Chapter 1
Overview

Chapter Table of Contents

Chapter 1
Overview

Introduction

The Analyst Application is a data analysis tool that provides easy access to basic statistical analyses. The application is intended for students and researchers as well as experienced SAS users and statisticians.

This interface takes a task-oriented approach to produce analyses and associated graphics. You can compute descriptive statistics, perform simple hypothesis tests, fit statistical models with regression and analysis of variance, and perform survival analysis as well as some multivariate analyses.

Most of the tasks provide access to analyses performed by SAS/STAT software, but some provide analyses not currently available with SAS/STAT procedures, such as certain hypothesis tests and basic sample size and power computations. In addition, you can produce many types of graphs, including histograms, box-and-whisker plots, probability plots, contour plots, and surface plots.

The Analyst Application enables you to input data in many ways, including importing data from other sources such as Excel files, inputting SAS data sets, or manually entering the data yourself. The data are displayed in a data table in which columns correspond to variables and rows correspond to observations or cases. You can edit individual elements in the data table, and you can create new columns and rows. You can also perform a number of other data manipulations such as subsetting the data, performing transforms, recoding, and stacking and splitting columns.

Once the data are ready, you specify tasks from pull-down menus, a customizable toolbar, or an index of commonly used task descriptions. The analysis results and plots are presented in separate windows and managed by a tree-list structure called a project tree. The underlying SAS code used to produce the results is available as a node in the project tree, and results can also be displayed in HTML form and viewed with a web browser. You can save projects and then recall them for further work.

Getting Started

In this example, you bring data into the Analyst data table, perform a regression analysis, and see the results in the project tree.

Bring Up Analyst and Create a Sample Data Set

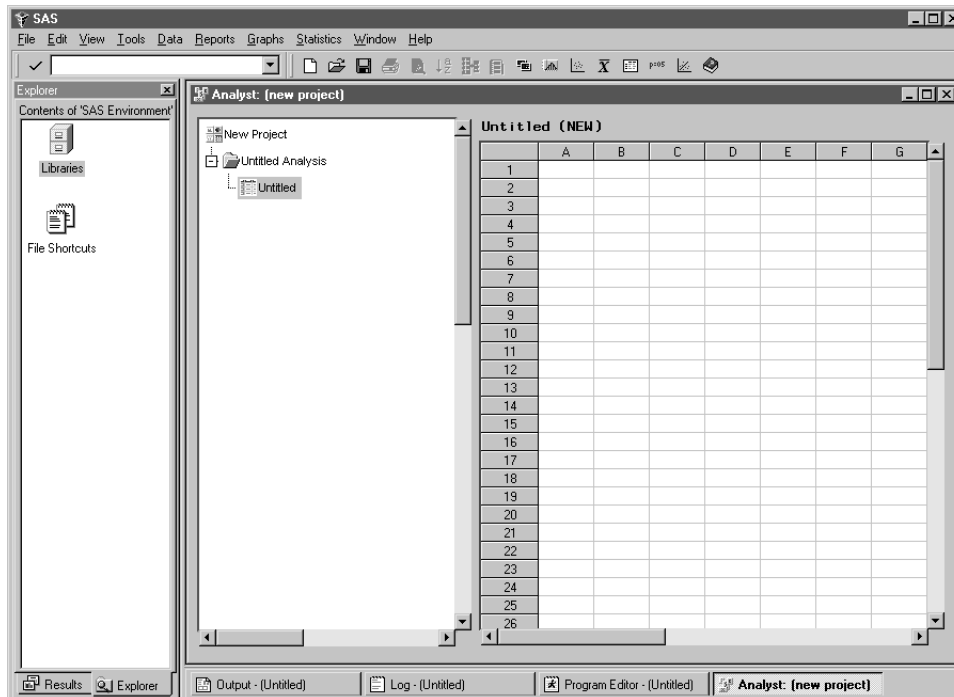Select **Solutions** → **Analysis** → **Analyst** from the main SAS menu.

Figure 1.1. The Analyst Application with Explorer and Results
Windows

You can close the Explorer and Results windows by clicking on the
close box in the upper right corner of the windows. These windows
are closed in the remaining examples in this book.

