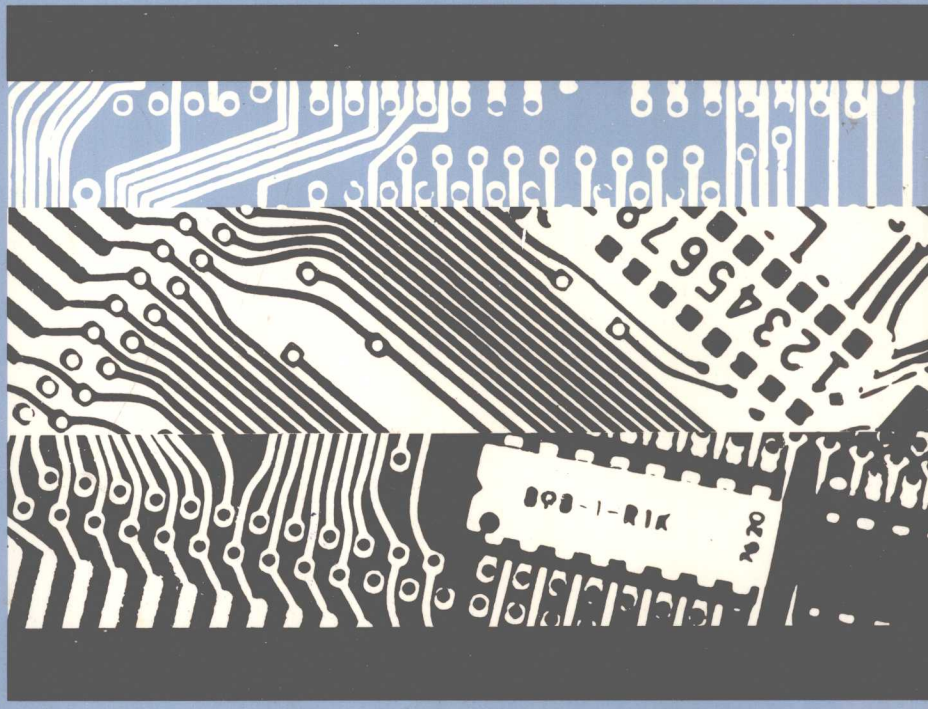
Through the Maze
Statistics with Computer Applications

Margaret Platt Jendrek



Through the Maze

Statistics with Computer Applications

Margaret Platt Jendrek

Miami University of Ohio

Wadsworth Publishing Company

Belmont, California

A Division of Wadsworth, Inc.

Sociology Editor: Sheryl Fullerton
Production: Mary Forkner, Publication Alternatives
Designer: Al Burkhardt
Copy Editor: Kathleen Engelberg
Technical Illustrator: Carl Brown
Cover: Al Burkhardt

Printed in the United States of America

3 4 5 6 7 8 9 10—89 88 87

Library of Congress Cataloging in Publication Data

Jendrek, Margaret P.
Through the maze.
Includes index.
1. Social sciences—Statistical methods—Data processing. 2. Statistics—Data processing.
3. SPSS (Electronic computer system) I. Title.
HA32.J46 1984 310'.28'54 84-13188
ISBN 0-534-03921-9

Through the Maze

Statistics with Computer Applications

preface

Social science students need to be able to interpret statistics and use statistical computer packages. In fact, some claim that by 1987 over eighty percent of all jobs will require a working knowledge of computers. No book provides students with instruction as well as extensive examples and quizzes in both statistical interpretation and computer use. Current texts cover either statistics, with an emphasis on computation, or computer programming. That's why I wrote this book! I wanted a book for my research methods students that would:

1. Present the conceptual foundations of basic statistics
2. Guide students in the interpretation of statistics
3. Present the computer as an important data analysis tool
4. Guide students in using a popular social science computer package, the *Statistical Package for the Social Sciences* (SPSS^x)
5. Guide students through computer projects that require statistical interpretations
6. Allow students to work at their own pace

This text was written to attain these objectives, and it is unique in four respects. First, it includes *many* examples and checkpoints. This approach allows students to learn materials at their own pace. Some students may need to work through the material several times, yet others may only need to review a few sections. In either case, the material is presented in such a way that students receive continuous feedback about their grasp of the subject.

Second, the book is designed to guide students through the interpretation of statistics and the use of the computer via a popular social science program, SPSS^x. Social science students are notorious for being "math, statistics, and computer shy." Yet their ability to interpret statistics and use computers is among their most marketable skills. This book teaches these skills and provides students with practice.

