APL-STAT

A Do-It-Yourself Guide to Computational Statistics Using APL

James B. Ramsey • Gerald L. Musgrave

APL-STAT

A Do-It-Yourself Guide to Computational Statistics Using APL

James B. Ramsey
New York University

Gerald L. Musgrave

LIFETIME LEARNING PUBLICATIONS
Belmont, California

A division of Wadsworth, Inc.

In preparing APL-STAT we were fortunate to have the help of many friends and colleagues. Rather than attempt to explain their individual contributions we simply list their names and express our thanks to each of them: Bert Alexander, Alea Curtis, Dorothy Dixson, David Edelman, John Hause, Robert Hessen, John Kassionas, Jan Kmenta, Alexander Kugushev, Charles Moore, Thomas Gale Moore, Jan Musgrave, Richard W. Parks, Virginia Perry, Alvin Rabushka, Grace Ramsey, Shannon Ramsey, Robert Rasche, Bernard Scheier, Bert Schoner, Andy Silver, Barbara Snarr, and Mike Sullivan.

Printed in the United States of America

1 2 3 4 5 6 7 8 9 10—85 84 83 82 81

Library of Congress Cataloging in Publication Data

Ramsey, James Bernard.
APL-STAT, a do-it-yourself guide to computational statistics using APL.

Includes index.
1. Statistics—Data processing. 2. Econometrics
3. Mathematical statistics—Data processing. 4. APL (Computer program language) I. Musgrave, Gerald L., joint author.
II. Title.
QA276.4.R35 519.5'028'5 80-15016
ISBN 0-534-97985-8

Preface

Please Read This Before Reading the Text!

This book explains how to perform both simple and complex statistical calculations using APL. "APL" is an acronym for "A Programming Language"—a computer programming language that is ideal for the computational work done in statistics.

The authors are both economists, and the content reflects their professional interests. However, political scientists, physicists, sociologists, industrial psychologists, public health and dental researchers, and others have used this book and found it helpful.

No previous knowledge of computers, computer programming, or methods involved in statistical computation will be needed to understand this book. You will start from the most elementary statistics and progress to more complicated procedures on a gradual step-by-step basis. The numerous examples, exercises, and statistical applications are drawn from a variety of fields. Emphasis is placed on how to obtain the statistical results with ease. Using this book you will be able to perform computations that otherwise would be so cumbersome or time-consuming that you would not do them. You also will be able to perform experiments and computer simulations with relatively little effort.

The APL statistical procedures presented are useful to researchers, analysts, managers, and anyone concerned with statistical calculations. We believe that when you have seen how easy it is to perform these computations, you will be as pleasantly surprised as we were. If you are familiar with computers here is a dramatic example of the simplicity of APL compared to the FØRTRAN statements used to compute the arithmetic mean. If you are a novice in these things don't be frightened—everything will be explained.

An Example of FØRTRAN and APL

```
      FØRTRAN                          APL
      DIMENSIØN X (1000)               X←⍞
      READ (5,99)N                     ⍞←AVE←(+/X)÷⍴X
99    FØRMAT (I4)
      READ (5,100) (X(I), I=1,N)
100   FØRMAT (9F8.0)
      SUM = 0.0
      DØ 10 J = 1,N
10    SUM = SUM + X(J)
      AVE = SUM/N
      WRITE (6,20)AVE
20    FØRMAT (F10.4)
      END
```

To estimate the parameters of $Y = B_1 + B_2 X_2 + B_3 X_3 + \cdots + B_N X_N + u$ via multiple regression, you could type in APL:

Use of ⌹ in Multiple Regression

```
B ← Y⌹X
```

In other computer languages an equivalent program might take 50 statements.

This book is not just an introduction to APL programming, although many people have learned APL from it. Certainly it is not a statistics textbook, but readers have commented that they never really understood certain statistical concepts until they "tried real numbers to see how the formulas worked." This book is a valuable aid to understanding statistics because it actually computes results and even displays probability distributions graphically. By the time you finish you will know a lot about APL programming. And after you spend a few hours at the computer, you will find that it is easier to program your own work than it is to learn to use the "canned" (FØRTRAN) routines available at the computer facility. More importantly, you will understand what you are doing and how the results are obtained. We have long maintained that the less you are asked to accept unquestioningly, the better is your intellectual health and the greater will be your interest in statistical subjects.

