ESTIMATION

Estimate: An estimate is a statement made to find an unknown population parameter

Estimator: The procedure or rule to determine an unknown population parameter is called an estimator. For ex. Sample mean $\bar{x}$ is an estimator of population mean μ

Types of Estimation:

Basically there are two kinds of estimates to determine the statistic of the population parameters

(a) Point Estimation : A point estimate of a parameter θ is a single numerical value, which is computed from a given sample.

(b) Interval Estimation: An interval estimation is given by two values between which the parameter may be considered to lie.

Interval Estimation of μ: The interval estimation of μ is given by the interval $\left(\bar{x}-E_{Max}, \bar{x}+E_{Max}\right)$ where $E_{Max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Interval Estimation of p (proportion): The interval estimation of μ is given by the interval $\left(p-E_{Max}, p+E_{Max}\right)$ where $E_{Max} = z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

$z_{\alpha/2}$ **values :** 1.96 for 95% confidence

2.58 for 99% confidence

1.64 for 90% confidence

1. A random sample of size 100 has a S.D. of 5. What can you say about the maximum error with 95% confidence.

Sol.

Given σ = 5, n = 100, $z_{\alpha/2}$ for 95% confidence = 1.96

We know that $E_{Max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\frac{5}{\sqrt{100}} = 0.98$

2. Assuming that σ = 20.0, how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3.0

Sol.

Given maximum error E = 3.0, and σ = 2.0, $z_{\alpha/2}$ = 1.96 for 95% confidence

We know that $E_{Max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2$

$$n=\left(\frac{1.96\,X\,20}{3}\right)^2=170.74\approx 171$$

3. In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs. 472.36 and the S.D. of Rs. 62.35. If $\bar{x}$ is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed Rs. 10

Sol.

Given n = 80, $\bar{x}$ = 472.36, σ = 62.35, E_{max} = 10

$$E_{Max}=z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Rightarrow z_{\alpha/2}=\frac{E_{Max}.\sqrt{n}}{\sigma}=\frac{10\sqrt{80}}{62.35}=\frac{89.4427}{62.35}=1.4345$$

$$z_{\alpha/2}=1.43$$

The area when $z_{\alpha/2}=1.43$ from tables is 0.4236

$1-\alpha=2\times0.4236=0.8472$

Confidence = $(1-\alpha)100\%=84.72$

Hence we are 84.72% confidence that the maximum error is Rs. 10

4. If we can assert with 95% that the maximum error is 0.05 and P=0.2, find the size of the sample.

Sol.

Given P=0.2, E = 0.05

We have Q = 1- P = 1 – 0.2 = 0.8 and $z_{\alpha/2}=1.96(\,for\,95\%)$

We know that maximum error, $E_{Max}=z_{\alpha/2}\sqrt{\frac{PQ}{n}}\Rightarrow 0.05=1.96\sqrt{\frac{0.2\times0.8}{n}}$

$$n=\frac{0.2\times0.8\times(1.96)^2}{(0.05)^2}=246$$

5. What is the size of the smallest sample required to estimate an unknown proportion to within a maximum error of 0.06 with at least 95% confidence

Sol.

Given E = 0.06, Confidence limit = 95%

i.e. $z_{\alpha/2}=1.96$

here P is not given, So we take $P=\frac{1}{2}\Rightarrow Q=\frac{1}{2}$

Hence $n=\left(\frac{z_{\alpha/2}}{E}\right)^2(PQ)\Rightarrow n=\left(\frac{1.96}{0.06}\right)^2\left(\frac{1}{2}.\frac{1}{2}\right)=266.78\approx 267$

6. The mean and S.D. of a population are 11,795 and 14054 respectively. What can one assert with 95% confidence about the maximum error if $\bar{x}$ = 11,795 and n = 50. And also construct 95% confidence interval for the true mean

Sol.

Given $\mu = 11795, \sigma = 14054, \bar{x} = 11795$, $n = 50, z_{\alpha/2} = 1.96$

$$E_{Max} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \Rightarrow 1.96.\frac{(14054)}{\sqrt{50}} = 3899$$

Confidence interval = $\left(\bar{x} - E_{Max}.\frac{\sigma}{\sqrt{n}}, \bar{x} + E_{Max}.\frac{\sigma}{\sqrt{n}},\right)$

$$(11795 - 3899, 11795 + 3899)$$

$$(7896, 15694)$$

7.(H.W) A sample of 10 cam shafts intended for use in gasoline engines has an average eccentricity of 1.02 and a S.D. of 0.044 inch. Assuming the data may be treated a random sample from a normal population, determine a 95% confidence interval for the actual mean eccentricity of the cam shaft?
(Ans) = (0.993, 1.047)

TEST OF HYPOTHESIS
(Large Samples)

Hypothesis: The assumptions (or guesses) about the population parameters is called Hypothesis. There are two types of hypothesis
(i) Null Hypothesis (ii) Alternative Hypothesis.

