Contents

Preface **xv**

1 Linear regression analysis **1**

1.1 Linear regression assumptions 1
1.2 Linear regression estimation 3
1.2.1 Residuals 7
1.2.2 Outliers 7
1.3 Linear regression R-square 8
1.4 Linear regression standardization 8
1.5 Example: Regression with one covariate 10
1.5.1 Individual residuals and outliers 13
1.5.2 Reporting results 15
1.6 Multiple covariates 17
1.6.1 Linear regression with two continuous covariates . . . 17
Interaction between two continuous covariates 18
1.6.2 Linear regression with one binary and one continuous covariate 20
Interaction between a binary and a continuous covariate 22
1.7 Example: Regression with two covariates 23
1.7.1 Reporting results 25
1.8 Example: Regression with an interaction 27
1.8.1 Reporting results 29
Presenting parameter estimates 30
Presenting results graphically 30
1.9 Special topics 34
1.9.1 Standardized coefficients greater than one 34
1.9.2 Standardized coefficients differing in significance from unstandardized coefficients 35
1.9.3 Two-group regression analysis 36

Example: Two-group regression analysis of an intervention study . . . 37
1.9.4 Bringing covariates into the model . . . 39
Missing data on x . . . 40
Example: Bringing a covariate into the model for the intervention example using a two-group analysis . . . 40
1.9.5 The Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) . . . 44
1.9.6 Heteroscedasticity modeling . . . 45
Example: Heteroscedasticity modeling of LSAY math data . . . 45
1.9.7 Random coefficient regression . . . 48
Example: Random coefficient regression for LSAY math data . . . 50

2 Mediation analysis **57**
2.1 A prototypical mediation model . . . 57
2.2 Mediation modeling techniques . . . 59
2.2.1 Estimation . . . 61
Indirect effect standard errors and confidence intervals 61
2.2.2 Standardization for mediation models . . . 63
2.2.3 Model testing . . . 65
2.3 Example: Sex discrimination . . . 66
2.3.1 Inspecting the data and reporting results . . . 68
2.4 Example: Head circumference . . . 72
2.4.1 Reporting results . . . 75
2.4.2 The saturated model . . . 80
2.5 Multiple mediators . . . 82
2.5.1 Example: Parallel mediators for media influence . . . 83
2.5.2 Example: Sequential mediators of socioeconomic status 87
2.6 Moderated mediation . . . 91
2.6.1 Case 1 (xz): Regression of y on x, m on x, both moderated by z . . . 91
2.6.2 Case 2 (mz): Regression of y on m moderated by z . 93
2.6.3 Case 3 (mx): Regression of y on m moderated by x . 95
2.6.4 Combined moderation case . . . 97
2.6.5 Example: Case 1 moderated mediation in an intervention of aggressive behavior in the classroom . . . 98

Testing significance of effects at specific moderator values . . . 99
Creating a plot with bootstrap confidence intervals for the effects at a range of moderator values . . 102
Combination of significance of effect at specific moderator values and plot of confidence intervals using MODEL INDIRECT . . . 106
2.6.6 Example: Case 2 moderated mediation for work team behavior . . . 108
2.6.7 Example: Case 3 moderated mediation of simulated data . . . 111
2.6.8 Example: Combined moderated mediation for sex discrimination . . . 117

3 Special topics in mediation analysis **121**
3.1 Monte Carlo simulation study of mediation . . . 121
3.1.1 Example: Monte Carlo study of indirect effects . . . 123
3.2 Monte Carlo studies of moderation . . . 130
3.2.1 Example: Moderation of the regression of m on x . . . 130
3.2.2 Example: Moderation of the regression of y on m . . . 136
3.3 Model misspecification . . . 138
3.3.1 Example: Omitted moderator . . . 138
3.3.2 Example: Omitted mediators . . . 141
3.3.3 Example: Confounders . . . 147
3.4 Instrumental variable estimation . . . 153
3.4.1 Example: IV estimation with mediator-outcome confounding . . . 155
Bias and coverage of IV estimation of the indirect effect 156
Comparing IV and ML standard errors and power . . 157
IV estimation dependence on the size of the x, m correlation . . . 158
Comparison of IV and maximum-likelihood estimation when the assumptions behind both approaches are violated . . . 159
3.5 Sensitivity analysis . . . 159
3.5.1 Example: Sensitivity analysis for an experimental study of sex discrimination in the workplace . . . 162
3.5.2 Example: Sensitivity analysis in a Monte Carlo study 165
3.6 Multiple-group mediation analysis . . . 169

3.6.1 Relating multiple-group parameters to interaction parameters . . . 170
3.6.2 Modification indices . . . 171
3.6.3 Example: Two-group analysis of moderated mediation for sex discrimination . . . 172
3.7 Measurement errors and latent variables . . . 176
3.7.1 Measurement error in an independent variable . . . 176
3.7.2 Measurement error in a mediator . . . 178
3.7.3 Example: Monte Carlo simulation study of measurement error in the mediator . . . 179
3.7.4 Known reliability . . . 182
3.7.5 Multiple indicators . . . 182
Reliability of a sum of indicators . . . 183
Structural equation modeling with a factor analysis measurement model . . . 185

