MIDTERM
(Take-home part)

Statistics 137
Spring Quarter, 2016
This take-home part covers 20% of the MIDTERM and it is due on **Wednesday, May 11**. You may form a group of up to 3 students registered in this course. You need to submit only one report. Please write down the names of the students in your group on the submitted work.

The file chicago.txt contains daily average receipts per theater for the movie Chicago (January 3, 2003 to April 18, 2003). Analyze the data set using time series methods. This series may have a trend as well as seasonality (day of the week). The goal of the analysis is to estimate the trend, seasonal components, model the stationary part and check how well the estimated model is able to do forecasting. Perform a complete analysis including transformations, estimation of trend, seasonal components and a model for the stationary part. Prepare a thorough report as you would do for a client. This report should include all the steps used in the analysis, their justifications (relevant plots, diagnostics etc.), computer codes and your conclusions. Also comment on the possible improvements (if needed) that can be made on your analysis. Please cut and paste the relevant portions from your computer commands and printouts.

Here are some steps you may follow and at each step summarize the findings.
(1) An introduction containing a brief description of the data, why it is a time series, and what can time series modeling achieve.
(2) Explain the nature of the variations in the data and what type of modeling scheme may be appropriate here.
(3) Plot the data as well as various Box-Cox transformations. Choose an appropriate transformation.
(4) For the transformed data, estimate the trend (polynomial model) as well as the seasonal components.
(5) Make appropriate plots for the various components (trend, seasonal, the rough etc.).
(6) Carry out diagnostics for the rough part (stationarity, independence, equal variance, normality, autocorrelations (ACF plot) etc.)
(7) For the stationary (rough) part, obtain an appropriate $AR(p)$ model using the AICC criterion. Do the necessary diagnostics (ACF plot) using the residuals of the selected AR model. [It is enough to consider $p = 0, ..., 5$.]
(8) Using your estimated model, predict the last 7 seven days using observations from the first 99 days. Plot the observed as well as the predicted values for the last seven days.
(9) Clearly state your final model and summarize your conclusion.