Topics:
- Describing Scatterplot Matrices and Multivariate Regression Analysis
- Lessons Covered: 46 - 50
- Textbook Chapter (Optional) : 13.4 and 13.5

Grading:
- Points are listed next to each question and should total 25 points.
- Late assignments will not be graded.

Deadlines:
- Final Submission: Monday, March 11th

Instructions:
- Download or view questions below.
- Clearly label and type answers, *without* question prompts, in word, google docs, PDF, or other word processing software.
- Insert diagrams or plots as a picture in an appropriate location.
- Math Formulas need to be typed with Math Type, LateX, or clearly using key board symbols such as +, -, *, /, sqrt() and ^
- Submit assignment Verify the correct document has been uploaded. If not, resubmit. You can submit up to three times.

Allowances:
- Any resources listed or posted in our class.
- You are encouraged to discuss the problems with other students, the instructor and TAs, however, all work must be your own words. Duplicate wording will be considered plagiarism.
- Outside Resources need to be cited. Websites such as Chegg, Coursehero, Koofers are discouraged, but if used need to be cited and within the boundaries of academic honesty.

The dataset ST314ExamData.csv represents the midterm and final exam grades for students in the ST314 online and campus courses, for the last three consecutive terms. Use this data to complete a multiple linear regression analysis in R and answer the following questions.

Final = Final Exam Score
Midterm = Midterm Exam Score,
Term = Has three levels W18, Sp18, F18
Format = Has two levels Campus and Online

Part 1. (6 points) Multivariate Visualization:

It is reasonable to consider that more than just midterm score may influence final exam score. Investigate the individual relationships between final exam score and the above explanatory variables. Use the R script Multivariate_Exam_Analysis.R to help you get started with the code.

    a.   Construct a scatterplot matrix including final and each of the explanatory variables.
        a.   (1 point) Paste the plot.
        b.   (1 point) Do any of the variables have a visual relationship with Final?
    b.   The scatterplot matrix is not all that helpful for the categorical variables Term and Format.
        a.   Create a side by side boxplot that looks at the relationship between Term and Final.
            i.   (1 point) Paste your plot.
            ii.   (1 point) Describe the relationship. Visually does Term seem to have a relationship with Final?
        b.   Create a side by side boxplot that looks at the relationship between Format and Final.
            i.   (1 point) Past your plot.
            ii.   (1 point) Describe the relationship. Visually does Format seem to have a relationship with Final?

Part 2. (7 points) Fit a Model

Fit a model that includes Term, Format and Midterm as explanatory variables for the response variable Final.

    a.   (1 point) Provide the R output of the model.
    b.   (2 points) State the least squares regression equation of your model.
    c.   The model without the variables Term and Format has an adjusted R^2 value of 0.4042.
        a.   (2 point) Does including the variables Term and Format improve the fit of the model?
        b.   (2 point) Interpret the adjusted R^2 value.

Part 3. (10 Points) Model interpretation.

Note: Model Interpretation can get tricky when there is more than two levels in a factor. For example, Term has three levels instead of two. The R output will designate this as VariableLevel, like "TermW18".

In the model, the coefficient for **TermW18** is **-2.45** this means that ***while the other variables in the model are held constant, a student taking the exam in the winter will score 2.45 points less on average than a fall 2018 student.*** We know to TermW18 is compared to fall, because F18 is the variable not included in the output. Meaning, fall is represented when spring and winter are both at 0.

    a. (2 point) Interpret each of the individual t tests by stating which variables are significant at 0.05, when the other variables are in the model.
    b. (2 points) Interpret in context $\beta_{FormatOnline}$ the coefficient for Format.
    c. (2 points) Interpret in context $\beta_{TermSp18}$ the coefficient for TermSp18.
    d. (2 point) Interpret in context $\beta_{midterm}$ the coefficient for the Midterm variable.
    e. (2 points) Calculate the 95% confidence interval for $\beta_{Midterm}$. Show work. Interpret the interval.

Part 4. (2 points) Prediction.

    a. (2 points) Use the least squares regression equation to predict final exam score for a winter, online student with a midterm score of 190.

(2 points) **Extra Credit.** Predict your final exam score based on your midterm score and whether you are an online or campus student. Consider W18 to represent you as a "winter" term student.