Test of significance for proportion

Test of significance of single proportion

- To test the significant difference between the sample proportion p and the Population proportion P we use the statistic
- $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$, n = sample size
- The formulated Null and Alternative hypothesis is, $H_0: P =$a specified value
- $H_1: P \neq$ a specified value

1. A coin is tossed 400 times and it turns up head 216 times. Discuss whether the coin may be regarded as unbiased one.

Solution:

Set the null hypothesis $H_0: P = \frac{1}{2}$

Set the alternative hypothesis $H_1: P \neq \frac{1}{2}$

Level of significance α= 0.05(5%)

The test Statistic $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$

Given $P = \frac{1}{2}, \; Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$

$$p = \frac{216}{400} = 0.54, \; n = 400$$

$$\Rightarrow Z = \frac{0.51 \; - 0.5}{\sqrt{\frac{0.5 \times 0.5}{400}}} = 1.6$$

Critical value: At 5% level, the tabulated value of Z_α is 1.96.

Conclusion: Since |Z| = 1.6 $<$ 1.96

Hence Null Hypothesis H_0 is accepted at 5% level of significance.

Hence the coin may be regarded as unbiased

2. In a city of sample of 500 people, 280 are tea drinkers and the rest are coffee drinkers. Can we assume that both coffee and tea are equally popular in this city at 5% Los.

Solution:

Set the null hypothesis $H_0: P = \frac{1}{2}$ (both coffee and tea drinkers are equally popular)

Set the alternative hypothesis $H_1: P \neq \frac{1}{2}$

Level of significance α= 0.05(5%)

The test Statistic $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$

Given $P = \frac{1}{2}$, $Q = 1 - P = 1 - \frac{1}{2} = \frac{1}{2}$

$$p = \frac{280}{500} = 0.56,\ n = 500$$

$$\Rightarrow Z = \frac{0.56 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{500}}} = 2.68$$

Critical value: At 5% level, the tabulated value of Z_α is 1.96.

Conclusion: Since |Z| = 2.68 > 1.96

Hence Null Hypothesis H_0 is rejected at 5% level of significance.

Hence both type of drinkers are not popular

3. **A manufacturing company claims that atleast 95% of its products supplied confirm to the specifications out of a sample of 200 products, 18 are defective. Test the claim at 5% Los.**

Solution:

Set the null hypothesis $H_0: P = 0.95$

Set the alternative hypothesis $H_1: P \neq 0.95$

Level of significance α= 0.05(5%)

The test Statistic $Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$

Given $P = 0.95,\ Q = 1 - P = 1 - 0.95 = 0.05$

$$p = \frac{200 - 18}{200} = 0.91,\ n = 200$$

$$\Rightarrow Z = \frac{0.91 - 0.95}{\sqrt{\frac{0.95 \times 0.05}{200}}} = -2.595$$

Critical value: At 5% level, the tabulated value of Z_α is 1.96.

Conclusion: Since |Z| = 2.595 > 1.96

Hence Null Hypothesis H_0 is rejected at 5% level of significance.

Test of significance of Difference between two sample proportion

- To test the significance of the difference between the sample proportions p_1 and p_2 we use the statistic
- $Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$
- where $P + Q = 1 \Rightarrow Q = 1 - P$, n_1, n_2 = sample sizes
- The formulated Null and Alternative hypothesis is,
- $H_0: P_1 = P_2$
- $H_1: P_1 \neq P_2$
- If P is not known, then
- $P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

1. **If a sample of 300 units of a manufactured product 65 units were found to be defective and in another sample of 200 units, there were 35 defectives. Is there significant difference in the proportion of defectives in the samples at 5% Los.**
   Solution:

   Set the null hypothesis $\mathrm{H_0: P_1 = P_2}$

   Set the alternative hypothesis $\mathrm{H_1: P_1 \neq P_2}$

   Level of significance α= 0.05(5%)

   The test statistic $Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$

   Given $n_1 = 300,\ n_2 = 200, p_1 = \frac{65}{300} = 0.2166,\ p_2 = \frac{35}{200} = 0.175$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(300)0.2166 + (200)0.175}{300 + 200} = 0.0799$$

$$\Rightarrow Q = 1 - P = 1 - 0.0799 = 0.9201$$

$$\Rightarrow Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.2166 - 0.175}{\sqrt{(0.0799)(0.9201)\ \left(\frac{1}{300}+\frac{1}{200}\right)}} = 1.233$$

   Critical value: At 5% level, the tabulated value of Z_α is 1.96.

   Conclusion: Since |Z| = 1.2333 < 1.96

   Hence Null Hypothesis H_0 is accepted at 5% level of significance.

   The difference in the proportion of defectives in the samples is not significant.

