Statistical inference about two population proportions

Session 23

08/05/2021

Statistical inference about two population proportions (P_1-P_2)

- Sometimes the researcher wishes to make inferences about the difference in two population proportions.
- This type of analysis has many applications in business, such as comparing the market share of a product for two different markets, studying the difference in the proportion of female customers in two different geographic regions, or comparing the proportion of defective products from one period to another.
- In making inferences about the difference in two population proportions, the statistic normally used is the difference in the sample proportions $(\widehat{p_1} - \widehat{p_2})$.

- The central limit theorem states that for large samples (each of $n_1\widehat{p_1}$, $n_1\widehat{q_1}$, $n_2\widehat{p_2}$, $n_2\widehat{q_2}$ >5 where $\hat{q} = 1 - \hat{p}$), the difference in sample proportions is normally distributed with a mean difference of

$$\mu_{\widehat{p_1}-\widehat{p_2}} = P_1 - P_2.$$

and a standard deviation of the difference of sample proportions

$$\sigma_{\widehat{p_1}-\widehat{p_2}} = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$

Z statistic

$$z = \frac{(\widehat{p_1} - \widehat{p_2}) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

Where $\widehat{p_1}$: proportion from sample 1

$\widehat{p_2}$:proportion from sample 2

n_1: size of sample 1

n_2: size of sample 2

P_1: proportion from population 1

P_2: Proportion from population 2

Q_1= 1-P_1

Q_2= 1-P_2

Confidence Interval to estimate $(P_1 - P_2)$

$$z = \frac{(\widehat{p_1} - \widehat{p_2}) - (P_1 - P_2)}{\sqrt{\frac{\widehat{p_1}\widehat{q_1}}{n_1} + \frac{\widehat{p_2}\widehat{q_2}}{n_2}}}$$

which on solving results

$$(\widehat{p_1} - \widehat{p_2}) - z_{\propto/2}\sqrt{\frac{\widehat{p_1}\widehat{q_1}}{n_1} + \frac{\widehat{p_2}\widehat{q_2}}{n_2}} \leq (P_1 - P_2) \leq$$

$$(\widehat{p_1} - \widehat{p_2}) + z_{\propto/2}\sqrt{\frac{\widehat{p_1}\widehat{q_1}}{n_1} + \frac{\widehat{p_2}\widehat{q_2}}{n_2}}$$

Example

- Managers of a supermarket chain want to determine the difference between the proportion of morning shoppers who are men and the proportion of after–5 P.M. shoppers who are men. Over a period of two weeks, the chain's researchers conduct a systematic random sample survey of 400 morning shoppers, which reveals that 352 are women and 48 are men. During this same period, a systematic random sample of 480 after–5 P.M. shoppers reveals that 293 are women and 187 are men. Construct a 98% confidence interval to estimate the difference in the population proportions of men.

Morning Shoppers	After–5 P.M. Shoppers
$n_1 = 400$	$n_2 = 480$
$x_1 = 48$ men	$x_2 = 187$ men
$\hat{p}_1 = .12$	$\hat{p}_2 = .39$
$\hat{q}_1 = .88$	$\hat{q}_2 = .61$

C.I Estimate

- The sample proportions are $\widehat{p_1} = \frac{48}{400} = 0.12$ and $\widehat{p_2} = \frac{187}{480} = 0.39$

- For α=0.02, $z_{\frac{\alpha}{2}}$=2.33
- Hence 98% C.I is

$$(.12 - .39) - 2.33\sqrt{\frac{(.12)(.88)}{400} + \frac{(.39)(.61)}{480}} \le p_1 - p_2$$

$$\le (.12 - .39) + 2.33\sqrt{\frac{(.12)(.88)}{400} + \frac{(.39)(.61)}{480}}$$

$$-.27 - .064 \ge p_1 - p_2 \ge -.27 + .064$$

$$-.334 \ge p_1 - p_2 \ge -.206$$

Output interpretation

```
TEST AND CI FOR TWO PROPORTIONS
Sample    X    N    Sample p
1        48   400   0.120000
2       187   480   0.389583
Difference = p(1) - p(2)
Estimate for difference: -0.269583
98% CI for difference: (-0.333692, -0.205474)
Test for difference = 0 (vs not = 0):
Z = -9.78  P-Value = 0.000
```

Modified formula

- The formula given includes the population proportions (P_1 and P_2) which usually unknown in real life situation. Hence a modified formula can be used as the test statistic for hypotheses about the difference in two population proportions.
- The denominator of formula is the standard deviation of the difference in two sample proportions and uses the population proportions in its calculations. However, the population proportions are unknown, so an estimate of the standard deviation of the difference in two sample proportions is made by using sample proportions as point estimates of the population proportions.

- The sample proportions are combined by using a weighted average to produce $\bar{p}$, which, in conjunction with $\bar{q}$ and the sample sizes, produces a point estimate of the standard deviation of the difference in sample proportions.
- We shall use this modified formula to test hypotheses about the difference in two population proportions.

$$z = \frac{(\widehat{p_1} - \widehat{p_2}) - (P_1 - P_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\widehat{p_1} + n_2\widehat{p_2}}{n_1 + n_2}$.

Test-8:Testing Hypotheses about two population Proportions

	Left tail test	Right tail Test	Two tail Test
Hypotheses	$H_0: P_1=P_2$ $H_1: P_1< P_2$	$H_0: P_1=P_2$ $H_1: P_1> P_2$	$H_0: P_1=P_2$ $H_1: P_1 \neq P_2$
Test Statistic	$z=\frac{(\widehat{p_1}-\widehat{p_2})-(P_1-P_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $\bar{p}=\frac{x_1+x_2}{n_1+n_2}=\frac{n_1\widehat{p_1}+n_2\widehat{p_2}}{n_1+n_2}$.		
Rejection Rule	Reject H_0 if $Z \leq -Z_\alpha$	Reject H_0 if $Z \geq Z_\alpha$	Reject H_0 if $\lvert Z\rvert \geq z_{\alpha/2}$

Example

- According to a study conducted for Gateway Computers, 59% of men and 70% of women say that weight is an extremely/very important factor in purchasing a laptop computer. Suppose this survey was conducted using 374 men and 481 women. Do these data show enough evidence to declare that a significantly higher proportion of women than men believe that weight is an extremely/very important factor in purchasing a laptop computer? Use a 5% level of significance.

Solution

- Population 1: Men & Population 2 : Women
- The sample proportions are $\widehat{p_1} = 0.59$ and $\widehat{p_2} = 0.70$
- n_1=374 and n_2= 481
- The null and alternative hypothesis are H_0: P_1-P_2=0

  H_1:P_1-P_2<0
- Given α=0.05
- The test statistic is

$$z = \frac{(\widehat{p_1} - \widehat{p_2}) - (P_1 - P_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\widehat{p_1} + n_2\widehat{p_2}}{n_1 + n_2}$.

- Substituting and simplifying we get, $\bar{p} = 0.65$ and $\bar{q} = 0.35$

- Calculated value of Z = $\frac{-0.11}{0.0328} = -3.35$

- For α=0.05, $Z_\alpha = -1.645$

- Rejection rule is reject H_0 if $Z \leq -Z_\alpha$

- Here $-3.35 < -1.645$ and hence we reject H_0 .

- Conclusion : There is enough evidence to conclude that a significantly higher proportion of women than men believe that weight is an extremely/very important factor in purchasing a laptop computer.