

The background of the cover features a dark blue field with a network of thin, light blue lines. Scattered across this field are several out-of-focus, glowing yellow circles of varying sizes, creating a bokeh effect.

Second Edition

Data Collection and Analysis

Edited by **Roger Sapsford** and **Victor Jupp**



C32
D232
E.2

DATA COLLECTION AND ANALYSIS

Second edition

edited by

Roger Sapsford and Victor Jupp



E200602713

 **SAGE Publications**

London • Thousand Oaks • New Delhi



in association with
The Open University

© The Open University 2006

First edition published 1996

Second edition published 2006

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.



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

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B-42, Panchsheel Enclave
Post Box 4109
New Delhi 110 017

British Library Cataloguing in Publication data

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

ISBN 0 7619 4362 5
ISBN 0 7619 4363 3 (pbk)

Library of Congress Control Number: 2005927680

Typeset by C&M Digital (P) Ltd., Chennai, India
Printed in India at Gopsons Papers Ltd, Noida
Printed on paper from sustainable resources

DATA COLLECTION
AND ANALYSIS

Preface

This book is about collecting and analyzing research data, but this does not mean that it teaches just the *techniques* of research. Research does indeed involve technical matters, but they are not in fact very difficult to grasp. What is more important to social science research than the technical questions is the frame of mind in which you approach it. Both of us regard empirical research as a part of theory-building, on the one hand, and of good professional practice on the other. Elsewhere (Smith, 1975), the prerequisite for good research has been described as 'the methodological imagination'.

Based originally on material prepared for Open University course DEH3I3 *Principles of Social and Educational Research* in the 1990s, this volume has been re-edited and in some places re-written for a wider market. The original material was targeted at the reader of research rather than the practitioner; its aim was to equip students to understand, evaluate and use evidence in their academic and professional work. The needs of students who have projects to carry out now receive greater emphasis, and each chapter ends with a 'running activity' which will help you to build up a plan or proposal for your own research. We have also suggested further reading in each chapter. The book is designed for students from a wide range of disciplines (including sociology, social psychology, social policy, criminology, health studies, government and politics) and practitioners and readers in a number of applied areas (for example, nurses and other medical practitioners, social workers and others in the caring professions, workers in the criminal justice system, market researchers, teachers and others in the field of education). The second edition has brought some of the examples up to date and added chapters on research issues and methods which have emerged as of importance over the past decade.

The first part of the book covers *design* issues which are closely bound up with how data are collected: the basic logic of different kinds of research studies, and the 'technology' of sampling. Part II deals with the basic research activities of observation and asking questions (using methods exhibiting varying degrees of structure) and the use of documents, published statistics and databases for research purposes. Part III covers (a) statistical analysis of quantitative data, from simple tabular analysis and the graphical presentation of data to complex multivariate techniques using regression or analysis of variance, and including a chapter on the preparation of data, and (b) the qualitative analysis of text and interview transcripts from both an 'ethnographic' and a 'critical' point of view. Two detailed examples of qualitative analysis can also be found on the website www.sagepub.co.uk/sapsford. The 'statistical' chapters do not assume any prior knowledge of statistical techniques. They also do

not require more than very simple calculations. Although it is not necessary for the work outlined in this book, we are assuming that those who go on to do research which requires extensive statistical calculation will arrange access to a computer (by far the most sensible way to do statistical analysis). Finally, Part IV looks at the broader ethical, political and conceptual issues that are important at all stages of research.

Acknowledgements

The material on which the first edition of this book was based was originally prepared for an Open University Course Team, and we should like to acknowledge the substantial amount contributed by the Team. Our thanks are due particularly to Martyn Hammersley, Kevin McConway, Keith Stribley and three colleagues who have died since the original course material was written and who are sorely missed – Judith Calder, Betty Swift and Michael Wilson.

Grateful acknowledgement is also made to the following sources for permission to reproduce material in this book:

Figures 1.2 and 1.3: D. Campbell (1969) 'Reforms as experiments', *American Psychologist*, vol. 24; © The American Psychological Association.

Figure 3.1: adapted from B.H. Junker (1960) *Field Work*, University of Chicago Press; © The University of Chicago.

Figures 8.1 and 8.9: P. Abbott (1988) *Material Deprivation and Health Status in the Plymouth Health Authority District*; © Professor Pamela Abbott, Glasgow Caledonian University.

Table 2.3: M.F. Weeks, B.L. Jones, R.E. Folsom Jr and C.H. Benrud C.H. (1984) 'Optimal times to contact sample households', *Public Opinion Quarterly*, vol. 48; © The Trustees of Columbia University.

