Statistics for Real-Life Sample Surveys

Non-Simple-Random Samples and Weighted Data

Samples used in social and commercial surveys, especially of the general population, are usually less random (often by design) than many people using them realise. Unless it is understood, this 'non-randomness' can compromise the conclusions drawn from the data. This book introduces the challenges posed by less-than-perfect samples, giving background knowledge and practical guidance for those who have to deal with them. It explains why samples are, and sometimes should be, non-random in the first place; how to assess the degree of non-randomness; when correction by weighting is appropriate and how to apply it; and how the statistical treatment of these samples must be adapted. Extended data examples show the techniques at work. This is a book for practising researchers. It is a reference for the methods and formulae needed to deal with commonly encountered situations and, above all, a source of realistic and implementable solutions.

The authors are from Roy Morgan Research, Australia's best-known market and social research, public opinion and political polling organisation.

Sergey Dorofeev is Technical Director of Roy Morgan International, the global arm of Roy Morgan Research.
Peter Grant was Research and Development Manager in the Melbourne head office.

Together they have more than 50 years' experience of commissioning and conducting survey research.

Statistics
for Real-Life Sample Surveys

Non-Simple-Random Samples and Weighted Data

Sergey Dorofeev and Peter Grant

Roy Morgan Research, Melbourne, Australia

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521674652

First published 2006

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

ISBN 978-0-521-85803-8 Hardback
ISBN 978-0-521-67465-2 Paperback

Contents

Preface

Yet another book on the statistics of samples: what can there be left to say? Indeed there is a very large choice of titles on statistics and sampling and most of what is in this book can be found somewhere in the published literature. There are two main kinds of book in this area. One kind presents the statistics of simple samples in a simplistic way, based on the application of pure probability theory, and is inclined to gloss over the fact that reality is rarely that simple. The other kind deals with the complexities of complex samples, but often in a way that anyone but a specialist mathematician will find daunting. Our purpose is to provide an appreciation of the statistical issues involved in real-life sample surveys and practical guidance on how to cope with them, without losing sight of the need to remain comprehensible to the non-mathematician.

Qualified professional statisticians, particularly those whose interest lies mainly in the assessment of sampling and non-sampling survey error, may find this book too superficial. But it is intended primarily for practising researchers whose main concern is to extract meaning from survey results and to apply it in the formulation of plans and policies, and who may have only basic formal statistical training. We assume a reasonable level of numeracy, but those who are not mathematically inclined should not be put off. Our first aim is to impart an understanding of the issues and the principles involved, for which only rudimentary mathematics is required. The more advanced formulae and the proofs are there for those who can make immediate use of them and as a source of reference for those who may need them in the future or who may need to refer others to them.

The content of this book is intentionally selective. We do not intend it to be a general introduction to the very broad subjects of sampling or statistics. We assume that readers are already familiar with the basic principles of the statistics of samples from other texts, but stand in need of some help in applying what they have learned in the more complicated 'real world', where practical problems loom so much larger than they did in the textbook examples.

Our experience provides us with two specific stimuli to write. Firstly, we have lost count of the researchers we have met, many of them with formal qualifications which include statistics, who were ignorant of some of the fundamental issues involved in any

deviation from simple random sampling because they had never seen them explained in practical terms. For these researchers a compact source of reference which covers the basic, and some of the not-so-basic, principles of sampling and evaluation of sample-based data can be of value. This should be especially true because we explain the mathematical and mechanical processes of sampling and data analysis, and try to show the reasoning behind them, some of their implications and some of the pitfalls.

Secondly, and more importantly, there are two subjects of considerable practical importance that are not well served by the more accessible literature. In general, our experience is that basic textbooks are inclined to present an unreasonably optimistic picture of the quality of the samples with which their readers will have to work. We aim to provide a review of the problems associated with the evaluation of real-world sample-based data and practical guidance on coping with them. In particular we examine the related questions of the treatment of samples other than simple random ones and the practice and implications of weighting data.

Weighting is a subject which is mentioned briefly if at all in the main textbooks and about which many researchers appear to know very little. However, in practice very many surveys have the data weighted at the analysis stage, for a variety of reasons. Weighting has important implications for the assessment of results, especially in the calculation of confidence intervals and in significance testing. Users of popular statistical software packages are often unaware of some of the things that these may do with weighted data.

This book brings together in a compact form the essential material for understanding and coping with the most salient problems of non-simple-random samples. Almost all of the theoretical material it contains is available from other sources, but much of that is to be found in specialist books and journals that may not be readily accessible, even to those who know what to look for. We aim to provide an awareness and understanding of the main factors in sample design and planning and the practical realities of sample selection and recruitment, their likely effects and the allowance that should be made for them in interpreting results. We particularly aim to provide the practising researcher with a source of reference for the methods and formulae necessary to deal with the most commonly encountered situations and above all with practical, realistic and implementable solutions.

We have also included a short review of some of the more common analytical tools used with survey data. If we seem slightly less than enthusiastic in our endorsement of these it is because experience has taught us to be cautious in applying sophisticated mathematical procedures to data that may be considerably less sophisticated. It is understandable that researchers should be keen to exploit any technique that will squeeze more information out of a data set amassed with much trouble and expense, but there are pitfalls to be avoided. We seek to temper that enthusiasm rather than to blunt it.

We could not have produced a book such as this unaided. We express our thanks first of all to Roy Morgan Research of Melbourne, Australia, for providing the support and

facilities which made this book possible. Parts of the text have been developed from material originally produced for in-house training and reference and we appreciate the opportunity to expand these in a wider ranging text. Results from a genuine large-scale survey are very valuable in providing illustrations and we are fortunate in being able to use examples from the Roy Morgan Single Source survey. This is a large-scale, wide-ranging, national continuous syndicated survey that has been operating in Australia for many years. It includes the Australian industry-currency print media survey and a number of other 'industry monitors'. The annual sample size is over 55 000 adults. It also operates in the USA, UK, New Zealand, and Indonesia.

We also thank the Australian Institute of Health and Welfare for their permission to use the case history of the 1998 National Drug Strategy Household Survey.

Like all practising researchers, we owe a great debt to the many other researchers who over the years have shared their accumulated experience. Above all, we are indebted to Michele Levine, CEO of Roy Morgan Research, for her unfailing encouragement of our efforts, and for her constant willingness to share her knowledge and experience with us. In addition, we must make special mention of George Rennie with whom we have worked on a number of projects and whose ideas and insights have often helped us to clarify our thinking.