = 6$	$\Sigma y_3 = 18$	$\Sigma(y_3 - \bar{y}_3)^2 = 8$

$n_1 = 5, n_2 = 4, n_3 = 6, \; N = n_1 + n_2 + n_3 = 15$

$\bar{y}_1 = 10, \bar{y}_2 = 13, \bar{y}_3 = 3, k = 3$

Grand mean or mean of sample means is

$$\bar{y} = \frac{T}{N} = \frac{\Sigma_{i=1}^{k} n_i \bar{y}_i}{N} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + n_3\bar{y}_3}{15} = \frac{50 + 52 + 18}{15} = 8$$

$$SSTr = \Sigma_{i=1}^{k} n_i (y_i - \bar{y})^2 = n_1(\bar{y}_1 - \bar{y})^2 + n_2(\bar{y}_2 - \bar{y})^2 + n_3(\bar{y}_3 - \bar{y})^2$$

$$= 5(10 - 8)^2 + 4(13 - 8)^2 + 6(3 - 8)^2 = 270$$

$$SSE = \Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \Sigma(y_1 - \bar{y}_1)^2 + \Sigma(y_2 - \bar{y}_2)^2 + \Sigma(y_3 - \bar{y})^2$$

$$= 18 + 6 + 8 = 32$$

Now $MSTr = \frac{SSTr}{k-1} = \frac{270}{3-1} = 135$ and $MSE = \frac{SSE}{N-k} = \frac{32}{15-3} = 2.667$

Analysis of variance Table for Cure Times

Source of Variation	*Degree of Freedom*	*Sum of Square(SS)*	*Mean Square(MS)*	*F-ratio*
Between samples (treatment)	$k - 1 = 3 - 1 = 2$	$SSTr = 270$	$MSTr = 135$	$F = \frac{MSTr}{MSE} = 50.6187$
Within Samples	$N - k = 15 - 3 = 12$	$SSE = 26$	$MSE = 2.667$	
Total	$15 - 1 = 14$	$SST = 302$		

Step 4. ***Degree of freedom:*** *d.f.* $= (k - 1, N - k) = (2, 12)$

Step 5. ***Critical value:*** Tabulated value of F at α=5% of significance for (2,12) *d.f.* is 3.89 i.e. $F_{(0.05),(2,12)} = 3.89$

Step 6. ***Decision:*** Since $F = 50.62 > F_{(0.05)(2,12)} = 3.89$

So, H_0 cannot be accepted. Hence, we conclude that there is significant difference in the mean curing times.

Example 47: Three randomly selected groups of chickens are fed on three different diets. Each group consists of five chickens. Their weight gains during a specified period of time are as follows.

Diet I	4	4	7	7	8
Diet II	3	4	5	6	7
Diet III	6	7	7	7	8

Test the hypothesis that mean gains of weights due to the three diets are equal.

Solution:

Step 1. Set up hypotheses

Null hypothesis: H_0: $\mu_1 = \mu_2 = \mu_3$; i.e., there is no significant difference in mean weight gains due to different diets.

Alternative hypothesis: H_1: $\mu_1 \neq \mu_2 \neq \mu_3$; i.e., there is significant difference in mean weight gains due to different diets.

Step 2. ***Level of significance*** (α): Since α is not given we take α = 5% = 0.05

Step 3. ***Test statistic:*** Under H_0, test statistic is

$$F = \frac{MSTr}{MSE} = \frac{SSTr/k-1}{MSE/N-k}$$

k = number of samples (treatments)

N = Total number of observations.

Calculation of *MSTr* and *MSE*

Diet I y_1	$(y_1-\bar{y}_1)^2$	*Diet II* y_2	$(y_2-\bar{y}_2)^2$	*Diet III* y_3	$(y_3-\bar{y}_3)^2$
4	4	3	4	6	1
4	4	4	1	7	0
7	1	5	0	7	0
7	1	6	1	7	0
8	4	7	4	8	0
$\sum y_1=30$	$\sum(y_1-\bar{y}_1)^2=14$	$\sum y_2=25$	$\sum(y_2-\bar{y}_2)^2=10$	$\sum y_3=35$	$\sum(y_3-\bar{y}_3)^2=2$