This book is not primarily a textbook. It is a book for the person who understands basic statistics, who wants a painless way to compute results, and yet wants to know what is really going on. We think that teachers of basic or applied statistics and especially econometrics will find our approach using APL to be an important part of a practical statistics course. Students are often assigned "artificial," "theoretical," or "academic" problems, situations, and exercises. These assignments are not made because the instructor thinks such things are important. Actually, most instructors understand the difficulty of tackling *real* statistics problems. Consequently, when the amount of computational pain the student (and teacher) must go through to get the statistical result is compared to the "statistics" that can be taught, a stress on pure theory almost always results. Thus, after a course (or even several courses), an individual may be unprepared to solve the first problem—how to perform the calculations! The use of APL minimizes these difficulties.

We think that when you complete APL-STAT you will agree—programming can be easy!

Because the text proceeds in a carefully structured sequence it is important that you follow it exactly and that you make sure you thoroughly understand each section before moving to the next. Later sections assume that prior sections have been mastered. You should do the exercises and check your answers in the back of the book. Above all, you can teach yourself a lot by experimenting, so try it.

Purpose of These Comments

If you forget something, the primitive function glossary at the back of the book will help you recall earlier material. If you need more information, the side of most pages has brief comments. These comments contain the name and symbol of the APL operator introduced on that page. You will be able to flip through the book quickly and locate what you want, using the comments. They also provide a quick visual guide to the major topics in any section.

We have a request. In the back of the book is an error sheet for recording *our* omissions, bad language (though never foul!), and other sins. We would be most obliged if you would send us this error sheet with your comments. The next edition will then be much better with your help.

JAMES B. RAMSEY GERALD L. MUSGRAVE

Note to Instructors

Instructors can assign much more meaningful examples and exercises using the procedures in this book than using either canned programs or hand calculation. Students will not be spending time in tedious calculation or in using the computer as a black box. Students will be able to perform calculations, including complex matrix algebra, know how they are done, and see the numerical results. They will be able to obtain results they understand. One example is where a multiple regression model requires the intercept to be "forced" through zero. It is surprising how simple the mathematics of this is (not having a column of ones in the regressor X matrix). It is also surprising how few preprogrammed packages allow this option. In APL you can modify your program to handle this change in a matter of moments.

Computer simulation and generation of distributions become a relatively trivial task in the hands of an APL-proficient student. We could enumerate a long list of such examples, and once you start you will see them too. Also, we have included our benchmark program data on the Longley regression problem in Appendix B. You may find it interesting to compare the computational accuracy of APL programs with the canned ones on your home computer or at your computer center.

In using this book as a text you might consider the following ideas. The titles of certain sections, e.g., *The Normal Distribution* in Chapter 6, are starred. These starred sections involve mathematical material which may be beyond the scope of an elementary course in statistics that doesn't have a mathematical prerequisite. Any APL instructions introduced in such sections will not be used anywhere else in the text without reexplanation. So starred sections can be dropped without fear of losing some important information about APL.

The book is carefully structured in that it follows the usual pattern of topics in the introductory statistics course and only uses as much APL as is needed to get the job done. Consequently, it is important that, except for the starred sections, the sequence be followed and sections are not skipped.

If you decide to alter the presentation of statistical subjects, have your

students read the APL-material in sequence, even if they skip the earlier presentations of the statistics. A number of readers have used this approach and found it to be satisfactory. In these cases the readers either knew statistics or were not interested in statistics per se. They wanted to learn APL and found this approach to be effective. One reason for this is that APL instructions are introduced to solve specific problems rather than presented in the abstract.

Each chapter has a large number of exercises and applications. The exercises help in exploring the use of APL concepts, functions, and symbols. The statistical applications help extend the depth and breadth of APL use. Throughout the book, experimentation is encouraged to expand and intensify interest and understanding.

An elementary nonmathematical course in statistics would usually stop at Chapter 9, which covers contingency tables, analysis of variance, and simple linear regression with one regressor. Chapters 10, 11, and 12 introduce various aspects of matrices and prepare the way for multiple linear regression analysis and topics that might be regarded as more "econometric." You may find that the use of APL will allow you to cover Chapters 10 through 13 as well. This is important since the rudiments of matrix algebra can be taught quickly using APL. The benefit will be that you can enable your students to master multiple linear regression and more complicated analysis of variance techniques more easily.

