Testing of Hypothesis: Large Sample Tests

Lesson 4: Test of Hypothesis for Difference of Two Population Proportions

If the test statistics are proportions R_1 and R_2 of samples of sizes n_1 and n_2, then the mean and the standard deviation of the sampling distribution of the differences of the two population proportions will be given by

$$\mu_{R_1-R_2} = \mu_{R_1} - \mu_{R_2} = p_1 - p_2$$

and

$$\sigma_{R_1-R_2} = \sqrt{{\sigma_{R_1}}^2 + {\sigma_{R_2}}^2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Here p_1 and p_2 are the proportions of the two populations from which the samples are drawn.

The test statistic will then be given by

$$z = \frac{(R_1 - R_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} - - - - (1)$$

The steps for testing are:

(i) Formulate null hypothesis, $H_0 : p_1 = p_2$: There is no significant difference between two population proportions

(ii) Formulate alternative hypothesis, $H_1 : p_1 \neq p_2$ (or otherwise) : There is a significant difference between two population proportions

(iii) Level of significance is : α

(iv) From the two tailed test, the critical region is : $|z| \geq z_{\alpha/2}$

(v) Compute the test statistic z from equation (1)

(vi) Reject H_0 if the computed value of z falls in the critical region, otherwise accept H_0

Note:

(i) If population proportions p_1 and p_2 are equal and known (= p), then equation (1) will be modified as

$$z = \frac{\mathrm{R}_1 - \mathrm{R}_2}{\sqrt{\mathrm{p}(1-\mathrm{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

(ii) If population proportions p_1 and p_2 are equal but unknown, then equation (1) will be changed to the following form

$$z = \frac{\mathrm{R}_1 - \mathrm{R}_2}{\sqrt{\mathrm{M}(1-\mathrm{M})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

where

$$\mathrm{M} = \frac{n_1\mathrm{R}_1 + n_2\mathrm{R}_2}{n_1 + n_2}$$

is called the unbiased estimator of the unknown population proportion.

Problems:

Ex.1. A machine produced 20 defective articles in a batch of 400. After overhauling, it produced 10 defectives in a batch of 300. Has the machine improved?

Solution:

Here in first batch of sample, $n_1 = 400$ and $R_1 = \frac{20}{400}$. After overhauling, in a second batch of sample, $n_2 = 300$ and $R_2 = \frac{10}{300}$.

Here the two population proportions are unknown, therefore we shall use

$M = \frac{n_1 R_1 + n_2 R_2}{n_1 + n_2} = \frac{3}{70}$ as an unbiased estimator of them.

(i) Null hypothesis, $H_0 : p_1 = p_2$: There is no significant difference between two population proportions, i.e, the machine has not improved

(ii) Alternative hypothesis, $H_1 : p_1 > p_2$, which means that the proportions of defective articles is reduced after overhauling, i.e, the machine has improved

(iii) $\alpha = 0.05$ (5% level)

(iv) The critical region at 5% level of significance is $z \geq 1.645$

(v) $$z = \frac{R_1 - R_2}{\sqrt{M(1-M)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\frac{1}{20}-\frac{1}{30}}{\sqrt{\frac{3}{70}\times\frac{67}{70}\left(\frac{1}{400}+\frac{1}{300}\right)}} = 1.08$$

(vi) As computed z < 1.645, i.e, the test statistic does not lie in the critical region, so H_0 is accepted

Hence, we conclude that the machine has not improved after overhauling.

Ans.

Ex.2. Suppose that a method A results in 20 unacceptable transistors out of 100 produced, whereas another method B results in 12 unacceptable transistors out of 100 produced. Can we conclude at 5% level of significance that the two methods are equivalent?

Solution:

Here in method A, $n_1 = 100$ and $\mathrm{R}_1 = \frac{20}{100}$. In method B, $n_2 = 100$ and $\mathrm{R}_2 = \frac{12}{100}$. Since the two population proportions are unknown, therefore we shall use $\mathrm{M} = \frac{n_1\mathrm{R}_1 + n_2\mathrm{R}_2}{n_1 + n_2} = \frac{4}{25}$ as an unbiased estimator of them.

(i) Null hypothesis, H_0 : $\mathrm{p}_1 = \mathrm{p}_2$, which means that the two methods are equivalent

(ii) Alternative hypothesis, H_1 : $\mathrm{p}_1 \neq \mathrm{p}_2$, i.e, the two methods are not equivalent

(iii) $\alpha = 0.05$

(iv) Critical region at 5% level of significance is $|z| \geq 1.96$

(v) $z = \frac{\mathrm{R}_1 - \mathrm{R}_2}{\sqrt{\mathrm{M}(1-\mathrm{M})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\frac{20}{100}-\frac{12}{100}}{\sqrt{\frac{4}{25}\times\frac{26}{25}\left(\frac{1}{100}+\frac{1}{100}\right)}} = 1.387$

(vi) Since computed $|z| < 1.96$, i.e, it does not lie in the critical region, so null hypothesis H_0 is accepted

Hence we conclude that the two methods are equivalent.

Ans.