CHAPTER - 10

INFERENCE

10.1 INTRODUCTION

From the definition of statistics, we know that any statistical investigation is concerned with the study of a collection or a totality of units (objects). Such a group or collection of objects is called **population** or **universe.** For example, heights and weights of students, marks scored in different subjects etc. Further a complete enumeration of the entire population associated with any statistical investigation. is highly impossible and at times not desirable also. For example, when the population is vast, accessing each and every unit is practically impossible. In such cases the investigation would be based on a sample (or a representative portion) of the population.

Definition : (sampling) A finite subset of a population is called a sample and the number of objects in a sample is called the **sample size.**

Definition : In order to determine some population characteristics, the objects in the sample are observed and the sample characteristics are used to approximately estimate the same for the entire population. The inherent and unavoidable error in any such approximation is known as **sample error.**

The investigation based on a sample (a representative portion of the population). is called **sample survey.** Suppose n units are selected from the population, these selected units form a sample of size n. If these units are selected by providing equiprobability to all the units of the population, the sample is called **random sample.**

Some of the important types of sampling are

i) **Purposive sampling**

ii) **Random sampling**

iii) **Simple sampling**

iv) **Stratified sampling**

For a variable in the population, suppose we find constants such as mean, standard deviation, etc., these constants are called **parameters** of the population. On the other hand, if we find mean, standard deviation, etc., of the sample, they are called **statistics.**

Parameter is a statistical constant of the population. Statistic is a function of the sample values.

Thus, mean height of students of a college is parameter. Whereas, mean height of 50 randomly selected students of the college is a statistic.

10.2 SAMPLING DISTRIBUTION AND STANDARD ERROR

Definition: Suppose a sample of size n is drawn from a population and the sample mean $\bar{x}$ is calculated. From the population, many such samples $\bar{x}$ of the same size can be drawn. For each of these samples mean can be computed. And so, there can be many values of $\bar{x}$.

The mean of diff. samples in a population

Suppose these different values are tabulated in the form of a frequency distribution, the resulting distribution is called **Sampling distribution** of $\bar{x}$. The standard deviation of this sampling distribution is called **Standard Error (S.E.)**

Definition: The distribution of values of a statistic for different samples of the same size is called sampling distribution of the statistic.

Definition: Standard Error (S.E.) of a statistic is the standard deviation of the sampling distribution of the statistic.

Sampling distributions of other status such as sample variance, sample median, etc., can also be computed, Obviously in each of these cases,. the corresponding standard deviation would be the standard error (S.E.) Thus the mean and variance of the sampling distribution of the statistic t are given by

$$\bar{t} = \frac{(t_1 + t_2 + \ldots\ldots + t_n)}{n}$$

$$\text{and} \quad \text{Var. } t = \frac{\sum(t_i - \bar{t})^2}{n}$$

Consider a population whose mean is μ and standard deviation is σ. Let a random sample of size n be drawn from this population. Then, 'the sampling distribution of $\bar{x}$ has mean μ and standard error $\dfrac{\sigma}{\sqrt{n}}$

Let a random sample of size n_1 be drawn from a population whose mean is μ_1 and standard deviation is σ_1. Also, let a random sample of size n_2 be drawn from another population whose mean is μ_2 and standard deviation is σ_2. Let $\bar{x}_1$ be the mean of the first sample and $\bar{x}_2$ be the mean of the second sample. Then, $(\bar{x}_1 - \bar{x}_2)$

has mean $(\mu_1 - \mu_2)$ and standard error $\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Example 1. Consider the collection of 25 samples of size 2 chosen from the population 1, 2, 3, 4, 5. Population mean

$$\mu = \frac{(1+2+3+4+5)}{5} = 3$$

$\frac{2}{2}$ $\frac{3}{2}$ $\frac{4}{2}$ $\frac{5}{2}$ $\frac{6}{2}$

The population variance is

$$\sigma^2 = \frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5} = 2$$

The sample mean of the 25 samples of size 2 are given in the table.

1	1.5	2	2.5	3
1.5	2	2.5	3	3.5
2	2.5	3	3.5	4
2.5	3	3.5	4	4.5
3	3.5	4	4.5	5

Mean of the sample means

$(1-3)^2 + (1.5-3)^2 +$
$(2-3)^2$

$$= \mu_{\bar{x}} = \frac{\text{sum of the sample means}}{25} = \frac{75}{25} = 3$$

We note that $\mu = \mu_{\bar{x}} = 3$

Similarly it can be verified that variance of the sampling distribution of means $\sigma_{\bar{x}}^2 = \dfrac{\sum(\bar{x}_i - \mu_{\bar{x}})^2}{n} = 1$ so that $\sigma^2 \neq \sigma_{\bar{x}}^2$

10.3 STANDARD ERROR OF PROPERTIES

Suppose in a population it is possible to make **dichotomous classification** (classification into two classes) of units as those which 'posses an attribute' and those which 'do not posses the attribute' For example:

1. Students of a college are classified as 'poor' and 'not poor'
2. People of a village are classified as 'literates' 'illiterates'.
3. Fruits are classified as 'ripe and 'unripe'
4. Male and females in an a audience.

In a population, let P be the proportion of units which 'posses the attribute'. From such a population, suppose a random sample of size n is drawn. Let x of these n units belong to the class 'possess the attribute'. Then $p = \dfrac{x}{n}$ is the sample proportion of the attribute.

Here, $p = \dfrac{x}{n}$ has mean P and standard error $\sqrt{\dfrac{PQ}{n}}$ where $Q = 1-P$.

Let a random sample of size n_1 be drawn from a population with proportion P_1 of an attribute. Let X_1 units in the sample possess the attribute. Then, the sample proportion is $p_i = \dfrac{x_1}{n_1}$. Let a random s~~i~~ample of size n_2 be drawn from a population with proportion P_2 ~~the~~ of an attribute. Let x_2 units in this sample possess the attribute. Then, the sample proportion is $p_2 = \dfrac{x_2}{n_2}$. Here, the difference of the sample proportions

$(p_1 - p_2)$ has mean $(p_1 - p_2)$ and standard error

$$\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}} \text{ where } Q_1 = 1 - P_1 \text{ and } Q_2 = 1 - P_2.$$

Here, if $P_1 = P_2 = P$, the standard error is $\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Thus, the following are the means and standard error of some statistics.

Statistic	Mean	Standard Error
$\bar{x}$ (sample mean)	m	$\frac{\sigma}{\sqrt{n}}$
$\bar{x}_1 - \bar{x}_2$ (difference of means)	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
p (sample proportion)	P	$\sqrt{\frac{PQ}{n}}$
$p_1 - p_2$ (difference of proportions)	$P_1 - P_2$	$\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$
$p_1 - p_2$ (when $P_1 = P_2 = P$)	0	$\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

Note 1. Standard error plays a very important role in the theory of large samples and forms the basis of hypothesis.

Note 2. If t is any statistic, then for large samples

$$Z = \frac{t - E(t)}{\sqrt{\text{var}(t)}} \sim N(0,1) \text{ asymptotically as } n \to \infty \text{ and hence}$$

$$Z = \frac{t - E(t)}{S.E} \sim N(0, 1) \text{ asymptotically as } n \to \infty$$

10.4 TESTING OF HYPOTHESES

The 'theory of statistical inference' has two branches, namely, 'theory of estimation' and 'theory of testing of hypotheses'. Theory of estimation, is a technique of estimation the population parameters by using simple statistics. Whereas, in testing of hypotheses, validity of presumptions regarding the parameters of the population are verified with the help of sample statics.

Definition : Hypothesis is a statement about the values of the population parameter. It is made on the basis of the information obtained by experimentation. Testing of hypothesis is a procedure for deciding whether to accept or reject the hypothesis. Procedures which enable us to decide whether to accept or reject hypothesis are called tests of hypothesis, also known as tests of significance.

In the theory of probability distribution for large n almost all the distributions like binomial, Poisson, Chi–square can be approximated very closely by a normal probability curve.

Suppose we presume that 'Mean income of a Mangalorian is Rs. 50 per day'. In order to test this presumption, some Mangalorian are randomly selected and their mean daily income is found. If this sample mean is in the close neighbourhood of Rs. 50, we conclude that 'mean income of a Mangalorian is Rs. 50 per day'. On the other hand, if the sample mean deviates much from Rs. 50 we conclude that 'mean income of a Mangalorian differs from Rs. 50 Thus if the sample mean is Rs. 50, since this value is close to Rs. 50, we conclude that 'mean income of a Mangalorian in Rs. 50 per day'. On the other hand, if the sample mean is Rs. 55.0 since this value is far away from Rs. 50 we conclude that 'mean Income of a Mangalorian differs from Rs. 50'.

10.5 STATISTICAL HYPOTHESIS

A statistical hypothesis is an assertion regarding the statistical distribution of the population. It is a statement regarding the parameters of the population.

Statistical hypothesis is denoted by H.

EXAMPLES

1. H_0: The population has mean $\mu = 25$
2. H: The population is normally distributed with mean $\mu = 25$ and standard deviation $\sigma = 2$.

In a test procedure, to start with, a hypothesis is made. The validity of this hypothesis is tested. If the hypothesis is found to be true, it is accepted. On the other hand, if it is found to be false, it is rejected.

The hypothesis which is being tested for possible rejection is called null hypothesis. The null hypothesis is denoted by H_0.

If the null hypothesis is found to be false another hypothesis which contradicts the null hypothesis is accepted. This hypothesis which is accepted when the null hypothesis is rejected is called alternative hypothesis. The alternative hypothesis is denoted by H_1.

For Example, (i) to test whether there is a difference between two populations we take the null hypothesis as H_0 ; there is no difference between the two populations. Rejecting this null hypothesis H_0 will mean that the two populations are different. Acceptance (non rejection) of the null hypothesis will mean that there is no difference in the population (populations are same).

(ii) If we want to test the null hypothesis that the population has specified mean μ_0 then the null hypothesis is $H_0 : \mu = \mu_0$. The alternative hypothesis H_1 will be (i) $H_1 : \mu \neq \mu_0$ (ii) $H_1\ \mu > \mu_0$ (iii) $H_1 : \mu < \mu_0$.

The alternative hypothesis in (i) is known as a two tailed alternative and the alternative hypotheses in (ii) and (iii) are known as right tailed alternative and left tailed alternative respectively.

10.6 TEST PROCEDURE

Suppose the mean of a population is not known. We want to test whether the mean is a given value μ_0. The test procedure is as follows.

The null hypothesis '$H_0 : \mu = \mu_0$ (population mean is μ_0)' is considered. A random sample of size n is drawn from the population. The sample mean $\bar{x}$ is found. If the difference between $\bar{x}$ and μ_0 is small (insignificant), the null hypothesis is accepted. If the difference its large (significant), the null hypothesis is rejected and an alternative hypothesis '$H_1 : \mu \neq \mu_0$ (population mean differs from μ_0)' is accepted. Thus, a constant k is chosen such that

if $|\bar{x} - \mu_0| > k \times (S.E.)$, H_0 is rejected.

That is, if $|\bar{x} - \mu_0| > k\dfrac{\sigma}{\sqrt{n}}$, H_0 is rejected.

That is, if $\dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} > k$, H_0 rejected.

Here, $|Z| = \dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}$ is called test statistic.

Test statistic is the statistic based on whose distribution testing is conducted.

The constant k is called **critical value.** The critical value is dependent on the distribution of the test statistic and level of significance.

10.7 CRITICAL REGION

It is possible to draw many samples of the same size n from a population drawn. Let S be the set of all such samples of size n that can be drawn from the population. Then, S is called **Sample space.** While testing a null hypothesis, among the samples which belong to S, some sample lead to the acceptance of the null hypothesis, whereas, some others lead to the rejection of the null hypothesis. The set of all those samples belonging to the sample space which lead to the rejected of the null hypothesis is called **critical region.** The critical region is denoted by ϖ. The critical region is also called **rejection region.** The set of samples which lead to the acceptance of the null hypothesis, is the **acceptance region.** It is $(S - \varpi)$

In fact, to decide whether the sample under consideration belongs to ϖ, the criteria $|Z| > k$ is adopted. And so, in effect, the critical region is defined by $|Z| > k$.

Test of significance level of sign

10.8 ERRORS OF THE FIRST AND THE SECOND KIND (Type I and type II errors)

There are two possible types of errors which may arise in testing a statistical hypothesis. This can be explained as follows :

While testing a null hypothesis against an alternative hypothesis, one of the following four situations arise :

Actual fact	Decision based on the sample	Error	
1 H_0 is true	accept H_0	correct decision	
2 H_0 is true	reject H_0	wrong decision	Type I
3 H_0 is not true	accept H_0	wrong decision	Type II
4 H_0 is not true	reject H_0	correct decision	

Here, in situations (2) and (3), wrong decisions are arrived at. These wrong decisions are termed as Error of the first kind (Type I error) and Error of the second kind (Type II error) respectively. Thus,

(1) Error of the first kind (Type I error) is taking a wrong decision to reject the null hypothesis when it is actually true.

(2) Error of the second kind (Type II error) is taking a wrong decision to accept the null hypothesis when it is actually not true.

The test procedure should be so framed as to safeguard against both these types of errors.

The probability of occurrence of the first kind of error is denoted by α. It is called **level of significance.**

Definition: The level of significance is the probability of rejection of the null hypothesis when it is actually true. Usually, the level of significance is fixed at 0.05 or 0.01. in other words, the level is fixed at 5% or 1%.

The probability of occurrence of the second kind of error is denoted by β.

The value (1 – β) is called power of the test.

Definition : Power of a test is the probability of rejecting H_0 when it is not true.

While testing, the level of significance α is decided in advance, so that the critical value k can be determined in a way that the power (1 – β) is maximum.

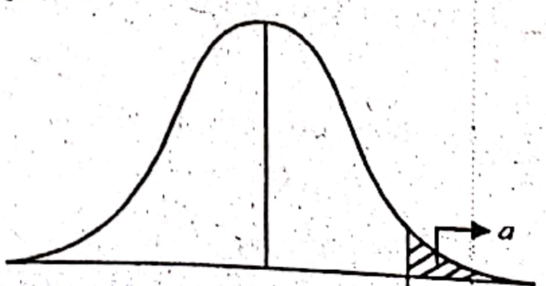
Thus it is observed that, the critical value k is based on the level of significance. For tests which are based on normal distribution, if $\alpha = 0.05$, the critical value is $k = 1.96$. If $\alpha = 0.01$, the critical value is $k = 2.58$.

