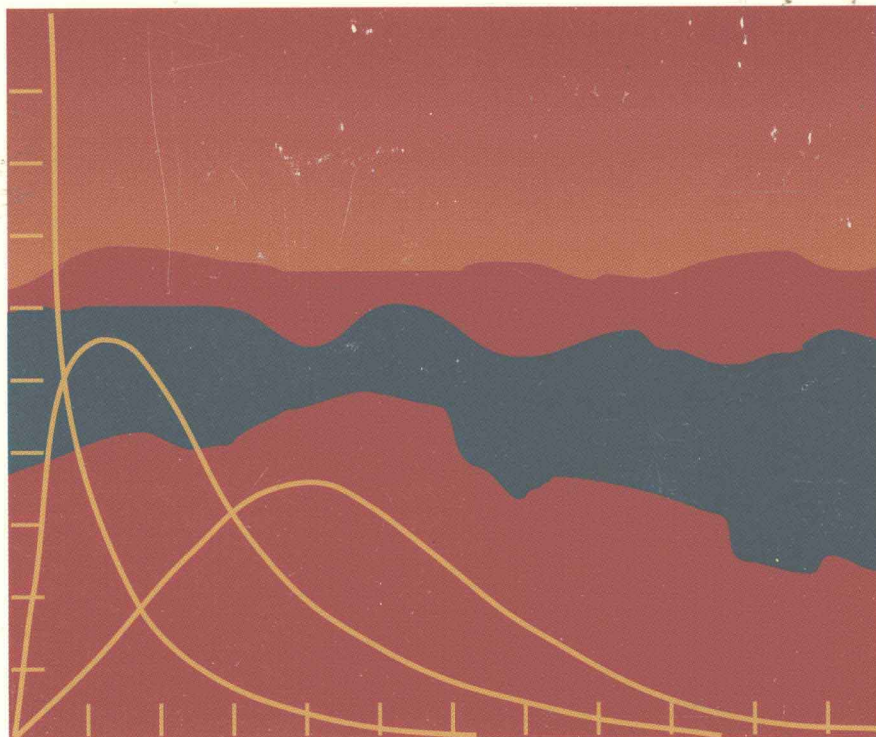


DOS EDITION



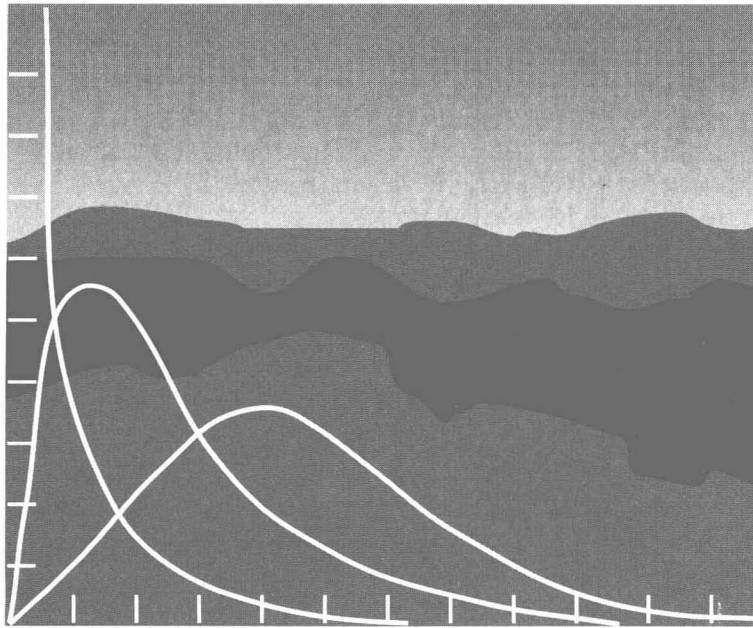
DATA ANALYSIS

WITH STUDENT SYSTAT



KENNETH N. BERK

D O S E D I T I O N



DATA ANALYSIS

WITH STUDENT SYSTAT

SYSTAT

KENNETH N. BERK

Illinois State University



Course Technology, Inc. One Main Street, Cambridge MA 02142

Data Analysis with Student SYSTAT DOS Edition is published by Course Technology, Inc.

Editorial Director Joseph B. Dougherty

Managing Editor Mac Mendelsohn

Product Manager Joan Carey

Production Manager Myrna D'Addario

Senior Production Editor Robin M. Geller

Production Assistants Erin Bridgeford, Christine Spillett

Desktop Publishing Supervisor Debbie Masi

Compositor Gex, Inc.

