Chapter 2
Introduction

Chapter Table of Contents

Chapter 2
Introduction

Overview of SAS/STAT Software

SAS/STAT software, a component of the SAS System, provides comprehensive statistical tools for a wide range of statistical analyses, including analysis of variance, regression, categorical data analysis, multivariate analysis, survival analysis, psychometric analysis, cluster analysis, and nonparametric analysis. A few examples include mixed models, generalized linear models, correspondence analysis, and structural equations. The software is constantly being updated to reflect new methodology.

In addition to 54 procedures for statistical analysis, SAS/STAT software also includes the Market Research Application (MRA), a point-and-click interface to commonly used techniques in market research. Also, the Analyst Application in the SAS System provides convenient access to some of the more commonly used statistical analyses in SAS/STAT software including analysis of variance, regression, logistic regression, mixed models, survival analysis, and some multivariate techniques.

About This Book

Since SAS/STAT software is a part of the SAS System, this book assumes that you are familiar with base SAS software and with the books *SAS Language Reference: Dictionary*, *SAS Language Reference: Concepts,* and the *SAS Procedures Guide*. It also assumes that you are familiar with basic SAS System concepts such as creating SAS data sets with the DATA step and manipulating SAS data sets with the procedures in base SAS software (for example, the PRINT and SORT procedures).

Chapter Organization

This book is organized as follows.

Chapter 1, "Changes and Enhancements to SAS/STAT Software in V7 and V8," contains information about the updates that are included in this release (Version 8). Chapter 2, this chapter, provides an overview of SAS/STAT software and summarizes related information, products, and services. The next ten chapters provide some introduction to the broad areas covered by SAS/STAT software. Subsequent chapters describe the SAS procedures that make up SAS/STAT software. These chapters appear in alphabetical order by procedure name.

The chapters documenting the SAS/STAT procedures are organized as follows:

- The *Overview* section provides a brief description of the analysis provided by the procedure.

- The *Getting Started* section provides a quick introduction to the procedure through a simple example.

- The *Syntax* section describes the SAS statements and options that control the procedure.

- The *Details* section discusses methodology and miscellaneous details.

- The *Examples* section contains examples using the procedure.

- The *References* section contains references for the methodology and examples for the procedure.

Following the chapters on the SAS/STAT procedures, Appendix A, "Special SAS Data Sets," documents the special SAS data sets associated with SAS/STAT procedures.

Typographical Conventions

This book uses several type styles for presenting information. The following list explains the meaning of the typographical conventions used in this book:

roman | is the standard type style used for most text.

UPPERCASE ROMAN | is used for SAS statements, options, and other SAS language elements when they appear in the text. However, you can enter these elements in your own SAS programs in lowercase, uppercase, or a mixture of the two.

UPPERCASE BOLD | is used in the "Syntax" sections' initial lists of SAS statements and options.

oblique | is used for user-supplied values for options in the syntax definitions. In the text, these values are written in *italic*.

helvetica | is used for the names of variables and data sets when they appear in the text.

bold | is used to refer to matrices and vectors.

italic | is used for terms that are defined in the text, for emphasis, and for references to publications.

monospace | is used for example code. In most cases, this book uses lowercase type for SAS code.

Options Used in Examples

Output of Examples

Most of the output shown in this book is produced with the following SAS System options:

```
options linesize=80 pagesize=200 nonumber nodate;
```

The template STATDOC.TPL is used to create the HTML output that appears in the online (CD) version. A style template controls stylistic HTML elements such as colors, fonts, and presentation attributes. The style template is specified in the ODS HTML statement as follows:

```
ODS HTML style=statdoc;
```

If you run the examples, you may get slightly different output. This is a function of the SAS System options used and the precision used by your computer for floating-point calculations.

Graphics Options

The examples that contain graphical output are created with a specific set of options and symbol statements. The code you see in the examples creates the color graphics that appear in the online (CD) version of this book. A slightly different set of options and statements is used to create the black and white graphics that appear in the printed version of the book.

If you run the examples, you may get slightly different results. This may occur because not all graphic options for color devices translate directly to black and white output formats. For complete information on SAS/GRAPH software and graphics options, refer to *SAS/GRAPH Software: Reference*.

The following GOPTIONS statement is used to create the online (color) version of the graphic output.

```
filename GSASFILE  '<file-specification>';

goptions gsfname=GSASFILE   gsfmode =replace
         fileonly
         transparency       dev     = gif
         ftext   = swiss    lfactor = 1
         htext   = 4.0pct   htitle  = 4.5pct
         hsize   = 5.625in  vsize   = 3.5in
         noborder           cback   = white
         horigin = 0in      vorigin = 0in ;
```

The following GOPTIONS statement is used to create the black and white version of the graphic output, which appears in the printed version of the manual.

```
filename GSASFILE   '<file-specification>';

goptions gsfname=GSASFILE   gsfmode =replace
         gaccess = sasgaedt fileonly
         dev     = pslepsf
         ftext   = swiss    lfactor = 1
         htext   = 3.0pct   htitle  = 3.5pct
         hsize   = 5.625in  vsize   = 3.5in
         border             cback   = white
         horigin = 0in      vorigin = 0in ;
```

In most of the online examples, the plot symbols are specified as follows:

```
symbol1 value=dot color=white height=3.5pct;
```

The SYMBOL*n* statements used in online examples order the symbol colors as follows: white, yellow, cyan, green, orange, blue, and black.