Null Hypothesis: The hypothesis which is a definite statement about the population parameter usually a hypothesis of no difference is called Null Hypothesis. It is denoted by H_0. Eg: H_0 : $\mu = \mu_0$

Alternative Hypothesis: Any Hypothesis which contradicts the Null Hypothesis is called an Alternative Hypothesis. It is denoted by H_1.

Eg: (i) H_1 : $\mu \neq \mu_0$ (two – tailed test)
(ii) H_1 : $\mu > \mu_0$ (Right – tailed test)
(iii) H_1 : $\mu < \mu_0$ (Left – tailed test)

Errors in Sampling: In practice we decide to accept or to reject the lot after examining a sample from it. As such we have two types of errors.

(i) **Type I error:** Reject H_0 when it is true.
If the Null Hypothesis H_0 is true but it rejected by test procedure, then the error made is called Type I error. It is denoted by α.
P(Reject H_0 when it is true) = P(Type I error) = α

(ii) **Type II error:** Accept H_0 when it is false.
If the Null Hypothesis H_0 is false but we accept H_1 by test procedure, then the error made is called Type II error. It is denoted by β.
P(Accept H_0 when it is false) = P(Type II error) = β

Level of Significance: The level of significance denoted by α is the confidence with which we rejects or accepts the Null hypothesis H_0 *i.e.* it is the maximum possible probability with which we are willing to risk an error in rejecting H_0 when it is true.

Test Statistic: Under a given hypothesis let the sampling distribution of a statistic t is approximately a normal distribution with mean E(t) then

$$z = \frac{t - E(t)}{S.E. of\ t}$$

Critical Values (Z_α) of Z

	Level of Significance		
	1%	5%	10%
Two – Tailed test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right – Tailed test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left – Tailed test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$

Test of Significance for large samples

There are 4 test of significance for large samples

(i) Test for single mean
(ii) Test for difference of means
(iii) Test for single proportion
(iv) Test for difference of proportion

I) Procedure for testing of Hypothesis for single mean

Suppose $\bar{x}$ is the mean of the sample with size n taken from a normal population with mean μ and S.D. σ

Step 1:Null Hypothesis : Define or set up a Null Hypothesis H_0

$$H_0 : \mu = \mu_0$$

Step 2:Alternative Hypothesis : Set up an Alternative Hypothesis H_1

(i) $H_1 : \mu \neq \mu_0$ (two – tailed test)

(ii) $H_1 : \mu > \mu_0$ (Right – tailed test)

(iii) $H_1 : \mu < \mu_0$ (Left – tailed test)

Step 3:Level of Significance: Select the appropriate level of significance α

Step 4:Test Statistic: Calculate the test statistic

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Step 5:Conclusion: Find the critical value Z_α at the level of significance α from normal table.

If $|Z_{cal}| < Z_\alpha$ accept H_0

If $|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

1. A sample of 64 students has a mean weight of 70kgs. Can this be regarded as a sample from a population with mean weight 56kgs and S.D. 25kgs

Sol.

Given $\bar{x}$ = 70, μ = 56, σ = 25 and n = 64

Step 1:Null Hypothesis : $H_0 : \mu = 56$

Step 2:Alternative Hypothesis :

(i) $H_1 : \mu \neq 56$ (two – tailed test)

Step 3:Level of Significance: α = 5%

Step 4:Test Statistic: The test statistic is

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{70 - 56}{25 / \sqrt{64}} = 4.48$$

Step 5:Conclusion: $|Z_\alpha|$ = 1.96 and Z_{Cal}=4.48

$|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

i.e. Samples cannot be regarded as one coming from given population

2. An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 level of significance

Sol.

Given $\bar{x}$ = 11, μ = 10, σ^2= 16 $\Rightarrow \sigma = \sqrt{16} = 4$ and n = 36

Step 1:Null Hypothesis : $H_0 : \mu = 10$

Step 2:Alternative Hypothesis :

(i) $H_1 : \mu < 10$ (left – tailed test)

Step 3:Level of Significance: α = 0.05

Step 4:Test Statistic: The test statistic is

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{11-10}{4/\sqrt{36}} = 1.5$$

Step 5:Conclusion: Z_α = 1.645 and Z_{Cal}=1.5

$|Z_{Cal}| < Z_\alpha$ accept H_0

3. A sample of 400 items is taken from a population whose S.D. is 10. The mean of the sample is 40. Test whether the sample has come from a population with mean 38. Also calculate 95% confidence interval for the population.