Copyeditor Joeth S. Barlas

Proofreader Ken Kimball

Technical Review Supervisor Mark Vodnik

Student Testers Jeremy Parker, Kevan Schultz, Rob Spadoni

Manufacturing Manager Elizabeth Martinez

Print Buyer and Planner Charlie Patsios

Text Designer Jean Hammond

Cover Designer Darci Mehall

Data Analysis with Student SYSTAT DOS Edition (c) 1994 by Course Technology, Inc.

Portions (c) 1994 by SYSTAT, Inc., used by permission.

All rights reserved. This publication is protected by federal copyright law. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, or be used to make any derivative work (such as translation or adaptation) without prior permission in writing from Course Technology, Inc.

Trademarks

Course Technology and the open book logo are trademarks of Course Technology, Inc.

SYSTAT and FASTAT are registered trademarks of SYSTAT, Inc.

Some of the product names used in this book have been used for identification purposes only and may be trademarks or registered trademarks of their respective manufacturers and sellers.

Disclaimer

Course Technology, Inc. reserves the right to revise this publication and make changes from time to time in its content without notice.

ISBN 1-878748-97-1 (text and 3½-inch software and data disk)

ISBN 1-56527-090-8 (text and 3½-inch data disk)

Printed in the United States of America.

10 9 8 7 6 5 4 3 2

Dear SYSTAT Student:

The academic community is SYSTAT's home. Making statistical software easier to use has always been one of our goals, but even more important to us is the practice of statistics. Practice means trying and learning. We hope this package will help you do that. Practice means making mistakes. We know you will make mistakes and hope that you will learn from them. Practice means, most of all, good use. It means using statistics to help people understand. You can use statistics software to do complex things, but that is not what statistics are about. Good statistical practice means summarizing information in ways that help others to understand patterns and make intelligent conclusions based on data.

We are not going to tell you that statistics are easy. Anyone who says that is trying to sell something. But learning to deal with and analyze uncertainty is at the core of life. Learn how to ask the right questions and you will not be fooled by "experts" who peddle bogus statistics. We hope you enjoy this text and look forward to working with you in the future.

Leland Wilkinson
President, SYSTAT, Inc.

From the Publisher

At Course Technology, Inc., we believe that technology will transform the way that people teach and learn. We are very excited about bringing you, college professors and students, the most practical and affordable technology-related products available.

The Course Technology Development Process

Our development process is unparalleled in the higher education publishing industry. Every product we create goes through an exacting process of design, development, review, and testing.

Reviewers give us direction and insight that shape our manuscripts and bring them up to the latest standards. Every manuscript is quality tested. Students whose backgrounds match the intended audience work through every keystroke, carefully checking for clarity, and point out errors in logic and sequence. Together with our own technical reviewers, these testers help us ensure that everything that carries our name is error-free and easy to use.

Course Technology Products

We show both *how* and *why* technology is critical to solving problems in college and in whatever field you choose to teach or pursue. Our time-tested, step-by-step instructions provide unparalleled clarity. Examples and applications are chosen and crafted to motivate students.

The Course Technology Team

This book will suit your needs because it was delivered quickly, efficiently, and affordably. In every aspect of our business, we rely on a commitment to quality and the use of technology. Every employee contributes to this process. The names of all of our employees are listed below:

Tom Atwood, David Backer, Stephen M. Bayle, Josh Bernoff, Erin Bridgeford, AnnMarie Buconjic, Jody Buttafoco, Jim Chrysikos, Susan Collins, John M. Connolly, David Crocco, Myrna D'Addario, Lisa D'Alessandro, Ann Deluca, Howard S. Diamond, Katie Donovan, Joseph B. Dougherty, Mary Jane Dwyer, Chris Elkhill, Don Fabricant, Kate Gallagher, Robin M. Geller, Suzanne Goguen, Laurie Gomes, Eileen Gorham, Michael Greene, Tim Hale, Roslyn Hooley, Nicole Jones, Matt Kenslea, Wendy Kincaid, Suzanne Licht, Elizabeth Martinez, Debbie Masi, Dan Mayo, Kathleen McCann, Mac Mendelsohn, Laurie Michelangelo, Kim Munsell, Paul Murphy, Amy Oliver, Kristine Otto, Jim Palmer, Debbie Parlee, Kristin Patrick, Charlie Patsios, Jim Pelis, Darren Perl, Kevin Phaneuf, George J. Pilla, Cathy Prindle, David Smith, Christine Spillett, Kathy Sutherland, Michelle Tucker, David Upton, Mark Valentine, Mark Vodnik, Jacqueline Winspear

Preface

This book has two goals: teaching SYSTAT and teaching data analysis. It can be used as a supplementary text for a first or second statistics course, or as the sole text for a data analytic course. For those already using SYSTAT or FASTAT, it should be a helpful hands-on supplement to the manuals.

