

Cambridge University Press

978-0-521-86311-7 - Collecting, Managing, and Assessing Data Using Sample Surveys

Peter Stopher

Frontmatter

[More information](#)

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

---

---

*Collecting, Managing, and Assessing Data Using Sample Surveys* provides a thorough, step-by-step guide to the design and implementation of surveys. Beginning with a primer on basic statistics, the first half of the book takes readers on a comprehensive tour through the basics of survey design. Topics covered include the ethics of surveys, the design of survey procedures, the design of the survey instrument, how to write questions, and how to draw representative samples. Having shown readers how to design surveys, the second half of the book discusses a number of issues surrounding their implementation, including repetitive surveys, the economics of surveys, Web-based surveys, coding and data entry, data expansion and weighting, the issue of nonresponse, and the documenting and archiving of survey data. The book is an excellent introduction to the use of surveys for graduate students as well as a useful reference work for scholars and professionals.

PETER STOPHER is Professor of Transport Planning at the Institute of Transport and Logistics Studies at the University of Sydney. He has also been a professor at Northwestern University, Cornell University, McMaster University, and Louisiana State University. Professor Stopher has developed a substantial reputation in the field of data collection, particularly for the support of travel forecasting and analysis. He pioneered the development of travel and activity diaries as a data collection mechanism, and has written extensively on issues of sample design, data expansion, nonresponse biases, and measurement issues.

Cambridge University Press

978-0-521-86311-7 - Collecting, Managing, and Assessing Data Using Sample Surveys

Peter Stopher

Frontmatter

[More information](#)

---

Cambridge University Press

978-0-521-86311-7 - Collecting, Managing, and Assessing Data Using Sample Surveys

Peter Stopher

Frontmatter

[More information](#)

# Collecting, Managing, and Assessing Data Using Sample Surveys

---

---

Peter Stopher



**CAMBRIDGE**  
UNIVERSITY PRESS

Cambridge University Press  
978-0-521-86311-7 - Collecting, Managing, and Assessing Data Using Sample Surveys  
Peter Stopher  
Frontmatter  
[More information](#)

---

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.

Cambridge University Press

978-0-521-86311-7 - Collecting, Managing, and Assessing Data Using Sample Surveys

Peter Stopher

Frontmatter

[More information](#)

---

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

Cambridge University Press

978-0-521-86311-7 - Collecting, Managing, and Assessing Data Using Sample Surveys

Peter Stopher

Frontmatter

[More information](#)

---

## 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</b>	<b>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</b>	<b>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</b>	<b>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</b>	<b>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



	Contents	ix
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 Focus groups</b>		<b>127</b>
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 Design of survey instruments</b>		<b>137</b>
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</b>	<b>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</b>	<b>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

	Contents	xi
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 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

xii	<b>Contents</b>	
	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>	251
	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>	265
	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

	Contents	xiii
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</b>	<b>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</b>	<b>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</b>	<b>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

	Contents	xv
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



	Contents	xvii
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>	<b>464</b>	
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>	<b>478</b>	
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	

xviii      **Contents**

<b>23 Documenting and archiving</b>	499
23.1 Introduction	499
23.2 Documentation or the creation of metadata	499
23.2.1 Descriptive metadata	500
23.2.2 Preservation metadata	503
23.2.3 Geospatial metadata	503
23.3 Archiving of data	506
<i>References</i>	511
<i>Index</i>	525

