
Book Review

Title: Doing data analysis with SPSS 10.0
Authors: Carver, Robert H, Nash, Jane G
Duxbury 2000, ISBN 0534374751

Reviewed by: Harper Gilmour
Email: harper@stats.gla.ac.uk

This book is aimed primarily at students, both undergraduate and postgraduate, who have to take an introductory course, or a service course in Statistics as part of their degree programme. Both authors have extensive experience of teaching such courses. Robert Carver is Professor of Business Administration and Jane Nash is Associate Professor of Psychology, both at Stonehill College in Massachusetts. One of the strengths of the book is that it uses real data throughout and, not surprisingly, many of these datasets are on problems of relevance to students in these two areas studying in the United States. There are however a number of datasets from other disciplines and some of non-American origin. A detailed description of the data sets is provided in an appendix at the end of the book, but the data sets themselves have to be downloaded from the Duxbury Press website. This was very easy and very quick to do since none of the data sets are unreasonably large.

It is important to realise that this book is intended as a supplementary text for an introductory course in statistics rather than a primary textbook on statistical methods. There is some material in each chapter explaining the rationale behind the methods being used but not enough for students with no prior knowledge of the subject. It is also important to realise that the book covers techniques available in SPSS base version 10. SPSS base does not cover more advanced topics such as repeated measures ANOVA, logistic regression or analysis of survival data all of which are, increasingly, covered briefly in many introductory service courses. There is a short section in the book on one factor repeated measures ANOVA which is included for completeness at the end of the ANOVA chapters, although the authors acknowledge that the procedure is not in SPSS base and may not be available to many students.

The book covers statistical topics in the traditional order. Chapters 1-5 introduce the basic features of SPSS and show how it can be used to create simple tables and graphs and provide descriptive statistics. Chapters 6-9 cover basic ideas in probability and show how to use SPSS to generate data from Binomial, Poisson and Normal probability distributions. Chapters 9 and 10 cover sampling distributions, the central limit theorem and confidence interval for a population mean, using a combination of simulations and applications to real data. Chapters 11-14 cover hypothesis tests for means for one and two samples (paired and independent samples) and one and two way ANOVA. Chapters 15-18 cover simple linear regression, multiple linear regression and an introduction to non linear models. Finally there is one chapter on each of chi-square tests, nonparametric tests, basic forecasting techniques and tools for quality.

Throughout the book, each time a new SPSS procedure is introduced a dialog box image is provided to illustrate the procedure, and detailed instructions are given on the completion of the dialog box. The instructions are

clearly written and easy to follow. Throughout the text there are questions about the datasets to encourage students to think about the interpretation of the output rather than just concentrating on the act of obtaining the output from the package. Each chapter ends with a section called 'Moving On' which contains a series of more challenging questions relating to the datasets. In these sections students have to work out for themselves how to use SPSS to answer the questions posed using the knowledge they have gained from earlier illustrative examples. A separate manual of solutions to these additional problems is available for teachers if required.

The authors give suggestions for a range of ways in which the material in this book could be used in an introductory statistics course. It would be possible to work through some of the chapters as a group with each student at a computer. This might work well with early chapters when students are unfamiliar with the package and require detailed guidance and help. As an alternative they suggest giving students a chapter with selected 'Moving On' questions as an assignment to be completed in their own time. The students would then come back to the class with questions about any difficulties they encountered. This might work well in the middle of the course after students are familiar with the basic operations in SPSS. Finally students might be asked to work through several chapters in preparation for a larger data analysis project. Many of the data files provided with the book contain additional variables, which are never used in any of the illustrations or exercises in the book, which could be used as the basis of such a project.

However, many teachers like to use their own data sets for projects and assignments and it would be relatively easy to do this. Students could be instructed to work through chapters of the book using the illustrative examples provided. Teachers could then replace some of the 'Moving On' problems with their own data sets or provide a larger project based on their own data.

On the whole the book is well written and covers material appropriate for an introductory course. There were one or two areas in which I thought it could be improved, in particular in the use of confidence intervals and in methods for testing for association (or independence) for categorical data.

The chapters of the book which deal with inference concentrate very much on the interpretation of p values rather than confidence intervals. There is a chapter which deals with confidence intervals, which is good as far as it goes, but it only covers the concept of a confidence interval and the interpretation of a confidence interval for a population mean. It is becoming more common for academic journals in a range of disciplines to expect the main results of a study to be expressed using confidence intervals rather than p values where it is appropriate to do so. I accept that this book is not intended as a primary textbook in statistics, but there are many procedures covered in the book where SPSS routinely produces a confidence interval as part of its default output yet this is not mentioned in the text nor in the accompanying questions. This occurs, for example, in the contexts of one, two and paired samples t tests and follow up comparisons after one way ANOVA. The only other place I could find in the text where confidence intervals are explicitly mentioned is in connection with prediction in linear regression.

The chapter on chi-square tests is disappointingly short. In my experience the chi-square test of association (or test of independence as it is called in the book) is the most commonly used of all hypothesis tests. Most students who use this test on their own project data encounter problems with the validity of the test at some stage due to small expected values. There is an illustration in the book of this problem, but no advice is given on how to tackle this. It would have been relatively easy to extend this example to cover the technique of grouping categories and cover the application of Fisher's exact test to 2 x 2 tables.

Despite these criticisms the book is well written and does what it sets out to do. In the authors' words "*This book provides a foot in the door [for a student wishing to learn how to use SPSS]. Interested students and other users can explore the software possibilities via the extensive Help system or other standard SPSS documentation.*"

Anyone who is responsible for running an introductory statistics course in which SPSS is the software package of choice should consider using this book, as a supplement to their primary textbook or lecture notes, in their analysis of data practical classes.