Data Analysis

This text teaches data analysis and SYSTAT usage at the same time. The examples used throughout the text are more than just vehicles for demonstrating SYSTAT. In each case, the objective is to discover something about the data. If the only purpose were to teach SYSTAT, it would suffice to show each kind of plot only once. Yet plots, such as the normal plot and the box plot, appear throughout the text because they are needed in the data analyses. For good data analysis, it is not enough to demonstrate the individual procedures, it is important to show how each tool is used in context. Perhaps because of insufficient space, software manuals tend to show minimal examples without appropriate plots or other diagnostic tools. Software users tend to copy these examples, missing out on important data analysis techniques that are needed to better understand their data. Fortunately, Course Technology, Inc. has provided enough space for complete data analysis examples. Students should learn from these examples that statistics are more than just a set of isolated techniques, and that these techniques can work together to make sense of data.

The majority of examples and exercises use interesting, real-life data that illustrate statistical principles and the use of SYSTAT. There are approximately fifty data sets included with the text on the Student SYSTAT Data Disk.

It is important that the data be meaningful to the reader, because the discovery process depends on having some intuition about the data. The data sets have been chosen to be understandable without much technical knowledge, so the student can tell what to expect from the data and appreciate the conclusions.

Controversy is one of the fun aspects of statistics. Two statisticians might not agree on what procedure to use and might not make the same decision. Disagreements are inevitable in any subject that has practical implications in areas such as nutrition, instruction, medication, and sales promotion. Contingency table analysis and Bonferroni adjustments are two areas of special controversy that are mentioned in Chapters 8 and 9. Statisticians might find some other areas of disagreement, and we encourage them to communicate their reasoning to CTI so that their ideas can be incorporated into future editions. The reviewers of the book have been extremely helpful in this regard.

Without a computer, the detective work of good data analysis is difficult. Unless your data set is small, it is a lot of work to do the plots and computations by hand. Without the diagnostic tools and computational power of a good statistical computing package, it is easy to get incorrect answers.

SYSTAT is a powerful package for doing the graphs and calculations needed in statistical analysis. With a wide variety of features described in more than 1,500 pages of documentation, SYSTAT is one of the major statistical packages. SYSTAT has done very well in magazine and journal reviews that compare major packages. Reviewers are especially impressed with the power and variety of the graphics available.

Student SYSTAT offers an excellent bridge to the full version of SYSTAT, because their basic features work the same way. This text applies not only to Student SYSTAT, but also to the full version of SYSTAT and FASTAT. Users of these larger packages might find the basic material more accessible in this text than in their software manuals.

Student SYSTAT, available in DOS, Macintosh, and Windows versions, contains features from the full program that are needed for first and second statistics courses. Student SYSTAT permits huge data sets—as many as 100 variables are allowed and the number of cases is limited only by the amount of disk space available. The procedures include: basic statistics, t test, contingency tables, correlations, multiple regression, analysis of variance, nonparametric statistics, and time series. There are high resolution graphics for one, two, or several variables, such as exploratory data analysis, normal probability plots, point cloud rotation, scatterplot matrixes, and 3-D plots. Other sophisticated options allow you to apply LOWESS smoothing, use another variable to set point size (bubble plot), or overlay several plots. It is easy to compare treatments with box plots or side-by-side density plots, and these can be overlaid for an effective display. Student SYSTAT has a data editor for entering, displaying, and modifying data. It also has an ASCII editor for working with command files and output files. The procedures can be controlled through menus, interactive commands, or command files.

Students who continue studying statistics can move with ease from Student SYSTAT to SYSTAT. The full version of SYSTAT includes many additional features. The analysis of variance procedure can handle complicated designs and it allows various contrasts and multiple comparisons. Multivariate tests are available for regression and analysis of variance. The multivariate procedures include canonical correlation, cluster analysis, discriminant analysis, factor analysis, and multidimensional scaling. Time series features include Box-Jenkins ARIMA analysis and Fourier analysis. There are also procedures for nonlinear regression and loglinear modeling of contingency tables.

Content and Organization

Chapter 1 introduces basic menu operation in SYSTAT; it is a prerequisite to the rest of the book and you should work through it carefully.

Chapter 2 discusses typed commands rather than menus. With a few exceptions, menus are sufficient in SYSTAT and they are easier for beginners. It is not necessary to work through all the details of Chapter 2 unless you are planning to use commands extensively; however, it would be beneficial to look over the first two-thirds of this chapter just to see how commands work. In later chapters most analysis is done with menus but the alternate command route is also provided. Expert use of SYSTAT requires commands.