Third, whereas some books examine several computer packages and others teach one computer language, few existing books teach the

basics of just one computer package. In one course students can meaningfully learn only one computer package.

Fourth, concept quizzes and computer application exercises appear throughout each chapter. *Concept quizzes* test students on definitions and concepts presented in the chapter. The "hands-on" *computer exercises* test both programming and interpretive skills. The "hands-on" exercises are based on a *data set* (on cards) provided to instructors who adopt the book. The data set includes variables of interest to people in education, political science, psychology and sociology.

Acknowledgments

My sincerest thanks to my colleague Ted Wagenaar. He read every chapter and offered constructive comments and moral support. The constructive comments offered by the following reviewers truly improved the book: Kristine L. Anderson, Florida Atlantic University; Kenneth J. Berry, Colorado State University; David W. Chilson, Bowling Green State University; Lawrence G. Felice, Baylor University; David B. Graeven, California State University, Hayward; Gary D. Hampe, University of Wyoming; Joseph Healey, Christopher Newport College; David A. Nordlie, Bemidji State University; and Wornie L. Reed, Morgan State University. I am grateful to the Literary Executor of the late Sir Ronald Fisher, F. R. S., to Dr. Frank Yates, F. R. S., and to the Longman Group Ltd., London, for permission to reprint Tables III and IV from their book, *Statistical Tables for Biological, Agricultural, and Medical Research* (6th edition, 1974). I do have a husband and daughter. I suspect there were days and weeks when they wondered if they had a wife and mother. I thank them both for giving me the undisturbed time I needed to create the book.

Margaret Platt Jendrek

To Gene and Emily

contents

chapter 1

An Introduction to the Computer

Welcome to the social scientist's computer world! For some of you this is a familiar world. For others, it's a new and magical world. This chapter familiarizes the new user with the computer and reviews computer terms for the old user.

After completing this chapter, you should be able to do the following:

1. Define selected computer terms
2. Use selected computer terms
3. Describe the four parts of the computer
4. Explain why computers cannot run the world
5. Explain why social scientists use computers

Why Use a Computer?

Read today's newspaper and notice the number of computer advertisements. Read a paper from 20, 10, or even 5 years ago, and you will not see computer advertisements. In recent years the computer has entered every phase of our lives (including our newspapers!). For example, schools computerize student lists; banks computerize customer statements; employers computerize payrolls; supermarkets computerize checkouts; and dating services computerize love and romance. Incredible? Yes and no. Yes, when we're speaking of the rapid developments in computer technology. No, when we're speaking of why we computerize.

The computer is a machine directed by a set of instructions to save, retrieve, and manipulate data. For the social scientist the computer is a valuable research tool. The computer makes it possible to handle large quantities of information quickly and accurately. For ex-

ample, adding the ages of 3000 respondents, a simple calculation, may take a person several hours. The same calculation may take a computer less than one second!

Remember, however, that the computer is simply a machine. An automobile is also a machine. When driving a car, you are responsible for selecting the route. If you select a direct route you quickly arrive at your destination. If you select an indirect route you arrive later. If you select a poor route you may never arrive at your destination at all. In a similar way, you direct the computer. A **program** is a set of sequential directions that you give the computer. A top-notch **programmer**, a person who writes computer programs, is someone who directs the computer to do a job in a quick and efficient manner. A poor program is written by a programmer who meanders through the task. A really bad program does not do its intended job; the programmer, *not* the computer, goofed in the instructions!

Another term for the programs that run the computer is **software**. The physical equipment that makes up the computer is termed **hardware**. You can think of your favorite chocolate-chip cookie recipe as a program, a set of sequential directions for baking cookies. Your blender and oven are the hardware needed to bake the cookies.

Before considering how a computer works, let's discuss the difference between **data** and **programs**. Data refers to the information that we want to manipulate or analyze. In the chocolate-chip cookie example, the chips, flour, salt, sugar, baking soda, and eggs are the data. Social scientists frequently collect data on a sample of individuals. Demographic data often include the age, income, sex, and education of respondents. If we want to know the average age of our respondents, we need a program that can calculate an average. When we process our age data through a program that calculates an average, we get the average age of our respondents. Notice that the same program can be used to calculate the average income of our respondents. Although the data may change, the program remains the same; it was written to calculate an average.