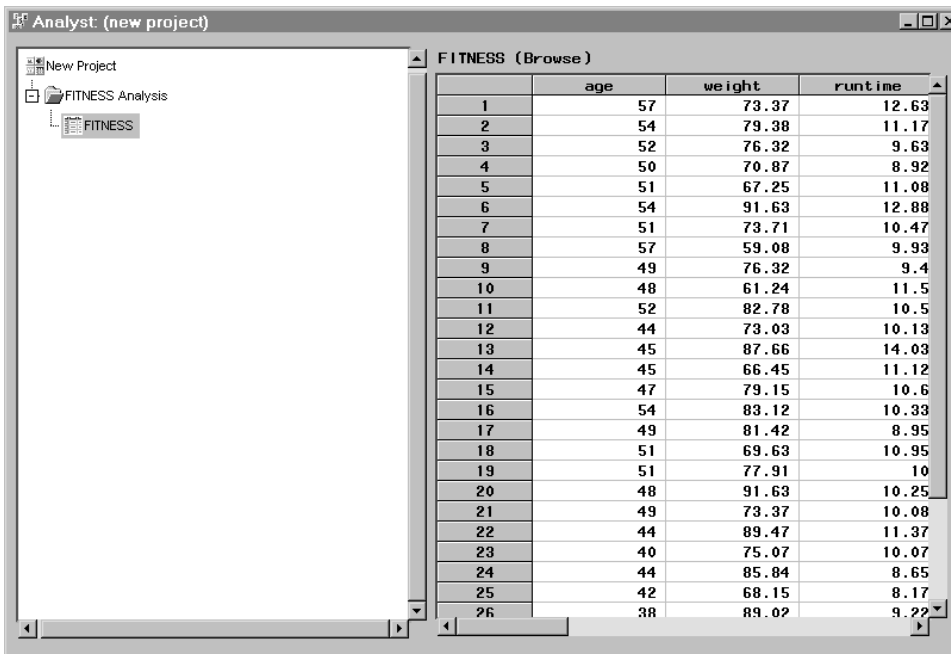To use one of the sample data sets for this example, follow these
steps:

1. Select **Tools** → **Sample Data** . . .

2. Select FITNESS.

3. Click **OK** to create the sample data set in your SASUSER
   directory.

Bring the Data into the Data Table

To bring the sample data into the data table, follow these steps:

1. Select **File** → **Open By SAS Name** . . .
2. Select **SASUSER** from the list of libraries.
3. Select FITNESS from the list of members.
4. Click **OK** to bring the sample data into the data table.

When you bring the FITNESS sample data into the data table, the
FITNESS Analysis folder, with a **FITNESS** table node representing
the data table, appears in the project tree.

Figure 1.2.   FITNESS Analysis Folder in Project Tree

Perform a Regression Analysis

To run a simple linear regression on the FITNESS data, follow these steps:

1. Select **Statistics** → **Regression** → **Simple** . . . to select the Simple Regression task.

2. In the Simple Linear Regression dialog, select oxygen from the list and click on the **Dependent** button to designate oxygen consumption as the dependent variable. Select runtime from the list and click on the **Explanatory** button to designate the amount of time to run 1.5 miles as the explanatory variable.
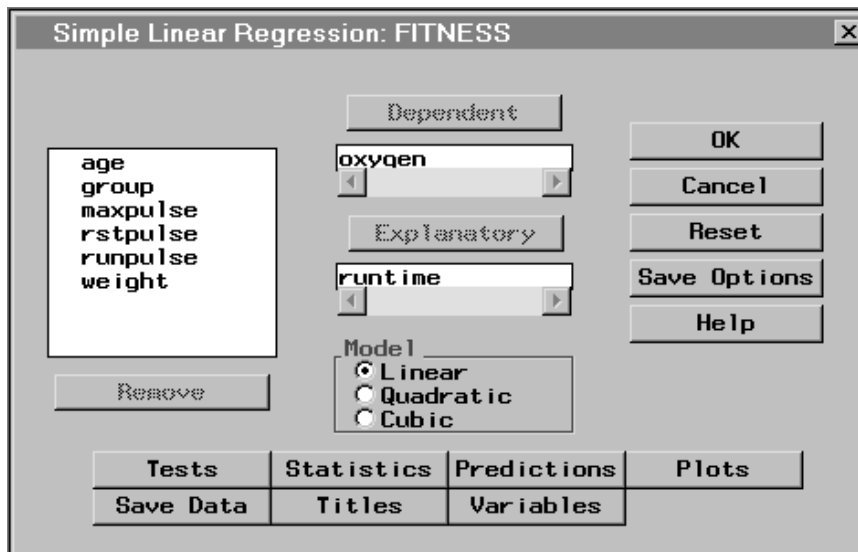
Figure 1.3.   Dependent and Explanatory Variables