Chapter 3 introduces basic graphics and statistics in SYSTAT. Chapter 4 introduces graphs that relate to more than one variable. Chapters 5 and 6 are about the basic distributions: normal, chi-square, t , and F . Simulations are introduced as a way to

demonstrate, for example, the Central Limit Theorem. It is not necessary to work through all of these commands, which involve a lot of typing, but you should read these two chapters. Even if you have learned about statistical distributions elsewhere, this material shows how SYSTAT can help you understand distributions. The remainder of the book uses the normal, chi-square, t , and F extensively.

Chapters 7 through 11 cover the basic techniques of statistical analysis: t tests, contingency tables, correlation, regression, and analysis of variance. Instead of being isolated in a separate chapter, nonparametric methods are introduced with the corresponding parametric procedures.

Chapter 12 is about time series, which should be useful for business and economics classes. Chapter 13, written by Jeffrey W. Steagall, covers quality control.

The Exercises form an integral part of the text. They demonstrate important principles and are intended to be interesting. Please read them, even if you do not have time do them all.

All of the features explained in Chapters 1 through 13, as well as additional features not covered in the text, are summarized in the Reference section.

SYSTAT Coupon

Students who purchase this text without the Student SYSTAT software can purchase their own copy of the software at a nominal price. Ordering information is contained at the end of the text.

Acknowledgments

The people at Course Technology, Inc. have worked very hard on this book. Thanks to those with direct involvement, including Kitty Pinard, Marcia Cole, Mac Mendelsohn, and Joan Carey. Thanks to Joe Dougherty for his managerial support, Mark Vodnik for his careful coordination of student testing and creation of the data disk, and Robin Geller for her expert navigation of this project through all the stages of the production process. Thanks also to the student validators, Jeremy Parker, Kevan Schultz, and Rob Spadoni who tried all of the steps and did all of the exercises; and to Joeth Barlas for her precise copyediting.

Thanks to Jeff Steagall for writing Chapter 13, Quality Control, and to Kathy Finnegan for writing the Reference section.

The editors at CTI deserve special thanks for getting good reviewers of the book. The reviewers have been extremely helpful in correcting errors and pointing out areas in need of improvement. The reviewers are:

Ray Faith, MIT; William Klein, Colby College; Walter R. Paczkowski, AT&T Bell Labs and Rutgers University; Chris Noble, Lawrence University; David Hildebrand, Wharton School, University of Pennsylvania; Larry Stonecipher, Sangamon State University; C. P. Kartha, University of Michigan; Justin Stolen, University of Nebraska at Omaha; Wayne Winston, Indiana University; Sue McFarland, Wharton School, University of Pennsylvania; and Andrew Seila, Terry College of Business at the University of Georgia

Data sets have come from family, friends, students, and others. Thanks to Laura Berk, Peter Berk, Robert Beyer, David Booth, Orlyn Edge, Stephen Friedberg, Richard Goldstein, Lotus Hershberger, Les Montgomery, and Diane Warfield. Joyce Nervades put together the SURVEY9.SYS data from the survey responses.

The people at SYSTAT, Inc. have made important contributions. MaryAnn Hill was a special reviewer, making many useful suggestions for improvement. Patrick Fleury helped with problems in running the software. Eve Goldman initiated the book and facilitated the project at every stage. Leland Wilkinson created SYSTAT and built it up to be one of the important statistical packages.

Finally, I would like to thank my wife and sons for their support in a time-consuming effort.

Kenneth N. Berk

Table of Contents

CHAPTER 1

GETTING STARTED 1

- How to Follow the Step Lists in This Book 2
- Installing Student SYSTAT 3
- Starting Student SYSTAT 4
- Changing the Default Directory 5
- Printer Installation 6
- Using the SYSTAT Edit Worksheet 7
 - Entering Variable Names 9
 - Entering Data 9
 - Moving Around the Data Worksheet 11
- Using the Edit Commands 11
 - Saving Your Data 11
 - Changing Variables or Data 12
- Re-expressing Data 12
 - The IF Statement 13
 - Useful Shortcuts in Typing Commands 14
- Exiting the Worksheet and Quitting SYSTAT 14
- Sorting Data 15
- Printing Output 17
 - The FEDIT Full-Screen Text Editor 17
- Creating Graphics 18
- Printing Plots 19
- Importing Data 21
- Renaming Variables 24
- Exercises 26

CHAPTER 2

COMMANDS 29

- Review Mode 30
- Command Files 32
- The Command Log 33
 - Saving Files from FEDIT 34
- Command Mode 35
 - The Command Prompt 35
 - Printing Graphs 37
 - FEDIT in Command Mode 38
 - The DATA Command Procedure 42
 - Importing ASCII Files in DATA 44
 - Additional Information on Importing ASCII Files 45
- Running SYSTAT 48
 - Menus and Commands in FEDIT 48
 - Interactive Commands and Commands in FEDIT 49
 - Commands in FEDIT 49
- Exercises 50

