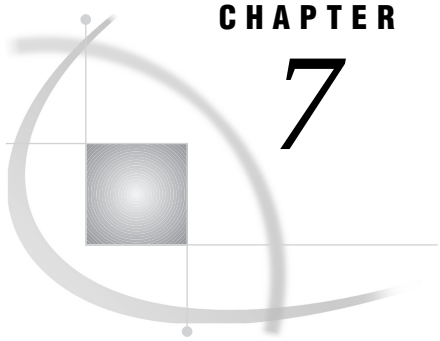


CHAPTER

7

Doing More with Data Analysis

Introduction

This chapter shows you how to use **Data Analysis** to generate a one-way frequency table, perform a linear regression, and perform an analysis of variance. You must have SAS/STAT software licensed to complete most of the tasks that use **Data Analysis**.

There are, of course, many more analyses that you can perform with **Data Analysis**, but there is not sufficient space to describe them all in detail here. However, the techniques for performing data analysis tasks are similar for all tasks. When you have mastered the techniques described in this chapter, you should have no trouble performing any of the other analyses.

For more information on using SAS software for statistical analysis, see the *SAS/STAT User's Guide* and the *SAS Procedures Guide*.

Generating a One-way Frequency Table

You can use the **Elementary** item on the Data Analysis menu to generate frequency tables. Frequency tables show the distribution of column values and can include frequencies, cumulative frequencies, percentages of the total frequency, and cumulative percentages for each *level*, or distinct value, of the analysis column. In this section you create a one-way frequency table showing the frequency of use of a seat belt in the automotive accidents recorded in the HIGHWAY table.

The one-way frequency table task can generate tables for both numeric and character columns. However, if you have a numeric column with many distinct values, the resulting table may not be informative.

Note: SAS/STAT software is *not* required to complete this task. △

Additional Information

For additional information on creating frequency tables, refer to "The FREQ Procedure" in the *SAS Procedures Guide*.

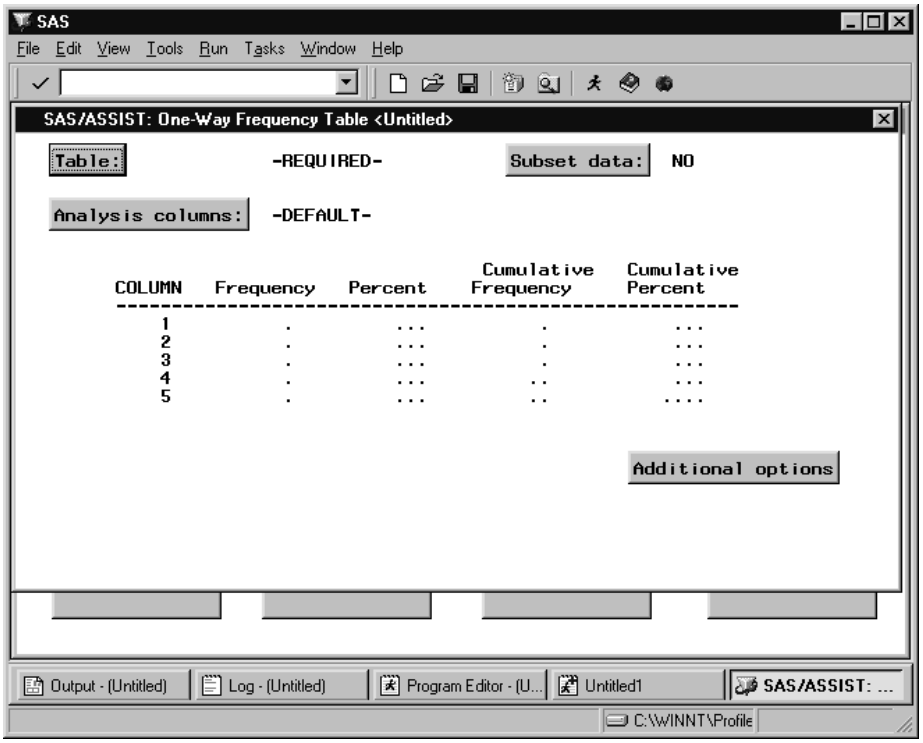
Instructions

1 Follow this selection path:

Tasks ► Data analysis ► Elementary ► Frequency tables ► Generate one-way frequency table

The One-Way Frequency Table window appears.

