



## **Section 6.2: Difference of Two Proportions**

Diez, D. M., Çetinkaya-Rundel, M., Barr, C. D. (2019). OpenIntro Statistics (4th ed.). OpenIntro.  
<https://www.openintro.org/book/os/> CC BY-SA 3.0

STAT 1201  
Introduction to Probability and Statistics

ONLINE AND DISTANCE EDUCATION

## 6.2 Difference of two proportions

We would like to extend the methods from Section 6.1 to apply confidence intervals and hypothesis tests to differences in population proportions:  $p_1 - p_2$ . In our investigations, we'll identify a reasonable point estimate of  $p_1 - p_2$  based on the sample, and you may have already guessed its form:  $\hat{p}_1 - \hat{p}_2$ . Next, we'll apply the same processes we used in the single-proportion context: we verify that the point estimate can be modeled using a normal distribution, we compute the estimate's standard error, and we apply our inferential framework.

### 6.2.1 Sampling distribution of the difference of two proportions

Like with  $\hat{p}$ , the difference of two sample proportions  $\hat{p}_1 - \hat{p}_2$  can be modeled using a normal distribution when certain conditions are met. First, we require a broader independence condition, and secondly, the success-failure condition must be met by both groups.

#### CONDITIONS FOR THE SAMPLING DISTRIBUTION OF $\hat{p}_1 - \hat{p}_2$ TO BE NORMAL

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

- *Independence, extended.* The data are independent within and between the two groups. Generally this is satisfied if the data come from two independent random samples or if the data come from a randomized experiment.
- *Success-failure condition.* 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.

### 6.2.2 Confidence intervals for $p_1 - p_2$

We can apply the generic confidence interval formula for a difference of two proportions, where we use  $\hat{p}_1 - \hat{p}_2$  as the point estimate and substitute the  $SE$  formula:

$$\text{point estimate} \pm z^* \times SE \quad \rightarrow \quad \hat{p}_1 - \hat{p}_2 \pm z^* \times \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

We can also follow the same Prepare, Check, Calculate, Conclude steps for computing a confidence interval or completing a hypothesis test. The details change a little, but the general approach remain the same. Think about these steps when you apply statistical methods.

**EXAMPLE 6.11**

We consider an experiment for patients who underwent cardiopulmonary resuscitation (CPR) for a heart attack and were 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 outcome variable of interest was whether the patients survived for at least 24 hours. The results are shown in Figure 6.1. Check whether we can model the difference in sample proportions using the normal distribution.

E

We first check for independence: since this is a randomized experiment, this condition is satisfied.

Next, we check the success-failure condition for each group. We have at least 10 successes and 10 failures in each experiment arm (11, 14, 39, 26), so this condition is also satisfied.

With both conditions satisfied, the difference in sample proportions can be reasonably modeled using a normal distribution for these data.

	Survived	Died	Total
Control	11	39	50
Treatment	14	26	40
Total	25	65	90

Figure 6.1: Results for the CPR study. Patients in the treatment group were given a blood thinner, and patients in the control group were not.

**EXAMPLE 6.12**

Create and interpret a 90% confidence interval of the difference for the survival rates in the CPR study.

We'll use  $p_t$  for the survival rate in the treatment group and  $p_c$  for the control group:

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

We use the standard error formula provided on page 217. As with the one-sample proportion case, we use the sample estimates of each proportion in the formula in the confidence interval context:

E

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

For a 90% confidence interval, we use  $z^* = 1.65$ :

$$\text{point estimate} \pm z^* \times SE \rightarrow 0.13 \pm 1.65 \times 0.095 \rightarrow (-0.027, 0.287)$$

We are 90% confident that blood thinners have a difference of -2.7% to +28.7% percentage point impact on survival rate for patients who are like those in the study. Because 0% is contained in the interval, we do not have enough information to say whether blood thinners help or harm heart attack patients who have been admitted after they have undergone CPR.

**GUIDED PRACTICE 6.13**

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 treatment groups. We'll consider heart attack outcomes in these patients:

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

Create a 95% confidence interval for the effect of fish oils on heart attacks for patients who are well-represented by those in the study. Also interpret the interval in the context of the study.<sup>16</sup>

**6.2.3 Hypothesis tests for the difference of two proportions**

A mammogram is an X-ray procedure used to check for breast cancer. Whether mammograms should be used is part of a controversial discussion, and it's the topic of our next example where we learn about 2-proportion hypothesis tests when  $H_0$  is  $p_1 - p_2 = 0$  (or equivalently,  $p_1 = p_2$ ).