CHAPTER 3

BASIC GRAPHICS AND STATISTICS 53

- Graphic Displays 54
 - Graphical Analysis of Data 54
 - Histograms 54
 - Box Plots 56
 - Stem-and-Leaf Plots 58
 - Transformation for Symmetry 61
- Basic Statistics 64
- Commands 67
- Exercises 68

CHAPTER 4

GRAPHICS 71

- Scatterplots 72
 - Labeling the Points 73
 - Plotting Several Y's 74
 - Can You Tell the Men from the Women? 76
- 3-D Plots 77
- Spinning in 3-D 79
- The Scatterplot Matrix 81
- Commands 83
 - Using a Third Variable for Symbol Size 84
- Bubble Plots 86
- Exercises 87

CHAPTER 5

THE NORMAL DISTRIBUTION 91

- Normal Distribution of Data 92
 - Normal Random Numbers 93
 - The Normal Probability Plot 96
 - The Population and the Sample 98
 - Commands 99
- Averages 100
 - The Average of Normal Data 101
 - Commands 104
- The Central Limit Theorem 105
- Exercises 108

CHAPTER 6

CHI-SQUARE, t , AND F 111

- The One-Sample t 112
- The Two-Sample t 114
- The Robustness of t 116

Chi-Square 120
The 2×2 Table 121
 Does It Have the Chi-Square Distribution 122
Can the Two-Group t Be Used? 125
The F Distribution 126
Probability Functions in SYSTAT 127
Exercises 127

CHAPTER 7

COMPARISONS 129

Dependent Comparisons 130
 A Nonparametric Approach 134 Commands 136
 Confidence Intervals 137
Independent Comparisons 138
 A Confidence Interval 141 A Nonparametric
 Test 141 Commands 142
Beneath the Surface 144
 Commands 150
Exercises 151

CHAPTER 8

TABLES 155

Variable Types: Continuous and Categorical 156
 Ordinal and Nominal Variables 156 Single
 Categorical Variables 156
Graphing Categorical Variables 156
 Pie Charts 157 Bar Charts 158
Frequency Tables 159
 Two-Way Table 159 Applying Pearson Chi-
 Square Test to a Modified Table 167 Evaluating
 p -Values 167 Other Test Statistics 168
Tables with Ordinal Variables 169
 Testing for a Relationship Between Two Ordinal
 Variables 169 Limitations of Pearson Chi-
 Square Test for Ordinal Variables 171 Ordinal
 Measures 171 Spearman Rho 172
Commands 172
Labeling Groups Within Categorical Variables 173
Starting from a Table 174
 Entering a Table into the Edit Worksheet 174
 Analyzing a Table 175
Combining Categories 177
 Defining a New Variable 178 Obtaining the
 Statistics with Combined Variable 178
Exercises 179

CHAPTER 9

CORRELATION AND SIMPLE REGRESSION 181

Simple Linear Regression 182
 Regression Line 182 Fitting the Regression
 Line 183 Regression Assumptions 183
 Straight-Line Model Problems 184
Calculating a Regression Line 184
 Plot the Data 185 Computing the Regression
 Equation by Hand 186
Correlation 187
 Correlation and Slope 187 Computing
 Correlation by Hand 187 Using SYSTAT to
 Compute Correlation 188 Interpreting
 Correlation Output 188
Regression in SYSTAT 189
 Storing Residuals and Other Diagnostics in a
 File 190 Satisfying the Four Regression
 Assumptions 191 Calculating a Confidence
 Interval for the Slope 191 Hypothesis Tests and
 Correlation 191 Tentative Conclusions 192
 Interpreting the Analysis of Variance Table 192
 The Independence Assumption 193
Interpreting the Residual File 193
 Interpreting the Diagnostics to Check for
 Assumption Violations 194 Plotting Student
 Residuals 195 Validating Conclusions from the
 Residual File 196
Commands for Correlation and Regression 196
Correlation and Plot Matrixes 197
 The Correlation Matrix for Relating Multiple
 Variables 197 Correlation of CALC with
 Others 198 Correlation with a Two-Valued
 Variable 199 Bonferroni 199 Bartlett 199
 Correlations among Predictors 200
 Correlation with CALCHS Implies
 Causation? 200
Measuring the Extent of Linear Relationships 200
 Scatterplot Matrix 200
Commands for Correlation and Plot Matrixes 202
Outlier Horrors 202
 Correlating Rating and Price with the
 Outlier 202 Correlating Rating and Price
 Without the Outlier 205
Commands 206
Other Horrors 207
Exercises 207

