

# Contents

	List of Tables	xvii
	List of Figures	xix
	Preface	xxv
	Multilevel and longitudinal models: When and why?	1
<b>I</b>	<b>Preliminaries</b>	<b>9</b>
<b>1</b>	<b>Review of linear regression</b>	<b>11</b>
	1.1 Introduction . . . . .	11
	1.2 Is there gender discrimination in faculty salaries? . . . . .	11
	1.3 Independent-samples t test . . . . .	12
	1.4 One-way analysis of variance . . . . .	17
	1.5 Simple linear regression . . . . .	19
	1.6 Dummy variables . . . . .	27
	1.7 Multiple linear regression . . . . .	30
	1.8 Interactions . . . . .	36
	1.9 Dummy variables for more than two groups . . . . .	42
	1.10 Other types of interactions . . . . .	48
	1.10.1 Interaction between dummy variables . . . . .	48
	1.10.2 Interaction between continuous covariates . . . . .	50
	1.11 Nonlinear effects . . . . .	52
	1.12 Residual diagnostics . . . . .	54
	1.13 ❖ Causal and noncausal interpretations of regression coefficients . .	56
	1.13.1 Regression as conditional expectation . . . . .	56
	1.13.2 Regression as structural model . . . . .	57

1.14	Summary and further reading . . . . .	59
1.15	Exercises . . . . .	60
<b>II</b>	<b>Two-level models</b>	<b>71</b>
<b>2</b>	<b>Variance-components models</b>	<b>73</b>
2.1	Introduction . . . . .	73
2.2	How reliable are peak-expiratory-flow measurements? . . . . .	74
2.3	Inspecting within-subject dependence . . . . .	75
2.4	The variance-components model . . . . .	77
2.4.1	Model specification . . . . .	77
2.4.2	Path diagram . . . . .	78
2.4.3	Between-subject heterogeneity . . . . .	79
2.4.4	Within-subject dependence . . . . .	80
	Intraclass correlation . . . . .	80
	Intraclass correlation versus Pearson correlation . . . . .	81
2.5	Estimation using Stata . . . . .	82
2.5.1	Data preparation: Reshaping to long form . . . . .	83
2.5.2	Using xtreg . . . . .	84
2.5.3	Using xtmixed . . . . .	85
2.6	Hypothesis tests and confidence intervals . . . . .	87
2.6.1	Hypothesis test and confidence interval for the population mean . . . . .	87
2.6.2	Hypothesis test and confidence interval for the between-cluster variance . . . . .	88
	Likelihood-ratio test . . . . .	88
	❖ Score test . . . . .	89
	F test . . . . .	92
	Confidence intervals . . . . .	92
2.7	Model as data-generating mechanism . . . . .	93
2.8	Fixed versus random effects . . . . .	95
2.9	Crossed versus nested effects . . . . .	97

2.10	Parameter estimation . . . . .	99
2.10.1	Model assumptions . . . . .	99
	Mean structure and covariance structure . . . . .	100
	Distributional assumptions . . . . .	101
2.10.2	Different estimation methods . . . . .	101
2.10.3	Inference for $\beta$ . . . . .	103
	Estimate and standard error: Balanced case . . . . .	103
	Estimate: Unbalanced case . . . . .	105
2.11	Assigning values to the random intercepts . . . . .	106
2.11.1	Maximum “likelihood” estimation . . . . .	106
	Implementation via OLS regression . . . . .	107
	Implementation via the mean total residual . . . . .	108
2.11.2	Empirical Bayes prediction . . . . .	109
2.11.3	Empirical Bayes standard errors . . . . .	113
	Comparative standard errors . . . . .	113
	Diagnostic standard errors . . . . .	114
2.12	Summary and further reading . . . . .	115
2.13	Exercises . . . . .	116
<b>3</b>	<b>Random-intercept models with covariates</b>	<b>123</b>
3.1	Introduction . . . . .	123
3.2	Does smoking during pregnancy affect birthweight? . . . . .	123
	3.2.1 Data structure and descriptive statistics . . . . .	125
3.3	The linear random-intercept model with covariates . . . . .	127
	3.3.1 Model specification . . . . .	127
	3.3.2 Model assumptions . . . . .	128
	3.3.3 Mean structure . . . . .	130
	3.3.4 Residual variance and intraclass correlation . . . . .	130
	3.3.5 Graphical illustration of random-intercept model . . . . .	131
3.4	Estimation using Stata . . . . .	131
	3.4.1 Using xtreg . . . . .	132