Sol.

Given $\bar{x}$ = 40, μ = 38, σ = 10 and n = 400

Step 1:Null Hypothesis : H_0 : $\mu = 38$

Step 2:Alternative Hypothesis :

(i) H_1 : $\mu = 38$ (two – tailed test)

Step 3:Level of Significance: α = 0.05

Step 4:Test Statistic: The test statistic is

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{40-38}{10/\sqrt{400}} = 4$$

Step 5:Conclusion: $|Z_\alpha|$ = 1.96 and Z_{Cal}=4

$|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

i.e. Samples cannot be regarded as one coming from given population with mean 38

95% confidence interval is $\left(\bar{x}-1.96\frac{\sigma}{\sqrt{n}}, \bar{x}+1.96\frac{\sigma}{\sqrt{n}}\right)$

$$\Rightarrow \left(\bar{x}-1.96\frac{10}{\sqrt{400}}, \bar{x}+1.96\frac{10}{\sqrt{400}}\right)$$

$$\Rightarrow (39.02, 40.98)$$

4(H.W). In 64 randomly selected hours of production, the mean and the standard deviation of the number of acceptance pieces produced by an automatic stamping machine are $\bar{x}$ = 1.038, σ = 1.46. At the 0.05level of significance does this enable us to reject the null hypothesis μ = 1.000 against the alternative hypothesis μ >1.000

Ans. Z_{cal} = 2.082, Z_α=1.645, Reject the null hypothesis

II) Procedure for testing of Hypothesis for difference of means

Suppose $\overline{x_1}$ and $\overline{x_2}$ be the means of the two samples with sizes n_1 and n_2 drawn from two population having means μ_1 and μ_2 and S.D. σ_1 and σ_2. To test whether the two population means are equal

Step 1:Null Hypothesis : Define or set up a Null Hypothesis H_0

H_0 : $\mu_1 = \mu_2$

Step 2:Alternative Hypothesis : Set up an Alternative Hypothesis H_1

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

Step 3:Level of Significance: Select the appropriate level of significance α

Step 4:Test Statistic: Calculate the test statistic

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

$$if\ \ \sigma_1^2 = \sigma_2^2 = \sigma^2\ \ then$$

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Step 5:Conclusion: Find the critical value Z_α at the level of significance α from normal table.

If $|Z_{cal}| < Z_\alpha$ accept H_0

If $|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

1. The means of two large samples of sizes 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the sample be regarded as drawn from the same population of S.D. 2.5 inches

Sol.

Given n_1 = 1000, n_2 = 2000 , $\overline{x_1}$ = 67.5, $\overline{x_2}$ = 68, σ = 2.5

Step 1:Null Hypothesis : H_0 : $\mu_1 = \mu_2$

Step 2:Alternative Hypothesis :

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

Step 3:Level of Significance: α=5%

Step 4:Test Statistic:

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{67.5 - 68}{2.5\sqrt{\frac{1}{1000} + \frac{1}{2000}}}$$

$$\Rightarrow z = \frac{-0.5}{0.0968} = -5.16$$

Step 5:Conclusion: $|Z_\alpha|$ = 1.96 and Z_{Cal}=-5.16

$|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

Hence the samples are not drawn from the same population of S.D. 2.5 inches

2. The mean height of 50 male students who participated in sports is 68.2 inches with a S.D. of 2.5. The mean height of 50 male students who have not participated in sports is 67.2 inches with a S.D. of 2.8. Test the hypothesis that the height of students who participated in sports is more than the students who have not participated in sports.

Sol.

Given n_1 = 50, n_2 = 50 , $\overline{x_1}$ = 68.2, $\overline{x_2}$ = 67.2, σ_1 = 2.5, σ_2 = 2.8

Step 1:Null Hypothesis : H_0 : $\mu_1 = \mu_2$

Step 2:Alternative Hypothesis :

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

Step 3:Level of Significance: α=5%

Step 4:Test Statistic:

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{68.2 - 67.2}{\sqrt{\frac{2.5^2}{50} + \frac{2.8^2}{50}}}$$

$$\Rightarrow z = 1.88$$

Step 5:Conclusion: $|Z_\alpha| = 1.96$ and $Z_{Cal} = 1.88$

$|Z_{Cal}| < Z_\alpha$ we accept H_0

Hence we conclude that there is no significance difference in the heights

3(H.W). At a certain large university a sociologist speculates that male students spend considerably more money on junk food than do female students. To test her hypothesis, the sociologist randomly select from the registrar's records the names of 200 students. Of there, 125 are men and 75 are women. The sample mean of the average amount spent on junk food per week by the men is Rs. 400 and S.D. is Rs. 100. For the women the sample mean is Rs. 450 and the sample S.D. is Rs. 150. Test the difference between the mean at 0.05 level

Ans. $Z_{cal} = 1.1547$, $Z_\alpha = 1.96$, Accept the null hypothesis

III) Procedure for testing of Hypothesis for single Proportion

Suppose p is proportion of the sample with size n taken from a normal population with proportion P

Step 1:Null Hypothesis : Define or set up a Null Hypothesis H_0

$H_0 : P = P_0$

Step 2:Alternative Hypothesis : Set up an Alternative Hypothesis H_1

(i) $H_1 : P \neq P_0$ (two – tailed test)

(ii) $H_1 : P > P_0$ (Right – tailed test)

(iii) $H_1 : P < P_0$ (Left – tailed test)

Step 3:Level of Significance: Select the appropriate level of significance α

Step 4:Test Statistic: Calculate the test statistic

$$z = \frac{p - P}{\sqrt{PQ/n}}$$

Step 5:Conclusion: Find the critical value Z_α at the level of significance α from normal table.

If $|Z_{cal}| < Z_\alpha$ accept H_0

If $|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

1. 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 equipment revealed that 18 were faulty. Test his claim at 5% level of significance

Sol.

Given n = 200, $p = \frac{182}{200} = 0.91, P = 95\% = 0.95$

Step 1:Null Hypothesis : $H_0 : P = 0.95$

Step 2:Alternative Hypothesis :

$H_1 : P < 0.95$ (Left – tailed test)

Step 3:Level of Significance: α = 5%

Step 4:Test Statistic: $z = \dfrac{p-P}{\sqrt{PQ/n}} = \dfrac{0.91-0.95}{\sqrt{\dfrac{0.92\times 0.05}{200}}} = -2.59$

Step 5:Conclusion: $|Z_\alpha|$ = 1.645 and Z_{Cal}= -2.59

$|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

i.e. The manufacturer claim is rejected

2. In a sample of 1000 people in Karnataka 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

Sol.

Given n = 1000, $p = \dfrac{540}{1000} = 0.54, P = \dfrac{1}{2} = 0.5$

Step 1:Null Hypothesis : H_0 : $P = 0.5$ *i.e.* Both rice and wheat are equally popular in the state

Step 2:Alternative Hypothesis :

H_1 : $P \neq 0.5$ (Two – tailed test)

Step 3:Level of Significance: α = 1%

Step 4:Test Statistic: $z = \dfrac{p-P}{\sqrt{PQ/n}} = \dfrac{0.54-.05}{\sqrt{\dfrac{0.5\times 0.5}{1000}}} = 2.532$

Step 5:Conclusion: $|Z_\alpha|$ = 2.58 and Z_{Cal}= 2.532

$|Z_{Cal}| < Z_\alpha$ accept H_0

i.e. Both rice and wheat are equally popular in the state

3(H.W). In a study designed to investigate whether certain detonators used with explosives in coal mining meet the requirement that at least 90% will ignite the explosive when charged. It is found that 174 of 200 detonators function properly. Test the null hypothesis P = 0.9 against the alternative hypothesis P ≠ 0.9 at 0.05 level of significance

Ans. Z_{cal} = -1.41, Z_α=1.645, Accept the null hypothesis

IV) Procedure for testing of Hypothesis for difference of proportions

Let p_1 and p_2 be the sample proportions in two large random samples of sizes n_1 and n_2 drawn from two populations having proportions P_1 and P_2.

Step 1:Null Hypothesis : Define or set up a Null Hypothesis H_0

H_0 : $P_1 = P_2$

Step 2:Alternative Hypothesis : Set up an Alternative Hypothesis H_1

H_1 : $P_1 \neq P_2$ (two – tailed test)

Step 3:Level of Significance: Select the appropriate level of significance α

Step 4:Test Statistic: Calculate the test statistic

(a) When the population proportions P_1 and P_2 are known

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{P_1Q_1}{n_1} + \dfrac{P_2Q_2}{n_2}}}$$

Where $Q_1 = 1 - P_1$ and $Q_2 = 1 - P_2$

(b) When the population proportions P_1 and P_2 are not known but sample proportions p_1 and p_2 are known

(i) **Method of Substitution:** In this method, sample proportions p_1 and p_2 are substituted for P_1 and P_2

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$

(i) **Method of Pooling:** Estimate the value of p from the two populations proportions as $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

So that q = 1 – p

Hence the test statistic is

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Step 5:Conclusion: Find the critical value Z_α at the level of significance α from normal table.

