Sri SAI RAM ENGINEERING COLLEGE

INSTITUTE OF TECHNOLOGY

West Tambaram, Chennai - 44

Sairam INSTITUTIONS

SAIRAM DIGITAL RESOURCES

YEAR II

SEM IV

MA8391

PROBABILITY AND STATISTICS (IT)

UNIT III

TESTING OF HYPOTHESIS

3.2 TESTING HYPOTHESIS:LARGE SAMPLE TESTS FOR PROPORTION AND DIFFERENCE OF PROPORTIONS

SCIENCE & HUMANITIES

4 QUALITY EDUCATION

NBA NATIONAL BOARD OF ACCREDITATION

NAAC NATIONAL ASSESSMENT AND ACCREDITATION COUNCIL A+

nirf NATIONAL INSTITUTIONAL RANKING FRAMEWORK

ISO 9001 : 2015 BUREAU VERITAS Certification

LARGE SAMPLE TESTS FOR PROPORTION AND DIFFERENCE OF PROPORTIONS

TEST FOR THE SIGNIFICANT DIFFERENCE BETWEEN SAMPLE PROPORTION AND POPULATION PROPORTION

PROCEDURE

A large sample of size n be taken from a population. Let P be the proportion of some attribute in the population. Let p be the proportion of the same attribute in the sample. We have to test whether p and P differ significantly or not. Let P_0 be the hypothetical value of the proportion in the population.

Null hypothesis $H_0 : P = P_0$

Alternative hypothesis $H_1 : P \neq P_0$ (two – tailed)

$$Q = 1 - P$$

Standard error of $p = \sqrt{\frac{PQ}{n}}$

Under H_0 , for large n , the sampling distribution of p is approximately normal with mean P and variance $\sqrt{\frac{PQ}{n}}$

$\therefore$ the test statistic is $Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$

Inference : If $|Z| < 1.96$, H_0 is accepted at 5 % level of significance.

$i.e,$ difference is not significant.

$|Z| > 1.96$, H_0 is rejected at 5 % level of significance.

$i.e,$ difference is significant.

$|Z| < 2.58$, H_0 is accepted at 1 % level of significance.

$i.e,$ difference is not significant.

$|Z| > 2.58$, H_0 is rejected at 1 % level of significance.

$i.e,$ difference is significant.

1. A die is thrown 400 times and a throw of 3 or 4 is observed 150 times. Test the hypothesis that the die is fair.

Solution : Given the die is thrown 400 times. $\therefore \quad n = 400$

Let p be the proportion of getting 3 or 4 in the sample.

Let P be the proportion of getting 3 or 4 in the population.

When the die is thrown 400 times 3 or 4 resulted in 150 times.

$$\therefore p = \frac{150}{400} = \frac{3}{8}$$

We want to test the hypothesis die is fair or not.

Null hypothesis H_0 : the die is fair (unbiased)

Alternative hypothesis H_1 : the die is not fair (biased)

$\therefore$ two tailed test.

Since die is unbiased $P = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$

$$\therefore \; Q = 1 - P = 1 - \frac{1}{3} = \frac{2}{3}$$

Under H_0 , the test statistic is

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{\frac{3}{8} - \frac{1}{3}}{\sqrt{\frac{1}{400} \cdot \frac{1}{3} \cdot \frac{2}{3}}} = \frac{(9-8)\,(3)\,(\sqrt{200})}{24} = \frac{1}{8}(14.14) = 1.767 = 1.77$$

The table value of Z at 5 % level is 1.96

$$\therefore |\, Z \,| < 1.96$$

Inference : Since $|Z| < 1.96$, H_0 is accepted at 5 % level of significance.

$i.e$, the data justifies the hypothesis that the die is unbiased.

2. In a sample of 400 parts produced by a factory, the number of defective parts was found to be 30. The company however claims that only 5 % of their products is defective. Is the claim tenable?

Solution : Given $n = 400$

No. of defectives in the sample $= 30$

$p =$ proportion of defective in the sample

$$p = \frac{30}{400} = 0.075$$

P = proportion of defective parts of the population

$$= 5\ \% = \frac{5}{100} = 0.05$$

$$Q = 1 - P = 1 - 0.05 = 0.95$$

We have to test whether the proportion in the population is $P = 5\ \%$ or not.

