IS 204 Research Design for Information Sciences

Final Project Guidelines

The objective of the final group project is to analyze a large dataset using a variety of statistical and machine methods covered in this course.

The goal of the final project is to not only conduct various methods, but also to draw applicable results and conclusions from your analyses.

The final group will consist of **3 students**. Each student in the final group will have to conduct at least 1 technique. Thus, a **3-member group** will conduct **at least 3 techniques** of their choosing.

Each group will write a single analysis, but each group member will be contributing to the project related to the technique of their choosing. Group members not completing their fair contribution to the final group project may receive a lower grade than the other members in their group.

The single analysis of the group will consist of:

1. **Journal Article (Technical Summary)**
    a. **7 typewritten double-spaced pages (not including tables, figures, and reference page)**
    b. Cited in **APA Style**
    c. Write a formal journal article in a clear and organized fashion for **a statistically literature audience** including the following sections:

a. **Abstract** of 250 words
b. **Introduction**
c. **Literature Review** (Background on the Application / **3-Peer Reviewed Journal Articles**)
d. **Methods** (see next page for further details)
e. **Discussion and Results** (also including limitations and possible future work)
    Results should be explained in terms of the application
f. **Conclusion**
    Conclusion should be in terms of the application.
g. **APA Reference** page
h. **Tables and Figures** (titled and numbered in a presentable format)
i. **Appendix** (include Syntax/Code)

Your **Methods section** should include the steps taken to **complete your statistical analyses**, which should the following:

1. Mention what **research questions** your Statistical or Machine Learning Techniques are trying to answer
2. Mention **types/groups of variables** being used to answer research questions (DO NOT LIST EACH INDIVIDUAL VARIABLE)
3. Mention any **transformations or formulas** used to create new variables
4. **Assumptions** of Statistical or Machine Learning Techniques (i.e., linearity, multicollinearity, normality, etc.)
5. **Options** of Statistical or Machine Learning Techniques, if the method has multiple ways of being answered (i.e., cross-validation, training/testing, 1-sample/2-sample, variable selection, etc.)
6. Where appropriate, provide **performance metrics** for regressions and machine learning techniques (i.e., R2, Adjusted R2, standard error, Mallow's Cp, Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, AIC, BIC, accuracy, sensitivity, specificity, Positive Predictive Value, Negative Predictive Value, Precision, Recall, F1-measure, AUC-ROC, etc.)
7. Define **Significance** (i.e., 1-side, 2-sided p-values)
8. Define **Alpha Value** (i.e., 0.05)
9. Specify **Analysis Program tools, Version, and Location**

Your **Results and Discussion section** should include the following:

1. Consider the practical (application) of the results, in addition to the statistical significance.

Tips for Final Paper

1. Write a clear and concise paper for the appropriate audience
2. Quote and cite any external sources in APA style format
3. Remember that statistical and methods can take multiple iterations to arrive at a final model
4. Understand that analyses may provide none, one, or more than one solution, specifically in regards to regression and machine learning models
    a. If models do not work, explain models tried, why they did not work, and how you attempted to correct issues.
    b. If more than one model, compare and contrast the various models and their performance metrics in predicting the outcome.
5. Tables and Figures make take multiple versions, in order to be clear and visually presenting.

IS 204 Final Project Grading Rubric	**Excellent**	**Good/Fair**	**Poor**
Layout & Organization	Report is: • Well-written, clear, organized with correct statistical terminology and grammar • Use headings and tables	Report is: • Uses correct statistical terminology • While mostly clear, not well-organized sections	Report is: • Layout cluttered • Not organized and/or clear • Major revision and editing required due to: • Multiple grammatical error
Exploratory Analysis	Checked assumptions, corrected any assumption issues	Did not check or correct all assumptions	Did not check necessary assumptions
Exploratory Graphs	Produced graphs and interpreted them correctly	Inaccurate or incomplete graphs	Missing or major flaws in graphs
Method Execution	Chose and properly conducted an appropriate method	Used in appropriate method or have minor issues in execution	Used inappropriate method and there are major problems in execution
Method Analysis	Appropriate post checks and performance analysis. Correctly analyzed and interpreted results.	Some issues or flaws with post statistical checks, performance analysis, or interpreting results.	Missing post checks and/or performance analysis. Major issues with analysis and interpretation of results.