3.4.2	Using xtmixed . . . . .	133
3.5	Coefficients of determination or variance explained . . . . .	134
3.6	Hypothesis tests and confidence intervals . . . . .	138
3.6.1	Hypothesis tests for regression coefficients . . . . .	138
	Hypothesis tests for individual regression coefficients . . . . .	138
	Joint hypothesis tests for several regression coefficients . . . . .	139
3.6.2	Predicted means and confidence intervals . . . . .	140
3.6.3	Hypothesis test for random-intercept variance . . . . .	142
3.7	Between and within effects of level-1 covariates . . . . .	142
3.7.1	Between-mother effects . . . . .	143
3.7.2	Within-mother effects . . . . .	145
3.7.3	Relations among estimators . . . . .	147
3.7.4	Level-2 endogeneity and cluster-level confounding . . . . .	149
3.7.5	Allowing for different within and between effects . . . . .	152
3.7.6	Hausman endogeneity test . . . . .	157
3.8	Fixed versus random effects revisited . . . . .	158
3.9	Assigning values to random effects: Residual diagnostics . . . . .	160
3.10	More on statistical inference . . . . .	164
3.10.1	❖ Overview of estimation methods . . . . .	164
3.10.2	Consequences of using standard regression modeling for clustered data . . . . .	167
3.10.3	❖ Power and sample-size determination . . . . .	168
3.11	Summary and further reading . . . . .	171
3.12	Exercises . . . . .	172
<b>4</b>	<b>Random-coefficient models</b> . . . . .	<b>181</b>
4.1	Introduction . . . . .	181
4.2	How effective are different schools? . . . . .	181
4.3	Separate linear regressions for each school . . . . .	182
4.4	Specification and interpretation of a random-coefficient model . . . . .	188
4.4.1	Specification of a random-coefficient model . . . . .	188

4.4.2	Interpretation of the random-effects variances and co-variances . . . . .	191
4.5	Estimation using xtmixed . . . . .	194
4.5.1	Random-intercept model . . . . .	194
4.5.2	Random-coefficient model . . . . .	196
4.6	Testing the slope variance . . . . .	197
4.7	Interpretation of estimates . . . . .	198
4.8	Assigning values to the random intercepts and slopes . . . . .	200
4.8.1	Maximum “likelihood” estimation . . . . .	200
4.8.2	Empirical Bayes prediction . . . . .	201
4.8.3	Model visualization . . . . .	203
4.8.4	Residual diagnostics . . . . .	204
4.8.5	Inferences for individual schools . . . . .	207
4.9	Two-stage model formulation . . . . .	210
4.10	Some warnings about random-coefficient models . . . . .	213
4.10.1	Meaningful specification . . . . .	213
4.10.2	Many random coefficients . . . . .	213
4.10.3	Convergence problems . . . . .	214
4.10.4	Lack of identification . . . . .	214
4.11	Summary and further reading . . . . .	215
4.12	Exercises . . . . .	216

**III Models for longitudinal and panel data 225**

**Introduction to models for longitudinal and panel data (part III) 227**

**5 Subject-specific effects and dynamic models 247**

5.1	Introduction . . . . .	247
5.2	Conventional random-intercept model . . . . .	248
5.3	Random-intercept models accommodating endogenous covariates . .	250
5.3.1	Consistent estimation of effects of endogenous time-varying covariates . . . . .	250

5.3.2	Consistent estimation of effects of endogenous time-varying and endogenous time-constant covariates . . .	253
5.4	Fixed-intercept model . . . . .	257
5.4.1	Using xtreg or regress with a differencing operator . . . . .	259
5.4.2	❖ Using anova . . . . .	262
5.5	Random-coefficient model . . . . .	265
5.6	Fixed-coefficient model . . . . .	267
5.7	Lagged-response or dynamic models . . . . .	269
5.7.1	Conventional lagged-response model . . . . .	269
5.7.2	❖ Lagged-response model with subject-specific intercepts .	273
5.8	Missing data and dropout . . . . .	278
5.8.1	❖ Maximum likelihood estimation under MAR: A simulation . . . . .	279
5.9	Summary and further reading . . . . .	282
5.10	Exercises . . . . .	283
<b>6</b>	<b>Marginal models</b>	<b>293</b>
6.1	Introduction . . . . .	293
6.2	Mean structure . . . . .	293
6.3	Covariance structures . . . . .	294
6.3.1	Unstructured covariance matrix . . . . .	298
6.3.2	Random-intercept or compound symmetric/exchangeable structure . . . . .	303
6.3.3	Random-coefficient structure . . . . .	305
6.3.4	Autoregressive and exponential structures . . . . .	308
6.3.5	Moving-average residual structure . . . . .	311
6.3.6	Banded and Toeplitz structures . . . . .	313
6.4	Hybrid and complex marginal models . . . . .	316
6.4.1	Random effects and correlated level-1 residuals . . . . .	316
6.4.2	Heteroskedastic level-1 residuals over occasions . . . . .	317
6.4.3	Heteroskedastic level-1 residuals over groups . . . . .	318
6.4.4	Different covariance matrices over groups . . . . .	321

