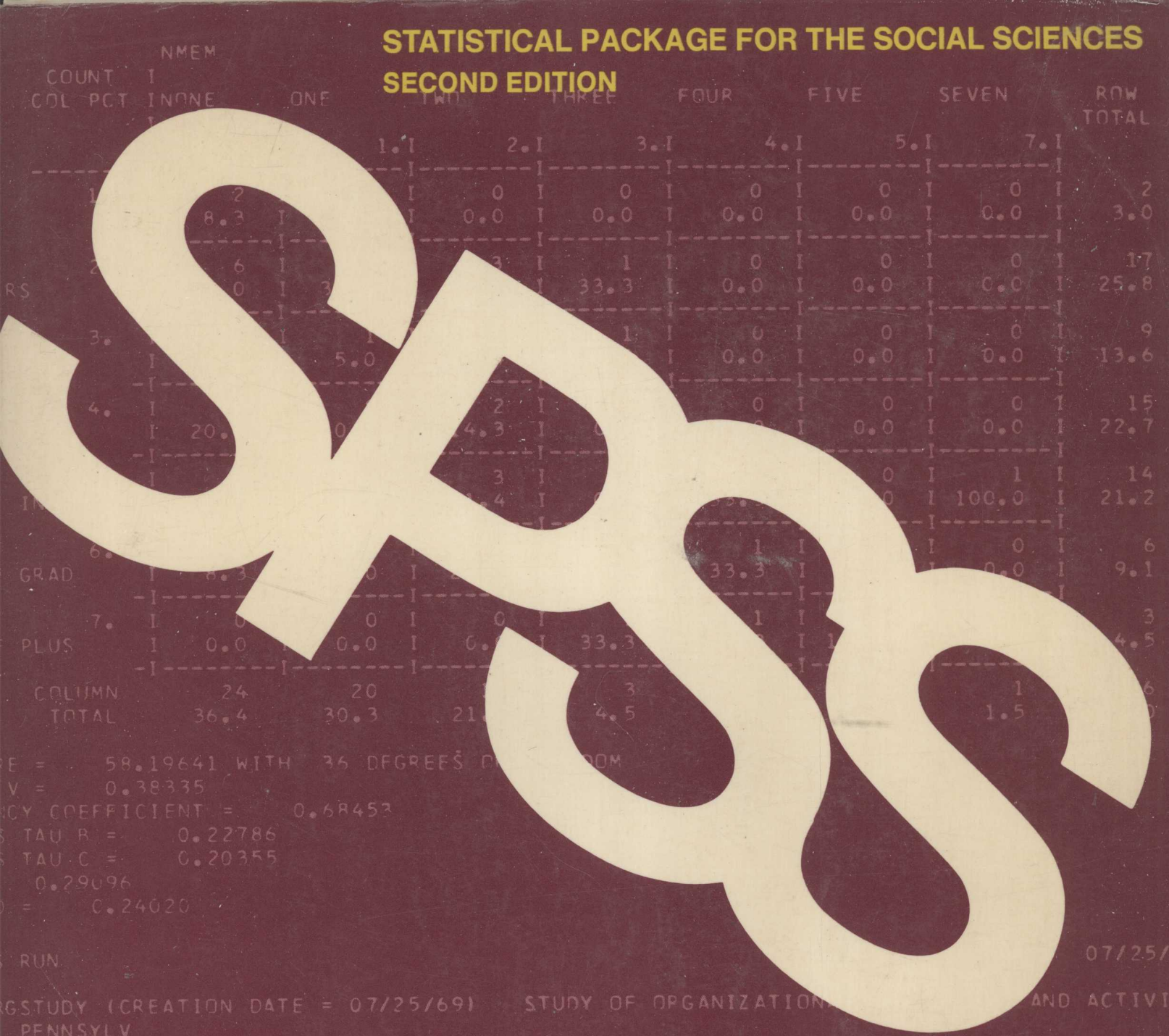
STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

SECOND EDITION



NORMAN H. NIE
C. HADLAI HULL
JEAN G. JENKINS
KARIN STEINBRENNER
DALE H. BENT

SPSS

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

SECOND EDITION

NORMAN H. NIE
Department of Political Science
and
National Opinion Research Center
University Of Chicago

C. HADLAI HULL
Computation Center
University of Chicago

JEAN G. JENKINS
National Opinion Research Center
University of Chicago

KARIN STEINBRENNER
National Opinion Research Center
University of Chicago

DALE H. BENT
Faculty of Business Administration
and
Computing Services
The University of Alberta

McGRAW-HILL BOOK COMPANY

New York
St. Louis
San Francisco
Auckland
Düsseldorf
Johannesburg
Kuala Lumpur
London
Mexico
Montreal
New Delhi
Panama
Paris
São Paulo
Singapore
Sydney
Tokyo
Toronto