Three administrative matters might be of interest. Many computer centers have only a few APL terminals. Don't let this apparent difficulty slow you down. First, if the terminals use a typing ball or a daisy wheel, the center can obtain APL balls or print wheels. They are easy to switch, are low in cost, and small adhesive labels are available for the keys. Second, if the terminals use a non-APL matrix printer or if the terminals are CRT's without APL characters, another solution is available. A Mnemonic character set that substitutes for the APL symbols is available. The multiple regression example in the preface was coded as

```
Y←Y⌹X
```

using the standard APL character set. In the Mnemonic character set it would be written as

```
Y←Y.DQX
```

Appendix C contains both the standard and Mnemonic character sets. Third, some computers have implemented only the monadic version of domino. In this case you simply enter the following two lines

```
∇YDQX
(⌹((⍉X)+.×X))+.×((⍉X)+.×Y)∇
```

when you enter `YDQX` the result is the same as if `Y⌹X` had been entered.

If in using the book you have any comments that would be helpful to others please pass them along to us and we will incorporate them in the next edition.

Contents

1

Introduction

1.1 Overview of APL

APL is a powerful and versatile computer programming language. When you use this language to communicate with the computer it will be as if you were personally operating the machine. APL is designed to operate on small microcomputers no larger than a typewriter, on minicomputers the size of one or two office desks, and on large maxicomputers the size of a truck. No matter how large or small the computer, once you log-on to the system it will appear from your perspective that you have a one-to-one relation with the computer. The APL contained in this book has been used on micro-, mini-, and maxicomputers produced by a variety of manufacturers. We found the APL language to be remarkably similar in all of these cases.

Administrative Procedures

The procedures used to log-on to the various systems that we have used vary greatly. Each computer center has its own administrative procedures, keywords, passwords, and account verification methods. In addition, you usually need to connect your computer terminal to the computer itself and this process can be mysterious at first. There is really nothing to this at all. Nevertheless, sometimes people who hang around computer centers make a big deal about the administrative and technical matters surrounding the use of the machine. The truth of the matter is that the procedure is much the same as getting a key for an office, registering for a class, or signing up for Little League. It's a hassle. Every organization thinks that there is only one way to do it, and yet every way is different. Appendix A contains a brief description of how it is done at the Stanford and NYU computer centers, and on an IBM 5120 desk-top computer. This description should

allow you to understand better the procedures that are used with your computer. In a short time the mystifying intricacies of gaining access to the computer become second nature. You type a few words and numbers and you are ready to go.

The APL Keyboard

We have included a few diagrams of typical APL keyboards in Appendix A. The alphabetic characters are in exactly the same position as they are on a standard typewriter. These letters are all capitals but (wouldn't you know it) they are in the lowercase positions. Holding the shift key down while pressing a specific key enters a special APL symbol. Each of these symbols performs a specific operation in APL. As you can see, the keyboards are almost identical, and in the very few instances where some minor differences do exist we will explain them. One of the most frightening things that the new APL programmer encounters is the APL character set. All those strange symbols are indeed foreboding. However, our experience has been that the symbols are easy to learn. They are not much more difficult to learn than the international road signs, especially if you take them one at a time in the context of an actual problem.

Some General Features of APL

CLEAR WS

Now suppose that you are sitting in front of the keyboard and you have logged-on. The computer has responded with the message *CLEAR WS*. The computer is indicating that you have been allocated a part of the computer—APL calls it a Work Space—named *CLEAR*. Now you can communicate with the computer, and it is in fact much like an electronic hand-held calculator except that it is much more powerful. To turn off the computer you simply type *)OFF*, for example, and log-off. You will soon see how APL can be used as a very powerful calculator in the immediate execution mode. However, it can do much more.

)OFF

You can define a set of instructions that will perform tasks such as balancing a checkbook; computing means, standard deviations, and regression coefficients; or directing the computer to simulate a Las Vegas casino game. In APL, the set of instructions is called a defined function. After the function has been defined you simply refer to it by name. The same instructions, operating on different data, can be used over and over again.

State Diagram

Figure 1.1 is a state diagram that represents these three APL modes. When you log-on you are given a clear work space, you are in immediate execution mode, and you have a powerful calculator at your disposal. You can enter data, process the data with a one-line APL expression, define an entire new work space with different functions and data, and test your functions on a line-by-line basis before you program the whole set of instructions.