Display 7.1 One-way Frequency Table Window



2 Use the **Table** button to select the SASUSER.HIGHWAY table.

3 Use the **Analysis columns** button to select SEATBELT as the analysis column.

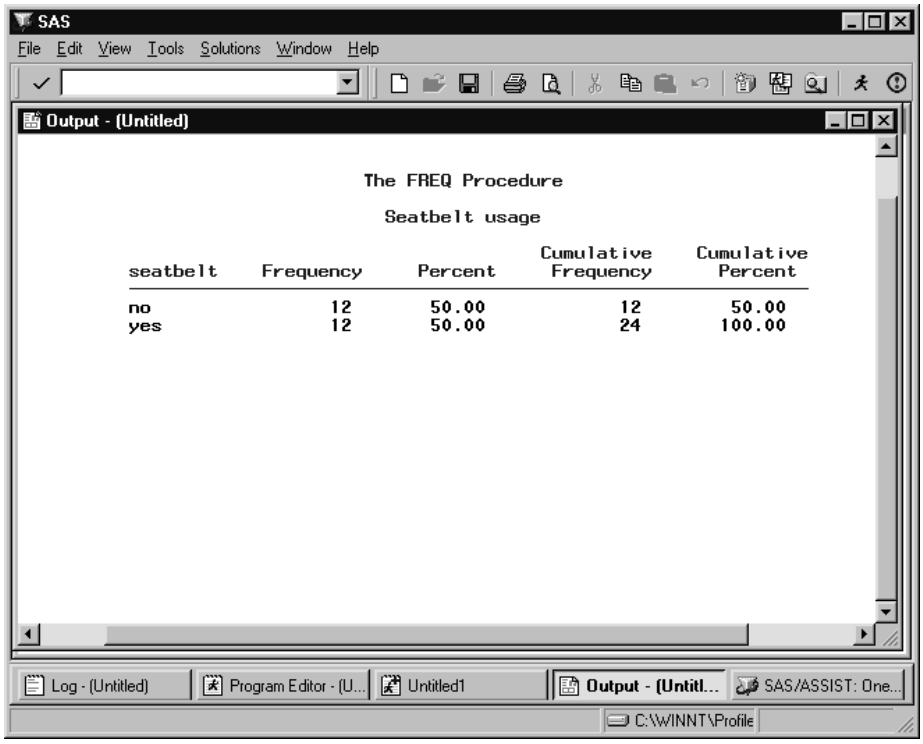
4 Follow this selection path:

Run ► Submit

Note: You can select as many analysis columns as you want, up to the number of numeric columns in the table. A frequency table is generated for each analysis column. Any columns selected as BY or weight columns are not displayed in the Select Table Variables window. △

The frequency table appears showing how often a seat belt was worn in the automotive accidents recorded in the HIGHWAY table.

Display 7.2 One-way Frequency Table Output



5 Return to SAS/ASSIST software from the Output window. See "Returning to SAS/ ASSIST Windows from the Output Window" on page 9 for more information.

Performing a Linear Regression

Regression analysis is an analysis of the relationship between one dependent column and one or more independent columns. You can use **Regression** on the Data Analysis menu to perform linear regression, logistic regression, and regression for correction with autocorrelation. When you use the **Linear Regression** item, you fit linear regression models by the method of least-squares. In addition, you can use **Linear Regression** to generate scatter plots and various diagnostic measures. You must have SAS/STAT software licensed to complete this task.

In this section, you perform a linear regression showing the relationship between the oxygen consumption rate of subjects while they run and the time it took for the subjects in the FITNESS table to run 1.5 miles. You use the 95% individual confidence interval option to display the 95% upper- and lower-confidence limits for an individual value.

To superimpose a regression line on a plot, refer to "Doing More with Plots" on page 108.

Additional Information

For additional information on performing linear regressions, refer to "The REG Procedure" chapter in the *SAS/STAT User's Guide*.

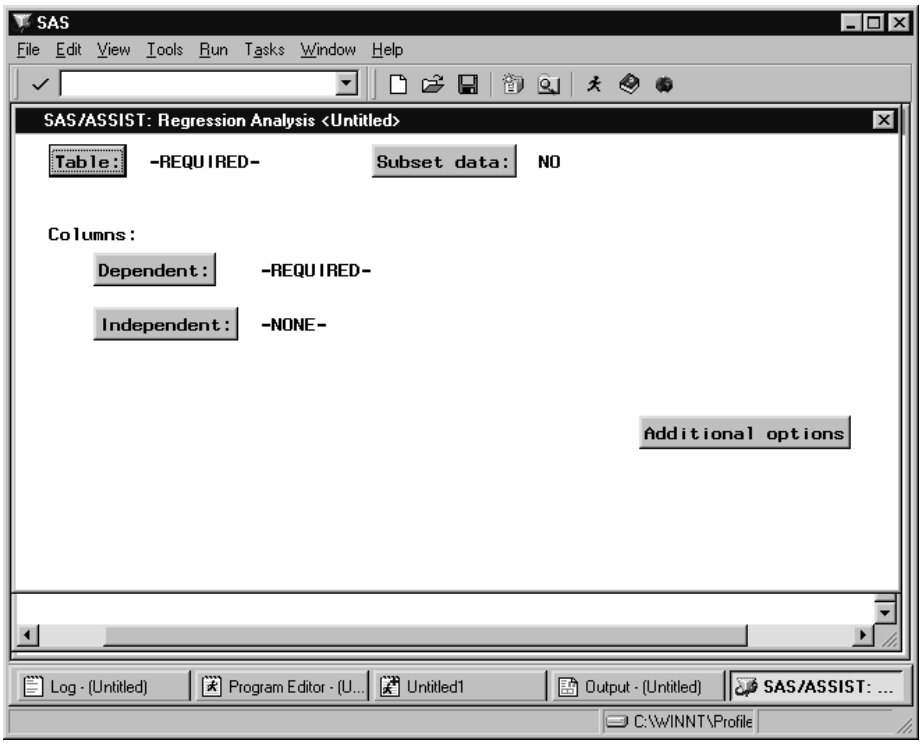
You can perform other types of regressions, in addition to the linear regression, with SAS/ASSIST software. For more information on logistic regression, which requires SAS/STAT software, refer to "The LOGISTIC Procedure" in *SAS/STAT User's Guide*. For information on regression with correction for autocorrelation, which requires SAS/ETS software, refer to "The AUTOREG Procedure" in *SAS/ETS User's Guide*.