The Parts of a Computer

To understand how a computer works one must be familiar with its four major parts: **input, storage, CPU,** and **output** devices. The input, storage, and output devices are called **peripheral devices**. They are external to the computer. Input devices enable us to feed information (programs or data) into the computer. Storage devices allow us to keep this information for later recall. The computer communicates the results of our program's execution to us through output devices. The CPU is the computer. The relationships among the four parts are illustrated in figure 1-1. Let's examine each device in more detail.

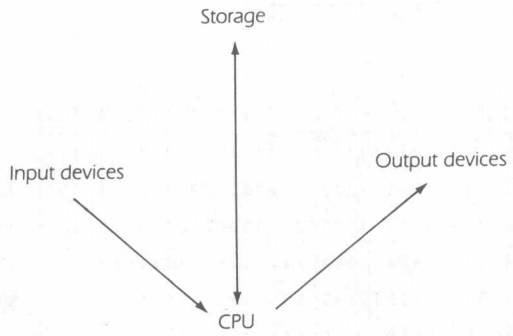


Figure 1-1 The four major parts of a computer.

Input Devices

Input devices enter data or programs into the computer. Typical input devices include the **card reader** and the **computer terminal**. The card reader inputs to the computer information contained on computer cards. As shown in figure 1-2, a computer card contains 80 columns. Each column is composed of the numbers 0 to 9. Information is punched onto cards by a **keypunch** machine, a machine that resembles a typewriter. The keypuncher types the program or data information at a keyboard, and the machine punches holes in the card's columns and rows. Information punched on cards is then input into the computer via a card reader. A typical card reader can process about 1000 cards per minute.

Alternatively, data or programs may be input from a terminal into the computer. Whereas a card reader acts as a mediating device between the computer card and the computer, the terminal is hooked directly to the computer. Two commonly used terminals are the **hardcopy** terminal and the **CRT** (cathode-ray tube). The hardcopy terminal looks like a typewriter. Users type their program or data at the terminal, sometimes called a console, and see the information printed on paper, the hardcopy. A CRT terminal resembles a typewriter with a small television screen. Users type their program or data at the terminal's console and see it displayed on a screen. Both types of terminals input data or programs directly into the computer. Working on a CRT tends to be faster than working on a hardcopy terminal because screen printing is more rapid than hardcopy printing. However, you don't usually get a hardcopy of your work at a CRT.

Storage Devices

Storage devices enable the computer to keep your programs and data. Without storage devices you would have to input your data or program into the computer every time you wanted to do a calculation or manip-

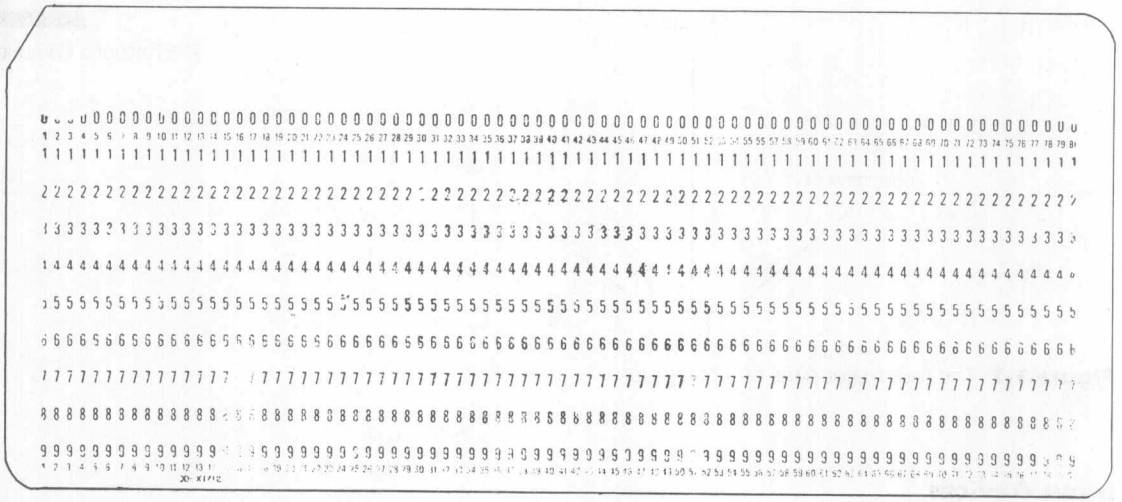


Figure 1-2 A computer card.

ulation. Storage devices permit data or programs to be stored compactly and reread quickly by the computer. Frequently used storage devices include **cards, disks, tapes,** and **core.** You can save and retrieve information from any one of these storage devices.

