SAS Essentials is the most commonly used statistical data analysis tool for scientific research worldwide. SAS enables researchers to perform data entry, retrieval, management, and mining; report writing and graphics; statistical analysis; and other essential applications.

SAS Essentials introduces a step-by-step approach to mastering SAS software for statistical data analysis. It’s also a valuable reference tool for any researcher currently using SAS. Designed for those new to SAS and filled with illustrative examples, the book shows how to read, write and import data; prepare data for analysis; use SAS procedures; evaluate quantitative data; analyze counts and crosstabulation tables; and compare means using the t-test. The book also provides instruction and examples on analysis of variance, correlation and regression, nonparametric analysis, logistic regression, creating graphs, controlling outputs using ODS, as well as advanced topics in SAS programming.

Written by Alan C. Elliott and Wayne A. Woodward—two experts in the field—SAS Essentials uses examples based on SAS 9.2 for Windows, although most of the examples will work with any Windows version of SAS.

Data sets used in the book’s examples are available on an accompanying web site.

The Authors
Alan C. Elliott is a biostatistician and faculty member in the Department of Clinical Sciences at the University of Texas Southwestern Medical Center at Dallas. A prolific writer, he is the author or coauthor of Directory of Microcomputer Statistical Software, Microcomputing with Applications, Getting Started in Internet Auctions, Statistical Analysis Quick Reference Guidebook, and other books.

Wayne A. Woodward is a professor of statistics and chair of the Department of Statistical Science at Southern Methodist University. He is a fellow of the American Statistical Association and was the 2004 recipient of the Don Owen award for excellence in research, statistical consulting, and service to the statistical community. During the last 30 years he has served as statistical consultant to a wide variety of clients in the scientific community.
SAS ESSENTIALS
A Guide to Mastering SAS for Research

ALAN C. ELLIOTT
WAYNE A. WOODWARD

JOSSEY-BASS
A Wiley Imprint
www.josseybass.com
<table>
<thead>
<tr>
<th>CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figures and Tables</td>
</tr>
<tr>
<td>Preface</td>
</tr>
<tr>
<td>Acknowledgments</td>
</tr>
<tr>
<td>The Authors</td>
</tr>
</tbody>
</table>

1 GETTING STARTED  
- Using SAS in a Windows Environment  
- Your First SAS Analysis  
- How SAS Works  
- Tips and Tricks for Running SAS Programs  
- Summary  
- Exercises  

2 GETTING DATA INTO SAS  
- Using the SAS Data Sets  
- Understanding SAS Data Set Structure  
- Rules for SAS Variable Names  
- Understanding Three SAS Variable Types  
- Methods of Reading Data into SAS  
- Going Deeper: More Techniques for Entering Data  
- Summary  
- Exercises  

3 READING, WRITING, AND IMPORTING DATA  
- Working with SAS Libraries and Permanent Data Sets  
- Reading and Creating Permanent SAS Data Sets Using the Windows File Name Technique  
- Reading Data from Permanent SAS Data Sets  
- Reading and Creating Permanent SAS Data Sets Using a SAS Library  
- Creating a SAS Library Using a Dialog Box  
Contents

Importing Data from Another Program 59
Going Deeper: More Ways to Manage Data 64
Going Deeper: Importing Microsoft Excel Data Using SAS Code 65
Discovering the Contents of a SAS Data Set 67
Summary 69
Exercises 70

4 PREPARING DATA FOR ANALYSIS 73
Labeling Variables with Explanatory Names 74
Creating New Variables 77
Using DROP and KEEP to Select Variables 89
Subsetting Data Sets 90
Using the SET Statement to Read an Existing Data Set 92
Using PROC SORT 94
Appending and Merging Data Sets 97
Going Deeper: Using PROC FORMAT 101
Summary 106
Exercises 106

5 PREPARING TO USE SAS PROCEDURES 109
Understanding SAS Support Statements 111
Understanding PROC Statement Syntax 116
Using the ID Statement in a SAS Procedure 120
Using the LABEL Statement in a SAS Procedure 121
Using the WHERE Statement in a SAS Procedure 123
Using PROC PRINT 123
Going Deeper: Introducing the SAS Output Delivery System (ODS) 126
Summary 128
Exercises 128

