

This document is meant to serve as both instruction and guidelines for what should be included in your final project and presentation. As a reminder, the final project is worth 20% of your grade, 10% of which is in the presentation and 10% of which is in the written portion.

General Rules: I am leaving you the option to work in groups of 1, 2, or 3. The expectations for groups scale with size however:

- Group of 1: Single multiple regression analysis
- Group of 2: Two multiple regression analysis using different response variables
- Group of 3: Two multiple regression analysis using different response variables and some form of an ANOVA test

You are free to either pick a dataset from a selection I supply or choose and find your own from a topic that you have an interest in. I'm happy to suggest or help brainstorm possible places to find data, but I can't find it for you.

From your dataset, you will be responsible for formulating a question(s) of interest and attempting to answer said questions utilizing multiple regression techniques or ANOVA testing. You will present both your work and results in a short presentation to the class and in a written document to me. Class presentations should be about 6 minutes for a solo presentation, 8 for a group of 2 and 10 for a group of 3. Paper length should be between 3 and 10 pages.

Topics to include: Both presentation and the paper should include similar elements. These would include:

- ☐ An Introduction: Introduce your dataset and explain the question or questions you are looking to answer. Show at least a portion of the dataset and describe available variables.
- ☐ Methods:
 - How was the study designed? You might need to investigate a bit to try to unearth. Are there potential sources of bias that you should be aware of when doing your analysis? What population could the results be generalized to? Can anything be said about causation?
 - Consider all or most of the variables in the dataset. Map out conceptually how variables may be inter-related, paying important attention to what relationships might be causal or where confounding variables may occur. I'll put together an example of this to help guide you.
 - If doing some hypothesis or ANOVA testing, how did you decide what sort of test to perform. Were all the needed conditions fulfilled?
- ☐ Multiple regression results:

- Fit your “full” model (your response and all explanatory variables that seem interesting. Show the regression summary so p-values can be seen for all variables. Comment on the adjusted R^2 and non-significant variables that seem worth dropping. Keep dropping non-significant variables one at a time until the adjusted R^2 stops increasing or only significant variables are left.
- Show your summary for the final reduced model (or models, if you found several interesting ones)
- Show your residuals plot and comment on if your model satisfied needed conditions.
- As a major results visual, construct one “pretty” plot that shows at least three variables at once. Your response variable on the y-axis, one explanatory on the x-axis, and others numeric or categorical explanatory variables corresponding to marker size, color, or symbol. Make sure the text is large enough for people in the back to read as part of your slides.
- You are telling a story, so make sure you choose enough graphs (and the right graphs) to tell the story of your dataset and your question well.

□ ANOVA results:

- Be clear about what categorical variable you are investigating. How many groups in your categorical variable will you be testing for differences?
- Provide comparison boxplots of the response variable for the groups of interest.
- Show your ANOVA summary results and comment on what they imply about statistical significance.
- If significant differences are found, perform 2 mean hypothesis tests to determine what groups are significantly different.

□ Conclusions: What have you learned about this issue and what follow-up study or analysis would you recommend. Make sure you discuss if the results make sense! Are there causal relationships present? Are your results muddled by potential confounding variables. Is there an easy fix that could be made to improve your model or results? If so, you should have already gone back and fixed that. Otherwise, discuss the issue and why it could not be addressed.

Misc: I’ll be randomly assigned presentation days once I get the finalized groups. You should plan to let me know by Wednesday, April 17, who you plan to work with or if you plan to work alone. All written materials will be due the last day of class by 5pm.
