Contents at a Glance

About the Authors xix

Acknowledgments xxi

Preface xxiii

■Part 1: Basics of SAS Programming for Analytics 1

■Chapter 1: Introduction to Business Analytics and Data Analysis Tools 3

■Chapter 2: SAS Introduction 29

■Chapter 3: Data Handling Using SAS 55

■Chapter 4: Important SAS Functions and Procs 95

■Part 2: Using SAS for Business Analytics 145

■Chapter 5: Introduction to Statistical Analysis 147

■Chapter 6: Basic Descriptive Statistics and Reporting in SAS 165

■Chapter 7: Data Exploration, Validation, and Data Sanitization 197

■Chapter 8: Testing of Hypothesis 261

■Chapter 9: Correlation and Linear Regression 295

■Chapter 10: Multiple Regression Analysis 351

■Chapter 11: Logistic Regression 401

■Chapter 12: Time-Series Analysis and Forecasting 441

■Chapter 13: Introducing Big Data Analytics 509

Index 541

Contents

About the Authors xix

Acknowledgments xxi

Preface xxiii

■Part 1: Basics of SAS Programming for Analytics 1

■Chapter 1: Introduction to Business Analytics and Data Analysis Tools 3

Business Analytics, the Science of Data-Driven Decision Making 3

Business Analytics Defined 3

Is Advanced Analytics the Solution for You? 5

Simulation, Modeling, and Optimization 6

Data Warehousing and Data Mining 7

What Can Be Discovered Using Data Mining? 7

Business Intelligence, Reporting, and Business Analytics 8

Analytics Techniques Used in the Industry 8

Regression Modeling and Analysis 8

Time Series Forecasting 10

Conjoint Analysis 11

Cluster Analysis 11

Segmentation 13

Principal Components and Factor Analysis 13

Correspondence Analysis 13

Survival Analytics 13

Some Practical Applications of Business Analytics 14
Customer Analytics 14
Operational Analytics 14
Social Media Analytics 14
Data Used in Analytics 15
Big Data vs. Conventional Business Analytics 15
Introduction to Big Data 15
Introduction to Data Analysis Tools 20
Main Parts of SAS, SPSS, and R 22
Selection of Analytics Tools 27
The Background Required for a Successful Career in Business Analytics 27
Skills Required for a Business Analytics Professional 27
Conclusion 28

■Chapter 2: SAS Introduction 29

Starting SAS in Windows 29
The SAS Opening Screen 31
The Five Main Windows 31
Editor Window 32
Log Window 34
Output Window 35
Explorer Window 40
Results Window 41
Important Menu Options and Icons 42
View Options 44
Run Menu 44
Solutions Menu 45
Shortcut Icons 45
Writing and Executing a SAS Program 46
Comments in the Code 47

Your First SAS Program 48
Debugging SAS Code Using a Log File 50
Example for Warnings in Log File 52
Tips for Writing, Reading the Log File, and Debugging 53
Saving SAS Files 53
Exercise 54
Conclusion 54
■Chapter 3: Data Handling Using SAS 55
SAS Data Sets 56
Descriptive Portion of SAS Data Sets 56
Data Portion of Data Set 57
SAS Libraries 58
Creating the Library Using the GUI 59
Rules of Assigning a Library 64
Creating a New Library Using SAS Code 64
Permanent and Temporary Libraries 65
Two Main Types of SAS Statements 68
Importing Data into SAS 68
Data Set Creation Using the SAS Program 68
Using the Import Wizard 70
Import Using the Code 77
Data Manipulations 80
Making a Copy of a SAS Data Set 80
Creating New Variables 82
Updating the Same Data Set 87
Drop and Keep Variables 88
Subsetting the Data 90
Conclusion 93