A 30-year study was conducted with nearly 90,000 female participants. During a 5-year screening period, each woman was randomized to one of two groups: in the first group, women received regular mammograms to screen for breast cancer, and in the second group, women received regular non-mammogram breast cancer exams. No intervention was made during the following 25 years of the study, and we'll consider death resulting from breast cancer over the full 30-year period. Results from the study are summarized in Figure 6.2.

If mammograms are much more effective than non-mammogram breast cancer exams, then we would expect to see additional deaths from breast cancer in the control group. On the other hand, if mammograms are not as effective as regular breast cancer exams, we would expect to see an increase in breast cancer deaths in the mammogram group.

	Death from breast cancer?	
	Yes	No
Mammogram	500	44,425
Control	505	44,405

Figure 6.2: Summary results for breast cancer study.

**GUIDED PRACTICE 6.14**

Is this study an experiment or an observational study?<sup>17</sup>

<sup>16</sup> Because the patients were randomized, the subjects are independent, both within and between the two groups. The success-failure condition is also met for both groups as all counts are at least 10. This satisfies the conditions necessary to model the difference in proportions using a normal distribution.

Compute the sample proportions ( $\hat{p}_{\text{fish oil}} = 0.0112$ ,  $\hat{p}_{\text{placebo}} = 0.0155$ ), point estimate of the difference ( $0.0112 - 0.0155 = -0.0043$ ), and standard error ( $SE = \sqrt{\frac{0.0112 \times 0.9888}{12933} + \frac{0.0155 \times 0.9845}{12938}} = 0.00145$ ). Next, plug the values into the general formula for a confidence interval, where we'll use a 95% confidence level with  $z^* = 1.96$ :

$$-0.0043 \pm 1.96 \times 0.00145 \rightarrow (-0.0071, -0.0015)$$

We are 95% confident that fish oils decreases heart attacks by 0.15 to 0.71 percentage points (off of a baseline of about 1.55%) over a 5-year period for subjects who are similar to those in the study. Because the interval is entirely below 0, the data provide strong evidence that fish oil supplements reduce heart attacks in patients like those in the study.

<sup>17</sup>This is an experiment. Patients were randomized to receive mammograms or a standard breast cancer exam. We will be able to make causal conclusions based on this study.

## GUIDED PRACTICE 6.15

G

Set up hypotheses to test whether there was a difference in breast cancer deaths in the mammogram and control groups.<sup>18</sup>

In Example 6.16, we will check the conditions for using a normal distribution to analyze the results of the study. The details are very similar to that of confidence intervals. However, when the null hypothesis is that  $p_1 - p_2 = 0$ , we use a special proportion called the **pooled proportion** to check the success-failure condition:

$$\begin{aligned}\hat{p}_{pooled} &= \frac{\# \text{ of patients who died from breast cancer in the entire study}}{\# \text{ of patients in the entire study}} \\ &= \frac{500 + 505}{500 + 44,425 + 505 + 44,405} \\ &= 0.0112\end{aligned}$$

This proportion is an estimate of the breast cancer death rate across the entire study, and it's our best estimate of the proportions  $p_{mgm}$  and  $p_{ctrl}$  if the null hypothesis is true that  $p_{mgm} = p_{ctrl}$ . We will also use this pooled proportion when computing the standard error.

## EXAMPLE 6.16

Is it reasonable to model the difference in proportions using a normal distribution in this study?

E

Because the patients are randomized, they can be treated as independent, both within and between groups. We also must check the success-failure condition for each group. Under the null hypothesis, the proportions  $p_{mgm}$  and  $p_{ctrl}$  are equal, so we check the success-failure condition with our best estimate of these values under  $H_0$ , the pooled proportion from the two samples,  $\hat{p}_{pooled} = 0.0112$ :

$$\begin{aligned}\hat{p}_{pooled} \times n_{mgm} &= 0.0112 \times 44,925 = 503 & (1 - \hat{p}_{pooled}) \times n_{mgm} &= 0.9888 \times 44,925 = 44,422 \\ \hat{p}_{pooled} \times n_{ctrl} &= 0.0112 \times 44,910 = 503 & (1 - \hat{p}_{pooled}) \times n_{ctrl} &= 0.9888 \times 44,910 = 44,407\end{aligned}$$

The success-failure condition is satisfied since all values are at least 10. With both conditions satisfied, we can safely model the difference in proportions using a normal distribution.

USE THE POOLED PROPORTION WHEN  $H_0$  IS  $p_1 - p_2 = 0$ 

When the null hypothesis is that the proportions are equal, use the pooled proportion ( $\hat{p}_{pooled}$ ) to verify the success-failure condition and estimate the standard error:

$$\hat{p}_{pooled} = \frac{\text{number of "successes"}}{\text{number of cases}} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$$

Here  $\hat{p}_1 n_1$  represents the number of successes in sample 1 since

$$\hat{p}_1 = \frac{\text{number of successes in sample 1}}{n_1}$$

Similarly,  $\hat{p}_2 n_2$  represents the number of successes in sample 2.