6.5	Comparing the fit of marginal models . . . . .	322
6.6	Generalized estimating equations (GEE) . . . . .	325
6.7	Marginal modeling with few units and many occasions . . . . .	327
6.7.1	Is a highly organized labor market beneficial for economic growth? . . . . .	328
6.7.2	Marginal modeling for long panels . . . . .	329
6.7.3	Fitting marginal models for long panels in Stata . . . . .	329
6.8	Summary and further reading . . . . .	332
6.9	Exercises . . . . .	333
<b>7</b>	<b>Growth-curve models</b>	<b>343</b>
7.1	Introduction . . . . .	343
7.2	How do children grow? . . . . .	343
7.2.1	Observed growth trajectories . . . . .	344
7.3	Models for nonlinear growth . . . . .	345
7.3.1	Polynomial models . . . . .	345
	Fitting the models . . . . .	346
	Predicting the mean trajectory . . . . .	349
	Predicting trajectories for individual children . . . . .	351
7.3.2	Piecewise linear models . . . . .	353
	Fitting the models . . . . .	354
	Predicting the mean trajectory . . . . .	357
7.4	Two-stage model formulation . . . . .	358
7.5	Heteroskedasticity . . . . .	360
7.5.1	Heteroskedasticity at level 1 . . . . .	360
7.5.2	Heteroskedasticity at level 2 . . . . .	362
7.6	How does reading improve from kindergarten through third grade? . . . . .	364
7.7	Growth-curve model as a structural equation model . . . . .	364
7.7.1	Estimation using sem . . . . .	366
7.7.2	Estimation using xtmixed . . . . .	371
7.8	Summary and further reading . . . . .	375

7.9	Exercises . . . . .	376
<b>IV</b>	<b>Models with nested and crossed random effects</b>	<b>383</b>
<b>8</b>	<b>Higher-level models with nested random effects</b>	<b>385</b>
8.1	Introduction . . . . .	385
8.2	Do peak-expiratory-flow measurements vary between methods within subjects? . . . . .	386
8.3	Inspecting sources of variability . . . . .	388
8.4	Three-level variance-components models . . . . .	389
8.5	Different types of intraclass correlation . . . . .	392
8.6	Estimation using xtmixed . . . . .	393
8.7	Empirical Bayes prediction . . . . .	394
8.8	Testing variance components . . . . .	395
8.9	Crossed versus nested random effects revisited . . . . .	397
8.10	Does nutrition affect cognitive development of Kenyan children? . .	399
8.11	Describing and plotting three-level data . . . . .	400
8.11.1	Data structure and missing data . . . . .	400
8.11.2	Level-1 variables . . . . .	401
8.11.3	Level-2 variables . . . . .	402
8.11.4	Level-3 variables . . . . .	403
8.11.5	Plotting growth trajectories . . . . .	404
8.12	Three-level random-intercept model . . . . .	405
8.12.1	Model specification: Reduced form . . . . .	405
8.12.2	Model specification: Three-stage formulation . . . . .	405
8.12.3	Estimation using xtmixed . . . . .	406
8.13	Three-level random-coefficient models . . . . .	409
8.13.1	Random coefficient at the child level . . . . .	409
8.13.2	Random coefficient at the child and school levels . . . . .	411
8.14	Residual diagnostics and predictions . . . . .	413
8.15	Summary and further reading . . . . .	418
8.16	Exercises . . . . .	419



<b>9</b>	<b>Crossed random effects</b>	<b>433</b>
9.1	Introduction . . . . .	433
9.2	How does investment depend on expected profit and capital stock?	434
9.3	A two-way error-components model . . . . .	435
9.3.1	Model specification . . . . .	435
9.3.2	Residual variances, covariances, and intraclass correlations	436
	Longitudinal correlations . . . . .	436
	Cross-sectional correlations . . . . .	436
9.3.3	Estimation using xtmixed . . . . .	437
9.3.4	Prediction . . . . .	441
9.4	How much do primary and secondary schools affect attainment at age 16? . . . . .	443
9.5	Data structure . . . . .	444
9.6	Additive crossed random-effects model . . . . .	446
9.6.1	Specification . . . . .	446
9.6.2	Estimation using xtmixed . . . . .	447
9.7	Crossed random-effects model with random interaction . . . . .	448
9.7.1	Model specification . . . . .	448
9.7.2	Intraclass correlations . . . . .	448
9.7.3	Estimation using xtmixed . . . . .	449
9.7.4	Testing variance components . . . . .	451
9.7.5	Some diagnostics . . . . .	453
9.8	❖ A trick requiring fewer random effects . . . . .	456
9.9	Summary and further reading . . . . .	459
9.10	Exercises . . . . .	460
<b>A</b>	<b>Useful Stata commands</b>	<b>471</b>
	<b>References</b>	<b>473</b>
	<b>Author index</b>	<b>485</b>
	<b>Subject index</b>	<b>491</b>