Expand | Print

Selecting the Best of the Candidate Models

After you use PROC GLMSELECT to generate candidate models, you typically want to select a final model. However, in some cases, you might need to base your selection of a model on additional considerations, such as the following: PROC GLMSELECT results likely do not include all possible candidate models.

To produce a fairly complete array of candidate models, you would need to specify all possible combinations of model selection methods and model selection statistics. A subject-matter expert might suggest the inclusion of certain variables in the model. Monetary considerations might affect model selection. Such considerations might include the cost to collect specific information, patterns of missing data, and management decisions.

Here are the model selection statistics for the three best models that PROC GLMSELECT generated for the cars analysis in the previous demonstration. The first is the three-variable model that contains **Hwympg**, **Hwympg^2**, and **Horsepower**. The two-variable model with **Horsepower** and **FuelTank** does not score well on R square, adjusted R square, or the AIC. The last model, an eight-variable model, has the highest value of R square (because it has the most variables) and adjusted R square. However, the AIC and SBC statistics suggest that it is a complex model. Further work might be needed.

Close