$n_1 = 5, n_2 = 5, n_3 = 5, \; N = n_1 + n_2 + n_3 = 15$

$\bar{y}_1 = 6, \; \bar{y}_2 = 5, \; \bar{y}_3 = 7, \; k = 3$

Grand mean or mean of sample means is

$$\bar{y} = \frac{T}{N} = \frac{\Sigma_{i=1}^{k} n_i \bar{y}_i}{N} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2 + n_3\bar{y}_3}{15} = \frac{30+25+35}{15} = 6$$

$$SSTr = \Sigma_{i=1}^{k} n_i (y_i - \bar{y})^2 = n_1(\bar{y}_1-\bar{y})^2 + n_2(\bar{y}_2-\bar{y})^2 + n_3(\bar{y}_3-\bar{y})^2$$
$$= 5(6-6)^2 + 5(5-6)^2 + 5(7-6)^2 = 10$$

$$SSE = \Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = \Sigma(y_1-\bar{y}_1)^2 + \Sigma(y_2-\bar{y}_2)^2 + \Sigma(y_3-\bar{y})^2$$
$$= 14 + 10 + 2 = 26$$

Now $MSTr = \frac{SSTr}{k-1} = \frac{10}{3-1} = 5$ and $MSE = \frac{SSE}{N-k} = \frac{26}{15-3} = 2.17$

One way *ANOVA* table

Source of Variation	*Degree of Freedom*	*Sum of Square(SS)*	*Mean Square(MS)*	*F-ratio*
Between samples (treatment)	$k-1=3-1=2$	$SSTr = 10$	$MSTr = \frac{10}{2} = 5$	$F = \frac{MSTr}{MSE} = 2.304$
Within Samples	$N-k = 15-3 = 12$	$SSE = 26$	$MSE = \frac{26}{12} = 2.17$	
Total	$15-1 = 14$	$SST = 36$		

Step 4. ***Degree of freedom:*** $d.f. = (k-1, N-k) = (2, 12)$

Step 5. ***Critical value:*** Tabulated value of F at α=5% of significance for (2,12) *d.f.* is 3.89 i.e. $F_{(0.05),(2,12)} = 3.89$

Step 6. ***Decision:*** Since $F = 2.304 < F_{(0.05),(2,12)} = 3.89$

So, H_0 is accepted. Hence, we conclude that there is no significant difference in mean weight gains due to different diets.

Example 48: (***Conducting a one-way analysis of variance***)

Suppose that each laboratory measures the tin-coating weight of 12 disks and that the results are as follows.

Laboratory A	*Laboratory B*	*Laboratory C*	*Laboratory D*
0.25	0.18	0.19	0.23
0.27	0.28	0.25	0.30
0.22	0.21	0.27	0.28
0.30	0.23	0.24	0.28
0.27	0.25	0.18	0.24
0.28	0.20	0.26	0.34
0.32	0.27	0.28	0.20
0.24	0.19	0.24	0.18
0.31	0.24	0.25	0.24
0.26	0.22	0.20	0.28
0.21	0.29	0.21	0.22
0.28	0.16	0.19	0.21

Construct an analysis of variance table.

Solution: The totals for the $k = 4$ samples all of sample size $n_i = 12$ are, respectively, 3.21, 2.72, 2.76 and 3.00; the grand total is $T = 11.69$; and the calculations required to obtain the necessary sums of squares are as follows:

$$N = N = \Sigma_{i=1}^{k} n_i = 12 + 12 + 12 + 12 = 48$$

$$C = \frac{T^2}{N} = \frac{(11.69)^2}{48} = 2.8470$$

$$SST = \Sigma_{i=1}^{k} \Sigma_{j=1}^{n_i} y_{ij}^2 - C$$
$$= (0.25)^2 + (0.27)^2 + \ldots + (0.21)^2 - 2.8470 = 0.0809$$

$$SSTr = \Sigma_{i=1}^{k} \frac{T_i^2}{n_i} - C = \frac{(3.21)^2}{12} + \frac{(2.27)^2}{12} + \frac{(2.76)^2}{12} + \frac{(3.00)^2}{12} - 2.8470 = 0.0130$$

$$SSE = SST - SSTr = 0.00809 - 0.0130 = 0.0679$$

Thus, we get the following analysis of variance table:

Source of variation	*Degree of freedom*	*Sum of Squares(SS)*	*mean square(MS)*	*F*
Laboratories	3	0.0130	0.0043	2.87
Error	44	0.0679	0.0015	
Total	47	0.0809		

Since the value obtained for F exceeds 2.82, the value of $F_{0.05}$ with 3 and 44 degrees of freedom, the null hypothesis can be rejected at the 0.05 level of significance; we conclude that the laboratories are not obtaining consistent results.