If you typed your data or program onto cards, you clearly retain the physical record of your work. Stored permanently on the cards, your program or data will be retained until you throw the cards away. Tape storage is similar to the storage of music on audio-cassettes. The information is retained until you erase or record over the tape. Disk storage is also similar to the use of audio-cassettes. Physically, however, disks resemble phonograph records. Information stored on disk is saved until you either erase or write over the disk.

Cards, disks, and tape are storage devices that are external to the computer; they are peripherals. There may be times when you want to store information in the computer itself, in what is called core. Core is usually a temporary form of storage. The information is kept in the computer's memory and not on an external device. Thus, you must tell the computer to set aside some memory for your data or program. In terms of speed, it is quicker to run information from core than from disk, from disk than from tape, and from tape than from cards. There is, however, a trade-off between speed and cost. Core costs more than disk. Core space is limited in size, and all computer programs must be run in core. Therefore, though limited in size, core is in high demand. Disk storage costs more than tape, and tape costs more than card storage.

Central Processing Unit

CPU stands for central processing unit. The CPU is the brain of the computer. It is the overall coordinator of a computer system. The CPU performs all the desired manipulations and calculations as directed by the program. For example, the CPU can retrieve information from storage, execute the steps in the program, and output the program's results to the proper device. The CPU ensures that the program is run and that the steps are followed in the sequence specified by the programmer.

Output Devices

Most input devices can also be used as output devices. After running a program, you want to see the results. Output devices allow you to view results. Typical output devices include cards, terminals, and line printers. You may tell the computer to do some manipulations or calculations and punch the results on cards. Or you may sit at a terminal and have the computer print the information on either a screen (CRT) or a line printer. If the line printer is not attached to your terminal, you will probably have to walk to a computer science building to pick up the hardcopy. A typical line printer prints 1000 lines per minute. Some printers produce more than 20,000 lines per minute! The printed output is called hardcopy or **printout**. Review figure 1-1 to see the links among the four parts of the computer: input devices, storage devices, CPU, and output devices.

Computer Programs

Throughout this chapter I've referred to computer programs, sets of sequential instructions telling the computer exactly what to do. How do you tell a computer what to do? Computers must be given instructions in a language they understand. In the United States, we speak English. If we ask people to do something in French or German, usually nothing happens, because the communication is not understood.

The same language barrier exists for our communications with the computer. Languages such as FORTRAN, COBOL, and BASIC were developed to allow us to communicate with the computer. Some of you may be thinking that you're having enough trouble learning French or Russian without having to learn a computer language too. Don't worry—you don't need to learn a computer language to communicate with the computer. A special kind of program, a prepackaged program, allows us to "talk" to the computer in English! We will be working with the prepackaged program called SPSS^{X}, the Statistical Package for the Social Sciences.

Before turning to SPSS^{X}, however, we need to examine the idea of data in more detail. After all, we write programs to store, retrieve, manipulate, and do calculations on data. The next chapter focuses on so-

cial science data: what it is and how to input it into the computer. In chapter 3 we begin SPSS^x programming.

CHECKPOINT

1. Match an entry in column A with one or more entries in column B. Answers to most checkpoints are located in the back of the book, but don't look! Try the exercise first.

A	B
a. Keypunch machine	**1.** Displays output on a screen
b. Card reader	**2.** An input device
c. CRT terminal	**3.** Punches holes in cards
d. Disk	**4.** A storage device
e. Tells the computer to calculate an average or print a table	**5.** Hardware
	6. Software

Check the one *best* answer for each of the following questions.

2. A person who writes software is called a(n)
   - **a.** author
   - **b.** programmer
   - **c.** program
   - **d.** writer

3. Which of the following devices is *not* external to the computer?
   - **a.** peripheral devices
   - **b.** disks
   - **c.** CPU
   - **d.** All of the above are external to the computer.

4. It is quicker to run information from
   - **a.** core than disk
   - **b.** disk than core
   - **c.** tape than core
   - **d.** card than disk

5. The part of the computer that actually performs the manipulations and calculations is called the
   - **a.** storage device
   - **b.** input device
   - **c.** output device
   - **d.** central processing unit