



# **Collecting, Managing, and Assessing Data Using Sample Surveys**

**Peter Stopher**



CAMBRIDGE



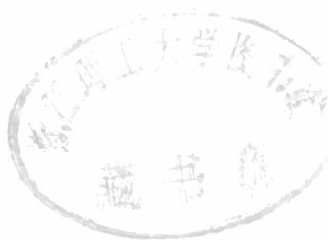
30807616

# Collecting, and Assessing Data Using Sample Surveys

---

---

Peter Stopher



**CAMBRIDGE**  
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS  
Cambridge, New York, Melbourne, Madrid, Cape Town,  
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press  
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521681872](http://www.cambridge.org/9780521681872)

© Peter Stopher 2012

This publication is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without the written  
permission of Cambridge University Press.

First published 2012

Printed in the United Kingdom at the University Press, Cambridge

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

ISBN 978-0-521-86311-7 Hardback

ISBN 978-0-521-68187-2 Paperback

Cambridge University Press has no responsibility for the persistence or  
accuracy of URLs for external or third-party internet websites referred to  
in this publication, and does not guarantee that any content on such  
websites is, or will remain, accurate or appropriate.

To my wife, Carmen, with grateful thanks for your faith in me and your continuing support and encouragement.

## *Acknowledgements*

---

As is always the case, many people have assisted in the process that has led to this book. First, I would like to acknowledge all those, too numerous to mention by name, who have helped me over the years, to learn and understand some of the basics of designing and implementing surveys. They have been many and they have taught me much of what I now know in this field. However, having said that, I would particularly like to acknowledge those whom I have worked with over the past fifteen years or more on the International Steering Committee for Travel Survey Conferences (ISCTSC), who have contributed enormously to broadening and deepening my own understandings of surveys. In particular, I would like to mention, in no particular order, Arnim Meyburg, Martin Lee-Gosselin, Johanna Zmud, Gerd Sammer, Chester Wilmot, Werner Brög, Juan de Dios Órtuzar, Manfred Wermuth, Kay Axhausen, Patrick Bonnel, Elaine Murakami, Tony Richardson, (the late) Pat van der Reis, Peter Jones, Alan Pisarski, Mary Lynn Tischer, Harry Timmermans, Marina Lombard, Cheryl Stecher, Jean-Loup Madre, Jimmy Armoogum, and (the late) Ryuichi Kitamura. All these individuals have inspired and helped me and contributed in various ways to this book, most of them, probably, without realising that they have done so.

I would also like to acknowledge the support I have received in this endeavour from the University of Sydney, and especially from the director of the Institute of Transport and Logistics Studies, Professor David Hensher. Both David and the university have provided a wide variety of support for the writing and production of this book, for which I am most grateful.

However, most importantly, I would like to acknowledge the enormous support and encouragement from my wife, Carmen, and her patience, as I have often spent long hours on working on this book, and her unquestioning faith in me that I could do it. She has been an enduring source of strength and inspiration to me. Without her, I doubt that this book would have been written.

As always, a book can see the light of day only through the encouragement and support of a publisher and those assisting in the publishing process. I would like to acknowledge Chris Harrison of Cambridge University Press, who first thought that this book might be worth publishing and encouraged me to develop the outline for

it, and then provided critical input that has helped to shape the book into what it has become. I would also like to thank profusely Mike Richardson, who carefully and thoroughly copy-edited the manuscript, improving immensely its clarity and completeness. I would also like to thank Joanna Breeze, the production editor at Cambridge. She has worked with me with all the delays I have caused in the book production, and has still got this book to publication in a very timely manner. However, as always, and in spite of the help of these people, any errors that remain in the book are entirely my responsibility.

Finally, I would like to acknowledge the contributions made by the many students I have taught over the years in this area of survey design. The interactions we have had, the feedback I have received, and the enjoyment I have had in being able to teach this material and see students understand and appreciate what good survey design entails have been most rewarding and have also contributed to the development of this book. I hope that they and future students will find this book to be of help to them and a continuing reference to some of those points that we have discussed.