In Example 6.16, the pooled proportion was used to check the success-failure condition.<sup>19</sup> In the next example, we see the second place where the pooled proportion comes into play: the standard error calculation.

<sup>18</sup> $H_0$ : the breast cancer death rate for patients screened using mammograms is the same as the breast cancer death rate for patients in the control,  $p_{mgm} - p_{ctrl} = 0$ .

$H_A$ : the breast cancer death rate for patients screened using mammograms is different than the breast cancer death rate for patients in the control,  $p_{mgm} - p_{ctrl} \neq 0$ .

<sup>19</sup>For an example of a two-proportion hypothesis test that does not require the success-failure condition to be met, see Section 2.3.

**EXAMPLE 6.17**

Compute the point estimate of the difference in breast cancer death rates in the two groups, and use the pooled proportion  $\hat{p}_{pooled} = 0.0112$  to calculate the standard error.

The point estimate of the difference in breast cancer death rates is

$$\begin{aligned}\hat{p}_{mgm} - \hat{p}_{ctrl} &= \frac{500}{500 + 44,425} - \frac{505}{505 + 44,405} \\ &= 0.01113 - 0.01125 \\ &= -0.00012\end{aligned}$$

The breast cancer death rate in the mammogram group was 0.012% less than in the control group. Next, the standard error is calculated *using the pooled proportion*,  $\hat{p}_{pooled}$ :

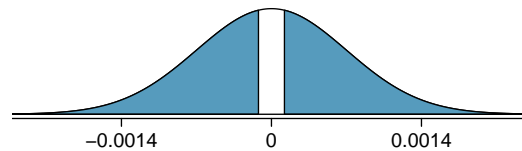
$$SE = \sqrt{\frac{\hat{p}_{pooled}(1 - \hat{p}_{pooled})}{n_{mgm}} + \frac{\hat{p}_{pooled}(1 - \hat{p}_{pooled})}{n_{ctrl}}} = 0.00070$$

**EXAMPLE 6.18**

Using the point estimate  $\hat{p}_{mgm} - \hat{p}_{ctrl} = -0.00012$  and standard error  $SE = 0.00070$ , calculate a p-value for the hypothesis test and write a conclusion.

Just like in past tests, we first compute a test statistic and draw a picture:

$$Z = \frac{\text{point estimate} - \text{null value}}{SE} = \frac{-0.00012 - 0}{0.00070} = -0.17$$



The lower tail area is 0.4325, which we double to get the p-value: 0.8650. Because this p-value is larger than 0.05, we do not reject the null hypothesis. That is, the difference in breast cancer death rates is reasonably explained by chance, and we do not observe benefits or harm from mammograms relative to a regular breast exam.

Can we conclude that mammograms have no benefits or harm? Here are a few considerations to keep in mind when reviewing the mammogram study as well as any other medical study:

- We do not accept the null hypothesis, which means we don't have sufficient evidence to conclude that mammograms reduce or increase breast cancer deaths.
- If mammograms are helpful or harmful, the data suggest the effect isn't very large.
- Are mammograms more or less expensive than a non-mammogram breast exam? If one option is much more expensive than the other and doesn't offer clear benefits, then we should lean towards the less expensive option.
- The study's authors also found that mammograms led to overdiagnosis of breast cancer, which means some breast cancers were found (or thought to be found) but that these cancers would not cause symptoms during patients' lifetimes. That is, something else would kill the patient before breast cancer symptoms appeared. This means some patients may have been treated for breast cancer unnecessarily, and this treatment is another cost to consider. It is also important to recognize that overdiagnosis can cause unnecessary physical or emotional harm to patients.

These considerations highlight the complexity around medical care and treatment recommendations. Experts and medical boards who study medical treatments use considerations like those above to provide their best recommendation based on the current evidence.

## 6.2.4 More on 2-proportion hypothesis tests (special topic)

When we conduct a 2-proportion hypothesis test, usually  $H_0$  is  $p_1 - p_2 = 0$ . However, there are rare situations where we want to check for some difference in  $p_1$  and  $p_2$  that is some value other than 0. For example, maybe we care about checking a null hypothesis where  $p_1 - p_2 = 0.1$ . In contexts like these, we generally use  $\hat{p}_1$  and  $\hat{p}_2$  to check the success-failure condition and construct the standard error.

