

Statistics for Social Research

This book is dedicated to
my mother
my late father
Alex, Tasos, Effi, Anne, Elli, Mimi, Costa, Anna
Erin, Ben, Luke, Sophie, Elli Rose, Jordan
Pamela, Ryan, Michelle
Timothy, Alana
Danielle, Christine, Leanne, Marie
Amanda, Lisa
Alexandra, Andrea, Christopher
Stacey, Chloe, Billie-Marie
Alexandra, Katherine, Evelyn
and the ones still to come

Statistics for Social Research

George Argyrous

*School of Social Science and Policy
The University of New South Wales, Australia*





© George Argyrous 1997

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No paragraph of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, 90 Tottenham Court Road, London, W1P 9HE.

Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

First published 1997 by
MACMILLAN PRESS LTD
Houndmills, Basingstoke, Hampshire RG21 6XS
and London
Companies and representatives
throughout the world

ISBN 978-0-333-73023-2

ISBN 978-1-349-14777-9 (eBook)

DOI 10.1007/978-1-349-14777-9

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

10 9 8 7 6 5 4 3 2 1
06 05 04 03 02 01 00 99 98 97



Contents

Introduction x

Part 1 Descriptive statistics

1 Variables and measurement 3

The conceptualization and operationalization of variables 4

Levels of measurement 7

Discrete and continuous variables 11

Summary 12

2 Basic tools of description: Tables 15

Types of descriptive statistics 16

Percentages and proportions 17

Frequency distributions 19

Class intervals 23

3 Graphs 28

Some general principles 29

Pie graphs 29

Bar graphs 30

Histograms 33

Frequency polygons 34

Common problems and misuses of graphs 35

4 Measures of central tendency and measures of dispersion 41

Measures of central tendency 41

Choosing a measure of central tendency 47

Measures of dispersion 50

Summary 57

5 Descriptive statistics on SPSS 60

Assigning variable labels: The Define Variable command 62

Defining values: The Value Labels command 63

Data entry 67
Saving a data file 70
Opening a data file 71
Frequency tables using SPSS 72
Valid cases and missing values 73
Measures of central tendency using SPSS 75
Measures of dispersion using SPSS 76
Summary 77

6 The normal curve 79

The normal distribution 79
z-scores 84
Using normal curves to describe a distribution 87
Normal curves on SPSS 92

Part 2 Inferential statistics

A Hypothesis testing: The one sample case

7 Sampling distributions 103

Random samples 105
The sampling distribution of a sample statistic 107
The central limit theorem 112
Generating random samples using SPSS 113
Summary 117

8 Estimation and confidence intervals 118

Estimation 119
Changing the confidence level 125
Changing the sample size 128
Choosing the sample size 131

9 Introduction to hypothesis testing: The z-test for a single mean 134

Hypothesis testing: The general idea 134
Hypothesis testing in detail 140
What do inference tests 'prove'? 150
The significance of statistical significance 151
A two-tail z-test for a single mean 152
A one-tail z-test for a single mean 155
Summary 158

10 The t-test for a single mean 160

The Student's *t*-distribution 160
Degrees of freedom 163
The one-sample *t*-test using SPSS 169

11 The z -test for a single proportion 174

Binomial variables 174

The sampling distribution of sample proportions 176

The z -test for a single proportion 177The z -test for a single proportion using SPSS 179

Estimating a population proportion 181

Inference using the confidence interval for a proportion 182

12 One sample non-parametric tests 185

Parametric and non-parametric tests 185

The chi-square goodness-of-fit test 186

The chi-square goodness-of-fit test using SPSS 190

The chi-square goodness-of-fit test for normality 193

Another non-parametric test: The runs test for randomness 196

The runs test using SPSS 201

Part 2 Inferential statistics**B Hypothesis testing for two or more independent samples****13 The t -test for the equality of two means 209**

Dependent and independent variables 210

The sampling distribution of the difference between two means 212

The two-sample t -test using SPSS 217

One-tail and two-tail tests 221

Appendix: The z -test for two means 223**14 Analysis of variance 226**

Hypothesis testing with more than two samples: The general idea 226

The one-way analysis of variance F -test 231

ANOVA using SPSS 237

15 Ordinal data tests for two or more samples 245

The Wilcoxon rank sum test 246

The Wilcoxon rank sum test using SPSS 251

Other non-parametric tests for two or more samples 253

Appendix: The Mann–Whitney U -test 254**16 The chi-square test for independence 257**

