# Contents

List of Tables xxi
List of Figures xxv
Preface xxxi

I Preliminaries 1

1 Review of linear regression 3
1.1 Introduction . . . 3
1.2 Is there gender discrimination in faculty salaries? . . . 3
1.3 Independent-samples t test . . . 4
1.4 One-way analysis of variance . . . 8
1.5 Simple linear regression . . . 11
1.6 Dummy variables . . . 18
1.7 Multiple linear regression . . . 20
1.8 Interactions . . . 26
1.9 Dummies for more than two groups . . . 29
1.10 Other types of interactions . . . 34
1.10.1 Interaction between dummy variables . . . 34
1.10.2 Interaction between continuous covariates . . . 35
1.11 Nonlinear effects . . . 36
1.12 Residual diagnostics . . . 38
1.13 Summary and further reading . . . 39
1.14 Exercises . . . 40

II Two-level linear models 49

2 Variance-components models 51

2.1 Introduction . . . 51
2.2 How reliable are peak-expiratory-flow measurements? . . . 52
2.3 The variance-components model . . . 54
2.3.1 Model specification and path diagram . . . 54
2.3.2 Error components, variance components, and reliability . . . 58
2.3.3 Intraclass correlation . . . 58
2.4 Fixed versus random effects . . . 61
2.5 Estimation using Stata . . . 62
2.5.1 Data preparation . . . 62
2.5.2 Using xtreg . . . 63
2.5.3 Using xtmixedt . . . 65
2.5.4 Using gllamm . . . 66
2.6 Hypothesis tests and confidence intervals . . . 68
2.6.1 Hypothesis test and confidence interval for the population mean . . . 68
2.6.2 Hypothesis test and confidence interval for the between-cluster variance . . . 69
2.7 More on statistical inference . . . 71
2.7.1 ∴ Different estimation methods . . . 71
2.7.2 Inference for *(3* . . . 72
Estimate and standard error: Balanced case . . . 72
Estimate: Unbalanced case . . . 74
2.8 Crossed versus nested effects . . . 75
2.9 Assigning values to the random intercepts . . . 77
2.9.1 Maximum likelihood estimation . . . 78
Implementation via OLS regression . . . 78
Implementation via the mean total residual . . . 79
2.9.2 Empirical Bayes prediction . . . 80
2.9.3 ∴ Empirical Bayes variances . . . 83
2.10 Summary and further reading . . . 85
2.11 Exercises . . . 86

