

STAT 478 Project Description – Spring 2021

Written project due date: Friday, April 30th, by 11:59 p.m. You may submit your paper on blackboard, or you may email me a copy of your project as a Word document or pdf file.

In this project, you are expected to perform a complete analysis of a dataset that interest you. You will submit a written exposition that thoroughly describes your approaches. This is an opportunity for you to amalgamate all of your time series knowledge and apply it to a real problem. Remember, an important part of a statistical analysis is clearly communicating it to others in writing and the written project is how you disseminate your work. **You may choose to work on this project individually or in group of two.**

Guidelines for Choosing a Dataset:

- Use other sources (e.g., data sets online, other research projects, etc.) to find an interesting dataset. The more interesting the better!
- The responses should be continuous in nature. Also, make sure you completely understand the sampling frequency; e.g., are the data collected every day? every week? every month? every year? every minute?
- Choose a dataset in an area you are interested in! You should be able to demonstrate a working knowledge of the subject.
- Don't choose a dataset that is very small. Ideally, we want $n > 75$ or so, but this is just a guideline.
- Choose *regularly spaced* data, meaning that there is essentially the same interval of time between each measurement.

Guidelines for Modeling and Forecasting:

It is important to keep in mind the four major steps of analyzing a time series data set.

1. Model specification: come up with a small set of candidate models for your data.
2. Parameter estimation: fit the models.
3. Model diagnostics: check the adequacy of the model fits and possibly revise the model.
4. Forecasting: choose a final model and forecast future observations.

Another recommendation: It might be a good idea to “withhold” some of the data from your series towards the end of it so that you can compare your forecasts to the actual values to see on how accurate/precise your forecasting is.

Grading Scale: (50 points total)

Writing (10 points): Organized, clearly written, comprehensible, and grammatically correct.

Analysis (25 points): Were the chosen models, graphs, and data analyses appropriate for the problem? Were the analyses carried out correctly? Were your statistical conclusions about the data set sensible and clearly justified by numerical or graphical evidence?

Context (15 points): Were the questions answered in terms of the variables of the dataset? Have you attempted to frame your conclusions and interpretations in a subject-matter context? Provide some background information about the dataset and why it is of interest.

Outline of the Written Project:

- **Title page and abstract:** You must prepare a title page with an appropriate title and abstract. The abstract should go on the title page. An abstract is a very high-level written summary of the entire project. Main points and findings only.
- **Introduction:** This part introduces the reader to the dataset and to the area to which it pertains. Basically, introduce the reader to the problem and why it is meritorious of investigation. This should be written at a very basic level. Remember your reader may not know anything about the area in which you are writing.
- **Model specification:** In this section, you want to describe, in clear detail, the data analysis used to specify your candidate models. Pretend as if you are taking the reader by the hand and leading him or her through your thought process which leads to your model selections.
- **Fitting and Diagnostics:** This part of the project should describe the model fitting and diagnostics techniques you used, with the goal of identifying a “final” model for forecasting. Identify also what possible deficiencies your final model has. Remember, no model is perfect.
- **Forecasting:** This section should describe the techniques you used to forecast future. Why is forecasting important? What impacts could your forecasting have?
- **Discussion:** Here you want to offer a summary of what you did in the project and draw your main conclusions. Also, it is a good idea to discuss here other issues related to the data analysis. For example, does your analysis have any shortcomings or lack of generalizability? What were the main problems you encountered? It is OK if your final model is not perfect. Real life data analysis is often more difficult than textbook problems.
- **Bibliography:** Cite all references (including the original source of your data) carefully.
- **Appendices (optional):** Use appendices to catalogue extra graphics/plots/output. Basically, I use an appendix to house information that I want the reader to have access to, but feel that it would interrupt the flow of the main body of the paper.

General Advice:

- Please make the formatting such that the report is easy to read. Break your report into sections. Each section should have a title. Use subsections if necessary.
- Integrate R graphics and output into the written text as you see fit. For example, if you want to show the time series itself, embed it into the written work. Look at the style of the way graphics are embedded in your textbook as a guide
- Edit and proof read your written project. You may get other people to read your project and offer comments/feedback.
- I don't have a specific target number of pages for the whole report. You should do enough to provide a full analysis of this dataset, with attention paid to each of the sections listed above. You should end up with around 8-10 pages.
- Have fun! This is an opportunity for you to amalgamate all of your time series knowledge and apply it to a real problem. Show me what you have learned. Remember, the written project is how you disseminate your work. I think the most important part of a statistical analysis is clearly communicating it to others in writing.

Project Rubric

Organization & clarity: 5pts
Grammar, spelling, and writing: 5pts
Model justification: 10pts
Analysis correctness: 10pts
Use of graphics and outputs: 5pts
Model interpretation: 5pts
Background context: 5pts
Completeness and Conclusion: 5pts