Null hypothesis $H_0 \ : \ P = 5\ \%$

Alternative hypothesis $H_1 \ : \ P \neq 5\ \%$ Two tailed test

Under H_0, the test statistic is

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.075 - 0.05}{\sqrt{\frac{(0.05)(0.95)}{400}}} = \frac{(-0.025)\,(20)}{\sqrt{0.0475}} = \frac{-0.5}{0.2179} = -2.29$$

$$|Z\,| = 2.29$$

The table value of $|Z\,|$ at 5 % level is 1.96

Inference : Since $|Z\,| > 1.96$, H_0 is rejected at 5 % level of significance.

$\therefore$ The claim is not tenable.

3. A quality – control engineer suspects that the proportion of defective units among certain manufactured items has increased from the set limit of 0.01. To test the claim, he randomly selected 100 of these items and found that the proportion of defective units in the sample was 0.02. Test the engineer's hypothesis at the 0.05 level of significance.

Solution : Given Sample size $n = 100$

Let $p =$ Proportion of defective units in the sample

$= 0.02$

$P =$ Proportion of defective units in the population

$= 0.01$

$Q = 1 - P = 1 - 0.01 = 0.99$

Null hypothesis $H_0 \ : \ P = 0.01$

Alternative hypothesis $H_1 \ : \ P > 0.01$ (Right tailed test)

Under H_0, the test statistic is

$$Z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{(0.02)-(0.01)}{\sqrt{\frac{(0.01)(0.99)}{100}}} = \frac{(0.01)\,(10)}{\sqrt{0.0099}} = \frac{0.1}{0.0994} = 1.006$$

For one-tailed test, the value of Z at 5 % level is 1.645

$\therefore \quad Z < 1.645$

Inference : The calculated value of $Z < 1.645$

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

The claim $P = 0.01$ is true.

TEST FOR THE SIGNIFICANT DIFFERENCE BETWEEN TWO PROPORTIONS IN TWO SAMPLES OR TEST FOR THE EQUALITY OF TWO PROPORTIONS

PROCEDURE

Let two large samples of sizes n_1, n_2 be drawn from a population with proportion P for some attribute A. Let p_1, p_2 be the proportions of the same attribute A in the samples respectively. We want to test the significant difference between p_1 and p_2 , $Q = 1 - P$.

Null hypothesis $H_0 : P_1 = P_2$

$i.e$, there is no significant difference between sample proportions.

Alternative hypothesis $H_1 : P_1 \neq P_2$

Two tailed test

Under H_0, the sampling distribution of $p_1 - p_2$ is approximately normal with mean P variance $PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

Inference : If $|Z| < 1.96$, H_0 is accepted at 5 % level of significance.

If $|Z| > 1.96$, H_0 is rejected at 5 % level of significance.

If $|Z| < 2.58$, H_0 is accepted at 1 % level of significance.

If $|Z| > 2.58$, H_0 is rejected at 1 % level of significance.

Case 2 : P is not known, $i.e.$ population proportion is not known.

If P is not known, then the unbiased estimate of P using p_1 and p_2 is

$$P = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$$

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

Inference : If $|Z| < 1.96$, H_0 is accepted at 5 % level of significance.

If $|Z| > 1.96$, H_0 is rejected at 5 % level of significance.

If $|Z| < 2.58$, H_0 is accepted at 1 % level of significance.

If $|Z| > 2.58$, H_0 is rejected at 1 % level of significance.

1. Random sample of 400 men and 600 women were asked whether they would like to have a fly-over near their residence 200 men and 325 women were in favour of it. Test the equality of proportion of men and women in the proposal.

Solution : Let p_1, p_2 be the proportions favourable to have a fly over among men and women respectively.

$$\therefore \quad p_1 = \frac{200}{400} \quad ; \quad p_2 = \frac{325}{600}$$

P is not known. $\therefore$ The best estimate of

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(400)\left(\frac{200}{400}\right) + (600)\left(\frac{325}{600}\right)}{400 + 600} = \frac{525}{100} = 0.525$$

$$Q = 1 - P = 1 - 0.525 = 0.475$$

Null hypothesis $H_0 : p_1 = p_2$

Alternative hypothesis $H_1 : p_1 \neq p_2$ (two tailed test)

Under H_0 , the test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.5-0.542}{\sqrt{(0.525)(0.475)\left(\frac{1}{400}+\frac{1}{600}\right)}} = \frac{-0.042}{\sqrt{(0.2494)\,(0.0025+0.00167)}}$$

$$= \frac{-0.042}{0.0324} = -1.30$$

$$|Z| = -1.30$$

The table value of Z at 5% level is 1.96.