2. **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 same defect. Is the difference between the proportions significant?**
   Solution:

Set the null hypothesis $H_0: P_1 = P_2$

Set the alternative hypothesis $H_1: P_1 \neq P_2$

Level of significance α= 0.05(5%)

The test statistic $Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$

Given $n_1 = 900,\ n_2 = 1600, p_1 = \frac{20}{100} = 0.2,\ p_2 = \frac{18.5}{100} = 0.185$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(900)0.2 + (1600)0.185}{900 + 1600} = 0.1904$$

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

$$\Rightarrow Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.2 - 0.185}{\sqrt{(0.1904)(0.8096)\ \left(\frac{1}{900} + \frac{1}{1600}\right)}} = 0.9169$$

Critical value: At 5% level, the tabulated value of Z_α is 1.96.

Conclusion: Since |Z| = 0.9169 < 1.96

Hence Null Hypothesis H_0 is accepted at 5% level of significance.

Hence there is no significant difference.

3. **Before an increase in excise duty on tea, 800 persons out of a sample of 1000 persons were found to be tea drinkers. After an increase is excise duty. 800 people were tea drinkers in a sample of 1200 people. Test whether there is a significant decrease in the consumption of tea after the increase in excise duty at 5% Los**

Solution:

Set the null hypothesis $H_0: P_1 = P_2$

Set the alternative hypothesis $H_1: P_1 > P_2$

Level of significance α= 0.05(5%)

The test statistic $Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$

Given $n_1 = 1000,\ n_2 = 1200, p_1 = \frac{800}{1000} = 0.8,\ p_2 = \frac{800}{1200} = 0.667$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(1000)0.8 + (1200)0.667}{1000 + 1200} = 0.7272$$

$$\Rightarrow Q = 1 - P = 1 - 0.7272 = 0.2728$$

$$\Rightarrow Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.8 - 0.667}{\sqrt{(0.7272)(0.2728)\ \left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 6.88$$

Critical value: At 5% level, the tabulated value of Z_α is 1.645.

Conclusion: Since |Z| = 6.88 $>$ 1.645

Hence Null Hypothesis H_0 is rejected at 5% level of significance.

There is a significance decrease in the consumption of tea due to increase in excise duty.

F-test for Variance Ratio Test

F – TEST

- This test is used to test the significance of two or more sample estimates of Population variance.
- To test whether if there is any significant difference between two estimates of population variance.
- To test if the two sample have come from the same population, we use F – test.
- The test statistic is given by, $F = \frac{S_1{}^2}{S_2{}^2}$
- Where $S_1{}^2 = \frac{n_1 s_1{}^2}{n_1 - 1}$ (n_1 − first sample size)
- $S_2{}^2 = \frac{n_2 s_2{}^2}{n_2 - 1}$ (n_2 − second sample size)
- And $S_1{}^2 > S_2{}^2$
- The degrees of freedom $(v_1, v_2) = (n_1 - 1, n_2 - 1)$
- Note: If $S_2{}^2 > S_1{}^2$ then $F = \frac{S_2{}^2}{S_1{}^2}$
- : If $S_1{}^2 > S_2{}^2$ then $F = \frac{S_2{}^2}{S_1{}^2}$ (Always $F > 1$)
- To test whether two independent samples have been drawn from the same population, we have to
- Equality of population means using t – test or z – test, test according to sample size.
- Equality of population variances using F – test.
-

Test of significance for equality of population variances

- **Working Rule:**
- Set the null hypothesis $H_0: \sigma_1{}^2 = \sigma_2{}^2$
- Set the alternative hypothesis $H_1: \sigma_1{}^2 \neq \sigma_2{}^2$

- Level of significance α at 5% (or) 1%
- The test statistic is given by, $F = \frac{S_1{}^2}{S_2{}^2}$

Where, $S_1{}^2 = \frac{1}{n_1 - 1}\sum(x_i - \bar{x})^2$, $S_2{}^2 = \frac{1}{n_2 - 1}\sum(y_i - \bar{y})^2$

Where $S_1{}^2 = \frac{n_1 s_1{}^2}{n_1 - 1}$

$$S_2{}^2 = \frac{n_2 s_2{}^2}{n_2 - 1}$$

If $|F| < F_{0.05}$, H_0 is accepted at 5% level of significance.

If $|F| > F_{0.05}$, H_0 is rejected at 5% level of significance.

1. **A sample of size 13 gave an estimated population variance of 3.0, while another sample of size 15 gave an estimate of 2.5. Could both sample be from populations with the same variance?**
   Solution:

   Given $n_1 = 8, n_2 = 7, S_1{}^2 = 3.0, S_2{}^2 = 2.5$

   Set the null hypothesis $H_0: \sigma_1{}^2 = \sigma_2{}^2$

   Set the alternative hypothesis $H_1: \sigma_1{}^2 \neq \sigma_2{}^2$

   Level of significance α at 5%

   The test statistic is given by, $F = \frac{S_1{}^2}{S_2{}^2}$

$$\Rightarrow F = \frac{3.0}{2.5} = 1.2$$

   Critical value: At 5% level, the tabulated value of F_α is 2.53 for the

   d.f= $(n_1 - 1, n_2 - 1) = (12,\ 14)$

   Conclusion: Since |F| = 1.2 < 2.53

   Hence Null Hypothesis H_0 is accepted at 5% level of significance.

   i.e., The two samples could have come from two normal populations with the same variance.

