Demo: Assessing Multicollinearity

In this video, we use the Bodyfat data to explore multicollinearity in JMP using all the potential predictors.

Recall that, in this scenario, we are interested in predicting %Fat as a function of several physical measurements.

We'll start by exploring the data. We'll use the Multivariate platform from Analyze, Multivariate Methods to produce both a correlation matrix and a scatterplot matrix. As expected, many of the bivariate correlations are extremely high.

Let's see if multicollinearity is a problem when we fit a regression model. We'll use Fit Model to fit a model for %Fat, with all the potential predictors, Age through Wrist. VIFs are available from the Parameter Estimates panel. To request VIFs, right-click on the panel and select Columns and then VIF. Some of the VIFs are extremely high. Recall that we use a cutoff of 5 or 10 to indicate that there is a problem with multicollinearity.

What happens if we remove BMI? As we discussed earlier, removing BMI might make sense, given that BMI is a function of both Weight and Height.

We'll remove BMI using the Effect Summary table. Let's look at the VIFs again to see whether this has addressed the collinearity. The VIF for Weight is still high.

Let's remove Weight from the model. Some of the VIFs are slightly over 10, but most of the VIFs are now much lower.

In this example, we decided to remove BMI and Weight from the model based on VIFs, and the issue with multicollinearity was, for the most part, resolved.

In practice, the decision of how to address multicollinearity should also be based on subject matter knowledge. And, as we discussed in the previous video, advanced statistical tools, such as penalized regression techniques, should be considered.

Statistical Thinking for Industrial Problem Solving

Close