

[See this page in the course material.](#)

Learning outcomes

- Construct a confidence interval to estimate the difference between two population proportions (or the size of a treatment effect) when conditions are met. Interpret the confidence interval in context.
- Interpret the meaning of a confidence level associated with a confidence interval and describe how the confidence level affects the margin of error.
- Given the description of a statistical study, evaluate whether conclusions are reasonable.

Confidence Interval for a Difference in Two Population Proportions: Beyond the Basics

For all confidence intervals, the margin of error is based on the standard error. We know from “Distributions of Differences in Sample Proportions” that the standard error for the sampling distribution of differences in sample proportions is:

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

Obviously, if we are trying to estimate the difference in population proportions, we will not know p_1 or p_2 . So we estimate these population proportions with our sample proportions. This is the same approach we used when had to estimate the standard error for the distribution of sample proportions in *Inference for One Proportion*. The estimated standard error becomes

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

This formula estimates the average error between a difference in sample proportions and the true difference in population proportions.

So a 95% confidence interval has the following formula:

$$(\text{difference in sample proportions}) \pm 2(\text{standard error})$$

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

We can use this formula only if a normal model is a good fit for the sampling distribution. Recall that this is true only if the expected number of successes and failures in each sample is at least 10. For those who like formulas, these conditions translate into the following inequalities.

$$n_1\hat{p}_1 \geq 10, n_1(1-\hat{p}_1) \geq 10, n_2\hat{p}_2 \geq 10, n_2(1-\hat{p}_2) \geq 10$$

We have to adjust these conditions because we do not know the population proportions p_1 and p_2 . We make the same adjustment we made in *Inference for One Proportion*. We require that the *actual* number of successes and failures in each sample is at least 10. For those who like formulas, these conditions translate into replacing p_1 and p_2 with the corresponding sample proportions. Luckily, this tweak works and the normal distribution still gives fairly accurate confidence levels for different critical z-scores.

Try It

Nicotine Replacement Therapy

The Centre for Addiction and Mental Health in Canada posted the following description of a clinical trial on clinicaltrials.gov in September 2011.

This study will examine the efficacy of mailed distribution of free Nicotine Replacement Therapy to smokers. Telephone numbers will be randomly selected from across Canada in order to recruit adult smokers interested in completing a smoking survey and willing to be interviewed again in 8 weeks and 6 months times. Study participants will be asked about their smoking history and a hypothetical question: would they be interested in receiving the nicotine patch if this were to be provided to them free of charge? Participants expressing interest will be randomly assigned to one of two groups. One group will be offered the opportunity to actually receive a program of 5 weeks of nicotine patch for free right away and the other group will not be offered the free nicotine patches. The proportions of smokers in the two groups who quit smoking by the 6-month interview will be compared.

This is an [Select] .

Check Answer

[See this interactive in the course material.](#)

Inference procedures require the use of random samples or random assignment. Will the researcher be able to conduct inference procedures using data from this study? Explain.

Write your essay resp

Check Answer

[See this interactive in the course material.](#)

The researcher will be able to use the data from this study to draw which type(s) of conclusions?

☐ a conclusion that makes an inference about the impact of free nicotine patches on all smokers in Canada

[See this interactive in the course material.](#)

When we downloaded this description, researchers had not started the study. So we made up some data to practice finding a confidence interval to estimate the treatment effect. Suppose that 300 smokers express interest in receiving the free nicotine patches. The researchers randomly assign 150 to the treatment group. These people receive the free patches. The other 150 do not receive free nicotine patches. After 6-months suppose that 15 of the treatment group and 11 of the control group have quit smoking.

Find the 95% confidence interval to estimate the treatment effect (based on made-up data). Interpret your interval in context.