Instructions

1 To display the Regression Analysis window, follow this selection path:

Tasks ▶ Data Analysis ▶ Regression ▶ Linear

Display 7.3 Regression Analysis Window



2 Use the **Table** button to then select the SASUSER.FITNESS table.

3 Use the **Dependent** button to select OXYGEN as the dependent column.

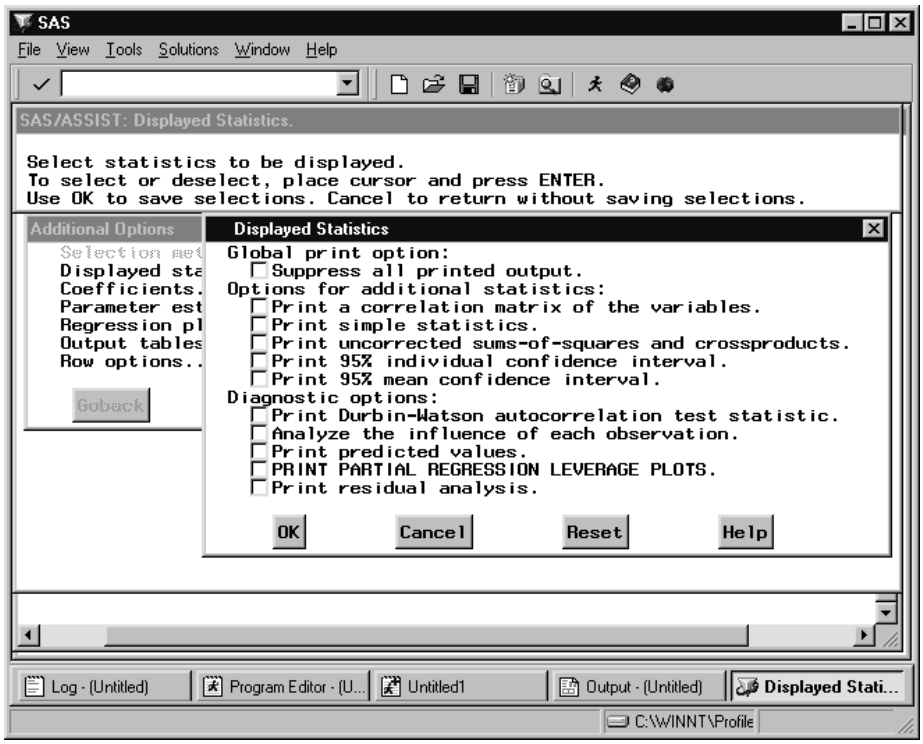
The dependent columns contain the observed values that the regression equation attempts to predict. The Select Table Variables window displays all the numeric columns in the FITNESS table except for any columns selected as BY or independent columns. A separate regression analysis is generated for each dependent column that you select.

4 Use the **Independent** button to select RUNTIME as the independent column.

The independent columns contain the values used to predict the dependent columns. The Select Columns window displays all the numeric columns in the FITNESS table except for any columns selected as BY or dependent columns.

5 Select **`Additional options`**, then **`Displayed statistics`**. The Displayed Statistics window appears.

Display 7.4 Displayed Statistics Window

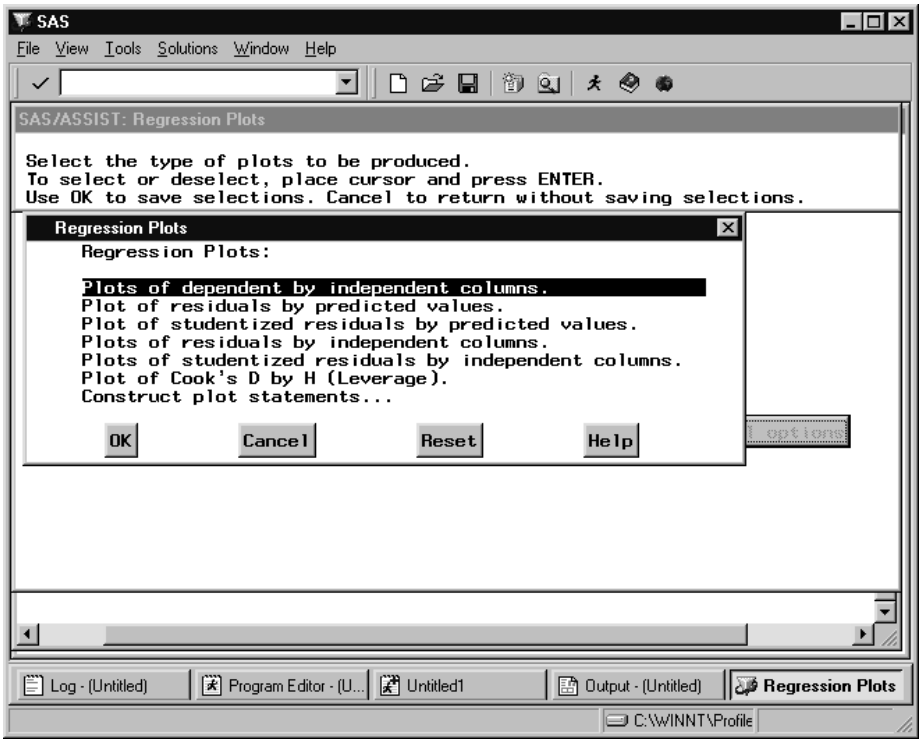


6 Select **`Print 95% individual confidence interval`**, then select **`OK`** to return to the Additional Options window.

