11.1: Introduction to the Lesson

The first two topics of this lesson explore how to use the framework associated with a linear regression model to fit data that are time-dependent. When data are time-dependent, the observations and errors are typically no longer independent. A common example would be stock prices at the end of each trading day. The data and error at day T are likely to be correlated with the data and error on previous days T-1, T-2, and so on. If we used our linear regression model in such a setting, our inference would be compromised: results from confidence intervals and hypothesis tests will no longer be reliable. To account for the correlation between data and error terms, we will use an autoregressive (AR) model for the error terms. After finding the appropriate AR model for the error terms, we can apply a simple transformation to our data, and then fit a linear regression model after the transformation.

In module 7, you were introduced to the general idea associated with model validation. The data are split into two parts; one part is used to fit the model, and the other part is used to test the predictive ability of the model. These parts are called the testing and validation sets respectively. In the last part of this lesson, we will compare a few different ways of splitting the data into these two parts, and you will learn about the pros and cons of these different ways to validate models.