## Figures

2.1	Scatter plot of odometer reading versus model year	page 12
2.2	Scatter plot of fuel type by body type	12
2.3	Pie chart of vehicle body types	13
2.4	Pie chart of household income groups	13
2.5	Histogram of household income	14
2.6	Histogram of vehicle types	14
2.7	Line graph of maximum and minimum temperatures for thirty days	15
2.8	Ogive of cumulative household income data from Figure 2.5	16
2.9	Relative ogive of household income	16
2.10	Relative step chart of household income	17
2.11	Stem and leaf display of income	22
2.12	Arithmetic mean as centre of gravity	24
2.13	Bimodal distribution of temperatures	25
2.14	Distribution of maximum temperatures from Table 2.4	29
2.15	Distribution of minimum temperatures from Table 2.4	30
2.16	Income distribution from Table 2.5	30
2.17	Distribution of vehicle counts	33
2.18	Box and whisker plot of income data from Table 2.5	36
2.19	Box and whisker plot of maximum temperatures	37
2.20	Box and whisker plot of minimum temperatures	37
2.21	Box and whisker plot of vehicles passing through the green phase	43
2.22	Box and whisker plot of children's ages	45
2.23	The normal distribution	45
2.24	Comparison of normal distributions with different variances	46
2.25	Scatter plot of maximum versus minimum temperature	52
2.26	A distribution skewed to the right	54
2.27	A distribution skewed to the left	54
2.28	Distribution with low kurtosis	55
2.29	Distribution with high kurtosis	55
3.1	Extract of random numbers from the RAND <i>Million Random Digits</i>	72
4.1	Example of a consent form	87

4.2	First page of an example subject information sheet	88
4.3	Second page of the example subject information sheet	89
5.1	Schematic of the survey process	92
5.2	Survey design trade-offs	103
6.1	Schematic of survey methods	113
8.1	Document file layout for booklet printing	162
8.2	Example of an unacceptable questionnaire format	164
8.3	Example of an acceptable questionnaire format	165
8.4	Excerpt from a survey showing arrows to guide respondent	168
8.5	Extract from a questionnaire showing use of graphics	169
8.6	Columned layout for asking identical questions about multiple people	171
8.7	Inefficient and efficient structures for organising serial questions	172
8.8	Instructions placed at the point to which they refer	173
8.9	Example of an unacceptable questionnaire format with response codes	175
9.1	Example of a sequence of questions that do not require answers	178
9.2	Example of a sequence of questions that do require answers	179
9.3	Example of a belief question	181
9.4	Example of a belief question with a more vague response	181
9.5	Two alternative response category sets for the age question	185
9.6	Alternative questions on age	186
9.7	Examples of questions with unordered response categories	187
9.8	An example of mixed ordered and unordered categories	188
9.9	Reformulated question from Figure 9.8	189
9.10	An unordered alternative to the question in Figure 9.8	189
9.11	Avoiding vague words in question wording	192
9.12	Example of a failure to achieve mutual exclusivity and exhaustiveness	194
9.13	Correction to mutual exclusivity and exhaustiveness	195
9.14	Example of a double negative	196
9.15	Example of removal of a double negative	196
9.16	An alternative that keeps the wording of the measure	197
9.17	An alternative way to deal with a double-barrelled question	197
10.1	Example of a qualitative question	200
10.2	Example of a qualitative question using number categories	200
10.3	Example of unbalanced positive and negative categories	201
10.4	Example of balanced positive and negative categories	201
10.5	Example of placing the neutral option at the end	202
10.6	Example of distinguishing the neutral option from 'No opinion'	202
10.7	Use of columned layout for repeated category responses	203
10.8	Alternative layout for repeated category responses	204
10.9	Statements that call for similar responses	204
10.10	Statements that call for varying responses	205
10.11	Rephrasing questions to remove requirement for 'Agree'/'Disagree'	206
11.1	Example of a postcard reminder for the first reminder	215

	<b>List of figures</b>	xxi
11.2	Framework for understanding respondent burden	241
14.1	Schematic of the four types of repetitive samples	338
14.2	Rotating panel showing recruitment, attrition, and rotation	353
18.1	An unordered set of responses requiring coding	402
18.2	A possible format for asking for an address	409
18.3	Excerpt from a mark-sensing survey	415
20.1	Illustration of the categorisation of response outcomes	436
20.2	Representation of a neural network model	459
23.1	Open archival information system model	508

