Constructing a Generalized Linear Model

You construct a generalized linear model by selecting the appropriate link function and response probability distribution based on the type of response variable(s) in your data. Predictor variables can include any combination of continuous and categorical variables, and may include interactions.

A number of link functions and probability distributions are available in SAS procedures such as GENMOD and GLIMMIX. The table shown here displays some common distributions in the exponential family and canonical link functions that are available in PROC GENMOD.

You know to use linear regression when the response variable is continuous with a normal distribution. The default link function for linear regression is the identity function. The identity function is the simplest link function and can sometimes be used in other types of generalized linear models. When we use the identity link function, we assume that the expected value of the response can be modeled by a linear combination of the predictors without any transformation (link function) needed. You can think of the identity function as the result of multiplying both sides of the equation by 1, so that both sides stay the same.

You use logistic regression when you have a dichotomous, or binary, response variable that uses the binomial distribution. In this instance, the canonical link function is the logit.

When your response variable is discrete with a Poisson distribution, you apply the log link function. You'll learn more about Poisson regression in the next section of this lesson.

For the gamma distribution, you use the inverse function, which is the default link function in PROC GENMOD. However, when practical or theoretical limitations indicate, analysts often use the log link function to constrain the predicted values to be positive. To learn more about these models, click the Information button.
