STATISTICS: textbooks and monographs volume 33

STATISTICAL COMPUTING

WILLIAM J. KENNEDY, Jr.
JAMES E. GENTLE

STATISTICAL COMPUTING

William J. Kennedy, Jr.
Department of Statistics
Iowa State University
Ames, Iowa

James E. Gentle
ISML, Inc.
Houston, Texas

MARCEL DEKKER, INC. New York and Basel

Library of Congress Cataloging in Publication Data

Kennedy, William J., Jr., [Date]
Statistical computing.

(Statistics, textbooks and monographs ; v. 33)
Includes index.
1. Mathematical statistics--Data processing.
I. Gentle, James E., [Date] joint author.
II. Title.
QA276.4.K46 519.5'028'54 80-10976
ISBN 0-8247-6898-1

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

The purpose of this book is to present material that is generally considered to be in the area called *statistical computing*. Many numerical methods and algorithms are discussed from a computational viewpoint, and techniques for implementing given algorithms in a computer are often considered. The diversity of statistical computer applications is reflected in the book through the various chapters, each of which deals with a relatively general but unique application area. This book is designed to serve both as a textbook at the beginning graduate level, and as a reference text for people interested in computer applications of statistics. Exercises are provided at the end of most chapters for the student, and an extensive list of references is given on each subject to benefit all readers.

Much of the material in this text forms the basis for the Statistics 580 and 581 courses that have been presented at Iowa State University for the past several years. This two-quarter course sequence is taken by graduate students in statistics, computer science, mathematics, and related disciplines. Emphasis in these courses, and in this book, is on numerical methods in statistical computing. Seminumerical and nonnumerical topics, such as computer graphics for example, are not discussed in detail.

The authors are indebted to many people for assistance which led to the completion of this text. All cannot be mentioned in this short space, but especially we wish to thank Professors H. O. Hartley, W. J. Hemmerle, and C. E. Gates for major contributions to our education in statistical computing. We also thank Dr. T. A. Bancroft and Dr. H. A. David for their continuing energetic support of statistical computing which has provided the atmosphere and resources needed for growth in this subject area at ISU. Thanks also go to Professors V. A. Sposito, T. J. Boardman, and W. J. Hemmerle who read parts of the manuscript and made valuable suggestions. Finally, we thank Valerie Engeltjes for the excellent typing support that she provided through several revisions of the original manuscript; and Joyce Johnson, Marlene Sposito, and Darlene Wicks for preparation of the final camera-ready copy.

William J. Kennedy, Jr.
James E. Gentle

CONTENTS

1 / INTRODUCTION

1.1 ORIENTATION

During the past few years a tremendous amount of computer software has been developed to support statistical computing requirements. One significant feature of the software which has evolved over the years is its ever-improving quality in terms of capability, efficiency, and reliability. Much of this improvement can be traced to the development of better numerical methods, algorithms, and programming techniques along with an awareness of the characteristics of computer floating-point arithmetic.

Top-quality computer software for statistical applications is not easily produced. Development begins with a careful choice of the best among competing numerical methods. Selected methods are then specified in detail, with an eye toward computer implementation, in the form of algorithms. Each algorithm must subsequently be carefully coded in a computer language to form the nucleus of the finished software package. Additional considerations such as ease of use and flexibility are important to the overall makeup of the software. Decisions in this area will usually lead to the use of other methods and associated algorithms, many of which may be semi- or nonnumerical in nature. These latter algorithms must also be carefully coded to insure optimal performance of the final product. Thus both numerical and nonnumerical methods and associated algorithms form a basis for computer software, and careful implementation of selected algorithms is a vitally important aspect of the developmental process.

The need for new and improved methods and algorithms is never-ending. Research workers in statistics and related sciences are continually producing new results which require computational support. In many cases an algorithm to support a new computing requirement can be formed by modifying and/or specializing some existing algorithms. In other cases an extension of some numerical method may lead to a desirable algorithm. The most difficult situation is one in which a totally new numerical method must be derived to satisfy the computing requirement. Fortunately, this is not the usual case.

1.2 PURPOSE

The purpose of this book is to present selected computational methods, algorithms, and other subject matter which is important to the effective use of the given methods and algorithms in the computer. Most of the computational methods and algorithms contained in this book have been used in one or more software systems. There is usually more than one "good" method, and possibly several algorithms for each method, for handling a given problem, and the reader will find this fact reflected frequently in the chapters of this book.

Since the literature in statistical computing is extensive, and since it is not possible in the space of this text to treat every subject area and all important aspects within the area, the references listed at the end of each chapter are an important part of the chapter. The authors have attempted to include an extensive list of publications (including many that are not cited in the body of the chapter) which will allow the reader to move quickly to an in-depth study in any of the major areas defined in this text. The literature in statistical computing is to be found primarily in journals on statistics, computer science, and numerical analyses. A list of the most frequently referenced journals in the field and the abbreviations used for each is given in Table 1.1. Proceedings of conferences on statistical computing also contain many relevant articles. The annual Symposium on the Interface of Statistics and

