ANOVA in SAS

Once the data is imported into SAS, a variety of analyses can be performed using SAS Analyst by selecting Solutions: Analysis: Analyst. After opening Analyst select File: Open by SAS Name to open your data.

In order to perform an analysis in SAS, the roles of the variables must be identified. Start by deciding what the ANOVA should be. The key to the correct analysis of variance with SAS is writing the correct model statement for the experiment based on the desired ANOVA. For the calculating machine example, the response or dependent variable is Time, the treatments are the two calculating Machines, and for each machine there are 10 Reps. The ANOVA table (source of variation and df) for this completely randomized design is:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>19</td>
</tr>
<tr>
<td>Between Machines</td>
<td>1</td>
</tr>
<tr>
<td>Error = Reps within Machines</td>
<td>18</td>
</tr>
</tbody>
</table>

This ANOVA defines the model used for the analysis; in this model the response depends only on the Machine and on random variation or error. To perform the ANOVA, select Statistics: ANOVA: Linear Models. This is the most general type of ANOVA, and can be used for one-way, two-way, factorial, split plot and other analyses. The variables appear in a box on the left of the window. Select Time and click on Dependent to identify this as the response variable. Select the treatment variable, Machine, and click on Class to identify this source of variation as a fixed independent variable. The Class box lists the source of variation in the ANOVA table except for the error and total. Any source of variation not specified in the model statement will be pooled in the error SS.

Note that holding down the Control or Shift key while selecting variables makes it possible to select several variables at once.

What about the Error (Reps within Machines) term? SAS Analyst will automatically calculate both the Total and the Error. The Error term is simply all of the variability that has not been explained by the model.

From the double row of buttons along the bottom of the window, select Plots. In this window, choose to Plot Dependent Means for Main Effects. Select the Residuals (residuals are the errors) button along the top and Plot Residuals vs Variables, then select Ordinary under Residuals and Independents under Variables. Select the OK button to save your choices.

Next select the Titles button along the bottom. The first line of the title will always be your name. The second line will identify the analysis, for example, calculating machine CRD.
Click OK to get back to the main ANOVA window and in the ANOVA window click OK to run the analysis.

**Analysis Results**

The results of the analysis will automatically open in a new window. At the top the SAS Procedure used is identified. GLM stands for General Linear Models. Next in the output the Class information gives the name of each class or fixed variable, the number of levels of each, and the values of each. Here the treatments (machines A and B) are listed as the two levels of Class. Following this the numbers of observations found and used are given. If there were missing values, it would be noted here. These lines provide an important check to ensure that SAS actually used the data that it was supposed to use.

The next section of the results provides the ANOVA in two tables. The first table has the sources of total variation (corrected total or total SS) partitioned into Model and Error. The Model is the total of all the sources of variation or variables included in the Class box. The mean square (variance) is the SS divided by the df. The observed F value is the ratio of the MS model/MS error. The P value (Pr > F) is the probability of obtaining an F-value this large or larger assuming that the null hypothesis is true.

This is followed the R-square, which is a measure of how much variation in the dependent variable (Time) is accounted for by the model. The Coefficient of Variation (CV) is the Std. Dev./mean*100%. The Root MSE estimates the standard deviation of the dependent variable (Time) and is computed as the square root of the MS error. (standard deviation). The Time Mean is the mean of the dependent variable.

The second ANOVA table subdivides the variance for each component or variable in the model, in this case only Machine. In this CRD example, the effect of Machine listed here is identical to the effect of the model listed in the first section of the ANOVA because Machine was the only variable in the model. The F value is obtained by testing with the MS error from the first table. The F test for Machine tests the hypothesis that the means for Machines A and B are equal.

To view the plots created, use the tree in the Results window or at the left of the Analyst window and find and select the Means Plot and the Residuals Plot. The Residuals Plot is very useful for checking for deviations from normality of the data.

If any problems arise, check the Log window for error messages.