Notes 11 - Confidence Intervals for Differences

STS 2300 Introduction to Data Analytics

Updated: 2025-04-14

Table of Contents

[Reading for Notes 11 1](#_Toc195553487)

[Learning Goals for Notes 11 1](#_Toc195553488)

[Constructing confidence intervals for 2](#_Toc195553489)

[Using a bootstrap distribution 3](#_Toc195553490)

[Using theory-based methods 4](#_Toc195553491)

[Constructing confidence intervals for 5](#_Toc195553492)

[Using a bootstrap distribution 5](#_Toc195553493)

[Using theory-based methods 7](#_Toc195553494)

[Interpreting confidence intervals for differences 8](#_Toc195553495)

[Situation 1 - Both endpoints are positive (or negative) 8](#_Toc195553496)

[Situation 2 - One positive and one negative endpoint 9](#_Toc195553497)

[Revisiting the Learning Goals for Notes 11 10](#_Toc195553498)

# Reading for Notes 11

The reading for Notes 09 - 11 is interwoven throughout [sections 8.4 - 8.7](https://moderndive.com/8-confidence-intervals.html#bootstrap-process) of the Modern Dive textbook.

# Learning Goals for Notes 11

* Be able to construct confidence intervals for differences in population proportions population means using bootstrap resampling (infer package) and using theory-based methods.
* Be able to accurately interpret confidence intervals for differences in population proportions and population means in context.

I’ll be using the following packages in this set of notes, so I’ll load them before I get started.

library(infer)  
library(dplyr)  
library(ggplot2)  
library(palmerpenguins)

# Constructing confidence intervals for

Let’s revisit the example from Activity 10 of the dog attempting to smell COVID-19. In the activity, we only considered cases where COVID-19 was present, but we might want to know if the dog was equally accurate for cases with and without COVID-19 present. This would mean we are interested in the following **parameter**:

= the true proportion of positive cases the dog can correctly identify minus the true proportion of negative cases the dog can correctly identify

The code below reads the full data into R.

* The ID variable tells us whether the dog said the sample was positive or negative for COVID-19.
* The actual variable tells us whether the sample really was positive or negative
* We are creating a variable called correct that will say "yes" or "no" based on whether the dog’s ID matches reality

covid\_dog <- data.frame(ID = c(rep("positive", 157), rep("negative", 792),  
 rep("positive", 33), rep("negative", 30)),  
 actual = c(rep("positive", 157), rep("negative", 792),  
 rep("negative", 33), rep("positive", 30))) |>   
 mutate(correct = ifelse(ID == actual, "yes", "no"))

## Using a bootstrap distribution

When we want to create a bootstrap distribution for a difference, we will have a variable that splits our data into two groups. This variable will go after the ~ in the specify() function. We will also need to use the order argument inside calculate() to say the order of subtraction for our difference.

In cases where we are interested in a difference in population proportions, our bootstrap distribution code will have the following form:

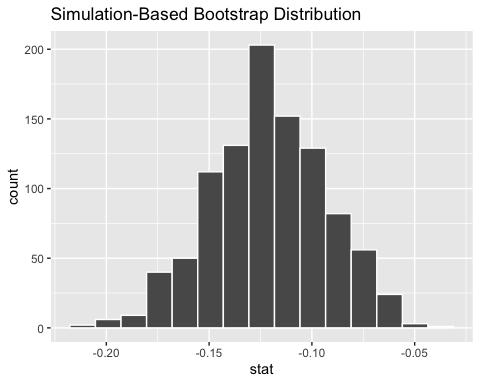
boot\_dist <- data |>   
 specify(formula = response ~ explanatory, success = "level") |>   
 generate(reps = 1000, type = "bootstrap") |>   
 calculate(stat = "diff in props", order = c("group 1", "group 2"))

In addition to changing boot\_dist, data, response, and level like we did in the single proportion case, we will also need to make the following changes:

* explanatory – the variable splitting our data into two groups
* group 1 – explanatory variable category to be considered our first group
* group 2 – explanatory variable category to be considered our second group

**Practice:** Try applying this to the COVID-19 sniffing dog example. Use positive samples as your first group and negative samples as your second group. (Note: I used seed 112300. You can use the same one to recreate my graph or choose your own to see how the graph and your corresponding intervals differ.)

set.seed(112300)  
dog\_boot <- covid\_dog |>   
 specify(formula = correct ~ actual, success = "yes") |>   
 generate(reps = 1000, type = "bootstrap") |>   
 calculate(stat = "diff in props", order = c("positive", "negative"))  
  
visualize(dog\_boot)



This bootstrap distribution represents 1,000 differences in sample proportions. Because my first group was positive samples and my second group was negative samples, the negative numbers I see mean the dog correctly identified a higher proportion of times when COVID-19 was **not** present than when it was. (Note: We will revisit this in more detail when we talk about interpreting intervals for differences.)

