Statistics for Research

With a Guide to SPSS

First edition published 2000
Second edition published 2005

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
32, M-Block Market
Greater Kailash – I
New Delhi 110 048

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 1 4129 1947 9
ISBN 1 4129 1948 7 (pbk)

Library of Congress Control Number: 2005929626

Typeset in Times New Roman
Printed on paper from sustainable sources
Printed in Great Britain by The Cromwell Press, Townbridge, Wiltshire

Contents

Part 3 Descriptive statistics: Numerical measures

Part 4 Inferential statistics: Tests for a mean

Part 5 Inferential statistics: Tests for frequency distributions

Part 6 Inferential statistics: Other tests of significance

Preface

This book is aimed at students and professionals who do not have any existing knowledge in the field of statistics. It is not unreasonable to suggest that most people who fit that description come to statistics reluctantly, if not with hostility. It is usually regarded as 'that course we had to get through'. I suspect that a sense of dread is also shared by many instructors when confronted with the prospect of having to teach the following material.

This book will hopefully ease some of these problems. It is written by a non-statistician for non-statisticians, for students who are new to the subject, and for professionals who may use statistics occasionally in their work. It is certainly not the only book available that attempts to do this. One might in fact respond with the statement 'not another stats book!' There are important respects, however, in which this book is different to the other numerous books in the field. When this book was first published by Macmillan Education as *Statistics for Social Research* it differentiated itself from other texts in three ways, each of which have been carried into this edition:

Communication of ideas. This book is written with the aim of communicating the basic ideas and procedures of statistical analysis to the student and user, rather than as a technical exposition of the fine points of statistical theory. The emphasis is on the explanation of basic concepts and especially their application to 'real-life' problems, using a more conversational tone than is often the case. Such an approach may not be as precise as others in dealing with statistical theory, but it is often the mass of technical detail that leaves readers behind, and turns potential users of statistical analysis away.

Integrated use of SPSS. This book integrates the conceptual material with the use of the main computer software package, **SPSS**. The development and availability of this software has meant that for most people 'doing stats' equals using a computer. The two tasks have converged. Unfortunately, most books have not caught up with this development and adequately integrated the use of computer packages with statistical analysis. They concentrate instead on the logic and formulas involved in statistical analysis and the calculation 'by hand' of problem solutions. At best other books have appendices that give brief introductions and guides to computer packages, but this does not bridge the gap between the hand calculations and the use of computer software. This book builds the use of SPSS into the text. The logic and application of various statistical techniques are explained, and then the examples are reworked on SPSS. Readers can link explicitly the traditional method of working through problems 'by hand' and working through the same problems on SPSS. Exercises also explicitly attempt to integrate the hand calculations with the use and interpretation of computer output.

To help readers along, a CD with all the data necessary to generate the results in the following chapters is included with this book, so that all the procedures described there can be replicated. You will need your own copy of SPSS to perform these procedures, and Chapter 2 lists a number of means by which you can obtain SPSS.

It is necessary, however, to point out that this is not a complete guide to SPSS. This book simply illustrates how SPSS can be used to deal with the basic statistical techniques that most researchers commonly encounter. It does not exhaust the full range of functions and options available in SPSS. For the advanced user, nothing will replace the *User's Guide* published by SPSS Inc. But for most people engaged in research, the following text will allow them to handle the bulk of the problems they will encounter.

For users of other statistics packages, the files are also saved in ASCII and Excel format so that they can be imported to these programs, along with a **Readme** file that explains the data definitions. All the files, and periodic minor updates and corrections, can be obtained at the following web site:

www.sagepub.co.uk/argyrous

Clear guide to choosing the appropriate procedures. This book is organized around the individual procedures (or sets of procedures) needed to deal with the majority of problems people encounter when analyzing quantitative data. Other texts flood the reader with procedure after procedure, which can be overwhelming. How to choose between the options? This book concentrates on just the most widely used techniques, and sorts through them by building the structure of the book around these options. Entire chapters are devoted to individual tests so that the situations in which a particular test is applied will not be confused with situations that call for other tests. Thus after working through the text, readers can turn to individual chapters as needed in order to address the particular problems they encounter.

The first version of this text proved to be popular in disciplines outside the social sciences, especially in the health sciences. As a result, the next version, published by Sage, UK, as *Statistics for Social and Health Research*, broadened its appeal to the health sciences through the inclusion of examples and exercises suited to their interests, but which were still intelligible to a non-specialist.

The second version of this text, as with the first, also found a broader audience than suggested by its title. This broad appeal suggested to me that a comprehensible 'generic' statistics textbook is of value to researchers in any field, and also desperately needed. Thus, this version drops from its title any reference to a specific discipline; its appeal is to all researchers who need some basic understanding of quantitative methods and the use of SPSS. Some specialized topics that are normally covered in certain fields and not others, such as the greater interest in small sample problems in the health sciences than in the social sciences, are not covered as a result. I have found, however, that instructors or students can supplement the basic techniques covered in this text with such specialized topics as required, especially given the vast amount of material now available on the internet.

