Probability & Statistics [MA 1006].

Unit - III - Test of Hypothesis

1. When we attempt to make decisions about the population on the basis of sample information, we've to make assumptions or guesses about the nature of the population involved or about the value of some parameter of the population. Such assumptions, which may or may not be true, are called statistical hypothesis.

A hypothesis which assumes that there is no significant diff between the sample statistic and the corresponding population parameter (or) between two sample statistics. Such a hypothesis of no difference is called a null hypothesis and is denoted by H_0.

A hypothesis that is different from complementary, the null hypothesis is called an alternative hypothesis and is denoted by H_1.

A procedure for deciding whether to accept (or) to reject a null hypothesis (to reject respectively) (or) to accept the alternative hypothesis is called the test of hypothesis.

Errors in Hypothesis Testing:-

The probability of rejecting a null hypothesis, when it is true, becomes greater. The error committed in rejecting H_0, when it is really true is called type-I error.

Type-I errors is also known as producer's risk.

The error committed in accepting H_0, when it is false, is called type-II error. type-II errors is also known is also known as Consumers' risk.

One-Tailed and Two-Tailed tests:-

If θ_0 is a population parameter and θ is the corresponding sample statistic and if we set up the null hypothesis $H_0 :- \theta = \theta_0$, then the alternative hypothesis which is complementary to H_0 can be any one of the following.

(i) $H_1 :- \theta \neq \theta_0$ i.e., $\theta > \theta_0$ (or) $\theta < \theta_0$.

(ii) $H_1 :- \theta > \theta_0$

(iii) $H_1 :- \theta < \theta_0$.

If H_1 is given, (i) is called two-tailed test.

If H_1 is given (ii) is called right-Tailed test.

If H_1 is given (iii) is called left-tailed test.

When H_0 is tested, while H_1 is a One-Tailed alternative, the test of hypothesis is called a one-tailed test.

When H_0 is tested, while H_1 is a Two-tailed alternative, the test of hypothesis is called a two-tailed test.

Level of Significance (LOS):-

Nature of Test	1 % LOS (0.01)	2 % LOS 0.02	5 % LOS 0.05	10 % LOS 0.1
Two-tailed	$\lvert Z_\alpha \rvert = 2.58$.	$\lvert Z_\alpha \rvert = 2.33$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right-tailed	$Z_\alpha = 2.33$	$Z_\alpha = 2.055$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
Left-tailed	$Z_\alpha = -2.33$	$Z_\alpha = -2.055$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$.

Procedure for testing of Hypothesis:-

1. Null hypothesis H_0 is defined.

2. Alternative hypothesis H_1 is defined.

also nature of the test is (whether One-tailed or two tailed) decided.

3. Let LOS α is fixed, and Z_α is noted.

4. Test-Statistic Z is computed.

5. Comparison is made between $\lvert Z \rvert$ and $\lvert Z_\alpha \rvert$.

5. If $|Z| < |Z_\alpha|$, H_0 is accepted and H_1 is rejected. i.e., It is concluded that the diff between t and $E(t)$ is not significant at $\alpha\%$ LOS.

If $|Z| > |Z_\alpha|$, H_0 is rejected and H_1 is accepted. i.e., It is concluded that the diff between t and $E(t)$ is significant at $\alpha\%$ LOS.

Tests of Significance for Large Samples:-

test-I:- Test of Significance of the diff between sample proportion and population proportion

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

$p \to$ sample proportion.
$P \to$ population proportion.
$Q \to 1-P$.
$n \to$ no of trials.

If $|Z| \leq Z_\alpha$, the diff between the sample proportion p and the population proportion P is not significant at $\alpha\%$ LOS.

Note:- 95% Confidence limits for P are then

$$\boxed{p - 1.96\sqrt{\frac{pq}{n}},\ p + 1.96\sqrt{\frac{pq}{n}}}$$

1. The fatality rate of typhoid patients is believed to be 17.26 %. In a certain year, 640 patients suffering from typhoid were treated in a metropolitan hospital and only 63 patients died. Can you consider the hospital efficient?