6 EVALUATING QUANTITATIVE DATA 130
Using PROC MEANS 131
Using PROC UNIVARIATE 138
Going Deeper: Advanced PROC UNIVARIATE Options 150
Summary 155
Exercises 156
## Contents

### 7 ANALYZING COUNTS AND TABLES
- Using PROC FREQ
- Analyzing One-Way Frequency Tables
- Creating One-Way Frequency Tables from Summarized Data
- Analyzing Two-Way Tables
- Going Deeper: Calculating Relative Risk Measures
- Going Deeper: Inter-Rater Reliability (Kappa)
- Summary
- Exercises

### 8 COMPARING MEANS USING T-TESTS
- Performing a One-Sample t-test
- Performing a Two-Sample t-test
- Performing a Paired t-test
- Summary
- Exercises

### 9 ANALYSIS OF VARIANCE
- Comparing Three or More Means Using One-Way Analysis of Variance
- Comparing Three or More Repeated Measures
- Going Deeper: Graphing Mean Comparisons
- Summary
- Exercises

### 10 CORRELATION AND REGRESSION
- Correlation Analysis Using PROC CORR
- Simple Linear Regression
- Multiple Linear Regression Using PROC REG
- Going Deeper: Calculating Predictions
- Going Deeper: Residual Analysis
- Summary
- Exercises
11 NONPARAMETRIC ANALYSIS 245
   Comparing Two Independent Samples Using NPAR1WAY 246
   Comparing k Independent Samples (Kruskal-Wallis) 249
   Comparing Two Dependent (Paired) Samples 251
   Comparing k Dependent Samples (Friedman’s Test) 253
   Going Deeper: Nonparametric Multiple Comparisons 255
   Summary 258
   Exercises 258

12 LOGISTIC REGRESSION 260
   Logistic Analysis Basics 261
   Performing a Logistic Analysis using PROC LOGISTIC 262
   Using Simple Logistic Analysis 264
   Multiple Binary Logistic Analysis 268
   Going Deeper: Assessing a Model’s Fit and Predictive Ability 271
   Summary 275
   Exercises 276

13 ANALYSIS OF VARIANCE, PART II 277
   Analysis of Covariance 278
   Going Deeper: Two-Factor ANOVA Using PROC MIXED 293
   Going Deeper: Repeated Measures with a Grouping Factor 296
   Summary 303
   Exercises 303

14 CREATING GRAPHS 305
   Creating Scatterplots and Line Graphs Using GPLOT 306
   Creating Bar Charts and Pie Charts 315
   Creating Stacked Bar Charts 319
   Creating Mean Bars Using GCHAR 321
   Creating Boxplots 323
   Going Deeper: ODS Graphics 326
   Summary 330
   Exercises 330
CONTROLLING OUTPUT USING ODS

Specifying the ODS Output Format and Destination
Specifying ODS Output Style
Using ODS to Select Specific Output Tables for Procedures
Going Deeper: Enhancing Graphics Using ODS and Creating Hyperlinks
Going Deeper: Capturing Information from ODS Tables
Extended ODS Features
Summary
Exercises

ADVANCED SAS PROGRAMMING TOPICS

Reading and Writing Data Using DDE
Using the RETAIN Statement
Arrays and DO Loops
Transposing Data Sets
Using SAS Macros
Summary
Exercises

APPENDIX A

SAS GRAPH OPTIONS REFERENCE

Using SAS Fonts
Specifying SAS Color Choices
Specifying Patterns for PROCS GPLOT and PROC UNIVARIATE
Bar and Block Patterns for Bar Charts, Pie Charts, and Other Graphics
SAS Line Styles
Using SAS Plotting Symbols

APPENDIX B

SAS FUNCTION REFERENCE

Using SAS Functions
Arithmetic/Mathematical Functions
Contents

Trignometric Functions 387
Date and Time Functions 388
Character Functions 391
Truncation Functions 394
Special Use Functions 394
Financial Functions 395
Working with Previous Observations 396
Miscellaneous Functions 396

APPENDIX C

CHOOSING A SAS PROCEDURE 397
Descriptive Statistics 398
Comparison Tests 399
Relational Analyses (Correlation and Regression) 401

