Using Splines for Nonlinear Relationships

Some nonlinear relationships between predictors and a response variable might be too complex for a simple polynomial function to model adequately. In a situation like this, modestly increasing the order of the polynomial might not substantially improve the situation. This problem can occur when the function behaves differently in different parts of the range of X. A common approach is to divide the range of X into segments and fit an appropriate curve in each segment.

Spline functions offer a useful way to perform this type of piecewise polynomial fitting and incorporate the complex nonlinear predictor-response relationships into a regression model. A spline is a smooth function consisting of piecewise polynomials joined at points called knots. In a regression model, a linear combination of polynomial basis functions is fit to the response, and the coefficients for this linear combination are estimated using ordinary least squares. Splines can take a number of forms depending on the degree of the polynomial and the number of knots that are used.

To create spline effects, you can use the EFFECT statement in PROC GLMSELECT and specify the SPLINE effect-type. By default, the EFFECT statement in PROC GLMSELECT uses a penalized B-spline basis (Eilers and Marx, 1996). Keep in mind that interpretation of the parameter estimates associated with a spline is not straightforward. In addition, models that contain spline effects are prone to overfitting. Thus, tuning the shape of a spline by adjusting the number and location of the knots, or both, can be important.

Close