Problems under Proportion(s)

Type	Formula
Single Proportion	$$\boldsymbol{Z = \frac{p' - p}{\sqrt{pq/n}}}$$
Difference of two Proportions	$$\boldsymbol{Z = \frac{p_1' - p_2'}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$ Where, $p = \frac{n_1 p_1' + n_2 p_2'}{n_1 + n_2}$ and $q = 1 - p$

Notations:

- p stands for population proportion and $q = 1 - p$.
- p' stands for sample proportion.
- p'_1 and p'_2 stands for proportions of sample 1 and sample 2, respectively.
- p_1 and p_2 stands for proportions of population 1 and population 2, respectively.

Problems
Sample Proportion (Single and Difference)

1. Before an increase in excise duty on tea, 800 people out of a 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 consumption of tea after the increase in duty. Also find the 95% confidence limits for difference of two proportions.

Solution:

Sample 1 (Before an increase in excise duty): $n_1 = 1000 \; and \; \boldsymbol{p_1'} = \frac{800}{1000} = 0.8$

Sample 2 (After an increase in excise duty): $n_2 = 1200 \; and \; \boldsymbol{p_2'} = \frac{800}{1200} = 0.67$

Population: $p = \frac{n_1 p_1' + n_2 p_2'}{n_1 + n_2} = \frac{800+800}{1000+1200} = 0.73 \; and \; q = 1 - p = 0.27$

Step 1:

H_0: There is no significant decrease in the consumption of tea after the increase in duty

(i.e.) $p_1' \leq p_2'$

H_1: There is a significant decrease in the consumption of tea after the increase in duty

(i.e.) $p_1' > p_2'$ [One tailed test]

Step 2: Choosing α the level of significance (LOS)

$$\alpha = 5\%$$

Step 3: Computing the test statistic Z using a formula

$$Z = \frac{p_1' - p_2'}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.8 - 0.67}{\sqrt{(0.73)(0.27)\left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 6.82$$

Step 4: Critical value Z_α

$$Z_\alpha = 1.645 \text{ (One tailed test)}$$

Step 5: Drawing the Conclusion

Here $|Z| = 6.82$ and $Z_\alpha = 1.645$.

Since $|Z| > Z_\alpha$, we reject H_0 at 5% level of significance.

Now to find the 95% confidence limits for difference of two population proportions.

$$p_1 - p_2 \in \left((p_1' - p_2') - Z_\alpha \left(\sqrt{\frac{p_1' q_1'}{n_1} + \frac{p_2' q_2'}{n_2}} \right), (p_1' - p_2') + Z_\alpha \left(\sqrt{\frac{p_1' q_1'}{n_1} + \frac{p_2' q_2'}{n_2}} \right) \right)$$

$$p_1 - p_2 \in \left((0.8 - 0.67) - 1.96 \left(\sqrt{\frac{(0.8)(0.2)}{1000} + \frac{(0.67)(0.33)}{1200}} \right), (0.8 - 0.67) + 1.96 \left(\sqrt{\frac{(0.8)(0.2)}{1000} + \frac{(0.67)(0.33)}{1200}} \right) \right)$$

$$p_1 - p_2 \in ((0.13) - 1.96(0.019), (0.13) + 1.96(0.019))$$

$$p_1 - p_2 \in (0.09276, 0.16724)$$

Note: This implies that we are 95% confidence that this confidence interval will contain the true unknown value of the difference of population proportions.

2. A manufacturing company claims that more than 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:

Sample: $n = 200 \; and \; p' = \frac{200-18}{200} = 0.91$

Population: $p = 0.95 \; and \; q = 1 - p = 0.05$

Step 1: Formulating the hypothesis H_0 & H_1 and decide whether it is a Two tailed test or a One tailed test with the help of H_1

H_0: At most 95% of the company products supplied confirm to the specifications

(i.e.) $p' \leq p$

H_1: More than 95% of the company products supplied confirm to the specifications

(i.e.) $p' > p$ [One tailed test]

Step 2: Choosing α the level of significance (LOS)

$\alpha = 5\%$ (given)

Step 3: Computing the test statistic Z using a formula

$$\boldsymbol{Z = \frac{p' - p}{\sqrt{pq/n}} = \frac{0.91 - 0.95}{\sqrt{(0.95)(0.05)/200}} = -2.56}$$

Step 5: Drawing the Conclusion

Here $|Z| = 2.56$ and $Z_\alpha = 1.645$.