The chi-square test and other tests of significance 257

Describing a sample: Crosstabulations 258

Crosstabulations using SPSS 261

Statistical independence and the relationships between variables 262

The chi-square test for independence 264

Degrees of freedom 268

The distribution of chi-square 269
The chi-square test using SPSS 272
Problems with small samples 275
Problems with large samples 276
Appendix: Hypothesis testing for two proportions 279

Part 2 Inferential statistics

C Hypothesis testing for two dependent samples

17 The *t*-test for the mean difference 287

Dependent and independent samples 287
The dependent samples *t*-test for the mean difference 288
The dependent samples *t*-test using SPSS 291

18 Non-parametric tests for dependent samples 298

The McNemar test for binomial distributions 298
The McNemar test using SPSS 301
The Wilcoxon signed-ranks test for ordinal data 303
The Wilcoxon signed-ranks test using SPSS 307

Part 3 Measures of association

19 Introduction to measures of association 313

Measures of association as descriptive statistics 313
If an association exists, how strong is it? 315
The pattern and/or direction of the association 316
Measures of association and tests of significance 317

20 Measures of association for nominal data 319

Cramer's *V* 319
Cramer's *V* using SPSS 321
Proportional reduction in error measures of association 324
Lambda 324
Lambda using SPSS 327
Limitations on the use of lambda 329
Appendix: Standardizing crosstabulations when lambda is zero 331

21 Measures of association for ordinal data 335

Gamma 335
Concordant and discordant pairs 336
Gamma using SPSS 341
Spearman's rank-order correlation coefficient 345
Spearman's rho using SPSS 348
Other measures of association for ordinal data 350

22 Regression and correlation for interval/ratio data 356

Scatter plots 356

Linear regression 358

Pearson's correlation coefficient (r) 366Explaining variance: The coefficient of determination (r^2) 368

Plots, correlation, and regression using SPSS 372

Testing for significance 378

The assumptions behind regression analysis 379

Summary 382

Appendix 386

Table A1 Area under the standard normal curve 386

Table A2 Distribution of t 387Table A3 Distribution of F 388

Table A4 Distribution of chi square 389

Key equations **390****Glossary 396****Answers 399****Index 420**



Introduction

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 stat’s book!’. However, there are important respects in which this book is different from the numerous other books in the field.

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. This is the **Statistical Package for the Social Sciences (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 or 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 appendixes 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 try to integrate the hand calculations with the use and interpretation of computer output.

To help readers along, a disk with all the data necessary to generate the results in the following chapters is provided, so that all the procedures described there can be replicated. Version 6.1 of SPSS is now available for both Macintosh and Windows platforms, and operates in virtually the same way in either format. There are slight differences in the appearance of some windows, but the basic menu structure is the same for both Windows and Macintosh environments. In fact, although the data were analysed on SPSS Version 6.1, users with SPSS Version 5.0 and Version 6.0 will still find that the basic procedures detailed in this book remain the same. Users of other statistical packages can download these data files in ASCII format from the following Web address: <ftp://www.arts.unsw.edu.au/Pub/>

However, it is necessary 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 social research, the following text will allow them to handle the bulk of the problems they will face.

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 in analysing quantitative data. Other texts flood the reader with procedure after procedure, which can be overwhelming. How to choose between the options? This book concentrates just on 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 the 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 confront.

Having noted the main features of this book as compared with 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 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 social research is a dangerous practice since it may give the impression that statistical analysis *is* social 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 in fact may 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 that detail the issues involved in social research, and the place of statistical analysis in the broader research process. An example is S. Sarantakos 1993, *Social Research*, published by Macmillan. 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 that warrant a self-contained treatment.

Many people have helped in bringing this book to press, although none should be implicated in any remaining errors or omissions. The students and my colleagues in the School of Social Science and Policy at the University of New South Wales have provided an invaluable sounding board for many of the ideas and forms of presentation that follow. Katrina Neal, Sandra Napoli, Jason Hecht, Rogelia Pe-Pua, and Frankie Leonard read various sections and provided helpful comments and suggestions; Carol Healey, Karen Tremayne, Cathy Deane, and Simon Kozlina read the entire manuscript and rescued it from many potential errors. Phuong and Joanne at the Sydney office of SPSS tolerated my queries about the use of the software with great patience and I am especially thankful for their help.

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 F.R.S., for permission to reproduce Tables III, IV and V from *Statistical Tables for Biological, Agricultural and Medical Research*, 6/e (1974) in Appendix Tables A2, A3 and A4; and to Professor A Hald for permission to reproduce in amended form Table 1 of *Statistical Tables and Formulas 1952* in Appendix A1.

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, 2052, Australia.
e-mail: g.argyrous@unsw.edu.au