**3 Random-intercept models with covariates 91**

3.1 Introduction . . . 91

3.2 Does smoking during pregnancy affect birthweight? . . . 91

3.3 The linear random-intercept model with covariates . . . 94

3.3.1 Model specification . . . 94

3.3.2 Residual variance and intraclass correlation . . . 96

3.4 Estimation using Stata . . . 97

3.4.1 Using xtreg . . . 97

3.4.2 Using xtmixed . . . 99

3.4.3 Using gllamm . . . 100

3.5 Coefficients of determination or variance explained . . . 102

3.6 Hypothesis tests and confidence intervals . . . 104

3.6.1 Hypothesis tests for regression coefficients . . . 104

Hypothesis tests for individual regression coefficients . . . . 105

Joint hypothesis tests for several regression coefficients . . . 105

3.6.2 Predicted means and confidence intervals . . . 107

3.6.3 Hypothesis test for between-cluster variance . . . 108

3.7 Between and within effects . . . 109

3.7.1 Between-mother effects . . . 109

3.7.2 Within-mother effects . . . 111

3.7.3 Relations among estimators . . . 113

3.7.4 Endogeneity and different within- and between-mother effects 114

3.7.5 Hausman endogeneity test . . . 122

3.8 Fixed versus random effects revisited . . . 124

3.9 Residual diagnostics . . . 125

3.10 More on statistical inference for regression coefficients . . . 129

3.10.1 Consequences of using ordinary regression for clustered data 129

3.10.2 ♣ Power and sample-size determination . . . 130

3.11 Summary and further reading . . . 132

3.12 Exercises . . . 132

**4 Random-coefficient models 141**

4.1 Introduction . . . 141

4.2 How effective are different schools? . . . 141

4.3 Separate linear regressions for each school . . . 142

4.4 Specification and interpretation of a random-coefficient model . . . . 146

4.4.1 Specification of random-coefficient model . . . 146

4.4.2 Interpretation of the random-effects variances and covariances 150

4.5 Estimation using Stata . . . 153

4.5.1 Using xtmixed . . . 153

Random-intercept model . . . 153

Random-coefficient model . . . 155

4.5.2 Using gllamm . . . 157

Random-intercept model . . . 157

Random-coefficient model . . . 157

4.6 Testing the slope variance . . . 159

4.7 Interpretation of estimates . . . 159

4.8 Assigning values to the random intercepts and slopes . . . 161

4.8.1 Maximum likelihood estimation . . . 161

4.8.2 Empirical Bayes prediction . . . 162

4.8.3 Model visualization . . . 164

4.8.4 Residual diagnostics . . . 165

4.8.5 Inferences for individual schools . . . 167

4.9 Two-stage model formulation . . . 168

4.10 Some warnings about random-coefficient models . . . 171

4.11 Summary and further reading . . . 173

4.12 Exercises . . . 173

**5 Longitudinal, panel, and growth-curve models 179**

**5.1** Introduction . . . 179

**5.2** How and why do wages change over time? . . . 180

5.3 Data structure . . . 181

5.3.1 Missing data . . . 181
5.3.2 Time-varying and time-constant variables . . . 181
5.4 Time scales in longitudinal data . . . 182
5.5 Random- and fixed-effects approaches . . . 185
5.5.1 Correlated residuals . . . 185
5.5.2 Fixed-intercept model . . . 186
Using xtreg . . . 186
Using anova . . . 189
5.5.3 Random-intercept model . . . 192
5.5.4 Random-coefficient model . . . 194
5.5.5 Marginal mean and covariance structure induced by random effects . . . 196
Marginal mean and covariance structure for random-intercept models . . . 196
Marginal mean and covariance structure for random-coefficient models . . . 198
5.6 Marginal modeling . . . 199
5.6.1 Covariance structures . . . 199
Compound symmetric or exchangeable structure . . . 200
Random-coefficient structure . . . 201
Autoregressive residual structure . . . 201
Unstructured covariance matrix . . . 201
5.6.2 Marginal modeling using Stata . . . 202
5.7 Autoregressive- or lagged-response models . . . 204
5.8 Hybrid approaches . . . 206
5.8.1 Autoregressive response and random effects . . . 206
5.8.2 ♠ Autoregressive responses and autoregressive residuals . . . 206
5.8.3 Autoregressive residuals and random or fixed effects . . . . 207
5.9 Missing data . . . 207
5.9.1 ♠ Maximum likelihood estimation under MAR: A simulation 207
5.10 How do children grow? . . . 210

5.10.1 Observed growth trajectories . . . 210
5.11 Growth-curve modeling . . . 211
5.11.1 Random-intercept model . . . 211
5.11.2 Random-coefficient model . . . 213
5.11.3 Two-stage model formulation . . . 216
5.12 Prediction of trajectories for individual children . . . 217
5.13 Prediction of mean growth trajectory and 95% band . . . 219
5.14 ♣ Complex level-1 variation or heteroskedasticity . . . 220
5.15 Summary and further reading . . . 222
5.16 Exercises . . . 223

III Two-level generalized linear models 229

**6 Dichotomous or binary responses 231**

6.1 Introduction . . . 231
6.2 Single-level models for dichotomous responses . . . 231
6.2.1 Generalized linear model formulation . . . 232
6.2.2 Latent-response formulation . . . 238
Logistic regression . . . 238
Probit regression . . . 239
6.3 Which treatment is best for toenail infection? . . . 242
6.4 Longitudinal data structure . . . 242
6.5 Population-averaged or marginal probabilities . . . 243
6.6 Random-intercept logistic regression . . . 247
6.7 Estimation of logistic random-intercept models . . . 248
6.7.1 Using xtlogit . . . 248
6.7.2 Using xtmelogit . . . 251
6.7.3 Using gllamm . . . 251
6.8 Inference for logistic random-intercept models . . . 252
6.9 Subject-specific vs. population-averaged relationships . . . 254
6.10 Measures of dependence and heterogeneity . . . 256

