Using the GLIMMIX Procedure to Fit a Lognormal Model

When your data exhibits nonconstant variance, you model the response using a probability distribution that accommodates that nonconstant variance (heteroscedasticity). You choose the appropriate distribution based on theoretical knowledge, previous research, both, or by examining the nature of the relationship between the mean and variance of the residuals.

One such distribution, which is used frequently to analyze cost or price data, is the lognormal distribution. A variable is said to follow the lognormal distribution when its logarithm follows a normal distribution. Recall that the cars data set includes the variable **Price**, as well as variables representing other car attributes. In the demonstration, you will see a lognormal regression model fitted to the **cars** data set. The lognormal distribution describes a response variable *Y* having the property that log (*Y*) [the log of Y] follows a normal distribution. The variance of *Y* is proportional to the square of the mean, so a lognormal model explicitly accounts for nonconstant variance.

You can use a generalized linear model to enable the distribution of the response variable to be something other than the normal distribution. In SAS, you can fit these models using PROC GENMOD or PROC GLIMMIX. You'll learn more about Generalized Linear Models in Lesson 5.

The lognormal distribution is available only in PROC GLIMMIX. The GLIMMIX procedure fits statistical models to data with correlations or nonconstant variance, as well as data in which the response variable is not necessarily normally distributed. The syntax for the GLIMMIX procedure is similar to that used in other SAS modeling procedures such as PROC GLMSELECT. In the PROC GLIMMIX statement, you specify the data set to be modelled and various options. In the EFFECT statement, you construct new effects for the model using predictor variables in the input data set. In the MODEL statement, you specify dependent variable and fixed effects. You use DIST=option to specify the built in probability distribution of the data. The output data set contains predicted values and residual diagnostics, computed after fitting the model. By default, all variables in the original data set are included in the output data set.

To fit a lognormal distribution, PROC GLIMMIX applies a log transformation to the response variable and models that transformed response using a normal distribution. As a result, the parameter estimates and standard errors reported by PROC GLIMMIX are on the log-transformed scale.

To make interpretation easier, statisticians usually prefer to obtain predicted means on the original scale of the data. This requires back-transformation of the values. If you were to apply the inverse log (exponential) function to the predicted values, back transformation would provide only unbiased estimates of the median rather than mean of the response variable on the original scale. To overcome this and to obtain low-biased estimates of the means on the original scale, you need to apply a low-bias adjustment factor ($0.5 * \sigma^2$, where σ^2 is mean squared error from the regression model). Thus, the formula for the mean of a lognormal distribution becomes:
E(Y)=~Exp (Xβ^ + $\sigma^{\wedge 2}$ / 2).

Close