Example 49: (*Comparing mean shear strength at three roof locations*)

As part of the investigation of the collapse of the roof a of a building, a testing laboratory is given all the available bolts that connected the steel structure at 3 different positions on the roof. The forces required to see each of these bolts (coded values) are as follows:

Position 1	90	82	79	98	83	91	–
Position 2	105	89	93	104	89	95	86
Position 3	83	89	80	94	–	–	–

Perform an analysis of variance to test at the 0.05 level of significance whether the differences among the sample means at the 3 positions are significant.

Solution:

Step 1. ***Null hypothesis***: $\mu_1 = \mu_2 = \mu_3$

Alternative hypothesis: The μ_i' s are not all equal.

Step 2. ***Level of significance***: $\alpha = 0.05$

Step 3. ***Criterion***: Reject the null hypothesis if $F > 3.74$, the value of $F_{0.05}$ for $k - 1 = 3 - 1 = 2$ and $N - k = 17 - 3 = 14$ degrees of freedom, where F is to be determined by an analysis of variance.

Step 4. ***Calculations***: Substituting $n_1 = 6$, $n_2 = 7$, $n_3 = 4$, $N = 17$, $T_1 = 523$, $T_2 = 661$, $T_3 = 346$, $T = 15, 30$ and

$$\Sigma\Sigma y_{ij}^2 = 138,638$$

into the computing formulas for the sums of squares, we get

$$SST = 138,638 - \frac{1530^2}{17} = 938$$

$$SSTr = \frac{523^2}{6} + \frac{661^2}{7} + \frac{346^2}{4} - \frac{1530^2}{17} = 234$$

and, $SSE = 938 - 234 = 704$

The remainder of the work is shown in the following analysis of variance table:

Source of variation	*Degree of freedom*	*Sum of Squares(SS)*	*mean square(MS)*	*F*
Positions	2	234	117	2.33
Error	14	704	50.3	
Total	16	938		

Step 5. ***Decision***: Since $F = 2.33$ does not exceed 3.74, the null hypothesis cannot be rejected. We cannot conclude that there are differences in the mean shear strengths of the bolts at the three different positions on the roof.

Example 50: The output of three varieties of wheat each grown on 4 plots of land is given below. Analyze the data and set up an *ANOVA* table. State if the variety difference are significant at α=0.05 level. **[TU, BE 2068 Bhadra]**

Varieties of wheat	*yeild tones /hactre*			
A	6	7	3	8
B	5	5	3	7
C	5	4	3	4

Solution: One way *ANOVA* Table

Source of variations	*d.f.*	*Sum of squares*	*mean squares*	*F ratio*
Treatment	$k-1=2$	$SSTr = 8$	$MSTr = 4$	
Error	$N-K=9$	$SSE = 24$	$MSE = 2.67$	$F = 1.498$
Total	11	$SST = 32$		

Since $F_{(0.05),(2,9)} = 4.265 > F$. H_0 is accepted ; i.e., no significance.

