University of San Diego
SHILEY-MARCOS SCHOOL OF ENGINEERING

ADS 506 Module 1 Exercises: Chapters 1-3

This assignment is due on Day 7 of the learning week. The assignment for this module consists of written work. Complete this entire assignment in R Markdown. You will need to include the question and number that you are answering within your submitted assignment. **Once completed, you will knit your deliverable to a Word/PDF file.**

Chapter 1: Approaching Forecasting (Page 23) #1-5

1. Is the goal of this study descriptive or predictive?
2. What is the forecast horizon to consider in this task? Are next-month forecasts sufficient?
3. What level of automation does this forecasting task require? Consider the four questions related to automation.
4. What does the meaning of t = 1,2,3 in the Air series? Which time period does t = 1 refer to?
5. What are the values for y_1, y_2, and y_3 in the Air series?

Chapter 2: Time Series Data (Page 43) #3

3. *Shipments of Household Appliance:* The file *ApplianceShipments.xls* contains the series of quarterly shipments (in millions of USD) of U.S. household appliances between 1985-1989.
   a) Create a well-formatted time plot of the data.
   b) Which of the four components (level, trend, seasonality, noise) seem to be present in this series?

Chapter 3: Performance Evaluation (Pages 67-68) #2 and 3

2. *Forecasting Shampoo Sales:* The file *ShampooSales.xls* contains data on the monthly sales of a certain shampoo over a three-year period. If the goal is forecasting sales in future months, which of the following steps should be taken? (choose one or more below).

   ☐ Partition the data into training and validation periods.

   ☐ Examine time plots of the series and of model forecasts only for the training period.

   ☐ Look at MAPE and RMSE values for the training period.

   ☐ Look at MAPE and RMSE values for the validation period.

   ☐ Compute naïve forecasts.

3. *Performance on Training and Validation Data:* Two different models were fit to the same time series. The first 100 time periods were used for the training period and the last 12 periods were treated as a validation period. Assume that both models make sense practically and fit the data reasonably well. Below are the RMSE values for each of the models:

	Training Period	Validation Period
Model A	543	690
Model B	669	675

   a) Which model appears more useful for retrospectively describing the different components of this time series? Why?
   b) Which model appears to be more useful for forecasting purposes? Why?