You use the 95% individual confidence interval to display the 95% upper- and lower-confidence limits for an individual value to reflect not only the variability in the predicted mean value, but also the variability in a single future observation.

7 Select **`Regression plots`**. The Regression Plots window appears.

Display 7.5 Regression Plots Window



8 Select **`Plots of dependent by independent columns`**.

By choosing to generate a plot for each dependent column with each independent column, you can detect a nonlinear relationship between columns in the regression model.

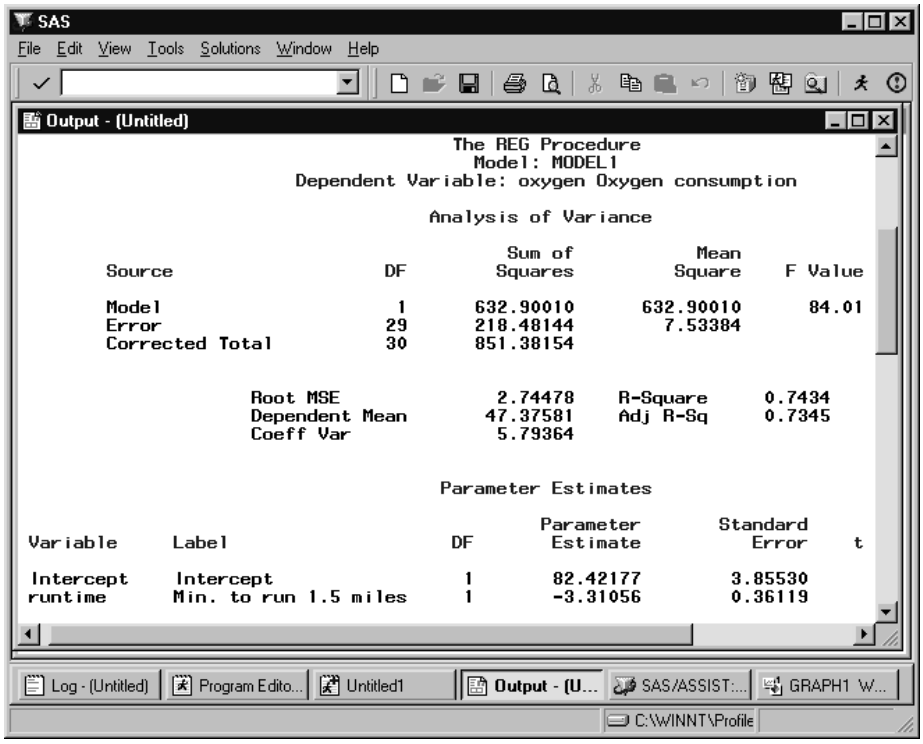
9 Select **`OK`** and then **`Goback`** to return to the Regression Analysis window.

10 Follow this selection path:

Run ► Submit

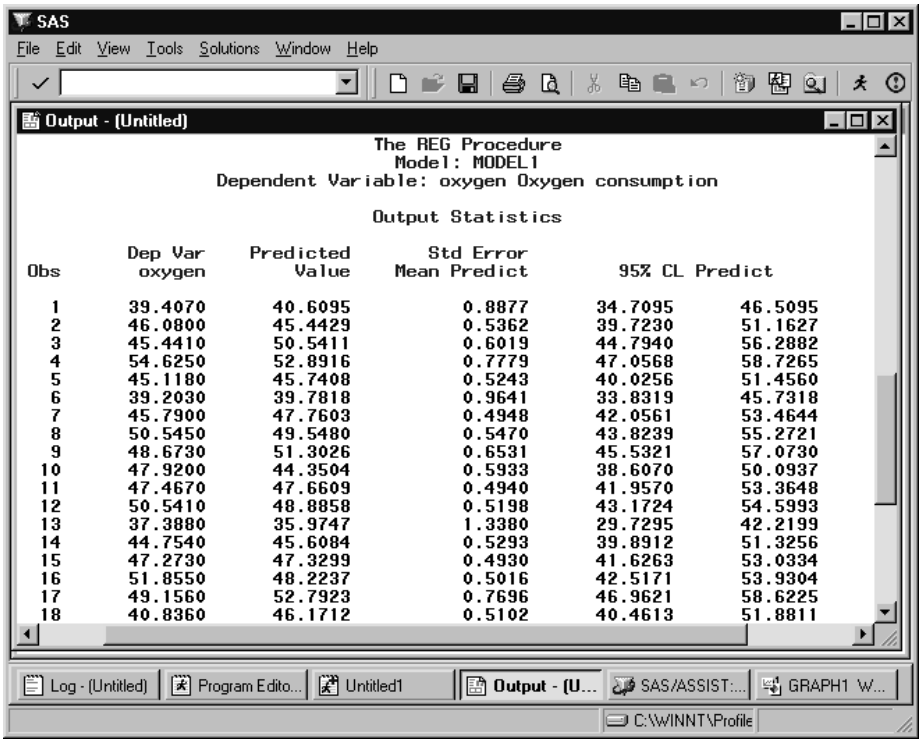
The analysis appears in the Output window.

