Difference of Two Proportions

Colby Community College

Example 1

Consider an experiment for patients who underwent cardiopulmonary resucitation (CPR) for a hear attack and we subsequently admitted to a hospital.

These patients were randomly divided into a treatment group, where they received a blood thinner or the control group where they did not receive a blood thinner.

The variable of interest was whether they survived for at least 24 hours.

We really have two samples, and hence two sample proportions:

- The proportion of the treatment group that survived: $\hat{p}_t$
- The proportion of the control group that survived: $\hat{p}_c$

How would we determine if blood thinners actually make a difference with these patients?

We can look at $\hat{p}_t - \hat{p}_c$.

But, what we really want to know is, if blood thinners have an effect of heart attack survival rates in the general population?

Note

The best point estimate for $p_1 - p_2$ is $\hat{p}_1 - \hat{p}_2$.

Conditions For The Sampling Distribution Of $\hat{p}_1 - \hat{p}_2$ To Be Normal

$\hat{p}_1 - \hat{p}_2$ can be modeled using a normal distribution when:

Independence: The data are independent within and between the two groups. Generally this is satisfied if the data come from a randomized experiment.

Success-Failure: The success-failure condition holds for both groups, where we check successes and failures in each group separately.

When these conditions are satisfied, the standard error of $\hat{p}_1 - \hat{p}_2$ is

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

where p_1 and p_2 represent the population proportions, and n_1 and n_2 represent the sample sizes.

Confidence Intervals for $\hat{p}_1 - \hat{p}_2$

When the independence and success-failure conditions are met, we can build confidence interval in the same general manner and before:

$$\begin{array}{ccc} \text{point estimate} & \pm & z^* \cdot SE \\ & \Downarrow & \\ \hat{p}_1 - \hat{p}_2 & \pm & z^* \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}} \end{array}$$

Example 2

We can summarize the results from the experiment in Example 1:

	Survived	Died	Total
Control	11	29	50
Treatment	14	26	40
Total	25	65	90

Is independence satisfied?

This is a randomized experiment, so yes.

Are the success-failure conditions satisfied?

The treatment group had 11 survivals and 29 deaths, and the control group had 14 survivals and 26 deaths. All are more than 10, so yes.

Example 2 (Continued)

Let us create a 90% confidence interval.

We first need to calculate the point estimate:

$$\hat{p}_t - \hat{p}_c = \frac{14}{40} - \frac{11}{50} = 0.35 - 0.22 = 0.13$$

Next, the standard error:

$$SE \approx \sqrt{\frac{0.35(1-0.35)}{40} + \frac{0.22(1-0.22)}{50}} = 0.095$$

Recall that the critical value for 90% confidence is 1.65.

The confidence interval is:

$$\hat{p}_t - \hat{p}_c \pm z^* \cdot SE \rightarrow 0.13 \pm 1.65 \cdot 0.095 \rightarrow (-0.027, 0.287)$$

We are 90% confident that blood thinners have a difference of -2.7% to 28.7% percentage point impact on survival rate for patients.

What can be conclude about whether blood thinners help or harm?

Since 0% is in the confidence interval, we don't have enough evidence to say if blood thinners had any impact.

Example 3

A 5-year experiment was conducted to evaluate the effectiveness of fish oils on reducing cardiovascular events, where each subject was randomized into one of two groups.

We'll consider heart attack outcomes in these patients:

	heart attack	no event	Total
fish oil	145	12788	12933
placebo	200	12738	12938