Test of significance of the difference between sample proportion and population proportion (single proportion)

Let X be the number of successes in n independent Bernoulli trials in which the probability of success for each trial is a constant $= P$ (say). Then it is known that X follows a binomial distribution with mean $E(X) = n\,P$ and variance $V(X) = n\,P\,Q$

When n is large, X follows $N\left(nP, \sqrt{nPQ}\right)$, i.e. a normal distribution with mean $n\,P$ and S.D. $\sqrt{nPQ}$, where $Q = 1 - P$. $\frac{X}{n}$ follows $N\left(\frac{Pn}{n}, \sqrt{\frac{nPQ}{n^2}}\right)$

Now $\frac{X}{n}$ is the proportion of successes in the sample consisting of n trials, that is denoted by p. Thus the sample proportion p follows $N\left(P, \sqrt{\frac{PQ}{n}}\right)$. Therefore test statistic

$$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1).$$

If $| z | \leq z_{\alpha}$, the difference between the sample proportion p and the population proportion P is not significant at α% LOS.

Note: When P is not known, the 95 percent confidents limits for P are given by

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

Problem 1: If 20 people were attacked by a disease and only 18 survived, will you reject the hypothesis that the survival rate if attacked by this disease is 85% in favor of the hypothesis that is more at 5% level.

Solution: Number of people survived= $x = 18$.

Size of the sample $= n = 20$.

$p =$ Proportion of the people survived$= \frac{x}{n} = \frac{18}{20} = 0.9$

It is given that= $P = 85\% = 0.85$. $Q = 1 - P = 1 - 0.85 = 0.15$

Null hypothesis: $H_0: P = 0.85$

Alternative hypothesis: $H_1: P > 0.85$

Level of significance $= \alpha = 0.05$

Test statistic: $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = 0.6265$.

Table value of $z = 1.645$.

Calculated value of z is less than the table value of z at 5% level of significance. Null hypothesis is accepted.

Problem 2: In a city, a sample of 1000 people were taken and out of them 540 are vegetarians and the rest are non-vegetarians. Can we say that both habits of eating (vegetarian or non-vegetarian) are equally popular in the city at 1% level of significance.

Solution: Number of people survived$= x = 540$.

Size of the sample $= n = 1000$.

$p =$ Proportion of the people survived$= \frac{540}{1000} = 0.54$

Population proportion(P) $= 0.5$. $Q = 1 - P = 1 - 0.5 = 0.5$

Null hypothesis: $H_0 : P = 0.5$

Alternative hypothesis: $H_1 : P \neq 0.5$

Level of significance $= \alpha = 0.01$

Test statistic: $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = 2.532$.

Table value of $z = 2.58$.

Calculated value of z is less than the table value of z at 5% level of significance. Null hypothesis is accepted.

Test of significance for difference of proportions

Suppose two samples of sizes n_1 and n_2 are drawn from two different populations. To test the significance of difference between the two proportions, we consider the following cases.

Case-I When the population proportions P_1 and P_2 are known:
In this case $Q_1 = 1 - P_1$ and $Q_2 = 1 - P_2$. The test statistic is

$$Z = \frac{P_1 - P_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Case-II When the population proportions P_1 and P_2 are not known but sample proportions p_1 and p_2 are known:

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

Problem 3: A machine puts out 16 imperfect articles in sample of 500. After the machine is overhauled, it puts out 3 imperfect articles in a batch of 100. Has the machine improved?

Solution : We are given n_1=500and n_2=100

p_1= Proportions of defective items before overhauling of machine =16/500=0.032

p_2= Proportions of defective items after overhauling of machine =3/100=0.03

H_0: $P_1=P_2$ i.e. the machine has not improved after overhauling.

H_1: $P_1>P_2$

$$Z = \frac{(p_1 - p_2)}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} = \frac{0.032 - 0.03}{\sqrt{\frac{(0.032)(0.968)}{500} + \frac{(0.3)(0.7)}{100}}} = \frac{0.002}{0.01878} = 0.106$$

Since Z<1.645 (Right tailed test), it is not significant at 5% level of significance. Hence we may accept the null hypothesis and conclude that the machine has not improved after overhauling.

Case-III Method of pooling:

In this method, the sample proportions p_1 and p_2 are pooled into a single proportion P, by using the formula:

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

The test statistic in this case is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ Q=1-P}$$

Case-IV Test the significance difference between p_1 and *P*:

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

The test statistic in this case is

$$Z = \frac{p_1 - P}{\sqrt{\frac{n_2 PQ}{n_1 + n_2}}}$$

Problem 4: In a random sample of 400 student of the university teaching department, it was found that 300 students failed in the examination. In another sample of 500 students of the affiliated colleges the number of failures in the same examination was found to be 300. Find out whether the proportion of failures in the university teaching departments significantly greater than the proportion of failures in the university teaching departments and affiliated colleges taken together.

Solution:

$$\text{Given: } n_1 = 400, n_2 = 500, p_1 = \frac{300}{400} = 0.75, p_2 = \frac{300}{500} = 0.6$$

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400\left(\frac{300}{400}\right) + 500\left(\frac{300}{500}\right)}{400 + 500} = 0.667, q = 0.333$$

Null hypothesis H_o : Assume that there is no significant difference between p_1 & p

Test Statistic : $Z = \frac{p_1 - p}{\sqrt{\frac{n_2 pq}{n_1(n_1+n_2)}}} = \frac{0.75 - .667}{\sqrt{\frac{500\times0.667\times0.333}{400(400+500)}}} = 4.74$

Tabulated value : z at 5% level of significance is 1.96

Conclusion : C.V > T.V

We reject the null hypothesis H_o.

Therefore, the proportion of failures in the affiliated colleges is greater than the proportion of failures in university departments and affiliated colleges taken together.