**HOMEWORK 6**

*Interactions in multiple regression, Kelley’s rule for reporting values*

Reading: This assignment focuses on content from your textbook, *STAT2: Building Models for a World of Data*, Chapter 3 Section 4. 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.

Last week we looked at additive models for the Jurors data set (question 3.29 – Jury duty in your book, starting on page 146). We finish that analysis now by considering the model with an interaction between period and the year indicator variable.

Write out the full fitted multiple regression equation for **the interaction model** using period and the year indicator variable to predict the percent who report for jury duty as described in part d of the book. Remember last week we only fit the additive model to these data. Now you need to include the period\*year indicator interaction. Your answer will be one equation, not two. Use 1998 as the reference/baseline group.

1. Coefficient for the intercept 76.455
2. Coefficient for period -.668
3. Coefficient for the year indicator 19.147
4. Coefficient for the interaction -.0972

Use the multiple regression coefficients to write out the fitted equation for year 2000

1. Coefficient for the year 2000 intercept 95.572
2. Coefficient for the year 2000 slope -.765

1. Test the null hypothesis that the two years have the same slope for period. Report the p-value for this test.

P-value: .6995

Note: You should be able to write a conclusion based on this test.

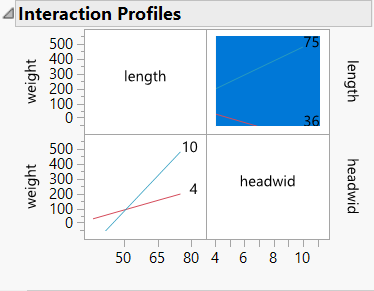
1. Report the value of the i2000 (year indicator) coefficient using Kelley’s rule to decide on the appropriate number of digits: 19
2. Imagine the standard error of the i2000 (year indicator) coefficient) was 1.973. What would you now report as the value of the i2000 coefficient: 19.1

Part 2: predicting the weight of a black bear. These questions give you practice interpreting quantitative x quantitative interactions. Weight is an important characteristic of a bear; it is also very difficult to measure on wild bears. That requires weighing objects weighing up to 400 pounds in the field. Other size measurements are much easier to take. The data in bear.jmp are for a simple random sample of bears in one particular area of Pennsylvania. We will focus on two explanatory variables: length = the body length of the bear and headwid = the width of the head.

1. Fit a model to predict weight using length, headwid and their interaction.

Report the estimated interaction coefficient.

1.941

1. Consider two bears, one (A) with a small headwidth (e.g. 5) and one (B) with a large headwidth (e.g. 9). Based on the interaction coefficient, which bear (A or B) has the larger length slope?
   1. Large bear. 9. B
2. Use JMP interaction plots to visualize how the relationship between weight and length changes with headwidth. Paste the relevant plot here:

Use the fitted multiple linear regression model to calculate the coefficients for the simple regression that uses length to predict weight for bears with headwidth = 5.

1. Intercept of the simple regression for headwidth=5: -225.068
2. Slope of the simple regression for headwidth=5: 6.2315

1. Test whether it is appropriate to include the interaction term in your model. Report the p-value.
   1. .0005

Evaluate the equal variance assumption and answer Yes or No:

1. Is it appropriate to assume equal variances? no

The investigators hear about a study done elsewhere that uses log length and log headwidth to predict log weight. Fit that regression model to these data. Include an interaction term in your model.

1. Test whether it is appropriate to include the interaction term in the model for log transformed data. Report the p-value (round to 2 digits). .73

Fit the additive model (no interaction) using log length and log headwin to predict log weight. Evaluate the linearity and equal variance assumptions for this model and answer Yes or No

1. Is it appropriate to assume linearity? Yes
2. Is it appropriate to assume equal variances? Yes

You will produce a report to your supervisors. You can only present one model. This model will be used to predict the weight of specific individual bears. Which model should you include? Your choices are A = the interaction model using untransformed variables or B = the additive model using log transformed variables. Check all appropriate statements

A because it assumes an interaction

A because logarithms are hard to interpret

A because B only predicts the median weight for a bear with certain measurements

B because it better satisfies the equal variance assumption

B because the model is simpler

B because additive models are easier to interpret