Thomas Jefferson HS for Science and Technology
Research Statistics 3
Final Project

Project due dates (5-point deduction for each day late):
- Part 1: Research Question, Literature Review, and Data Set due by midnight Friday, May 23
- Part 2: Final Paper, and RStudio Markdown file, due by midnight, Friday, May 30
- Part 3: Class Presentations: Due on Final Exam Day

Final Presentation Due Dates:

Mon., June 2 - 7th period (SENIORS ONLY)
Tues., June 3 – 5th and 6th (SENIORS ONLY)
Wed., June 4 – 1st and 2nd period
Thurs. June 5 – 3rd and 4th period
Fri., June 6 – 5th and 6th period
Mon., June 9 – 7th period

The purpose for this project is for you to have the opportunity to use what you have learned this semester and apply it in a real research context. Your final product will be a manuscript that is high quality (could be published). You will also do a presentation of your findings that will occur on the final exam day. As you transition to the university, or begin a summer internship, you may have the opportunity to participate in research with your professor(s) or mentors. You may even have the opportunity to publish your findings. Take a look at recent university undergraduate research that has been published at the undergraduate level https://our.unc.edu/share/publish/

Step 1: Begin by selecting a data base from a topic of interest. This is the data that you will use for the project so make sure you select variables that will help you demonstrate your knowledge of inference for linear regression. You must have **at least 6 quantitative variables** to demonstrate your ability to analyze multiple linear regression models. Aggregate data can be found in several open access sites and accessed without logging in for limited exploration and reuse, such as:

> Collection of data sites on the TJ Library site
> https://sites.google.com/fcpsschools.net/tjstatistics/datasets

You may NOT use a project or data set that you have used for another class!!!

You may NOT use Kaggle, this is a Data Science Competition platform!

Step 2: Find out what previous research has been conducted on your topic of interest by reviewing the literature on your topic. You should search for peer reviewed journal articles. These could be previous studies looking at the same condition/problem. Focus on gaps in the literature, inconsistencies, and variations from what you will be studying in addition to trends.

You will write a review of **three** peer reviewed articles. Be sure to use APA format in your writing and include a reference list. Here's the direct link to TJs eJournals – a great way to look for specific journals or browse by category, https://sites.google.com/fcpsschools.net/tjlibraryresources/home

Step 3: Write an original research question that could be answered using the data set you select then analyze the data using the content you learned about multiple linear regression in RS3.

Step 4: Analyze your selected data based on your research question then document your results in the methods section of your research paper. This should include the item in the grading rubric below.

Step 5: Explain your results. This should include charts and tables to organize your findings in addition to a detailed explanation of your study. What were the study limitations, ethical consideration, benefits to participants (if any)?

Step 6: Write a <u>Research Report</u> that includes the following:

Part 1 – Literature Review
- Describe the Design of the study.
- Describe participants or sample (age, gender, selection criteria, etc.) This should come from the documentation related to your data set.
- Measures - Description of data collected to answer the research questions such as (temperature readings, interviews, surveys, etc.)
- Analysis - A description of the methods/approaches taken in analysis of your data.
- Results/findings from each article.

Part 2 – Analyzing Your Data Set

Data Analysis. Analyze your data set, be sure to consider sample size, justify sampling procedure. Statistical methodology should be explained. Be sure to explain how variables were selected for your model, any transformations you used, and any outliers/influential points you identified. You should create and document your model, then complete a <u>significance test</u> and <u>confidence interval</u>. You should use transformations when the conditions are not met. Work should be shown for each step.

A discussion of the results of your data analysis procedures aligned to your research question. This should include models generated, diagrams, graphs, tables etc. You must include graphs to show conditions for inference have been met. Be sure to show all work for calculating the test statistic and confidence interval. You will need to submit the RMarkdown file showing your ability to select and measure the effectiveness of your model.

Part 3 - Reflection
Discussion of how the skills and knowledge gained during this project could be used to assist you in being a better researcher.

References (APA style)

Step 7: Prepare a short presentation (using power point or other technology). Formal Presentations of Research Projects will be conference-like presentations of approximately **5 minutes**, followed by a brief question and answer session among the class members.