Since $|Z| > Z_\alpha$, we reject H_0 at 5% level of significance.

3. The fatality rate of typhoid patients is believed to be less than 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?. Also find the 95% confidence limits for proportion.

Solution:

Sample: $n = 640 \; and \; p' = \frac{63}{640} = 0.098$

Population: $p = 0.1726 \; and \; q = 1 - p = 0.8274$

Step 1:

H_0: The hospital is inefficient.

(i.e.) $p' \geq p$

H_1: The hospital is efficient.

(i.e.) $p' < p$ [One tailed test]

Step 2: Choosing α the level of significance (LOS)

$$\alpha = 5\%$$

Step 3: Computing the test statistic Z using a formula

$$Z = \frac{p' - p}{\sqrt{pq/n}} = \frac{0.0984 - 0.1726}{\sqrt{(0..1726)(0.8274)/640}} = -4.96$$

Step 4: Critical value Z_α

$$Z_\alpha = 1.645 \text{ (One tailed test)}$$

Step 5: Drawing the Conclusion

Here $|Z| = 4.96$ and $Z_\alpha = 1.645$.

Since $|Z| > Z_\alpha$, we reject H_0 at 5% level of significance.

Now to find the 95% confidence limits for population proportion.

$$p \in \left(p' - Z_\alpha\left(\sqrt{p_1 q_1/n}\right), p' + Z_\alpha\left(\sqrt{p_1 q_1/n}\right)\right)$$

$$p \in \left(0.0984 - 1.96\left(\sqrt{\frac{(0.0984)(0.9016)}{640}}\right), 0.0984 + 1.96\left(\sqrt{\frac{(0.0984)(0.9016)}{640}}\right)\right)$$

$$p \in (0.0984 - 1.96(0.012), 0.0984 + 1.96(0.012))$$

$$p \in (0.07488, 0.12192)$$

Note: This implies that we are 95% confidence that this confidence interval will contain the true unknown value of the population proportions.

Here, $p = 0.1726 \notin (0.07488, 0.12192)$, thus we have H_0 is rejected.

Module 6: Hypothesis Testing II

Small sample tests- Student’s t-test, F-test- chi-square test- goodness of fit – independence of attributes- Design of Experiments - Analysis of variance – One way-Two way-Three way classifications - CRD-RBD-LSD.

Test of Hypothesis – Small Sample

Small Sample (i.e sample size is <= 30) test is done by using **F-test** and **t-test**.

F-Test

To test the significant difference btn population variances.

$$\boxed{F = \frac{\text{Greater variance}}{\text{Smaller variance}}}$$

Eg: $F = \dfrac{\sigma_1^2}{\sigma_2^2}$ (if $\sigma_1^2 > \sigma_2^2$)

or

$F = \dfrac{\sigma_2^2}{\sigma_1^2}$ (if $\sigma_2^2 > \sigma_1^2$).

If σ_1^2 and σ_2^2 are unknown

Sample variance

Thus $\sigma_1^2 = \dfrac{n_1 s_1^2}{n_1 - 1}$ and $\sigma_2^2 = \dfrac{n_2 s_2^2}{n_2 - 1}$

Population variance   Population variance.

<u>Degree of freedom</u>, $\nu = (n_1 - 1, n_2 - 1)$ if $\sigma_1^2 > \sigma_2^2$

and $\nu = (n_2 - 1, n_1 - 1)$ if $\sigma_2^2 > \sigma_1^2$.