<u>Soln:-</u>

H_0 : $p = P$ ∴ the hospital is not efficient.

H_1 : $p < P$.

One-tailed test is to be used.

Let us assume that LOS = 1 %. ∴ $z_\alpha = -2.33$

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} \quad , \quad p = \frac{63}{640} = 0.0984$$

and $P = 0.1726$ and

$Q = 1 - P = 0.8274$.

$$\therefore z = \frac{0.0984 - 0.1726}{\sqrt{\frac{0.1726 \times 0.8274}{640}}} = -4.96$$

$|z| > |z_\alpha|$.

∴ The diff between p and P is significant.

∴ H_0 is rejected and H_1 is accepted.

i.e., The hospital is efficient in bringing down the fatality rate of typhoid patients.

2. Experience has shown that 20% of a manufactured Product is top quality. In one day's Production of 400 articles, only 50 are of top quality. Show that either the production of the day chosen was not a representative sample or the hypothesis of 20% was wrong. Based on the particular's day's production, find also the 95% confidence limits for the percentage of top quality product.

Soln:- H_0 :- $P = \frac{1}{5}$ {i.e., 20% of the products manufactured is of top Quality.}

H_1 :- $P \neq \frac{1}{5}$. {Two-tailed test is to be Used}

p :- proportion of top Quality Products in the sample.

$p = \frac{50}{400} = \frac{1}{8}$. $\{\because q = 1 - \frac{1}{8} = 7/8\}$.

Let us assume that, LOS = 5% $\therefore z_\alpha = 1.96$.

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{1}{8} - \frac{1}{5}}{\sqrt{\frac{1}{5} \times \frac{4}{5} \times \frac{1}{400}}} = -\frac{3}{40} \times 50 = -3.75$$

$\therefore |z| = 3.75 > 1.96$

$\therefore$ The diff between p and P is significant at 5% LOS.

Also, Ho is rejected, i.e., the production of the particular day chosed is not a representative sample.

To find the 95% Confidence limits:-

$$\frac{p - P}{\sqrt{\frac{pq}{n}}} \leq 1.96 = p - \sqrt{\frac{pq}{n}} \times 1.96 \leq P \leq p + \sqrt{\frac{pq}{n}} \times 1.96$$

i.e., $0.125 - \sqrt{\frac{1}{8} \times \frac{7}{8} \times \frac{1}{400}} \times 1.96 \leq P \leq 0.125 + \sqrt{\frac{1}{8} \times \frac{7}{8} \times \frac{1}{400}} \times 1.96$

i.e., $0.093 \leq P \leq 0.157$

$\therefore$ 95% confidence limits for the percentage of top quality are 9.3 and 15.7.

---

3. The salesman in a departmental store claims that, almost 60% of the shoppers entering the store leaves without making a purchase. A random sample of 50 shoppers showed that 35 of them left without making a purchase. Are there sample results consistant with the claim of the salesman? Use a LoS of 0.05.

Soln:- Let P and p denote the population and same proportions of shoppers not making a purchase.

$H_0 :: p = P$.

$H_1 :: p > P$ $\{ \because p = 0.7$ and $P = 0.6 \}$.

$\{$one-tailed test is to be used.$\}$

$\therefore$ LoS $= 5\%$ and $z_\alpha = 1.645$.

$$z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.7 - 0.6}{\sqrt{\frac{0.6 \times 0.4}{50}}} = 1.443$$

i.e., $|z| < |z_\alpha|$

Conclusion:- $\therefore$ H_0 is accepted and H_1 is rejected.

i.e., The diff between p and P is not significant at 5% LoS.

$\therefore$ The sample results are consistant with the claim of the salesman.$\}$

Test of Significance of the diff between two sample proportion :-

Let p_1 and p_2 be the proportion of successes in two large samples of sizes n_1 and n_2 resply drawn from the same population (or) from two population with the same proportion P.

$\therefore$ Test Statistic $$z = \frac{p_1 - p_2}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}.$$

If P is not known, an Unbiased estimate of P based on both samples given by.