4 Causal inference for mediation **187**
4.1 Causal assumptions . . . 188
4.2 Potential outcomes and counterfactuals . . . 189
4.2.1 Example: Hypothetical potential outcome data . . . 190
4.3 Basics of counterfactually-defined effects . . . 191
4.3.1 Example: Hypothetical mediation potential outcomes 193
4.4 Direct and indirect effects . . . 196
4.4.1 Direct effects . . . 197
4.4.2 Indirect effects . . . 198
4.4.3 Total effect decomposition . . . 199
4.4.4 Example: Hypothetical mediation data analysis . . . 199
4.5 Causal effect formulas . . . 200
4.5.1 Example: Effects in the simple mediation case . . . 203
4.5.2 Example: Effects with moderation of Y regressed on M (case 3) . . . 204
4.5.3 Example: Effects in the combined moderation case . . 206
4.5.4 Example: Effects combining case 1 and case 2 moderation . . . 207
4.5.5 Multiple mediators . . . 208
4.6 Summary . . . 209

5 Categorical dependent variable **211**
5.1 Basic concepts for categorical variables . . . 211
5.1.1 Binary variables . . . 214

5.2 Binary dependent variable 216
5.2.1 Example: OLS, logistic, and probit regression of coal miner respiratory problems 217
5.2.2 Modeling with a logistic regression function 220
5.2.3 Modeling with a probit regression function 222
5.2.4 Estimation of the logistic and probit regressions . . . 224
5.2.5 Probability curve formulation versus a latent response variable formulation 224
5.2.6 R^2 and standardization 226
R^2 for a binary outcome 227
Standardization 227
5.2.7 Example: Logistic and probit regression of British coal miner data 228
Logistic regression 228
Probit regression 232
Computation of estimated probabilities 232
Comparing logistic and probit regression coefficients . 234
Comparing the logistic and probit regression models by BIC 234
5.2.8 Logistic and probit regression with one binary and one continuous x 234
5.2.9 Logistic regression and adjusted odds ratios 235
5.2.10 Example: Adjusted odds ratios for alcohol survey data 237
5.2.11 Example: Adjusted odds ratios for educational achievement data 239
5.3 Ordinal dependent variable 240
5.3.1 Probability curve formulation of ordinal dependent variable regression 240
5.3.2 Latent response variable formulation of ordinal dependent variable regression 243
5.3.3 Example: Sample probits for drinking related to age and income 245
5.3.4 Testing the parallel probability curve assumption behind the ordinal regression model 245
5.3.5 Example: Ordinal logistic regression of mental impairment 247
Estimated odds ratio 248
Estimated probabilities 249
Odds ratio with an interaction 250
5.4 Nominal dependent variable 252

5.4.1 Example: Multinomial logistic regression of antisocial behavior 253

6 Count dependent variable 259

6.1 Poisson model 259
6.2 Poisson model with a random intercept 261
6.3 Zero-inflated Poisson model 261
6.4 Negative binomial model 262
6.5 Zero-inflated negative binomial model 262
6.6 Two-part (hurdle) model with zero-truncation 263
6.7 Varying-exposure model 263
6.8 Comparing models 264
6.9 Example: Count regression of marital affairs 264
6.9.1 Poisson, Poisson with a random intercept, and negative binomial models 266
Negative binomial model 268
6.9.2 Zero-inflated Poisson and zero-inflated negative binomial models 269
6.9.3 Two-part (hurdle) modeling 272
6.9.4 Conclusion for marital affairs analyses 276
6.10 Example: Poisson with varying exposure 276

7 Censored dependent variable 279

7.1 Basic concepts for a censored variable 279
7.2 Censored-normal (tobit) regression 280
7.3 Censored-inflated regression 282
7.4 Sample selection (Heckman) regression 283
7.4.1 Example: Simulated sample selection data 285
7.5 Two-part regression 288
7.6 Example: Methods comparison on alcohol data 290
7.6.1 Analysis results for the four models 293
Loglikelihood and BIC comparisons of the four models 294
Comparing the results for the censored-normal (tobit) and censored-inflated models 295
Comparing the results for the sample selection (Heckman) and two-part models 296
Comparing the results for the censored-inflated and two-part models 297

Comparing the fit for estimated probabilities and means for the censored-inflated and two-part models . . . 300
7.7 Switching regressions . . . 302
7.7.1 Example: Monte Carlo simulation of switching regressions . . . 302