3. To create a scatter plot of your results, click on the **Plots** button. In the **Predicted** tab, select **Plot observed vs independent**.
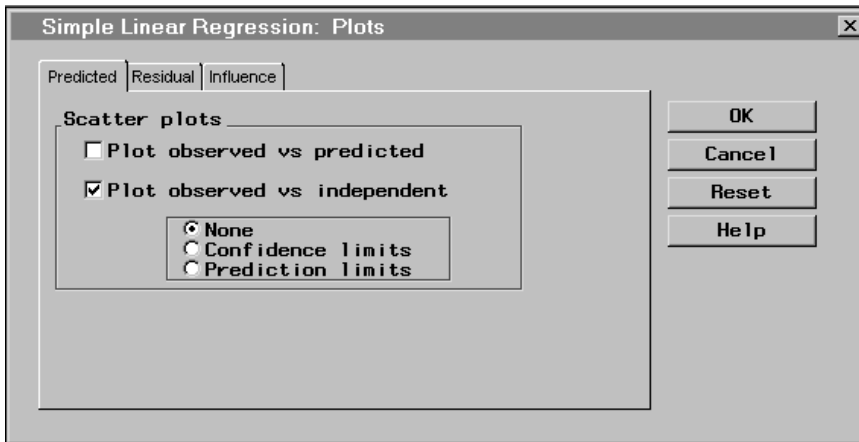
Figure 1.4. Scatter Plot

Click **OK**.

4. In the Simple Linear Regression dialog, click on the **Titles** button to specify a title for your results. In the first field of the Titles dialog, type **Speed vs. Oxygen Consumed**.

Figure 1.5. Title

Click **OK**.

5. Click **OK** in the Simple Linear Regression dialog to generate the results.

Your Results

When you click **OK** in the Simple Linear Regression dialog, the output from a simple linear regression is automatically displayed in the Analysis window. Drag the borders of the Analysis window until you can see all of the output.


```
Analysis                                                              _ □ ×
                         Speed vs. Oxygen Consumed

                             The REG Procedure
                              Model: MODEL1
                   Dependent Variable: oxygen Oxygen consumption

                             Analysis of Variance

                                   Sum of          Mean
        Source              DF     Squares        Square    F Value    Pr > F

        Model                1    632.90010     632.90010     84.01    <.0001
        Error               29    218.48144       7.53384
        Corrected Total     30    851.38154


                 Root MSE              2.74478   R-Square    0.7434
                 Dependent Mean       47.37581   Adj R-Sq    0.7345
                 Coeff Var             5.79364


                             Parameter Estimates

                                  Parameter      Standard
        Variable   Label          DF   Estimate       Error    t Value   Pr > |t|

        Intercept  Intercept       1    82.42177     3.85530     21.38    <.0001
        runtime    Min. to run 1.5 miles  1  -3.31056   0.36119     -9.17    <.0001
```
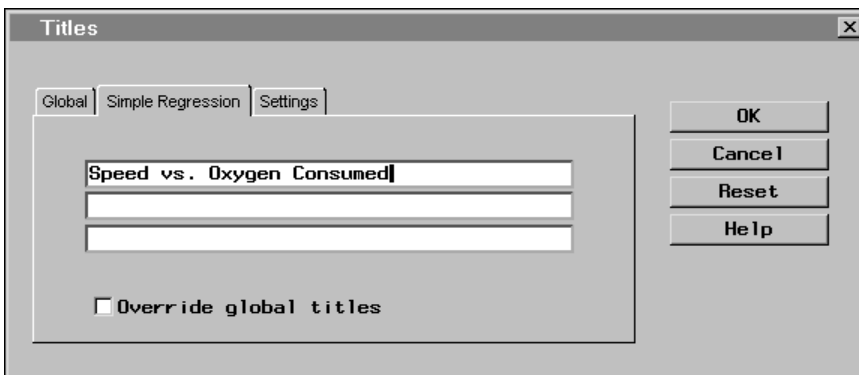
Figure 1.6.   Simple Regression Output

This model might be considered minimally adequate with an R-square value of 0.7434; the negative coefficient for runtime indicates that the linear relationship between oxygen consumption and running time is a negative one.