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, \quad Q = 1 - P.$$

If $|z| \le |z_\alpha|$ $\therefore$ the diff between the two sample proportions p_1 and p_2 is not significant at α % of LOS.

1. In a large city A, 20 % of a random sample of 900 school boys had a slight physical defect. In another large city B, 18.5 % of a random sample of 1600 school boys had the ~~skarab~~ defect. Is the diff between the proportion significant?

Soln:- Given $p_1 = 0.2$, $p_2 = 0.185$, $n_1 = 900$ and $n_2 = 1600$.

Soln:-

$H_0 :: p_1 = p_2$. $\{\because p_1 = 0.2,\ p_2 = 0.185,$

$H_1 :: p_1 \neq p_2$ $\quad n_1 = 900,\ n_2 = 1600\}$.

$\therefore$ Two-tailed test is to be used.

LOS = 5% $\Rightarrow z_\alpha = 1.96$

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

Since, the population proportion P is not given,

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{180 + 296}{900 + 1600} = 0.1904$$

$$\therefore Q = 1 - P = 1 - 0.1904 = 0.8096$$

Sub in (1)

$$z = \frac{0.2 - 0.185}{\sqrt{0.1904 \times 0.8096 \left(\frac{1}{900} + \frac{1}{1600}\right)}}$$

after simplifying, $\boxed{z = 0.92}$.

$\therefore |z| \le |z_\alpha|$

$\therefore$ The the diff between p_1 and p_2 is not significant at 5% LOS.

2. Before an increase in excise duty on tea, 800 people out of sample of 1000 were consumers of tea. After the increase in duty, 800 people were consumers of tea in a sample of 1200 persons. Find whether there is significant decrease in the consuption of tea after the increase in duty.

Soln:- Let p_1 and p_2 be the proportion of the consumers before and after the increase in duty resply.

$$p_1 = \frac{800}{1000} = \frac{4}{5}.$$

$$p_2 = \frac{800}{1200} = \frac{2}{3}$$

H_0:- $p_1 = p_2$.

H_1 :- $p_1 > p_2$. ∴ One-tailed test is to be used.

LOS = 1% = 2.33.

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

The population proportion is not given.

$$\therefore P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{800+800}{2200} = 0.7273.$$

$$Q = 1 - P = 1 - 0.7273 = 0.2727.$$

$$= \frac{0.8 - 0.67}{\sqrt{0.7273 \times 0.2787\left(\frac{1}{1000}+\frac{1}{1200}\right)}}$$

$$= \frac{0.13 \times \sqrt{1000 \times 1200}}{\sqrt{0.7273 \times 0.2727 \times 2200}} = 6.82.$$

∴ $|z| > |z_\alpha|$ => H_0 is rejected and H_1 is accepted.

∴ The diff between p_1 and p_2 is significant at 1% LOS.

There is significant decrease in the consumption of tea after the increase in duty.

15.5 % of a random sample of 1600 undergraduates were smokers, whereas 20% of random sample of 900 postgraduates were smokers in a State. Can we conclude that less number of undergraduates are smokers than the postgraduates?

Soln:- $p_1 = 0.1555$ and $p_2 = 0.2$, $n_1 = 1600$, $n_2 = 900$.

H_0 :- $p_1 = p_2$.

H_1 :- $p_1 < p_2$

One-tailed test is to be used.

$\therefore$ LOS = 5 % = -1.645.

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad \text{where } P = \frac{n_1 p_1 + n_2 p_2}{n_1+n_2} = 0.1712$$

$$Q = 1 - 0.1712 = 0.8288$$

$$\therefore z = \frac{0.155 - 0.2}{\sqrt{0.1712 \times 0.8288\left(\frac{1}{1600}+\frac{1}{900}\right)}} = \frac{-0.045 \times 1200}{\sqrt{0.1712 \times 0.8288 \times 2500}}$$

$$= -2.87$$

$$|z| > |z_\alpha|$$

$\therefore$ The diff between p_1 and p_2 is significant

i.e., H_0 is rejected and H_1 is accepted.

$\therefore$ The habit of smoking is less among the undergraduates than among the postgraduates.