Tables 8.1, 8.5 and 8.8: P. Abbott (1988) *Material Deprivation and Health Status in the Plymouth Health Authority District*; © Professor Pamela Abbott, Glasgow Caledonian University.

Table 9.2: M. Conway, G. Cohen and N. Stanhope (1991) 'On the very long-term retention of knowledge acquired through formal education: twelve years of cognitive psychology', *Journal of Experimental Psychology*, vol. 120, no. 4, December; © 1991 by the American Psychological Association. Reprinted by permission.

Table 9.7: G. Cohen, N. Stanhope and M. Conway (1992) 'Age differences in the retention of knowledge by young and elderly students', *British Journal of Developmental Psychology*, vol. 10, Part 2; © 1992 by the British Psychological Society.

Box 3.1: N. Flanders (1970) *Analyzing Teaching Behavior*, Addison-Wesley; © Ned A. Flanders.

Box 10.1: M. Hammersley and P. Atkinson (1983) *Ethnography: Principles in Practice*, Tavistock Publications; © Martyn Hammersley and Paul Atkinson.

Box 10.3: A. Strauss and J. Corbin (1990) *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, Sage Publications Inc.

Box 11.1: J. Scott (1990) *A Matter of Record: Documenting Sources in Social Research*, Polity Press.

List of Contributors

Roger Sapsford is Reader in Social Research Methods at the University of Teesside and before that he taught research methods at the Open University. He is author of *Life Sentence Prisoners: reaction, response and change* and (with Pamela Abbott) of *Community Care for Mentally Handicapped Children* and *Research Methods for Nurses and the Caring Professions*. He was also lead editor of *Theory and Social Psychology* and *Researching Crime and Criminal Justice*.

Victor, Jupp is a principal lecturer at the University of Northumbria at Newcastle on Tyne, with special responsibility for criminology and social research methods. He was previously a lecturer in research methods at the Open University. He is author of *Methods of Criminological Research*, co-author of *Invisible Crime: their victims and their regulation*, and lead editor of *Doing Criminological Research*.

Pamela Abbott is Vice-Rector of the Kigali Institute of Science, Technology and Management in Rwanda and Professor of Social Policy at Glasgow Caledonian University. She has written a large number of books and articles but is best known as the author of *Women and Social Class*, *The Family and the New Right* and *Introduction to Sociology: feminist perspectives*.

David Boulton was a principal lecturer in the Department of Applied Community Studies at Manchester Metropolitan University at the time when this chapter was written. His teaching was entirely focused on undergraduate and postgraduate research methods, and his research on ways in which the construction of data influences policy outcomes, with special reference to issues surrounding policing.

Judith Calder was a senior lecturer in research methods with the Institute of Educational Technology at the Open University. She contributed to undergraduate and postgraduate courses in research methods and statistics and undertook research consultancies for the British Council, the ODA, the DfE, the Council of Europe and the World Bank. It was with deep regret that we learned of her death a few years before this second edition of the book was published.

Ruth Finnegan was Professor of Comparative Social Institutions at the Open University at the time when her chapter was written. Her publications include *Oral Literature in Africa*, *Literacy and Oracy: studies in the technology of communication*,

The Hidden Musicians, and a jointly edited CD-ROM entitled *Project Reports in Family and Community History*.

Peter Foster was a senior lecturer in education at Manchester Metropolitan University at the time when this chapter was written. He is author of *Practice in Multiracial and Antiracist Education* and (with Roger Gomm and Martyn Hammersley) *Constructing Educational Inequality*.

Martyn Hammersley is Professor of Education and Social Research at the Open University. His substantive research has mostly been in the sociology of education, but much of his more recent work has been concerned with methodological issues. He is author of *The Dilemma of Qualitative Method, Reading Ethnographic Research, What's Wrong with Ethnography?*, *The Politics of Social Research* and (with Paul Atkinson) *Ethnography: Principles and Procedures*.

William (Bill) Schofield was a developmental psychologist and statistician at the Cambridge University Department of Experimental Psychology at the time when this chapter was written and now works at the University of Sydney. He is author of a large number of reports on research into physical and mental growth, morbidity and malnutrition, is responsible for the WHO/FAG equations for the prediction of basal metabolic rate and has been consultant to international organisations and major drug companies.

Betty Swift was a lecturer in research methods in the Institute of Educational Technology at the Open University and author of a large number of reports on the University's teaching. She died in 1993, not long after the original unit was written on which this chapter is based, and she is sorely missed.