### GUIDED PRACTICE 6.19

G

A quadcopter company is considering a new manufacturer for rotor blades. The new manufacturer would be more expensive, but they claim their higher-quality blades are more reliable, with 3% more blades passing inspection than their competitor. Set up appropriate hypotheses for the test.<sup>20</sup>



Figure 6.3: A Phantom quadcopter.

Photo by David J (<http://flic.kr/p/oiWLNu>). CC-BY 2.0 license.

This photo has been cropped and a border has been added.

<sup>20</sup> $H_0$ : The higher-quality blades will pass inspection 3% more frequently than the standard-quality blades.  $p_{highQ} - p_{standard} = 0.03$ .  $H_A$ : The higher-quality blades will pass inspection some amount different than 3% more often than the standard-quality blades.  $p_{highQ} - p_{standard} \neq 0.03$ .

**EXAMPLE 6.20**

The quality control engineer from Guided Practice 6.19 collects a sample of blades, examining 1000 blades from each company, and she finds that 899 blades pass inspection from the current supplier and 958 pass inspection from the prospective supplier. Using these data, evaluate the hypotheses from Guided Practice 6.19 with a significance level of 5%.

First, we check the conditions. The sample is not necessarily random, so to proceed we must assume the blades are all independent; for this sample we will suppose this assumption is reasonable, but the engineer would be more knowledgeable as to whether this assumption is appropriate. The success-failure condition also holds for each sample. Thus, the difference in sample proportions,  $0.958 - 0.899 = 0.059$ , can be said to come from a nearly normal distribution.

The standard error is computed using the two sample proportions since we do not use a pooled proportion for this context:

$$SE = \sqrt{\frac{0.958(1 - 0.958)}{1000} + \frac{0.899(1 - 0.899)}{1000}} = 0.0114$$

In this hypothesis test, because the null is that  $p_1 - p_2 = 0.03$ , the sample proportions were used for the standard error calculation rather than a pooled proportion.

Next, we compute the test statistic and use it to find the p-value, which is depicted in Figure 6.4.

$$Z = \frac{\text{point estimate} - \text{null value}}{SE} = \frac{0.059 - 0.03}{0.0114} = 2.54$$

Using a standard normal distribution for this test statistic, we identify the right tail area as 0.006, and we double it to get the p-value: 0.012. We reject the null hypothesis because 0.012 is less than 0.05. Since we observed a larger-than-3% increase in blades that pass inspection, we have statistically significant evidence that the higher-quality blades pass inspection *more than 3%* as often as the currently used blades, exceeding the company's claims.

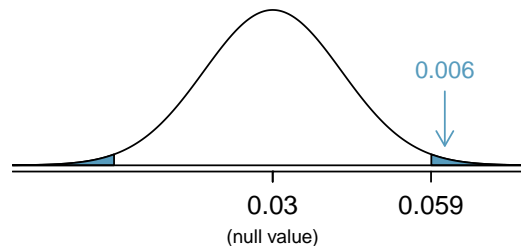


Figure 6.4: Distribution of the test statistic if the null hypothesis was true. The p-value is represented by the shaded areas.



## 6.2.5 Examining the standard error formula (special topic)

This subsection covers more theoretical topics that offer deeper insights into the origins of the standard error formula for the difference of two proportions. Ultimately, all of the standard error formulas we encounter in this chapter and in Chapter 7 can be derived from the probability principles of Section 3.4.

The formula for the standard error of the difference in two proportions can be deconstructed into the formulas for the standard errors of the individual sample proportions. Recall that the standard error of the individual sample proportions  $\hat{p}_1$  and  $\hat{p}_2$  are

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

The standard error of the difference of two sample proportions can be deconstructed from the standard errors of the separate sample proportions:

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

This special relationship follows from probability theory.

### GUIDED PRACTICE 6.21

Prerequisite: Section 3.4. We can rewrite the equation above in a different way:

$$SE_{\hat{p}_1 - \hat{p}_2}^2 = SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2$$

Explain where this formula comes from using the formula for the variability of the sum of two random variables.<sup>21</sup>

<sup>21</sup>The standard error squared represents the variance of the estimate. If  $X$  and  $Y$  are two random variables with variances  $\sigma_x^2$  and  $\sigma_y^2$ , then the variance of  $X - Y$  is  $\sigma_x^2 + \sigma_y^2$ . Likewise, the variance corresponding to  $\hat{p}_1 - \hat{p}_2$  is  $\sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2$ . Because  $\sigma_{\hat{p}_1}^2$  and  $\sigma_{\hat{p}_2}^2$  are just another way of writing  $SE_{\hat{p}_1}^2$  and  $SE_{\hat{p}_2}^2$ , the variance associated with  $\hat{p}_1 - \hat{p}_2$  may be written as  $SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2$ .