NOTE TO THE EIGHTH PRINTING: Since the writing of this manual, which documents Release 6, SPSS Inc. has issued Releases 7 and 8, adding new facilities to the package and implementing significant improvements. New subprograms include MULT RESPONSE for multiple-response variables, NPAR TESTS for nonparametric data, RELIABILITY for computing reliability coefficients for multiple-item scales, SURVIVAL for life tables and associated graphs and statistics, and REPORT for generating publishable documents. Major improvements include faster operation or more precise calculations in most subprograms, enhanced data-management capabilities, and a more flexible facility for sorting cases. In addition, new facilities have been added for interaction between SPSS and SCSS, the SPSS Inc. conversational system. These and other changes are documented in SPSS UPDATE MANUAL, available from McGraw-Hill Book Company.

The statistical algorithms used in SPSS programming are now fully documented in SPSS STATISTICAL ALGORITHMS, available from SPSS Inc., Suite 3300, 444 North Michigan Avenue, Chicago, IL 60611. A POCKET GUIDE TO SPSS syntax, options, and statistics, also available from SPSS Inc., is a valuable companion to the larger manuals for users already familiar with SPSS.

SPSS and SCSS are the trademarks of SPSS, Inc. for its proprietary computer software. No materials describing such software may be produced or distributed without the written permission of the owners of the trademark and license rights in the software and the copyrights in the published materials.

Library of Congress Cataloging in Publication Data
Nie, Norman H

SPSS: statistical package for the social sciences.

1. SPSS (Electronic computer system) 2. Electronic data processing—Social sciences. I. Title. II. Title: Statistical package for the social sciences.

HA33.N48 1975 029.7 74-16223
ISBN 0-07-046531-2 (pbk)
ISBN 0-07-046532-0 (hbk)

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES

11 12 13 14 15 16 17 18 WCWC 7 9 8 7 6 5 4 3 2 1 0

This book was set in Times Roman by Creative Book Services, division of McGregor & Werner, Inc. The editors were Kenneth J. Bowman and Matthew Cahill; the designer was Joseph Gillians; the production supervisor was Thomas J. LoPinto. New drawings were done by J & R Services, Inc.
Webcrafter, Inc. was printer and binder.

PREFACE

Statistical Package for the Social Sciences (SPSS), the system of computer programs described in this volume, represents nearly a decade of systems design and programming and documentation on the part of the authors and others. When the first edition of this manual was published in 1970, SPSS was being used at approximately 60 installations. It is now being used at nearly 600 installations, including conversions to almost 20 different operating systems and computers. The current version of SPSS has almost double the amount of statistical procedures and data-management facilities as were documented in the first edition. Needless to say, we are thankful for the opportunity to continue developing SPSS and to serve an ever-increasing audience.

SPSS has been a success, much more so than the authors ever expected. We feel that the major reason for this success is that SPSS was developed through the close cooperation of three types of specialists: practicing social science researchers, computer scientists, and statisticians. At each stage in the development of SPSS we have attempted to satisfy these criteria:

1 That the statistical procedures be mathematically and statistically correct
2 That the program design and code be computationally efficient
3 That the logic and syntax of the system parallel the way in which social scientists approach data analysis
4 That the system provide statistical procedures and data-management facilities tailored to the particular needs of empirical social researchers

Without the contribution of experts in each field SPSS could not have effectively satisfied these goals.

Vying for importance in the success of SPSS has been our continuing concern for documentation. While many of our users are very sophisticated social scientists, they are not necessarily sophisticated statisticians and computer specialists, nor should they be expected to have the time or training necessary to master these fields. We feel that SPSS documentation must be comprehensive enough to enable students and researchers alike to use SPSS accurately

and efficiently. We hope that even someone who has no experience with computers will be able to run SPSS jobs successfully by using this manual. Furthermore, since the SPSS system is used in many social science methods courses, we feel it is important that the manual not only explain how to invoke the statistical procedures but also provide brief discussions of the various statistical techniques available. Of course, in attempting to satisfy these various needs the manual has grown to be a very large book indeed. However, we hope that the organization of the manual as well as the inclusion of an index will enable users to find what they need conveniently.

Finally, the availability within SPSS of a wide variety of data-management facilities and statistical procedures simplifies the process of data analysis. It avoids many difficult, time-consuming, and generally unrewarding tasks involved in using a variety of single-purpose computer programs, each with its own idiosyncratic control-card syntaxes and input data formats. The researcher thus spends less time as a data-preparation clerk and more time as a social scientist analyzing substantive results.