6.10.1 Conditional or residual intraclass correlation of the latent responses . . . 256
6.10.2 Median odds ratio . . . 257
6.11 Maximum likelihood estimation . . . 258
6.11.1 ♣ Adaptive quadrature . . . 258
6.11.2 Some speed considerations . . . 261
Advice for speeding up gllamm . . . 263
6.12 Assigning values to random effects . . . 264
6.12.1 Maximum likelihood estimation . . . 264
6.12.2 Empirical Bayes prediction . . . 265
6.12.3 Empirical Bayes modal prediction . . . 266
6.13 Different kinds of predicted probabilities . . . 267
6.13.1 Predicted population-averaged probabilities . . . 267
6.13.2 Predicted subject-specific probabilities . . . 268
Predictions for hypothetical subjects . . . 268
Predictions for the subjects in the sample . . . 269
6.14 Other approaches to clustered dichotomous data . . . 271
6.14.1 Conditional logistic regression . . . 271
6.14.2 Generalized estimating equations (GEE) . . . 273
6.15 Summary and further reading . . . 274
6.16 Exercises . . . 275
7 Ordinal responses 287
7.1 Introduction . . . 287
7.2 Single-level cumulative models for ordinal responses . . . 287
7.2.1 Generalized linear model formulation . . . 287
7.2.2 Latent-response formulation . . . 288
7.2.3 Proportional odds . . . 292
7.2.4 •> Identification . . . 292
7.3 Are antipsychotic drugs effective for patients with schizophrenia? . . . 295
7.4 • Longitudinal data structure and graphs . . . 295

7A.1 Longitudinal data structure . . . 296
7.4.2 Plotting cumulative proportions . . . 296
7.4.3 Plotting estimated cumulative logits and transforming the time scale . . . 298
7.5 A single-level proportional odds model . . . 299
7.5.1 Model specification . . . 299
7.5.2 Estimation using Stata . . . 300
7.6 A random-intercept proportional odds model . . . 303
7.6.1 Model specification . . . 303
7.6.2 Estimation using Stata . . . 303
7.7 A random-coefficient proportional odds model . . . 305
7.7.1 Model specification . . . 305
7.7.2 Estimation using gllamm . . . 305
7.8 Different kinds of predicted probabilities . . . 307
7.8.1 Predicted population-averaged probabilities . . . 307
7.8.2 Predicted patient-specific probabilities . . . 309
7.9 Do experts differ in their grading of student essays? . . . 312
7.10 A random-intercept probit model with grader bias . . . 312
7.10.1 Model specification . . . 312
7.10.2 Estimation . . . 313
7.11 Including grader-specific measurement error variances . . . 315
7.11.1 Model specification . . . 315
7.11.2 Estimation . . . 315
7.12 Including grader-specific thresholds . . . 317
7.12.1 Model specification . . . 317
7.12.2 Estimation . . . 318
7.13 Summary and further reading . . . 323
7.14 Exercises . . . 324
**8 Discrete-time survival 331**
8.1 Introduction . . . 331

8.1.1 Censoring and truncation . . . 331
8.1.2 Time-varying covariates and different time-scales . . . 333
8.1.3 Discrete- versus continuous-time survival data . . . 334
8.2 Single-level models for discrete-time survival data . . . 334
8.2.1 Discrete-time hazard and discrete-time survival . . . 334
8.2.2 Data expansion for discrete-time survival analysis . . . 336
8.2.3 Estimation via regression models for dichotomous responses 338
8.2.4 Including covariates . . . 342
Time-constant covariates . . . 342
Time-varying covariates . . . 346
8.2.5 Handling left-truncated data . . . 350
8.3 How does birth history affect child mortality? . . . 352
8.4 Data expansion . . . 352
8.5 ∴ Proportional hazards and interval censoring . . . 354
8.6 Complementary log-log models . . . 356
8.7 A random-intercept complementary log-log model . . . 360
8.7.1 Model specification . . . 360
8.7.2 Estimation using Stata . . . 361
8.8 ∴ Marginal and conditional survival probabilities . . . 362
8.9 Summary and further reading . . . 366
8.10 Exercises . . . 367
9 Counts 373
9.1 Introduction . . . 373
9.2 What are counts? . . . 373
9.2.1 Counts versus proportions . . . 373
9.2.2 Counts as aggregated event-history data . . . 374
9.3 Single-level Poisson models for counts . . . 374
9.4 Did the German health-care reform reduce the number of doctor visits? . . . 376
9.5 ' Longitudinal data structure . . . 377

9.6 Single-level Poisson regression . . . 378
9.6.1 Model specification . . . 378
9.6.2 Estimation using Stata ; . . . 378
9.7 Random-intercept Poisson regression . . . 380
9.7.1 Model specification . . . 380
9.7.2 Estimation using Stata . . . 381
Using xtpoisson . . . 382
Using xtmepoisson . . . 383
Using gllamm . . . 384
9.8 Random-coefficient Poisson regression . . . 385
9.8.1 Model specification . . . 385
9.8.2 Estimation using Stata . . . 386
Using xtmepoisson . . . 386
Using gllamm . . . 387
9.8.3 Interpretation of estimates . . . 388
9.9 Overdispersion in single-level models . . . 389
9.9.1 Normally distributed random intercept . . . 389
9.9.2 Negative binomial models . . . 390
Mean dispersion or NB2 . . . 391
Constant dispersion or NB1 . . . 392
9.9.3 Quasilikelihood or robust standard errors . . . 392
9.10 Level-1 overdispersion in two-level models . . . 394
9.11 Other approaches to two-level count data . . . 395
9.11.1 Conditional Poisson regression . . . 395
9.11.2 Conditional negative binomial regression . . . 396
9.11.3 Generalized estimating equations . . . 396
9.11.4 Marginal and conditional estimates when responses are MAR 397
• Simulation . . . 397
9.12 How does birth history affect child mortality? . . . 401
9.12.1 Simple piecewise exponential survival model . . . 401

