SAS Procedures for ANCOVA

Let's look at the SAS procedures that you can use for ANCOVA. Remember that ANCOVA is a combination of regression analysis and ANOVA. As stated in previous lessons, PROC REG is used primarily for regression, in which the independent variables are continuous. PROC GLM is used primarily for ANOVA, in which the independent variables are categorical. You can use either of these procedures for ANCOVA. So, how do you choose?

Your choice is based on the purpose of your ANCOVA analysis. If your purpose is to analyze the effect of the covariate on the response variable, while controlling for differences in the levels of a grouping variable, PROC REG might be more appropriate. On the other hand, you might be interested primarily in analyzing the effect of a grouping variable while controlling for a covariate. In this case, PROC GLM might be more appropriate. You can also use PROC GLMSELECT to perform ANCOVA. The results of ANCOVA from PROC GLMSELECT and PROC GLM are very similar except that PROC GLM produces an additional plot.

Let's make a quick comparison of PROC REG and PROC GLM for ANCOVA. Remember that PROC REG and PROC GLM both use ordinary least squares to fit general linear models of the type $Y = X\beta + \varepsilon$ to your data, where the response variable is continuous. Also remember two assumptions. First, the observations are independent. Second, the errors are normally distributed with a mean of zero and a constant variance of σ^2. This assumption is often checked using the residuals from the fitted model. Both procedures use the ordinary least squares method to obtain parameter estimates and the standard errors.

The main difference between these two procedures has to do with the creation of design variables for categorical predictors. Design variables can also be called dummy variables or indicator variables. PROC GLM supports the CLASS statement, which creates design variables for you based on the specified parameterization method. However, PROC REG does not support the CLASS statement. This means that you must write a DATA step to create any design variables that you need. Depending on the number of categorical variables, and the number of levels in the categorical variables, this process could become tedious. With PROC GLM, it is also a bit easier to work with interaction terms in the model. When you use PROC GLM, you can reference an interaction term directly in the MODEL statement. However, when you use PROC REG, you must first create any interaction terms in a DATA step before you reference them in the MODEL statement. Finally, each procedure has some unique options and plots that are tailored to its particular approach. This lesson focuses on the use of PROC GLM to perform ANCOVA.