■Chapter 4: Important SAS Functions and Procs 95
SAS Functions 95
Numeric Functions 96
Character Functions 101
Date Functions 105
Important SAS PROCs 108
The Proc Step 108
PROC CONTENTS 108
PROC SORT 112
Graphs Using SAS 120
PROC gplot and Gchart 121
PROC SQL 125
Data Merging 129
Appending the Data 129
From SET to MERGE 131
Blending with Condition 132
Matched Merging 134
Conclusion 143
■Part 2: Using SAS for Business Analytics 145
■Chapter 5: Introduction to Statistical Analysis 147
What Is Statistics? 147
Basic Statistical Concepts in Business Analytics 149
Population 149
Sample 149
Variable 150
Variable Types in Predictive Modeling Context 151
Parameter 151
Statistic 152
Example Exercise 152

Statistical Analysis Methods 160
Descriptive Statistics 160
Inferential Statistics 160
Predictive Statistics 161
Solving a Problem Using Statistical Analysis 161
Setting Up Business Objective and Planning 161
The Data Preparation 161
Descriptive Analysis and Visualization 161
Predictive Modeling 162
Model Validation 162
Model Implementation 162
An Example from the Real World: Credit Risk Life Cycle 163
Business Objective and Planning 163
Data Preparation 163
Descriptive Analysis and Visualization 163
Predictive Modeling 164
Model Validation 164
Model Implementation 164
Conclusion 164
■ Chapter 6: Basic Descriptive Statistics and Reporting in SAS 165
Rudimentary Forms of Data Analysis 165
Simply Print the Data 165
Print and Various Options of Print in SAS 165
Summary Statistics 168
Central Tendencies 169
Calculating Central Tendencies in SAS 173
What Is Dispersion? 177
Calculating Dispersion Using SAS 182
Quantiles 185
Calculating Quantiles Using SAS 187

Box Plots 189
Creating Boxplots Using SAS 192
Bivariate Analysis 196
Conclusion 196

Chapter 7: Data Exploration, Validation, and Data Sanitization 197
Data Exploration Steps in a Statistical Data Analysis Life Cycle 197
Example: Contact Center Call Volumes 198
Need for Data Exploration and Validation 201
Issues with the Real-World Data and How to Solve Them 204
Missing Values 204
The Outliers 205
Manual Inspection of the Dataset Is Not a Practical Solution 205
Removing Records Is Not Always the Right Way 205
Understanding and Preparing the Data 206
Data Exploration 206
Data Validation 206
Data Cleaning 207
Data Exploration, Validation, and Sanitization Case Study: Credit Risk Data 207
Importing the Data 210
Step 1: Data Exploration and Validation Using the PROC CONTENTS 211
Step 2: Data Exploration and Validation Using Data Snapshot 214
Step 3: Data Exploration and Validation Using Univariate Analysis 221
Step 4: Data Exploration and Validation Using Frequencies 232
Step 5: The Missing Value and Outlier Treatment 239
Conclusion 259

Chapter 8: Testing of Hypothesis 261
Testing: An Analogy from Everyday Life 261
What Is the Process of Testing a Hypothesis? 262
State the Null Hypothesis on the Population: Null Hypothesis (H_0) 266
Alternate Hypothesis (H_1) 266

Sampling Distribution 267
Central Limit Theorem 269
Test Statistic 272
Inference 274
Critical Values and Critical Region 279
Confidence Interval 280
Tests 283
T-test for Mean 283
Case Study: Testing for the Mean in SAS 283
Other Test Examples 286
Two-Tailed and Single-Tailed Tests 287
Conclusion 293
■ Chapter 9: Correlation and Linear Regression 295
What Is Correlation? 295
Pearson's Correlation Coefficient (r) 297
Variance and Covariance 297
Correlation Matrix 298
Calculating Correlation Coefficient Using SAS 298
Correlation Limits and Strength of Association 301
Properties and Limitations of Correlation Coefficient (r) 306
Some Examples on Limitations of Correlation 306
Correlation vs. Causation 312
Correlation Example 313
Correlation Summary 318
Linear Regression 318
Correlation to Regression 320
Estimation Example 322
Simple Linear Regression 325
Regression Line Fitting Using Least Squares 325
The Beta Coefficients: Example 1 327