Remark: In fact, a decision to accept H_0 is based only on the given data. And so, rather than making an assertive statement 'H_0 is accepted', we would make a statement 'H_0 is not rejected'. However, at the level of this book, we will not bother about the subtle difference between these statements.

10.9 TWO-TAILED AND ONE-TAILED TESTS

(Two–sided and one–sided tests)

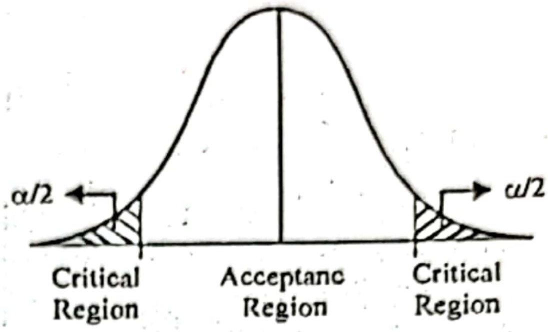
null hypothesis



While testing H_0, if the critical region is considered at one tail of the sampling distribution of the test statistic, the test is one–tailed test.

One the other hand, if the

critical region is considered at both the tails of the sampling distribution of the test statistic, the test is two–tailed test.



10.10 LARGE SAMPLE TESTS

In the previous sections, we have discussed the means and standard deviations (standard errors) of the sampling distributions sample mean, sample proportion, etc. It can be shown that for large samples, these statistic have normal distribution. And so, while testing hypothesis concerning means and proportions, the critical values can be obtained by using normal distribution.

10.11 TEST FOR MEAN

Suppose the mean of a population is not known. We want to test whether the mean is the given value μ_0. The null hypothesis is $H_0 : \mu = \mu_0$. The alternative hypothesis is $H_1 : \mu \neq \mu_0$.

For large random sample of size n from the population, under

H_0, $Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ is $N(0,1)$

Therefore, the test statistic is $|Z| = \dfrac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}$

For the sample, if the calculated value $|Z|_{cal} > k$, H_0 is rejected.

On the other hand, if $|Z|_{cal} \leq k$ H_0 is accepted.

For the level of significance $\alpha = 0.05$, the critical value is $k = 1.96$. However, for $\alpha = 0.01$, the critical value is $k = 2.58$.

Note: Here, if σ is not known, the test statistic is $|Z| = \dfrac{|\bar{x} - \mu_0|}{s/\sqrt{n}}$

where s is the sample standard deviation.

10.12 TEST FOR EQUALITY OF MEANS

Suppose we are interested in testing whether means of two populations are equal.

The null hypothesis is $H_0 : \mu_1 = \mu_2$ (the means of the two populations are equal). The alternative hypothesis is $H_1 : \mu_1 \neq \mu_2$

Let μ be the common mean and let σ_1 and σ_2 be the standard deviations of the two populations.

Let a random sample of size n_1 be drawn from the first population. Let the sample mean be $\bar{x}_1$. Also, let a random sample of size n_2 be drawn from the second population. Let the mean of this sample be $\bar{x}_2$.

Then, $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

And so, the test statistic is $$|Z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

For the samples, if $|Z|_{cal} > k$, H_0 is rejected.

On the other hand if $|Z|_{cal} \leq k$, H_0 is accepted.

For the level of significance $\alpha = 0.05$, the critical value is $k = 1.96$. However, for $\alpha = 0.01$, the critical value is $k = 2.58$. Here, if

and σ_1 and σ_2 are not known, the test statistic is $$|Z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where s_1 and s_2 are the sample standard deviations.

10.13 TEST FOR PROPORTION

Suppose the proportion of an attribute in a population is not known, we want to test whether the proportion is a given value P_0. The null hypothesis is H_0: $P = P_0$. The alternative hypothesis is $H_1 : P \neq P_0$.

In a large random sample of size n from the population, let x units posses the attribute. Then, the sample proportion is $p = \frac{x}{n}$

And so, $Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}}$ is $N(0,1)$

Therefore, the test statistic is $|Z| = \frac{|p - P_0|}{\sqrt{\frac{P_0 Q_0}{n}}}$

For the sample, if $|Z|_{cal} > k$, H_0 is rejected.

On the other hand, if $|Z|_{cal} \leq H_0$ is accepted.

For the level of significance $\alpha = 0.05$, the critical value is $k = 1.96$. However, for $\alpha = 0.01$, the critical value is $k = 2.58$.

10.14 TEST FOR EQUALITY OF PROPORTIONS

Suppose there are two populations with unknown proportions, and we wish to test whether the proportions (of certain attribute) in the two populations are equal. The null hypothesis is H_0: $P_1 = P_2$ (the proportions are equal). The alternative hypothesis is $H_1: P_1 \neq P_2$.

Let P be the common proportion. Let a large random sample of size n_1 be drawn from the first population. Among these n_1 units, let x_1 units possess the attribute, so that the sample proportion is $p_1 = \frac{x_1}{n_1}$.

Also, let a large random ~~ramp;~~ sample of size n_2 be drawn from the second population. Among the units, let x_2 units possess the attribute, so that the sample proportion is $p_2 = \frac{x_2}{n_2}$

Then, $$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ is $N(0, 1)$

And so, the test statistic is

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Generally, the common proportion P will not be known. And so, it is estimated from the samples.

The estimate is $P = \dfrac{x_1 + x_2}{n_1 + n_2}$

Also, $P = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

For the sample, if $|Z|_{cal} > k$, is rejected

On the other hand, if $|Z|_{cal} \leq k$, H_0 is accepted.

For the level for significance $\alpha = 0.05$, the critical value is $k = 1.96$. However, for $\alpha = 0.01$, the critical value is $k = 2.58$.

The following is a procedure for testing a statistical hypothesis.

Step I Set up the null hypothesis H_0.

Step II Set up the alternative hypothesis H_1. This will determine whether we have to use right tailed or left tailed or two tailed test.

Step III Compute the test statistic under the null hypothesis H_0.

Step IV Choose the appropriate level of significance α (generally 5% or 1%).

Step V Compare the computed value of $|Z|$ in Step IV with the critical value k at the required level of significance α.

Conclusion : If $|Z| < k$, then accept the null hypothesis.

If $|Z| > k$, then reject the null hypothesis.

Tests of significance for Large Samples

We discuss the test of significance for large samples. (For practical purposes a sample may be regarded as large if $n > 30$.)

I. Tests for proportion or percentage.

(A) Single proportion (B) Difference of proportions.

II. Tests for means.

(A) (i) Test for single mean if standard deviation of the population is known. (i.e.) $H_0 : \mu = \mu_0$, σ is known.

(ii) Test for single mean if σ is not known $H_0 : \mu = \mu_0$, σ is unknown.

(B) (i) Test for equality of means of 2 normal populations with known standard deviations. (i.e., $H_0 : \mu_1 = \mu_2$; σ_1, σ_2 given $\sigma_1 = \sigma = \sigma_2$.

III. Tests for standard deviations.

(A) Test for single standard deviation $H_0 : \sigma = \sigma_0$.

(B) Test for equality of 2 standard deviations (i.e.) $H_0 : \sigma_1 = \sigma_2$.

Example 1. The mean and standard deviation of weight of boys of a college are 47 kgs. and 3.1 kgs. respectively. The mean and standard deviation of weight of girls of the college are 45 kgs. and 2.8 kgs. From the college, 16 boys and 9 girls are randomly selected.

i) Find the mean and standard deviation (S.E.) of mean weight of the 16 selected boys.

ii) Find the mean and standard deviation of mean weight of the 9 selected girls.

iii) Find the mean and standard deviation of the difference of the mean weight of the selected boys and the mean weight of the selected girls.

Solution : Here, $\mu_1 = 47$ kgs. $\sigma_1 = 3.1$ kgs., $n_1 = 16$,

$\mu_2 = 45$ kgs., $\sigma_2 = 2.8$ kgs. and $n_2 = 9$

Let $\bar{x}_1$ be the mean weight of the selected boys.

Let $\bar{x}_2$ be the mean weight of the selected girls.

i) Mean of $\bar{x}_1 = \mu_1 = 47$ kgs.

$$S.E(\bar{x}_1) = \frac{\sigma_1}{\sqrt{n_1}} = \frac{3.1}{\sqrt{16}} = 0.775 \text{ kgs.}$$

ii) Mean of $\bar{x}_2 = \mu_2 = 45$ kgs.

$$S.E(\bar{x}_2) = \frac{\sigma_2}{\sqrt{n_2}} = \frac{2.8}{\sqrt{9}} = 0.933 \text{ kgs.}$$

iii) Mean of $(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$

$$= 47 - 45 = 2 \text{ kgs.}$$

$$S.E.(x_1 - x_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$= \sqrt{\frac{(3.1)^2}{16} + \frac{(2.8)^2}{9}}$$

$$= 1.213 \text{ kgs.}$$

Example 2. A coin is tossed 144 times and a person gets 80 heads. Can we say that the coin is unbiased one?

Solution : Set the null hypothesis H_0 : the coin is unbiased.

Given $n = 144$. Probability of getting a head in a toll $P = \frac{1}{2}$

Hence $Q = \frac{1}{2}$

Let X = number of successes = number of getting heads = 80.

$$\therefore \quad Z = \frac{80-\left(144/2\right)}{\sqrt{144.\frac{1}{2}.\frac{1}{2}}} = \frac{80-72}{\sqrt{36}} = \frac{8}{6} = 1.33 < 1.96$$

Since $|Z| < 1.96$ we accept the hypothesis at 5% level of significance. Hence the coin is unbiased.

Example 3. In city A, 32% voters voted for Congress. In city B, 29% voters voted for Congress.

i) Among 70 randomly selected voters from city A, if p_1 is the proportion of voters who voted for Congress, find the standard error of p_1.

ii) Among 60 randomly selected voters from city B, if p_2 is the propor-tion of voters who voted for Congress, find the standard error of p_2.

iii) Find the mean and standard error of $(p_1 - p_2)$.

Solution :

Here, $P_1 = \frac{32}{100} = 0.32$ and $P_2 = \frac{29}{100} = 0.29$

$n_1 = 70$ and $n_2 = 60$

i) $$S.E(p_1) = \sqrt{\frac{P_1 Q_1}{n_1}}$$

$$= \sqrt{\frac{0.32 \times 0.68}{70}}$$

$$= 0.05575$$

ii) $$S.E.(p_2) = \sqrt{\frac{P_2 Q_2}{n_2}}$$

$$= \sqrt{\frac{0.29 \times 0.71}{60}}$$

$$= 0.05858$$

iii) $$S.E.(p_1 - p_2) = \sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$$

$$= \sqrt{\frac{0.32 \times 0.68}{70} + \frac{0.29 \times 0.71}{60}}$$

$$= 0.08087$$

Example 4. A coin is tossed 800 times and a person gets 350 heads. Can we say that he has made a random tossing each time. (equivalently can we say that the coin is unbiased one.)

Solution : Let the null hypothesis H_0 : random tossing is made.

Given $n = 800$. Probability of getting head in a toss $P = \frac{1}{2}, Q = \frac{1}{2}$.

Let X = number of getting heads = 350.

$$Z = \frac{350 - 800(\frac{1}{2})}{\sqrt{800(\frac{1}{4})}}$$

$$= \frac{350 - 400}{\sqrt{200}} = \frac{-50}{10\sqrt{2}} = \frac{-5\sqrt{2}}{2}$$

$$= -3.535.$$

$\therefore$ $|Z| > 3$. Hence H_0 is rejected and conclude that the coin is not randomly tossed. (equivalently the coin may not be unbiased).

Example 5. The poportion of women in a society is 0.48. Among 64 randomly selected people of the society, let p_1 be the proportion of women. In another selection of 86 people, let p_2 be the proportion of women. Find –

i) Standard error of p_1

ii) Standard error of p_2

iii) Standard error of the difference $(p_1 - p_2)$

Solution :

Here, $P_1 = P_2 = 0.48 = P$ (say)

$n_1 = 64$ and $n_2 = 86$

i) $$S.E.(p_1) = \sqrt{\frac{PQ}{n_1}}$$

$$= \sqrt{\frac{0.48 \times 0.52}{64}}$$

$$= 0.06245$$

ii) $$S.E.(p_2) = \sqrt{\frac{PQ}{n_2}}$$

$$= \sqrt{\frac{0.48 \times 0.52}{86}}$$

$$= 0.05387$$

iii) $$S.E.(p_1 - p_2) = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{0.48 \times 0.52\left(\frac{1}{64} \times \frac{1}{86}\right)}$$

$$= \sqrt{\frac{0.48 \times 0.52 \times (86 + 64)}{64 \times 86}}$$

$$= 0.08248$$

Example 6. It is required test the hypothesis that on an average Punjabis are 180 cms. tall. For this, a random sample containing 50 Punjabis is considered. The mean and standard deviation of heights of these are found to be 178.9 cms. and 3.3 cms. Based on this data, that would you conclude? (Use 5% level of significance).

Solution: Here, $\mu = 180$ cms, $n = 50$, and $s = 3.3$ cms.

H_0: Average height of Punjabis is 180 cms.

H_1: Average height of Punjabis differs from 180 cms.

The test statistic is

$$|Z| = \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}}$$

$$= \frac{|178.9 - 180|}{3.3/\sqrt{50}} = 2.36$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$.

Since $|Z|_{cal} > 1.96$, H_0 is rejected.

Conclusion: Average height of Punjabis differs from 180 cms

Example 7. A machine is designed so as to fill bottles with 200 ml. of a medicine. A sample of 100 bottles when measured, had a mean content of 201.3 ml. If the standard deviation of the fillings is known to be 5ml. test whether the machine is functioning properly. Use 5% level of significance.

Solution: Here $\mu_0 = 200$ ml., $\sigma = 5$ ml, $n = 100$ and $\bar{x} = 201.3$ ml.

H_0: The machine is functioning properly.
(The machine on an average fills 200ml. medicine.)

H_1: The machine is not functioning properly.

The test statistic is –

$$|Z| = \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}}$$

$$= \frac{1201.3 - 2001}{5/\sqrt{100}} = 2.6$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$. H_0 is rejected.

Conclusion: The machine is not functioning properly. It is not filling on an average 200ml. medicine.

