

Contents

Online Resources	xi
About the Authors	xiii
Preface	xv
Acknowledgements	xvi
How to Use This Book	xvi
Common Error Messages in R to Know About	xviii
Regression Approach to ANOVA	xxi
Authors' Contact	xxii
1 Introduction to R	1
1.1 What Is R and Why Should You Use It?	2
1.2 What Is RStudio?	6
1.3 How to Install R and RStudio?	7
1.3.1 R for Windows	7
1.3.2 R for Mac	7
1.3.3 R for Linux	8
1.3.4 RStudio for Windows, Mac OS, and Linux	8
1.4 Getting to Know RStudio	9
1.4.1 Script pane	10
1.4.2 Console pane	11
1.4.3 Environment pane	12
1.4.4 Plot pane	13
1.5 Internet Resources for R	13
1.6 Contributed R Packages	16
1.7 Updating R, RStudio, and Contributed Packages	18
1.8 Conclusion	19
2 Importing and Working With Data in R	23
2.1 How Is a Dataset Represented in R?	24
2.2 Importing Data Into R	25
2.3 Entering Data in R	32
2.4 How Do We Work With a Dataset in R?	36
2.5 Data Types	39
2.6 Conclusion	42

3	How Does R Work?	45
3.1	The Way R Works	46
3.2	What Is a Function?	47
3.3	What Is an Object?	50
3.3.1	Vectors	53
3.3.2	Data frames	55
	Extracting rows and columns from data frames	57
	Adding observations and variables	59
	Renaming variables	61
3.3.3	Matrices	62
3.3.4	Lists	63
3.4	Conclusion	65
4	Data Management	69
4.1	Data Management of Variables	71
4.1.1	Generating new variables	71
4.1.2	Recoding variables	74
4.1.3	Replacing values	76
4.1.4	Renaming variables	79
4.1.5	Exploring missing values	80
4.1.6	Generating dummy variables	84
4.1.7	Changing the data types of variables	86
4.1.8	Labelling variables	88
4.1.9	Tidying up categorical variables	88
4.2	Data Management of Datasets	90
4.2.1	Selecting and excluding variables	90
4.2.2	Selecting observations	92
4.2.3	Merging datasets by variables	94
4.2.4	Merging datasets by observations	96
4.2.5	Sorting a dataset	97
4.2.6	Reshaping a dataset	98
4.2.7	Ordering variables	100
4.2.8	Drawing random samples from a dataset	101
4.2.9	<i>Piping</i> (chaining)	102
4.3	Conclusion	104
5	Data Visualization With <code>ggplot2</code>	109
5.1	The Role of Visualization in Data Analysis	110
5.2	Understanding <code>ggplot2</code>	112
5.2.1	Structure of a layer	113
	Data	114
	Mapping	114
	Geometric objects	115
	Statistical transformations	118
	Positions	122

5.2.2	Additional components affecting all layers	124
	Scales	124
	Coordinate systems	125
	Labels and annotations	127
	Themes	128
	Facets	131
5.3	Examples in R	132
5.3.1	Univariate graphs	133
	One continuous variable	133
	One categorical variable	138
5.3.2	Bivariate graphs	140
	One categorical x and one continuous y	140
	One continuous x and one continuous y	142
5.3.3	Multivariate	143
	Continuous X_1 and Y and one categorical X_2	143
5.4	Conclusion	144
6	Descriptive Statistics	149
6.1	Describing a Single Variable	151
6.1.1	Central tendency of a distribution	153
6.1.2	Measures of spread	156
6.1.3	Skew and kurtosis	160
6.1.4	Discrete distributions	162
6.1.5	Quick descriptive analysis	165
	summarytools package	166
	skimr package	168
6.2	Describing Relationships Between Variables	170
6.2.1	Correlation coefficient	171
6.2.2	Cross-tabulation	174
6.3	Summarizing Variables Across Groups	176
6.4	Conclusion	179
7	Simple (Bivariate) Regression	183
7.1	What Is Regression Analysis?	184
7.2	Simple Linear Regression Analysis	185
7.2.1	Ordinary least squares	188
7.2.2	Goodness of fit	190
	Residual standard deviation	190
	Coefficient of determination (R^2)	191
7.2.3	Hypothesis test for slope coefficient	193
	The p -value approach	193
	The confidence interval approach	195
7.2.4	Prediction in linear regression	195
7.3	Example in R	196
7.4	Conclusion	200