How Good Is My Model? 328
Regression Assumptions 335
When Linear Regression Can't Be Applied 344
Simple Regression: Example 345
Conclusion 349

■ Chapter 10: Multiple Regression Analysis 351
Multiple Linear Regression 351
Multiple Regression Line 353
Multiple Regression Line Fitting Using Least Squares 354
Multiple Linear Regression in SAS 355
Example: Smartphone Sales Estimation 355
Goodness of Fit 357
Three Main Measures from Regression Output 358
Multicollinearity Defined 383
How to Analyze the Output: Linear Regression Final Check List 395
Double-Check for the Assumptions of Linear Regression 395
F-test 395
R-squared 395
Adjusted R-Squared 395
VIF 396
T-test for Each Variable 396
Analyzing the Regression Output: Final Check List Example 396
Conclusion 399

■ Chapter 11: Logistic Regression 401
Predicting Ice-Cream Sales: Example 401
Nonlinear Regression 404
Logistic Regression 407
Logistic Regression Using SAS 408

SAS Logistic Regression Output Explanation 410
Output Part 1: Response Variable Summary 410
Output Part 2: Model Fit Summary 412
Output Part 3: Test for Regression Coefficients 412
Output Part 4: The Beta Coefficients and Odds Ratio 413
Output Part 5: Validation Statistics 415
Individual Impact of Independent Variables 415
Goodness of Fit for Logistic Regression 416
Chi-square Test 416
Concordance 417
Prediction Using Logistic Regression 419
Multicollinearity in Logistic Regression 419
No VIF Option in PROC LOGISTIC 421
Logistic Regression Final Check List 421
Loan Default Prediction Case Study 422
Background and Problem Statement 422
Objective 422
Data Set 422
Model Building 426
Final Model Equation and Prediction Using the Model 438
Conclusion 440

■Chapter 12: Time-Series Analysis and Forecasting 441

What Is a Time-Series Process? 441
Main Phases of Time-Series Analysis 445
Modeling Methodologies 445
Box–Jenkins Approach 446
What Is ARIMA? 446
The AR Process 446

The MA Process 448
ARMA Process 450
Understanding ARIMA Using an Eyesight Measurement Analogy 452
Steps in the Box–Jenkins Approach 453
Step 1: Testing Whether the Time Series Is Stationary 454
Step 2: Identifying the Model 465
Step 3: Estimating the Parameters 497
Step 4: Forecasting Using the Model 501
Case Study: Time-Series Forecasting Using the SAS Example 503
Checking the Model Accuracy 506
Conclusion 507

■ **Chapter 13: Introducing Big Data Analytics 509**

Traditional Data-Handling Tools 509
Walmart Customer Data 509
Facebook Data 510
Examples of the Growing Size of Data 510
What Is Big Data? 511
The Three Main Components of Big Data 511
Applications of Big Data Analytics 513
The Solution for Big Data Problems 514
Distributed Computing 514
What Is MapReduce? 515
Map Function 515
Reduce Function 515
What Is Apache Hadoop? 517
Hadoop Distributed File System 517
MapReduce 519
Apache Hive 520

Apache Pig 521
Other Tools in the Hadoop Ecosystem 521
Companies That Use Hadoop 523
Big Data Analytics Example 524
Examining the Business Problem 524
Getting the Data Set 525
Starting Hadoop 525
Looking at the Hadoop Components 527
Moving Data from the Local System to Hadoop 529
Viewing the Data on HDFS 530
Starting Hive 534
Creating a Table Using Hive 535
Executing a Program Using Hive 536
Viewing the MapReduce Status 537
The Final Result 539
Conclusion 540

Index 541