

w271: Homework 6 (Due: Week 7)

Professor Jeffrey Yau

Due: 4pm Pacific Time on the Day of the Live Session (i.e. Monday) of Week 7

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.
 - 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 convention; 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 submit 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 this homework, you are asked to conduct Time Series EDA, develop a Time Trend Model, conduct model diagnostic analysis, and use the model to make forecasts.

Load the file “**bls_unemployment.csv**” This file contains the monthly unemployment rate in the United States from January 2007 to January 2017. *Note that these data are NOT seasonally adjusted.*

1. Load the csv file into a data.frame, calling it *df*, and examine the structure of the “raw” series after you load it into a data.frame.
2. Convert it into a R time-series object, and examine the structure of the series after you convert the *df* into a time series object. What is the difference between the two structures?

Suppose we want to approximate the unemployment rate between 2010 and 2017 using a linear time trend model. For this exercise, feel free to modify this dataset but be sure that you explain what you did and why.

3. Conduct EDA on the series.
4. Create a linear time trend model by regressing the unemployment rate on time. Interpret the model results.
5. Examine the residuals of this model as you would if it were a classical linear model. In addition, generate ACF and PACF plots of the residuals.
6. Use this model to predict the unemployment rate in 2018 January (i.e. 12 months from the end of the sample). Do the result make sense? How about a prediction of the unemployment rate in 2020 December?