Inference : Since $Z < 1.96$, H_0 is accepted at 5% level of significance.

The difference is not significant $i.e.$ men and women are equally favourable for fly near their residence.

2. Before increase in excise duty on tea, 400 people out of a sample of 500 persons were found to be tea drinkers. After an increase in duty, 400 people were tea drinkers out of sample of 600 people. Using the standard error of proportion state whether there is a significant decrease in the consumption of tea?

Solution : Let p_1, p_2 be the proportion of tea drinkers before and after the increase in excise duty.

Given $n_1 = 500, \qquad n_2 = 600$

$$p_1 = \frac{400}{500} = 0.8 \; ; \; p_2 = \frac{400}{600} = 0.67$$

$\therefore$ P is not given.

The best estimate of P is given by

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{(500)\left(\frac{400}{500}\right) + (600)\left(\frac{400}{600}\right)}{500 + 600} = \frac{800}{1100} = \frac{8}{11}$$

$$Q = 1 - P = 1 - \frac{8}{11} = \frac{3}{11}$$

Null hypothesis H_0 : $p_1 = p_2$

Alternative hypothesis H_1 : $p_1 > p_2$ $\quad\therefore$ one tailed test

Under H_0 , the test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.8-0.67}{\sqrt{\left(\frac{8}{11}\right)\left(\frac{3}{11}\right)\left(\frac{1}{500}+\frac{1}{600}\right)}} = \frac{(0.13)\ (11)}{\sqrt{(24)\left(\frac{1}{5}+\frac{1}{6}\right)\left(\frac{1}{100}\right)}}$$

$$= \frac{1.43}{(24)\ (0.0037)} = 4.822$$

$$Z = 4.82$$

The table value of Z at 5% level for one tailed test is 1.645 and for 1% level is 2.33.

$\therefore\ Z > 1.645$ and $Z > 2.33$

Inference: Since $Z > 1.645$ and $Z > 2.33$, H_0 is rejected at 5% as well as 1% level of significance.

$\therefore$ the difference is highly significant. $\quad\therefore\ p_1 > p_2$ is true

$\therefore$ there is a significant decrease in the consumption of tea drinkers after the increase in excise duty.

3. In a random sample of 1000 people from city A, 400 are found to be consumers of wheat. In sample of 800 from city B, 400 are found to be consumers of wheat. Does this data give a significant difference between the two cities as far as the proportion of wheat consumers is concerned?

Solution : Let p_1 be the proportion of consumers of wheat in city A.

Let p_2 be the proportion of consumers of wheat in city B.

Given $n_1 = 1000,\qquad n_2 = 800$

$$p_1 = \frac{400}{1000} = 0.4 \quad ; \quad p_2 = \frac{400}{800} = 0.5$$

P is not known.

The best estimate of P is given by

$$P = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{(1000)\left(\frac{400}{1000}\right) + (800)\left(\frac{400}{800}\right)}{1000 + 800} = \frac{800}{1800} = \frac{4}{9}$$

$$Q = 1 - P = 1 - \frac{4}{9} = \frac{5}{9}$$

Null hypothesis $H_0 \ : \ p_1 = p_2$

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

Under H_0, the test statistic is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.4-0.5}{\sqrt{\left(\frac{4}{9}\right)\left(\frac{5}{9}\right)\left(\frac{1}{1000}+\frac{1}{800}\right)}} = \frac{(-0.13)\ (10)\ (9)}{\sqrt{(20)(0.225)}}$$

$$= \frac{-9}{2.1213} = 4.24$$

$$|Z| = 4.24$$

The table value of Z at 5% level is 1.96. $\therefore \ |Z| > 1.96$

Inference : Since $|Z| > 1.96$ and $|Z| > 2.58$, H_0 is rejected at 5% level of significant.

$\therefore$ H_0 is rejected even at 1% level of significance.

$\therefore$ there is significant difference between the consumers of wheat in the two cities A and B.

YOUTUBE LINK

https://www.youtube.com/watch?v=2WZVzjbmmLQ

https://www.youtube.com/watch?v=QonO9IGgdME