If $|Z_{cal}| < Z_\alpha$ accept H_0

If $|Z_{Cal}| > Z_\alpha$ reject H_0 and accept H_1

1. Random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same, at 5% level.

Sol.

Given $n_1 = 400$, $n_2 = 600$, $p_1 = \frac{200}{400} = 0.5, p_2 = \frac{325}{600} = 0.541$

Step 1:Null Hypothesis : Assume that there is no significant difference between the option of men and women as far as proposal of flyover is concerned.

$H_0: p_1 = p_2$

Step 2:Alternative Hypothesis :

$H_1: p_1 \neq p_2$ (Two – tailed test)

Step 3:Level of Significance: α = 5%

Step 4:Test Statistic: $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400 \times 0.5 + 600 \times 0.541}{400 + 600} = \frac{525}{1000} = 0.525$

$and\ q = 1 - p = 1 - 0.525 = 0.475$

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.5 - 0.541}{\sqrt{0.525 \times 0.475\left(\frac{1}{400} + \frac{1}{600}\right)}} = -1.28$$

Step 5:Conclusion: $|Z_\alpha| = 1.96$ and $Z_{Cal} = -1.28$

$|Z_{Cal}| < Z_{\alpha}$ accept H_0

i.e. there is no significant difference between the option of men and women as far as proposal of flyover is concerned.

2. In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat consumers is concerned?

Sol.

Given $n_1 = 1000$, $n_2 = 800$, $p_1 = \frac{400}{1000} = 0.4,\ p_2 = \frac{400}{800} = 0.5$

Step 1:Null Hypothesis : H_0 : $p_1 = p_2$ *i.e.* there is no difference

Step 2:Alternative Hypothesis :

H_1 : $p_1 \neq p_2$ (Two – tailed test)

Step 3:Level of Significance: $\alpha = 5\%$

Step 4:Test Statistic: $z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400 \times 0.4 + 400 \times 0.5}{1000 + 800} = \frac{360}{1800} = 0.2$

$$and\ q = 1 - p = 1 - 0.2 = 0.8$$

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.4 - 0.5}{\sqrt{0.2 \times 0.8\left(\frac{1}{800} + \frac{1}{1000}\right)}} = -5.2705$$

Step 5:Conclusion: $|Z_{\alpha}| = 1.96$ and $Z_{Cal} = -5.2705$

$|Z_{Cal}| > Z_{\alpha}$ reject H_0 and accept H_1

i.e. there is a significant difference between town A and town B as the proportion of wheat consumers is concerned.

3.(H.W). 100articles from a factory are examined and 10 are found to be defective. 500 similar articles from a second factory are found to be 15 defective. Test the significance between the difference of two proportions at 5% level.

Ans. $Z_{cal} = 3.18$, $Z_{\alpha} = 1.96$, Reject the null hypothesis

TEST OF HYPOTHESIS
(Small Samples)

The following are some important tests for small samples

(i) Student's 't' test, (ii) F-test (iii) χ^2 - test

DEGREES OF FREEDOM (d.f.) : The number of independent variates which make up the statistic is known as the degree of freedom (d.f.). It is denoted by ν (the letter 'Nu' of the Greek alphabet). If n is the sample size then $\nu = n-1$. If there are 2 samples of sizes n_1 and n_2 then $\nu = n_1 + n_2 - 2$

Student's 't' test:

(I) Test for Single mean

Let $\bar{x}$ = Mean of a sample
n = Size of the sample
σ = S.D. of the population
μ = Mean of the population supposed to be normal
s^2 = the unbiased estimate of population variance σ^2

Step 1:Null Hypothesis : Define or set up a Null Hypothesis H_0

$$H_0 : \mu = \mu_0$$

Step 2:Alternative Hypothesis : Set up an Alternative Hypothesis H_1

(i) $H_1 : \mu \neq \mu_0$ (two – tailed test)
(ii) $H_1 : \mu > \mu_0$ (Right – tailed test)
(iii) $H_1 : \mu < \mu_0$ (Left – tailed test)

Step 3:Level of Significance: Select the appropriate level of significance α

Step 4:Test Statistic: Calculate the test statistic

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

Step 5:Conclusion: Find the critical value t_α at the level of significance α with n-1 d.f. from t-distribution table.

