w271: Homework 8 (Due: Week 9)

Professor Jeffrey Yau

Due: By 4 p.m. Pacific Time on the Day of the Live Session in Week 9

Instructions (Please Read it Carefully!):

- **Page limit of the pdf report: None, but please be reasonable**
- Page setup:
  - Use the following font size, margin, and linespace:
    - fontsize=11pt
    - margin=1in
    - line_spacing=single
- Submission:
  - Homework needs to be completed individually; this is not a group project, but you are welcome to discuss with classmates.
  - Each student submits his/her homework to the course github repo by the deadline; submission and revision made after the deadline will not be graded
  - Submit 2 files:
    1. A pdf file that details your answers. Include all the R codes used to produce the answers. *Please do not suppress the codes in your pdf file.*
    2. R markdown file used to produce the pdf file
  - Use the following file-naming convensation; fail to do so will receive 10% reduction in the grade:
    - StudentFirstNameLastName_HWNumber.fileExtension
    - For example, if the student's name is Kyle Cartman for homework 1, name your files as
      - KyleCartman_HW1.Rmd
      - KyleCartman_HW1.pdf
  - Although it sounds obvious, please write your name on page 1 of your pdf and Rmd files.
  - For statistical methods that we cover in this course, use only the R libraries and functions that are covered in this course. If you use libraries and functions for statistical modeling that we have not covered, you have to (1) provide an explanation of why such libraries and functions are used instead and (2) reference to the library documentation. **Lacking the explanation and reference to the documentation will result in a score of zero for the corresponding question.** For data wrangling and data visualization, you are free to use other libraries, such as dplyr, ggplot2, etc.
  - For mathematical formulae, type them in your R markdown file. **Do not write them on a piece of paper, snap a photo, and either insert the image file or sumbit the image file separately. Doing so will receive a** 0 **for that whole question.**
  - Students are expected to act with regards to UC Berkeley Academic Integrity.

In the last live session, I demonstrated a complete worked-out example and had you repeated the process using another series. In this week's homework, you will continue with the exercise using another data series `series3.csv`. Your task is to (1) build a time series model using the `arima()` function and the class of ARMA model, which includes AR, MA, and ARMA models and (2) conduct a **3-step** ahead foreast.

The ARMA time series model building process, which I outlined both in the async lecture and demonstrated in the last live session (for the AR and MA models), typically includes (1) a time series EDA, which requires an examination of the stationarity of the series and a determination of whether the class of ARMA model is a "reasonable" model as a starting point, (2) model estimation (perhaps a few models need to be attempted), (3) model selection based on some metrics, say AIC, (4) model diagnostic, and (5) model forecast (after a valid model is found). You need to explain why certain AR/MA/ARMA model is chosen as a starting point based on your time series EDA.