Exercise 8

Theoretical Questions

1. Discuss the difference between Null Hypothesis and alternative hypothesis with example. Also explain the type of errors in testing hypothesis. ***[TU, BE, 2058 Bhadra/ 061 Ashwin / 062 Jestha/ 064 Shrawan/ 065 Kartik/ 065 Chaitra]***
2. Discuss the two tailed and one tailed test of hypothesis and give the situation when we use these tests? ***[TU, BE, 2063 Ashwin/ 065 Chaitra/ 067 Mangsir]***
3. Discuss the major steps to be adopted by researchers in testing of hypothesis
   or Explain clearly the procedure generally followed in the testing of hypothesis. ***[TU, BE, 2056 Bhadra/ 057 Bhadra/ 062 Baishakh/ 063 Kartik/ 064 Shrawan/ 065 Kartik]***
4. (a) Explain the following terms: (i) Statistic and parameter;
   (ii) Level of significance; (iii) Critical region and value ***[TU BE 064 Poush/ 065 Chaitra/ 067 Shrawan]***
   (b) Differentiate: (i) critical and acceptance region
   (ii) Type I and II errors ***[TU BE, 2068]***
5. What condition must be met so that z-test can be used to test a hypothesis concerning a population mean? Also write errors arising in testing hypothesis. ***[TU, BE, 2064 Poush]***
6. Discuss the errors of hypothesis with example. Which error is more risky? ***[TU, BE, 2066 Magh]***
7. What are the assumptions for t-test? Describe the procedures of testing mean for small sample case with population standard deviation unknown? ***[TU, BE, 2067 Mangsir]***
8. Distinguish between level of significance and confidence limits. ***[TU, BE, 2067 Mangsir/2068 Bhadra]***
9. Write down the steps for testing hypothesis of population mean for the large sample size. ***[TU, BE, 2068 Bhadra]***
10. What are assumptions for Z-test? Describe the procedures of testing proportion ***[TU, BE 2067 Mangsir]***
11. What are assumptions of t-test? Describe the procedure of test of significance between two means for small sample. ***[TU, BE, 056 Bhadra/ 057 Bhadra/ 058 Shrawan/ 062 Jestha]***
12. Discuss the difference between t-test and z-test. Discuss the test of significance of mean of large sample. ***[TU, BE, 2061 Ashwin/ 062 Baishakh/ 2067 Shrawan]***

What are assumptions of t-test? Describe the procedure of the test of significance of a mean for sample case. *[TU, BE, 2062 Bhadra/ 064 Shrawan/2068 Magh]*

Discuss the large sample test for the equality of two population means. *[TU, BE, 2062 Bhadra/ 2063 Kartik]*

On what basis we give decision of test of hypothesis? Describe the test procedure of pair t-test. *[TU, BE, 063 Ashadh/ 2064 Shrawan/ 2067 Mangsir]*

Discuss with suitable example at what circumstances researcher should apply pair t-test? *[TU, BE, 2063 Kartik]*

Explain the assumptions of t-test and distinguish between large sample and small sample test of significance. *[TU, BE, 2064 Poush]*

When and for what purpose 't' test of significance is used? *[TU, BE, 2065]*

Describe the procedure of test of significance between two means for small sample. *[TU, BE, 2065 Chaitra / 2067 Mangsir]*

Explain the procedure of test of significance for the difference between two means for large samples. *[TU, BE 2068 Jestha/2068 Magh (Back)]*

NOVA

Why do we need ANOVA? Explain the weaknesses of Z- test and t -test.

Explain briefly about one way ANOVA.

Numerical Problems

est for single mean

All cigarettes presently on the market have an average nicotine content of at least 1.6 mg per cigarette. A firm that produces cigarettes claims that it has discovered a new way to cure tobacco leaves that will result in the average nicotine content of a cigarette being less than 1.6 *mg*. To test this claim, a sample of 20 of the firm's cigarettes were analyzed. If it is known that the standard deviation of a cigarette's nicotine content is 0.8 mg, what conclusion can be drawn, at 5 percent level of significance, if the average nicotine content of the 20 cigarettes is 1.54 *mg* ?

[**Hint:** $H_0 : \mu \geq 1.6$ versus $H_1 : \mu < 1.6$, Reject H_0]

A moped manufacturer hypothesized that the mean miles per gallon for its moped is 115.2. It takes a sample of 49 mopeds and finds that sample mean to be 117.4 per gallon. If the population standard deviation is known to be 8.4, test the hypothesis that the true mean miles per gallon is 115.2 against alternative hypothesis that it is greater than 115.2 using the 0.05 significance level.

[*Ans* : $|z|$ = 1.83, *Reject* H_0] *[TU, MBA 2052]*

A sample of 900 members has a mean of 3.4 cm and standard deviation 2.61 cm. Can the sample be regarded as one drawn from a population with mean 3.25 cm. Using the level of significance as 0.05, is the claim acceptable? Also calculate the 95% confidence limits for the population mean.

