Multiple Linear Regression versus Simple Linear Regression

Why would you perform multiple linear regression instead of a series of simple linear regressions? The biggest advantage is that multiple regression enables you to determine the relationship between a predictor and response while controlling for all other predictors included in the model. Sometimes a hidden relationship can be revealed or a strong relationship can disappear when additional predictors are accounted for by including them in the regression model. You can determine whether a relationship exists between the response variable and several predictor variables simultaneously. You can also test for interactions just like in ANOVA.

What about the disadvantages? The more predictors you have, the more complicated interpreting the model becomes. Consider an example with one response variable and seven potential predictor variables. For this number of variables, you have 127 possible models with at least one predictor variable! The increased complexity makes it more difficult to interpret the models, and to decide which model to use. Later we'll discuss techniques to choose a "best" model, a choice that might depend on the purposes of the analysis and subject-matter expertise.

Overall, the advantage of performing multiple linear regression over a series of simple linear regression models far outweighs the disadvantages. In practice, the response often depends on multiple factors that might interact in some way.

When do you use multiple regression? It's actually a powerful tool for both explanatory analysis and for prediction.

In explanatory analysis, you develop a model to test the statistical significance of the parameter coefficients to determine whether a relationship exists between the response variable and the predictor variables. For example, does increasing the number of police officers affect the crime rate? In a situation like this, you're not necessarily concerned about predicting crime. Instead, you're trying to understand what relationship certain factors have on the crime rate. When you interpret the parameters, you take the magnitudes and signs of the coefficients into account.

In contrast, when you use multiple regression for prediction, your focus is the predictive power of the model. For example, suppose you want to estimate or predict a person's percentage of body fat. To do this, you might fit a predictive regression model that uses skin-fold measurements from different parts of the body to predict the true percentage of body fat. You're less interested in the values of the parameter coefficients (the βs) and their statistical significance (the p-values), although you might use these statistics to help you choose between several candidate models. The predictive model that you choose might have terms that aren't even significant, but it's the model that you've determined best predicts future values of the response variable.

We'll use multiple regression for both explanatory and predictive analyses.