

9.17: Hypothesis Test for Difference in Two Population Proportions (5 of 6)

Learning Objectives

- Given the description of a statistical study, evaluate whether conclusions are reasonable.

Thinking Critically about Conclusions from Statistical Studies

It is not uncommon to see debate over the conclusions and implications of statistical studies. When we read summaries of statistical studies, it is important to evaluate whether the conclusions are reasonable. Here we discuss two common pitfalls in drawing conclusions from statistical studies.

1. The conclusion is not appropriate to the study design.
2. The conclusion confuses statistical significance with practical importance.

We discuss these pitfalls in general, then look at examples that involve an inference about the difference between two population proportions or two treatments. But these pitfalls can happen with conclusions drawn from any inference procedure.

The Conclusion Is Not Appropriate to the Study Design

Here are several examples of this common pitfall.

1. **The study makes an inference based on nonrandom data.** If the data come from a sample that is not randomly selected or from groups that are not randomly assigned, we should not use the data in inference procedures. Why? Well, all inference procedures are based on probability. We can make probability statements only about random events, so the data must come from randomly selected or randomly assigned individuals if we want to make a statement about the population on the basis of the data. With nonrandom data, our main option is to analyze the data using exploratory data analysis (the ideas from Modules 2, 3, and 4).
2. **The study makes inappropriate cause-and-effect conclusions.** We can make cause-and-effect conclusions only with data from a randomized comparative experiment. If data comes from a single observational study, we cannot make cause-and-effect conclusions.
3. **The study overgeneralizes its conclusions.** If researchers randomly assign individuals to one of two treatments in an experiment, statistically significant results suggest the treatment is effective. This is an appropriate causal conclusion if the experiment is well designed. But if the original group of individuals is not randomly selected, then we should be cautious about generalizing this conclusion to a broader population.

The following table summarizes these ideas.

Learn By Doing

Do Energy Drink “Cocktails” Lead to Increased Injury Risk?

The following excerpt is from “Energy Drink ‘Cocktails’ Lead to Increased Injury Risk, Study Shows” (*Science Daily*, www.sciencedaily.com, Nov. 6, 2007).

College students who drink alcohol mixed with so-called “energy” drinks are at dramatically higher risk for injury and other alcohol-related consequences, compared to students who drink alcohol without energy drinks....The researchers found that students who consumed alcohol mixed with energy drinks were twice as likely to be hurt or injured, twice as likely to require medical attention, and twice as likely to ride with an intoxicated driver.

The study collected data from a Web-based survey of 4,271 students at 10 North Carolina colleges and universities.

Following are two summaries of this study. Are these summaries appropriate? Explain.

<https://assessments.lumenlearning.co...sessments/3945>

<https://assessments.lumenlearning.co...sessments/3946>

The Conclusion Confuses Statistical Significance with Practical Importance

Is a statistically significant difference always large enough to be important on a practical level? The answer is no.

Recall that when a P-value is less than the level of significance, we say the results are statistically significant. It means that the results are not due to chance. In the case of a difference in sample proportions, we are saying that the observed difference is larger

than we expect to see in random samples from populations with the same population proportions. But this does not necessarily mean the difference is large enough to be important in real life.

We also know that the P-value depends on the size of the sample. Results from large samples vary less, so an observed difference is more likely to be statistically significant if the samples are large. This means that a very small difference in population parameters can be detected by the hypothesis test as statistically significant. In this case, the population difference may be too small to be important to decisions we might make in real life. On the other hand, small random samples can have a lot of variability in their results. In this case, a large population difference may go undetected by the hypothesis test because a large sample difference may not be statistically significant. We have to be cautious that we don't confuse statistical significance with practical importance.

Example

Controversy about HPV Vaccine

Recall the earlier example about the debate between Republican presidential candidates in 2011. Michele Bachmann, one of the candidates, implied that the vaccine for human papillomavirus (HPV) is unsafe for children and can cause mental retardation. In response, *USA Today* published an article on September 19, 2011, titled "No Evidence HPV Vaccines Are Dangerous." The article describes two studies by the Centers for Disease Control and Prevention (CDC) that track the safety of the vaccine. Here is an excerpt from the article.

First, the CDC monitors reports to the Vaccine Adverse Event Reporting System, a database to which anyone can report a suspected side effect. CDC officials then investigate to see whether reported problems could possibly be caused by vaccines or are simply a coincidence. Second, the CDC has been following girls who receive the vaccine over time, comparing them with a control group of unvaccinated girls....Again, the HPV vaccine has been found to be safe.

We now examine "fake" data to demonstrate a couple of points. Suppose the CDC conducts a clinical trial to study the safety of the vaccine. Researchers select a random sample of girls and assign girls randomly to two groups: 1,000 girls get the vaccine, and 1,000 girls do not. Suppose that 6 girls in the vaccinated group develop serious health problems, and 1 girl in the unvaccinated group develops serious health problems.

Is this difference statistically significant at the 5% level?

Yes. If we use a statistical software package to find the P-value, we get a P-value of about 0.03. So the data supports the claim that the proportion of serious side effects is greater in the vaccine group (P is about 0.03).

Is the difference of practical importance? We investigate ways to think about this next.

Learn By Doing

Controversy about HPV Vaccine

Suppose a headline summarizing this experiment says "Vaccine Leads to Significantly Higher Risk of Serious Health Problems for Girls." Indicate whether each critique of this headline is valid or invalid.

<https://assessments.lumenlearning.co...sessments/3947>

<https://assessments.lumenlearning.co...sessments/3638>

<https://assessments.lumenlearning.co...sessments/3641>

Someone reading the headline "Vaccine Leads to Significantly Higher Risk of Serious Health Problems for Girls" might think that it is very risky to have the vaccine. Practical importance here involves how someone judges the risk of vaccination.

Here are two headlines that make statements about the risk associated with the vaccine. Indicate whether the statements are valid or invalid.

<https://assessments.lumenlearning.co...sessments/3642>

<https://assessments.lumenlearning.co...sessments/3643>

<https://assessments.lumenlearning.co...sessments/3948>

CC licensed content, Shared previously

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

This page titled [9.17: Hypothesis Test for Difference in Two Population Proportions \(5 of 6\)](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Bill Pelz](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.