[Hint $H_0 : \mu = 3.25$ versus $H_1 \neq 3.25$, $|z|$ = 1.72, Accept H_0. *C.I.* for μ = (3.23, 3.57)]

An insurance agent claims that the average age of policy holders who insure through him is less than the average age for all the agents, which is 30 years. A random sample of 100 policy holders who had insured through him gave him the following age distribution:

Age (yrs)	16–20	21–25	26–30	31–35	36–40
No. of persons	12	22	20	30	16

Calculate the mean and *s.d.* of the sample and use these variables to test his claim at 5% level of significance.

[**Hint:** $\bar{X}$ = 28.8, s = 6.35, $H_0 : \mu = 30$ versus $H_1 : \mu < 30$, $|z|$ = 1.89 Reject H_0]

5. A sample of 50 pieces of a certain type of string was tested. The mean breaking strength turned out to be 14.5 kgs. Test whether the sample is from a batch of strings having a mean breaking strength of 15.6 kgs and standard deviation of 2.2 kgs. [**Hint:** $H_0 : \mu = 15.6$ versus $\mu \neq 15.6$, $|z|$ = 3.53, Reject H_0] *[TU, BIE, 2067 Mangsir/ MBA 2039]*

6. The mean breaking strength of cable supplied by a manufacturer is 1800 with standard deviation 100. By a new technique in the manufacturing process, it is claimed that the breaking strength of the cables have been increased. In order to test this claim a sample of 50 cables is tested. It is found that the mean breaking strength is 1850. Can we support the claim at 0.01 level of significance?

[**Hint:** $H_0 : \mu$ 1800 versus $H_1 : \mu > 1800$, H_0 is rejected] *[Pokhara Uni, BE, 2001]*

7. Suppose that we want to investigate whether on the average men earn more than $20 per week more than women in a certain industry. If sample data show that 60 men earn on the average $\bar{X}_1$ = $585 per week with a *s.d.* of s_1 = $31.20, while 60 women earn on the average $\bar{X}_2$ = $532.20 per week with a standard deviation of s_2 = $36.40, what can you conclude at 0.01 level of significance?

[Ans: z = 2.07, the data fail to confirm the hypothesis that the men earn in excess of $20/ week more than women]

Z–test for difference

1. An examination was given to two classes consisting of 40 and 50 students respectively. If the first class mean mark was 74 with standard deviation of 8, while in the second class the mean mark was 78 with *s.d.* of 7. Is there a significance difference between the performance of two classes at a level of significance of (a) 0.05 and (b) 0.01. Also construct the confidence interval for $(\mu_1 - \mu_2)$, the difference of means, at $\alpha = 0.05$.

[Ans : (a) $|z|$ = 2.49, Reject H_0; (b) $|z|$ = 2.49, Accept H_0; *C.I.* = (–7.148, –0.85)]

2. To investigate a possible "built–in" sex bias in a graduate school entrance examination, 50 male and 50 female graduate students who were rated as above average graduate students by their professors were selected to participate in the study by a actually taking this test. Their test results on this examination are summarized in the following table: *[TU, BE, 2064 Poush]*

	Males	*Females*
$\bar{X}$	720	693
s^2	104	85
n	50	50

Do these data indicate that males will, on the average, score higher than females of the same ability on this exam? Use $\alpha = 0.05$

[**Hint:** $H_0 : \mu_1 = \mu_2$ versus $\mu_1 > \mu_2$, $|z|$ = 13.89, Reject H_0]

3. A study on expenditure behavior of tourists in Nepal revealed that the average expenditure of 100 European tourist per day is NRs. 9200 with *s.d.* of NRs. 600. Also, the average expenditure of 100 American tourists per day is NRs. 10,000 with

standard deviation of NRs. 500. Can you conclude that the expenditure behavior of European and American tourists are same?

[Hint: take $\alpha = 5\%$, $|z| = 10.242$, Reject H_0] [*TU, MBS, 2059*]

4. In selecting a sulphur concrete for roadway construction in regions that experience heavy frost, it is important that the chosen concrete have a low value of thermal conductivity to minimize subsequent damage due to changing temperatures. Suppose two types of concrete, a graded aggregate and a no-fines aggregate, are being considered for a certain road. Following table summarizes data from an experiment carried out to compare the two types of concrete. Does this information suggest that true average conductivity for the graded concrete exceeds that for the no-fines concrete? Use a test at $\alpha = 0.01$.