Peter Stopher  
Blackheath, New South Wales  
August 2011

# Contents

---

<i>List of figures</i>	page xix
<i>List of tables</i>	xxii
<i>Acknowledgements</i>	xxv
<b>1 Introduction</b>	1
1.1 The purpose of this book	1
1.2 Scope of the book	2
1.3 Survey statistics	4
<b>2 Basic statistics and probability</b>	6
2.1 Some definitions in statistics	6
2.1.1 Censuses and surveys	7
2.2 Describing data	8
2.2.1 Types of scales	8
<i>Nominal scales</i>	8
<i>Ordinal scales</i>	9
<i>Interval scales</i>	9
<i>Ratio scales</i>	10
<i>Measurement scales</i>	10
2.2.2 Data presentation: graphics	11
2.2.3 Data presentation: non-graphical	16
<i>Measures of magnitude</i>	17
Frequencies and proportions	17
Central measures of data	21
Measures of dispersion	34
<i>The normal distribution</i>	45
<i>Some useful properties of variances and standard deviations</i>	46
Proportions or probabilities	47
Data transformations	48
Covariance and correlation	50
Coefficient of variation	51
	vii

Other measures of variability	53
<i>Alternatives to Sturges' rule</i>	62
<b>3 Basic issues in surveys</b>	64
3.1 Need for survey methods	64
3.1.1 A definition of sampling methodology	65
3.2 Surveys and censuses	65
3.2.1 Costs	66
3.2.2 Time	67
3.3 Representativeness	68
3.3.1 Randomness	69
3.3.2 Probability sampling	70
<i>Sources of random numbers</i>	71
3.4 Errors and bias	71
3.4.1 Sample design and sampling error	73
3.4.2 Bias	74
3.4.3 Avoiding bias	78
3.5 Some important definitions	78
<b>4 Ethics of surveys of human populations</b>	81
4.1 Why ethics?	81
4.2 Codes of ethics or practice	82
4.3 Potential threats to confidentiality	84
4.3.1 Retaining detail and confidentiality	85
4.4 Informed consent	86
4.5 Conclusions	89
<b>5 Designing a survey</b>	91
5.1 Components of survey design	91
5.2 Defining the survey purpose	93
5.2.1 Components of survey purpose	94
<i>Data needs</i>	94
Comparability or innovation	97
Defining data needs	99
Data needs in human subject surveys	99
<i>Survey timing</i>	100
<i>Geographic bounds for the survey</i>	101
5.3 Trade-offs in survey design	102
<b>6 Methods for conducting surveys of human populations</b>	104
6.1 Overview	104
6.2 Face-to-face interviews	105
6.3 Postal surveys	107



6.4	Telephone surveys	108
6.5	Internet surveys	111
6.6	Compound survey methods	112
6.6.1	Pre-recruitment contact	112
6.6.2	Recruitment	113
	<i>Random digit dialling</i>	115
6.6.3	Survey delivery	117
6.6.4	Data collection	118
6.6.5	An example	119
6.7	Mixed-mode surveys	120
6.7.1	Increasing response and reducing bias	123
6.8	Observational surveys	125
<b>7</b>	<b>Focus groups</b>	127
7.1	Introduction	127
7.2	Definition of a focus group	128
7.2.1	The size and number of focus groups	128
7.2.2	How a focus group functions	129
7.2.3	Analysing the focus group discussions	131
7.2.4	Some disadvantages of focus groups	131
7.3	Using focus groups to design a survey	132
7.4	Using focus groups to evaluate a survey	134
7.5	Summary	135
<b>8</b>	<b>Design of survey instruments</b>	137
8.1	Scope of this chapter	137
8.2	Question type	137
8.2.1	Classification and behaviour questions	138
	<i>Mitigating threatening questions</i>	139
8.2.2	Memory or recall error	142
8.3	Question format	145
8.3.1	Open questions	145
8.3.2	Field-coded questions	146
8.3.3	Closed questions	147
8.4	Physical layout of the survey instrument	150
8.4.1	Introduction	150
8.4.2	Question ordering	153
	<i>Opening questions</i>	153
	<i>Body of the survey</i>	154
	<i>The end of the questionnaire</i>	158
8.4.3	Some general issues on question layout	159
	<i>Overall format</i>	160

