PSTAT 126 - Regression Analysis

Fall 2017

Data Project Guidelines

**Note:** These course materials are the sole property of Dr. Todd M. Gross. They are strictly for use by students enrolled in a course taught by Dr. Gross. They may not be altered, excerpted, distributed, or posted to any website or other document-sharing service.

**Due**: Thursday, Dec 7 in Lecture

**Value**: 15% of course grade

**Instructions**: You will complete this project with ONE partner and will turn in ONE report together.

**Description of Data**

A researcher is interested in predicting what makes people happy. He asks 100 volunteers to rate their happiness on a 10-point scale (1 = very unhappy, 10 = very happy). He also asks them their gender (0 = male, 1 = female), the number of hours they work each week, and to rate the quality of their love relationship(s) on a 10-point scale (1 = very lonely, 10 = deeply in love).

The dataset is contained in the **projdata.txt** file on Gauchospace. The four variables are named **happy, sex, workhrs,** and **relationship.**

**Assignment**

Find the most appropriate regression model for these data, and describe your results in a brief report.  You will be graded on your use of all the techniques learned in this course. These include:

1. Exploring the data with Scatterplots
2. Fitting the 1st order linear model to the data
3. Evaluating possible interactions through hypothesis testing
4. Using hypothesis tests to determine model fit, test individual regression coefficients, and to compare models
5. Using model fitting techniques to arrive at a final model
6. Using residuals to diagnose possible violations of assumptions of your final model, and remedying violations if they are found.
7. Describing your final model, including interpretation of regression coefficients, statistical fit of the model, and variance explained by the model.

**Report/Write-Up**

Your report must 1) describe the data and the research questions under study, 2) explain the statistical methods you used to analyze these data, 3) summarize your findings, 4) include appropriately labeled plots, and 5) provide a clear description of your final model.  Your report should be 3-5 pages long, not including an appendix which must contain your R code and output.

**Report Structure**

Names and Lab Section(s) – List the names and lab sections of both partners on the first page of your report.

Introduction (1-2 paragraphs)

Briefly describe the dataset. State your research questions. That is, predict the relationships you expect to see between the variables.

Method (1-2 paragraphs)

Describe the regression analyses you performed. Describe how you fit the initial model, how you diagnosed potential violations of the model assumptions, how you selected the best-fitting model, and how you evaluated the fit of that model.

Results (1 -2 pages)

Provide the results of your analysis including all appropriate descriptive and inferential statistics. Explain any plots you reviewed. Clearly describe your final model.

Discussion (1-2 paragraphs)

Explain the conclusions you can draw from the model, and interpret the meaning of the regression coefficients in terms of the variables being measured. Identify any limitations to the model, and any questions for future research.

Plots – Provide appropriate plots. Plots may be placed in the section where they are discussed, or in the appendix.

Appendix – Include the R code you used to analyze your data and the R output.

Your report needs to be double-spaced, 12pt font size, with 1-inch margins, and be

3 - 5 pages long (Note: maximum 5 pages, not including your Appendix). Your R code and output can be in 10pt font.