Type	*Sample size*	*Sample average conductivity*	*Sample S.D.*
Graded	42	0.486	0.187
No fines	42	0.359	0.158

[Hint: $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$; $|z| = 3.36$ Reject H_0]

t-test for a single mean

1. A random sample of 16 values from a normal population showed a mean of 41.5 inches and sum of squares of deviations from this mean equal to 135 square inches. Show that the assumption of a mean of 43.5 inches for the population is not reasonable. Obtain 95% and 99% fiducial limits for the same.

[Ans: $|t| = 2.667$, H_0 is rejected, 95% limit for μ = (39.902, 43.098); 99% limits for μ = (39.29, 43.71)] [*TU 2053*]

2. A company claims that the mean life time of its electric bulbs is 28 months. A random sample of 10 bulbs has the following life in months: 24, 26, 32, 28, 20, 20, 23, 24, 30 and 43. Test the claim of the company at 5% level of significance and obtain 95% fiducial limit for the population mean.

[Ans: H_0 is accepted, (22.92, 33.08)] [*TU Model*]

3. The National Bureau of Standard has previously reported the value Se (Selenium) in *NBS* orchard leaves to be 0.080. A random sample of size $n = 6$ give the following determinations 0.072, 0.073, 0.080, 0.078, 0.088, 0.080. Does the data contradict the previously reported value? [Given: for 5 *d.f.* $P(|t| > 2.571) = 0.05$]

[Hint: $H_0 : \mu = 0.08$ versus $\mu \neq 0.08$, $|t| = 0.6356 < 2.571$, Accept H_0]

[*Pokhara, uni, BE, 2002*]

4. The manufacture of Shilpa electric bulbs claims that have a mean life of 25 months. A random sample of 9 such bulbs gave the following values:
Life in months: 24, 26, 32, 28, 20, 20, 23, 27 and 34. Can you regard the manufacturer's claim to be valid at 5% level of significance?

[Hint: $\bar{x} = 26$, $s^2 = 23.75$, $t = 0.62$, Accept H_0] [*TU, 2052 MBA*]

5. A random sample of size 16 showed a mean of 52 with a standard deviation 4. Test the hypothesis that the mean of the population is 50.

[Ans: $|t| = 2 < t_{\alpha/2, n-1} = 2.131$, H_0 is accepted] [*TU, MBS 2065*]

6. The height of 10 adult males selected at random from a given locality had mean 158 cm and variance 39.06 cm. Test at 5% level of significance the hypothesis that the adult male of the given locality are on the average less than of 162.5 cm tall. (Given for 9 degree of freedom $t = \pm 1.83$ for one tailed test) [*TU, BE, 2056 Bhadra*]

[Hint: Set $\mu = 162.5$ versus $\mu < 162.5$ and solve]

7. The following radiation reading was obtained from television display area in a sample of 10 department stores: 0.40, 0.43, 0.60, 0.15, 0.50, 0.80, 0.36, 0.16, 0.89 the recommended limit from this type of radiator exposure is 0.5 m/hr. Assuming that the observation come from a normal distribution with mean μ the true average amount of radiation in television display area in all department stores, test $H_0 : \mu$ = 0.5 versus $H_1 : \mu > 0.5$ using level 0.05% level of significance.

[Hint: Set $H_0 : \mu = 0.5$ versus $H_1 : \mu > 0.5$] [*TU, BE, 2062 Bhadra*]

8. The specification for a certain kind of nylon ribbon call for a mean breaking strength of 200 pounds. If 15 pieces of such nylon ribbon (randomly selected from different rolls) have a mean breaking strength of 198.2 pounds with a standard deviation of 4.2 pounds. Could this evidence support for required specification? Test it at 0.05 level of significance. **[Hint:** Set $H_0 : \mu = 200$ versus $H_1 : \mu < 200$][*TU, BE 2062 Baishakh*]

9. Given a random sample of 5 pints from different production lots, we want to test whether the fat content of a certain kind of ice cream exceeds 14%. What can we conclude at the 0.01 level of significance about the null hypothesis $\mu = 14\%$ if the sample has the mean $\bar{x}$ = 14.9% and the standard deviation σ = 0.42%.

[Ans : $t = 4.79$, Reject H_0]

t–test for difference of means

1. Two different types of drugs D_1 and D_2 were applied on certain patients for increasing weight at interval of one week time period. From the following observation, can you conclude that the second drug is more effective in increasing weight, use 1% level of significance.

