Homework 8

your name here

Due Mon, Apr 18, 2022 at 11:00 PM in D2L

Instructions

Use this .Rmd file as a template for your homework. Please use D2L to turn in both the Knitted PDF output and your R Markdown file. Your .Rmd file should compile on its own if it is downloaded by your instructor.

Load packages and data

```
library(tidyverse)
library(openintro)
```

Exercises

Baby weight vs. smoking

The following problems use the `ncbirths` data set explored in Lab 08. For these exercises, we will use the entire data set (not just white mothers as in the lab).

Consider the possible relationship between a mother's smoking habit and the weight of her baby.

1. Create a version of the `ncbirths` dataset omitting observations where there are NAs for `habit`. You can call this version `ncbirths_habitgiven`.
2. Plotting the data is a useful first step because it helps us quickly visualize trends, identify strong associations, and develop research questions. Create an appropriate plot displaying the relationship between `habit` and `weight`. What does the plot highlight about the relationship between these two variables?
3. Calculate the mean `weight` in each of the `habit` groups. How do these two means compare?

There is an observed difference in sample means, but is this difference statistically significant? That is, if there really was no relationship between a mother's smoking habit and the weight of her baby, is this difference in sample means unlikely to occur just by chance? In order to answer this question you will conduct a hypothesis test.

4. Using standard statistical notation, write the hypotheses for testing if the average weights of babies born to smoking and non-smoking mothers are different. Hint: Mathematical notation can be created in your .Rmd file using LaTeX code, which is surrounded by `$`. For instance, the alternative hypothesis displayed before Exercise 3 in Lab 08 was created with the code `$H_A: \mu \neq 7.43$`.
5. Use bootstrapping to construct a 95% confidence interval for the difference between the average weights of babies born to smoking and non-smoking mothers.
6. Use your interval from the previous exercise to state a conclusion to the hypothesis test.

Extra Credit

Write and run code that will carry out a randomization test for these data. If you are unfamiliar with a randomization test for this scenario, or you just want a refresher, review Section 6.3.1 of the STAT 216 online textbook.

You may write code using only base R functions, or you can explore the `tidymodels` or `infer` packages, which include functions to run simulation-based hypothesis tests.