Write your essay resp

Check Answer

[See this interactive in the course material.](#)

Try It

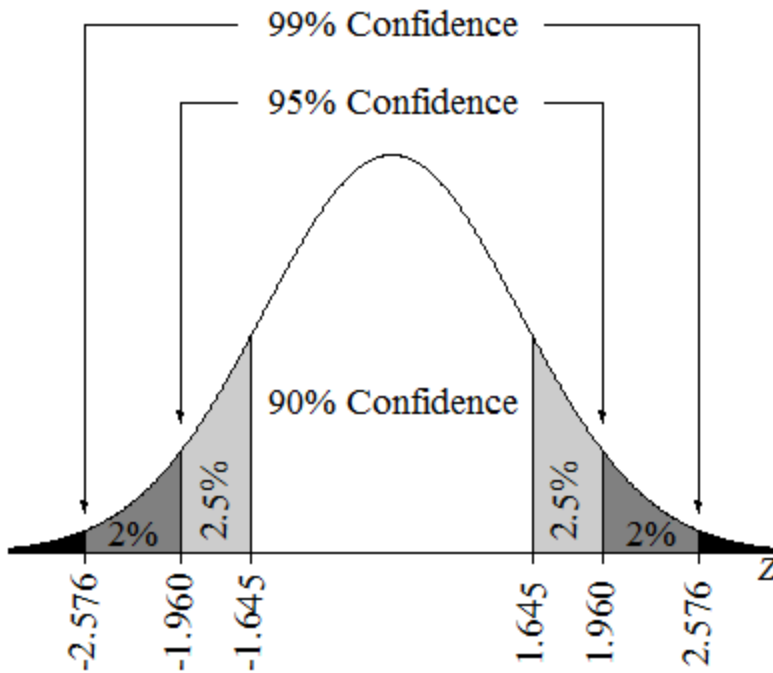
Suppose that we replicate this experiment a few times with varying sample sizes. Which confidence interval for the difference in the proportion of smokers who quit smoking ("treatment" minus "control") suggests that the treatment is effective?

☐ (0.011, 0.112)

[See this interactive in the course material.](#)

Other Levels of Confidence

In *Inference for One Proportion*, we saw that we can create confidence intervals for other levels of confidence. Changing the level of confidence changes the critical z-score. The following image shows the three most commonly used confidence levels and their critical z-scores.



The following table summarizes the critical values for the most commonly used confidence levels.

Confidence Level	Critical Value Z_c
90%	1.645
95%	1.960
99%	2.576

Note: A more exact value for the margin of error of a 95% confidence interval uses $Z_c = 1.96$ instead of 2 standard errors.

Try It

Recall the Abecedarian Early Intervention Project. In this experiment researchers selected 111 “high risk” infants based on mother’s education, family income and other factors. They randomly assigned 57 infants to receive 5-years of high quality pre-school. The remaining 54 infants were a control group. All children received nutritional supplements, social services, and health care to control the effects of these confounding factors on the outcomes of the experiment. By the age of 21 a much higher percentage of the treatment group enrolled in college, 42% vs. 20%.

Estimate the size of the treatment effect with a 90% confidence interval. (Be sure to check conditions for use of a normal model first!) Interpret your confidence interval in the context of this experiment.

Note: The standard error is approximately 0.085.

Write your essay resp

Check Answer

[See this interactive in the course material.](#)

Try It

What Is the Effect of Increasing the Confidence Level on the Margin of Error?

In an article titled “The Patriotism God Gap: Is the U.S. the Greatest Country in the World?” (*Christianity Today*, August 5, 2011), Tobin Grant cites data from the Pew Research Center. Here is an excerpt from the article:

About 40 percent of other Christians [non-evangelicals] said the U.S. stands alone as the greatest country. Those with no religion stand out as being much less likely to see the U.S. as the greatest country. Only 20 percent said the U.S. was the best country in the world.

The article does not give the sample sizes for these two groups. For this activity, let’s suppose the data describes random samples of 500 from the populations of “other Christians” and those with “no religion.” With samples this large, we can safely model the sampling distribution of

sample differences with a normal curve.

Use the simulation to find the margin of error for the 90%, 95%, and 99% confidence intervals. (Note: Conditions for use of the confidence interval formula are met because the sample size is large.)

Describe the effect of increasing the confidence level on the margin of error. Explain why the margins of error behave as they do.

Write your essay resp

Check Answer

[See this interactive in the course material.](#)

Licenses and Attributions

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by:** Open Learning Initiative. **Located at:** <http://oli.cmu.edu>. **License:** [CC BY: Attribution](#)

</div