Michael Wilson was a senior lecturer in research methods in the Faculty of Social Sciences at the Open University, and before that a researcher and lecturer at Imperial College London, the University of Leeds and the University of Cardiff. His death in 1995 is deeply regretted by all his colleagues.

Contents

List of Figures	xi
List of Tables	xii
List of Boxes	xiii
Preface	xiv
Acknowledgements	xvi
List of Contributors	xviii
PART I DESIGN ISSUES	1
1 Validating Evidence	1
<i>Roger Sapsford and Victor Jupp</i>	
Counting Cases: Measurement and Case Selection	2
Examples from UK Government Statistics	3
Three Kinds of Crime Survey	4
A Study of Mothers	9
Comparing Groups: the Logic of Design	13
Quasi-experimental Analysis: Road Traffic Fatalities	13
Experiments in HIV Counselling	16
Establishing Boundaries in Qualitative Studies	20
Conclusion	22
Key Terms	22
Further Introductory Reading	23
Summary Activity: Preparing Your Research Proposal	23
2 Survey Sampling	26
<i>William Schofield</i>	
Sampling	26
Defining the Population to be Sampled	27
Sampling Units	28
The Sampling Frame	28
Selecting a Sample	29
Simple Random Sampling	30
Stratified Random Sampling	32
Cluster Sampling	34
Quota Sampling	36
Estimation of Population Parameters	38

Means, Variance and Standard Deviations	39
How Sample Means are Distributed	43
Error, Sample Size and Non-response	47
Sample Size	48
Non-response	49
Conclusion	51
Key Terms	52
Further Reading	53
Answers to Activities	53
Research Proposal Activity 2	54
PART II DATA COLLECTION	57
3 Observational Research	57
<i>Peter Foster</i>	
Using Observation	57
Advantages and Limitations	59
Structure in Observation	60
Relationships and Roles	64
Negotiating Access	64
Developing Relationships	69
The Researcher's Role	73
Managing Marginality	77
What to Observe	77
Focusing Research	77
Representative Sampling	78
Recording Observations	80
More-structured Observation	80
Less-structured Observation	82
Using Technology	85
Assessing the Validity of Observations	86
Threats to Validity	87
Ways of Assessing Validity	88
Conclusion	90
Key Terms	91
Further Reading	92
Research Proposal Activity 3	92
4 Asking Questions	93
<i>Michael Wilson and Roger Sapsford</i>	
Comparison of Methods	93
Interview and Questionnaire Methods	93
Procedural/Structural Factors	94
Contextual Factors	95
Choosing Between Methods	97

Highly Structured Methods	97
Interview Schedules	99
Open-ended Questions in Standardized Interview Schedules	101
Self-administered Questionnaires	102
Piloting	103
Asking More Complex Questions	107
Factual Questions	108
Retrospective Questions	108
Composite Measurement	109
Less Structured Methods	112
Observing and Interviewing	113
An Example of Less Structured Interviewing	114
Interviewer Control in Less Structured Interviews	116
Reactivity and Validity in Research Asking Questions	118
Conclusion	120
Key Terms	120
Further Reading	122
Research Proposal Activity 4	122
5 Research and Information on the Net	124
<i>Roger Sapsford</i>	
Searching the Literature	124
Information on the Net	127
Sampling on the Net	129
Questionnaires on the Net	130
Qualitative Interviewing on the Net	132
Participating and Observing	133
Conclusion	134
Key Terms	135
Further Reading	136
Research Proposal Activity 5	136
6 Using Documents	138
<i>Ruth Finnegan</i>	
Documentary Sources	139
Some Initial Distinctions	142
Primary Versus Secondary Sources	142
'Direct' and 'Indirect' Uses	143
How Documentary Sources Come into Being	144
Further Questions to Ask	146
Conclusion	149
Key Terms	150
Further Reading	150
Research Proposal Activity 6	151

PART III DATA ANALYSIS	153
7 Preparing Numerical Data <i>Betty Swift</i>	153
Establishing the Nature of the Problem	154
Prior Structure, Method of Recording and Sample Size	157
Prior Structure	157
Method of Recording	160
Sample Size	161
Getting the Data into Shape for Analysis	162
Getting the Pre-coding Right	165
Coding Open-ended Questions	165
Coding Data from Unstructured Interviews	168
Three Approaches to Coding	170
Correcting Errors	173
Data Manipulation During Analysis	174
Coping with Missing Values	174
Re-coding Variables	175
Conclusion	177
Key Terms	178
Further Reading	179
Answers to Activities	179
Appendix	181
Research Proposal Activity 7	182
8 Extracting and Presenting Statistics <i>Roger Sapsford</i>	184
The Right Figure for the Job	184
Frequencies	185
Two-way Distributions	192
Introducing a Third Variable	196
Postscript: the Element of Chance	202
Key Terms	203
Further Reading	205
Answers to Activities	205
Research Proposal Activity 8	206
Appendix: Data-set	207
9 Statistical Techniques <i>Judith Calder and Roger Sapsford</i>	208
Different Kinds of Data	208
Approaches to Analysis	209
Looking at Variables	209
Two Types of Statistics	211