Example 8. A random variable has variance 81. Its mean is not known. 144 random observations of the variable have mean 112.6. Test the hypothesis that the mean of variable is 114.5. Use level of significance a = 0.01 and also a = 0.05.

Solution : Here, $\mu_0 = 114.5$ $\sigma = \sqrt{81} = 9$, $n = 144$ and $\bar{x} = 112.6$

H_0 : The mean of the variable is 114.5

H_1 : The mean of the variable differs from 114.5

The test statistic is

$$|Z| = \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}}$$

$$= \frac{|112.6 - 114.5|}{9 / \sqrt{144}} = 2.53$$

i) The level of significance is $\alpha = 0.01$

The critical value is $k = 2.58$

Since $|Z|_{cal} < 2.58$, H_0 is accepted.

ii) The level of significance is $\alpha = 0.05$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$, H_0 is rejected.

Conclusion : At 1% level of significance, we conclude that the mean of the variable is 114.5 But, at 5% level of significance, we conclude that the mean differs from 114.5

Remark : A test at 5% level of significance is more powerful than a test at 1% level of significance.

Example 9. A die was thrown 9000 times and of these 3220 yielded a 3 or 4. Is this consistent with the hypothesis that the die was unbiased.

Solution : Given $n = 9000$.

P = probability of getting 3 or 4 in a throw of a die $= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

$$\therefore \quad Q = \frac{2}{3}$$

The number of successes $X = 3220$.

Set the null hypothesis H_0 : The die is unbiased.

Under the null hypothesis the test statistic is

$$|Z| = \frac{3220 - 9000 \times \left(\frac{1}{3}\right)}{\sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}}$$

$$= \frac{3220 - 3000}{\sqrt{2000}} = \frac{220}{\sqrt{2000}}$$

$$= \frac{220}{44.72} = 4.92$$

Since $|Z| = 4.92 > 3$, H_0 is rejected and we conclude that almost certainly the die is biased.

Example 10. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?

Solution : Here $n = 600$ and $p = \frac{325}{600} = .5417$.

Set the null hypothesis H_0 : The number of smokers and non smokers are equal in the city.

$\therefore$ P = population proportion of smokers in the city $= 0.5$ and $Q = 0.5$.

Alternative hypothesis H_1: $P > 0.5$.(right tailed test). Under H_0 we have

$$|Z| = \frac{.5417 - .5}{\sqrt{\frac{.5 \times .5}{600}}} = \frac{.0417}{.0204} = 2.04.$$

We note $Z = 2.04 > 1.645$ (under 5% level of significance for right tailed test). Hence the difference is significant and H_0 is rejected at 5% level and H_1 is accepted. Hence we conclude that the majority of men in the city are smokers.

$Z = 2.04 < 2.33$ (under 1% level of significance for right tailed test). Hence the difference is not significant at 1% level of significance. Hence H_0 may be accepted at 1% level of significance.

Note : We cannot come to solid conclusion from this random sample data. We may try for some other sample data for deciding whether majority of men try for some other sample data for deciding whether majority of men in that city are smokers.

Example 11. A manufacturer claimed that at least 95% of the equipment which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipments revealed that 18 were faulty. Test his claim at a significance level of (i) 5% (ii) 1%.

Solution : Out of sample of 200 equipments 18 were faulty.

$\therefore$ X = number of pieces conforming to specifications in the samples = 200 − 18 = 182.

p = sample proportion conforming to specifications = $\frac{182}{200} = .91$

Set the null hypothesis $H_0 : P = .95$ (the proportion of pieces conforming to specification in the population is 95%). Hence $Q = .05$.

Alternative hypothesis $H_1 : P < .95$ (at least 95% conformed to the specification – left tailed alternative).

$$\text{Test statistic } |Z| = \frac{.91 - .95}{\sqrt{\frac{.95 \times .05}{200}}} = \frac{.04}{.0154} = -2.6.$$

$\therefore$ Z is significant as it lies in the critical region. Hence we reject the null hypothesis at 5% level of significance. and

H_0 is rejected at 1% level also.

Example 12. From a population with mean 836, a random sample containing 225 observations is taken. The sample had mean 840.5 and standard deviation 45. Test whether the sample mean differs significantly from the population mean.

Solution : Here, $\mu_0 = 836$, $n = 225$, $\bar{x} = 840.5$ and $s = 45$.

H_0 = The sample mean does not differ significantly from the population mean.

H_1 = The sample mean differs from the population mean.

The test statistic is

$$|Z| = \frac{|\bar{x} - \mu_0|}{s / \sqrt{n}}$$

$$= \frac{|840.5 - 836|}{45 / \sqrt{225}} = 1.5$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is accepted.

Conclusion : The sample mean does not differ significantly from the population mean.

Example 13. Given that on the average 4% of insured men of age 65 die within a year and that 60 of a particular group of 1000 such men (age 65) died within a year. Can this group be regarded as a representative sample?

Solution : Here $n = 1000$. $p = \frac{60}{1000} = .06$

P = proportion of deaths in the population = .04. Hence Q = 96.

Let $H_0 : p = P$. Then $= \frac{.06 - .04}{\sqrt{\frac{.04 \times .96}{1000}}} = \frac{.02}{.0062} = 3.23.$

Here $Z > 3$. Hence H_0 is rejected and hence we can conclude that the group chosen is not a representative sample.

Example 14. It is known that IQ boys has standard deviation 10 and that IQ of girls has standard deviation 12. Mean IQ of 200 randomly selected boys is 99 and mean IQ of 300 randomly selected girls is 97. Can it be concluded that on an average boys and girls have the same IQ? (Use 1% level of significance.)

Solution : Here, $\sigma_1 = 10$ $n_1 = 200$ $\bar{x}_1 = 99$

$\sigma_2 = 12$ $n_2 = 300$ $\bar{x}_2 = 97$

H_0 = boys and girls have the same IQ.

H_1 = Boys and girls have different IQ.

The test statistic is —

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

Here, its calculated value is —

$$|Z| = \frac{|99 - 97|}{\sqrt{\frac{100}{200} + \frac{144}{300}}} = 2.02$$

The level of significance is $\alpha = 1\%$

The critical value is $k = 2.58$

Since $|Z|_{cal} < 2.58$, H_0 is accepted.

Conclusion : Boys and girls have tghe same IQ.

Example 15. A firm manufactures resistors which are known to have resistance with standard deviation 0.02 ohms. A random sample of 64 resistors had mean resistance 1.39 ohms. Can we conclude that the mean resistnace of the resistors manufactured by the firm is 1.4 ohms?

Solution : Here, $\mu_0 = 1.4$ ohms, $\sigma = 0.02$ ohms, $n = 64$ and

H_0 = The mean resistance of the resistors is 1.4 phms.

H_1 : The mean resistance of the resistors is not equal to 1.4 ohms.

The test statistic is —

$$|Z| = \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}}$$

$$= \frac{11.39 - 1.41}{0.02 / \sqrt{64}} = 4$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$, H_o is rejected.

Conclusion : The mean resistance of the resistors is not equal to 1.4 ohms.

Note : In this question, the level of significance is not mentioned. In such situations, it is customary to take it as 5%.

Example 16. A random sample of 500 apples was taken from a large consignment of apples and 65 were found to be bad. Find the S.E. of the population of bad ones in a sample of this size. Hence deduce that the percentage of bad apples in the consignment almost certainly lies between 8.5 and 17.5

Solution : Given $n = 500$. Let X = number of bad apples in the sample = 65.

p = proportion of bad apples in the sample $= \frac{65}{500} = .13$

Hence $q = .87$

S.E. of proportion $= \sqrt{\frac{.13 \times .87}{500}} = .015.$

Limits for proportions of bad apples in the population is

$$P \pm 3 \text{ S.E.} = .13 \pm 3(.015) = .13 \pm .045$$

$$= .175 \text{ and } .085 = 17.5\% \text{ and } 8.5\%$$

Example 17. A manufacturer claims that only 4% of his products supplied by him are defective. A random sample of 600 products contained 36 defectives. Test the claim of the manufacturer.

Solution : p = sample propertion of defectives $= \dfrac{36}{600} = .06.$

P = the proportion of defectives in the population $= \dfrac{4}{100} = .04$ so that $Q = 1 - .04 = .96.$

Set $H_0 : P = 0.4$ is true. The test statistic is

$$Z = \frac{\dfrac{(0.06 - 0.04)}{\sqrt{666}}}{\sqrt{0.04 \times .96}} = \frac{0.02 \times 24.49}{0.2} = 2.45$$

Thus $Z = 2.45$

If we set the alternative hypothesis $H_1 : P \neq 0.04$ we have to use two tailed test.

$\therefore \mid Z \mid = 2.5 > 1.96$. Hence we reject the hypothesis at 5% level of significance.

If we set $H_1 : P > 0.04$ we have to apply right tailed test.

$\therefore \mid Z \mid = 2.5 > 1.645$. Here also we reject the hypothesis at 5% level. Thus in either case the manufacturer's claim cannot be accepted.

Example 18. A sample of 900 days is taken from meteorological records of a certain district and 100 of them are found to be foggy. What are the probable limits, in percentage, of foggy days in the district?

Solution : Given $n = 900$. Let X = number of foggy days in the meteorological records = 100.

p = proportion of foggy days in the entire population of meteorology records P is not known. Hence we can take $P = p = \dfrac{1}{9}$ and $Q = q = \dfrac{8}{9}$ and $N = n = 900.$

$\therefore$ S.E. of proportion of foggy days $= \sqrt{\dfrac{1/9 \times 8/9}{900}} = .0105.$

Limits for the proportion of foggy days in percentage in the population.

$$= P \pm 3 \text{ S.E.} = \frac{1}{9} \pm 3\,(.0105) = .1111 \pm .0315 = .1426;\ .0796.$$

Hence the percentage of foggy days lies between 14.26 and 8.

Example 19. A study of systolic blood pressure of a randomly selected groupo of 36 patients suffering from a disease and another group of 36 persons who do not suffer from the disease gave the following results.

	Suffering	Not suffering
Sample size :	36	36
Mean systolic pressure :	178	161
Standard deviation :	24	12

Test whether the average systolic pressure of the persons of the two groups differ significantly.

Solution : Here, $n_1 = 36$ $\quad \bar{x}_1 = 178$ $\quad s_1 = 24$

$n_2 = 36$ $\quad \bar{x}_2 = 161$ $\quad s_2 = 12$

H_0 : Average systolic pressure of the two groups of persons do not differ significantly.

H_1 : Average systolic pressure of the two groups of differ significantly.

The test statistic is

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

$$= \frac{|178-161|}{\sqrt{\frac{24^2}{36}+\frac{12^2}{36}}} = 3.80$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is rejected.

Conclusion : Average systolic pressure of the two groups differ significantly.

Example 20. During a country wide investigation the incidence of TB was found to be 1%. In a college of 400 strength 5 were reported to be affected whereas in another college of 1200 strength 10 were reported to be affected. Does this indicate any significant difference?

Solution : Given $P = .01$ Hence $Q = .99$.

$$n_1 = 400,\ p_1 = \frac{5}{400} = \frac{1}{80} = .0125.$$

$$n_2 = 1200,\ p_2 = \frac{10}{1200} = \frac{1}{120} = .0083.$$

Set the null hypothesis $H_0 : P_1 = P_2$.

Test statistic $$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0.1)$$

$$Z = \frac{.0125 - .0083}{\sqrt{.01 \times .99\left(\frac{1}{400} \times \frac{1}{1200}\right)}}$$

$$= \frac{.0042}{\sqrt{.01 \times .99 \left(\frac{1}{300}\right)}} = .7368.$$

$\therefore\ Z < 1.96$

Hence the difference is not significant on 5% level of significance.

Example 21. 500 articles from a factory are examined and found to be 2% defective. 800 similar articles from a second factory are found to have only 1.5% defectives. Can it be reasonable concluded that the products of the first factory are inferior to those of the second.

Solution : Given $n_1 = 500$; $n_2 = 800$.

Proportion of defectives in the first factory $p_1 = \frac{2}{100} = .02$

Proportion of defectives in the second factory $p_2 = \frac{1.5}{100} = .015$

Since proportion P of the population is not given it can be estimated as

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{10+12}{1300} = 0.017$$

$\therefore\quad Q = 0.983.$

Set the null hypothesis $H_0 : P_1 = P_2$ (products do not differ in equality; (i.e.) two factories are producing similar products.) Under this null hypothesis the test statistics.

$$|Z| = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.02 - 0.015}{\sqrt{0.017 \times 0.983 \left(\frac{1}{500} + \frac{1}{800}\right)}}$$

$$= \frac{0.005}{\sqrt{0.017(0.0020+0.0013)}}$$

$$= \frac{0.005}{\sqrt{0.017 \times 0.0033}}$$

$$= \frac{0.005}{0.0075} = 0.67 < 1.96$$

The difference of proportions is not significant on 5% level. Hence the hypothesis is accepted.

$\therefore$ The factories are producing similar products which do not differ in quality. Hence we conclude that one is not inferior to the other.

Exercise 22: Intelligence test on two groups of boys and girls gave the following results.

	Mean marks	Standard deviation	Sample size
Girls	75	15	150
Boys	70	20	250

Is there significant difference in the mean marks obtained by the boys and the girls? Test at 1% level of significance.

Solution: Here, $n_1 = 150$ $s_1 = 15$

$n_2 = 250$ $s_2 = 20$

H_0: The difference is mean marks is not significant.

H_1: The difference in mean marks is significant.

The test statistic is

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

$$|Z|_{cal} = \frac{|75-70|}{\sqrt{\frac{(15)^2}{150}+\frac{(20)^2}{250}}} = 2.84$$

The level of significance is $\alpha = 1\%$

The critical value is $k = 2.58$

Since $|Z|_{cal} > 2.58$. H_0 is rejected

Conclusion: The difference in mean marks of boys and girls is significant.

Exercise 23. Doordarshan anticipated 65% viewer ship for the recent India Vs Pakistan cricket match. Among 300 viewers of DD who where contacted. 217 had viewed the match. Does this figure support Doordarshan's anticipated viewership?

Solution: Here, $P_0 = \frac{65}{100} = 0.65$ $\quad n = 300$ and $x = 217$

Therefore, $p = \frac{x}{n} = \frac{217}{300} = 0.723$

H_0: The viewership for the cricket match is 65%.