If $|t_{cal}| < t_\alpha$ accept H_0
If $|t_{Cal}| > t_\alpha$ reject H_0 and accept H_1

Note: The confidence limits are $\left(\bar{x} - t_\alpha \frac{s}{\sqrt{n}}, \bar{x} + t_\alpha \frac{s}{\sqrt{n}}\right)$

1. The average breaking strength of the steel rods is specified to be 18.5 thousand pounds. To test this sample of 14 rods were tested. The mean and S.D. obtained were 17.85 and 1.955 respectively. Is the result of experiment significant?

Sol.

Given $\bar{x}$ = 17.85, n = 14, μ = 18.5, S.D. (s) = 1.955

Step 1:Null Hypothesis : $H_0 : \mu = 18.5$

Step 2:Alternative Hypothesis : $\mu \neq 18.5$ (two – tailed test)

Step 3:Level of Significance: α = 0.05

Step 4:Test Statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n-1}} = \frac{17.85 - 18.5}{1.955/\sqrt{14-1}} = -1.199$$

Step 5:Conclusion: t_α = 2.16 at 5% level of significance with n-1 = 13 d.f. from t-distribution table.

If $|t_{cal}| < t_\alpha$ accept H_0

i.e. the result of the experiment is significant.

2(H.W.). A mechanist is making engine parts with axle diameters of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a S.D. of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specification at 0.05 level of significance.

Ans. t_{cal} = 3.15, t_α = 2.26, Reject H_0

3. A random sample of 10 boys had the following I.Q's : 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100

(a) Do these data support the assumption of a population mean I.Q. of 100?

(b) Find a reasonable range in which most of the mean I.Q. Values of samples of 10 boys lie.

Sol.

(a) Here S.D. and mean of sample is not given directly.

We have to determine these S.D. and mean as follows.

Mean $\bar{x} = \frac{\sum x}{n} = \frac{972}{10} = 97.2$

x	x- $\bar{x}$	(x- $\bar{x}$)^2
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
972		1833.60

We know that $S^2 = \frac{1}{n-1}\sum_i \left(x_i - \bar{x}\right) = \frac{1833.60}{9} = 203.73$

$\therefore$ S.D. $S = \sqrt{203.73} = 14.27$

Step 1:Null Hypothesis : The data support the assumption of a population mean I.Q. of 100 in the population. *i.e.* H_0 : $\mu = 100$

Step 2:Alternative Hypothesis : $\mu \neq 100$ (two – tailed test)

Step 3:Level of Significance: α = 0.05

Step 4:Test Statistic:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{97.2 - 100}{14.27/\sqrt{10}} = -0.62$$

Step 5:Conclusion: t_α = 2.26 at 5% level of significance with n-1 = 10-1 = 9 d.f. from t-distribution table.

If $|t_{cal}| < t_\alpha$ accept H_0

i.e. The data support the assumption of a population mean I.Q. of 100 in the population.

(b) The 95% confidence limits are given by

$$\left(\bar{x}-t_\alpha \frac{s}{\sqrt{n}}, \bar{x}+t_\alpha \frac{s}{\sqrt{n}}\right)=(97.2-2.26\times 4.512,\ 97.2+2.26\times 4.512)=(87,\ 107.40)$$

(II) Test for Difference of means

Let $\bar{x}$ and $\bar{y}$ be the means of the two small samples of sizes n_1 and n_2 drawn from two normal populations having means μ_1 and μ_2. s_1^2 and s_2^2 are the variances of two samples. To test whether the two population means are equal.

Step 1:Null Hypothesis : H_0 : $\mu_1 = \mu_2$

Step 2:Alternative Hypothesis : H_1 : $\mu_1 \neq \mu_2$ (two – tailed test)

Step 3:Level of Significance: Select the appropriate level of significance α

Step 4:Test Statistic: Calculate the test statistic

$$t=\frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Where $S^2=\frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2-2}$

Step 5:Conclusion: Find the critical value t_α at the level of significance α with n_1+n_2-2 d.f. from t-distribution table.

If $|t_{cal}| < t_\alpha$ accept H_0

If $|t_{Cal}| > t_\alpha$ reject H_0 and accept H_1

1. Samples of two types of electric light bulbs were tested for length of life and following data were obtained

Type I	Type II
Sample size n_1 = 8	n_2 = 7
Sample mean $\bar{x}$ = 1234 hours	$\bar{y}$ = 1036 hrs
Sample S.D. s_1 = 36 hrs	s_2 = 40 hrs

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life.

Sol.