<i>Appearance of the survey</i>	161
<i>Front cover</i>	162
<i>Spatial layout</i>	163
<i>Choice of typeface</i>	164
<i>Use of colour and graphics</i>	166
<i>Question numbering</i>	169
<i>Page breaks</i>	170
<i>Repeated questions</i>	171
<i>Instructions</i>	172
<i>Show cards</i>	174
<i>Time of the interview</i>	174
<i>Precoding</i>	174
<i>End of the survey</i>	175
<i>Some final comments on questionnaire layout</i>	176
<b>9 Design of questions and question wording</b>	177
9.1 Introduction	177
9.2 Issues in writing questions	178
9.2.1 Requiring an answer	178
9.2.2 Ready answers	180
9.2.3 Accurate recall and reporting	181
9.2.4 Revealing the data	182
9.2.5 Motivation to answer	183
9.2.6 Influences on response categories	184
9.2.7 Use of categories and other responses	185
<i>Ordered and unordered categories</i>	187
9.3 Principles for writing questions	188
9.3.1 Use simple language	189
9.3.2 Number of words	190
9.3.3 Avoid using vague words	191
9.3.4 Avoid using 'Tick all that apply' formats	193
9.3.5 Develop response categories that are mutually exclusive and exhaustive	193
9.3.6 Make sure that questions are technically correct	195
9.3.7 Do not ask respondents to say 'Yes' in order to say 'No'	196
9.3.8 Avoid double-barrelled questions	196
9.4 Conclusion	197
<b>10 Special issues for qualitative and preference surveys</b>	199
10.1 Introduction	199
10.2 Designing qualitative questions	199
10.2.1 Scaling questions	200

10.3	Stated response questions	206
10.3.1	The hypothetical situation	206
10.3.2	Determining attribute levels	207
10.3.3	Number of choice alternatives or scenarios	207
10.3.4	Other issues of concern	208
	<i>Data inconsistency</i>	208
	<i>Lexicographic responses</i>	209
	<i>Random responses</i>	209
10.4	Some concluding comments on stated response survey design	210
<b>11</b>	<b>Design of data collection procedures</b>	<b>211</b>
11.1	Introduction	211
11.2	Contacting respondents	211
11.2.1	Pre-notification contacts	211
11.2.2	Number and type of contacts	213
	<i>Nature of reminder contacts</i>	213
	<i>Postal surveys</i>	215
	<i>Postal surveys with telephone recruitment</i>	216
	<i>Telephone interviews</i>	217
	<i>Face-to-face interviews</i>	219
	<i>Internet surveys</i>	220
11.3	Who should respond to the survey?	221
11.3.1	Targeted person	221
11.3.2	Full household surveys	223
	<i>Proxy reporting</i>	224
11.4	Defining a complete response	225
11.4.1	Completeness of the data items	226
11.4.2	Completeness of aggregate sampling units	228
11.5	Sample replacement	229
11.5.1	When to replace a sample unit	229
11.5.2	How to replace a sample	233
11.6	Incentives	235
11.6.1	Recommendations on incentives	236
11.7	Respondent burden	240
11.7.1	Past experience	241
11.7.2	Appropriate moment	242
11.7.3	Perceived relevance	242
11.7.4	Difficulty	243
	<i>Physical difficulty</i>	243
	<i>Intellectual difficulty</i>	244
	<i>Emotional difficulty</i>	245
	<i>Reducing difficulty</i>	246

11.7.5	External factors	246
	<i>Attitudes and opinions of others</i>	246
	<i>The 'feel good' effect</i>	247
	<i>Appropriateness of the medium</i>	248
11.7.6	Mitigating respondent burden	248
11.8	Concluding comments	250
<b>12</b>	<b>Pilot surveys and pretests</b>	<b>251</b>
12.1	Introduction	251
12.2	Definitions	252
12.3	Selecting respondents for pretests and pilot surveys	255
12.3.1	Selecting respondents	255
12.3.2	Sample size	258
	<i>Pilot surveys</i>	258
	<i>Pretests</i>	261
12.4	Costs and time requirements of pretests and pilot surveys	262
12.5	Concluding comments	264
<b>13</b>	<b>Sample design and sampling</b>	<b>265</b>
13.1	Introduction	265
13.2	Sampling frames	266
13.3	Random sampling procedures	268
13.3.1	Initial considerations	268
13.3.2	The normal law of error	269
13.4	Random sampling methods	270
13.4.1	Simple random sampling	271
	<i>Drawing the sample</i>	271
	<i>Estimating population statistics and sampling errors</i>	273
	Example	276
	Sampling from a finite population	279
	Sampling error of ratios and proportions	279
	<i>Defining the sample size</i>	281
	Examples	283
13.4.2	Stratified sampling	285
	<i>Types of stratified samples</i>	285
	Study domains and strata	287
	Weighted means and variances	287
	<i>Stratified sampling with a uniform sampling fraction</i>	289
	Drawing the sample	289
	Estimating population statistics and sampling errors	290
	Pre- and post-stratification	291
	Example	293

