MAT 443: Final Project
Due Monday, June 5 by 8:15 PM

Find a suitable dataset to work with for the final project **AS SOON AS POSSIBLE!**. See the end of this PDF for possible places to find good data sets.

You can work **ALONE** or work in groups of **TWO**. When working on the project if you decide to work with a partner, I must know what each of the partners contributed to the paper and have a copy of each person's objectives and R/SAS code.

Requirements of the Dataset: The data must have a minimum of 15 variables with at least 12 quantitative variables and a minimum of 200 cases. If you plan on running discriminant analysis, one of your variables must be categorical. If you find an interesting dataset with a good design that does not meet these minimal requirements, e-mail me and I may let you use it.)

As soon as you find your dataset, e-mail it to me so I can see if it will be sufficient for the project. I will then be happy to help you to outline you procedures and make suer your research objectives are met. Do not hesitate to ask me for help.

Background Research for Project Topic: Every paper should begin with appropriate background information about the topic. References used in obtaining this information should be cited in the paper. You should have at least five references dealing with previous studies about your dataset or references about the topic of your paper and possible studies completed even though different than yours.

Statement of Research Objectives: Once you find a suitable dataset to work with, you must formulate a set of research hypotheses that your project will address. The research hypotheses that you create will determine what tests and methodology will be needed from the course to address these needs. The statements of your research hypotheses can be included as pat of your purpose or after your descriptive statistics in the body of your report. As soon as you formulate these objectives, e-mail them to me so I can make sure they are suitable.

Statistical Procedures to Use in the Report: The procedures to use besides the numerical descriptive statistics should include at least a couple of the following techniques:

- Canonical correlation
- Discriminant analysis
- Logistic regression for classification
- Principal component analysis
- Factor analysis
- Multivariate analysis of variance
- Cluster analysis (Week 10)

Guidelines for Submission of Written Report

Be sure to include the title and source of the data.

- **Abstract** – An abstract is usually necessary because your report may have multiple readers. Some of them will need to know the details of your report, including the supporting data on which you are basing your conclusions and recommendations. Other readers will not need as many details but will want only the preliminary study information and main conclusions. The abstract or executive summary should cover the general subject of the research, the scope of the research, identification of the type of methodology used, and conclusions. You can describe the project's objective, what you have done, and general findings. It should be limited to no more than 200 words and is often confined to less than 100 words.

- **Background and Purpose of Study** – This is an introduction to the report. Review of literature should be included here, too. The brief literature review is mandatory for the final project. You most likely have many goals in mind and you can state these or your research hypotheses here. It is these research hypotheses you will be addressing in the paper.

- **Sample Collection Techniques, Methodology, & Design (if applicable)** – Describe the data collection process, if you know it. If not, tell us how and when you acquired the data. Describe the size of the

sample and information about the variables of interest. Include information about margin of error for percentages quoted in report based on sample sizes used. Discuss the design, and if applicable, the power of the test.

- **List of Variables, Responses, and Codebook of Variables in the Analysis (if applicable)** – If this is very lengthy, this codebook may appear in the Appendix.

- **Routines Used in SAS or R** – This can be in the introduction or scattered throughout the report. This is up to you.

- **The Body of the Report** – The body is the bulk of your report. It is the major part of your paper. In a logical order, present the results.

  - Begin with the tabular and numeric descriptive statistical techniques. These descriptive statistics findings should be supported by tables or graphs. This is followed by a summary of the descriptive findings. Be sure to include margin of error for all percentages quoted in the reported.

  - The next part summarizes the inferential procedures used. These should correspond to the research hypotheses previously stated. Comment on the rationale for using certain procedures. For all statistical tests contained in the report, be sure to state the hypotheses, assumptions, test statistics, P-values, and conclusions found. All results should be in narrative form and supported by tables or graphs when needed to enhance the explanation. Do not be afraid to be creative!

- **Summary/Conclusions & Recommendations** – This section contains a brief summary of your findings. The conclusions are generally stated first and are the outcomes and decisions based on your research results. Recommendations, on the other hand, are suggestions for how to proceed based on your conclusions. List any limitations, if they exist, and give suggestions for future research.

- **Limitations (if any) and Other Variables You Wish You Wish You Would Have Included in the Study** – Be specific about your findings. How did they compare with your research hypotheses? What other variables would you have tried to include in the study, and were there variables you would not use in further studies that were contained in this study?

- **References Section** – Include any sources where you may have done a search on background information on the topic. Include the source of the data reference site (if not confidential) and other texts or software you may have used in the report.

- **Appendix** – If you used survey data, a copy of the survey should be included in the appendix. Some people also put the codebook in the appendix.

Where to Find Datasets?

The following is a list of sites where you might be able to find datasets. If you have a dataset from your place of employment that fits the criteria for this project, feel free to use it.

- `https://datadryad.org`

- `https://www.kaggle.com/datasets`

- `http://www.statsci.org/datasets.html`

- OzDASL – `http://www.statsci.org/data/index.html`

- DASL – `http://lib.stat.cmu.edu/DASL/`

- *JSE* Data Archive – `https://ww2.amstat.org/publications/jse/jse_data_archive.htm`

- Datasets from UCLA's Department of Statistics – `http://www.stat.ucla.edu/cases/`

- More datasets from UCLA's Department of Statistics – `http://www.stat.ucla.edu/projects/datasets/`

- U.S. Government Datasets – `https://www.data.gov/`