H_1: The viewership for the cricket match differs from 65%

The test statistic is –

$$|Z| = \frac{|p-P_0|}{\sqrt{\frac{P_0\,Q_0}{n}}}$$

$$|Z|_{cal} = \frac{|0.723-0.65|}{\sqrt{\frac{0.7\times 0.3}{300}}} = 2.759$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$, H_0 is rejected.

Conclusion: The viewership for cricket match differs from 65%

Example 24. A sample of 1000 products from a factory are examined and found to be 2.5% defective. Another sample of 1500 similar products from another factory are found to have only 2% defective. Can we conclude that the products of the first factory are inferior to those of the second?

Solution : Given $n_1 = 1000$, $n_2 = 1500$.

Proportion of defectives in the first factory $p_1 = \dfrac{25}{1000} = .025.$

Proportion of defectives in the second factory $p_2 = \dfrac{30}{1500} = .02$

Since the proportion of the population is not given it can be estimated as

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{25 + 30}{1000 + 1500} = 0.22.$$

Hence $Q = 1 - P = 1 - .022 = .978.$

Null hypothesis $H_0 : P_1 = P_2$ (two factories are producing similar products).

$$\text{Test hypothesis } Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0.1)$$

$$Z = \frac{.025 - .020}{\sqrt{0.22 \times .978 \times \left(\frac{1}{1000} + \frac{1}{1500}\right)}}$$

$$= \frac{.005}{\sqrt{\frac{.022 \times .978}{600}}} = .83 < 1.96$$

$\therefore$ The difference of proportion is not significant on 5% level. Hence this hypothesis is accepted and the two factories are producing similar products. Hence one is not inferior to the other.

Example 25. A machine puts out 16 imperfect articles in a sample of 500 articles. After the machine is overhauled it puts out 3 defective articles in a sample of 100. Has the machine improved?

Solution : Given $n_1 = 500, n_2 = 100.$

$$p_1 = \text{proportion of defectives in the first sample} = \frac{16}{500} = .032$$

$$p_2 = \text{proportion of defectives in the second sample} = \frac{3}{100} = .03$$

Set the null hypothesis $H_0 : P_1 = P_2$ (there is no significant difference in the machine before and after overhauling. (i.e.) the machine has not improved after overhauling.) Alternative hypothesis is $H_1 : P_1 > P_2$ (right tailed alternative).

Since the proportion P of the population is not known we can estimate it as

$$P = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{16+3}{500+100} = \frac{19}{600} = .032;$$

$$Q = 1 - P = 1 - .032 = .968.$$

The test statistic $\quad Z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

$$= \frac{.032 - .03}{\sqrt{.032 \times .968\left(\dfrac{1}{500} + \dfrac{1}{100}\right)}}$$

$$= \frac{.002}{\sqrt{.032 \times \frac{6}{500}}} = \frac{.002}{.019} = .105.$$

Since $Z < 1.645$ (right tailed test) it is not significant at 5% level of significance. Hence, we can accept the null hypothesis and conclude that the machine has not been improved.

Example 26. Certain dose of an analgesic (pain reliever) when administered to each of 32 women patients, the average duration of pain relief was 3.5 hours. The same dose when administered to 36 men patients, the average duration of pain relief was 4 hours. By past experience, if it is known that the standard deviation of duration of pain relief is 0.5 hours, test whether on an average the duration of pain relief is the same among men and women.

Solution: Here, $n_1 = 32$ $\bar{x}_1 = 3.5$

$n_2 = 36$ $\sigma_1 = \sigma_2 = 0.5$

H_0: The average duration of pain relief is the same among men and women.

H_1: The average duration of pain relief is not the same among man and women.

The test is –

$$|Z| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$|Z|_{cal} = \frac{|3.5 - 4|}{\sqrt{\frac{(0.5)^2}{32} + \frac{(0.5)^2}{36}}} = 33.88$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$, H_0 is rejected.

Conclusion: The average duration of pain relief is not the same among men and women.

Example 27: A survey of 225 randomly selected students of I BCA from all over Bangalore University revealed that 89.4% of them used the book 'Statistics by Srimani + Vinayakamoorthy' as text. Can we conclude at 1% level of significance that the book is being used by 90% of the students.

Solution: Here, $P_0 = \dfrac{90}{100} = 0.9$ $p = \dfrac{89.4}{100} = 0.894$ and $n = 225$

H_0: The book is being used 90% of the students.

H_1: The percentage of students using the book differs from 90%..

The test statistic is $|Z| = \dfrac{|p - P_0|}{\sqrt{\dfrac{P_0 Q_0}{n}}}$

$$= \frac{|0.894 - 09|}{\sqrt{\dfrac{0.9 \times 0.1}{225}}} = 0.3$$

The level of significance is $\alpha = 1\%$

The critical value is $k = 2.58$

Since $|Z|_{cal} < 2.58$, H_0 is accepted.

Conclusion: The book is being used by 905 of the students,

Example 28. Manufacturer of VM–cars is of the opinion that 35% of the cars plying on the roads of Bangalore are VM–cars. A person decided to verify this by conducting a survey. He counted the cars plying on K.G. Road between 4 p.m. and 4.10 p.m. and noted the make. Among 347 cars which he counted 107 were VM–cars. Does this figure support the manufacturer's claim?

Solution : Here, $P_0 = \frac{35}{100} = 0.35$ $n = 347$ and $x = 107$

Therefore, $p = \frac{x}{n} = \frac{107}{347} = 0.308$

H_0 : The proportion of VM–cars is 0.35.

H_1 : The proportion of VM–cars differs from 0.35.

The test statistic is —

$$|Z| = \frac{|p - P_0|}{\sqrt{\frac{P_0 Q_0}{n}}}$$

$$= \frac{|0.308 - 0.35|}{\sqrt{\frac{0.35 \times 0.65}{347}}} = 1.64$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, is accepted.

Conclusion : The proportion of VM-cars is 0.35.

Example 29. The manufacturers of Brand R pens contend that the proportion of college students of Bangalore who use Band R pens is 0.3. In order to test this contention, 40 students were randomly picked and questioned in this regard. Among these 40 students, 10 were found to use Brand R pens. At 0.05 level of significance, test whether the manufacturers' contention is acceptable.

Solution : Here, $P_0 = 0.3$, $n = 40$, $x = 10$

Therefore, the sample proportion is $p = \frac{x}{n} = \frac{10}{40} = 0.25$

H_0 = The proportion of users of Brand R pens is 0.3

H_1 : The proportion of users of Brand R pens differs from 0.3.

The test statistic is —

$$|Z| = \frac{|p - P_0|}{\sqrt{\frac{P_0 Q_0}{n}}}$$

$$= \frac{|0.25 - 0.3|}{\sqrt{\frac{0.3 \times 0.7}{40}}} = 0.69$$

The level of significance is $\alpha = 0.05$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is accepted.

Conclusion : The proportion of users of Brand R pens is 0.3.

Example 30. In a year there are 956 births in a town A of which 52.5% were males while in town A and B combined this proportion in total of 1406 births was .496. Is there any significant difference in the proportion of male births in the two towns?

Solution : Given $n_1 = 956$. p_1 = proportion of male birth = .525. p_2 and n_2 are not given; but $n_1 + n_2 = 1406$ so that $n_2 = 450$.

Given $P = .496$ and $Q = 1 - .496 = .504$.

We have $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \Rightarrow .496 = \frac{956 \times .525 + 450 \times p_2}{1406}$

$\Rightarrow p_2 = .434$ (verify).

Set the null hypothesis $H_0 = P_1 = P_2$ (no difference of births of males in the towns).

$$\therefore \quad Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{.525 - .434}{\sqrt{.496 \times .504\left(\frac{1}{956} + \frac{1}{450}\right)}} = 3.14$$

$Z > 3$. Hence there is significant difference in the proportion of male births in the towns A and B.

Example 31. It is required to verify whether a coin is biased. The coin is tossed 32 times and the results are noted, 19 of the 32 tosses resulted in the occurrence of head. Can we conclude that the coin is biased?

Solution: We know that the probability of head for an unbiased coin is 0.5. Therefore, we test whether $P = 0.5$

Thus, $P_0 = 0.5$ $n = 32$ and $x = 19$

Therefore, $p = \frac{x}{n} = \frac{19}{32} = 0.5938$

H_0: The coin is unbiased. That is, $P = 0.5$

H_1: The coin is biased.

The test statistic is –

$$|Z| = \frac{|p - P_0|}{\sqrt{\frac{P_0 Q_0}{n}}}$$

$$= \frac{|0.5938 - 0.5|}{\sqrt{\frac{0.5 \times 0.5}{32}}} = 1.06$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is accepted.

Conclusion: The coin is unbiased.

Example 32. It is required to test whether a coin is baised.

i) Suppose the coin is tossed 40 times and the proportion of tosses resulting in head is 0.4, what is the conclusion?

ii) Suppose the coin is tossed 100 times and the proportion of tosses resulting in head is 0.4, what is the conclussion?

Solution :

i) Thus, $n = 40$, $p = 0.4$ and $P_0 = 0.5$

H_0: The coin is unbiased. $(P = 0.5)$

H_1: The coin is biased. $(P \neq 0.5)$

The test statistic is –

$$|Z| = \frac{|p - P_0|}{\sqrt{\frac{P_0 Q_0}{n}}}$$

$$|Z|_{cal} = \frac{|0.4 - 0.5|}{\sqrt{\frac{0.5 \times 0.5}{40}}} = 1.2649$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is accepted.

ii) Here, $n = 100$, $p = 0.4$ and $P_0 = 0.5$

$$\therefore \quad |Z|_{cal} = \frac{|0.4 - 0.5|}{\sqrt{\frac{0.5 \times 0.5}{100}}} = 2$$

Since $|Z|_{cal} < 1.96$, H_0 is rejected.

Conclusion: The coin is unbiased.

Example 33. According to probability theory, the probability that in a family which ias two children, both the children are sons is 0.25. In a locality, among 136 families which have 2 children each, 46 families have 2 sons. Does this information subscribe to the theory?

Solution : Here, $P_0 = 0.25$ $n = 136$ and $x = 46$

Therefore, $$p = \frac{x}{n} = \frac{46}{136} = 0.3382$$

H_0: The coin is unbiased. ($P = 0.5$)

H_1: The coin is biased. ($P \neq 0.5$)

The test statistic is –

$$|Z| = \frac{|p - P_0|}{\sqrt{\frac{P_0 Q_0}{n}}}$$

$$= \frac{|0.3382 - 0.25|}{\sqrt{\frac{0.25 \times 0.75}{136}}} = 2.375$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is rejected.

Conclusion: The proportion of families with 2 sons differs from 0.25.

Example 34. From the following data, test whether the difference between the proportions in the two samples is significant.

	Size	Proportion
Sample I	1000	0.02
Sample II	1200	0.01

Solution : Here, $n_1 = 1000,$ $p_1 = 0.02$

$n_2 = 1200,$ $p_2 = 0.01$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{1000 \times 0.02 + 1200 \times 0.01}{1000 + 1200} = 0.0146$$

H_0 : The proportions do not differ significantly.

H_1 : The proportions differ significantly.

The test statistic is —

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{0.02 - 0.01}{\sqrt{0.0146 \times 0.9854\left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 1.9471$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is accepted.

Conclusion : The proportions do not differ significanctly.

Example 35. In a random sample consisting of 60 students of a college, 2 were smokers. In a random sample consisting of 17 lecturers of the college, 5 were smokers. Test at 5% level of significance that the proportion of smokers among the students is the same as that among lecturers.

Solution :

Here, $n_1 = 60,$ $\bar{x}_1 = 2,$ $p_1 = \frac{x_1}{n_1} = \frac{2}{60} = 0.0333$

$n_2 = 17$, $\quad \bar{x}_2 = 5$, $\quad p_2 = \frac{x_2}{n_2} = \frac{5}{17} = 0.2941$

H_0 : The proportion of smokers among students is the same as the proportion among lecturers.

H_1 : The proportion of smokers among students differs from the proportion among lecturers.

The test statistic is —

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{2+5}{60+17} = 0.0909$$

$$|Z|_{cal} = \frac{|0.0333 - 0.2941|}{\sqrt{0.0909 \times 0.9091\left(\frac{1}{60} + \frac{1}{17}\right)}}$$

$$= \frac{0.2608}{\sqrt{\frac{0.0909 \times 0.9091 \times (17+60)}{60 \times 17}}} = 3.302$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$, H_0 is rejected.

Conclusion : The proportion of smokers among students differs from the proportion among lecturers.

Example 36. Among 326 scooters which crossed Mysore Bank junction at Bangalore during a period of one hour, 143 were Bank B scooters

Among 213 scooters which crossed Shivaji Statue junction at Poona during a period of one hour, 137 were Brand B scooters. Test whether the proportion of Brand B scooters on the roads of Bangalore differs from the proportion of Poona.

Solution:

Here, $n_1 = 326,\quad \bar{x}_1 = 143,\quad p_1 = \frac{x_1}{n_1} = \frac{143}{326} = 0.4387$

$n_2 = 213,\quad \bar{x}_2 = 137,\quad p_2 = \frac{x_2}{n_2} = \frac{137}{213} = 0.6432$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{143 + 137}{326 + 213} = 0.5195$$

H_0 : The proportion of Brand B Scooters on the roads of Bangalore is the same as the proportion in Poona.

H_1: The proportion of Brand B Scooters on the roads of Bangalore differs from the proportion in Poona.

The test statistic is –

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$|Z|_{cal} = \frac{|0.4387 - 0.6432|}{\sqrt{0.5195 \times 0.4805\left(\frac{1}{326} + \frac{1}{213}\right)}} = 4.646$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} > 1.96$. H_0 is rejected.

Conclusion: The proportion of Brand *B* scooters on the roads of Bangalore differs from the proportion in Poona.

Example 37. A survey of 80 men and 80 women who were aged 50 or more was conducted. Among the 80 men, 8 had high B.P. Among the 80 women, 6 had high B.P. Test whether the proportion of men with high B.P. differs from the proportion of women with high B.P.

Solution: Here, $n_1 = 80, \quad \bar{x}_1 = 8, \quad p_1 = \frac{x_1}{n_1} = \frac{8}{80} = 0.1$

$$n_2 = 80, \quad \bar{x}_2 = 6, \quad p_2 = \frac{x_2}{n_2} = \frac{6}{80} = 0.075$$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{8+6}{80+80} = 0.0875$$

H_0: The proportion of persons with high B.P. is the same among men and women.

H_1: The proportion of persons with high B.P. is different among men and women.

The test statistic is –

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{|0.1 - 0.075|}{\sqrt{0.0875 \times 0.9125\left(\frac{1}{80} + \frac{1}{80}\right)}} = 0.5596$$

The level of significance is $\alpha = 5\%$

The critical value is $k = 1.96$

Since $|Z|_{cal} < 1.96$, H_0 is accepted.

Conclusion: The proportion of persons with high B.P. is the same among men and women.

Exercise 38. In batch A is PUC class of a college, out of 78 students, 67 passed in an examination. in batch B of PUC class of the same college out of 72 students, 69 passed in the examination. Can you conclude that the performance of the students of the two batches is the same? Use 1% level of significance

Solution: Here, $n_1 = 78$, $\bar{x}_1 = 67$, $p_1 = \frac{x_1}{n_1} = \frac{67}{78} = 0.859$

$n_2 = 72$, $\bar{x}_2 = 69$, $p_2 = \frac{x_2}{n_2} = \frac{69}{72} = 0.9583$

$$P = \frac{x_1 + x_2}{n_1 + n_2} = \frac{67+69}{78+72} = 0.9067$$

H_0: The performance of the students of the two batches is the same.

H_1: The performance of the students of the two batches are different.

The test statistic is –

$$|Z| = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{|0.859 - 0.9583|}{\sqrt{0.9067 \times 0.0933\left(\frac{1}{78} + \frac{1}{72}\right)}}$$

$$= \frac{0.0993}{\sqrt{\frac{0.9067 \times 0.0933 \times (72+78)}{78 \times 72}}} = 2.089$$

The level of significance is $\alpha = 1\%$

The critical value is $k = 2.58$

Since $|Z|_{cal} < 2.58$, H_0 is accepted.

Conclusion : The performacne of the students of the two batches is the same.

Remark : However, here, the value $|Z|_{cal} = 2.089 > 1.96$, Therefore, at 5% level H_0 will be rejected.

PROBLEMS FOR PRACTICE – I

1) What do you mean by sampling distribution of a statistic?

2) What is meant by Standard Error (S.E)?

3) Write down the mean and standard errors of-
   i) sample mean $\bar{x}$
   ii) difference of sample means $(\bar{x}_1 - \bar{x}_2)$
   iii) sample proportion p.
   iv) difference of sample proportions $(p_1 - p_2)$.

4) Explain
   i) Statistical hypothesis, ii) Null hypothesis.
   iii) Alternative hypothesis.

5) Explain the procedure of testing a hypothesis.

6) Define
   i) Test statistic ii) Critical region.
   iii) Errors of type I and type II iv) Level of significance
   v) Power of a test procedure vi) One-tailed and two-tailed tests.

7) Explain the procedure of testing whether a population has a given mean using a large sample.

8) Using large samples, how do you test equality of means of two populations?

9) Explain the procedure of testing whether a population has given proportion of units possessing an attribute.

10) Explain how equality of proportions in two populations are tested.

11) The standard deviation of weight of children of K.G. class is 2.2. kgs. What is the standard error of mean weight of 16 children ?

[Ans. 0.55 kgs]

12) A random sample of size 64 is drawn from a population whose variance is 25. Write down the standard error of the sample mean.

[Ans. 0.625]

13) Mean and standard deviation of hights of II BCA students are 158cms. and 3cms. respectively. Mean and standard deviation of heights of I BCA students are 152 cms. and 4 cms. respectively.

i) If, $\bar{x}_1$ is the mean height of 12 randomly selected II BCA students, write down the mean and stnadard error of $\bar{x}_1$.

ii) If is the mean height of 15 randomly selected I BCA students, write down the mean and standard error of $\bar{x}_2$.

iii) Also write down the mean and standard error of $(\bar{x}_1 - \bar{x}_2)$.

[158,0866,152, 1.0328, 6, 1.3478 cms.]

14) The proportion of vegetarians in a city is 0.46. Find the mean and standard error of the proportion of vegetarians in a random sample of size 14.

[0.1332]

15) There are 1700 students in a college. 595 of these students are girls.30 students are randomly selected. Find the standard error of proportion of girls in the selection.

[0.08708]

16) It is known that 15% students who pass P.U.C. go for college education. Among 24 students who passed P.U.C. from a school, what proportion of students is expected to go for college education. Find the standard error of the proportion. [0.15, 0.07289]

17) Suppose the U.P.S.C. Board results for the state is 33% pass. Find the mean and stnadard deviation of the proportion p_1 of passes in a school where 36 candidates appeared. If p_2 is the proportion of

passes in another school where 47 candidates appeared, find the standard error of p_2. Also, find the standard error of $(p_1 - p_2)$.

[0.33, 0.07837, 0.06859, 0.1041]

18) Among those in nursing servioce, the proportion of women is 0.85. If p_1 is the proportion in 13 mursing staff of Hospital A and p_2 is the proportion in 23 nursing staff of Hospital B, find (i) S.E. (p_1) (ii) S.E. (p_2) and (iii) S.E. $(p_1 - p_2)$.

[0.09903, 0.07445, 0.1239]

19) A machine is designed so as to fill plastic bags with 500 ml. milk. A sample of 400 bags when checked had an average content 498 ml. milk. The standard deviation was 8 ml. At 5% level of significance, test whether the machine is functioning properly.

20) A drill drills holes with standard deviation of depth 0.03 cms. It is adjusted to drill holes of depth 5.5 cms. For 50 holes drilled, the mean depth is 5.503 cms. Test whether the adjustment is correct.

[rejected]

21) The standard deviation of marks scored by BCA students in an examination is 10.5. The mean marks scored by 64 students of a college is 38. Can we conclude at 1% level of significance that mean marks of BCA students is 40?

22) A sample of 100 students is found to have height 64 inches. Can it be reasonably regarded as a sample from a large population with mean height 66 inches and standard deviation 4 inches? (Test at 1% and 5% levels of significance)

[Ans. 5, rejected at 1% and 5% levels of significance]

23) A random sample of 200 tins of vanaspathi has a mean weight 4.97 kgs. and standard deviation 0.2kgs. Test at 1% level of significance, that the tins have 5kgs. vanaspathi.

[Ans. 2.12 accepted]

24) A random sample of 100 rods drawn from a lot of rods had mean length 32.7 cms and standard deviation 1.3 cms. Can it be concluded that the lot has mean 32 cms.?

25) From population with mean 21.3 and standard deviation 0.4, a random sample of 625 observations is taken. If the sample mean is 21.33, test whether it differs significantly from the population mean. [1.88, accepted]

26) A random sample of size 60 from a population with unknown distribution has mean 103.4 and standard deviation 4. Test whether the population mean is 105.

27) 300 random observations of a variable have mean 81.3 and sgtandard deviation 12.4. Test whether we can take the mean of the variable to be 81. [Ans. 0.42 accepted]

28) A random variable has variance 64. Its mean is not known. 100 random observations of the variable have mean 147.3. Test whether the mean of the variable is 150.
[Ans. 3.38 accepted]

29) The mean and standard deviation of heights of 100 randomly selected boys are 163 cms and 3 cms, respectively. The mean and standard deviation of heights of 80 randomly selected girls are 161 cms and 2 cms, respectively. Can it be concluded at 1% level of significance that boys and girls are equally tall? [5.35, rejected]

30) The standard deviation of length of fibre manufactured by process A is 0.5 cms, and stnadard deviation of length of fibre manufactured by process B is 0.6 cms. A sample of 40 randomly selected fibres from process A has mean length 16.7 cms. A sample of 60 randomly selected fibres from process B has mean length 16.4 cms. Test whether process A and process B differ with regard to length of fibre manufactured by them. [2.71 rejected]

31) Samples of electric lamps manufactured by two firms gave the following results –

	Firm A	Firm B
Sample size :	80	60
Average buring hours :	1348	1270
Standard deviation :	108	96

Test whether the average life of bulbs manufactured by the firm are the same. [Ans. 4.52, rejected]

32) From the following data test whether the means differ significantly.

	I sample	II sample
Size :	300	200
Mean :	75.4	74.3
Variance :	65.6	57.8

[Ans. 1.54 accepted]

33) Past experience shows that among borewells dug by a firm. 78% are successful. The firm digs 65 borewells in a district. Among them 58 were successful. Can we conclude that these figures agree with the past experience? [Test both at 5% and 1% levels]

[Ans. 2.19, 5% rejected, 1% accepted]

34) There is a contention that birth to male child and birth to female child are equiprobable. To test this, sex of 460 new born babies were recorded. Among the 460 babies, 283 were male. Conduct the test at 5% level. [Ans. 4.94, rejected]

35) Among 80 potatoes which are randomly pulled out from the field, 11 are infested. Test the hypothesis that 25% potatoes are infested.

[Ans. 2.32, rejected]

36) There is a claim that a bombing systgem has 0.9 probability of hitting ground targets. To test the validity of this claim, 70 dummy bombs are dropped on a target. Among them 49 hit the target. What is your conclusion at 5% level of significance? What would be the conclusion if the level of significance is 1%?

[5.57, rejected at 5% as well as 1% levels.]

37) A survey of 1000 college girls revealed that 23% of them viewed cricket match. Two years hence, another survey of 800 college girls revealed that 24% of them viewed cricket match. Test the hypothesis that the proportion of college girls viewing cricket match has differed in the 2 years. [Ans. 0.5, accepted]

38) The proportion of substandard crackers among 400 crackers manufactured by firm A is 0.12. The proportion among 500

crackers manufactured by firm B is 0.08. Test at 1% level of significance that the proportion is the same among the products of the two firms.

[Ans. 2.01 rejected at 1%]

39) In a random selectioon of 85 workers of a facotry, 18 were unmarried. Can we conclude that 20% workers of the factory are unmarried?

40) Among 37 educated youths, 3 were smokers. Among 26 uneducated youths, 4 were smokers. Test whether the proportion of smokers is the same among the educated and among the uneducated.

41) Among 80 electric bulbs manufactured by Process A, 3 were defective. Among 130 electric bulbs manufactured by process B, 2 were defective. Test whether the proportion of defectives in the two process differ.

42) From the following data, test whether the difference between the proportions in the two samples is significant.

	Size	Proportion
Sample I	300	0.45
Sample II	800	0.49

II

1) A die is thrown 9000 times and a throw of 3 or 4 observed 3240 times. Show that the die cannot be regarded as an unbiased one.

2) A coin is tossed 10,000 times and it turns up head 5195 times. Discuss whether the coin may be regarded as unbiased one?

3) A coin is tossed 400 times and head turns up 216 times. Discuss whether the coin may be unbiased one.

4) A coin is tossed 960 times and it falls with head upwards 184 times. Is the coin biased?

5) 500 eggs are taken from a large consignment and 50 are found to be bad. Estimate the percentage of bad eggs in the consignment and assign the limits within which the percentage probably lies.

6) A manufacturer claimed that at least 95% of the equipments which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipments revealed that 18 were faulty. Test his claim at a significant level of (i) .05 and (ii) .01.

7) A social worker believes that fewer than 25% of the couples in a certain area are ever used any form of birth control. A random sample of 120 couples was contracted. Twenty of them said they had used some method of birth control. Comment on the social worker's belief.

8) A manufacturer of bulbs claims that on the average 2 percents or less of all the bulbs manufactured by his firm are defective. A random sample of 400 bulbs contained 13 defective bulbs. On the evidence of this sample do you support the manufacture's claim? Assume that the maximum risk you wish to run of falsely rejecting the manufacturer's claim has been set at 5%.

9) In a sample of 400 parts manufactured by a factory the number of defective parts was found to be 30. The company however claimed that only 5% of their product is defective. Is the claim acceptable?

10) 10 dice are thrown 300 times each and a throw of an odd number is reckoned as asuccess. Are the dice unbiased if 1650 throws of odd numbers have been received?

11) In a sample of 500 people in Karnataka 280 are tea drinkers, the next are coffee drinkers. Can we assume that both coffee and tea are equally popular in Karnataka at 1% level of significance?

12) In a sample of 1000 people in Tamil Nadu 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?

13) 20 people were attacked by a disease and only 18 survived. Will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favour of the hypothesis that is more at 5% level?

14) In a sample of 600 men from a certain large city 450 are found to be

smokers. In one of 900 from another city 450 are smokers. Do the data indicate that the cities are significantly different with respect to prevalence of smoking among men?

15) A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Obtain the 98% confidence limits for the percentage number of bad apples in the consignment.

16) A candidate is an election claims 90% of support of all voters in a locality. Verify his claim at 5% level of significance if in a random sample of 400 voters from the locality 320 supported his candidate.

Additional Problems on Test of Significance for Means

Example 1. A sample of 900 men is found to have a mean height of 64 inch. If this sample has been drawn from a normal population with S.D. 20 inch, find the 99% confidence limits for the mean height of the men in the population.

Solution : Given $n = 900$; $\bar{x} = 64$; $\sigma = 20$. The 99% confidence limits for the population mean μ is

$$\bar{x} - 2.58\left(\frac{\sigma}{\sqrt{n}}\right) \text{ and } \bar{x} + 2.58\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\Rightarrow \quad 64 - 2.58\left(\frac{20}{\sqrt{900}}\right) \text{ and } 64 + 2.58\left(\frac{20}{\sqrt{900}}\right)$$

$$\Rightarrow \quad 64 - \frac{51.6}{30} \text{ and } 64 + \frac{51.6}{30}$$

$$\Rightarrow \quad 64 - 1.72 \text{ and } 64 + 1.72. \qquad \Rightarrow 62.28 \text{ and } 65.72$$

$\therefore$ The 99% confidence limits for the mean height is (62.28, 65.72).

Scanned with OKEN Scanner

Example 2. An insurance agent has claimed that the average age of policy holders who insure through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had insure through him reveal that the mean and S.D. are 28.8 years and 6.35 years respectively. Test his claim at 5% level of significance.

Solution : Given $\mu = 30.5$; $n = 100$; $\bar{x} = 28.8$; $s = 6.35$

Set the null hypothesis $H_0 : \mu = 30.5$

The alternative hypothesis is $H_1 : \mu < 30.5$ (left tailed test).

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{28.8 - 30.5}{6.35/\sqrt{100}} = \frac{-1.7 \times 10}{6.35} = -2.68$$

$\therefore \quad |Z| = 2.68 > 1.96$

It is significant at 5% level of significance. Hence H_0 is rejected and H_1 is accepted at 5% level. Hence the claim of the agent is valid.

Example 3. A normal population has a mean of 6.48 and S.D. of 1.5. In a sample of 400 members mean is 6.75. Is the difference significant?

Solution : Given $\mu = 6.48$; $\sigma = 1.5$; $n = 400$; $\bar{x} = 6.75$.

Let $H_0 : \mu = \bar{x}$ Under the hypothesis the test statistic is

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{6.75 - 6.48}{1.5/\sqrt{400}}$$

$$= \frac{.27 \times 20}{1.5} = \frac{5.4}{1.5} = 3.6 > 1.96.$$

$\therefore$ The difference is significant.

Example 4. A sample of 400 individuals in found to have a mean of 67.47. Can it be reasonably regarded as a sample from a large population with mean 67.39 and S.D. 1.3?

Solution : Given $\mu = 67.39$; $\sigma = 1.3$; $n = 400$ and

Let H_0 = Under this H_0 the test statistic is

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{67.47 - 67.39}{1.3/\sqrt{400}} = \frac{.08 \times 20}{1.3} = 1.23$$

$$|Z| = 1.23 < 1.96.$$

$\therefore$ At 5% level of significance the difference is not significant. Hence H_0 is accepted at 5% level of significance. Hence we can conclude that the sample could have come from the large population with mean 67.39 and S.D. 1.3.

Example 5. An entrance examination test was given to a large group of boys and girls who appeared for the selection of professional course in Tamil Nadu who scored on the average 84.5 marks. The same test was given to a sample group of 400 boys and girls. They scored an average of 82.5 marks with a S.D. 22.5 marks. Examine whether the difference is significant. Can we conclude that the sample is from the population?

Solution : Given $\mu = 84.5$; $n = 400$; $= 82.5$ and $s = 22.5$. Here the S.D. of the population is not given. Let $H_0 : \mu$ Under this H_0 the test statistic is

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{82.5 - 84.5}{22.5/\sqrt{400}} = \frac{-2 \times 20}{22.5} = 1.778$$

$\therefore$ $|Z| = 1.778 < 1.96.$

$\therefore$ The difference is not significant on 5% level of significance. Hence we can conclude that the sample is from the population.

PROBLEMS FOR PRACTICE

1) A lady stenographer claims that she can take dictation at the rate of 120 words per minute. Can we reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 words with a S.D. of 15 words?

I Sem. B.C.A. – Prob...

2) A sample of 100 iron bars is said to be drawn from a large number of bars whose lengths are normally distributed with mean 4 feet and S.D. 0.6 feet. If the sample mean is 4.2 fee can the sample be regarded as a truly random sample?

3) In a sample of 50 individuals the average income was found to be Rs. 2172 per annum. It is known that the S.D. of the distribution of income is 90.11. Can we conclude that average annual income is Rs. 2200?

4) It is claimed that a random sample of 100 tyres (with a mean life of 15269 km) is drawn from a population of tyres which has a mean life of 15200 km and a standard deviation of 1248 km. Test the validity of this claim.

5) A sample of size 400 was drawn and the sample mean was found to be 99. Test whether this sample could have come from a normal population with mean 100 and variance 64 at 5% level of significance.

6) A sample of 100 students is found to have a mean height of 162.6 cm. Can it be reasonably regarded as a simple sample from a large population with mean height of 167.6 cm and S.D. 10 cm?

7) A sample of 900 members has a mean 3.4 cm and S.D. 2.61 cm. Is the sample from a large population of mean 3.25 cm and S.D. 2.61 cm? If the population is normal and its mean is unknown find the 95% and 98% confidence limits of true mean?

8) A random sample of 400 flower stems has an average length of 10cm. Can this be regarded as a sample from a large population with mean of 10.2 cm and a S.D. of 2.25 cm?

9) An ambulance service claims that it takes on the average 8.9 minutes to reach its destination in emergency calls. To check on this claim the agency which issues license to ambulance services has them timed on 50 emergency calls getting a mean of 9.2 minutes with a S.D. of 1.6 minutes. What can they conclude at the 5% level of significance?

10) A large group of athletes is found to have a mean weight of 140 lbs can S.D. 5 lbs. One student from a college was found to weight 120 lbs. Can it be reasonably concluded that he was not a athlete?

11) A sample of 5000 students from Madurai Kamaraj University was taken and their average weight was found to be 62.5 Kgs. with a s.d. of 22 Kgs. Find the 95% confidential limits of the average weight of the students in the entire university

12) The average sale of a toilet soap in a particular locality in a particular shop with an average 320 and S.D. 40. An attractive display of advertisement for the soap in local TV increased in 36 days the sale by 70 in that soap in a day. Can we say that the advertisement has helped very much?

13) A random sample of 400 items is found to have mean of 82 and S.D. of 18. Find 95% confidence limits for the mean of the population from which the sample is drawn.

14) A company manufacturing electric bulbs claims that the average life of its bulbs is 1600 hours. The average life and standard deviation of a random sample of 100 such bulbs were 1570 hours and 120 hours respectively. Should we accept the claim of the company.

15) Daily sales figures of 40 shop keepers showed that their average sales and standard deviation were Rs. 528 and 600 respectively. Is the assertion that daily sale on the average is Rs. 400 contradicted at 5% level of significance by the sample?

(2) ADDITIONAL PROBLEMS ON TEST OF SIGNIFICANCE FOR DIFFERENCE OF SAMPLE MEANS

Example 1. The following data gives the means of two samples taken from a population. Examine whether there is any significant difference between the two samples. $n_1 = 1000$; $n_2 = 2000$; $\bar{x} = 67.5$; $\bar{x}_2 = 68$; $\sigma = 2.5$.

Solution : Set the null hypothesis $H_0 : \mu_1 = \mu_2$.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}} = \frac{-.5\sqrt{2000}}{2.5\sqrt{3}} = -5.16.$$

$\therefore \quad |Z| = 5.16 > 3$. Hence the difference is highly significant.

Example 2. Two populations have their means equal but the s.d. of one is twice the other. (i) Show that in the sample of size 2000 drawn from each under simple sampling conditions the difference of means will in all probability not exceed. 15σ where σ is the smaller s.d. (ii) Find the probability that the difference will exceed half this amount.

Solution : Given $\mu_1 = \mu_2$ and $\sigma_1 : \sigma_2 = 1 : 2$ or $\sigma_2 = 2\sigma_1$ or σ, 2σ are the standard deviations. $n_1 = 2000 = n_2$.

(i) $$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma^2}{2000} + \frac{(2\sigma^2)}{2000}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{5}{2000}}}$$

$$\therefore \quad |Z| = \frac{|\bar{x}_1 - \bar{x}_2|}{.05\sigma}$$ Under simple sampling conditions

$$|Z| < 3 \quad \Rightarrow \quad |\bar{x}_1 - \bar{x}_2| < 3 \times .05\sigma$$

$$\Rightarrow \quad |\bar{x}_1 - \bar{x}_2| < .15\sigma.$$

Hence (i) is proved.

(ii) We have to find $\quad P\left(|\bar{x}_1 - \bar{x}_2| > \frac{.15\sigma}{2}\right)$

(i.e.) to find $P(|Z|(.05\sigma) > .075\sigma)$ (i.e.) to find $P(|Z| > 1.5)$

$$\text{Now } P(|Z| > 1.5) = 1 - P(|Z| < 1.5)$$
$$= 1 - 2P(0 < Z < 1.5)$$
$$= 1 - 2\,(.4332) \text{ (from normal table)}$$
$$= 0.1336.$$

Example 3. The number of accidents per day were studied for 144 days in Madras city and for 100 days in Delhi city. The mean numbers of accidents and the s.d were respectively 4.5 and 1.2 for Chennai city and 5.4 and 1.5 for Delhi city. Is Chennai more prone to accidents than Delhi?

Solution : Given $n_1 = 144;\quad \bar{x}_1 = 4.5;\quad \sigma_1 = 1.2$

$n_2 = 100;\quad \bar{x}_2 = 5.4;\quad \sigma_2 = 1.5$

Set the null hypothesis $H_0 : \mu_1 = \mu_2$. (the two cities have the same accident rates). Since the population standard deviations are not given we take the sample s.d. as the population s.d. and apply the test statistic.

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{4.5 - 5.4}{\sqrt{\frac{1.2^2}{144} + \frac{1.5^2}{100}}} = -4.99$$

$\therefore\ |Z| = 4.99 > 3$ are reject the hypothesis that the two cities have the same accident rates. However since Delhi has higher rate of accidents than Chennai city. We can say that Delhi is more prone to accidents.

Example 4. The mean height of 50 male students who showed above average participation in college was 68.2" with s.d. of 2.5"; while 50 male students who showed no interest in such participation had mean height 67.5" with s.d. 2.8". Test the hypothesis that male students who participate in college sports are taller than other male students.

Solution : Given $n_1 = 50;\quad \bar{x}_1 = 68.2;\quad s_1 = 2.5$

$n_2 = 50;\quad \bar{x}_2 = 67.5;\quad s_2 = 2.8$

Set the null hypothesis $H_0 : \mu_1 = \mu_2$ (there is no significant difference between the mean height of the male students who participated and who did not participate in college sports). The alternative hypothesis is $H_1 : \mu_1 > \mu_2$ (right tailed test).

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{68.2 - 67.5}{\sqrt{\frac{2.5^2}{50} + \frac{2.8^2}{50}}}$$

$$= \frac{0.7\sqrt{50}}{\sqrt{6.25 + 7.84}} = \frac{0.7 \times 7.07}{3.75} = 1.32$$

Since $Z = 1.32 < 1.96$ it is not significant at 5% level and we accept the null hypothesis and we conclude that the average height of the male students who participate in college sports is same as the average height of the other male students in the college.

Example 5. The mean yields of rice from two places in a district were 210 kgs and 220 kgs per acre from 100 acres and 150 acres respectively. Can it be regarded that the sample were drawn from the same district which has the s.d. of 11 kgs per acre?

Solution : $n_1 = 100$; $\bar{x}_1 = 210$; $\sigma = 11$; $n_2 = 150$; $\bar{x}_2 = 220$

Set the null hypothesis $H_0 : \mu_1 = \mu_2$ (samples have been drawn from the same population of s.d. 11 kgs.) Hence $H_1 : \mu_1 \neq \mu_2$.

$\therefore$ The test statistics $= Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$$Z = \frac{210 - 220}{11\sqrt{\frac{1}{100} + \frac{1}{150}}} = \frac{-10}{11\sqrt{\frac{250}{15000}}} = -7.04$$

$\therefore |Z| = 7.04 > 3$. The value is highly significant and hence we reject the null hypothesis. Hence we conclude that the samples are certainly not from the same district with the s.d. 11.

Example 6. The average income of persons was Rs. 210 with S.D. of Rs. 10 in a sample of 100 people of a city. For another sample of 150 persons the average income was Rs. 220 with s.d. of Rs. 12. The S.D. of incomes of the people of the city was Rs. 11. Test whether there is any significant difference between the average income of the localities.

Solution : Given $n_1 = 100$; $\bar{x}_1 = 210$; $s_1 = 10$.

$n_2 = 150$; $\bar{x}_2 = 220$; $s_2 = 12$.

$\sigma = 11$.

Set the null hypothesis $H_0 : \mu_1 = \mu_2$ (there is no significant difference between the average income of the localities).

$H_1 : \mu_1 > \mu_2$

Under the null hypothesis H_0 the test statistic is

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

$$= \frac{210 - 220}{\sqrt{\frac{10^2}{100} + \frac{12^2}{150}}} = \frac{-10}{\sqrt{1 + 0.96}} = \frac{-10}{\sqrt{1.96}} = \frac{-10}{1.4} = -7.14$$

$\therefore$ $|Z| = 7.14 > 1.96$. The difference is significant.

$\therefore$ H_0 is rejected at 5% level of significance.

Example 7. The following data are got from an investigation.

Samples	No. of cases	Mean Wages	S.D. of the Wages
Sample I	400	Rs. 47.4	Rs. 3.1
Sample II	900	Rs. 50.3	Rs. 3.3

Find out whether the two mean wages differ significantly.

Solution : Given $n_1 = 400$; $\bar{x}_1 = 47.4$; $s_1 = 3.1$

$$n_2 = 900; \quad \bar{x}_2 = 50.3; \quad s_2 = 3.3$$

Set the null hypothesis H_0 : (the difference in mean wages is not significant).

$$Z = \frac{47.4 - 50.3}{\sqrt{\frac{3.1^2}{400} + \frac{3.3^2}{900}}} = -15.26$$

$\therefore |Z| = 15.26 > 1.96$. Hence the difference is significant at 5%.

PROBLEMS FOR PRACTICE

1) A college conducts both day and night classes intended to be identical. A sample of 100 days students yields examination result mean 72.4 and S.D. 14.8. A sample of 200 night students yields examination result mean 73.9 and S.D. 17.9. Are the two means statistically equal at 5% level?

2) Mean yields of milks in a month for two sets of cows and their variability of yields are as follows. Examine whether the differences in mean yield of milk of two sets of cows are significant.

	Set of 60 cows	Set of 90 cows
Mean yield of milk per cow	150	170
S.D. per cow	15	13

3) Given the following data. Test whether the means of the samples are significantly different.

Sample	Size	Mean
I	50	140
II	60	150

S.D. of the population = 10.

4) A simple sample of heights of 6400 English men has a mean of 67.85 inches and S.D. 2.56 inches while a simple sample of heights of 1600 Australians has a mean of 68.55 and a S.D. of 2.52 inches. Do the data indicate that Australians are on the average taller than English men?

5) In a random sample of 500 the mean is found to be 20. In another independent sample of 400 the mean is 15. Could the sample have been drawn from the same population with s.d. 4?

6) A random sample of 1000 men from Northern India gives their mean wage to be Rs. 30 per day with a S.D. of Rs. 1.50. A sample of 1500 men from Southern India gives a mean wage of Rs. 32 per day with S.D. of Rs. 2. Discuss whether the mean rate of wages varies between the two regions.

7) The means of simple samples of 1000 and 2000 are 67.5 and 68.8 inches respectively. Can the samples be regarded as drawn from the same population of s.d. 2.5 inches.

8) Given the following information relating to two classes of workers *A* and *B*. Test whether there is any significant difference between their mean wages.