Equal allocation	294
Summary of proportionate sampling	295
<i>Stratified sampling with variable sampling fraction</i>	295
Drawing the sample	295
Estimating population statistics and sampling errors	296
Non-coincident study domains and strata	296
Optimum allocation and economic design	297
Example	298
Survey costs differing by stratum	300
Example	301
Practical issues in drawing disproportionate samples	303
Concluding comments on disproportionate sampling	305
13.4.3 Multistage sampling	305
<i>Drawing a multistage sample</i>	306
<i>Requirements for multistage sampling</i>	307
<i>Estimating population values and sampling statistics</i>	308
Example	309
<i>Concluding comments on multistage sampling</i>	314
13.5 Quasi-random sampling methods	314
13.5.1 Cluster sampling	316
<i>Equal clusters: population values and standard errors</i>	317
Example	319
The effects of clustering	321
<i>Unequal clusters: population values and standard errors</i>	322
Random selection of unequal clusters	324
Example	325
Stratified sampling of unequal clusters	326
Paired selection of unequal-sized clusters	327
13.5.2 Systematic sampling	328
<i>Population values and standard errors in a systematic sample</i>	328
Simple random model	329
Stratified random model	329
Paired selection model	329
Successive difference model	330
Example	330
13.5.3 Choice-based sampling	333
13.6 Non-random sampling methods	334
13.6.1 Quota sampling	334
13.6.2 Intentional, judgemental, or expert samples	335
13.6.3 Haphazard samples	335
13.6.4 Convenience samples	336
13.7 Summary	336

<b>14 Repetitive surveys</b>	337
14.1 Introduction	337
14.2 Non-overlapping samples	338
14.3 Incomplete overlap	339
14.4 Subsampling on the second and subsequent occasions	341
14.5 Complete overlap: a panel	342
14.6 Practical issues in designing and conducting panel surveys	343
14.6.1 Attrition	344
<i>Replacement of panel members lost by attrition</i>	345
<i>Reducing losses due to attrition</i>	346
14.6.2 Contamination	347
14.6.3 Conditioning	348
14.7 Advantages and disadvantages of panels	348
14.8 Methods for administering practical panel surveys	349
14.9 Continuous surveys	352
<b>15 Survey economics</b>	356
15.1 Introduction	356
15.2 Cost elements in survey design	357
15.3 Trade-offs in survey design	359
15.3.1 Postal surveys	360
15.3.2 Telephone recruitment with a postal survey with or without telephone retrieval	361
15.3.3 Face-to-face interview	362
15.3.4 More on potential trade-offs	362
15.4 Concluding comments	363
<b>16 Survey implementation</b>	365
16.1 Introduction	365
16.2 Interviewer selection and training	365
16.2.1 Interviewer selection	365
16.2.2 Interviewer training	368
16.2.3 Interviewer monitoring	369
16.3 Record keeping	370
16.4 Survey supervision	372
16.5 Survey publicity	373
16.5.1 Frequently asked questions, fact sheet, or brochure	374
16.6 Storage of survey forms	374
16.6.1 Identification numbers	375
16.7 Issues for surveys using posted materials	377
16.8 Issues for surveys using telephone contact	377
16.8.1 Caller ID	378
16.8.2 Answering machines	378

