# Homework 2

### **BSTA** 512/612

2024-02-01

Important

THIS PAGE IS UNDER CONSTRUCTION!! It's likely that I will be making changes to this assignment at this time!

#### **Directions**

- Download the .qmd file here.
- You will need to download the datasets. Use this link to download the homework datasets needed in this assignment. If you do not want to make changes to the paths set in this document, then make sure the files are stored in a folder named "data" that is housed in the same location as this homework .qmd file.
- Please upload your homework to Sakai. Upload both your .qmd code file and the rendered .html file
- For each question, make sure to include all code and resulting output in the html file to support your answers.
- Show the work of your calculations using R code within a code chunk. Make sure that both your code and output are visible in the rendered html file. This is the default setting.
- If you are computing something by hand, you may take a picture of your work and insert the image in this file. You may also use LaTeX to write it inline.
- Write all answers in complete sentences as if communicating the results to a collaborator. This means including a sentence summarizing results in the context of the research study.
  - Questions not requiring a sentence are

- \* Ch 7 # 1, 2, 5
- \* Ch 6 # 5, 6
- \* Ch 14 # 2, 12, 14



It is a good idea to try rendering your document from time to time as you go along! Note that rendering automatically saves your qmd file and rendering frequently helps you catch your errors more quickly.

### Question 1 (chapter 6)

Use the data from Chapter 5 Question 9 to answer the following questions. Use the log-transformed values as given in the dataset.

Note: the question numbers below do not refer to questions from the textbook. Complete the problems below instead of the ones in the book.

(1)

Create a scatterplot of the dependent and independent variables, and in words describe the their relationship. Is it reasonable to use a linear regression to model the relationship?

(2)

Find the correlation coefficient between the two variables. Is the value consistent with your description of the relationship in the previous question? Why or why not?

(3)

Test whether the two variables are significantly correlated. Do this using the formula and then check your work with R's test for correlations. Make sure to include the hypotheses and a conclusion.

(4)

Calculate the confidence interval for  $\rho$  using the formula and verify that it matches the confidence interval in R's test output. Include an interpretation of the confidence interval and also explain why the confidence interval is consistent with the p-value.

### **(5)**

Calculate the coefficient of determination using the ANOVA table output, and confirm that it matches the value in the R output (what R output shows this and what is it labeled as?).

### (6)

What is another way to calculate the coefficient of determination? Do the calculation and verify that you have the same answer.

## **(7)**

Give an interpretation of the coefficient of determination in the context of the study.

Note: the question numbers below do not refer to questions from the textbook. Complete the problems below instead of the ones in the book.