Class	Mean wages (Rs.)	S.D. (Rs.)	No. of workers
A	47	28	1000
B	49	49	1000

9) A random sample of 1000 men from city *A* shows mean wage of Rs. 250 with s.d. of Rs. 150. A sample of 1500 men from city *B* has a mean wage of Rs. 268 with s.d. of Rs. 200. Discuss the statement "the wages very between city *A* and city *B*" at 5% level of significance.

10) 60 new entrants in a given university are found to have a mean weight of 68.6 kgs and 50 seniors have a mean weight of 69.51 kgs. Is the evidence conclusive that mean weight of the seniors is greater than that of the new entrants? Assume the s.d. of the weights to be 2.48 kgs.

ADDITIONAL PROBLEMS ON TEST FOR STANDARD DEVIATION

Example 1. A manufacturer of electric bulbs, by some process, finds the s.d. of the life of the lamps to be 100 hrs. He wants to change the process if the new process results in even smaller variation in the life of lamps. In adopting the new process a sample of 150 bulbs gave the S.D. of 95 hrs. Is the manufacturer justified in changing the process?

Solution : Given $\sigma = 100$; $s = 95$; $n = 150$.

Set the null hypothesis $H_0 : \sigma = s$

$$Z = \frac{s - \sigma}{\frac{\sigma}{\sqrt{2n}}} = \frac{95 - 100}{\frac{100}{\sqrt{300}}} = \frac{-5\sqrt{300}}{100}$$

$$= \frac{-5 \times 17.32}{100} = \frac{-86.6}{100} = .866$$

$\therefore \; | Z | = .866 < 1.96$. Hence the deviation is not significant at 5% level and hence the hypothesis is accepted.

So the manufacturer find no justification in changing the process on this evidence alone.

Example 2. The mean production of wheat of a sample of 100 plots is 200 kgs per acre with s.d. of 10 kgs. Another sample of 150 plots gives the mean production of wheat at 220 kgs with s.d. of 12 kgs. Assuming the s.d. of the 11 kgs for the universe find at 1% level of significance, whether the two results are consistent.

Solution : Given $\sigma = 11$ and

	Size	Mean	S.D.
Sample 1	$n_1 = 100$	$\bar{x}_1 = 200$	$s_1 = 10$
Sample 2	$n_2 = 150$	$\bar{x}_2 = 220$	$s_2 = 12$

Set the null hypothesis $H_0 : \mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$ (the two results are consistent).

For $H_0 : \mu_1 = \mu_2$ the test statistic is

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{200 - 200}{11\sqrt{\frac{1}{100} + \frac{1}{150}}} = \frac{-20}{11\sqrt{\frac{10}{600}}} = \frac{-155}{11}$$

$= -14.1$ (approximately)

$\therefore \ |Z| = 14.1 > 3$. Hence the two means differ significantly at 5% level and even at 1% level.

For $H_0 : \sigma_1 = \sigma_2$ the test statistic is

$$Z = \frac{s_1 - s_2}{\sigma\sqrt{\frac{1}{2n_1} + \frac{1}{2n_2}}} = \frac{10 - 12}{11\sqrt{\frac{1}{200} + \frac{1}{300}}} = 1.99$$

$\therefore \quad 19.6 < |Z| < 2.58.$

Hence the difference of s.d. is significant at 5% level and not significant at 1% level.

$\therefore$ At 1% level the difference between s.d. is not significant but between means it is significant.

Hence we can conclude that at 1% level the two results are not consistent.

PROBLEMS FOR PRACTICE

1) The s.d. of a random sample of 900 members is 4.6 and that of another independent sample of 1600 is 4.8. Examine if the S.D. are significantly different.

2) Random samples drawn from two countries gave the following data

relating to the heights of adult males.

	Country A	Country B
Mean height in inches	67.42	67.25
Standard deviation	2.58	2.5
Number in samples	1000	1200

(i) Is the difference between the means significant?

(ii) Is the difference between the standard deviations significant?

3) From the experience of a manufacturer of battery cells according to a technique the S.D. of the life of battery cells was 150 days. He is interested to introduce a new technique in manufacturing batteries with less variation in life of batteries. In his new technique with a sample of 200 batteries he got the S.D. of the life of battery cells as 140 days. Is the manufacturer justified in changing the technique?

4) The S.D. of the height of all students in Bangalore University is 4". Two samples are taken. The S.D. of the height of 100 B.Sc. students is 3.5" and the height of 100 B.C.A. Students is 4.5". Test the significance of the difference of S.d. of the samples.

5) In a survey of incomes of two classes of workers of two random samples gave the following details. Examine whether the difference between (i) means and (ii) the s.d. are significant.

Sample	Size	Mean Annual Income in Rs.	standard deviation in Rs.
I	100	582	24
II	100	546	28

10.15 TESTS OF SIGNIFICANCE BASED ON T-DISTRIBUTION (Small Samples)

In the earlier sections, we discussed certain tests of significance based on the theory of normal distribution. These tests are valid only for large samples. When the sample is small the sampling distribution in most cases may not be normal.

Small sample test hold good for large as well as small samples. However the large sample test can not always be applied to small samples. For practical purposes a sample of size n is taken to be small if $n < 30$.

Consider a normal population with mean μ and s.d. σ. Let $x_1, x_2, \ldots\ldots, x_n$ be a random sample of size n with mean $\bar{x}$ and standard deviation σ. We know that $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ is a standard normal variate $N(0,1)$.

If the standard deviation σ of the population is not known then we have to estimate it from the sample standard deviation s. It has been proved that $s\sqrt{\dfrac{n}{n-1}}$ can be taken as an estimate of σ. If n is large it is close to s and hence in large sampling theory s was taken as an estimate of σ. If n is small $s\sqrt{\dfrac{n}{n-1}}$ is the estimate for σ. Hence the test statistic in small sample becomes.

$$Z = \frac{\bar{x}-\mu}{\dfrac{\left(s\sqrt{\dfrac{n}{n-1}}\right)}{\sqrt{n}}} = \frac{\bar{x}-\mu}{\dfrac{s}{\sqrt{n-1}}}$$

Now let us recall the definition :

$$t = \frac{\bar{x}-\mu}{\dfrac{s}{\sqrt{n-1}}} \qquad \ldots\ldots(1)$$

where (1) follows Student's t–distribution with $n-1$ degrees of freedom.

Like normal distribution this t–distribution has also extensive applications. Like the table of area for normal distribution, tables of area of this distribution are available. The table values at 5% level and 1% level are generally referred to as $t_{.05}$ and $t_{.01}$ respectively.

The following are some tests of significance based on Student's t–distribution.

I. Test for the difference between the mean of a sample and that of a population.

Under the null hypothesis $H_0 : \mu =$ the test statistic

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

which can be tested at any level of significance with $n-1$ degrees of freedom.

II. Test for the difference between the means of two samples.

II(a) If $\bar{x}$ and $\bar{x}_2$ are the means of two independent samples of sizes n_1 and n_2 from a normal population with mean μ and standard deviation σ it is found that

$$\frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

If σ is not known we estimate it from the samples. If s_1 and s_2 are the standard deviations of the samples then σ can be estimated as

$$\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$ Hence we can get

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

which following a t–distribution with d.f.v $= (n_1 + n_2 - 2)$.

Setting the null hypothesis $H_0 = \mu_1 = \mu_2$ the above test statistic 't' in (1) can be tested at any level of significance.

II(b). Suppose the sample size are equal; i.e., $n_1 = n_2 = n$. Then we have n pairs of values. Further we assume that the n pairs are independent. Then the test statistic t in (1) becomes

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n\left(s_1^2 + s_2^2\right)}{2n - 2}\left(\frac{2}{n}\right)}}$$

$\therefore \quad t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\left(s_1^2 + s_2^2\right)}{(n-1)}}}$ is a Student's t variate with $\nu = n + n - 2 = 2n - 2$.

II(c) Suppose the sample sizes are equal and if the n pairs of values in this case are not independent. Then the pairs of values are in some way associated. Hence we can not adopt the above method. We find the differences on the associated pairs of values and apply the test statistic $t = \dfrac{\bar{x} - \mu}{s / \sqrt{n-1}}$ to test whether the means of differences is significantly different from zero. In this case the d.f. is $n - 1$.

Confidence limits (Fiducial limits). If σ is not known and n is small then

(i) 95% confidence limits for μ is $\left(\bar{x} - \dfrac{s t_{.05}}{\sqrt{n-1}},\ \bar{x} + \dfrac{s t_{.05}}{\sqrt{n-1}}\right)$

(ii) 99% confidence limits for μ is $\left(\bar{x} - \dfrac{s t_{.01}}{\sqrt{n-1}},\ \bar{x} + \dfrac{s t_{.01}}{\sqrt{n-1}}\right)$

Illustrative Examples

Example 1. Ten specimen of copper wires drawn from a larger lot have the following breaking strengths in kgs 578, 572, 570, 568, 572, 575, 570, 572, 569, 548. Test whether the mean breaking strengths of the lot may be taken to be 578 kg weight.

Solution : Given $n = 10$; $\mu = 578$. We use the t–test. We require to find mean $\bar{x}$ and standard deviation s.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5694}{10} = 569.4$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$= \frac{1}{10}\,[8.6^2 + 2.6^2 + 0.6^2 + (-1.4)^2 + 2.6^2 + 5.6^2 + 0.6^2 + 2.6^2 + (-0.4)^2 + (-21.4)^2]$$

$$= \frac{1}{10}\,[73.96 + 6.76 + 0.36 + 1.96 + 6.76 + 31.36 + 0.36 + 6.76 + 0.16 + 457.96]$$

$$s^2 = \frac{1}{10}[586.4] = 58.64$$

$$\therefore \quad s = 7.66$$

Set the null hypothesis $H_0 : \mu = 578$.

Under this null hypothesis the test statistics.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{569.4 - 578}{\frac{7.66}{\sqrt{9}}} = \frac{-8.6 \times 3}{7.66} = \frac{-25.8}{7.66} = -3.37.$$

$$\therefore \quad |t| = 3.37 > t_{.05}.$$

$\therefore$ The difference is significant at 5% level of significance.

$\therefore$ H_0 is rejected at 5% level of significance. The mean break strength of copper wires cannot be taken as 578.

Example 2. A random sample of 10 boys has the following I.Q (Intelligent Quotients). 70, 120, 110, 101, 88, 83, 95, 98, 107, 110. Do these data support the assumption of a population mean I.Q. of 100?

Solution : Given $n = 10$; $\mu = 100$. Set $H_0 : \mu = 100$. Under H_0 test statistic $t = \dfrac{\bar{x} - \mu}{s / \sqrt{n-1}} \sim t_{(n-1)}$ where $\bar{x}$ and s can be calculated from the sample data as $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{972}{10} = 97.2$ and

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{1833.60}{10} = 183.36. \quad \text{Hence } s = 13.54$$

$$\therefore \quad t = \frac{97.2 - 100}{13.54 / \sqrt{9}} = \frac{-2.8 \times 3}{13.54} = -.6204$$

$\therefore \quad |t| = .62$ (nearly).

The table value for 9 d.f. at 5% level of significance is $t_{.05} = 2.26$,

$\therefore \quad |t| = .62 < t_{.05}$.

Hence the difference is not significant at 5% level. Hence H_0 may be accepted at 5% level. Hence the data support the assumption of population mean 100.

Example 3. A group of 10 rabits fed on a diet A and another group of 8 rats fed on a different diet B recorded the following increase in weights in gms.

Diet A	5	6	8	1	12	4	3	9	610
Diet B	2	3	8	1	10	2	8	–	–

Test whether diet A is superior to diet B.

Solution : Given $n_1 = 10$; $n_2 = 8$.

Mean of the first sample $\bar{x}_1 = \dfrac{5+6+......+10}{10} = \dfrac{64}{10} = 6.4$

Mean of the second sample $\bar{x}_2 = \dfrac{2+3+......+8}{8} = \dfrac{40}{8} = 5.0$

Standard deviations s_1 and s_2 of the first and second samples can be found as $s_1^2 = \dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2 = 10.24$ and $s_2^2 = 10.25$

Set the null hypothesis $H_0 : \mu_1 = \mu_2$ (There is no significant difference between mean weight due to A and B.)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{n_1 s_2^2 + n_2 s_2^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

$$= \frac{6.4 - 5}{\sqrt{\dfrac{10 \times 10.24 + 8 \times 10.25}{10+8-2}\left(\dfrac{1}{10} + \dfrac{1}{8}\right)}}$$

$$= \frac{1.4}{\sqrt{11.525(.1 + .125)}} = \frac{1.4}{1.6} = .875.$$

Table value for t at 5% level of significance for $(n_1 + n_2 - 2) = 16$ d.f. is $t_{.05} = 2.12$. Since $t = .875 < t_{05}$ the difference is not significant at 5% level of significance. Hence the null hypothesis may be accepted. Hence we may conclude that the 2 diets do not differ significantly as far as the effect on increase in weights is concerned.

∴ We can not say that the diet A is superior to the diet B.

Example 4. The following table gives the lengths of 12 samples of Egyptian cotton taken from a consignment.

48, 46, 49, 46, 52, 45, 43, 47, 47, 46, 45, 50.

Test if the mean length of the consignment can be taken as 46.

Solution : Given $n = 12$. $\mu = 46$. Set $H_0 : \mu = 46$.

Under H_0 the test statistic $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n-1}}$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{564}{12} = 47.$$

$$\therefore \quad s^2 = \frac{\sum (x_i - 47)^2}{12} = \frac{1+1+4+1+25+4+16+1+4+9}{12}$$

$$= \frac{66}{12} = 5.5.$$

$$\therefore \quad s = 2.35.$$

$$\therefore \quad t = \frac{47-46}{2.35/\sqrt{11}} = \frac{\sqrt{11}}{2.35} = 1.41.$$

The table value for $\nu = 11$ d.f. at 5% level of significance is $t_{.05} = 2.20$. Hence $|t| = 1.41 < t_{.05}$.

$\therefore$ The difference is not significant at 5% level. Hence we can accept the hypothesis and we can conclude that we can take the mean length of the consignment to be 46.

Example 5. The table gives the biological values of protein from 6 cows' milk and 6 buffaloes' milk. Examine whether the differences are significant.

Cow's milk	1.8	2.0	1.9	1.6	1.8	1.5
Buffalo's milk	2	1.8	1.8	2.0	2.1	1.9

Solution : We can find :

Mean value of protein of cow's milk = 1.8.