D_1	8	12	13	9	3	8	10	9
D_2	10	8	12	15	6	11	12	12

[Hint: Set $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$ $|t| = 1.208 < 1.761$, H_0 is accepted]

2. For a random sample of 10 pigs fed on diet A, the increases in weight (in lbs) in a certain period were 10, 17, 13, 12, 9, 8, 14, 15, 6 and 16. For another random sample of 12 pigs fed on diet B, the increase in weight in the same period were 14, 18, 8, 21, 23, 10, 17, 12, 22, 15, 7 and 13. Test whether diets A and B differ significantly as regards their effect on increase in weight. [*TU, 2054*]

[Hint: $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_2 \neq \mu_2$, $|t| = 1.50 < 2.086$, H_0 is accepted]

3. The monthly advertising expenditure of a company for two products *A* and *B* are as follows:

	Expenditure in Rs	
Month	*Product A*	*Product B*
January	100	175
February	120	200
March	125	250
April	145	225
May	150	200
June	140	150
July	200	200

Is there sufficient evidence to conclude that the average expenditure on advertising on product B is more than that on product A? **[Hint:** Set $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$ (Left tailed test) $|t| = 3.52 > 1.782$, H_0 is rejected] [*TU 2058 MBS*]

4. Sample of sales in similar shops in towns A and B regarding a new product yielded the following information:

For town $A : \bar{X}_1 = 3.45, \Sigma X_1 = 38, \Sigma X_1^2 = 228, n_1 = 11$

For town $B : \bar{X}_2 = 4.44, \Sigma X_2 = 40, \Sigma X_2^2 = 222, n_2 = 9$

Is there any evidence of difference in sales in the two towns? Test at 5% level of significance. [**Hint:** Set $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$, $|t| = 0.788$, H_0 is accepted]

5. Two horses A and B are tested according to the time (in seconds) to run a particular track with the following results:

Horse	*A*	28	30	32	33	33	29	34
Horse	*B*	29	30	30	24	27	29	

(i) Test whether you can discriminate between two horses.

(ii) Would you conclude that the horse A has greater running capacity than the horse B?

[Ans: (i) $t = 2.45$, Reject H_0. Hence we can discriminate between two horses; (ii) $t = 2.45$ Reject H_0 and hence accept H_1]

6. Measuring specimens of nylon yarn taken from two spinning machines, it was found that 8 specimens from the first machine had a mean denier of 9.67 with a standard deviation of 1.81, while 10 specimens from the second machine had a mean denier of 7.43 with a *s.d.* of 1.48. Assuming that the populations sampled are normal and have the same variance, test the null hypothesis $\mu_1 - \mu_2 = 1.5$ against the alternative hypothesis $\mu_1 - \mu_2 > 1.5$ at the 0.05 level of significance.

[Ans: $t = 0.96$, Accept H_0]

7. The heights of six randomly chosen sailors are inches: 63, 65, 68, 69, 71 and 72. Those of 9 randomly chosen soldiers are 61, 62, 65, 66, 69, 70, 71, 72 and 73. Discuss in the light that these data throw on the suggestion that sailors are on the average taller than soldiers.

[**Hint:** $H_0 : \mu_1 > \mu_2$, $|t| = 0.09 < 1.76$, Accept H_0] [*TU, MBA 2050/ P.U. BE 2002/ 2009*]

Paired t-test

1. Marks of 8 students before and after tuition is given below:

	1	2	3	4	5	6	7	8
Before tuition	50	54	52	53	48	51	53	54
After tuition	54	57	54	55	52	56	56	55

Can you conclude that tuition has benefited the students?

[**Hint:** set $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 < \mu_2$. $t| = 6.48 > 1.895$, H_0 is rejected]

2. In a manufacturing company the new modern manager is in a belief that music enhances the productivity of the workers. He made observation on six workers for a week and recorded the production before and after the music was installed. From the data given below, can you conclude that the productivity has indeed changed due to music: [*TU, MBA 2056*]

Employee	1	2	3	4	5	6
Week without music	219	205	226	198	209	216
week with music	235	186	240	203	221	205

[Ans: $|t| = 0.477$, H_0 is accepted]

3. A special coaching class on Mathematics subject in a group of 10 students field the

following change in score : 8, 10, −2, 0, −5, −1, 9, 12, 6, 5. Test at 5% level of significance whether the coaching class was effective or not.

