Using PROC GLMSELECT to Build a Multiple Regression Model

Let's look at the basic syntax of PROC GLMSELECT for developing a multiple linear regression model. Note that the EFFECT, MODEL, and OUTPUT statements shown here are a subset of the statements that are available for use in PROC GLMSELECT. In the PROC GLMSELECT statement, you can specify various options.As in other procedures, the DATA= option specifies the data set to use for the regression. The OUTDESIGN= option creates a data set that contains the design matrix. By default, this output data set (or design matrix) contains the predictor variables in the selected model. You can then use this output data set in later analyses. The OUTDESIGN= option also has its own options, which are not discussed in this course.

The EFFECT statement enables you to construct new effects, based on predictor variables in the input data set, that can be used as predictors in the model. These effects are referred to as constructed effects to distinguish them from the original variables in the analysis. Several types of constructed effects are available; examples include polynomial effects and spline effects. When you use the OUTDESIGN= option in the PROC GLMSELECT statement, the design matrix data set also contains any constructed effects that the EFFECT statement creates. You learn more about using the EFFECT statement later in this lesson.

In the MODEL statement, you specify the dependent variable (or response variable) and the model-effects (that is, predictor variables). By default, the stepwise method is used in automatic model selection. However, if you want to specify a different method, or no method, you can add the SELECTION= option.

The OUTPUT statement creates a new SAS data set that saves diagnostic measures calculated for the fitted model. You use the OUT= option to specify the data set name. If you don't specify anything else in this statement, the only diagnostic included in the output data set is the predicted response. For each statistic that you want to include in the data set, you specify its keyword. For example, you use the keyword R to specify a residual. For a list of available statistics keywords in the OUTPUT statement, see the SAS documentation. Following the keyword, if you want to specify a variable name other than the one assigned by default to that statistic, you can also specify an equal sign followed by the variable name.

Copyright © 2017 SAS Institute Inc., Cary, NC, USA. All rights reserved.