Print

Summary: Lesson 4: Analysis of Covariance

This summary contains topic summaries, syntax, and sample programs.

Topic Summaries

To go to the movie where you learned a task or concept, select a link.

Introduction to Analysis of Covariance (ANCOVA)

Analysis of covariance is the regression between a continuous response (or Y variable) and a combination of continuous and categorical predictors (or X variables). In an ANCOVA model, the continuous predictor is usually called a covariate. The categorical predictors are usually referred to as classification, class, or grouping variables.

An ANCOVA model integrates ANOVA with regression by combining the following: a regression of the response versus the continous predictor (or predictors) while holding the categorical predictors constant, and an ANOVA comparison of the mean response values for different categories or groups of the categorical variable while holding the continuous predictors constant.

In analysis of covariance, the relationship of the response variable to the covariate (that is, the continuous predictor variable) is quantified and used to increase power.

ANCOVA is a combination of regression analysis and ANOVA. 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 most 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 most appropriate.

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 response follows a normal distribution with an expected value $X\beta$ and a constant variance σ^2, given values of the predictor variables.

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.

A mathematical representation of the ANCOVA model is shown here: $Y_{ij} = \mu + \tau_i + \beta_1 X_{ij} + \phi_i X_{ij} + \varepsilon_{ij}$.

Let's see how to code the ANCOVA model in PROC GLM. The CLASS statement specifies one or more classification (or grouping) variables and creates corresponding indicator variables in the model. In the MODEL statement, the predictors include any interaction terms that involve the classification variable.

To enable the values of classification variables to be used in the analysis, PROC GLM creates a numeric indicator variable—that is, a design variable—for each level of each classification variable specified in the CLASS statement. So, how does the model defined in PROC GLM relate to the mathematical model of ANCOVA? PROC GLM uses the ordinary least squares method to fit the general linear model $Y = X\beta + \varepsilon$ to your data.

Let's look at possible outcomes of ANCOVA for the clinical trials scenario. This scenario uses the simplest possible ANCOVA analysis, with one continuous predictor (**BaselineBP**) and one categorical predictor (**Treatment**).

We will explore five different relationships between the variables, as represented in the following models:

- equal slopes and intercepts
- equal intercepts but unequal slopes

- equal slopes but unequal intercepts
- unequal slopes and intercepts, and
- equal slopes where all slopes are zero

One possible relationship between variables is that both the slopes and the intercepts for the three treatments are the same. n this second possible model, the three levels of the classification variable **Treatment** have the same intercept, but different slopes. In the third possible model, the slopes are the same but the intercepts are different for the treatments. Another possibility is that both the slopes and the intercepts are different for at least two of the treatments. Finally, the rate of change of the Y variable (**BPChange**) is zero for all three treatments with a change in the X variable (**BaselineBP**).

Least Squares Means for ANCOVA Models

You can use the LSMEANS statement to compute least squares means for ANCOVA as well, to adjust for a covariate (a continuous predictor variable). And, just as you can in ANOVA, you can specify options in the LSMEANS statement to perform multiple comparison tests in order to determine which group means are different.

In ANCOVA, your main interest is generally to compare group means. Specifically, you compare mean response values for different categories or groups of the categorical variable (or the CLASS variable) while holding the covariate X constant. To generate least squares means, you can use the LSMEANS statement in PROC GLM.

LSMEANS *effects / options*;

By default, the LSMEANS statement calculates the least squares mean of the dependent variable for each group at the mean value of the covariate In the AT option, you specify a covariate (in the syntax shown here, this is referred to as variable), an equal sign, and a value of the covariate.

AT *variable* **=** *value*

As in ANOVA, you can use multiple comparison tests to determine which group means are different. So, you can specify multiple values of the covariate in your program.

Diagnostics and Remedial Measures for ANCOVA Models

You can use either PROC REG or PROC GLM to perform diagnostics on your ANCOVA model. However, PROC REG has more diagnostic features than PROC GLM, so we will focus on PROC REG.

PROC REG and PROC GLM both have options that request a panel of summary diagnostic plots. PROC REG also has the following features that PROC GLM does not offer: multicollinearity diagnostics and plots of the DFBETA and DFFITS statistics. These last two statistics are useful for identifying influential observations. Unlike PROC REG, PROC GLM supports the CLASS statement to create the design variables needed for the categorical variables in an ANCOVA model.

Sample Programs

Conducting ANCOVA Using PROC GLM

```
proc glm data=mydata.trials;
   class treatment;
   model bpchange = treatment baselinebp
                    treatment*baselinebp / solution;
title 'Analysis of Covariance';
run;
quit;
```

Using Least Squares Means and Multiple Comparison Tests in an ANCOVA Model

```
proc glm data=mydata.trials;
   class treatment;
   model bpchange = treatment|baselinebp;
   lsmeans treatment / pdiff=all adjust=tukey;
   lsmeans treatment / at baselinebp=90 pdiff=all adjust=tukey;
   lsmeans treatment / at baselinebp=100 pdiff=all adjust=tukey;
title 'Least Squares Means for ANCOVA Model';
run;
quit;
```

Performing Diagnostics and Remedial Measures on an ANCOVA Model

```
ods select none;
proc glmselect data=mydata.trials outdesign=design;
   class treatment;
   model bpchange = treatment|baselinebp / selection=none;
run;
%put macro variable _glsmod=&_glsmod;

ods select ParameterEstimates DiagnosticsPanel
    DFFITSPlot DFBETASPanel;
proc reg data=design plots=(dfbetas dffits);
   model bpchange=&_glsmod / vif influence;
title 'Check Collinearity on ANCOVA Model';
run;
quit;

proc stdize data=mydata.trials method=mean
            out=trials2c (rename=(baselinebp=baselinebpc));
   var baselinebp;
run;

ods select none;
proc glmselect data=trials2c outdesign=design2c;
   class treatment;
   model bpchange = treatment|baselinebpc / selection=none;
title 'Check Collinearity on Centered ANCOVA Model';
run;

ods select ParameterEstimates DiagnosticsPanel
    DFFITSPlot DFBETASPanel;
proc reg data=design2c plots=(dfbetas dffits);
   model bpchange=&_glsmod / vif influence;
run;
quit;

proc sgplot data=mydata.trials;
   reg y=bpchange x=baselinebp / group=treatment;
   refline 95 / axis=x;
title 'Original Data';
run;

proc sgplot data=trials2c;
  reg y=bpchange x=baselinebpc / group=treatment;
  refline 0 / axis=x;
title 'Centered Data';
run;
```

Close