[Ans: $|t| = 2.29$ H_0 is rejected] [*TU 2059*]

4. Sales of new electronic item in six stores before and after special promotional program are observed as follows: [*TU 2061*]

Store	1	2	3	4	5	6
sales before campaign	50	30	31	48	55	42
sales after campaign	52	29	30	52	56	45

can you judge the special promotional program a success? ($\alpha = 0.01$) [Ans: $|t| = 1.58$, H_0 is accepted].

5. The sales figure of a item in six shops before and after an advertisement is given as

Before	53	28	31	48	50	42
After	58	29	30	55	56	45

Test whether the advertisement was effective? [*TU, BE 2068 Jestha*]

ANOVA

1. Two horse A and B were tested according to the time (in second) to run the particular track with the following results. [*TU 2053*]

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	-

Test whether the two horses are equally consistent or not?

[Ans: $F = 1.03$, H_0 is accepted]

2. The following table represents the sales of three salesmen in four different districts.

District	*Sales figure (000) sales persons*		
	A	*B*	*C*
Kathmandu	14	20	16
Lalitpur	12	23	15
Bhaktpur	10	20	10
Palpa	8	18	12

Test whether there is any significant difference in the sales of different districts.

[Ans: $F = 0.55$, H_0 is accepted] [*TU 2057*]

3. Test whether two populations have the same variance or not from the following.

Sample I	*Sample II*
$n_1 = 7$ $\Sigma(y_1 - \bar{y})^2 = 320$	$n_2 = 6$ $\Sigma(y_2 - \bar{y})^2 = 350$

[Ans: $F = 131$, H_0 is accepted] [*TU 2057*]

4. The following are the numbers of mistakes made in 5 successive days for 4 technicians working for a photographic laboratory:

Technician I	*Technician II*	*Technician III*	*Technician IV*
6	14	10	9
14	9	12	12
10	12	7	8
8	10	15	10
11	14	11	11

Test at the level of significance $\alpha = 0.01$, whether the differences among the 4 sample means can be attributed to chance. [Ans. $F = 0.68$, not significant]

5. Given the following observations collected according to the one-way analysis of variance design,

Treatment I	6	4	5	–	–
Treatment II	13	10	13	12	–
Treatment III	7	9	11	–	–
Treatment IV	3	6	1	4	1

a) Decompose each observation y_{ij} as

$y_{ij} = \bar{y} + (\bar{y}_i - \bar{y}) + (y_i - \bar{y}_i)$ and obtain the sum of squares and degrees of freedom for each component.

b) Construct the analysis of variance table and test the equality of treatments using $\alpha = 0.05$ [Ans: (a) $SSTr = 204$, with 3 degrees of freedom; $SSE = 34$ with 11 degrees of freedom; $SST = 238$ with 14 degrees of freedom (b) $F = 22.0$, significant at $\alpha = 0.05$]

6. A completely randomized design experiment with 10 plots and 3 treatments gave the following results. Perform the analysis of variance for the significance of treatment effects.

Plot no.	1	2	3	4	5	6	7	8	9	10
Treatment	*B*	*B*	*C*	*A*	*C*	*C*	*A*	*B*	*A*	*B*
Yield	10	4	12	3	8	7	5	6	4	7

[Hint

Observation	*Treatment*			
	A	*B*	*C*	*Total*
1	3	10	12	25
2	5	4	8	17
3	4	6	7	17
4	–	7	–	7
Total	12	27	27	60

[Ans: $F = 3.81$, H_0 is accepted]

7. To find the best arrangement of instrument on a control panel of an airplane, 3 different arrangements were tested by simulating an emergency condition and observing the reaction time required to correct the condition. The reaction times (in tenths of a second) of 28 pilots (randomly assigned to the different arrangements) were as follows.

Arrangement 1	14	13	9	15	11	13	14	11				
Arrangement 2	10	12	9	7	11	8	12	9	10	13	9	10
Arrangement 3	11	5	9	10	6	8	8	7				

Test at the level of significance $\alpha = 0.01$ whether we can reject the null hypothesis that the differences among the arrangements have no effect.

[Ans: $F = 11.3$, significant at 0.01 level]