CHAPTER 10

MULTIPLE REGRESSION 211

Regression Models 212

Regression Assumptions 212

Uses of Regression 212

Regression Example: Predicting Grades 213

The Analysis of Variance Table 215 Multiple Correlation 216 An Error Estimate 216 Coefficients and the Prediction Equation 216 Tolerance and Multicollinearity 217 Tests for the Coefficients 218 Warnings about High Leverage 218 The Independence Assumption 219

Using Residuals to Verify Regression Assumptions 219 Interpreting the Residual Output 219

Useful Regression Plots 220

Dependent Variable vs. Predicted Values 220

Plotting STUDENT Residual vs. Estimate 221

Normal Errors and the Normal Plot 223

Summary of CALC Analysis 224

Commands 224

Regression Example: Sex Discrimination 224

Regression on Just the Men 225 Using a SPLOM to See Relationships 225 Correlation Matrix 226 Multiple Regression 227 Interpreting the Regression Output 228 Warnings 229 Significant Durbin-Watson? 229

Residual Analysis of Discrimination Data 230

Constructing a Variable to Number Cases 230

STUDENT Residuals vs. Predicted 231 Normal

Plot of STUDENT Residuals 232

Are Women Underpaid? 233

Obtaining Statistics on Residuals 233 Drawing Conclusions 235

Selecting a Subset of Predictor Variables 235

Forward Selection 235 Backward

Elimination 235

Forward Selection Example: Calculus Data 236

Selecting Predictors for the Calculus Data 236

Interpreting Stepwise Regression Results 238

Partial Correlation 239 Improving the Error

Sum of Squares 239 Tolerance 240 Building a

Model in Three Steps 240 Multiple Regression

Summary 240 Are the F Tests Valid? 241

Commands for Forward Stepping 242

Validation with Random Splitting (Optional) 242

Randomly Splitting the Data 242 Performing

Forward Selection with Randomly Split Data 243

Exercises 245

CHAPTER 11

ANALYSIS OF VARIANCE 249

One-Way Analysis of Variance 250

Analysis of Variance Assumptions 250 One-Way Example: Comparing Hotel Prices 250 Graphing the Data to Verify Assumptions 251 Computing the Analysis of Variance 252 What Is the Analysis of Variance? 254 Means and Standard Deviations 255 Multiple Comparisons: Tukey Test 256 A Nonparametric Approach: Kruskal-Wallis Test 256 The Dallal Plot 257

Commands 258

Two-Way Analysis of Variance 259

Two-Way Example: Cola 259 The Model 260 Assumptions 261 Density Plot 262 Equal Variances? Outliers? 262 Interaction? 263 Main Effects 263 Hypothesis Tests 263 The Analysis of Variance Table 265 Cell Means Standard Deviations 266 Dallal Plot 266 Checking for Outliers with a Normal Plot 267 The Interaction Plot 268 Summary 270

Commands for Two-Way ANOVA 270

Analysis of Covariance 271

Assumptions for Analysis of Covariance 271 A Graphic Illustration 272 The Graph Shows Increased Precision 272 Adjusted Means from the Graph 273 Bias Adjustment 273 Nobody Said It Would Be Easy 273 Analysis of Covariance Example: Hotels 273 Plotting the Variables to Check for a Linear Relationship 274 Sneak Peek at the Plot with Parallel Lines 275 Performing the Analysis of Covariance 276 Interpreting the Analysis of Covariance Output 277 Outliers 278 Putting Parallel Lines on the Plot 278 Regression to Compare NY with Others 279 Testing for Parallel Lines in the Analysis of Covariance 281 Two-Way ANOVA Without Interaction 282 Comparing Two Models 283 Two-Way ANOVA with Missing Cells 284

Commands for ANCOVA Hotel Example 284

Examples of Reacting to Interaction: The Glue Study 285 Use By for Several One-Way ANOVAs 287 Dealing with Interaction 289

Commands for the Glue Analysis 290

Exercises 290

CHAPTER 12

TIME SERIES 295

Time Series Concepts 296

Percentage Change and Logarithms 296 The
Lag Function 296

The Dow in the 80s 297

Transforming for Constant Variance 297
Plotting the Dow Jones Time Series 298 Plotting
the Changes 299 Plotting Transformations to
Minimize the Variance 300

Commands 301

Using the Series Submenu on the Statistics Menu 302

Plotting Using the Series Submenu 302 The
Autocorrelation Function 304 Autocorrelation of
Differences of Logarithms 305 Transformations
in Series 307 ACF of Percent Change 307 Using
Clear to Undo a Transformation 308