You can save and print your results from this window. You can also copy your output to the Program Editor window, where you can copy it to the clipboard and paste it into other applications.

Close the Analysis window to see the project tree. In addition to output, a scatter plot and the SAS code used to perform the regression and create the scatter plot are displayed as nodes in the project tree by default.
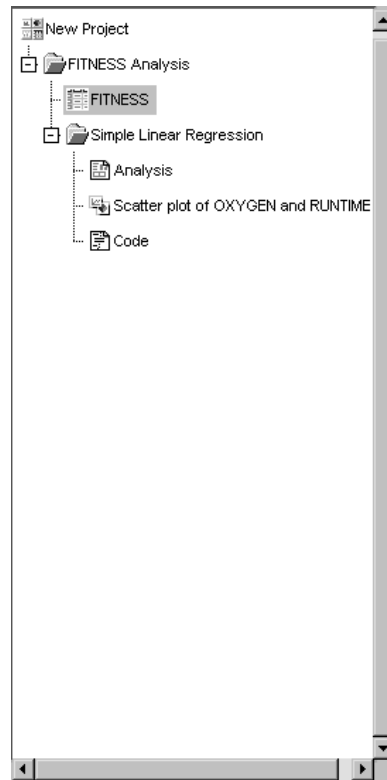

Figure 1.7. Results in Project Tree

Double-click on the **Scatter plot** node to view the scatter plot that you have created.
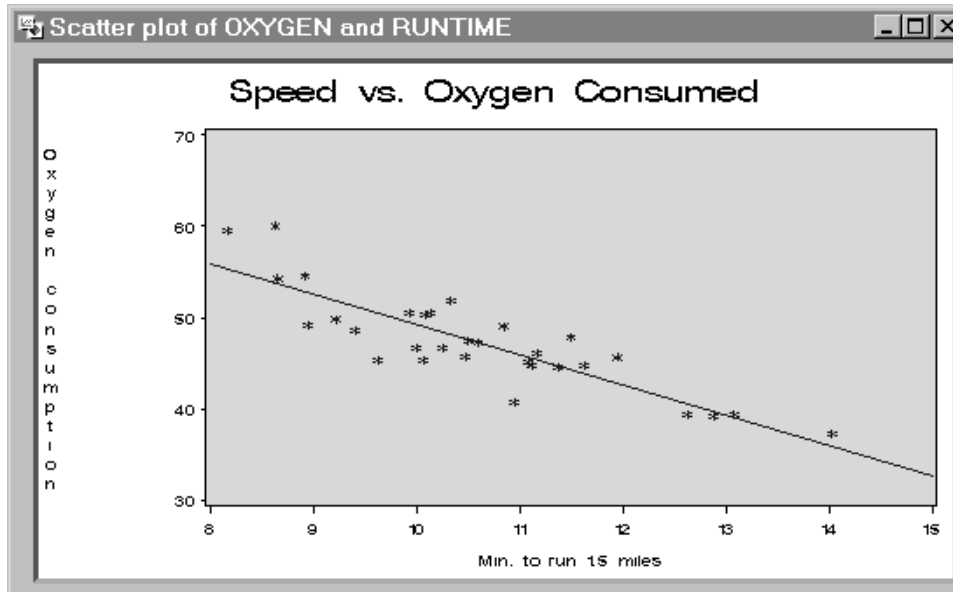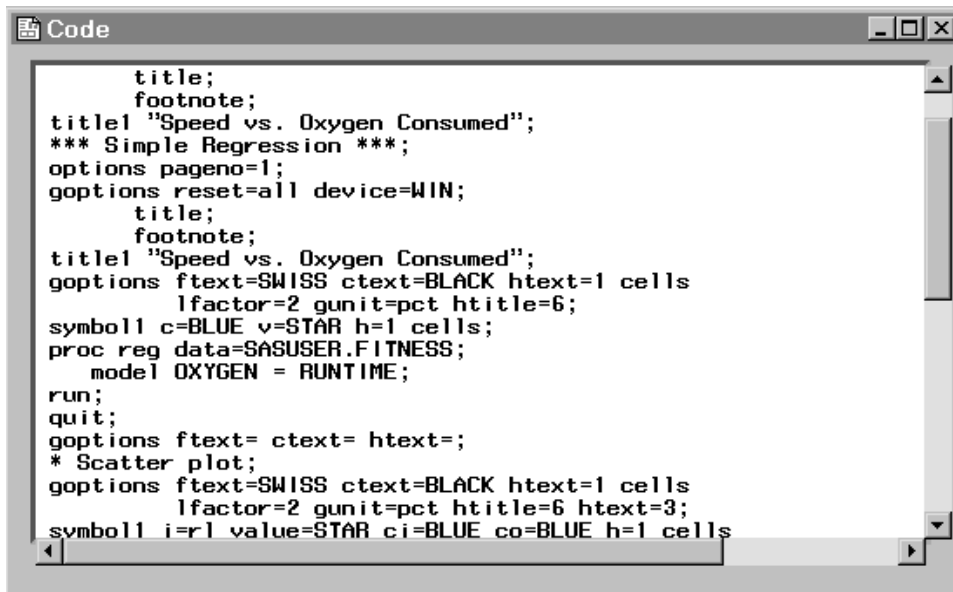
Figure 1.8. Scatter Plot