8 Mediation non-continuous variables 307
8.1 Binary outcome, continuous mediator . . . 307
8.1.1 A simple hypothetical mediation model . . . 309
8.1.2 Total, indirect, and direct effects in terms of differences in probabilities . . . 310
8.1.3 Causal effect formulas for a continuous M and a binary Y . . . 311
8.1.4 Causal effect formulas applied to a simple mediation model with a binary outcome and a continuous mediator 314
8.1.5 Causal effect formulas defined on the odds ratio scale 315
Odds ratio effects assuming a rare outcome . . . 315
8.1.6 Causal effects with multiple mediators and a binary outcome . . . 316
8.1.7 Example: Intention to use cigarettes . . . 316
Probit regression . . . 318
Logistic regression . . . 321
8.1.8 Example: HPV vaccination trial . . . 323
No intervention-mediator interaction . . . 324
Intervention-mediator interaction . . . 324
Analysis results . . . 327
8.2 Count outcome, continuous mediator . . . 330
8.2.1 Causal effect formulas for a count outcome . . . 330
8.2.2 Example: Aggressive behavior and school removal . . 332
Estimated count probabilities . . . 333
8.3 Two-part outcome, continuous mediator . . . 337
8.3.1 Causal effect formulas for a two-part outcome . . . 337
8.3.2 Example: Two-part mediation analysis of economic stress data . . . 339
Causal effects for two-part modeling . . . 340
Causal effects for regular modeling with log y . . . 343
Causal effects for regular modeling without log y . . . 344
8.4 Binary and ordinal mediator . . . 346
8.4.1 Causal effect formulas for a binary mediator . . . 346

8.4.2 Ordinal mediator 348
8.4.3 Latent response variable mediator 348
8.4.4 Estimation 349
8.4.5 Example: Hypothetical data from the potential outcome example with a binary mediator and a continuous outcome 350
8.4.6 Example: Ordinal mediator for intention to use cigarettes 352
8.4.7 Example: Pearl's artificial 2 x 2 x 2 example 357
8.5 Nominal mediator 361
8.5.1 Causal effect formulas for a nominal mediator 361
8.5.2 Estimation 362
8.5.3 Example: Hypothetical data with a nominal mediator and a binary outcome 362
8.6 Mediator with measurement error 368
8.6.1 Example: A Monte Carlo simulation study for a mediator measured with error 369
8.6.2 Example: An intervention study of aggressive behavior in the classroom and juvenile court record 374

9 Bayesian analysis **381**
9.1 Prior, likelihood, and posterior 381
9.1.1 Posterior distribution for the mean of a normal distribution 383
9.1.2 Types of priors 385
9.1.3 Non-normality of parameter distributions 386
9.2 Markov Chain Monte Carlo (MCMC) 387
9.2.1 Example: Bayesian estimation of a mean with missing data for LSAY math 388
Input for Bayesian analysis 391
9.2.2 Plots 393
Trace plot 393
Autocorrelation plot 395
Posterior distribution plot 397
9.2.3 Convergence checking 399
9.3 Model fit 400
9.4 Bayes versus ML intervals 402
9.5 Example: Mediation model for media influence 404
The Potential Scale Reduction (PSR) convergence criterion 406
Trace plot 406

Autocorrelation plot 406
Inspecting model fit and parameter estimates 408
9.6 Example: Mediation model for firefighter data 413
9.6.1 Non-informative priors 413
9.6.2 Informative priors 413
9.7 Example: Model testing of direct effects 416
9.8 Example: High school dropout and missing data 418
9.8.1 Missing data on the mediator ($n = 2,213$) 418
Bayesian analysis 419
Maximum-likelihood analysis 422
9.8.2 Missing data on the control variables ($n = 2,898$) . . . 423
Assuming normality for all covariates 424
Acknowledging that some control variables are binary 425

10 Missing data 427
10.1 Example: Missing data information 427
10.2 MCAR, MAR, and NMAR 429
10.2.1 MCAR: Missing completely at random 430
10.2.2 MAR: Missing at random 430
10.2.3 NMAR: Not missing at random 431
10.3 MAR for bivariate normal variables (H_1 case) 431
10.3.1 Listwise versus ML 432
10.3.2 Maximum-likelihood estimation in the bivariate case with missing on one variable 434
10.3.3 The EM algorithm 436
10.3.4 Multiple imputation 438
10.3.5 Example: Estimating sample statistics for intervention data 440
10.4 MAR for regression (H_0 case) 443
10.4.1 Missing data and selection on x or y 443
10.4.2 Regression analysis with missing data 445
10.4.3 Technical aspects of ML assuming MAR 448
10.4.4 Example: MAR simulated data analysis 449
10.4.5 Missing data correlates 456
Example: Simulation study with missing data correlate of missing on y 456
10.5 NMAR 464
10.5.1 Example: Simulated NMAR data with missing influenced by the latent outcome 465

10.5.2 Example: Selection modeling versus ML assuming MAR when MAR holds . . . 470
10.6 Example: Comparing missing data methods . . . 472
10.7 Missing data on covariates . . . 476
10.7.1 Example: Simulation study of missing on binary covariates . . . 476
Inputs for generating and analyzing data assuming MAR 476
Inputs for generating and analyzing data under NMAR 480
Simulation results . . . 480

Appendices **489**

A Covariance algebra **491**
A.1 Definition of an expectation . . . 491
A.1.1 Rules for an expectation . . . 492
A.2 Definition of covariance and variance . . . 492
A.2.1 Rules for variance . . . 493
A.3 Functions of random variables . . . 493
A.4 Example: Covariance algebra rules applied to linear regression 493
A.5 Example: Derivation of the slope attenuation . . . 494
A.6 Example: Variable measured with error . . . 495

References **497**