
Statistics 5550: Project
Due Monday, April 26th, 2021 by 5pm

The course project will consist of an analysis of a time series data set of your choice, subject to the guidelines described below. The goals of the project are to demonstrate the ability to

- i. Use exploratory data analysis methods to discover important features of the data;
- ii. Appropriately handle trend and seasonality in non-stationary time series;
- iii. Identify, estimate and assess the fit of appropriate ARMA, ARIMA and SARIMA models to model time dependence;
- iv. Compare competing models or approaches to modeling for a data set;
- v. Construct optimal forecasts for future data and quantify uncertainty about the forecasts;
- vi. Communicate your results in a report that is written clearly and easy to read, and that accurately describes your analysis.

The written report should be 6-7 pages in length, including figures. You may include extra figures in an appendix if you wish. You should submit your report as a **pdf** file to Carmen (an assignment has been set up for this). You should also submit a separate text file with your R code as well as a text file with your data.

First Deadline: Monday, April 12, 2021

The following information is required to be submitted by the first project deadline of Monday, April 12, 2021:

Data Info: You will need to identify and obtain a data set to use for this project. Your data should conform to the requirements below. No later than the end of the day on **Monday, April 12, 2021**, you should upload your chosen data to Carmen (there is an assignment set up for this). I will review and approve the data set (or suggest alternatives if the data do not seem appropriate). Please also upload a brief description of the data so that I know what I am looking at. The earlier you do this the better!

The time series data you select for the project should

1. exhibit both non-constant trend and seasonal behavior,
2. have enough data to cover several seasonal cycles (i.e., not just one year of monthly data),
3. should have a stationary component that exhibits dependence through time,
4. *not* come from an R package.

Main Deadline: Monday, April 26, 2021

Analysis Requirements

1. Provide an exploratory analysis of the data as discussed at the beginning of the semester. Describe any non-stationary features of the data, discuss methods for eliminating any non-stationary behavior, discuss potential transformations of the data (if appropriate) and provide and discuss graphical summaries.
2. Perform an analysis of the data where you explicitly estimate the trend and seasonal components (e.g. via regression techniques or smoothing), detrend and deseasonalize the data, and then use ARMA modeling techniques to analyze the detrended, deseasonalized data. You should identify and fit a small number of plausible ARMA models for the detrended, deseasonalized data. Be sure to check the model fits and compare the models using the methods discussed in class. Choose and report a final fitted model for your data and explain why you made this choice. Be sure to write down the final model. For example, if it is a seasonal means model with a linear trend plus an AR(1) stationary process, you might write that the estimated model is

$$y_t = 23.2 t + \hat{s}_t + x_t, \quad x_t = 0.653x_{t-1} + w_t, \quad w_t \sim \text{iid } N(0, 2.37),$$

and provide the estimated seasonal effects, along with details about how they were estimated. Forecast the original time series (the series containing trend and seasonality) out several time periods and provide prediction intervals.

3. Perform a second analysis of the data where you handle the trend and seasonal components by appropriately differencing the data and fitting an appropriate SARIMA model. Fit several competing SARIMA models and compare their fits using residual diagnostics and other methods discussed in class. Choose and report a final fitted model for your data and explain why you made this choice. Be sure to write down the final model. For example, if it is an ARIMA(1, 0, 0) \times (0, 1, 1)₁₂ model, you might write that the fitted model is

$$(1 - 0.8B)(1 - B^{12})x_t = (1 + 0.3B^{12})w_t, \quad w_t \sim \text{iid } N(0, 1.25).$$

Use the fitted model to forecast over the same time period as you did for your other model above, and provide prediction intervals.

4. Compare the models you fit in parts (2) and (3) above. How are they similar? How are they different? Which model do you prefer, and why?

Write-Up Requirements

1. The project should be written up in the form of a report in paragraph style and typeset.
2. Provide a clear description of your analysis so that anyone who is familiar with time series modeling could understand what you have done. For example, don't say "I detrended the data"; say specifically what you have done. Provide equations where appropriate.

3. The write-up should have five sections:
 - i. Introduction—provides a clear description of the data and where it came from. The goals of the analysis are described. An exploratory data analysis is performed.
 - ii. Models for Trend and Seasonality—the results of your approach to item (2) under “Analysis Requirements” above.
 - iii. SARIMA Modeling—the results of your approach to item (3) under “Analysis Requirements” above.
 - iv. Model Comparison—compare your two models as described above.
 - v. Conclusions—summarize the results.
4. Do not include “screen shots” of computer output. All figures and tables must be appropriately formatted and labeled.
5. Do not cut and paste R output—extract the appropriate information from the output and include it in a concise and well-formatted manner. (I have copy-and-pasted R output in some of the handouts I have prepared for you this semester. This was mainly so that you could match up your work with the output in the handout. For your report, don’t use this copy-and-paste approach.)

R Code

You will need to turn in a text file with your R code to Carmen. Make sure your R code works and that it is mostly clear what you are doing. Delete extraneous code or code from analyses that didn’t make it into the final write-up.