ECON 3818: Introduction to Statistics with Computer Applications

Kyle Butts

November 4, 2020

Chapter 23: Comparing Two Proportions

Notation

We will use notation similar to that used in our study of two-sample t-statistics.

Population	Pop. Proportion	Sample Size	Sample Proportion
1	p_1	n_1	$\hat{p}_1$
2	p_2	n_2	$\hat{p}_2$

Sampling Distribution of $\hat{p}$ Review

$X \sim B(1, p)$ is the underlying variable.

$$\hat{p} = \frac{\sum \# \text{ of successes}}{n}$$

The sample distribution of $\hat{p}$ with population proportion p_0:

$$\hat{p} \sim N(p_0, \frac{p_0(1-p_0)}{n})$$

Sampling Distribution of a Difference between Proportions

To use $\hat{p}_1 - \hat{p}_2$ for inference we use the following information:

- When the samples are large, the distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal
- The mean of the sampling distribution is: $p_1 - p_2$
- Assuming the two populations are independent, the standard deviation of the distribution is:

$$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Normal Distribution Review

If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are normally distributed and indepdent, then $X_1 - X_2$ is normally distributed,

$$E\{X_1 - X_2\} = \mu_1 - \mu_2,$$

$$Var(X_1 - X_2) = \sigma_1^2 + \sigma_2^2$$

Large-Sample Confidence Intervals for Comparing Proportions

Using the equation for standard error:

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

The confidence interval is constructed as:

$$\hat{p}_1 - \hat{p}_2 \pm Z^* SE$$

Example

Construct a 95% confidence interval for the following difference in proportions:

Population	**# Successes**	**Sample Size**	**Sample Proportion**
1	75	100	$\hat{p}_1 = 0.75$
2	56	100	$\hat{p}_2 = 0.56$

Construct a 95% confidence interval for the following difference in proportions:

Population	# Successes	Sample Size	Sample Proportion
1	75	100	$\hat{p}_1 = 0.75$
2	56	100	$\hat{p}_2 = 0.56$

$$SE = \sqrt{\frac{(0.75)(0.25)}{100} + \frac{(0.56)(0.44)}{100}} = 0.0659$$

Confidence interval $=$

$(0.75 - 0.56) \pm (1.96)(0.0659) \implies 0.06$ to 0.32

Significance Tests for Comparing Proportions

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 \neq 0$$

In order to test the hypothesis, we must first calculated the pooled sample proportion

$$\hat{p} = \frac{\text{number of successes in } \textit{both samples combined}}{\text{number of individuals in } \textit{both samples combined}}$$

Then we use the following z-statistic:

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Example

Population	# Successes	Sample Size	Sample Proportion
1	212	616	$\hat{p}_1 = 0.344$
2	7	49	$\hat{p}_2 = 0.143$

Example

Population	# Successes	Sample Size	Sample Proportion
1	212	616	$\hat{p}_1 = 0.344$
2	7	49	$\hat{p}_2 = 0.143$

- Calculate

$$\hat{p} = \frac{212 + 7}{616 + 49} = 0.329$$

- Calculate Z-statistic

$$Z = \frac{0.344 - 0.143}{\sqrt{(0.329)(0.671)\left(\frac{1}{616} + \frac{1}{49}\right)}} = 2.88$$

Example – continued

The z-statistic was 2.88, and we have a two-tailed alternative hypothesis. Therefore:

$$\text{p-value } = 2 \cdot P(Z > 2.88) = 2 \cdot 0.002 = 0.004$$

Therefore we reject null at $\alpha = 0.05$