

# Homework 3

P8130 Fall 2025

Due: October 14, 2025 at 11:59pm

## 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.

REMINDER: You are encouraged to collaborate on homework, explain things to each other, and test each other's knowledge. But **everyone must complete their own assignment and write their own solutions.**

Use the `birthwt` data set from the `{MASS}` package for Problems 1-3. See `?birthwt` for variable definitions.

## Problem 1 (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 using a 5% significance level.
- b) Given your answer from part (a), what kind of hypothesis test will you perform to compare means?
- 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.
- d) Construct a 95% confidence interval for the difference in means and provide a proper interpretation.

## **Problem 2 (10 points)**

Researchers are interested in exploring whether rates of uterine irritability in the group of pregnant women who smoke vs those who do not smoke? (Use columns `ui` and `smoke`.)

- a) Construct a 95% confidence interval for the rate of uterine irritability among pregnant women who smoke and interpret the interval.
- b) Construct a 95% confidence interval for the difference in the rate between those who smoke and those who do not. Interpret the interval.
- c) Conduct a hypothesis test for a difference in proportions at the  $\alpha = 0.01$  level. What can we conclude?

### **Problem 3 (10 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 can you claim are different from which? Interpret your results.

### Problem 4 (10 points)

A public-health team is evaluating a new community program intended to reduce average systolic blood pressure (SBP) among participants. The current population guideline level is  $\mu_0 = 130$  mmHg. Past studies indicate that the population standard deviation of SBP is  $\sigma = 15$  mmHg.

- a) The investigators will use a one-sample Z test and a 5% significance level to test  $H_0: \mu = 130$  vs.  $H_a: \mu < 130$ .
- b) If the true mean after the program is 125 mmHg, calculate the power of the test for  $n = 30$ . What is the power if the true mean is 122?
- c) What should the sample size  $n$  be to ensure at least 80% power if the true mean after the program is 124? What would it need to be if we needed 90% power?