General Linear Mixed Model

Before we introduce the general linear mixed model, let's review the general linear model. To fit a general linear model, you can use PROC GLM, PROC GLMSELECT, or PROC REG. The general linear model can also be called the fixed-effects-only linear model. Let's review the components of this model.

Y is the vector of observed response data values. X is the known design matrix for the fixed effects, which is specified in the MODEL statement. β is the vector of unknown fixed-effect parameters. And ε is the vector of random errors. Remember the assumption that ε is normally distributed with a mean of zero and constant variance σ^2. The fixed-effects-only linear model (that is, the general linear model) is actually a special case of the general linear mixed model. The linear mixed model accommodates random effects as well as fixed effects. Let's compare the two models.

In both the general linear model and the general linear mixed model, the dependent variable Y must have a normal distribution given the predictors. This means that the response must be a continuous variable. As you learned in the previous lesson, when Y has a nonnormal distribution given the predictors, the general linear model is extended to the generalized linear model. Likewise, when the response has a nonnormal distribution given the predictors, the general linear mixed model is extended to the generalized linear mixed model.

A response with a nonnormal distribution is typically, but not always, a categorical variable. For example, as you learned earlier, the generalized linear model applies also to Gamma regression, in which the continuous response has a nonnormal distribution given the predictors. In this lesson, however, we focus on the general linear mixed model.

So, how is the general linear mixed model different from the general linear model? The general linear mixed model extends the general linear model in two ways. It includes random effects, and it allows a more flexible specification of the covariance matrix of the random errors. For example, the linear mixed model allows for both correlated error terms and error terms with heterogeneous variances. The general linear mixed model contains an additional term, $Z\gamma$. Z is the known design matrix for the random effects. γ is the vector of unknown random-effect parameters. In the linear mixed model, unlike the linear fixed-effects-only model, the random error terms in ε do not need to be independent and homogeneous.

Let's look at the three main assumptions that apply to general linear mixed models. The first assumption is that the random effects and random errors are normally distributed with a mean of zero and covariance matrices G and R, respectively. Covariance matrices are also known as variance-covariance matrices. So, how do the covariances relate to the model? Because normal data can be modeled entirely in terms of the means and covariances, the two sets of parameters in a linear mixed model actually specify the complete probability distribution of the data. The parameters of the mean model are referred to as fixed-effect parameters, that is, the betas in the term $X\beta$. The parameters of the variance-covariance model are referred to as covariance parameters. The covariance parameters consist of the two covariance matrices: G and R. The G matrix describes the relationship among the gammas in the random-effects term $Z\gamma$. The R matrix describes the relationship among the errors in the epsilon term. The errors are also known as residual errors. The random effects have a covariance matrix because they are not required to be independent of each other. The random errors have a covariance matrix for the same reason.

The second assumption for linear mixed models is that the random effects and random errors are independent of each other. The third assumption is that the means (that is, the expected values) of the responses are linearly related to the predictor variables. In other words, the means are linear in terms of the fixed-effect parameters. In the $X\beta$ term, the betas must be linear; they cannot be raised to a power. For example, a linear mixed model cannot have the term β^2. However, the Xs can be raised to a power, as in polynomial models. Remember that polynomials are linear models because they are linear in the mean parameters (that is, the fixed-effect parameters or the betas).