The development of SPSS began in 1965 at Stanford University as a result of the frustration felt by several of the authors in trying to serve the research and teaching needs of the political science department and the Institute for Political Studies by a library of single-purpose data analysis programs. We and the users suffered the annoyance of having to learn how to operate many different programs and endured the time-consuming task of transferring data and results between essentially noncompatible programs. Furthermore, the documentation for many of these programs ranged from cryptic to nonexistent. That and the fact that the programs were written in many different languages made them extremely difficult to maintain. Finally, while many statistical analysis programs were available, we were forced to write our own programs to perform even the simplest recoding, data transformation, file editing, and other routine housekeeping chores that are essential in social science data analysis. As a result of our experiences we began to design an integrated system that would automate the routine tasks of data processing and around which a series of statistical programs could be built. The basic data-modification, file-handling, and data-description facilities were programmed, and as time permitted statistical analysis procedures were added.

The system developed into SPSS as it was described in the first edition of the manual. Since that time we have followed our plan of incremental development. New facilities have gradually been added and many flaws in the original design have been corrected. SPSS users have submitted suggestions and even complete procedures which have been incorporated into the system. The update manuals that we published each year finally became so unwieldy that we were compelled to provide this second edition of the manual so that SPSS would once again be fully described in a single volume.

The current version of SPSS still has a number of deficiencies, but we feel that it meets a great many of the needs of social science data analysis. SPSS has always been considered an open-ended system, and we have again included a programmer's guide to SPSS (Appendix I) in the hopes that users who wish to add statistical procedures to the system will do so. Because of the design of SPSS, all future programs incorporated into the system can take complete advantage of the capabilities for file maintenance and data handling which exist in the package; this should be an incentive to those contemplating the addition of other statistical programs.

For some time now we have felt that one of the most serious drawbacks of the current SPSS system lies not in the lack of a wider variety of statistical techniques but in that it operates only as a batch program. We and others have had a great deal of success running SPSS through a text-editing and remote batch-entry system, enabling the user to prepare and enter all jobs from a remote terminal and to retrieve and print small jobs directly on the terminal. However, though this adds convenience, it in no way substitutes for a true conversational statistical analysis package. Consequently, the SPSS project staff is now developing a conversational version of SPSS.

The difference between performing analysis in a batch system and a conversational system can be compared to the difference between interacting with another person by letter and by phone. With a conversational system the researcher is brought into close contact with the data being analyzed; results from a statistical procedure are returned on the researcher's terminal in a

few seconds and may be used immediately to guide the next step of the analysis. This facilitates the iterative process of investigative research in which ideas are tested on the data, the results suggest new ideas and modifications of old ideas, the new ideas are then tested, and so forth. A train of thought which could take days (and many trips to a perhaps distant computer center) to explore in a batch system can be explored in a single interactive session. Thus the researcher's concentration and interest are focussed on the relationships being investigated. At the same time that a conversational system facilitates pursuing a single line of thought, the limited printing speed on remote terminals discourages the grand fishing expeditions and the production of excessive amounts of output that plague research in a batch environment. Finally, a good conversational system is somewhat self-teaching; it can guide the user in formulating the correct requests, and the almost instant feedback that facilitates the use in performing statistical analysis is also invaluable in identifying syntax errors and enabling the user to correct them immediately. In short, we feel that a conversational statistical analysis program could help revolutionize empirical social research.

Conversational SPSS, in addition to including conversational statistical analysis procedures, will feature interactive data- and file-definition capabilities, expanded data-transformation facilities, and ultimately the capacity to process hierarchical and other complex file structures. Although we are very enthusiastic about the prospects for conversational data analysis, we do not expect it to completely supplant batch processing. While a conversational system is ideal for investigative research, hypothesis testing, and rapid retrieval of a small number of specific statistics, it cannot be used efficiently for lengthy routine jobs and/or processing files that contain a very large number of data cases. Thus we expect that researchers will wish to use both the batch and conversational versions of SPSS. This will be reflected in the design of conversational SPSS as well as the evolution of batch SPSS—we are attempting to make the two systems useful as a team. Conversational SPSS will be capable of reading and producing batch-SPSS-format system files, and the control statement syntax for conversational SPSS will be as compatible as possible with batch-SPSS conventions.

We are all very excited about the future of SPSS, and we thank you, the users, for the opportunity to continue its development. Your support, enthusiasm, and constructive criticism have been both impetus and reward to us. We dedicate this book to you in the hope that it will make data analysis an easier task, and that in some small way, through your research, we are contributing to the development of the social sciences.