Display 7.6 Regression Analysis Output



11 Use the scroll bars or the FORWARD command or function key to display the page of the analysis that shows the 95% individual confidence interval in the **`95% CL Predict`** columns.

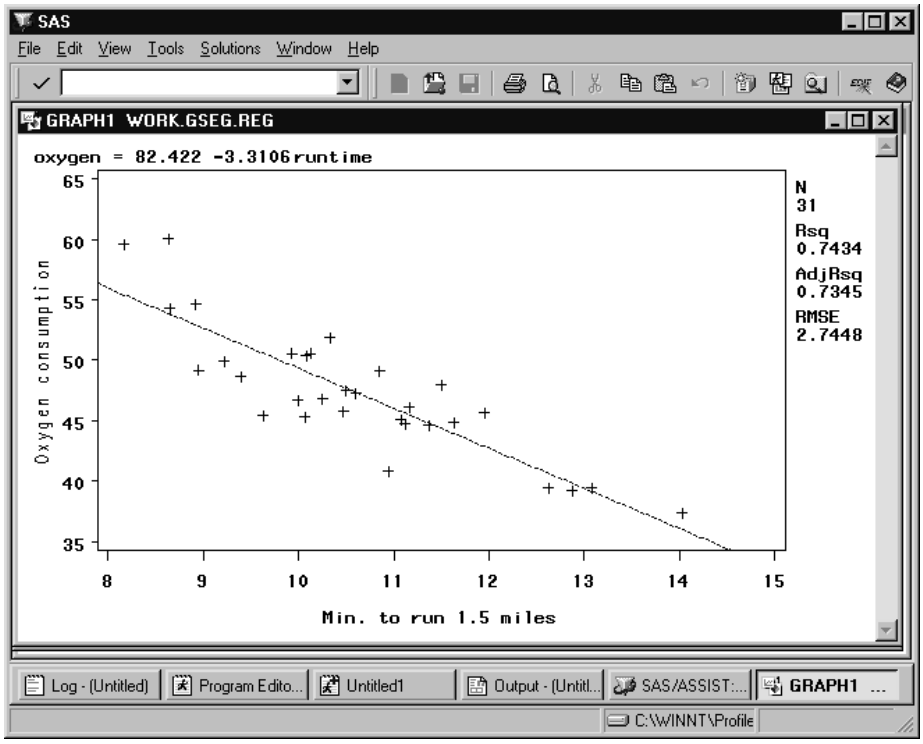
Display 7.7 Regression Analysis Output (continued)



Individual confidence intervals are referred to as prediction intervals, hence the word **`Predict`** in the output.

12 Access the GRAPH window to see the plot of the dependent by independent column.

Display 7.8 Regression Plot Output



13 Return to SAS/ASSIST software from the Output window. See "Returning to SAS/ASSIST Windows from the Output Window" on page 9 for more information.

Performing an Analysis of Variance

You can use the **Anova** item on the Data Analysis menu to perform an analysis of variance (ANOVA). You can construct a simple analysis of variance, analysis of covariance (ANCOVA), multivariate analysis of variance (MANOVA), and repeated measures ANOVA. You must have SAS/STAT software licensed to complete this task.

In this section, you perform an analysis of variance on the VENEER table to show the relationship between the brand of veneer and the amount of veneer that was worn away during testing.

Additional Information

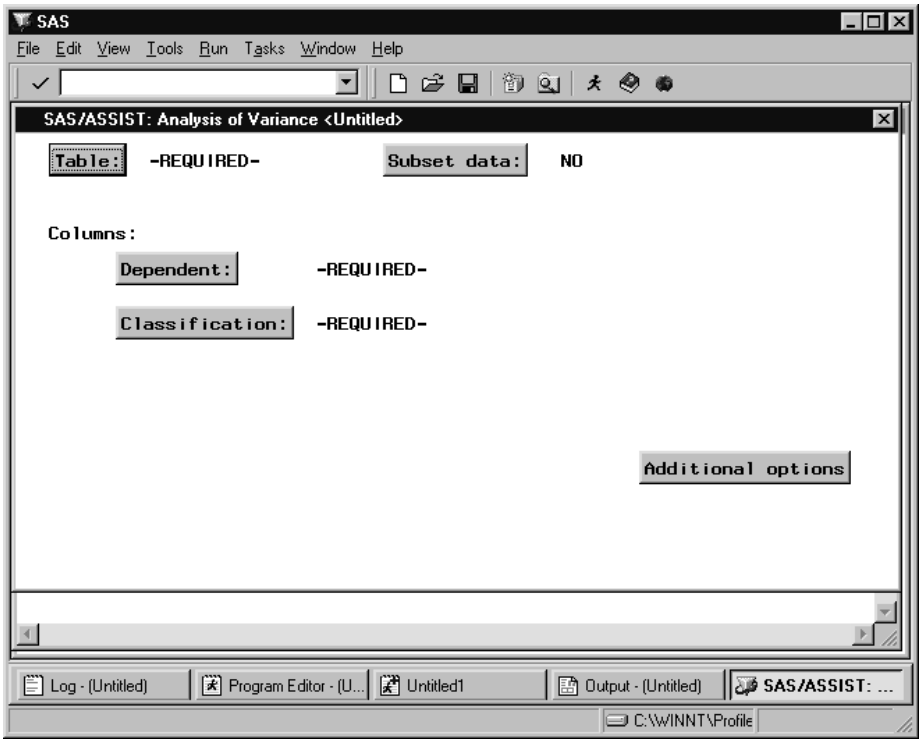
For additional information on performing an analysis of variance, refer to the GLM and ANOVA procedures in the *SAS/STAT User's Guide*.

