**HOMEWORK 5**

*Inference for Simple Linear Regression, Multiple regression*

Reading: This assignment focuses on content from your textbook, *STAT2: Building Models for a World of Data*, Chapter 2 Sections 1-2, 4 and Chapter 3 sections 1-3. Read these sections of your textbook.

Notes:

* You are required to use your own words in answering all homework questions. You cannot copy information from the book or other sources.
* Round all numbers to 3 decimal places unless otherwise specified.

Answer the following questions based on question 2.15 – Breakfast cereal in your book, on page 78.

The data set is cereal.jmp

1. Test (part a): report only the p-value: .0013
2. Choose the appropriate conclusion:

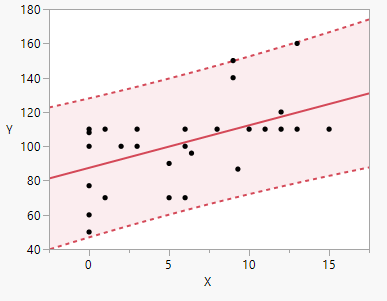
There is strong evidence of an association between sugar content and calories

1. Interval (part b):
   1. report lower bound 1.043
   2. And upper bound 3.918

No need for an interpretation of the interval.

1. Find the 95% confidence interval for the mean calories when sugar content is 6 grams.
   1. Lower bound 95.773
   2. Upper bound 108.853
2. Find the 95% prediction interval for calories when sugar content is 6 grams.
   1. Lower bound: 62.608
   2. Upper bound 142.012

1. Here is a plot of some data, with unnamed intervals around predicted Y values. Does this plot show prediction intervals (for individual values) or confidence intervals (for mean values)?



1. Explain your choice (free answer)

Individual values. You can see it tries to include all the data. The prediction interval for the mean is much narrower.

Here is an ANOVA table for a new data set with some values filled in:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | degrees of freedom | sum-of-squares | mean square | F statistic |
| Model | 1 |  | 125.1 |  |
| Error |  | 10.3 |  |  |
| Total | 52 | 135.4 |  |  |

1. Calculate the error degrees of freedom:

Your answer: 51

1. Calculate the Model sum-of-squares:

Your answer: 125.1

1. Calculate the F statistic. There is only one, for the Model:

Your answer: 655.864

Answer the following questions based on question 3.29 – Jury duty in your book, starting on page 146.

The data set is jurors.jmp

Create the graph described in part a of the book. Based on this graph, answer each as True or False. The questions about regression slopes are for the multiple regression predicting pctReport from the period number and indicator variable for year (1 = 2000 and 0 = 1998).

1. Later periods have a higher predicted percent report. True
2. The difference between 1998 and 2000 is about the same for all periods True
3. The regression slope for period is negative True
4. The regression slope for period is about the same value in both years True
5. The regression slope for the year indicator variable is positive True

Fit the multiple regression equation as described in part c of the question 3.29. Report the estimated

1. Slope for period: -.717
2. Slope for Year: 17.835

Consider the 4 regression conditions / assumptions. Assess them for the additive regression model:

1. Is it appropriate to assume independence Yes, residual by row
2. Is it appropriate to assume linearity Yes, looks linear
3. Is it appropriate to assume equal variances Yes, residual by predicted
4. Is it appropriate to assume normality Yeah, residual normal quantile

Note: By able to explain your answers to each, but that isn’t necessary for this HW.

1. Test the null hypothesis of no difference in predicted pctReport between 1998 and 2000. Report the p-value. If JMP reports < 0.0001, enter 0

0

1. What is the predicted percent who reported for jury duty in the 5th period in year 1998 based on the additive model? 73.497