Just like in the previous chapter, I can use the percentile or standard error methods with the get\_ci() function to calculate my confidence interval for .

**Practice:** Calculate a 99% confidence interval using the percentile method.

**99% confidence interval:** (-0.191, -0.0467)

## Using theory-based methods

The prop.test() function can be used to calculate theory-based confidence intervals for a difference in proportions. The function will require us to specify the same three arguments (x, n, conf.level) that we did for a single proportion. However, now x and n will both be vectors of length 2. We can use the c() function to enter the numbers as a vector. In other words:

* x will consist of the number of successes for both groups
* n will consist of the sample size for each group

**Practice:** I’ve used the table() function to help find values for x and n. Use the results to make a 99% theory-based confidence interval for the difference in the proportion of correct identifications between samples with and without COVID-19.

table(covid\_dog$correct, covid\_dog$actual)

##   
## negative positive  
## no 33 30  
## yes 792 157

prop.test(x = c(157, 792), n = c(157+30, 792+33), conf.level = 0.99)

**99% confidence interval:** (-0.195, -0.0458)

Remember that theory-based methods have assumptions that must be met for the intervals to be valid. For the interval for the difference in proportions, we have the following assumptions:

* The data is a random sample.
* The sample size is “large enough” for both groups (usually at least 5 successes and 5 failures).

# Constructing confidence intervals for

For this example, we will revisit the penguins data from the palmerpenguins package. In Notes 10, we looked at the population mean bill length of Antarctic Penguins. Here we will compare the population mean bill length between female and male Adelie penguins.

The code below creates a data frame called adelie that subsets the penguins data from the palmerpenguins package to only include Adelie penguins and to remove all penguins missing information on their sex.

adelie <- filter(penguins, species == "Adelie",  
 !is.na(sex))

## Using a bootstrap distribution

In cases where we are interested in a difference in population means, our bootstrap distribution code will have the following form:

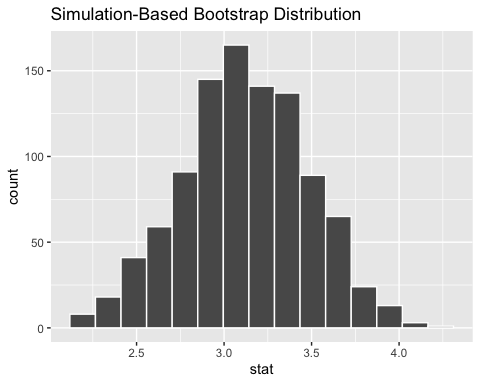
boot\_dist <- data |>   
 specify(formula = response ~ explanatory) |>   
 generate(reps = 1000, type = "bootstrap") |>   
 calculate(stat = "diff in means", order = c("group 1", "group 2"))

From this template, we will need to make the following changes:

* boot\_dist – replace with name for our object (if desired)
* data – replace with name of our data frame we are using
* response – replace with our *quantitative* variable of interest
* explanatory – replace with our *categorical* variable that determines our two groups
* group 1 and group 2 – replace with the two categories we are comparing

**Practice:** Fill in the blanks below to create a bootstrap distribution for a difference in population mean bill length between male and female Adelie penguins.

set.seed(8395722)  
adelie\_boot <- adelie |>   
 specify(formula = bill\_length\_mm ~ sex) |>   
 generate(reps = 1000, type = "bootstrap") |>   
 calculate(stat = "diff in means", order = c("male", "female"))  
  
visualize(adelie\_boot)



This bootstrap distribution represents 1,000 differences in sample means. Because my first group is male and my second group is female, positive numbers mean male penguins had a higher mean bill length than female penguins.

As with our other examples, I can use the percentile or standard error methods with the get\_ci() function to calculate a confidence interval for .

**Practice:** Calculate a 90% confidence interval using the SE method. (Note: Remember you will need to first calculate your point estimate.)

**90% confidence interval:** (2.55, 3.72)

## Using theory-based methods

The t.test() function, which we used to calculate theory-based confidence intervals for a mean, can also be used for differences in means. There are two ways we can use the function.

* Option 1 specifies our data frame in a data argument then references the column names in the formula notation.
* Option 2 ignores the data argument and instead uses $ notation to directly reference the vectors for our two variables.

# Option 1  
t.test(response ~ explanatory, data = dataname)  
  
# Option 2  
t.test(dataname$response ~ dataname$explanatory)

**Practice:** Use t.test() to calculate a 90% theory-based confidence interval for the difference in population mean bill lengths for female versus male Adelie penguins.