Instructions

1 To display the Analysis of Variance window, follow this selection path:

Tasks ▶ Data Analysis ▶ Anova ▶ Analysis of Variance

Display 7.9 Analysis of Variance Window



2 Use the **Table** button to select the SASUSER.VENEER table.

3 Use the **Dependent** button to select **WEAR** as the dependent column.

The dependent column contains the measurements to be analyzed. The Select Table Variables window displays all the numeric columns in the VENEER table except any columns selected as BY or classification columns. If you have more than one dependent column, you can generate a separate ANOVA model for each dependent column or one multivariate model for the set of dependent columns.

4 Use the **Classification** button to select **BRAND** as the classification column.

The classification columns identify analysis groups. The variation in the dependent column values is analyzed within and across the classification groups to determine whether or not the classification columns are significant sources of variation. The Select Table Variables window displays all the columns in the VENEER table except for any columns selected as BY or dependent columns.

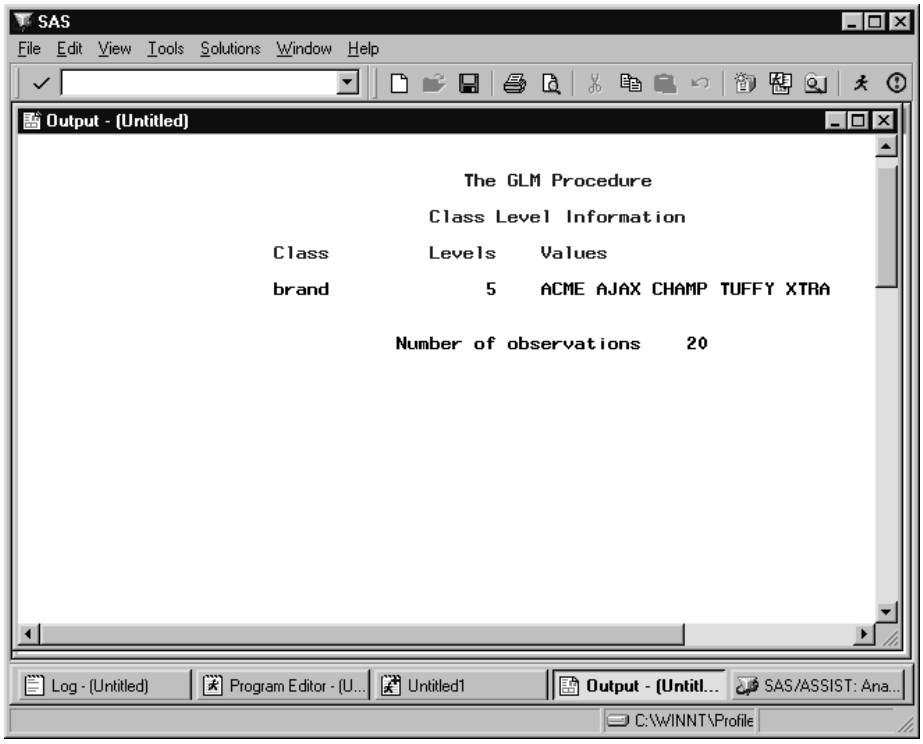
Each classification column is treated as a main effect in the model. You can modify the ANOVA model to add interactions, nested effects, covariates, and random effects by using the **Additional options** item in the Analysis of Variance window.