In the examples appearing in the printed manual, symbol statements specify COLOR=BLACK and order the plot symbols as follows: dot, square, triangle, circle, plus, x, diamond, and star.

The %PLOTIT Macro

Examples that use the %PLOTIT macro are generated by defining a special macro variable to specify graphics options. See Appendix B, "Using the %PLOTIT Macro," for details on the options specified in these examples.

Where to Turn for More Information

This section describes other sources of information about SAS/STAT software.

Accessing the SAS/STAT Sample Library

The SAS/STAT sample library includes many examples that illustrate the use of SAS/STAT software, including the examples used in this documentation. To access these sample programs, select **Help** from the pmenu and select **SAS System Help**. From the Main Contents list, choose **Sample SAS Programs and Applications**.

Online Help System

You can access online help information about SAS/STAT software in two ways. You can select **SAS System Help** from the **Help** pmenu and then select **SAS/STAT Software** from the list of available topics. Or, you can bring up a command line and issue the command **help STAT** to bring up an index to the statistical procedures, or issue the command **help CATMOD** (or another procedure name) to bring up the help for that particular procedure. Note that the online help includes syntax and some essential overview and detail material.

SAS Institute Technical Support Services

As with all SAS Institute products, the SAS Institute Technical Support staff is available to respond to problems and answer technical questions regarding the use of SAS/STAT software.

Related SAS Software

Many features not found in SAS/STAT software are available in other parts of the SAS System. If you don't find something you need in SAS/STAT software, try looking for the feature in the following SAS software products.

Base SAS Software

The features provided by SAS/STAT software are in addition to the features provided by base SAS software. Many data management and reporting capabilities you will need are part of base SAS software. Refer to *SAS Language Reference: Concepts*, *SAS Language Reference: Dictionary*, and the *SAS Procedures Guide* for documentation of base SAS software.

SAS DATA Step

The DATA step is your primary tool for reading and processing data in the SAS System. The DATA step provides a powerful general purpose programming language that enables you to perform all kinds of data processing tasks. The DATA step is documented in *SAS Language Reference: Concepts*.

Base SAS Procedures

Base SAS software includes many useful SAS procedures. Base SAS procedures are documented in the *SAS Procedures Guide*. The following is a list of base SAS procedures you may find useful:

CORR	compute correlations
RANK	compute rankings or order statistics
STANDARD	standardize variables to a fixed mean and variance
MEANS	compute descriptive statistics and summarizing or collapsing data over cross sections

| TABULATE | print descriptive statistics in tabular format |
| UNIVARIATE | compute descriptive statistics |

SAS/ETS Software

SAS/ETS software provides SAS procedures for econometrics and time series analysis. It includes capabilities for forecasting, systems modeling and simulation, seasonal adjustment, and financial analysis and reporting. In addition, SAS/ETS software includes an interactive time series forecasting system.

SAS/GRAPH Software

SAS/GRAPH software includes procedures that create two- and three-dimensional high resolution color graphics plots and charts. You can generate output that graphs the relationship of data values to one another, enhance existing graphs, or simply create graphics output that is not tied to data.

SAS/IML Software

SAS/IML software gives you access to a powerful and flexible programming language (Interactive Matrix Language) in a dynamic, interactive environment. The fundamental object of the language is a data matrix. You can use SAS/IML software interactively (at the statement level) to see results immediately, or you can store statements in a module and execute them later. The programming is dynamic because necessary activities such as memory allocation and dimensioning of matrices are done automatically. SAS/IML software is of interest to users of SAS/STAT software because it enables you to program your methods in the SAS System.

SAS/INSIGHT Software

SAS/INSIGHT software is a highly interactive tool for data analysis. You can explore data through a variety of interactive graphs including bar charts, scatter plots, box plots, and three-dimensional rotating plots. You can examine distributions and perform parametric and nonparametric regression, analyze general linear models and generalized linear models, examine correlation matrixes, and perform principal component analyses. Any changes you make to your data show immediately in all graphs and analyses. You can also configure SAS/INSIGHT software to produce graphs and analyses tailored to the way you work.

SAS/INSIGHT software may be of interest to users of SAS/STAT software for interactive graphical viewing of data, editing data, exploratory data analysis, and checking distributional assumptions.

SAS/OR Software

AS/OR software provides SAS procedures for operations research and project planning and includes a point-and-click interface to project management. Its capabilities include the following:

- solving transportation problems
- linear, integer, and mixed-integer programming
- nonlinear programming
- scheduling projects
- plotting Gantt charts
- drawing network diagrams
- solving optimal assignment problems
- network flow programming

SAS/OR software may be of interest to users of SAS/STAT software for its mathematical programming features. In particular, the NLP procedure in SAS/OR software solves nonlinear programming problems, and it can be used for constrained and unconstrained maximization of user-defined likelihood functions.

SAS/QC Software

SAS/QC software provides a variety of procedures for statistical quality control and quality improvement. SAS/QC software includes procedures for

- Shewhart control charts
- cumulative sum control charts
- moving average control charts
- process capability analysis
- Ishikawa diagrams
- Pareto charts
- experimental design

SAS/QC software also includes the ADX interface for experimental design.