
Contents

Preface	VII
1 ANOVA and Ordinary Regression	1
1.1 Analysis of Variance	1
1.1.1 Synthetic Estimation	4
1.2 Ordinary Regression	7
1.2.1 Prediction	12
1.3 Model Diagnostics	16
1.3.1 Simulation-Based Diagnostics	19
1.4 Toward Causal Inference	23
1.5 Designing Regression Studies	24
1.6 Observational Studies and Experiments	26
1.6.1 Human Subjects	28
1.6.2 Observational Studies and Estimation of Effects	29
Suggested Reading	31
Problems and Exercises	31
2 Maximum Likelihood Estimation	37
2.1 Likelihood	37
2.1.1 Consistency	40
2.1.2 Asymptotic Efficiency and Normality	41
2.2 Sufficient Statistics	44
2.3 Synthetic Estimation	48
2.4 Model Selection	51
2.4.1 Hypothesis Testing	54
2.4.2 Inference Following Model Selection	55
2.5 Model Selection Criteria Related to Likelihood	57
Suggested Reading	62
Problems and Exercises	62

3	Sampling Methods	67
3.1	Preliminaries	67
3.2	Horvitz–Thompson Estimator	69
3.2.1	Simple Random Sampling	72
3.2.2	Systematic Sampling Designs	73
3.2.3	Some Other Sampling Designs	75
3.3	Stratification	76
3.4	Clustering	77
3.4.1	Two-Stage Clustered Sampling	78
3.5	Planned and Realised Sampling Designs	80
3.5.1	Adjusting and Trimming the Weights	81
3.6	Sample Size Calculation	85
3.7	Using Auxiliary Information	87
3.7.1	Fitting Models to Survey Data	88
3.7.2	Regression Estimators of the Population Mean	90
3.8	Small-Area Estimation	92
	Suggested Reading	96
	Problems and Exercises	97
4	The Bayesian Paradigm	103
4.1	The Updating Mechanism	103
4.1.1	Setting the Prior	104
4.1.2	Conjugate Priors	109
4.2	Computational Issues	109
4.2.1	Sampling from Multivariate Distributions	112
4.3	Coherence	114
4.4	Bayesian Study Design	118
4.5	Model Diagnostics and Prediction	121
	Suggested Reading	123
	Problems and Exercises	124
5	Incomplete Data	129
5.1	Terminology and Notation	129
5.2	Dealing with Incomplete Data	135
5.3	Nonresponse Mechanisms	138
5.3.1	Models for Incomplete Data	140
5.4	EM Algorithm	141
5.5	Multiple Imputation	144
5.5.1	Modelling Nonresponse	146
5.5.2	Working with Plausible Values	146
5.5.3	Monotone Response Patterns and Chained Equations	148
5.5.4	Sensitivity Analysis with Respect to NMAR	149
5.6	MI for Categorical and Longitudinal Data	150
5.7	Other Applications of the Missing-Data Idea	152
5.7.1	Finite Mixtures	153

5.7.2	Outliers and Contaminants: Data Editing	155
5.7.3	Balanced Complete Data	156
	Suggested Reading	156
	Problems and Exercises	157
6	Imperfect Measurement	163
6.1	The Measurement Process	163
6.1.1	Information About the Measurement Process	167
6.2	Attributes of a Good Manifest Variable	168
6.2.1	Impartiality	168
6.3	Linear Regression with Manifest Variables	170
6.3.1	Latent Outcome Variable	170
6.3.2	Latent Regression Variable	172
6.3.3	Estimating the Latent Values	174
6.4	Categorical Manifest Variables	175
6.5	Measurement Error as Incompleteness	177
6.6	Simulation–Extrapolation	183
6.7	Coarse Data	186
6.7.1	Inference with Coarse Variables	187
6.7.2	Bootstrap	190
	Suggested Reading	194
	Problems and Exercises	195
7	Experiments and Observational Studies	201
7.1	Comparing Treatments	201
7.1.1	Experimental Design	203
7.1.2	Assignment and Sampling Designs	207
7.1.3	SUTVA Assumptions and Cluster-Randomisation	207
7.1.4	The Scale for Comparison	208
7.2	Block-Randomisation	209
7.2.1	When and How to Block	211
7.2.2	From Sample to Population	212
7.3	Observational Studies	213
7.3.1	Matching	218
7.4	Imperfect Experiments	220
7.5	Comparing Institutions	222
7.5.1	Comparing Performances over Time	225
7.6	Regression Methods	226
	Suggested Reading	227
	Problems and Exercises	227