ACKNOWLEDGMENTS

For the past five years the SPSS project has been located at the National Opinion Research Center at the University of Chicago. Throughout this critical period of our development NORC has provided us with financial, organizational, and intellectual support. This support has indeed been generous and has been a major factor responsible for our rapid growth and development. We can think of no other organization which would have provided a better home for us. In particular, we are grateful to Norman Bradburn and James Davis, the past and present directors of NORC, and Associate Director Shelby Orrell, without whose backing we might have floundered in the early years when our financial situation could only be described as precarious. We also wish to thank the University of Chicago Computation Center for donating computer time and allowing C. Hadlai Hull the flexibility to continue his work on the SPSS project.

The name of Patrick Bova, the librarian of NORC, is very familiar to many SPSS users. Mr. Bova has organized and directs the distribution of the SPSS system and all associated literature, and while maintaining our financial records, had thus far managed to keep us both solvent and honest. We are all very grateful for his efforts in what has become a very difficult, complicated, and time-consuming effort. The SPSS project is also very fortunate in the recent addition of ViAnn Beadle, who has joined the staff as user consultant. Ms. Beadle gives competent and conscientious attention to whatever questions and problems users may have. She has taken an enormous burden off our programming staff, providing us with the time for, among other things, the completion of this manual.

Special thanks must be given to Jae-On Kim, who has contributed greatly to SPSS over a period of years; Professor Kim helped design the factor analysis and analysis of variance routines, wrote several chapters of this manual, and regularly consults with us about statistical matters. We also wish to thank William R. Klecka, who contributed to the development of the discriminant function analysis procedure and wrote the associated chapter; William C. Mitchell, who designed the multiple-regression processing algorithms; and Bill Reynolds of the University of North Carolina, who designed the SCATTEGRAM procedure. For various improvements and corrections we are thankful to David Muxworthy of the University of Edinburgh, David Specht of Iowa State University, Jonathan B. Fry and Michael Klein of Akron University, Alfred J. Tuchfarber of the University of Cincinnati, Philip Burns of the University of Illinois, Cleve Moler of the University of Michigan, Harold W. Gugel of General Motors, and William L. McKeown of Stanford University. Martin L. Levin of Emory University and Kisun Han of DUALabs have developed SOS, the SPSS-Override System, which enables SPSS to process hierarchical files of the structure found in the Public Use Sample files.[1] Finally, we thank all those installations and users who have contributed suggestions and procedures. Even those which have not been implemented have provided us with ideas and thus have influenced the development of SPSS.

We are very appreciative of the work of Susan Hull, who prepared all of the more than 500 control-card and output examples appearing in this manual, and who implemented the integer version of the BREAKDOWN procedure. We must add that she performed these tasks as a volunteer, yet was as dedicated and painstaking as any other member of the staff. Another large contribution to the completion of this manual was made by Karin Donker, who typed and retyped until she nearly went mad, but who nevertheless cheerfully worked evenings and weekends when we had deadlines to meet.

We are indebted to the Vogelback Computing Center at Northwestern University for continued support of SPSS and in particular to James Tuccy of that center. He not only manages the CDC-6000 version of SPSS but also designed and programmed a number of the facilities described in this manual, including the CANCORR, DISCRIMINANT, ONEWAY, and T-TEST procedures. Space prevents us from individually mentioning each of the other SPSS conversions and the conversion managers, but we are grateful to them, as, we are sure, are their users.

We also express our appreciation to those at Stanford University who supported SPSS during its early development there: the political science department and the Institute of Political Studies, Professor Sidney Verba, Director of the Cross-National Program on Political and Social Change, and Professor Heinz Eulau and Kenneth Prewitt of the City Council Research Project. We thank the Stanford Computer Center for donating computer time. We are also grateful to the many students and researchers both at Stanford and at the University of Chicago who patiently used test versions of SPSS.

A system like SPSS is, of course, never developed in a vacuum, and we would like to acknowledge the contributions made to our general thinking by several previous systems: Data Text, developed at Harvard University, and BMD, developed at the University of California at Los Angeles, were especially significant. The factor analysis subprogram was adapted from a program developed at the University of Alberta. The Guttman Scale subprogram borrowed heavily from a program originally designed by Professor Ronald Anderson. The output format and some of the table statistics used in subprogram CROSSTABS were directly borrowed from the Data Text system. While we wish to acknowledge our great indebtness to those who developed these programs, we alone are responsible for whatever errors or mistakes are present in their implementation.

Norman H. Nie
C. Hadlai Hull
Jean G. Jenkins
Karin Steinbrenner
Dale H. Bent

[1]SOS is currently operational for IBM 360/370 and Univac Series 70 computers. Further information may be obtained from Kisun Han of DUALabs, Suite 900, 1601 North Kent Street, Arlington, Virginia 22209.

CONTENTS