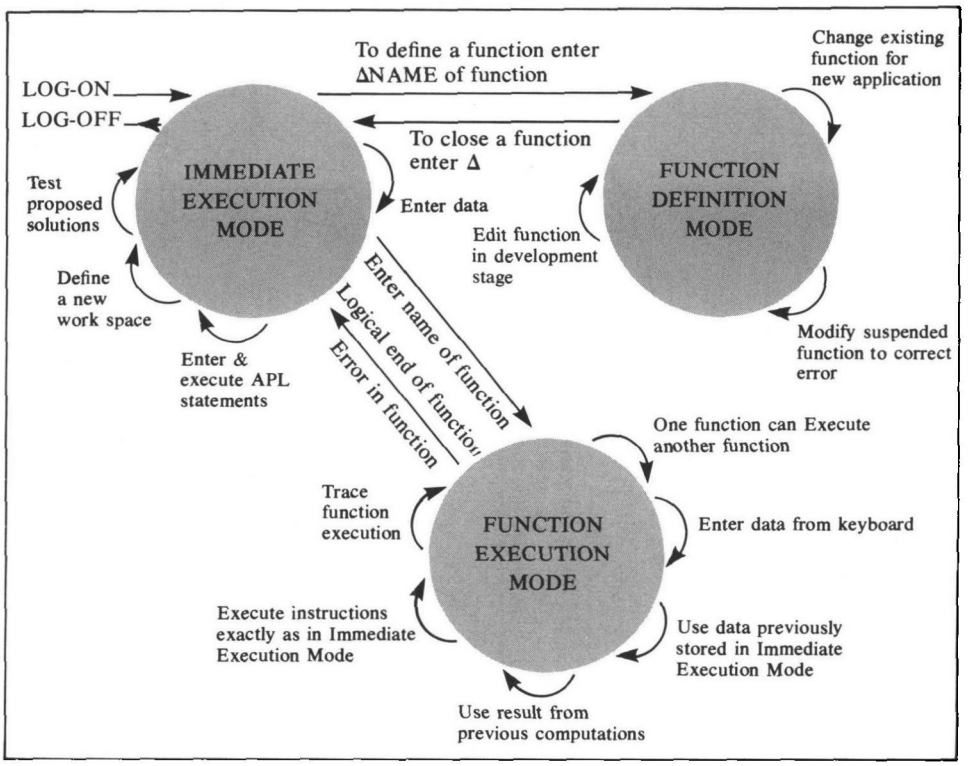

Figure 1.1

Then you can define your own function, edit any part of it, or modify it for a particular application. Also, should a function stop because of a programming error and further processing thereby be suspended, you can correct the error by editing the function and then resume the function's execution from the point of suspension. You need not start from the beginning if your previous calculations were correct.

A function is executed by simply entering its name. You can specify the particular data set to be processed, and your function can call other functions, request data, and produce results for use by other functions. In addition, you can trace the execution of your function by having the results of any line or group of lines displayed—all of this without having to write any output statements. When your function's execution is completed it returns you to immediate execution mode where you began. We hope that this sounds simple, straightforward, and like something you can do—because it is!

1.2 Road Map of Where We Are Going and How We Will Get There

In the next chapter you will learn how to use APL as a calculator. After these basics are under your belt, the general presentation is to explain a statistical problem and then to solve it using APL. On the way to the solution the various APL functions and programming methods are presented and explained. We first discuss the sample mean and median, stan-

