Introduction to Statistical Methods in Political Science

Lesson Week 5: Sampling Distribution for the Difference of Two Proportions

Ignacio Urbina

Example: Public Opinion

One common application of comparing two sample proportions is in analyzing public opinion across different demographic groups. For instance:

- Consider two groups within a population: those under 30 and those over 60.
- We survey both groups to assess their support for the current president.
- Let $\hat{p}_1$ be the proportion of support among those under 30 and $\hat{p}_2$ among those over 60.

This analysis helps us understand how support for the president varies with age, which can be crucial for policy making and election strategies.

Example: Psychological Experiment

In psychological research, comparing sample proportions is key to evaluating the effect of different treatment conditions:

- Imagine a study testing the effect of two therapeutic interventions on patient anxiety reduction.
- One group receives cognitive behavioral therapy (CBT) while another receives mindfulness-based stress reduction (MBSR).
- Let $\hat{p}_1$ be the proportion of patients who report a significant reduction in anxiety with CBT, and $\hat{p}_2$ with MBSR.

This comparison can provide insights into which intervention is more effective, guiding treatment decisions and future research.

Introduction to Sample Proportions

Consider two independent samples where:

- Sample 1: n_1 observations with proportion $\hat{p}_1$
- Sample 2: n_2 observations with proportion $\hat{p}_2$

We are interested in the statistic $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions.

Review of Expectations and Variances

Expectations:

- For any random variables X and Y, and constants a, b:

$$E(aX + bY) = aE(X) + bE(Y)$$

Variances:

- For independent random variables X and Y, and constants a, b:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

- If X and Y are not independent, the covariance must be considered:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

Expectation of the Statistic

Using the rules for expectations of linear combinations of random variables:

$$E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = p_1 - p_2$$

Where p_1 and p_2 are the true population proportions.

Variance of the Statistic

Using the rules for variances of linear combinations of *indepedent* random variables:

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2)$$

Given $\hat{p}_1$ and $\hat{p}_2$ are independent,

$$\text{Var}(\hat{p}_1) = \frac{p_1(1-p_1)}{n_1}, \quad \text{Var}(\hat{p}_2) = \frac{p_2(1-p_2)}{n_2}$$

Thus,

$$\text{Var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Standard Error of the Statistic

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\text{Var}(\hat{p}_1 - \hat{p}_2)}$$

Substituting the variances,

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Normal Approximation of the Sampling Distribution

Under the Central Limit Theorem, for large sample sizes n_1 and n_2,

$$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

This normal approximation allows us to perform hypothesis testing and construct confidence intervals for $p_1 - p_2$.