Step 1:Null Hypothesis : H_0 : $\mu_1 = \mu_2$

Step 2:Alternative Hypothesis : H_1 : $\mu_1 \neq \mu_2$ (two – tailed test)

Step 3:Level of Significance: α = 5%

Step 4:Test Statistic: $S^2=\frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2-2}=\frac{8(36)^2+7(40)^2}{8+7-2}=1659.08$

$$t=\frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}=\frac{1234-1036}{\sqrt{1659\left(\frac{1}{8}+\frac{1}{7}\right)}}=9.39$$

Step 5:Conclusion: $t_\alpha = 2.16$ at 5% level of significance with $n_1+n_2-2 = 8+7-2 = 13$ d.f. from t-distribution table.

If $|t_{cal}| > t_\alpha$ Reject H_0 and accept H_1

i.e. The two types I and II of electric bulbs are not identical.

2(H.W.). The means of two random samples of sizes 9 and 7 are 196.42 and 198.82 respectively. The sum of the squares of the deviations from the mean are 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population.

Ans. $t_{cal} = -2.63$, $t_\alpha = 2.63$, Reject H_0

3. Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results.

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether the two horses have the same running capacity.

Sol.

Given $n_1 = 7$, $n_2 = 6$

We first compute the sample means and S.D.

$\bar{x}$ = mean of first sample = $\frac{1}{7}(28+30+32+33+33+29+34) = 31.286$

$\bar{y}$ = mean of second sample = $\frac{1}{6}(29+30+30+24+27+29) = 28.16$

x	$(x-\bar{x})$	$(x-\bar{x})^2$	y	$(y-\bar{y})$	$(y-\bar{y})^2$
28	-3.286	10.8	29	0.84	0.7056
30	-1.286	1.6538	30	1.84	3.3856
32	0.714	0.51	30	1.84	3.3856
33	1.714	2.94	24	-4.16	17.3056
33	1.714	2.94	27	-1.16	1.3456
29	-2.286	5.226	29	0.84	0.7056
34	2.714	7.366			
219		31.4358	169		26.8336

Now $S^2 = \frac{1}{n_1+n_2-2}\left[\sum\left(x_i-\bar{x}\right)^2+\sum\left(x_i-\bar{x}\right)^2\right] = \frac{31.4358+26.8336}{11} = 5.23$

Step 1:Null Hypothesis : H_0 : $\mu_1 = \mu_2$

Step 2:Alternative Hypothesis : H_1 : $\mu_1 \neq \mu_2$ (two – tailed test)

Step 3:Level of Significance: α = 5%

Step 4:Test Statistic:

$$t = \frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{31.286-28.16}{\sqrt{5.23\left(\frac{1}{7}+\frac{1}{6}\right)}} = 2.443$$

Step 5:Conclusion: $t_\alpha = 2.2$ at 5% level of significance with $n_1+n_2-2 = 6+7-2 = 11$d.f. from t-distribution table.

If $|t_{cal}| > t_\alpha$ Reject H_0 and accept H_1

i.e. both horses A and B do not have the same running capacity.

SNEDECOR'S F – TEST OF SIGINIFICANCE

F – test is used to test whether there is any significance difference between the variances of 2 samples.

Let two independent random samples of sizes n_1 and n_2 be drawn from two normal populations.

To test the hypothesis that the two population variances σ_1^2 and σ_2^2 are equal.

Step 1:Null Hypothesis : Define or set up a Null Hypothesis H_0

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

Step 2:Alternative Hypothesis : Set up an Alternative Hypothesis H_1

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \text{ (two – tailed test)}$$

Step 3:Level of Significance: Select the appropriate level of significance α

Step 4:Test Statistic:

$$S_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{\sum\left(x - \bar{x}\right)^2}{n_1 - 1} \text{ and } S_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{\sum\left(y - \bar{y}\right)^2}{n_2 - 1}$$

Where s_1^2 and s_2^2 are the variances of the tow samples.

Calculate the test statistic $F = \frac{S_1^2}{S_2^2} \, or \, F = \frac{S_2^2}{S_1^2}$ according as $S_1^2 > S_2^2 \, or \, S_2^2 > S_1^2$

Step 5:Conclusion: Find the critical value F_α at the level of significance α with $\nu_1 = n_1 - 1 \, and \, \nu_2 = n_2 - 1$ d.f. from F-distribution table.

If $|F_{cal}| < t_\alpha$ accept H_0

If $|F_{Cal}| > t_\alpha$ reject H_0 and accept H_1

1. In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have the same variance.

Sol.

