Northeastern University
Khoury College of
Computer Sciences

Practice Problems 2

For this set of practice problems you should not use any forecasting packages. Instead write the code yourself as demonstrated in the lessons. However, you may wish to compare your answers to those of your peers and to those from functions in packages.

1. (15 pts) The built-in dataset USArrests contains statistics about violent crime rates in the US States. Determine which states are outliers in terms of murders. Outliers, for the sake of this question, are defined as values that are more than 1.5 standard deviations from the mean.
2. (15 pts) For the same dataset as in (1), is there a correlation between urban population and murder, *i.e.*, as one goes up, does the other statistic as well? Comment on the strength of the correlation. Calculate the Pearson coefficient of correlation in R.
3. (3 x 10 pts) Based on the data on the growth of mobile phone use in Brazil (you'll need to copy the data and create a CSV that you can load into R or use the *gsheet2tbl()* function from the **gsheet** package), forecast phone use for the next time period using a 2-year weighted moving average (with weights of 5 for the most recent year, and 2 for other), exponential smoothing (alpha of 0.4), and linear regression trendline.
4. (20 pts) Calculate the squared error for each model, *i.e.*, use the model to calculate a forecast for each given time period and then the squared error. Finally, calculate the average (mean) squared error for each model. Which model has the smallest mean squared error (MSE)?
5. (20 pts) Calculate a weighted average forecast by averaging out the three forecasts calculated in (3) with the following weights: 4 for trend line, 2 for exponential smoothing, 1 for weighted moving average. Remember to divide by the sum of the weights in a weighted average.

Submission Details

- Practice Problems are for learning and practice and therefore are not graded and no submission is required. You are encourage to discuss and review them with your peers. Additionally, they are reviewed during weekly recitations. If you desire, you may ask for individual feedback from the instructional staff during office hours. Completing practice problems will prepare you for the graded practicums and their completion is critical to doing well on the practicums and the final project.

Useful Resources

- Correlation Coefficient in R

Learning

Blackboard

Lynda.com

Data Camp

Support

Contact Instructor

Virtual Office

Book Appointment

Search

Created by Martin Schedlbauer, PhD