Table 1.1 Journal Titles and Abbreviations

Abbreviation	Journal Title
JACM	*Journal of the Association for Computing Machinery (ACM)*
CACM	*Communications of the ACM*
TOMS	*ACM Transactions on Mathematical Software*
Comp. Rev.	*ACM Computing Reviews*
Comp. Sur.	*ACM Computing Surveys*
Numer. Math.	*Numerische Mathematik*
SIAM J.[a]	*Journal of the Society for Industrial and Applied Mathematics (SIAM)*
SIAM J. Appl. Math.	*SIAM Journal on Applied Mathematics*
SIAM Rev.	*SIAM Review*
SIAM J. Num. Anal.	*SIAM Journal on Numerical Analysis*
MTAC[b]	*Mathematical Tables and Other Aids to Computation*
Math. Comp.	*Mathematics of Computation*
JSCS	*Journal of Statistical Computation and Simulation*
Ann. Stat.	*Annals of Statistics*
AMS	*Annals of Mathematical Statistics*
JASA	*Journal of the American Statistical Association*
Amer. Stat.	*The American Statistician*
Techno.	*Technometrics*
Comm. Stat. (A)	*Communications in Statistics, Part A*
Comm. Stat. (B)	*Communications in Statistics, Part B*
JRSS (A)	*Journal of the Royal Statistical Society, Series A*
JRSS (B)	*Journal of the Royal Statistical Society, Series B*
Appl. Stat.	*Applied Statistics (Also called JRSS, Series C)*
Comp. J.	*Computer Journal*

Table 1.1 (Continued)

Abbreviation	Journal Title
BIT	*Nordisk Tidskrift for Informationbehandling*
Computing	*Computing*
J.O.T.A.	*Journal of Optimization Theory and Applications*
J. Inst. Maths. Applics.	*Journal of the Institute of Mathematics and its Applications*

[a] The journal title *SIAM Journal* was used until 1966 when the name was changed to *SIAM Journal on Applied Mathematics*.

[b] This title was used from 1943 through 1959. In 1960 the title *Mathematics of Computation* was adopted.

Computer Science, the proceedings of which are published, has been a major factor in determining the growth of the field of statistical computing. The Statistical Computing Section of the American Statistical Association sponsors sessions at the annual meetings of the ASA, and the proceedings of these sessions are also published.

Exercises are given at the end of most chapters. The purpose of each exercise will be either to reinforce or to extend the material given in the chapter. None of the exercises is intended to be extremely difficult to solve.

1.3 PREREQUISITES

This book is designed for use by readers who have background in statistics equivalent to undergraduate training in statistical methods and theory; mathematical training in elementary calculus, matrix algebra, and numerical analysis; and a familiarity with the FORTRAN computer-programming language.

A complete discussion of an algorithm and its implementation on the computer often requires consideration of details which are dependent on computer hardware characteristics, the computer language used, and in some cases the language compiler. In situations in which computer hardware specifications are pertinent to the discussion, this text assumes use of the IBM 370 series computer. Needed hardware specifications for this machine are given in Chapter 2. This reduces what would otherwise be a much larger prerequisite in applied computer science, and it gives the user of machines other than the IBM 370 (or equivalents) a basis for translating the text material to another machine-compatible form. There are only a few cases where the discussion is affected by the specific machine, the major ones being pseudorandom number generation and error analysis.

1.4 PRESENTATION OF ALGORITHMS

Many different computational methods will be described and discussed in succeeding chapters. In some instances an algorithm associated with some given computational method will be provided. When such is the case, and the algorithm is not stated in the form of a FORTRAN program, then the general form of the statement of the algorithm will be as follows. The first line will contain the word *"ALGORITHM"* followed by the name given to the algorithm and a FORTRAN-like argument list, enclosed in parentheses, containing input and output quantities. For example, a first line might be

```
    ALGORITHM  RAND(I,J,X)
```

Following the first line, in some cases, will be a short set of comments designed to facilitate overall understanding of the algorithm. Then one or more numbered steps will be given along with a description of the operations to be performed in each step. These will take the form

```
Step        Description
 1.      Set I = I*J
 .              .
 .              .
 .              .
```

Comments will also appear between the steps in the algorithm as often as they are needed to explain the process.

When algorithms are stated in a programming language, such as FORTRAN, the meaning of the "=" symbol as an assignment of value is clear. However, there is some possibility of confusion when the "=" sign is used in the statement of algorithms using other than a programming language. The question may arise, in a specific situation, as to whether the "=" symbol is specifying an assignment of value or whether a statement of equality is intended. Sometimes the symbol "← " or ":=" is used to denote assignment of value. This will not be done in this text. The context in which the symbol "=" is used will serve to define whether an assignment of value or a statement of equality is intended. If there is room for doubt, a comment will be made to clarify the situation.

2 / COMPUTER ORGANIZATION

2.1 INTRODUCTION

The purpose of this chapter is to describe the computer in general terms and consider some of the internal computer operations which are of special interest to the user in scientific applications. The computer is a tool to be used in performing calculations, and a basic understanding of how this tool operates is essential.

2.2 COMPONENTS OF THE DIGITAL COMPUTER SYSTEM

The computer system may be thought of as being made up of five components: (1) the main storage, (2) a control component, (3) an arithmetic and logic component, (4) an input component, and (5) an output component. Figure 2.1 depicts the way in which these components are interconnected.

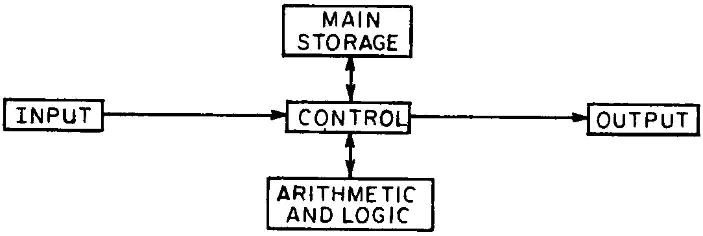

Figure 2.1 Components of the computer system.