Given $n_1 = 8$, $n_2 = 10$, $\sum\left(x_i - \bar{x}\right)^2 = 84.4, \sum\left(y_i - \bar{y}\right)^2 = 102.6$

$$S_1^2 = \frac{\sum\left(x - \bar{x}\right)^2}{n_1 - 1} = \frac{84.4}{7} = 12.057 \text{ and } S_2^2 = \frac{\sum\left(y - \bar{y}\right)^2}{n_2 - 1} = \frac{102.6}{9} = 11.4$$

Step 1:Null Hypothesis : $H_0 : \sigma_1^2 = \sigma_2^2$

Step 2:Alternative Hypothesis : $H_1 : \sigma_1^2 \neq \sigma_2^2$ (two – tailed test)

Step 3:Level of Significance: α = 5%

Step 4:Test Statistic:

$$F = \frac{S_1^2}{S_2^2} = \frac{12.057}{11.4} = 1.057$$

Follows F – distribution with (n_1-1, n_2-2) degrees of freedom

Step 5:Conclusion: $F_{0.05}\left(7,9\right) = 3.29$

$|F_{cal}| < t_\alpha$ accept H_0

i.e. the populations have the same variance.

2(H.W.). Two random samples reveal the following results :

Sample	Size	Sample Mean	Sum of Squares of deviations from the Mean
1	10	15	90
2	12	14	108

Test whether the samples came from the same normal population with equal variances.
Ans. F_{cal} = 1.018, $F_{0.05}(9,11) = 2.90$, AcceptH_0

χ^2 TEST AS A TEST OF GOODNESS OF FIT

We use this test to decide whether the discrepancy between theory and experiment is significant or not.

Test Statistic:

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

Conclusion: If the calculated value of χ^2 > tabulated value of χ^2 at α level, the Null Hypothesis H_0 is rejected. Otherwise H_0 is accepted

1. The number of automobile accidents per week in a certain community are as follows: 12, 8, 20, 2, 14,10,15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.
Sol.

Expected frequency of accidents each week = $\frac{100}{10} = 10$

Step 1:Null Hypothesis : H_0 : The accident conditions were the same during the 10 week period.
Step 2:Alternative Hypothesis : H_1 : The accident conditions were the not same during the 10 week period.
Step 3:Level of Significance: α = 5%
Step 4:Test Statistic:

Observed Frequency (O_i)	**Expected Frequency (E_i)**	$(O_i - E_i)$	$\frac{(O_i - E_i)^2}{E_i}$
12	10	2	0.4
8	10	-2	0.4
20	10	10	10.0
2	10	-8	6.4
14	10	4	1.6
10	10	0	0.0
15	10	5	2.5
6	10	-4	1.6
9	10	-1	0.1
4	10	-6	3.6
100	100		26.6

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right] = 26.6$$

Step 5:Conclusion: $\chi^2_\alpha = 16.9$ at 5% level of significance at n-1 = 10-1 = 9 d.f.

$|\chi^2_{cal}| > \chi^2_\alpha$ Reject H_0

i.e. the accident conditions were not the same during the 10 week period

2. A sample analysis of examination results of 500 student was made. It was found that 220 students had failed, 170 had secured a third class, 90 were placed in second class and 20 got a first class. Do these figures commensurate with the general examination result which is in the ratio of 4 : 3 : 2 : 1 for the various categories respectively.

Sol.

Step 1:Null Hypothesis : H_0 : The observed results commensurate with the general examination results.

Step 2:Alternative Hypothesis : H_1 : The observed results not commensurate with the general examination results.

Step 3:Level of Significance: α = 5%

Step 4:Test Statistic:

Expected frequencies are in the ratio of 4 : 3 : 2 : 1

Total frequency = 500

If we divide the total frequency 500 in the ratio 4 : 3 : 2 : 1, we get the expected frequencies as 200, 150, 100, 50

Observed Frequency (O_i)	Expected Frequency (E_i)	$(O_i - E_i)$	$\frac{(O_i-E_i)^2}{E_i}$
220	200	20	2.00
170	150	20	2.667
90	100	-10	1.000
20	50	-30	18.00
500	500		23.667

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i-E_i)^2}{E_i}\right] = 23.667$$

Step 5:Conclusion: $\chi^2_\alpha = 7.81$ at 5% level of significance at n-1 = 4-1 = 3 d.f.

$|\chi^2_{cal}| > \chi^2_\alpha$ Reject H_0

i.e. The observed results not commensurate with the general examination results.

3.(H.W). A die is thrown 264 times with the following results. Show that the die is biased

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	28	58	54	52

Ans. $\chi^2_\alpha = 11.07$, $\chi^2 = 17.6362$ Reject H_0