Descriptive Measures	211
Measures of Central Tendency	212
Measures of Spread	212
Measures of Location	213
Inferential Statistics	214
Types of Error	214
Chi-square	217
z and t tests	220
Analysis of Variance	221
Measuring Association	225
Correlation Coefficients	225
Simple Linear Regression	226
Multivariate Statistics	227
Analysis of Variance and Related Techniques	230
Regression Techniques	234
Further Multivariate Approaches	237
Key Terms	238
Further Reading	241
Answer to Activity 9.4	241
Research Proposal Activity 9	242
10 Analysis of Unstructured Data	243
<i>David Boulton and Martyn Hammersley</i>	
Complementary Perspectives of Reader and Researcher	244
Types of Qualitative Data	245
The Process of Analysis	246
Data Preparation	246
Starting the Analysis	250
Reflexivity and the Assessment of Validity	256
Conclusion	257
Key Terms	257
Further Reading	258
Research Proposal Activity 10	258
11 Discourse Research	260
<i>Roger Sapsford</i>	
Ideology and Discourse	260
Ideology	260
Discourse	262
Discursive Structures	264
Discourse Processes	266
Conclusion	268
Key Terms	269
Further Reading	270
Research Proposal Activity 11	271

12 Documents and Critical Research	272
<i>Victor Jupp</i>	
Documents, Texts and Discourse	273
Types of Document	274
A Typology of Documents	277
Critical Social Research	277
Critical Analysis of Documents	279
Case Study 1: A Proposal for Critical Analysis	281
Case Study 2: Critical Analysis of a Public Document	284
Case Study 3: Critical Analysis of Decision-making	285
Case Study 4: Critical Analysis of a Research Report	286
Conclusion	289
Key Terms	289
Further Reading	289
Research Proposal Activity 12	290
PART IV CONCLUSION	291
13 Ethics, Politics and Research	291
<i>Pamela Abbott and Roger Sapsford</i>	
Ethics and Research	293
Politics and Research	298
Intelligence and the Politics of 'Race'	298
Intelligence, Achievement and the Politics of Class	302
Politics, Ideology and 'Research Style'	305
Overview of Issues	309
Conclusion	310
Further Reading	311
Research Proposal Activity 13	311
References	313
Index	323

List of Figures

1.1	Diagrammatic representation of simple action research	13
1.2	Connecticut traffic fatalities, 1951–1959	14
1.3	Control series design comparing Connecticut fatalities with those in four comparable states	15
1.4	Diagrammatic representation of comparative research using a treatment group and a control group	16
2.1	Distribution of heights in a sample of 1,052 mothers	39
2.2	Mothers' heights: histogram and superimposed normal curve	40
2.3	Mothers' heights expressed in deviation and z-score units	41
2.4	Distribution of mothers' heights in a single simulated random sample, mean = 62.49, s.d. = 2.45	44
2.5	Distribution of mean heights from a hundred simulated random samples, mean = 62.498, s.d. = 0.074	44
2.6	Distribution of mean heights from a hundred simulated random samples, scaled for more detailed display	45
3.1	Theoretical social roles for fieldwork	72
8.1	Distribution of wards of the Plymouth Health Authority District on the Material Deprivation Index	188
8.2	Bar graph of Material Deprivation Index scores	189
8.3	Line graph of Material Deprivation Index scores	189
8.4	Effects of varying scale on a line graph	190
8.5	Pie chart showing share of votes in a fictional constituency	192
8.6	Health Index by Material Deprivation Index	197
8.7	Health and Material Deprivation Indexes (%)	197
8.8	Deaths from lung cancer, cerebrovascular disease (CVD) and ischaemic heart disease (IHD) in St Peter's and South Hams	198
9.1	Distribution of retention interval frequencies in a study of memory	212
9.2	Low variability (a) and high variability (b)	213
9.3	Standard normal distribution: each shaded area represents $\alpha/2$	216
9.4	Lines of best fit	227
9.5	Linear regression line and the regression coefficient	228
9.6	Linear regression line showing the residual	228
9.7	Correlation as a measure of variance explained	229
9.8	A model of multiple regression	235
9.9	Proportions of variance explained	235