Test - III

Test of significance of the difference between sample mean($\bar{x}$), and population mean (μ).

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

where, $\bar{x}$ - sample Mean.
μ - Population Mean.
σ - variance.
n - no of sample size.

Note :- 95 % confidance limits for μ are given by $\frac{|\mu - \bar{x}|}{\sigma/\sqrt{n}} \leq 1.96$

ie., $\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$.

A sample of 100 students is taken from a large population. The mean height of the student in this sample is 160 cm. Can it be reasonable regarded that, in the population, the mean heights is 165 cm, and the S.D is 10 cm?

Soln:- $\bar{x} = 160$, $n = 100$, $\mu = 165$, $\sigma = 10$.

H_0 :- $\bar{x} = \mu$ {the diff between $\bar{x}$ and μ is not significant}.

H_1 :- $\bar{x} \neq \mu$ {Two-tailed Test is to used}

Let LOS is 1% = 2.58.

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{160 - 165}{\frac{10}{\sqrt{100}}} = -5.$$

$|Z| > Z_\alpha$.
$\therefore H_0$ is rejected.

Hence, the diff between $\bar{x}$ and μ is significant at 1% LOS. it is not statistically correct to assume that $\mu = 165$.

2. The mean breaking strength of the cables supplied by a manufacturer is 1800 with a S.D of 100. By a new technique in the manufacturing process, it is claimed that the breaking strength of the cable has increased. In order to test this claim, a sample of 50 cables is tested and it is found that the mean breaking strength is 1850. Can we support the claim at 1% LOS?

Soln:- $\bar{x} = 1850$, $n = 50$, $\mu = 1800$, $\sigma = 100$.

H_0 :- $\bar{x} = \mu$.

H_1 :- $\bar{x} > \mu$.

∴ One-Tailed test is to be used.

ie, LOS = 1% ⇒ $z_\alpha = 2.33$.

$$\therefore z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1850 - 1800}{\frac{100}{\sqrt{50}}} = 3.54.$$

$|z| > |z_\alpha|$

∴ The diff between $\bar{x}$ and μ is significant at 1% LOS.

ie, H_0 is rejected and H_1 is accepted.

∴ Based on the sample data, we may support the claim of increas in breaking strength.

The mean value of a random sample of 60 items was found to be 145 with a S.D of 40. Find the 95% confidence limits for the population mean. What size of the sample is required to estimate the population mean within five of it's actual value 95% or more confidence, using the sample mean?

Soln:- 95% confidence limits for μ are given by

$$\frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq 1.96$$

Since, the population S.D (σ) is not given, we have to use $\frac{\mu - \bar{x}}{\frac{s}{\sqrt{n}}} \leq 1.96$.

ie., $\bar{x} - 1.96 \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{s}{\sqrt{n}}$.

$$145 - 1.96 \frac{\times 40}{\sqrt{60}} \leq \mu \leq 145 + \frac{1.96 \times 40}{\sqrt{60}}$$

$$134.9 \leq \mu \leq 155.1$$

Test - IV : Test of Significance of the difference between the Means of two samples :-

Let $\bar{x}_1$ and $\bar{x}_2$ be the Mean of two large Samples of sizes n_1 and n_2 drawn from two populations with the same mean μ and variances σ_1^2 and σ_2^2 resply.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{(or)} \quad \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \longrightarrow (1)$$

{$\because$ in place of σ_1 by s_1 and σ_2 by s_2}.

Note :- (1) If the samples are drawn from the same population, $\sigma_1 = \sigma_2 = \sigma$ then.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \longrightarrow (2)$$

Note :- (2) If σ_1 and σ_2 are not known and $\sigma_1 = \sigma_2 = \sigma$ is approximated by

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}. \quad \text{sub in (2).}$$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}.$$

after simplifying,

$$\boxed{z = \frac{\bar{x}_1 - \bar{x}_2}{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} \longrightarrow (3)$$

1. A simple sample of heights of 6400 English men has a mean of 170 cm and a S.D of 6.4 cm, while a simple sample of heights of 1600 Americans has a mean of 172 cm and S.D of 6.3 cm. Do the data indicate that Americans are on the average, taller than the Englishmen?

