Overview

This is the last unit projects that you will complete this semester. All unit projects are group projects. You will continue to work with the dataset that you chose for previous unit projects.

***IMPORTANT: ALL ANALYSES MUST BE SUBMITTED BY THE DUE DATE. WE WILL NOT BE ABLE TO GRANT EXTENSIONS FOR THIS PROJECT DUE TO THE VOLUME OF GRADING THAT IS NECESSARY AT THE END OF THE SEMESTER**.**

Unit Project 3 Instructions

Unit Project 3 has the same general setup: you will write a final briefing memo to your supervisor, Dr. Random. Dr. Random has asked that you to conduct analyses using techniques such a chi-square, ANOVA, and to find predictive models using simple and multiple linear analysis. You will then conclude with a set of recommendations.

The format for your briefing memo is similar to Unit1 and 2, but also includes a reflection section. Your memo, including all written analysis, R-code, and output from your code will be submitted as **an R Markdown file with HTML output.** A template will be provided, but you are welcome to make your own.

EACH PERSON IN YOUR GROUP IS RESPONSIBLE FOR RUNNING AND WRITING UP ONE OF THE DATA ANALYSES. The name of the responsible person should be listed with the analysis.

Table A lists the detailed requirements for the analyses that will be included in your briefing memo. EACH TEAM MEMBER IS INDIVIDUALLY RESPONSIBLE FOR ONE ANALYSIS. You will decide among yourselves how to divide the work. Each of you will use R for calculations and graphs. ***The individual responsible for the code and analysis must be clearly indicated.***

Format of Briefing Memo

1. Header: Create a header for your memo entitled "Briefing Memorandum", followed by "To:", "From:", and "Subject:" lines

2. Background/Recap: Provide a brief (one paragraph, please that 1) recaps the key features of your dataset, 2) states the objective of your current task as a follow-up of your previous data exploration.

3. Analysis: ***Each team member is individually responsible for the written analysis in the team report which corresponds to his/her designated statistical analysis from Part A.*** The written analysis should include any summary numerical and graphical output for the analysis, neatly displayed and labelled. Be sure to respond to the questions given with each analysis in Part A. **Use full sentences.** Remember, you are explaining your results to Dr. Random. *As you discuss each analysis, be sure to interpret what you are finding in the context of your particular data situation.*

4. Recommendations: In this section, you will briefly summarize the conclusions of your estimates and tests. What statistically significant findings did you have? What limitations are there in your findings? If you coordinated analyses for A3, A4, and/or A5, discuss what the findings tell you about a best fit model. What limitations are there in your findings? Discuss if/how this is relevant in the context of your data – does it give us new insights ? Help us to better understand some phenomenon? Or not?

5. Reflection: Write a group reflection responding to the following questions:

a. What impact did the Unit Projects have on your understanding of statistics?
b. What new knowledge did you gain about your statistical analyses? Briefly explain.
c. What did you learn by creating the unit projects?
d. What challenges did you face completing the unit projects?

Examples of reflections and thoughts on writing your reflections:
http://www.slcc.edu/eportfolio/reflection.aspx

Individual Analyses (Part 3 of Briefing Memo)

Table A presents the list of the analyses that will be included in your briefing memo. EACH TEAM MEMBER IS INDIVIDUALLY RESPONSIBLE FOR ONE ANALYSIS. You will decide among yourselves who will do what. Each person on the team will develop the listed R analysis and discussion/interpretation text for their assigned analysis. The R code and discussion/interpretation text will be incorporated into the SINGLE R Markdown file for the entire team , properly annotated and commented where necessary. ***The individual responsible for the code and analysis must be clearly indicated on the template.***

Required Individual Analyses

	Analysis Using R	Discussion/Interpretation
A1. Chi-square Test of Independence	• Select and identify a pair of potential explanatory and response variables that are both categorical. Each variable must have at ***least two but no more than four categories***. You may combine categories to reduce them to four. **Be sure to eliminate all missing or NA values before doing your analysis. DO NOT REPLACE THEM WITH ZEROS.** • Show any transformations done to your variable(s) that you have done to reduce it to two to four categories. • Create a contingency table for your variables. • Create an appropriate visualization (graph) for your variables. • Determine whether your data meets the conditions for performing a chi-square test of independence. • Run the chi-square test and show the output table.	• Clearly identify explanatory and response variables. • State the conditions and tell how each condition is met. Reference any calculations or visualizations that you use to justify this. • In words, state a null and alternative hypothesis *in the context of your data*. • In your written analysis, properly identify your test statistic and p-value, and state whether or not you reject the null hypothesis. • Then give a conclusion *in the context of the problem*. • Discuss any limitations to your analysis due to conditions not met, sampling bias, outliers in the data, or other issues that you notice. Reference your output from the R analysis, as needed.