16.8.3 Repeated requests for callback	380
16.9 Data on incomplete responses	381
16.10 Checking survey responses	382
16.11 Times to avoid data collection	383
16.12 Summary comments on survey implementation	383
<b>17 Web-based surveys</b>	<b>385</b>
17.1 Introduction	385
17.2 The internet as an optional response mechanism	388
17.3 Some design issues for Web surveys	389
17.3.1 Differences between paper and internet surveys	389
17.3.2 Question and response	390
17.3.3 Ability to fill in the Web survey in multiple sittings	392
17.3.4 Progress tracking	393
17.3.5 Pre-filled responses	394
17.3.6 Confidentiality in Web-based surveys	395
17.3.7 Pictures, maps, etc. on Web surveys	395
<i>Animation in survey pictures and maps</i>	396
17.3.8 Browser software	396
<i>User interface design</i>	396
<i>Creating mock-ups</i>	397
<i>Page loading time</i>	398
17.4 Some design principles for Web surveys	398
17.5 Concluding comments	399
<b>18 Coding and data entry</b>	<b>401</b>
18.1 Introduction	401
18.2 Coding	402
18.2.1 Coding of missing values	402
18.2.2 Use of zeros and blanks in coding	403
18.2.3 Coding consistency	404
<i>Binary variables</i>	404
<i>Numeric variables</i>	404
18.2.4 Coding complex variables	405
18.2.5 Geocoding	406
<i>Requesting address details for other places than home</i>	408
<i>Pre-coding of buildings</i>	409
<i>Interactive gazetteers</i>	410
<i>Other forms of geocoding assistance</i>	410
<i>Locating by mapping software</i>	411
18.2.6 Methods for creating codes	412
18.3 Data entry	413
18.4 Data repair	416

<b>19</b>	<b>Data expansion and weighting</b>	418
19.1	Introduction	418
19.2	Data expansion	419
19.2.1	Simple random sampling	419
19.2.2	Stratified sampling	419
19.2.3	Multistage sampling	420
19.2.4	Cluster samples	420
19.2.5	Other sampling methods	421
19.3	Data weighting	421
19.3.1	Weighting with unknown population totals	422
	<i>An example</i>	423
	<i>A second example</i>	424
19.3.2	Weighting with known populations	426
	<i>An example</i>	427
19.4	Summary	429
<b>20</b>	<b>Nonresponse</b>	431
20.1	Introduction	431
20.2	Unit nonresponse	432
20.2.1	Calculating response rates	432
	<i>Classifying responses to a survey</i>	433
	<i>Calculating response rates</i>	435
20.2.2	Reducing nonresponse and increasing response rates	440
	<i>Design issues affecting nonresponse</i>	440
	<i>Survey publicity</i>	442
	<i>Use of incentives</i>	442
	<i>Use of reminders and repeat contacts</i>	443
	<i>Personalisation</i>	444
	<i>Summary</i>	445
20.2.3	Nonresponse surveys	445
20.3	Item nonresponse	450
20.3.1	Data repair	450
	<i>Flagging repaired variables</i>	451
	<i>Inference</i>	452
	<i>Imputation</i>	452
	Historical imputation	453
	Average imputation	454
	Ratio imputation	454
	Regression imputation	455
	Cold-deck imputation	456
	Hot-deck imputation	457
	Expectation maximisation	457



Multiple imputation	458
Imputation using neural networks	458
Summary of imputation methods	460
20.3.2 A final note on item nonresponse	460
<i>Strategies to obtain age and income</i>	461
Age	461
Income	462
<b>21 Measuring data quality</b>	464
21.1 Introduction	464
21.2 General measures of data quality	464
21.2.1 Missing value statistic	465
21.2.2 Data cleaning statistic	466
21.2.3 Coverage error	467
21.2.4 Sample bias	468
21.3 Specific measures of data quality	469
21.3.1 Non-mobility rates	469
21.3.2 Trip rates and activity rates	470
21.3.3 Proxy reporting	471
21.4 Validation surveys	472
21.4.1 Follow-up questions	473
21.4.2 Independent measurement	475
21.5 Adherence to quality measures and guidance	476
<b>22 Future directions in survey procedures</b>	478
22.1 Dangers of forecasting new directions	478
22.2 Some current issues	478
22.2.1 Reliance on telephones	478
<i>Threats to the use of telephone surveys</i>	479
<i>Conclusions on reliance on telephones</i>	481
22.2.2 Language and literacy	481
<i>Language</i>	481
<i>Literacy</i>	483
22.2.3 Mixed-mode surveys	486
22.2.4 Use of administrative data	487
22.2.5 Proxy reporting	488
22.3 Some possible future directions	489
22.3.1 A GPS survey as a potential substitute for a household travel survey	493
The effect of multiple observations of each respondent on sample size	495