Commands 309

Lagged Averages 309

Exponential Smoothing 309 Residuals from
Exponential Smoothing 311 Correlating Buy
Signal Strength with the Next Observation 312
Graphing the Relationship 313 Unweighted
Averages 314 The Smoothed File in Edit 315
Merging the Two Smoothed Files 316
Comparing Buy Signals and the Month's
Change 316 Summary 317

Commands 317

Seasonality 319

Multiplicative Seasonality 319 Additive
Seasonality 319

Seasonal Example: Beer Production 320

High-Resolution Plot 320 A Box Plot to Show
Seasonality 321 Outliers 323 Series Plot 323
Seasonal ACF 324 Seasonal Adjustment 325
Smoothing for Level, Growth, and Seasonality 327
Additive Model 328 Multiplicative Model 328
Smoothing Constants 328 Specifying the

Model 328 Forecasts 329 Seasonal

Indexes 330 Future Forecasts 331

Graphing the Forecasts 332

Commands 334

Exercises 336

CHAPTER 13

QUALITY CONTROL 339

Statistical Quality Control 340

Two Types of Variation 340

Control Charts 341

Control Charts and Hypothesis Testing 344

Variables and Attributes 344

Mean Charts and General Theory of Control
Charts 344

Understanding the Basic Mean Chart 346

The Mean Chart When Sigma Is Unknown 348

The Range Chart 353

The P-Chart 356

Using Commands to Create Control Charts 358

Exercises 359

SYSTAT REFERENCE 361

The Utilities Menu 362

The Files Menu 367

The Data Menu 374

The Graph Menu 384

The Statistics Menu 404

Commands 418

The DATA Command Procedure 426

Summary of Data Sets 471

Bibliography 474

Index 475

CHAPTER 1

GETTING STARTED

OBJECTIVES

In this chapter you will learn to:

- Install Student SYSTAT
- Use the menus
- Enter data
- Import data
- Sort data
- Create graphics
- Print results

Student SYSTAT is a powerful program designed to help you look at data graphically and numerically. Student SYSTAT can perform many numerical statistical procedures, as well as a variety of graphical procedures. Although you might initially plan to use a certain statistical procedure, it is important to explore the data from more than one angle, because often there are unexpected twists. Sometimes there is more going on in the data than you could have imagined ahead of time, and you must do some detective work to find out what is really happening. This book uses examples with real data to give you a feel for this discovery process.

This first chapter covers the basics of Student SYSTAT, without using fancy statistics or graphics. Nevertheless, there is much to be discovered with these displays.

How to Follow the Step Lists in This Book

In each chapter of this book there are step-by-step instructions. Menu steps that you should perform are numbered within a grey bar, for example:

1 Press [Esc] to return to the main menu.

Within numbered lists are the following types of instructions:

- Keys that you press are in brackets and in boldface: Press **[Esc]**.
- The letters, numbers, and special characters that you type are in boldface: Type **quit**.
- Key combinations that you press simultaneously are in brackets and boldface and adjacent: Press **[Ctrl][Y]**. This indicates that while pressing and holding the first key, you press the second key, then release both.
- Menu items that you select are in boldface: Select **Utilities**. This indicates that you highlight the menu item with the cursor and then press **[Enter]**.

Command steps that you should perform are displayed next to a grey bar, for example:

To obtain a new variable that measures change from 1980 to the second half of the 80s, type:

LET DIFF = AVER8589 - DAYS_80

and press **[Enter]**.

- Characters that you type are boldface and in a different type:

LET DIFF = AVER8589 - DAYS_80

- Keys that you press are in brackets: Press **[Enter]**.

In addition to these two types of step lists, the book also contains sets of commands. They are presented as reference material for the student. For example:

Returning to the command prompt via the Escape key, you could alternatively use the AVG command to form the average:

LET AVER8589 = AVG(DAYS_85,DAYS_86,DAYS_87,DAYS_88,DAYS_89)

Installing Student SYSTAT

Installing Student SYSTAT is straightforward. All you need to do is back up your disks, run the installation routine, and start working.

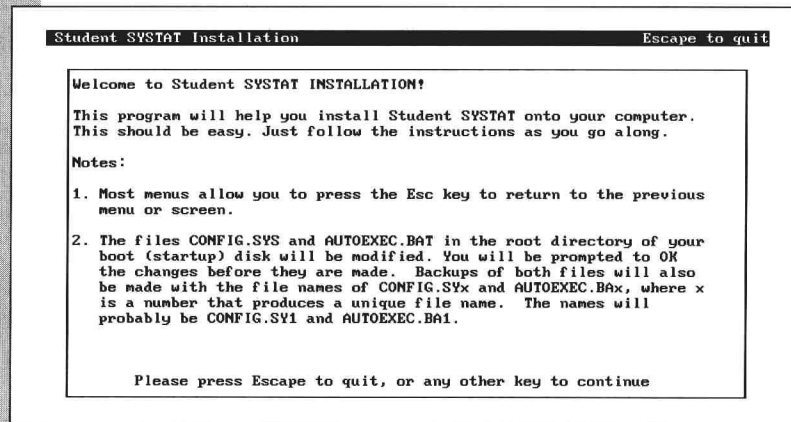
- Use the DOS command COPY or DISKCOPY to make a backup of all the Student SYSTAT disks. (See your DOS manual for more information.) Place the original disks in a safe place, and use the backup disks to install SYSTAT.