**Code / Answer:** t.test(bill\_length\_mm ~ sex, data = adelie, conf.level = 0.9)

90% CI: (-3.72, -2.54)

**Question:** What is different about how these intervals compared to my bootstrap intervals above? Why do you think this happened? (Hint: Think about our groups)

**Answer:** Our theory-based interval was entirely negative numbers. This is because the default is to put the groups in alphabetical order (female – male). We want to make sure we’re aware of what order corresponds with each interval since it will impact our interpretations.

# Interpreting confidence intervals for differences

When we’re interpreting an interval for a difference in two numbers, we have to consider whether our interval contains positive numbers, negative numbers, or a mix of the two. This is because the values within our interval represent differences.

Recall that:

* When a difference is positive, it indicates the first group being compared has the larger value
* When a difference is negative, it indicates the second group being compared has the larger value

Let’s start by considering an interval that contains all positive numbers or all negative numbers (instead of a mix of the two).

## Situation 1 - Both endpoints are positive (or negative)

If our confidence interval has endpoints that are either both positive or both negative, we could give an interpretation like this:

**Generic Interpretation:** We are xx% confident that the population \_\_\_ is between \_\_\_ and \_\_\_ bigger for our group with the higher mean/proportion than for our group with the smaller mean/proportion

Parameter described in words.

We will need to put this into context for each specific example.

My 90% confidence interval for the difference in Adelie penguin population mean bill lengths (male - female) was: (2.538, 3.727). Because all of the values in my interval are positive, I’m confident that males (group 1) have a longer population mean bill length than females (group 2). My interpretation would look like:

**Diff in means interpretation in context:** We are 90% confident that the population mean bill length for Adelie penguins is between 2.538 mm and 3.727 mm longer for male penguins than for female penguins.

Note: we will not use negative numbers in our interpretations. We will convert them to positive and add words like “longer” or “shorter” to convey what the negative represented.

**Practice:** Try writing an interpretation for our COVID-smelling dog example (positive - negative). My interval was (-0.2, -0.057). (Hint: First figure out which group we think the dog was better at identifying.)

**Answer:** We are 99% confident that the population proportion of cases that the dog can successfully identify is between 0.057 and 0.2 less for cases where COVID-19 was present than for cases where COVID-19 was not present.

Alternative version: We are 99% confident that the dog is successful in identification between 5.7 and 20 percentage points more often for cases without COVID-19 than for cases with it.

[We use the phrase “percentage points” instead of “percent” when we are talking about a difference or adding two numbers.]

## Situation 2 - One positive and one negative endpoint

In some cases, we might have a confidence interval where the lower endpoint is negative and the upper endpoint is positive. This suggests that we aren’t confident either group has a higher population mean/proportion. However, we can still provide context about *how big* any difference might reasonably be.

**Generic Interpretation:** We cannot be xx% confident that either group has a bigger population mean/proportion. However, we are xx% confident that the population \_\_ is anywhere from \_\_ bigger for group 1 to \_\_ bigger for group 2.

Let’s revisit the penguin bill length example (male - female). Consider the following hypothetical confidence intervals.

* (-0.1, 0.2) – In this case, I’m very confident that the population mean bill lengths are very similar even if they aren’t exactly the same.
* (-4, 3) – In this case, I don’t really know how these population mean bill lengths compare. They could be the same or either sex could have a much longer bill length than the other.
* (-0.1, 5) – In this case, **IF** one of the population mean bill lengths is much longer than the other, I’m pretty sure it’s group 1.

**Question:** How might our understanding of Adelie penguin bill lengths practically differ based on these three intervals?

**Answer:** See above.

**Practice:** Pick one of these and write an interpretation in context of the example.

**Interpretation:** Below are a couple possible ways to write the interpretation. For each one, you could replace the numbers with one of the other intervals above.

* We cannot be 99% confident that the population mean bill length of Adelie penguins is different between male and female penguins. However, we can be 99% confident that the population mean bill length is anywhere from 0.1 mm longer for female Adelie penguins to 0.2 mm longer for male Adelie penguins.
* We cannot be 99% confident that the population mean bill length of Adelie penguins is different between male and female penguins. However, we can be 99% confident that the population mean bill length is anywhere from 0.1 mm shorter to 0.2 mm longer for male Adelie penguins than for female Adelie penguins.

# Revisiting the Learning Goals for Notes 11

* Be able to construct confidence intervals for differences in population proportions population means using bootstrap resampling (infer package) and using theory-based methods.
* Be able to accurately interpret confidence intervals for differences in population proportions and population means in context.