An Effective Modeling Cycle

You've learned a lot about developing a good regression model for your data. Let's summarize the steps involved in building a model.

The first step is to perform a preliminary analysis. You plot your data, calculate descriptive statistics, and perform correlation analysis. This initial analysis helps you to identify the variables that might be useful in the regression model.

Next, you check for multicollinearity among the variables that you identified in step 1 by using the VIF statistic, condition indices, and variation proportions.

In the third step, you use the information gathered in steps 1 and 2, together with model selection options in PROC GLMSELECT to identify one or more candidate models. You can evaluate potential models by comparing the adjusted coefficients of determination, Mallows' C_p, and information criteria statistics. You can also use PROC REG to produce the plot of residuals versus the predicted values, plots of the residuals versus the predictor variables, and the R-F spread plot to assess the model fit.

Fourth, you need to check and validate your assumptions by examining plots of the residuals versus predicted values and performing other statistical tests, including tests for normality of the residuals, constant variance, and independent observations. You can detect the presence of influential observations by examining plots of RSTUDENT residuals, Cook's D statistics, DFFITS statistics, DFBETAS statistics, covariance ratio statistics, leverage statistics, and partial leverage plots.

If the testing you perform in step 4 indicates a problem, then in step 5 you revise your model and generate a new model. This is the part of the modeling cycle that we'll look at in detail in this section of the course. Based on the nature of the problem you identified in step 4, you might need to return to step 3 and identify new candidate models. When you generate a new model, steps 3, 4 and 5 form an iterative cycle.

The final step is to evaluate the model's predictive capability with data that was not used to build the model. If possible, you build the model with part of your data and use the remainder of the data to determine how well the model fits the data. This final step is not covered in this course. You can learn more about predictive modeling by taking additional SAS training. A list of related SAS courses is available in the Help and Resources section of this course.

Close