

Contents

Preface	xv
1. Preview	1
1.1 Introduction	1
1.2 Different types of research study	1
1.3 Bias and confounding in observational studies	4
1.4 The sample survey	12
1.5 Data analysis	12
1.6 Experiments and quasi-experiments	13
1.7 Ethical considerations	15
1.8 Planning and carrying out a research study	16
2. The sample survey	17
2.1 Introduction	17
2.2 Simple random sampling	19
2.3 The distribution of sample means	20
2.4 Estimation of totals and proportions	23
2.5 Determining sample sizes with simple random sampling	26
2.6 Stratified sampling	28
2.7 Sample size allocation for stratified samples	30
2.8 Stratified sampling for proportions	32
2.9 Poststratification	33
2.10 Quota sampling	34
2.11 Ratio and regression estimation	34
2.12 Cluster sampling	37
2.13 Systematic sampling	38
2.14 Multi-stage sampling	39
2.15 Non-sampling errors	40
2.16 Biased procedures	41
2.17 Bias introduced during data collection	42
2.18 Processing errors	43
2.19 Methods of data collection	43

CONTENTS

2.20	Designing a questionnaire	44
2.21	Summary and further reading	46
	Exercises	47
	Case Study 1: The 1989 Mayoral election in Dunedin	47
	Appendix: Checklist for planning a sample survey	57
3.	Other sampling designs	58
3.1	Introduction	58
3.2	Mark–recapture sampling	58
3.3	Removal sampling	62
3.4	Transect sampling	66
3.5	Size-biased sampling	66
3.6	Sampling positions of objects in space	68
3.7	Distance-based methods	70
3.8	Quadrat counts	75
3.9	Sampling of quadrats	77
3.10	Summary and further reading	78
	Exercises	79
	Case Study 2: Estimating the size of a population of mice	79
	Case Study 3: Site location in prehistoric Thailand	82
4.	The linear regression model	87
4.1	Introduction	87
4.2	The simple linear regression model	87
4.3	Estimation and tests with simple linear regression	88
4.4	Multiple linear regression	92
4.5	The extra sum of squares principle	94
4.6	Assessing the assumptions of a regression model	98
4.7	Residual plots	100
4.8	Tests on residuals	107
4.9	The Durbin–Watson test	108
4.10	Summary and further reading	112
	Exercises	113
	Case Study 4: The loss of sales in a store due to a fire	120
5.	Experimental designs to assess the effect of a treatment	132
5.1	Introduction	132

CONTENTS

5.2	External and internal validity	133
5.3	Pseudoreplication	135
5.4	The one-group pretest–posttest design	137
5.5	The static group comparison	138
5.6	The importance of a comparable control group	139
5.7	Regression to the mean	140
5.8	The comparative posttest design	142
5.9	The comparative change design	144
5.10	The comparative change design without randomization	148
5.11	The regression discontinuity design	148
5.12	Regression adjustments for confounding variables	152
5.13	Matching	156
5.14	Summary and further reading	161
	Exercises	161
	Case Study 5: The Lanarkshire milk experiment	163
6.	Interrupted time series	167
6.1	Introduction	167
6.2	Analysis with the interrupted time series design	168
6.3	Analysis with the comparative time series design	175
6.4	Summary and further reading	177
	Exercises	178
	Case Study 6: Fluoridation in Birmingham	182
7.	More advanced experimental designs	188
7.1	Introduction	188
7.2	Factorial experiments and the analysis of variance	188
7.3	Some general considerations	194
7.4	Transformation of data	199
7.5	Experiments with three or more factors	201
7.6	Higher order interactions	204
7.7	Blocking	205
7.8	Latin squares	208
7.9	Nested factors	209
7.10	Split-plot experiments in agricultural research	210
7.11	Other uses of the split-plot design	213
7.12	Repeated measures experiments	215
7.13	The assumption of compound symmetry	217

CONTENTS

7.14	Analysis of variance by multiple regression	219
7.15	Ancillary analyses	224
7.16	Summary and further reading	225
	Exercises	225
	Case Study 7: An experiment on lockstitch seams	227
8.	Some special types of data	234
8.1	Introduction	234
8.2	The analysis of count data	236
8.3	Chi-squared tests on count data	237
8.4	McNemar's test for correlated proportions	242
8.5	Log-linear models	244
8.6	Extraneous variance	252
8.7	Using base rates	253
8.8	The analysis of proportion data	255
8.9	Survival data	261
8.10	Generalized linear modelling	263
8.11	Multivariate data	264
8.12	Summary and further reading	266
	Exercises	267
	Case Study 8: The location of prehistoric Maya sites	270
9.	Computer intensive statistics	286
9.1	Introduction	286
9.2	Randomization tests	288
9.3	Determining the number of randomizations required	292
9.4	Confidence limits by randomization	293
9.5	Monte Carlo tests	294
9.6	Jackknifing	298
9.7	Bootstrapping	302
9.8	Summary and further reading	304
	Exercise	305
	Case Study 9: The solar activity cycle in the past	305
10.	Ethical considerations	312
10.1	Introduction	312
10.2	Obligations to society	312
10.3	Obligations to funders and employers	313

CONTENTS

10.4 Obligations to colleagues	314
10.5 Obligations to human subjects	314
10.6 Experiments on animals	316
10.7 Summary and further reading	316
Case Study 10: A clinical trial on ECMO	317
11. Synthesis: carrying out a research study	321
11.1 Introduction	321
11.2 Deciding on objectives	321
11.3 Planning the collection of data	324
11.4 Sample size determination	325
11.5 Collecting data	334
11.6 Analysing data	334
11.7 The research report	335
References	336
Author index	345
Subject index	348