In developing this new edition, I have also made some substantial changes (improvements!), while still retaining the three broad objectives set for the first version.

Reordering chapters around classes of descriptive techniques rather than levels of measurement. The previous editions were criticized, rightly I believe, for being too rigid in their emphasis on the limits placed on analysis by levels of measurement. When people analyze data they usually think in terms of classes of statistics first, such as central tendency, frequency tables, or correlation. The level at which variables are measured is an important consideration, but does not correspond to the way researchers 'think' about the problems they want to address. To accommodate this, chapters have been organized around the mainly used descriptive techniques, with data considerations (including levels of measurement) forming an element in the exposition of those techniques.

Reference to material available on the internet. The material now available on the internet is extensive and growing all the time. The lack of 'quality control', however, can make the use of such material fraught with perils. I have drawn on internet tools where appropriate and where I have been able to assess the quality of the information and resources presented. I have given the address for these internet sources in the text, but the reader should be aware that the maintenance of these sites is out of the control of myself or Sage, UK.

Streamlining of the five-step hypothesis testing procedure. I have dramatically altered the five-step hypothesis testing procedure, eliminating the calculation of critical scores and the pre-setting of alpha levels. This is in response to the current trends in academic journals, especially in the health sciences, which seek a less prescriptive approach to decision-making than has been the case in the past. It also reduces the calculations needed to arrive at a conclusion; a great relief to many students.

Greater emphasis on reporting results. I have found that researchers are often at a loss as to how to communicate their findings. I therefore have built into the five-step hypothesis testing procedure an explication of how to report findings. Getting results is one thing, but unless these can be communicated, especially to a general audience, their importance is lost.

Reference to the literature on statistical methods. Textbooks are always a lie. They present a field of knowledge as uncontroversial, when in fact it is usually a terrain of hot debate. This is no less the case with statistics textbooks, including previous incarnations of this one. Rather than continue the lie, I have introduced at various places some important points of debate and references to the literature where those interested can pursue the debates further.

Having noted the main features of this book as compared to others in the field, it is also worth noting what this book is not. This book looks at the analysis of quantitative data, and only the analysis of quantitative data. It makes no pretence to being a comprehensive guide to social or health research. Issues relating to the selection of research problems, the design of research methods, and the procedures for checking the validity and reliability of results are not covered. Such a separation of statistics from more general considerations in the design of research is a dangerous practice since it may give the impression that statistical analysis *is* research. Yet, nothing could be further from the truth. Statistical analysis is one way of processing information, and not always the best. Nor is it a way of proving anything (despite the rhetorical language it employs). At best it is evidence in an ongoing persuasive argument. The separation of statistics from the research process in general may in fact be responsible for the over-exalted status of statistics as a research tool.

Why then write a book that reinforces this separation? First, there is the simple fact that no single book can do everything. Indeed, other books exist which detail the issues involved in research, and the place of statistical analysis in the broader research process. Rather than duplicating such efforts this book is meant to sit side by side with such texts, and provide the methods of statistical analysis when required. Second, statistical analysis is hard. It raises distinct issues and problems of its own which warrant a self-contained treatment.

To the researcher or students using this book I have included other material on the CD that accompanies this book, especially chapters on detailed SPSS procedures that were too specialized for the actual text but which may be of interest. I have also placed this material on the website www.sagepub.co.uk/argyrous, to which I will be adding more material (including a list of corrections to any errors that may be discovered) over time, so you may wish to check this site periodically for such new material.

To the instructor, I have a wealth of material available for you at your request. This includes PowerPoint slides, Flash presentations, complete web pages for use in on-line courses, and a database of over 500 WebCT quiz questions that can be used for testing students and also for providing tutorial exercises. Please feel free to contact me at the address below and I will forward to you any (or all) of this material.

In the preparation of this edition I have been greatly assisted by the comments of many people who read the previously published versions of this text, and my thanks go to them. I do wish to specifically thank Ji In Lee for a thorough reading of the previous edition and

suggestions, to Punitha Arjunan for compiling the index and comments on the manuscript, and to Paul Francis and Peta Kennedy for comments on the manuscript of this edition. I am indebted to the Longman Group UK Ltd, on behalf of the Literary Executor of the late Sir Ronald Fisher and Dr Frank Yates FRS, for permission to reproduce Tables III, IV, and V from *Statistical Tables for Biological, Agricultural, and Medical Research*, 6/e (1974) in the Appendix, and to Professor A. Hald for permission to reproduce in amended form Table 1 of *Statistical Tables and Formulas 1952* in the Appendix.

Lastly, to the reader, I welcome any comments and criticisms, which can be passed on to me at the following address:

School of Social Science and Policy
University of New South Wales
NSW 2052, Australia
email: g.argyrous@unsw.edu.au

PART 1

An introduction to statistical analysis