The scatter plot illustrates the results of your simple linear regression: higher oxygen consumption rates are associated with lower running times. You can change the graph to fit the window, edit the graph, or save it to a different format, such as JPEG, by selecting **File** → **Save As** . . .

Close the Scatter Plot window to view the **Code** node in the project tree.

Double-click on the **Code** node to view the SAS code that was used to perform the simple linear regression and create the scatter plot.

```
       title;
       footnote;
title1 "Speed vs. Oxygen Consumed";
*** Simple Regression ***;
options pageno=1;
goptions reset=all device=WIN;
       title;
       footnote;
title1 "Speed vs. Oxygen Consumed";
goptions ftext=SWISS ctext=BLACK htext=1 cells
         lfactor=2 gunit=pct htitle=6;
symbol1 c=BLUE v=STAR h=1 cells;
proc reg data=SASUSER.FITNESS;
   model OXYGEN = RUNTIME;
run;
quit;
goptions ftext= ctext= htext=;
* Scatter plot;
goptions ftext=SWISS ctext=BLACK htext=1 cells
         lfactor=2 gunit=pct htitle=6 htext=3;
symbol1 i=rl value=STAR ci=BLUE co=BLUE h=1 cells
```

Figure 1.9.　Code

Projects

A project is a collection of results from analyses performed on one or more data sets.

A project is displayed as a project tree that contains folders of the different data tables, reports, code, and other results that are associated with the project. Results are presented as nodes in this tree. The folder for each data set contains the results for that data set.
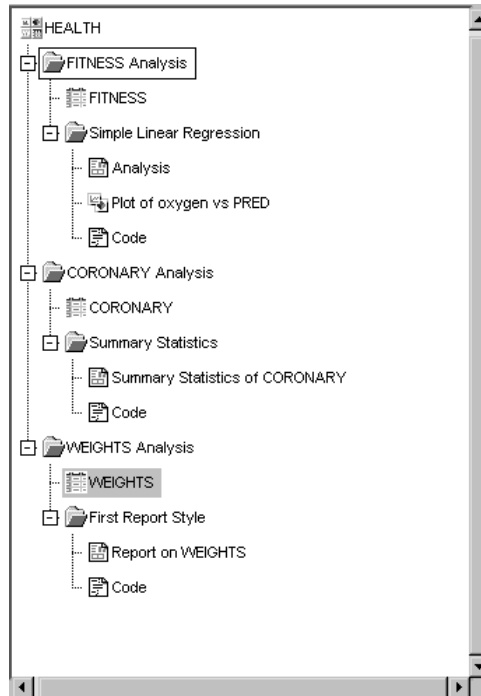
Figure 1.10. Project Tree

Project Folders

You can open and close a folder by clicking the plus (+) or minus (-) sign next to it, by double-clicking on the folder, or by selecting the folder and selecting **Expand** or **Collapse** from the pop-up menu.