5 Follow this selection path:

Run ▶ Submit

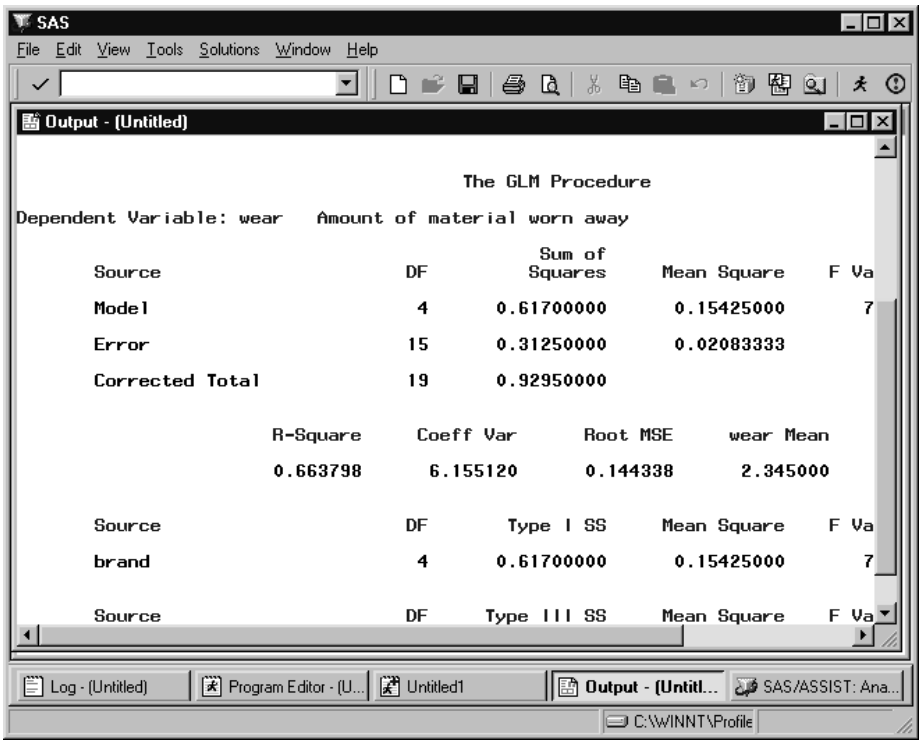
The first page of the analysis appears showing the class level information.

Display 7.10 Analysis of Variance Output



6 Use the scroll bars or the FORWARD command or function key to display the next page of the analysis.

Display 7.11 Analysis of Variance Output (continued)



Refer to the GLM and ANOVA procedures in the *SAS/STAT User's Guide* for information about interpreting the statistics in this report.

7 Return to SAS/ASSIST software from the Output window. See "Returning to SAS/ASSIST Windows from the Output Window" on page 9 for more information.

Performing a Repeated Measures ANOVA

Repeated measures designs are characterized by recording several measurements over time or space on the same experimental unit. This section shows how to perform a repeated measures ANOVA with SAS/ASSIST software. The first part of the process is data entry. For this particular example, you apply a logarithmic function to the data and then you perform the repeated measures ANOVA.

This example uses the data from Example 7 in "The GLM Procedure" chapter of *SAS/STAT User's Guide*.

Entering the Data

In this section, you enter the raw data using the **`Create Data`** task. For more information on this task, refer to the "Entering Data Interactively" chapter in *Getting Started with the SAS System Using SAS/ASSIST Software*.

1 Follow this selection path:

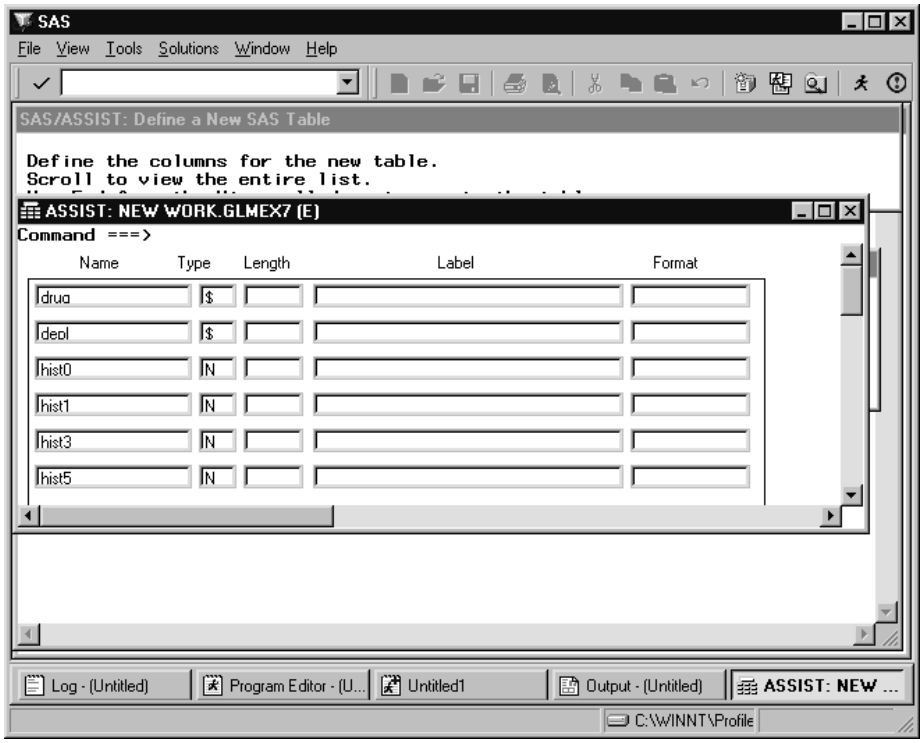
Tasks ► Data Management ► Create data ► Interactively ► Enter data in tabular form

The Select a New SAS Table to Create window appears.

2 In the **Table** field, type **GLMEX7**. If desired, select **Permanent**, select a library in which to store the table, and select **OK.**. Select **Goback**. The Define a New SAS Table window appears.

3 Define the columns in the new table by editing this window so that it looks like the following display.

Display 7.12 Define a New SAS Table Window With Columns Defined



You might need to resize the window to see all six rows.

4 Follow this selection path:

File ▶ Close

If prompted, select **Yes** to save the changes. The FSEDIT window appears.