Soln:- $n_1 = 6400$, $\bar{x}_1 = 170$ and $S_1 = 6.4$

$n_2 = 1600$, $\bar{x}_2 = 172$ and $S_2 = 6.3$

H_0:- $\mu_1 = \mu_2$ (or) $\bar{x}_1 = \bar{x}_2$.

ie., the samples have drawn from Two different populations with the same mean.

H_1:- $\bar{x}_1 < \bar{x}_2$ (or) $\mu_1 < \mu_2$ {∵ left-tailed test is to be used}.

∴ LOS be 1% $\Rightarrow Z_\alpha = -2.33$

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{170 - 172}{\sqrt{\frac{(6.4)^2}{6400} + \frac{(6.3)^2}{1600}}} = -11.32$$

$\therefore |Z| = 11.32$.

$|Z| > |Z_\alpha|$

ie., the diff between $\bar{x}$ and $\bar{x}_2$ are significant at 1% LOS.

∴ H_0 is rejected and H_1 is accepted.

$\Rightarrow$ The Americans are on the average, taller than the Englishmen.

Test the significance of the difference between the means of the samples, draw from two normal population with the same S.D from the following data.

	Size	Mean	S.D.
Sample 1	100	61	4
Sample 2	200	63	6

H_0 :- $\bar{x}_1 = \bar{x}_2$ (or) $\mu_1 = \mu_2$.

H_1 :- $\bar{x}_1 \neq \bar{x}_2$ (or) $\mu_1 \neq \mu_2$.

$\therefore$ { Two-tailed test is to be used. }

$\therefore$ LOS $= z_\alpha = 5\% = 1.96.$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{61-63}{\sqrt{\frac{4^2}{200} + \frac{6^2}{100}}} = -3.02.$$

$|z| = 3.02 \Rightarrow |z| > |z_\alpha|.$

i.e., The difference between $\bar{x}_1$ and and $\bar{x}_2$ is significant at 5% Level.

$\therefore$ H_0 is rejected and H_1 is accepted.

i.e., The two normal population, from which the samples are drawn may not have the same mean, though they may have the same S.D.

In a random sample of size 500, the mean is found to be 20, In another independent sample of size 400, the mean is 15, Could the samples have been drawn from the same population with S.D 4?

Soln:- Given $n_1 = 500$, $n_2 = 400$, $\bar{x}_1 = 20$, $\bar{x}_2 = 15$, $\sigma = 4$.

H_0:- $\bar{x}_1 = \bar{x}_2$. ie., {Samples have been from same population}

H_1:- $\bar{x}_1 \neq \bar{x}_2$

∴ Two-tailed test is to be Used.

∴ LOS = Z_α 1% = 2.58.

$$\therefore Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{20-15}{4\sqrt{\frac{1}{500}+\frac{1}{400}}} = 18.6$$

ie., $|Z| > |Z_\alpha|$ ∴ H_0 is rejected.

⇒ the diff between $\bar{x}_1$ and $\bar{x}_2$ is significant at 1% LOS.

∴ The samples have not been drawn from the same population.

4. The average marks scored by 32 boys is 72 with S.D of 8, while that for 36 girls is 70 with S.D of 6. Test at 1% LOS whether the boys perform better than girls.

Soln:- H_0: $\bar{x}_1 = \bar{x}_2$ or $\{\mu_1 = \mu_2\}$

H_1: $\bar{x}_1 > \bar{x}_2$ ∴ Right-tailed is to be used.

LOS = 1% ⇒ $z_\alpha = 2.33$.

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

$$= \frac{72-70}{\sqrt{\frac{8^2}{32} + \frac{6^2}{36}}} = 1.154$$

∴ $|z| < |z_\alpha|$

ie., The diff between $\bar{x}_1$ and $\bar{x}_2$ is not significant at 1% LOS.

∴ H_0 is accepted and H_1 is rejected.

ie., statistically, we cannot conclude that the boys perform better than girls.