9.12.2 Piecewise exponential survival model with covariates and frailty . . . 404
9.13 Which Scottish counties have a high risk of lip cancer? . . . 407
9.14 Standardized mortality ratios . . . 407
9.15 Random-intercept Poisson regression . . . 410
9.15.1 Model specification . . . 410
9.15.2 Estimation using gllamm . . . 410
9.15.3 Prediction of standardized mortality ratios . . . 411
9.16 V Nonparametric maximum likelihood estimation . . . 414
9.16.1 Specification . . . 414
9.16.2 Estimation using gllamm . . . 414
9.16.3 Prediction . . . 419
9.17 Summary and further reading . . . 419
9.18 Exercises . . . 420

IV Models with nested and crossed random effects 429

**10 Higher-level models with nested random** effects **431**

10.1 Introduction . . . 431
10.2 Do peak-expiratory-flow measurements vary between methods? . . . 432
10.3 Two-level variance-components models . . . 433
10.3.1 Model specification . . . 433
10.3.2 Estimation using xtmixed . . . 433
10.4 Three-level variance-components models . . . 436
10.4.1 Model specification . . . 436
10.4.2 Different types of intraclass correlation . . . 438
10.4.3 Three-stage formulation . . . 439
10.4.4 Estimation using xtmixed . . . 439
10.4.5 Empirical Bayes prediction using xtmixed . . . 442
10.5 Did the Guatemalan immunization campaign work? . . . 443
10.6 A three-level logistic random-intercept model . . . 444
10.6.1 Model specification . . . 444

10.6.2 Different types of intraclass correlations for the latent responses . . . 445
10.6.3 Different kinds of median odds ratios . . . 445
10.6.4 Three-stage formulation . . . 446
10.7 Estimation of three-level logistic random-intercept models using Stata 446
10.7.1 Using gllamm . . . 446
10.7.2 Using xtmelogit . . . 450
10.8 A three-level logistic random-coefficient model . . . 453
10.9 Estimation of three-level logistic random-coefficient models using Stata . . . 454
10.9.1 Using gllamm . . . 454
10.9.2 Using xtmelogit . . . 458
10.10 Prediction of random effects . . . 459
10.10.1 Empirical Bayes prediction . . . 459
10.10.2 Empirical Bayes modal prediction . . . 460
10.11 Different kinds of predicted probabilities . . . 461
10.11.1 Predicted marginal probabilities . . . 461
10.11.2 Predicted median or conditional probabilities . . . 461
10.11.3 Predicted posterior mean probabilities . . . 462
10.12 Summary and further reading . . . 463
10.13 Exercises . . . 463
11 Crossed random effects 473
11.1 Introduction . . . 473
11.2 How does investment depend on expected profit and capital stock? . 474
11.3 A two-way error-components model . . . 475
11.3.1 Model specification . . . 475
11.3.2 Residual intraclass correlations . . . 476
11.3.3 Estimation . . . 476
11.3.4 Prediction . . . 478
11.4 How much do primary and secondary schools affect attainment at age 16? . . . 481

11.5 An additive crossed random-effects model . . . 483
11.5.1 Specification . . . 483
11.5.2 Estimation using xtmixed . . . 484
11.6 Including a random interaction . . . 485
11.6.1 Model specification . . . 485
11.6.2 Intraclass correlations . . . 485
11.6.3 Estimation using xtmixed . . . 486
11.6.4 Some diagnostics . . . 488
11.7 ∴ A trick requiring fewer random effects . . . 489
11.8 Do salamanders from different populations mate successfully? . . . . 493
11.9 Crossed random-effects logistic regression . . . 495
11.10 Summary and further reading . . . 500
11.11 Exercises . . . 501

**A Syntax for gllamm, eq, and gllapred: The bare essentials 509**

**B Syntax for gllamm 515**

**C Syntax for gllapred 527**

**D Syntax for gllasim 531**

**References 535**

**Author index 549**

**Subject index 555**