5 Enter the following data into the table. Note that the **hist5** column has a missing value for row 6 (**drug=Morphine, depl=N**).

Table 7.1 Data for Repeated Measures ANOVA Example

Row (Obs)	drug	depl	hist0	hist1	hist3	hist5
1	Morphine	N	0.04	0.20	0.10	0.08
2	Morphine	N	0.02	0.06	0.02	0.02
3	Morphine	N	0.07	1.40	0.48	0.24
4	Morphine	N	0.17	0.57	0.35	0.24
5	Morphine	Y	0.10	0.09	0.13	0.14

Row (Obs)	drug	depl	hist0	hist1	hist3	hist5
6	Morphine	Y	0.12	0.11	0.10	
7	Morphine	Y	0.07	0.07	0.06	0.07
8	Morphine	Y	0.05	0.07	0.06	0.07
9	Trimethaphan	N	0.03	0.62	0.31	0.22
10	Trimethaphan	N	0.03	1.05	0.73	0.60
11	Trimethaphan	N	0.07	0.83	1.07	0.80
12	Trimethaphan	N	0.09	3.13	2.06	1.23
13	Trimethaphan	Y	0.10	0.09	0.09	0.08
14	Trimethaphan	Y	0.08	0.09	0.09	0.10
15	Trimethaphan	Y	0.13	0.10	0.12	0.12
16	Trimethaphan	Y	0.06	0.05	0.05	0.05

6 Follow this selection path to exit the FSEDIT window:

File ► Close

If prompted, select **Yes** to save the changes.

Calculating Logarithms

In order to minimize correlation between the mean and the variance of the data, the logarithm of the data needs to be calculated.

1 Follow this selection path:

Tasks ► Data management ► Subset/Copy

The Subset or Copy a Table window appears.

2 If the active table is not GLMEX7, use the **Table** button to select it from the WORK library, or, if you chose to save the table permanently, the permanent library where you saved the table.

3 Use the **Output Table** button. In the Output Table or View window, specify GLM7OUT as the name of the output table. Select **Temporary** or **Permanent**, as desired, before selecting **Goback**.

4 Select **Define new columns**. The Define or Modify a Column window appears.

5 In the **Column** field, type **LHIST0**.

6 Select **Initialize**. The Enter Numeric Expression window appears. Select **Function**, then **Mathematical Functions**. The Select Data window appears.

7 Select **LOG(n)**. The **Specify Arguments to a Function** window appears.

8 In the **Value for parameter** field, type **HIST0** and select **OK**,

9 Select **OK** twice. The Define New Columns window appears.

10 Select **Add new column**.

Repeat steps 5 through 10 to create the LHIST1, LHIST3, and LHIST5 columns, using the HIST1, HIST3, and HIST5 columns, respectively, as parameters. When finished, select **OK** (instead of **Add New Column**) from the Define New Column to return to the Subset or Copy a Table window.

11 Select **Submit** from the **Run** menu. If desired, select **OK**, then **Goback** to view the table; otherwise, select **Cancel**, then **Goback**.

Performing the Analysis

You are now ready to perform the repeated measures ANOVA.

1 Follow this selection path:

Tasks ▶ Data Analysis ▶ Anova ▶ Analysis of Variance

The Analysis of Variance window appears.

2 If the active table is not GLM7OUT, use the **Table** button to select the GLM7OUT table.

3 Use the **Dependent** button to select the LHIST0, LHIST1, LHIST3, and LHIST5 columns.

4 Use the **Classification** button to select the DRUG and DEPL columns.

5 Follow this selection path:

Additional options ▶ Model effects ▶ Interactions

The Interactions window appears.

6 Select **DRUG**, then *****, then **DEPL** to construct the **DRUG*DEPL** interaction. Select **OK**, then **Goback**.

7 Select **Analysis type**, then **Repeated measures**, then **Factor names and levels**. The Repeated measures Factors window appears.

8 In this window, you specify the repeated factor and the number of levels associated with that factor. In the first row, type **TIME** for **Factor Name** and **4** for **Number of Levels**. Select **OK**.

9 Select **Factor Values**. The Repeated measures Factor Levels window appears.

10 In this window, you specify the intervals for the repeated factor. In the four active spaces under **Level Values**, type **0**, **1**, **3**, and **5**. Select **OK**.

11 Select **Factor Transformations**. The Repeated Measures Transformations window appears. In this window, you specify a single-degree-of-freedom contrast. For descriptions of each transformation, select **Help**.

12 Select **Polynomial**, then **OK**.

13 Select **Options**. The Repeated Measures Options window appears.

14 Select **Test within-subject effects** to produce an analysis-of-variance table for each contrast defined by the within-subject factors. Select **OK**.

15 Select **Goback** three times to return to the Analysis of Variance window.

16 Select **Submit** from the **Run** menu. The results of the analysis appear in the Output window.

17 If desired, compare these results to those in Example 7 in "The GLM Procedure" chapter of *SAS/STAT User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *Doing More With SAS/ASSIST Software, Version 8*, Cary, NC: SAS Institute Inc., 1999. pp. 282.

Doing More With SAS/ASSIST Software, Version 8

ISBN 1–58025–520–5

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999