Grading Rubric for Research Paper

Name: _____

Title of Project: _____

	Needs Improvement	Satisfactory	Exemplary	Points Earned
Part 1: Research Question and Data Set				
Review of Literature	**(3 points)** Lack of clarity and/or flow, brief or incomplete discussion of related research literature, or inaccurate representations of research literature	**(6 points)** Review of literature may either lack clarity or flow or include brief or inaccurate or representations of research literature.	**(10 points)** Clearly written, logical review of significant research literature related to at least 3 articles related to this topic.	/10pts.
Research Questions	**(1 points)** Research questions not stated, not clear, or not related to purpose and literature review	**(2 points)** Research questions not stated in a clear manner or are not logically connected to the purpose and literature reviewed	**(3 points)** Research questions stated succinctly and are logical follow-ups to stated purpose and literature reviewed	/3 pts.
Data Representation: Raw Data	**(0 points)** Raw data used in the study are not provided.	**(4 points)** Raw data are given, but presentation is incomplete (Fewer than 6 quantitative variables) or disorganized.	**(7 points)** All raw data are included and well organized. Data set is complete and contains a minimum of 6 quantitative variables.	/7 pts.
				Total points **/20**
Part 2: Final Paper, and RStudio Markdown file				
	(1 point)	**(3 points)**	**(6 points)**	
Participants/ Sample	Much information is not provided relative to methods of identifying	Some information on procedure for selecting participants is	Complete description of procedures for selecting participants and detailed	/6 pts.

	participants, or missing significant demographic information on participants	unclear, or some demographic information on participants is not provided.	demographic information on participants	
Data Selection and Cleaning	Acceptable data. Limited cleaning or preprocessing.	Good quality dataset. Proper basic cleaning. Some room for improvement in outlier handling, etc.	High-quality dataset. Excellent handling of missing values, outliers, and data types.	/6 pts.
Exploratory Data Analysis (EDA)	Basic EDA with minimal plots/stats. Some relevant aspects missing.	Good EDA with appropriate plots and stats. Covers main features.	Thorough EDA with insightful visualizations and summary stats. Variables well-understood.	/6 pts.
Model Building	Few predictors justified. Minimal or incomplete assumption testing.	Appropriate predictors. Most assumptions tested. Some explanation.	Model includes well-justified predictors. All assumptions (linearity, multicollinearity, etc.) tested and addressed.	/6 pts.
Model Interpretation	Misinterprets coefficients or ignores interpretation entirely.	Coefficients interpreted correctly with moderate context.	Coefficients well-interpreted in context. Insightful understanding of statistical significance and practical impact.	/6 pts.
Model Performance	Incorrect or missing performance evaluation.	Uses correct metrics. Some attempt at model validation.	Uses appropriate metrics (R^2, Adjusted R^2, RMSE, etc.). Evaluates model using cross-validation or test set.	/6 pts.
Results/Data Representation	Results are not resented for all the data collected, and/or discussions of results are difficult to follow or understand.	Analysis and discussion of results is incomplete, lacks some details, and/or is difficult to follow or understand.	All results of the analysis of data are described clearly and in detail and presented logically. Graphs are included where appropriate.	/6 pts.
Connections to Literature	In discussing answers to the research questions, connections to the literature reviewed in Part 1 are not	In discussing answers to the research questions, connections to the literature reviewed in Part 1 are	In discussing answers to the research questions, connections to the literature reviewed in Part 1 are made. Areas in which	/6 pts.

	presented or inaccurate. The relationships between results of this study and previous research findings are not included.	incomplete. The relationships between the results of this study and previous research studies is presented without detail or clarity.	the results of this study match and do not match previous research findings are identified and discussed.	
Implications for further research	The discussion does not provide suggestions for further research on the topic, and/or recommendations are not provided for the applications of the findings.	There is a brief or incomplete discussion of suggestions for further research on the topic, and/or makes brief or incomplete recommendations for the applications of the findings.	The discussion includes suggestions for further research that is needed on the topic. Recommendations for how the findings can be applied are identified and discussed in detail.	/6 pts.
Reflection	Missing or non-reflective discussion.	A discussion of how skills and knowledge gained during this project could be used to assist in becoming a better researcher.	A reflective discussion of how specific skills and knowledge gained during this project could be used to assist in becoming a better researcher, including examples.	/6 pts.
References	Reference list is not provided, brief, or not in APA format.	List of at least 3 peer reviewed journal articles related to the research provided in APA format. Some references from the paper are not included in the reference list.	Analysis of 6-10 peer reviewed journal articles related to the research and discussed in the literature review/discussion provided in APA format.	/6 pts.
Code/Appendix	**(0 Points)** Missing RMarkdown file or unreadable code.	**(7 Points)** RMarkdown file contains some code included but lacks clarity.	**(14 Points)** RMarkdown file contains well-documented code, reproducible analysis, clear workflow.	14 pts.
				Total Points **/80**

Part 3: Class Presentations				
	Minimum	Elaborated	Exemplary	Points Earned
Presentation of Research	**(2 points)** Presentation lacked clarity and sufficient detail.	**(5 points)** Research questions, literature reviewed, Participants and methodology, Findings, and implications	**(10 points)** The Research questions, literature reviewed, methodology, Findings, and implications. Presentation was supported by a power point presentation with clear graphics illustrating the study conducted.	
				Total Points **/10**

Comments: