**Brooke Wheeler**

**2/9/22**

**Assignment #4**

**Due:** **Wednesday February 9**, 2022 by **5 PM ET**

**Please sign** your name to the **appropriate** space below**.** Remember that you are permitted to receive (and provide) authorized assistance but must acknowledge it if you do.

I received assistance on this assignment and/or discussed it with fellow classmates or a tutor.



**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_recieved help from TA\_\_\_\_\_\_\_\_\_\_\_**



I received no assistance on this assignment and/or did not discuss it with anyone other than Professor Miller.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**



**Part 1: The Regression Parameters**

**Instructions:** In these problems we are going to focus on the regression parameters, and , and their estimators, and . Provide complete justification for each problem. You may type your work or include work done by hand.

1. You are given observations for the explanatory variable and observations for the response variable. The relationship between the response variable and the explanatory variable can be described by the linear model .
2. For this model, what is the formula for the residual, ?

Timeline

Description automatically generated

1. Write the expression for the Error Sum of Squares, for this model.

Timeline

Description automatically generated

1. Derive the formula for the least squares estimate of . (Find the value of which minimizes the SSE. This is like what we derived in the notes, except in this case we only have to minimize the SSE with respect to one variable, b.)

Timeline

Description automatically generated

**Part 2: Inferences in Regression**

**Instructions:** For each question below, you must show all work. Use R to find p-values and critical values. You may type your work or include work done by hand.

1. (Continued from Assignment 3) The time it takes to transmit a file always depends on the file size. Suppose you transmitted 30 files, with the average size of 126 Kbytes and the standard deviation of 35 Kbytes. The average transmittance time was 0.04 seconds with the standard deviation of 0.01 seconds. The correlation coefficient between the time and the size was 0.86.

In the previous homework, we fit a regression model that predicted the time it will take to transmit a 400 Kbyte file. According to this model, the standard deviation of responses is estimated by

.

* 1. Construct a 95% confidence interval for the regression slope.

Diagram

Description automatically generated

The 95% confidence interval is from .00019 to .00039.

* 1. Based on this interval, is the slope significant at the 5% level?
     1. Based on the interval the slope is significant at the 5% level. Since zero is not included in the confidence interval we can reject the null hypothesis at the level of .05 that there is not linear association between file size and time it takes to send the file. Meaning we can support that there is a linear association.
  2. State the null and alternative hypotheses that would be used in b). Calculate the test statistic and the p-value.

Text, letter

Description automatically generated

* 1. When you answered questions b) and c), it was correct to conduct a two-sided test. However, in this given example, why does it make more sense to consider an upper-tail alternative?
     1. In this example it would make sense that file size effects the time it takes to send the file and it would make sense that the bigger the file size the longer it takes. Above we only found evidence to support the hypothesis that there is a linear association between file size and time it takes to send the file. If we did a upper-tail test our alternative hypothesis would be that the slope is greater than 0, meaning that there is a positive slope, that the larger the file size the longer it takes to send the file.

1. (Continued from Assignment 3) At a gas station, 180 drivers were asked to record the mileage of their cars and the number of miles per gallon. The results are summarized in the table.

|  |  |  |
| --- | --- | --- |
|  | Sample mean | Standard deviation |
| Mileage | 24,598 | 14,634 |
| Miles per gallon | 23.8 | 3.4 |

The sample correlation coefficient is *r* = −0*.*17. In the previous assignment, we fit a regression model that described how the number of miles per gallon depends on the mileage. According to this model, the standard deviation of responses is estimated by

1. Do the given data present a significant evidence that cars with higher mileage are less economic? Formulate appropriate null hypothesis and alternative and conduct the test.

Diagram, timeline

Description automatically generated

Since the p-value is less than alpha=.05 we reject the null hypothesis and support that there is a negative association between mileage and mpg. Meaning that as cars have higher mileage they get less mpg.

**Part 3: Mini-Project**

**Instructions:** You will need R to answer the following questions. For each, provide any relevant R output and clearly state a conclusion with full support for your answer.

1. (*based on text* ***p. 90-91:******2.4***)

**Grade point average** (This data set was already used in Assignment 3 and is posted on Canvas).

* 1. Obtain a 99% confidence interval for . Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?

Graphical user interface, text, application, email

Description automatically generated

* 1. Test whether or not a linear association exists between student’s ACT score and GPA at the end of the freshman year . Use a level of significance of 0.01. State the hypotheses, p-value, decision, and conclusion.

Text, letter

Description automatically generatedGraphical user interface, text, application, chat or text message

Description automatically generated

Since the p-value is less than alpha = .01 we reject the null hypothesis that there is no association between gpa and ACT scores and support that there is an association between ACT scores and GPA at the end of freshman year.