Homework 3

P8130 Fall 2022

Due: October 23, 2022 at midnight EST

P8130 Guidelines for Submitting Homework

- Your homework must be submitted through Courseworks. No email submissions!
- Only one PDF file should be submitted, including all derivations, graphs, output, and interpretations. When handwriting is allowed (this will be specified), scan the derivations and merge ALL PDF files (http: //www.pdfmerge.com/).
- You are encouraged to use R for calculations, but you must show all mathematical formulas and derivations. Please include the important parts of your R code in the PDF file but also submit your full, commented code as a separate R/RMD file.
- To best follow these guidelines, we suggest using Word (built in equation editor), R Markdown, Latex, or embedding a screenshot or scanned picture to compile your work.

DO NOT FORGET: You are encouraged to collaborate on homeworks, explain things to each other, and test each other's knowledge. But Do NOT hand out answers to someone who has not done any work. Everyone ought to have ideas about the possible answers or at least some thoughts about how to probe the problem further. Write your own solutions!

Use the `birthwt` data set from the `{MASS}` package for Problems 1-5. See `?birthwt` for variable definitions.

Problem 1 (10 points)

Some medical professionals claim that the average weight of American women is 171 pounds. The column `lwt` holds the mother's weight (in pounds) at last menstrual period, i.e. her pre-pregnancy weight. Use this column for the following questions.

a) Construct a 95% confidence interval of true mean weight of American women.

b) Interpret the confidence interval.

c) Comment on the validity of the statement above ("Some medical professionals claim that the average weight of American women is 171 pounds"). In other words, what can we say about this statement given our confidence interval from part a?

Problem 2 (10 points)

In this data set, we have a variable (`smoke`) indicating the smoking status of the mothers during pregnancy. Some doctors believe that smoking status is related to weight. Using the columns `smoke` and `lwt`, test this claim. *(Note: a value of 1 indicates the mother is in the "smoking" group.)*

a) Test for the equality of variances between the two groups. (Use a 5% significance level.)

b) Given your answer from part a, what kind of hypothesis test will you perform?

c) Conduct your chosen hypothesis test from part b at the 10% significance level. What is your decision regarding the null? Interpret this result in the context of the problem.

Problem 3 (6 points)

According to the CDC, approximately 20% of pregnant American women suffer from hypertension. Do our data support this claim? (Use column `ht` - a value of 1 means the mother is suffering from hypertension.)

a) Conduct a 99% confidence interval and interpret the results. What can we conclude about the CDC's claim from this interval?

b) Conduct a one-sided hypothesis test at the $\alpha = 0.1$ level. In this test, we want to see if the true proportion is indeed less than the claimed 20%. What can we conclude about the CDC's claim?

Problem 4 (4 points)

Is there a difference between uterine irritability in the group of pregnant women who smoke vs the group of pregnant women that don't smoke? (Use columns `ui` and `smoke`.)

Conduct a hypothesis test at the $\alpha = 0.01$ level. What can we conclude about the proportions of women with uterine irritability between the smoking groups?

Problem 5 (12 points)

Is race related to birth weight? (Use columns `race` and `bwt`.)

a) What test would be most appropriate to answer this question?

b) What assumptions are we making by using this test? Are all assumptions met?

c) Conduct the test at the 5% significance level and interpret your results. Be sure to write the hypotheses you are testing.

d) Perform multiple comparisons - which races are significantly different? Interpret your results.