## Tables

---

2.1	Frequencies and proportions of vehicle types	<i>page</i> 18
2.2	Frequencies, proportions, and cumulative values for household income	19
2.3	Minimum and maximum temperatures for a month (°C)	20
2.4	Grouped temperature data	21
2.5	Disaggregate household income data	22
2.6	Growth rates of an investment fund, 1993–2004	26
2.7	Speeds by kilometre for a train	27
2.8	Measurements of ball bearings	29
2.9	Number of vehicles passing through the green phase of a traffic light	32
2.10	Sorted number of vehicles passing through the green phase	32
2.11	Number of children by age	34
2.12	Deviations from the mean for the income data of Table 2.5	38
2.13	Outcomes from throwing the die twice	40
2.14	Sorted number of vehicles passing through the green phase	43
2.15	Deviations for vehicles passing through the green phase	44
2.16	Values of variance and standard deviation for values of $p$ and $q$	47
2.17	Deviations for vehicles passing through the green phase raised to third and fourth powers	57
2.18	Deviations from the mean for children's ages	58
2.19	Data on household size, annual income, and number of vehicles for forty households	59
2.20	Deviations needed for covariance and correlation estimates	61
3.1	Heights of 100 (fictitious) university students (cm)	76
3.2	Sample of the first and last five students	76
3.3	Sample of the first ten students	76
3.4	Intentional sample of ten students	77
3.5	Random sample of ten students (in order drawn)	77
3.6	Summary of results from Tables 3.2 to 3.5	77
6.1	Internet world usage statistics	112

	<b>List of tables</b>	xxiii
6.2	Mixed-mode survey types (based on Dillman and Tarnai, 1991)	121
11.1	Selection grid by age and gender	222
13.1	Partial listing of households for a simple random sample	272
13.2	Excerpt of random numbers from the RAND <i>Million Random Digits</i>	273
13.3	Selection of sample of 100 members using four-digit groups from Table 13.2	274
13.4	Data from twenty respondents in a fictitious survey	276
13.5	Sums of squares for population groups	286
13.6	Data for drawing an optimum household travel survey sample	299
13.7	Optimal allocation of the 2,000-household sample	299
13.8	Optimal allocation and expected sampling errors by stratum	300
13.9	Results of equal allocation for the household travel survey	300
13.10	Given information for economic design of the optimal allocation	301
13.11	Preliminary sample sizes and costs for economic design of the optimum allocation	301
13.12	Estimation of the final sample size and budget	302
13.13	Comparison of optimal allocation, equal allocation, and economic design for \$150,000 survey	302
13.14	Comparison of sampling errors from the three sample designs	303
13.15	Desired stratum sample sizes and results of recruitment calls	305
13.16	Distribution of departments and students	310
13.17	Two-stage sample of students from the university	311
13.18	Multistage sample using disproportionate sampling at the first stage	313
13.19	Calculations for standard error from sample in Table 13.18	315
13.20	Examples of cluster samples	316
13.21	Cluster sample of doctor's files	320
13.22	Random drawing of blocks of dwelling units	326
13.23	Calculations for paired selections and successive differences	332
18.1	Potential complex codes for income categories	406
18.2	Example codes for use of the internet and mobile phones	407
19.1	Results of an hypothetical household survey	424
19.2	Calculation of weights for the hypothetical household survey	424
19.3	Two-way distribution of completed surveys	424
19.4	Two-way distribution of terminated surveys	425
19.5	Table 19.3 expressed as percentages	425
19.6	Sum of the cells in Tables 19.3 and 19.4	425
19.7	Cells of Table 19.6 as percentages	426
19.8	Weights derived from Tables 19.7 and 19.5	426
19.9	Results of an hypothetical household survey compared to secondary source data	427
19.10	Two-way distribution of completed surveys by percentage (originally shown in Table 19.5)	427
19.11	Results of factoring the rows of Table 19.10	428

19.12	Second iteration, in which columns are factored	428
19.13	Third iteration, in which rows are factored again	429
19.14	Weights derived from the iterative proportional fitting	429
20.1	Final disposition codes for RDD telephone surveys	439
23.1	Preservation metadata elements and description	504



## *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

xxvi      **Acknowledgements**

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