Nodes

You can view a node within a folder by double-clicking on the node, or by selecting the node and selecting **View** from the pop-up menu. You can view the same node in a new window by selecting the node and selecting **View in New Window** from the pop-up menu.

From the pop-up menu, you can also delete, print, save, and, except in the case of data tables, rename the node.

If the node is a table, you can view the table in a new window by selecting **View** from the pop-up menu, or you can open the table for analysis by selecting **Open** from the pop-up menu. Also, you can select **Interactive Analysis** from the pop-up menu to invoke SAS/INSIGHT software (if it is installed) to perform interactive exploratory analyses.

Data Table

When you open a data file or SAS library member in Analyst, the data are brought into a data table where you can view and edit the data, perform numerous data transformations, and create new variables.

You can save your data by overwriting your original data source, or you can create a new data table by combining, summarizing, transposing, or taking samples of existing data tables.

Figure 1.11. Data Table

Using the Mouse

You can use the mouse to open project nodes and to select variables in Analyst.

Opening Nodes

Double-click on a node in the project tree to display its contents. Double-clicking on a data set node displays a view of the data set.

To open the data set into the data table, select the data set node and select **Open** from the pop-up menu.

Selecting and Removing Variables

In a task dialog, you can select one or more variables for analysis.

To select variables for analysis, double-click on each variable name or highlight the names and click on the analysis button. To select more than one contiguous variable, press the Shift key while clicking the mouse on the first and last variable that you want to select. All the variables between the first and last variables will be automatically selected. To select noncontiguous variables, press the Ctrl key while clicking the mouse on each variable.

To remove variables from a variable list, double-click on each variable name, or select the variables and click on the **Remove** button.

A C in front of a variable name indicates that it is a character variable.

Accessing Tasks and Help

You can access tasks and help in Analyst by using the menus, the index, and the toolbar on Windows, or the toolbox on other operating systems.

Menus

You can use the menus in Analyst to accomplish a variety of tasks. Click on the pull-down menus at the top of the window, and select items with the mouse, or click the right mouse button within a window to display a pop-up menu. Most items are available from both the pop-up and pull-down menus.

- From the **File** menu, you can access and save projects and data sets, and print reports.

- From the **Edit** menu, you can switch data between Browse and Edit mode, and add, duplicate, and delete columns and rows.

- From the **View** menu, you can move and hide columns, and view the attributes of the data table.

- From the **Tools** menu, you can set the titles for your results, create sample data sets, specify viewer and graph preferences, and assign a new SAS library.

- From the **Data** menu, you can filter, sort, summarize, concatenate, merge, transpose, and apply calculations to your data.

- From the **Reports** menu, you can create detailed and summary reports.

- From the **Graphs** menu, you can generate charts, plots, and histograms.

- From the **Statistics** menu, you can choose statistical analyses and use the index to search for tasks or statistics.

- From the **Help** menu, you can display help for Analyst and the rest of the SAS System.

Index

Use the index to access statistical and graphical tasks through commonly used terms. Select **Statistics** → **Index** . . . to display the Index of Tasks dialog. Click on a term to open a task, or type a term in the field to find it in the list.
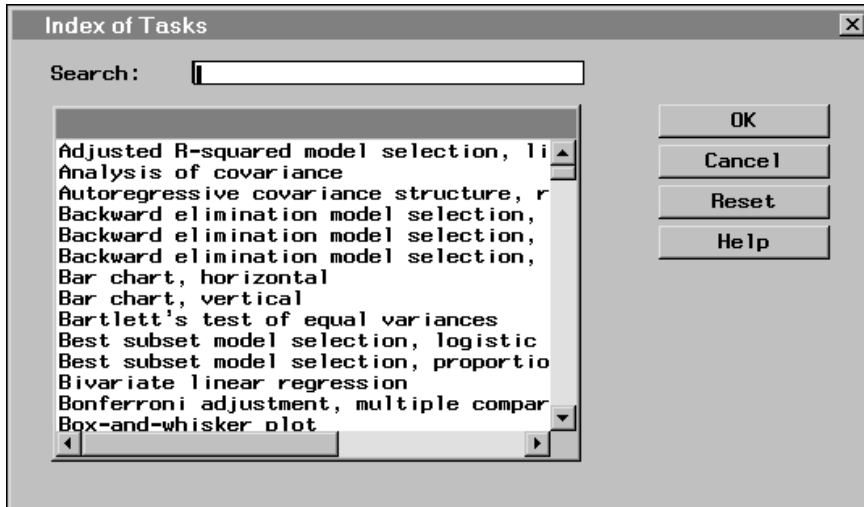
Figure 1.12. Index

