

Programming Assignment 6

STAT 311

Please complete the following problems and submit a file named `STAT311-HW7.R` to Gradescope. You should start from the provided `STAT311-HW7.R` file on Canvas.

Overview

Address each of the following questions. As always, you should start from the provided skeleton code.

Question 1 - (6 points)

You are working as a data analyst for a private firm, and receive an email from your supervisor.

“Dear Employee” (You’d think he’d known your name by now)

“I sent you over some data from experiments on the tensile strength of different steel samples. I need you to determine if there is evidence that the treated steel has a tensile strength higher than $\mu = 377.9095$. Test it at the usual $\alpha = 5\%$ level.”

This data is saved as `tensileData`

- a) What is the difference between the observed and expected mean under the assumed hypothesis? Save your answer in the variable `q1.a`
- b) What is the estimated standard error of this test? Save your answer in the variable `q1.b`
- c) What is the test statistic T for this test? Save your answer in the variable `q1.c`
- d) What is the degrees of freedom for this test? Save your answer in the variable `q1.d`
- e) What is the p-value for this test? Save your answer in the variable `q1.e`
- f) TRUE or FALSE? Based on this p-value we should reject the null and conclude that the treated steel has a higher mean than specified. Save your answer in the variable `q1.f`

Question 2 - (7 points)

Your supervisor responds to your email about the analysis, but your not sure that he read it.

“Employee”

“Ignore my previous email, I actually needed you to compare the control and treatment groups of the data I sent you to see if the treatment group has a higher mean than the control group. The researchers tell me that they aren’t sure if the treatment impacted the variability of tensile strength, so don’t assume equal variances. This needs to be on my desk by end of day.”

(Note: Your test should look for the difference between $\mu_{treatment} - \mu_{control}$.)

- a) What is the difference between the observed difference of means and the expected difference of means under the assumed hypothesis? Save your answer in the variable `q2.a`
- b) What is the estimated standard error of this test? Save your answer in the variable `q2.b`
- c) What is the test statistic T for this test? Save your answer in the variable `q2.c`
- d) What is the conservative estimation of degrees of freedom for this test? Save your answer in the variable `q2.d`
- e) What is the p-value for this test using the conservative estimation of the degrees of freedom? Save your answer in the variable `q2.e`
- f) What is the p-value for this test using the accurate estimation of the degrees of freedom in the `t.test` function? Save your answer in the variable `q2.f`

- g) TRUE or FALSE? Based on this p-value we should reject the null and conclude that the treated steel has a higher mean than the control steel. Save your answer in the variable `q2.g`

Question 3 - (5 points)

Your supervisor again responds to your email about the analysis, but definitely didn't read it.

"Employee"

"It should have been obvious," (it wasn't) "but the values in this data set are paired together, with each row representing steel samples from different manufacturers. I'm going to need you to stay late and fix the error in your analysis to account for this paired design." (Maybe you should updated your linkedin profile later)

(Note: Your test should look for the difference between $\mu_{treatment} - \mu_{control}$.)

- a) What is the estimated standard error of this test? Save your answer in the variable `q3.a`
- b) What is the test statistic T for this test? Save your answer in the variable `q3.b`
- c) What is the degrees of freedom for this test? Save your answer in the variable `q3.c`
- d) What is the p-value for this test? Save your answer in the variable `q3.d`
- e) TRUE or FALSE? Based on this p-value we should reject the null and conclude that the treated steel has a higher mean than the control steel. Save your answer in the variable `q3.e`

Question 4 - (4 points)

You finally got a new job in a botany lab, doing data analysis for their team. (It's got better pay and benefits, but mostly you just wanted to get away from that supervisor). You receive the following email from your new supervisor.

"Hey new person! (Sorry, I don't know your name, but I promise to learn it soon!)"

"I sent some measurement data on some flowers that we were doing, and we wanted to know if you could do a regression analysis to determine how the sepal length of these irises differs based on the petal width. No rush on this, we're not running a sweat shop here! The attached document should have all the details."

This data is saved as `flowerData`

- a) What is the correlation between sepal length and petal width? Save your answer in the variable `q4.a`
- b) Using R functions or by hand, fit a regression equation to predict sepal length from petal width. What is the intercept of that equation? Save your answer in the variable `q4.b`
- c) What is the slope term of the above defined equation? Save your answer in the variable `q4.c`
- d) Using your estimated model, what average sepal length do you predict for an iris with a petal width of 1.75? Save your answer in the variable `q4.d`