INSTALL—a menu-driven installation program—guides you through the installation of Student SYSTAT. INSTALL copies Student SYSTAT to your computer's hard drive simply and safely. Just follow the steps. In these installation instructions, it is assumed that the name of your hard drive is C and the name of your diskette drive is A. If necessary, substitute the correct letter designations.

To start INSTALL:

- 1 Insert Disk 1 into drive A.
- 2 At the C> prompt, type A:INSTALL and press [Enter]. The installation program begins and the first screen appears. See Figure 1-1.

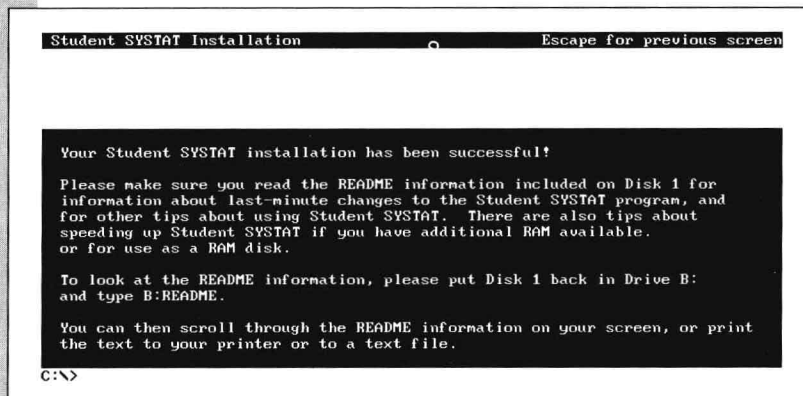
Figure 1-1
Welcome to Student
SYSTAT INSTALLATION!



- 3 After reviewing this screen, press [Enter] or any other key to continue.
- 4 Type A or the letter of the diskette drive where you will be inserting SYSTAT disks during installation.
- 5 At the prompt, enter the directory on your hard drive where you want to install Student SYSTAT. You can put it in any directory, but we recommend that you create a directory by typing a name for it at the DOS prompt. For example, if you are installing Student SYSTAT on your hard drive, you would type **SYSTAT** or **STUSTAT** (or some other appropriate name) at the prompt c:\ and press [Enter]. It is very important that you type a name for your directory in order to avoid installation into your root directory.

- 6 With Disk 1 in drive A, press [Enter] or any other key to continue. After a few moments, you will be prompted to insert the next disk into the drive.
 - 7 At the prompt, insert Disk 2 into drive A. Follow the same procedure for the remaining Student SYSTAT disks. (If you insert the wrong disk, you will be prompted to replace it with the correct one.) If you have a problem, see your instructor or technical support person for help.
 - 8 After installing the last disk, you will be prompted to specify the drive from which your computer boots, and to specify whether it is a hard drive or a diskette drive. Remove the last installation disk from your drive.
 - 9 As Student SYSTAT examines your CONFIG.SYS and AUTOEXEC.BAT files, it might recommend changes. You should accept the changes that SYSTAT recommends. Type Y to accept or N to keep them unchanged. If you are using a computer in a lab, check with your technical support person before changing the CONFIG.SYS and AUTOEXEC.BAT files.
- When INSTALL is complete, the screen in Figure 1-2 appears, letting you know that your Student SYSTAT installation was successful, and giving you instructions on how to access the README file.

Figure 1-2
Your Screen After
Successfully Installing
Student SYSTAT



- 10 At the prompt type **COPY A:.* C:\SYSTAT** (or whatever you named your directory) and press [Enter] to copy the Data Disk into the directory where you installed Student SYSTAT. If you need to keep your data on a diskette, see the section Changing the Default Directory.

Starting Student SYSTAT

- To start Student SYSTAT:
- 1 Go to the drive where Student SYSTAT resides (this will usually be drive C).
 - 2 At the prompt type **cd \SYSTAT** (or whatever you have named your directory) and press [Enter].
 - 3 Type **STUSTAT** and press [Enter]. Neither DOS nor SYSTAT is case-sensitive except in certain situations where it is explicitly indicated.
 - 4 After the logo appears press [Esc] or any other key to start Student SYSTAT.