8	Clinical Trials	233
8.1	The Context	233
8.2	Models and Inference	234
8.2.1	Comparing Two Groups	237
8.3	Crossover Design	239
8.3.1	Estimation	241
8.3.2	Minimax Estimation	243
8.4	Treatment Heterogeneity	246
8.5	Bioequivalence	247
8.6	Incorporating Utilities	253
8.7	Example	255
	Suggested Reading	260
	Problems and Exercises	261
9	Random Coefficients	265
9.1	Introduction	265
9.2	Patterns of Variation	267
9.2.1	Invariance with Respect to Linear Transformations	268
9.3	Maximum Likelihood Estimation	270
9.3.1	Restricted Maximum Likelihood	276
9.3.2	Residuals	277
9.3.3	Borrowing Strength	278
9.3.4	EM Algorithm—A Connection to Missing Data	279
9.3.5	Technical Details	280
9.4	Model Validity	282
9.5	Inference About Variation	285
9.5.1	Confounding in Cluster-Level Variation	286
9.6	Multilevel Models and Other Extensions	287
9.7	Estimating Many Quantities	290
9.8	Some Applications	292
9.8.1	Small-Area Estimation	292
9.8.2	Performance Assessment of Institutions	293
9.8.3	Progression over Time	293
9.8.4	Studying Families	294
	Suggested Reading	295
	Problems and Exercises	295
10	Generalised Linear Models	301
10.1	Introduction	301
10.2	Examples of GLMs	306
10.3	Maximum Likelihood Estimation	307
10.3.1	Overdispersion	309
10.3.2	Model Selection	311
10.4	Residuals	315
10.4.1	Overall Assessment of Fit	317

10.5 Random Coefficients	318
10.5.1 Fitting GLMM	321
10.5.2 Exact Derivatives of the Log-Likelihood	323
10.5.3 Laplace Approximation	324
10.5.4 Some Generalisations and Alternatives	325
Suggested Reading	329
Problems and Exercises	330
11 Longitudinal and Time-Series Analysis	335
11.1 Introduction	335
11.2 Markov Property	340
11.3 Time Series	342
11.3.1 Moving Average	343
11.4 Targets, Designs, and Models	344
11.4.1 Panels	348
11.5 Irregular Time Points and Time-Specific Covariates	349
11.6 Analysis	352
11.6.1 Missing Values	354
11.6.2 Multivariate and Cluster-Longitudinal Data	355
11.6.3 Mixture Models	356
11.7 Example: House Prices in New Zealand	359
11.7.1 Adjustment for Rating Valuation	364
Suggested Reading	366
Problems and Exercises	367
12 Meta-Analysis and Estimating Many Quantities	371
12.1 Introduction	371
12.2 Studies with Identical Targets	373
12.3 Study-Specific Targets	375
12.4 Maximum Likelihood Estimation	378
12.5 Publication Bias	380
12.5.1 Funnel Plot	381
12.6 Inference for a Particular Context	384
12.6.1 Prediction for an Unrealised Context	387
12.7 Multivariate Meta-Analysis	387
12.8 Estimating Many Quantities	389
Suggested Reading	392
Problems and Exercises	392
Appendix. A Refresher	395
A.1 Populations and Variables	395
A.2 Replications and Randomness	399
A.2.1 Efficiency	400
A.3 Notation	402
A.4 Distributions	404

XVI Contents

A.4.1 Describing Distributions	406
A.4.2 Approximating the Distribution by a Histogram	410
A.5 Sampling Design	411
A.5.1 Complex Sampling Designs	413
A.5.2 Sampling Frame	413
A.5.3 The Planned and Realised Sampling Processes	414
A.6 Measurement Processes	416
A.7 Infinite Populations	419
A.7.1 Continuous Distributions	421
A.7.2 Superpopulations: Models	423
A.8 Distributions	424
A.8.1 Simulations	427
A.9 Classes of Distributions and Models	428
A.10 Normal Distributions	429
A.10.1 Log-Normal Distributions	431
A.11 Uniform Distributions	432
A.12 Beta and Gamma Distributions	433
A.13 Classes of Discrete Distributions	434
A.13.1 Discrete Uniform Distributions	435
A.14 Discrete Bivariate Distributions	436
A.14.1 Conditional Distributions	437
A.15 Bivariate Continuous Distributions	438
A.16 Operating with Bivariate Distributions	440
A.17 Random Samples	443
A.18 Regression	444
A.19 Multivariate Distributions	445
A.19.1 Multivariate Normal Distributions	446
A.19.2 Regression with Normally Distributed Variables	447
A.20 Formulating Inferences	448
A.20.1 Confidence Intervals	449
Problems and Exercises	451
References	459
Index	469



<http://www.springer.com/978-0-387-98735-4>

Studying Human Populations
An Advanced Course in Statistics
Longford, N.T.
2008, XVI, 474 p., Hardcover
ISBN: 978-0-387-98735-4