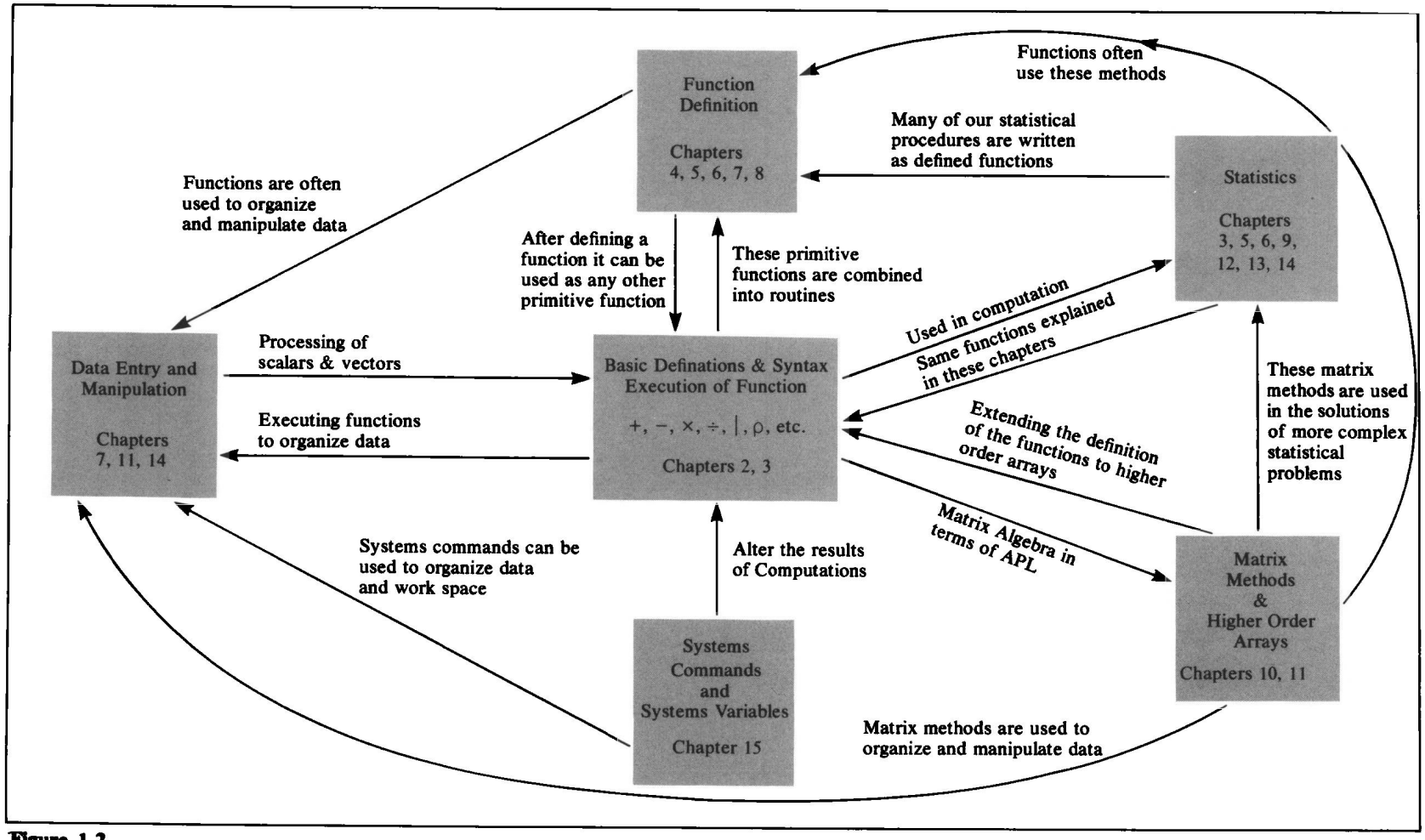
Function Definition
Chapters 4, 5, 6, 7, 8
Functions often use these methods
Many of our statistical procedures are written as defined functions
Statistics
Chapters 3, 5, 6, 9, 12, 13, 14
Functions are often used to organize and manipulate data
After defining a function it can be used as any other primitive function
These primitive functions are combined into routines
Used in computation
Same functions explained in these chapters
Data Entry and Manipulation
Chapters 7, 11, 14
Processing of scalars & vectors
Executing functions to organize data
Basic Definations & Syntax Execution of Function
+, −, ×, ÷, |, ρ, etc.
Chapters 2, 3
Extending the definition of the functions to higher order arrays
These matrix methods are used in the solutions of more complex statistical problems
Matrix Algebra in terms of APL
Systems commands can be used to organize data and work space
Alter the results of Computations
Matrix Methods & Higher Order Arrays
Chapters 10, 11
Systems Commands and Systems Variables
Chapter 15
Matrix methods are used to organize and manipulate data

Figure 1.2

dard deviation, covariance, and higher order moments. Then we investigate a number of the most prominent statistical distributions, including the binomial, Poisson, and normal density and cumulative distribution functions. After you learn how to handle more complex data structures in APL and to write more general and powerful functions you will learn how to diagnose and correct programming errors. After you go through a case study using APL in a research project, we present an introduction to elementary linear correlation and regression, analysis of variance, and the chi-square and F distributions. Next we show how to do matrix algebra in APL, including the operation of matrix inversion, which is performed with one symbol, ⌹. Multidimensional arrays are discussed in Chapter 12, where the various APL functions are explained in relation to these higher order arrays. The final chapters concentrate on computational statistics related to multiple linear regression, two-stage least squares, instrumental variables, Aitken estimators, Durbin's First Order Autoregressive Models, and K-class estimators including limited information maximum likelihood estimators.

Don't let this impressive sounding jargon put you off. The first half of the book has been understood by good high-school students, and they were able to write APL programs after only a few hours of study. The later chapters have been used in both undergraduate and graduate classes. Also, the statistical routines have been used by a number of our colleagues in their statistical research. So you can see that while much of the material is technical, it progresses at a measured rate. Figure 1.2 is a schematic representation of APL-STAT. It might help you to visualize how the various components of APL are related.

We can summarize our position this way:

APL
TRY IT—YOU'LL LIKE IT

So turn the page and let's go . . .