For example, if you want to use the Brown-Forsythe test, click on **Brown-Forsythe test of equal variances,** click **OK**, and the One-Way ANOVA task is opened. The Brown-Forsythe test is available in the Tests dialog.

Toolbar

You can select an Analyst task from the toolbar or toolbox. Click on the icon for a task to select it. By default, the toolbar displays a range of tasks, from opening a file to performing a linear regression. You can display a description for each icon by dragging the mouse cursor over the toolbar. You can also add tasks to the toolbar. See Chapter 17, "Details," for more information.

Figure 1.13. Toolbar

Getting Help

You can get help in Analyst in three ways:

- the **Help** menu

- the **Help** button on a dialog

- the **Help** icon in the toolbar

Help Menu

When Analyst is open, select **Help** → **Using This Window** to display help on Analyst. From the main window, **Using This Window** displays the table of contents for Analyst help. From other windows, **Using This Window** displays the help for that particular window. If you are on the Windows operating system, you can go to another help topic through the table of contents or the index.
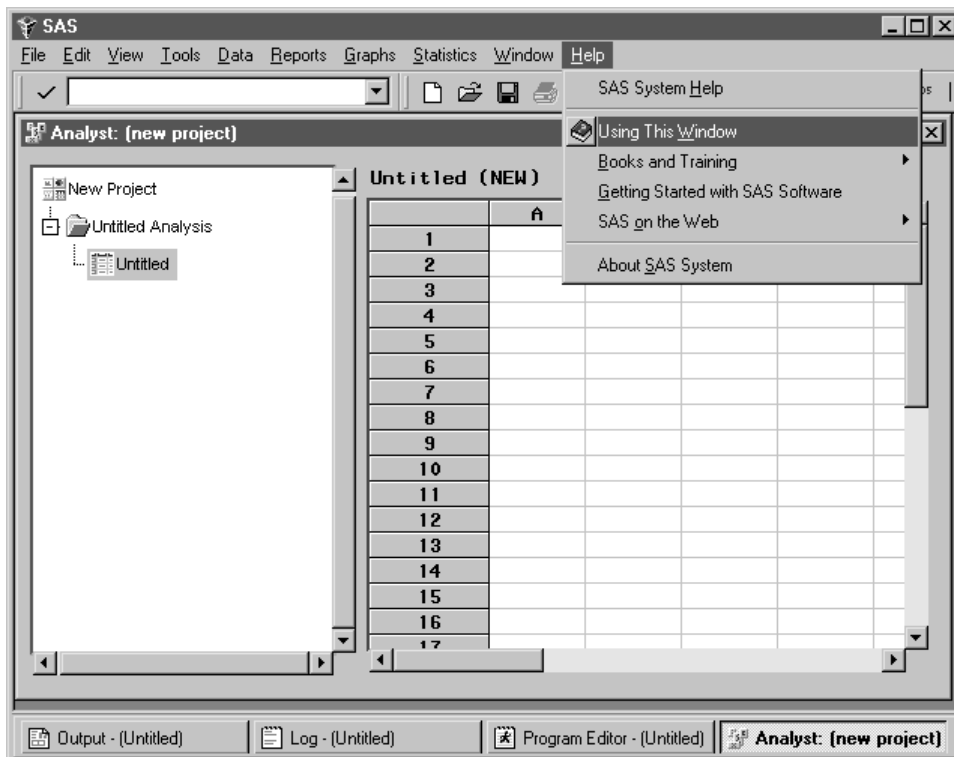
Figure 1.14. Help Menu

Help Button in Dialogs

Each dialog in Analyst has a help button that you can click on to display the help for that task.

Help Icon on the Toolbar

Click on the **Help** icon on the toolbar to display the help table of contents for Analyst.