APPENDIX D

QUICK REFERENCE 403
REFERENCES 409
INDEX 411
FIGURES AND TABLES

FIGURES

1.1 Initial SAS screen 3
1.2 The FIRST.SAS file opened 6
1.3 Output for FIRST.SAS 7
1.4 How SAS works 10
1.5 Help file display 16
3.1 Reading and storing data in SAS 52
3.2 New Library dialog box 56
3.3 The SAS Library icon 57
3.4 SAS Viewtable 58
3.5 Excel data ready for import 60
3.6 Initial screen of the SAS Import Wizard 60
3.7 Selecting an Excel file name 61
3.8 Selecting the library and member name 62
4.1 Creating a new data set using the SET statement 93
4.2 Example output from merged data sets 101
5.1 Using SAS 110
5.2 Example output 110
6.1 Plots to assess normality from PROC UNIVARIATE 148
6.2 Histogram with normal curve superimposed 150
6.3 Multiple histograms from PROC UNIVARIATE 152
6.4 PROC UNIVARIATE histogram output using two groups 155
6.5 Multiple Histograms from the CAR dataset 158
7.1 Example data records from the RASH dataset 175
8.1 Plots produced by PROC TTEST 191
8.2 Plots produced by PROC TTEST 197
8.3 Plots produced by PROC TTEST using the PAIRED option 199
9.1 Side-by-side boxplots produced by ODS GRAPHICS 214
Figures and Tables

9.2  Side-by-side dot plots produced by GPLOT 215
10.1  Matrix of scatterplots 221
10.2  Example simple linear regression using PROC REG 228
10.3  Scatterplot and regression line for final model 237
10.4  ODS GRAPHICS output for PROC REG 240
10.5  Prediction limits 241
10.6  Washington, D.C., assaults, 1978–93 242
11.1  Side-by-side boxplots for the weight-gain data 258
12.1  A logistic plot from a simple logistic model 268
12.2  ROC curve 275
13.1  Regression lines for ANCOVA data 283
13.2  Parallel lines for ANCOVA data, for second model assuming parallel lines 283
13.3  Graphical representation of Scheffe results 284
13.4  Example of an interaction effect 285
13.5  ODS Graphics interaction plot for the two-factor ANOVA example (SAS version 9.2) 290
13.6  Interaction plots for two-way ANOVA 292
13.7  Plots for repeated measures analysis 302
14.1  Default scatterplot using PROC GPLOT 307
14.2  PROC GPLOT with dots and line specified 310
14.3  PROC GPLOT with dots and lines specified 313
14.4  PROC GPLOT showing error bars 314
14.5  Horizontal bar chart using PROC GCHART 316
14.6  PROC GCHART showing groups bars 318
14.7  Stacked bar chart 320
14.8  Means with error bars 323
14.9  Boxplots showing highway MPG by cylinders 325
14.10  Boxplots with enhancements 326
14.11  Graphics output from ODS 328
14.12  ODS graphics output from PROC CORR 330
15.1  ODS TRACE output 339
15.2  Drill-down plot, first run 341
TABLES

16.1 Excel data to import into SAS via the clipboard 353
16.2 Data written to an Excel spreadsheet from SAS 357
16.3 Data from the COMPLICATIONS data set 367
A.1 Standard SAS fonts 378
A.2 Example crosshatch pattern 380
A.3 Example fill patterns 382
A.4 Line styles 382
A.5 Select line style 383
A.6 Standard SAS symbols 384

TABLES

1.1 SAS Function keys 13
2.1 Overview of the SAS DATA step 20
2.2 SAS Data set (table) structure 22
2.3 A sample SAS data set 26
2.4 Example SAS FORMATs and INFORMATs 36
3.1 SAS DBMS data sources 66
4.1 SAS sorting sequence 95
4.2 SAS output showing sort ascending and descending 97
6.1 Output for PROC MEANS 135
6.2 PROC MEANS output using the BY option 137
6.3 PROC MEANS output using the CLASS option 137
6.4 Moments output for PROC UNIVARIATE 143
6.5 Basic statistical measures from PROC UNIVARIATE 144
6.6 Tests for location from PROC UNIVARIATE 144
6.7 Quantiles from PROC UNIVARIATE 145
6.8 Extreme values output from PROC UNIVARIATE 146
6.9 Test for normality from PROC UNIVARIATE 148
7.1 Sample table statements 161
7.2 Simple output for PROC FREQ 164
7.3 PROC FREQ output using the ORDER FREQ option 165
7.4 Output from PROC FREQ using the ORDER=FORMATTED option 166
7.5 Proc FREQ output for summarized counts 168