8	Multiple Linear Regression	203
8.1	Multiple Regression Analysis	204
8.1.1	Estimation	205
8.1.2	Goodness of fit and the F test	206
8.1.3	Adjusted R^2	207
8.1.4	Partial slope coefficient	208
8.1.5	Prediction in multiple regression	209
8.1.6	Standardization and relative importance	210
8.1.7	Regression assumptions and diagnostics	211
8.2	Example in R	213
8.3	Conclusion	225
9	Dummy-Variable Regression	229
9.1	Why Dummy-Variable Regression?	230
9.1.1	Creating dummy variables	231
9.1.2	The logic behind dummy-variable regression	233
9.2	Regression With One Dummy Variable	233
9.2.1	Example in R	234
9.3	Regression With One Dummy Variable and a Covariate	236
9.3.1	Example in R	237
9.4	Regression With More Than One Dummy Variable	239
9.4.1	Example in R	241
9.4.2	Comparing the included groups	242
	Changing the reference group	243
	Linear combination	244
9.4.3	Pairwise multiple comparison adjustment	247
9.5	Regression With More Than One Dummy Variable and a Covariate	249
9.5.1	Example in R	250
9.6	Regression With Two Separate Sets of Dummy Variables	251
9.6.1	Example in R	253
9.7	Conclusion	256
10	Moderation/Interaction Analysis Using Regression	259
10.1	Interaction/Moderation Effects	260
10.2	Product-Term Approach	262
10.3	Interaction Between a Continuous Predictor and a Dummy Moderator	264
10.3.1	Example in R	266
10.4	Interaction Between a Continuous Predictor and a Continuous Moderator	268
10.4.1	Example in R	269
10.5	Interaction Between a Dummy Predictor and a Dummy Moderator	273
10.5.1	Example in R	273
10.6	Interaction Between a Continuous Predictor and a Polytomous Moderator	276
10.6.1	Example in R	277

10.7 Additional Considerations	282
10.7.1 Significant versus non-significant interactions	282
10.7.2 Centring and standardization	282
10.8 Conclusion	283
11 Logistic Regression	287
11.1 Simple Logistic Regression in R	292
11.1.1 Interpretation of coefficients in logistic regression	295
11.1.2 Goodness-of-fit and model selection	299
11.2 Multiple Logistic Regression	301
11.3 Logistic Regression for Classification	311
11.4 Conclusion	317
12 Multilevel and Longitudinal Analysis	321
12.1 Representation of Nested Data Structures	324
12.1.1 Converting between wide and long format	326
12.2 Complete, Partial, and No Pooling	329
12.3 Significance Testing for Linear Mixed Models	336
12.3.1 Mixing fixed and random effects	340
12.4 Model Comparison for Longitudinal Mixed Models	343
12.5 Conclusion	348
13 Factor Analysis	351
13.1 What Is Factor Analysis?	352
13.1.1 What is factor analysis used for?	354
13.2 The Factor Analysis Process	355
13.2.1 Determining the number of factors	355
Eigenvalue rule	356
Scree test	356
Parallel analysis	356
Theoretical sensitivity	356
13.2.2 Extracting the factors	357
13.2.3 Rotating the factors	359
13.2.4 Refining and interpreting the factors	362
13.3 Composite Scores and Reliability Tests	363
13.4 Example in R	364
13.4.1 Determining the number of factors	365
13.4.2 Extracting with rotation	367
13.5 Conclusion	371
14 Structural Equation Modelling	375
14.1 What Is Structural Equation Modelling?	376
14.1.1 Types of structural equation modelling	377

14.2	Confirmatory Factor Analysis	379
14.2.1	Model specification	379
14.2.2	Model identification	381
14.2.3	Parameter estimation	383
14.2.4	Model assessment	383
	Interpreting parameter estimates	383
	Model fit indices	387
	Chi-squared test	387
	Standardized root mean square residual	389
	Root mean squared error of approximation	390
	Comparative fit index	390
	Tucker–Lewis index	391
14.2.5	Model modification	392
14.3	Latent Path Analysis	394
14.3.1	Specification of the LPA model	395
14.3.2	Measurement part	395
14.3.3	Structural part	399
14.4	Conclusion	402
15	Bayesian Statistics	407
15.1	Bayesian Data Analysis	410
15.2	Bayesian Data Analysis in R	412
15.3	Example Analysis in R	414
15.3.1	Model diagnostics	415
15.3.2	Bayesian estimation of regression coefficients	417
15.3.3	Bayesian model selection	422
15.3.4	Model checking	427
15.3.5	Choice of prior distribution	429
15.4	Conclusion	432
	Bibliography	437
	Index	443