🚀 **Conducting ANCOVA Using PROC GLM**

This demonstration shows how to conduct an analysis of covariance using PROC GLM.

We want to determine the effect of treatment and baseline blood pressure on the change in blood pressure. In PROC GLM, the data set is trials. The CLASS statement specifies the categorical variable **Treatment**. The MODEL statement specifies the model: the response variable is **BPChange** and the three predictors are **Treatment**, **BaselineBP**, and the interaction of **Treatment** and **BaselineBP**. Notice the asterisk in the interaction term. PROC GLM does not produce the solutions for the parameter estimates unless you request them. We want to see the table of parameter effects so, at the end of the MODEL statement, we specify the SOLUTION option.

By default, if you specify an ANCOVA model with one or two CLASS variables and one continuous variable, PROC GLM produces an ANCOVA plot of the response values versus the covariate values, and lines representing the fitted relationship within each classification level. We'll want to look at the plot first, to get a sense of the relationship among the variables.

Let's run the code.

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

At the bottom of the results is the Analysis of Covariance plot for **BPChange**. This plot shows a scatter plot of the data plus the fitted models of the three treatment levels. It is not surprising that the *Placebo* line is almost flat; the placebo is not likely to have much effect on blood pressure. However, the other two treatments (*Approved Drug* and *New Drug*) have a definite downward slope, which indicates an increase in the reduction of blood pressure. As the baseline blood pressure goes up, the change in blood pressure goes down. These two treatments appear to have a similar slope although it looks like the approved drug might be slightly more effective than the new drug. By looking at the plot alone, however, we cannot determine whether the intercepts for *Approved Drug* and *New Drug* are significantly different from each other.

Now let's look at the numbers. At the top of the results, the Class Level Information table shows the three treatment levels. This is followed by the number of observations read and used.

In the table that shows the Type III sums of squares, the last row, the F statistic for the **BaselineBP** and **Treatment** interaction term tests the null hypothesis that the slopes of **BaselineBP** are equal across all treatments. The *p*-value is less than 0.0001. Therefore, you reject the null hypothesis and conclude that the slopes for the treatments are not all equal to each other. The unequal slopes model is needed for this data set.

The last table shows the solution, or the parameter estimates which can be used to write the regression equations for each treatment. Let's look at each of these parameter estimates, starting with the intercept, in the first row.

The intercept term estimates the intercept for the last level of the grouping variable, in this case *Placebo*. The parameter estimate for the *Approved Drug* treatment estimates the difference between the intercept of *Approved Drug* and the intercept of the last group, *Placebo*. Likewise, the parameter estimate for the *New Drug* treatment estimates the difference between the intercept of *New Drug* and the intercept of *Placebo*. The parameter estimate for the *Placebo* treatment is zero because it estimates the difference between the intercept of *Placebo* and the intercept of the last group, *Placebo*. The parameter estimate for **BaselineBP** is the estimate of the slope for the last level of the grouping variable, *Placebo*. The parameter estimate for the interaction between **BaselineBP** and the *Approved Drug* treatment estimates the difference between the slope of *Approved Drug* and the slope of the last group, *Placebo*. The parameter estimate for the interaction between **BaselineBP** and the *New Drug* treatment estimates the difference between the slope of *New Drug* and the slope of the last group, *Placebo*. The parameter estimate for the interaction between **BaselineBP** and the *Placebo* treatment is zero because it estimates the difference between the slope of *Placebo* and itself.

To see the set of equations that represents the regression model for each of the three treatments, click the Information button.

Close