A2. **Analysis of Variance**	• Select and identify a potential categorical explanatory variable and a quantiative response variable. The categorical variable much have at least two but no more than four categories. You may combine categories to reduce them to four. **Be sure to eliminate all missing or NA values before doing your analysis. DO NOT REPLACE THEM WITH ZEROS.** • Show any transformations done to your variable(s) that you have done to reduce it to two to four categories. • Create an appropriate summary table for your variables. • Create an appropriate visualization (graph) for your variables. • Determine whether your data meets the conditions for performing an analysis of variance test. Show any calculations or visualizations that you use to justify this. • Run the analysis of variance test. • Calculate the power of this ANOVA test.	• Clearly identify explanatory and response variables. • State the conditions and tell how each condition is met. Refer to any calculations or visualizations that you use to justify this. • State the null and alternative hypothesis of this test in words *in the context of your data.* • In your written analysis, properly state your test statistic and p-value, and state whether or not you reject the null hypothesis. • Then give a conclusion *in the context of your data*. • Comment on whether/how the power is relevant, given the conclusion of your test. • Discuss any limitations to your analysis due to conditions not met, sampling bias, outliers in the data, or other issues that you notice. Reference your output from the R analysis, as needed.
A3. Simple Linear Regression	• Select and identify potential quantitative explanatory and response variables. **Be sure to eliminate all missing or NA values**	• Clearly identify explanatory and response variables. • State the conditions and tell how each condition is met. Refer to any

***Special note**: You may wish to coordinate with the analysts for A4 and A5. You each could use the same response variable and test combinations of predictors to see which works best. That would be reported in your final summary and recommendations.*	**before doing your analysis. DO NOT REPLACE THEM WITH ZEROS.** • Show any transformations done to your variable(s), if appropriate. • Create a scatterplot and find the correlation for your variables. • Determine whether your data meets the conditions for a linear model. • State the conditions and perform all appropriate diagnostics. Show any calculations or visualizations that you use to justify this. • Run the linear regression. Interpret the slope and intercept, if appropriate. • Display a plot of the linear model overlaying the scatterplot	calculations, visualizations or diagnostics that you use to justify this. • State the null and alternative hypothesis of this analysis in words *in the context of your data.* • Write the linear regression model and interpret the slope and intercept, if appropriate. • In your written analysis, properly state your test statistic(s) and p-value, and state whether or not you reject the null hypothesis. State and interpret the R^2 • Then give a conclusion *in the context of your data.* • Discuss any limitations to your analysis due to conditions not met, sampling bias, outliers in the data, or other issues that you notice. Reference your output from the R analysis, as needed.
A4. **Multiple Linear Regression 1 (Two quantitative explanatory variables)** ***Special note**: You may wish to*	• Select and identify a potential quantitative response variable and at least TWO explanatory quantitative variables. You may use the same response variable as A3 and/or A5, but at least one of your explanatory variables should be different.	• Clearly identify explanatory and response variables. • State the conditions and tell how each condition is met. Refer to any

coordinate with the analysts for A4 and A5. You each could use the same response variable and test combinations of predictors to see which works best. That would be reported in your final summary and recommendations.	Explanatory variables should all be quantitative. **Be sure to eliminate all missing or NA values before doing your analysis. DO NOT REPLACE THEM WITH ZEROS.** • Show any transformations done to your variable(s), if appropriate. • Create table(s) of correlation and covariance for your variables. • Determine whether your data meets the conditions for a linear model. State the conditions and perform all appropriate diagnostics. Show any calculations or visualizations that you use to justify this. • Run the regression.	calculations, visualizations or diagnostics that you use to justify this. • State the null and alternative hypothesis of this analysis in words *in the context of your data.* • Write the regression model. Interpret each estimated coefficient, if appropriate. • In your written analysis, properly identify the test statistics and p-value for each coefficient , and state whether or not you reject the null hypothesis. • State and interpret the adjusted R^2 • Then give a summary conclusion *in the context of your data.* • Discuss any limitations to your analysis due to conditions not met, sampling bias, outliers in the data, or other issues that you notice. Reference your output from the R analysis, as needed.
A5. **Multiple regression 2 (one quant/one cat explanatory variables)** ***Special note**: You may wish to coordinate with the*	• Select and identify a potential quantitative response variable and at least TWO explanatory quantitative variables. You may use the same response variable as A3 and/or A5. One of your explanatory	• Clearly identify explanatory and response variables. • State the conditions and tell how each condition is met. Refer to any

analysts for A4 and A5. You each could use the same response variable and test combinations of predictors to see which works best. That would be reported in your final summary and recommendations.	variables must be quantitative and the other categorical with two categories. **Be sure to eliminate all missing or NA values before doing your analysis. DO NOT REPLACE THEM WITH ZEROS.** • Show any transformations done to your variable(s), if appropriate. • Create table(s) of correlation and covariance for your variables. • Determine whether your data meets the conditions for a linear model. State the conditions and perform all appropriate diagnostics. Show any calculations or visualizations that you use to justify this. • Run the regression.	calculations, visualizations or diagnostics that you use to justify this. • State the null and alternative hypothesis of this analysis in words *in the context of your data.* • Write the regression model. Interpret each estimated coefficient, if appropriate. • In your written analysis, properly identify the test statistics and p-value for each coefficient , and state whether or not you reject the null hypothesis. • State and interpret the adjusted R^2 • Then give a summary conclusion *in the context of your data.* • Discuss any limitations to your analysis due to conditions not met, sampling bias, outliers in the data, or other issues that you notice. Reference your output from the R analysis, as needed.