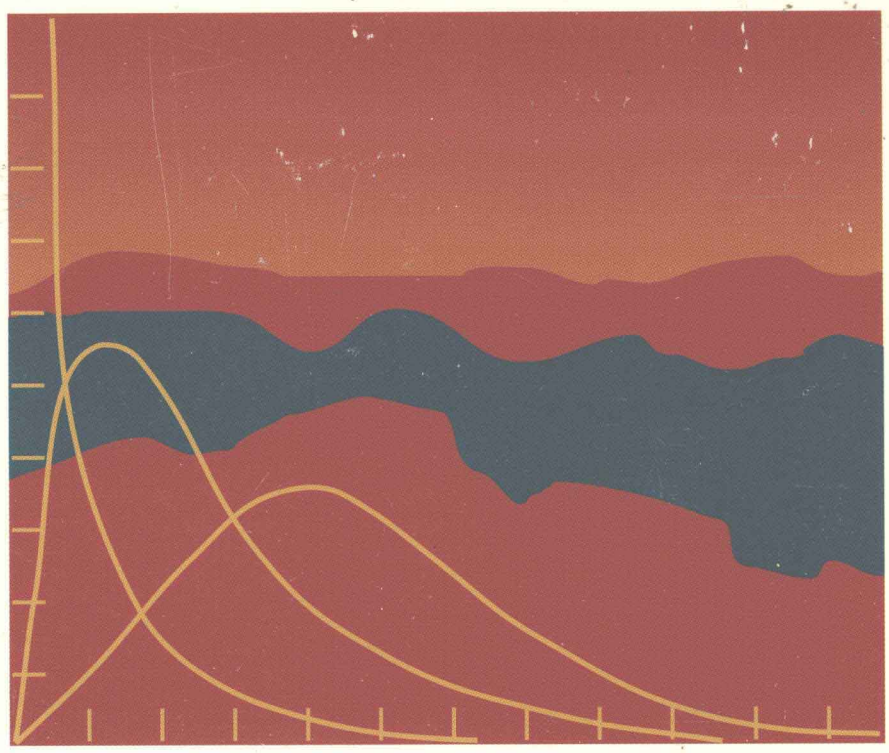
DOS EDITION



# DATA ANALYSIS

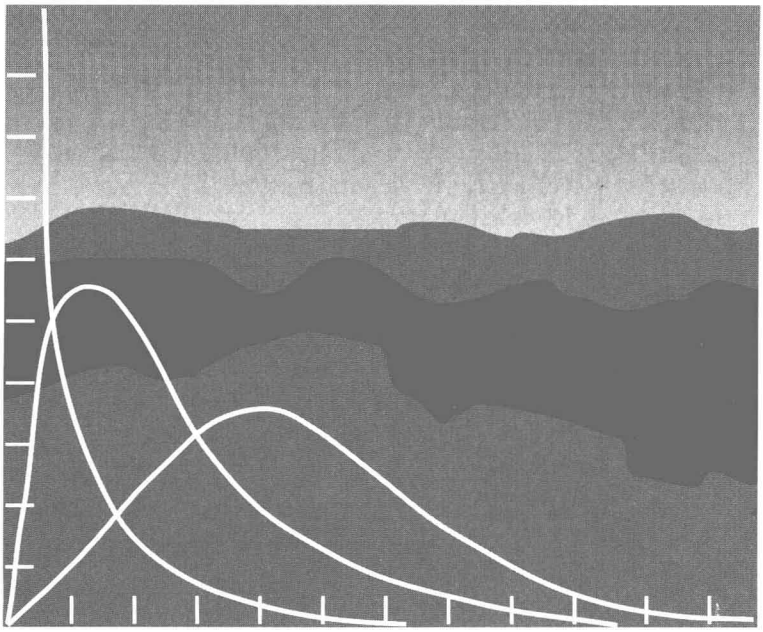
## WITH STUDENT SYSTAT

SYSTAT®

KENNETH N. BERK

Course TECHNOLOGY

DOS EDITION



# 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



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 furture.

*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

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

## 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.

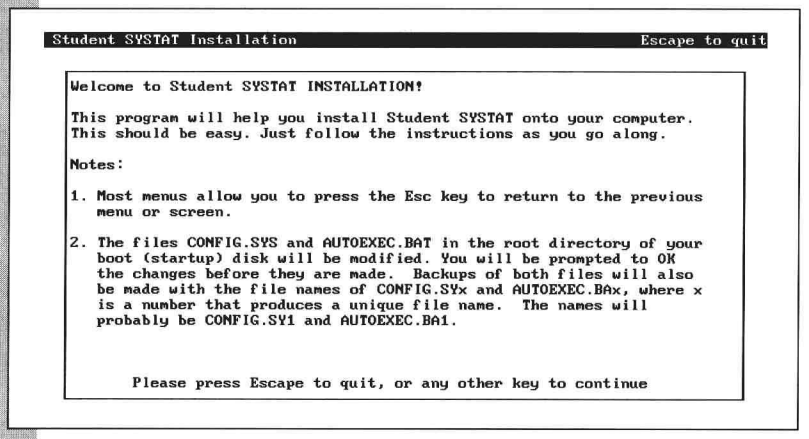



**Figure 1-1**
Welcome to Student SYSTAT INSTALLATION!

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.

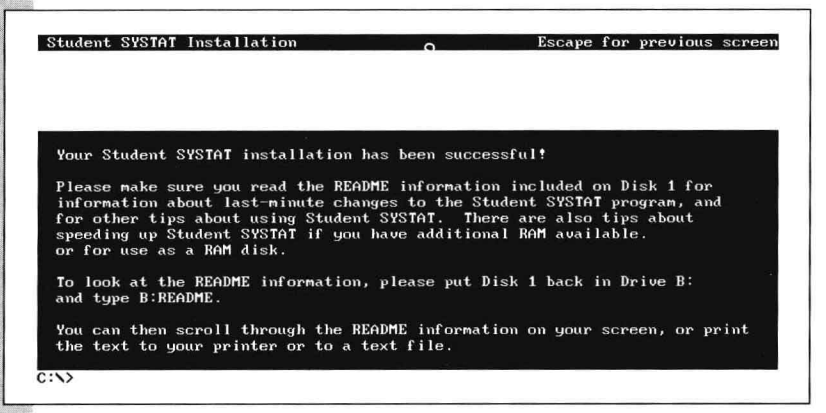



**Figure 1-2**
Your Screen After Successfully Installing Student SYSTAT

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.