Mean value of protein on buffalo's milk = 1.9.

Variance of protein of cow's milk = .03.

Variance of protein of buffalo's milk = .01.

We notice that the two sets of observations are independent. Given $n_1 = n_2 = 6$; $\bar{x}_1 = 1.8$; $\bar{x}_2 = 1.9$; $s_1^2 = .03$; $s_2^2 = .01$

Set the null hypothesis $H_0 : \mu_1 = \mu_2$. Under this null hypothesis the test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2 + s_2^2}{n-1}}} = \frac{-.1}{\sqrt{\dfrac{.03+.01}{5}}} = \frac{-1}{\sqrt{\dfrac{.04}{5}}} = \frac{-.1}{.09} = -1.11.$$

and the d.f. $\nu = 2n - 2 = 10$.

The table value for $\nu = 10$ d.f. at 5% level of significance is 2.23.

$\therefore$ $|t| = 1.11 < 2.23$. Hence the difference is not significant. Hence the hypothesis is accepted.

Example 6. Test whether the sample having the values 63, 63, 64, 55, 66, 69, 70, 70, 71 has been chosen from a population with mean 65 at 5% level of significance.

Solution : Given $n = 9$; $\mu = 65$. Set $H_0 : \mu = 65$.

Under H_0, the test statistic $t = \dfrac{\bar{x} - \mu}{s / \sqrt{n-1}}$ where $\bar{x}$ and s are to be calculated from the given sample.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{591}{9} = 65.67$$

$$\bar{x} = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2 = 4335.22 - 4312.55 = 22.67 \text{ (verify)}.$$

$\therefore \quad s = 4.8$

$$\therefore \quad t = \frac{65.67 - 65}{4.8/\sqrt{8}} = \frac{(0.67)/\sqrt{8}}{4.8} = 0.39$$

$\therefore \quad |t| = 0.39.$

The table value for $\nu = 8$ d.f. at 5% level of significance is $t_{.05} = 2.31$.

$\therefore |t| < t_{.05}$. Hence the difference is not significant at 5% level. Hence we can accept the hypothesis that the sample could have chosen from the population with mean 65.

Example 7. To compare the prices of a certain product in two cities ten shops were selected at random in each town. The prices noted are given below.

City I	61	63	56	63	56	63	59	56	44	61
City II	55	54	47	59	51	61	57	54	64	58

Test whether the average prices can be said to be same in the two cities.

Solution : $n = 10$. Mean price of City I is $\bar{x}_1 = \frac{\sum x_i}{n} = \frac{582}{10} = 58.2$

Mean price of City II is $\bar{x}_2 = \frac{\sum x_i}{n} = \frac{560}{10} = 56.$

For City I, $s_1^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2 = 3417.4 - 3387.24 = 30.16$

Similarly $s_2^2 = 3157.8 - 3136 = 21.8$

Set the null hypothesis $H_0 : \mu_1 = \mu_2$ (i.e.) there is no difference in average prices.)

Under this null hypothesis the test statistic is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{n-1}}} = \frac{58.2 - 56}{\sqrt{\frac{30.16 + 21.8}{9}}} = \frac{6.6}{\sqrt{51.96}} = \frac{6.6}{7.2} = 0.92.$$

The table value of t for $\nu = 2n - 2 = 18$ d.f. is $t_{.05} = 2.10$ at 5% level.

$|t| = 0.92 < t_{0.05}$.

∴ The null hypothesis is accepted.

∴ We can conclude that the average price is same in both the cities.

Example 8. It was found that a machine has produced pipes having a thickness .50 mm. To determine whether the machine is in proper working order a sample of 10 pipes is chosen for which the mean thickness is .53 mm and s.d. is .03 mm. Test the hypothesis that the machine is in proper working order using a level of significance of (i) .05 (ii) .01.

Solution : Given $\mu = .50$; $\bar{x} = .53$; $s = .03$; $n = 10$.

Set the null hypothesis $H_0 : \mu = .50$ (the machine is in proper working order.)

Under the null hypothesis the test statistic is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{.53 - .50}{.03} \times \sqrt{9} = \frac{.03 \times 3}{.03} = 3.$$

(i) The table value for $\nu = 9$ d.f. at 5% level of significance is $t_{.05} = 2.26$; (i.e.) $|t|$ $t_{.05}$.

∴ The difference is significant at 5% level of significance.

∴ The null hypothesis is rejected at 5% level of significance.

(ii) The table value for $\nu = 9$ d.f at 1% level of significance is $t_{.01} = 3.25$. Hence $|t| = 3 < t_{.01}$.

∴ The difference is not significant at 1% level of significance. Hence the null hypothesis is accepted at 1% level of significance.

Note : Since we can reject H_0 at 5% level but not at 1% level of significance we can conclude to check the machine or take at least another sample.

Example 9. The average breaking strength of steel rods is specified to be 18.5 thousand pounds. To test this a sample of 14 rods was tested. The mean and standard deviation obtained were 17.85 and 1.955 respectively. Is the result of the experiment significant with 95% confidence?

Solution : Given $\mu = 18.5$; $\bar{x} = 17.85$; $s = 1.955$; $n = 14$.

Set the null hypothesis $H_0 : \mu = 18.5$

Under the null hypothesis the test statistics is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{17.85 - 18.5}{1.955/\sqrt{13}} = \frac{-0.65 \times 3.61}{1.955} = \frac{-2.35}{1.955} = -1.20.$$

$$\therefore \quad |t| = 1.20$$

Here $v = n - 1 = 13$. The table value for $v = 13$ d.f. at 5% level of significance is $t_{.05} = 2.16$ $|t| < t_{.05}$.

∴ The difference is not significant at 5% level of significance. Hence H_0 is accepted.

Hence the hypothesis may be accepted at 5% level of significance.

Example 10. Find 95% confidence limits for the mean of a normally distributed population from which the following sample was taken :

15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

Solution : Here $n = 10$. $\bar{x} = \frac{130}{10} = 13$.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{10} = \frac{4+16+9+25+9+16+36+4+0+1}{10} = \frac{120}{10}$$

$s^2 = 12$

Hence $s = 3.40$.

Table value for 9 d.f. at 5% level of significance is 2.26.

95% confidence limits for the population mean μ is

$$= \left(\bar{x} - \frac{st_{.05}}{\sqrt{n-1}},\ \bar{x} + \frac{st_{.05}}{\sqrt{n-1}}\right)$$

$$= \left(13 - \frac{3.46 \times 2.26}{\sqrt{9}},\ 13 + \frac{3.46 \times 2.26}{\sqrt{9}}\right)$$

$$= (13 - 2.6,\ 13 + 2.6) = (10.4,\ 15.6).$$

PROBLEMS FOR PRACTICE

1) 10 students are selected at random in a university and their heights are measured in inches as 64, 65, 65, 67, 67, 69, 70, 72, 72. In light of these data discuss the suggestion that he mean height of the students in the university is 66 inches.

2) Ten specimen of iron bars drawn from a large lot have the following breaking strength in kg. weight : 578, 572, 570, 568, 572, 578, 570, 572, 596, 584. Test whether the mean breaking strength of the lot may be taken to be 578 kg.

3) The mean life time of electric bulbs produced by a company has in the past been 1120 hrs. with the a standard deviation of 125 hrs. A sample of 8 electric bulbs chosen from a supply of newly produced bulbs showed a mean life time of 1070 hrs. Test the hypothesis

that the mean life time of the bulbs has not changed (using 5% level of significance).

4) A soap manufacturing company was distributing a particular brand of soap through a large number of retail shops. Before a heavy advertisement campaign the mean sales per week per shop was 140 dozens. After the campaign a sample of 26 shops was taken and the mean sale was found to be 147 dozens with standard deviation 16. Can you consider the advertisement effective?

5) Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh in kgs as follows. 50, 49, 52, 44, 45, 48, 46, 45, 49, 45. Test whether the average packing can be taken to be 50 kgs.

6) A machine is expected to produce nails of length 7 cm. A random sample of 10 nails were found to measure in cm 7.2, 7.3, 7.1, 6.9, 6.8, 6.5, 6.9, 6.8, 7.1, 7.2 respectively. On the basis of this sample what can we say about the reliability of the machine?

7) Prices of shares of a company on the different days in a month were found to be 66, 65, 69, 70, 69, 71, 70, 63, 64, 68. Discuss whether the mean prices of the shares in the month is 65.

8) A random sample of 8 envelops is taken from a letter box of a post office and their weights in grams are found to be 12.1, 11.9, 12.4, 12.3, 11.9, 12.1, 12.4, 12.1. Does this sample indicate an average weight of the envelopes received at that post office is 12.35 grams at 1% level of significance.

9) 2 types of batteries are tested for their length of life and the following data is obtained. Is there a significant difference in the 2 means?

	No. of samples	Mean life in hrs	Variance
Type A	9	600	121
Type B	8	640	144

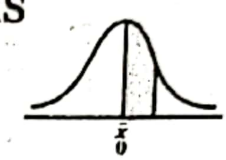
10) In an examination in Mathematics 12 students in one class had mean mark of 78 with a s.d. of 6 while 15 students in another class had a mean mark 74 with a s.d. of 8. Determine whether the first group is superior to the second group using a significance level of 5%.

11) A group of 5 patients treated with medicine A weight 42, 39, 48, 60, 41 kgs. Second group of 7 patients from the same hospital treated with medicine B weight 38, 42, 56, 64, 68, 69, 62 kgs. Do you agree with the claim that medicine B increases the weight significantly?

12) Find 98% confidence limits for the mean of a normally distributed population from which the following sample was taken : 19, 16, 15, 15, 14, 13, 12, 10, 9.

APPENDIX - NORMAL DISTRIBUTION - AREAS

Area under the standard normal curve from 0 to z



z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	3078	.3106	.3133
0.9	.4192	.3186	.3212	.3238	.3264	.3289	.3315	3315	.3365	.3389
1.0	.3413	.3438	.3461	.3845	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4713	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4916
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4949	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4951	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

X^2-DISTRIBUTION-CRITICAL VALUES

d.f.	ONE–TAILED TEXT (UPPER TAIL)		TWO-TAILD TEST		TWO-TAILED TEST	
	5%	1%	5%		1%	
	k	k	k_1	k_2	k_1	k_2
1	3.84	6.63	0.098	5.02	0.039	7.88
2	5.99	9.21	0.05	7.38	0.01	10.60
3	7.81	11.34	0.22	9.335	0.07	12.84
4	9.49	13.28	0.48	11.14	0.21	14.86
5	11.07	15.09	0.83	12.83	0.41	16.75
6	12.59	16.81	1.24	14.45	0.68	18.55
7	14.07	18.48	1.69	16.01	0.99	20.26
8	15.51	20.09	2.18	17.53	1.34	21.96
9	16.92	21.67	2.70	19.02	1.73	23.59
10	18.31	23.21	3.25	20.48	2.16	25.19
11	19.68	24.72	3.82	21.92	2.60	26.76
12	21.03	26.22	4.40	23.34	3.07	28.30
13	22.36	27.69	5.01	24.74	3.57	29.82
14	23.68	29.14	5.63	26.12	4.07	31.32
15	25.00	30.58	6.26	27.49	4.60	32.80
16	26.30	32.00	6.91	28.85	5.14	34.27
17	27.59	33.41	7.56	30.19	5.70	35.72
18	28.87	34.81	8.23	31.53	6.26	37.16
19	30.14	36.19	8.91	32.85	6.84	38.58
20	31.41	37.57	9.59	34.17	7.43	40.00
21	32.67	38.93	10.26	35.48	8:03	41.40
22	33.92	40.29	10.98	36.78	8.64	42.80
23	35.17	41.64	11.69	38.08	9.26	44.18
24	36.42	42.98	12.40	39.36	9.89	45.56
25	37.65	44.31	13.12	40.65	10.22	46.93
26	38.89	45.64	13.84	41.92	11.16	48.29
27	40.11	46.96	14.57	43.19	11.81	49.64
28	41.34	48.28	15.31	44.46	12.46	50.99
29	42.56	49.59	16.05	45.72	13.12	52.34
30	43.77	50.89	16.79	46.98	13.79	53.67
40	55.76	63.69	24.43	59.34	20.71	66.77
60	79.08	88.38	8.303	83.30	35.53	91.95
80	101.88	112.33	106.63	106.63	51.17	116.32
100	124.34	135.81	129.56	129.56	67.33	140.17

POISSON DISTRIBUTION - EXPONENTS

m	e^{-m}
0.00	1.0000
0.01	0.9900
0.02	.9802
0.03	.9704
0.04	.9608
0.05	.9512
0.06	.9418
0.07	.9324
0.08	.9231
0.09	.9139
0.1	.9048
0.2	.8187
0.3	.7408
0.4	.6703
0.5	.6065
0.6	.5488
0.7	.4966
0.8	.4493
0.9	.4066
1.0	.3679
1.1	.3329
1.2	.3012
1.3	.2725
1.4	.2466
1.5	.2231
1.6	.2019
1.7	.1827
1.8	.1653
1.9	.1496
2.0	.1353

m	e^{-m}
2.1	.1225
2.2	.1108
2.3	.1003
2.4	.0907
2.5	.0821
2.6	.0743
2.7	.0672
2.8	.0608
2.9	.0660
3.0	.0498
3.1	.0450
3.2	.0408
3.3	.0369
3.4	.0334
3.5	.0302
3.6	.0273
3.7	.0247
3.8	.0224
3.9	.0202
4.0	.0183
4.1	.0166
4.2	.0150
4.3	.0136
4.4	.0123
4.5	.0111
4.6	.0101
4.7	.0091
4.8	.0082
4.9	.0074
5.0	.0067

m	e^{-m}
5.2	.0055
5.4	.0045
5.6	.0037
5.8	.0030
6.0	.0025
6.2	.0020
6.4	.0017
6.6	.0014
6.8	.0011
7.0	.00091
7.2	.00075
7.4	.00061
7.6	.00050
7.8	.00041
8.0	.00034
8.2	.00027
8.4	.00023
8.6	.00018
8.8	.00015
9.0	.00012
9.2	.000101
9.4	.000083
9.6	.000068
9.8	.000055
10.0	.000045
11.0	.000017
12.0	.000006

Help: $e^{-9.72} = e^{-9} \times e^{-0.7} \times e^{-0.02}$

$= 0.00012 \times 0.4966 \times 0.9802 = 0.0000584120784$