Graphs

Analyst enables you to create several different kinds of graphs:

- bar charts
- pie charts
- histograms
- box plots
- probability plots
- scatter plots
- contour plots
- surface plots

Use the **Graphs** menu to select the type of graph you want to create.

You can apply settings to all graphs that you produce with Analyst by selecting **Graph Settings** . . . from the **Tools** menu.

Reports

In Analyst, you can create a simple listing of your data or a summary report.

Listing Reports

Select **Reports** → **List Data** . . . to produce a listing of your data.

Summary Reports

You can create a summary report in any one of five table styles. Select **Reports** → **Tables** . . . and select one of the styles that are illustrated.
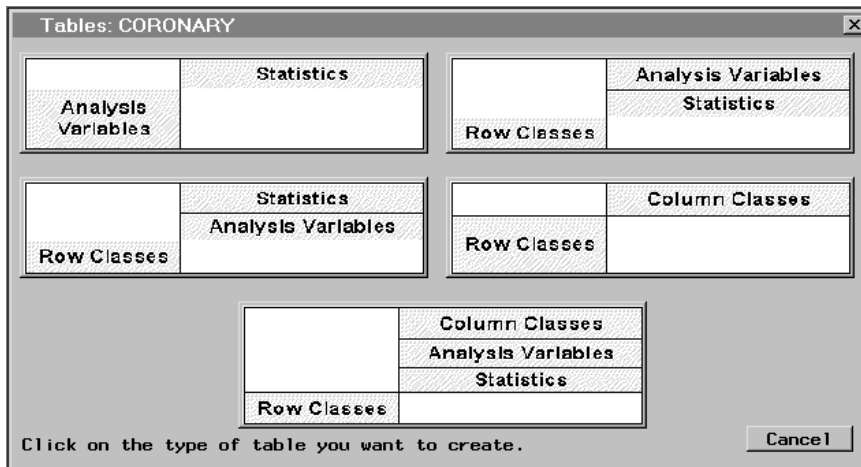
Figure 1.15. Table Styles

Statistical Tasks

Analyst contains a wide range of statistical tasks. You can compute descriptive statistics, perform simple hypothesis tests, and fit models with analysis of variance and regression analysis. There are also tasks for survival analysis, mixed models, repeated measures analysis, and multivariate techniques. Analyst also provides basic sample size and power computations. Graphics are included in most analytical tasks, and you can request many types of graphs directly from the **Graphs** menu.

Descriptive

The descriptive statistical tasks that you can perform on your data include

- summary statistics
- distributions
- correlations
- frequency counts

Table Analysis

In the Table Analysis task, you can create and analyze 2–way to n-way frequency tables.

Hypothesis Tests

The hypothesis tests that you can perform on your data include

- one-sample Z-test for a mean
- one-sample t-test for a mean
- one-sample test for a proportion
- one-sample test for a variance
- two-sample t-test for means
- two-sample paired t-test for means
- two-sample test for proportions
- two-sample test for variances

ANOVA

You can perform one-way and factorial analysis of variance (ANOVA). The one-way task includes the one-way nonparametric ANOVA. You can also fit the general linear model, perform repeated measurements ANOVA, and fit basic mixed models.

Regression

The Linear Regression task provides linear and multiple linear regression analysis. The Simple Linear Regression task predicts a dependent variable from a single independent quantitative variable.

The Logistic Regression task investigates the relationship between a binary outcome (such as success and failure) or an ordinal outcome (such as mild, moderate, and severe) and a set of explanatory variables.

Multivariate

The Principal Components task computes principal components from a set of variables.

The Canonical Correlation task describes the relationship between two sets of variables by finding a small number of linear combinations from each set of variables that have the highest possible between-set correlations.

Survival

The Life Tables task computes nonparametric estimates of the survival distribution of data that may be right censored due to withdrawals or study termination. This task computes rank tests and a likelihood ratio test for testing homogeneity of survival functions across strata.

The Proportional Hazards task performs regression analysis of survival data based on the Cox proportional hazards model.

Sample Size

The Sample Size tasks enable you to determine the power of a test, given the sample size, or the sample size required to obtain a specified power. These calculations can be made for a variety of situations, including t-tests, confidence intervals, tests of equivalence, and one-way ANOVA. These are prospective power and sample size computations; retrospective power computations are provided for some of the analytical tasks.