2. **Two random samples of sizes 8 and 11, drawn from two normal populations are characterized as follows**

	Size	Sum of observations	Sum of square of observations
Sample I	**8**	**9.6**	**61.52**
Sample II	**11**	**16.5**	**73.26**

   You are to decide if the two populations can be taken to have the same variance.

   Solution:

   Given $n_1 = 8, n_2 = 11, \sum(x_i - \bar{x})^2 = 61.52,\ \sum(y_i - \bar{y})^2 = 73.26$

   Set the null hypothesis $H_0: \sigma_1{}^2 = \sigma_2{}^2$

Set the alternative hypothesis $H_1: {\sigma_1}^2 \neq {\sigma_2}^2$

Level of significance α at 5%

The test statistic is given by, $F = \frac{{S_1}^2}{{S_2}^2}$

${S_1}^2 = \frac{1}{n_1-1}\sum(x_i - \bar{x})^2 = \frac{1}{8-1}(61.52) = 8.78,$

$${S_2}^2 = \frac{1}{n_2 - 1}\sum(y_i - \overline{y})^2 = \frac{1}{11-1}(73.26) = 7.326$$

$$F = \frac{{S_1}^2}{{S_2}^2} = \frac{8.78}{7.326} = 1.1984$$

Critical value: At 5% level, the tabulated value of F_α is 3.14 for the

d.f= $(n_1 - 1, n_2 - 1) = (7,\ 10)$

Conclusion: Since |F| = 1.1984 < 3.14

Hence Null Hypothesis H_0 is accepted at 5% level of significance.

i.e., The two samples could have come from two normal populations with the same variance.

3. **Two random samples gave the following data**

	Size	Mean	variance
Sample I	**8**	**9.6**	**1.2**
Sample II	**11**	**16.5**	**2.5**

Can we conclude that the two samples have been drawn from the same normal population?

Solution:

The two samples have been drawn from the same population means we have to check

The variance of the population do not differ significantly by F – test.

The sample means (and hence the population means) do not differ significantly by t – test

F – test:

Given $n_1 = 8, n_2 = 11, \overline{x_1} = 9.6\ , \overline{x_2} = 16.5\ , {s_1}^2 = 1.2, {s_2}^2 = 2.5$

Set the null hypothesis $H_0: {\sigma_1}^2 = {\sigma_2}^2$

Set the alternative hypothesis $H_1: {\sigma_1}^2 \neq {\sigma_2}^2$

Level of significance α at 5%

$$S_1{}^2 = \frac{n_1 s_1{}^2}{n_1 - 1} = \frac{8 * 9.6}{8 - 1} = 1.37$$

$$S_2{}^2 = \frac{n_2 s_2{}^2}{n_2 - 1} = \frac{11 * 2.5}{11 - 1} = 2.75$$

The test statistic is given by, $F = \frac{S_1{}^2}{S_2{}^2}$

$$\Rightarrow F = \frac{2.75}{1.37} = 2.007$$

Critical value: At 5% level, the tabulated value of F_α is 3.63 for the

d.f= $(n_1 - 1, n_2 - 1) = (10,\ 7)$

Conclusion: Since |F| = 2.007 < 3.63

Hence Null Hypothesis H_0 is accepted at 5% level of significance.

i.e., The two samples could have come from two normal populations with the same variance.

T – test: (Equality of means)

Given $n_1 = 8, n_2 = 11, \overline{x_1} = 9.6\ , \overline{x_2} = 16.5\ , s_1{}^2 = 1.2, s_2{}^2 = 2.5$
Set the null hypothesis H_0: $\mu_1 = \mu_2$

Set the alternative hypothesis H_1: $\mu_1 \neq \mu_2$

Level of significance α= 0.05(5%)

The test Statistic is $t = \frac{\overline{x_1} - \overline{x_2}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

$$S = \sqrt{\frac{n_1 s_1{}^2 + n_2 s_2{}^2}{n_1 + n_2 - 2}} = \sqrt{\frac{8 * 1.2 + 11 * 2.5}{8 + 11 - 2}} = 1.4772$$

$$t = \frac{9.6 - 16.5}{1.4772\sqrt{\frac{1}{8} + \frac{1}{11}}} = -10.0525$$

$$|t| = 10.0525$$

- **Critical value:** At 5% level, the tabulated value of t_α is 2.110 for

$$v = n_1 + n_2 - 2 = 8 + 11 - 2 = 17$$

- **Conclusion:** Since $|t|$= 10.0525 $>$ 2.110
- Hence Null Hypothesis H_0 is rejected at 5% level of significance.

i.e., the two samples could not have been drawn from the same normal population