

Contents

<i>Acknowledgments</i>	x
<i>0. Preface: Approach and How to Use This Book</i>	xii
0.1. <i>Strategy of the Book</i>	xii
0.2. <i>Why R?</i>	xiii
0.3. <i>Why the Tidyverse?</i>	xiv
0.4. <i>R Packages Required for This Book</i>	xv
0.5. <i>What This Book Is Not</i>	xv
0.6. <i>How to Use This Book</i>	xv
0.7. <i>Information for Teachers</i>	xvi
1 Introduction to R	1
1.1. <i>Introduction</i>	1
1.2. <i>Baby Steps: Simple Math with R</i>	2
1.3. <i>Your First R Script</i>	4
1.4. <i>Assigning Variables</i>	5
1.5. <i>Numeric Vectors</i>	7
1.6. <i>Indexing</i>	9
1.7. <i>Logical Vectors</i>	10
1.8. <i>Character Vectors</i>	11
1.9. <i>Factor Vectors</i>	12
1.10. <i>Data Frames</i>	13
1.11. <i>Loading in Files</i>	16
1.12. <i>Plotting</i>	19
1.13. <i>Installing, Loading, and Citing Packages</i>	20
1.14. <i>Seeking Help</i>	21
1.15. <i>A Note on Keyboard Shortcuts</i>	22
1.16. <i>Your R Journey: The Road Ahead</i>	23
2 The Tidyverse and Reproducible R Workflows	27
2.1. <i>Introduction</i>	27
2.2. <i>tibble and readr</i>	28
2.3. <i>dplyr</i>	30
2.4. <i>ggplot2</i>	34
2.5. <i>Piping with magrittr</i>	36
2.6. <i>A More Extensive Example: Iconicity and the Senses</i>	37

2.7.	<i>R Markdown</i>	44
2.8.	<i>Folder Structure for Analysis Projects</i>	45
2.9.	<i>Readme Files and More Markdown</i>	46
2.10.	<i>Open and Reproducible Research</i>	47
3	Descriptive Statistics, Models, and Distributions	53
3.1.	<i>Models</i>	53
3.2.	<i>Distributions</i>	53
3.3.	<i>The Normal Distribution</i>	54
3.4.	<i>Thinking of the Mean as a Model</i>	57
3.5.	<i>Other Summary Statistics: Median and Range</i>	58
3.6.	<i>Boxplots and the Interquartile Range</i>	59
3.7.	<i>Summary Statistics in R</i>	60
3.8.	<i>Exploring the Emotional Valence Ratings</i>	64
3.9.	<i>Chapter Conclusions</i>	67
4	Introduction to the Linear Model: Simple Linear Regression	69
4.1.	<i>Word Frequency Effects</i>	69
4.2.	<i>Intercepts and Slopes</i>	71
4.3.	<i>Fitted Values and Residuals</i>	72
4.4.	<i>Assumptions: Normality and Constant Variance</i>	74
4.5.	<i>Measuring Model Fit with R^2</i>	74
4.6.	<i>A Simple Linear Model in R</i>	77
4.7.	<i>Linear Models with Tidyverse Functions</i>	82
4.8.	<i>Model Formula Notation: Intercept Placeholders</i>	83
4.9.	<i>Chapter Conclusions</i>	84
5	Correlation, Linear, and Nonlinear Transformations	86
5.1.	<i>Centering</i>	86
5.2.	<i>Standardizing</i>	87
5.3.	<i>Correlation</i>	89
5.4.	<i>Using Logarithms to Describe Magnitudes</i>	90
5.5.	<i>Example: Response Durations and Word Frequency</i>	94
5.6.	<i>Centering and Standardization in R</i>	98
5.7.	<i>Terminological Note on the Term ‘Normalizing’</i>	101
5.8.	<i>Chapter Conclusions</i>	101
6	Multiple Regression	103
6.1.	<i>Regression with More Than One Predictor</i>	103
6.2.	<i>Multiple Regression with Standardized Coefficients</i>	105
6.3.	<i>Assessing Assumptions</i>	109
6.4.	<i>Collinearity</i>	112
6.5.	<i>Adjusted R^2</i>	115
6.6.	<i>Chapter Conclusions</i>	116

7	Categorical Predictors	117
7.1.	<i>Introduction</i>	117
7.2.	<i>Modeling the Emotional Valence of Taste and Smell Words</i>	117
7.3.	<i>Processing the Taste and Smell Data</i>	119
7.4.	<i>Treatment Coding in R</i>	122
7.5.	<i>Doing Dummy Coding 'By Hand'</i>	123
7.6.	<i>Changing the Reference Level</i>	124
7.7.	<i>Sum-coding in R</i>	125
7.8.	<i>Categorical Predictors with More Than Two Levels</i>	127
7.9.	<i>Assumptions Again</i>	129
7.10.	<i>Other Coding Schemes</i>	130
7.11.	<i>Chapter Conclusions</i>	131
8	Interactions and Nonlinear Effects	133
8.1.	<i>Introduction</i>	133
8.2.	<i>Categorical * Continuous Interactions</i>	134
8.3.	<i>Categorical * Categorical Interactions</i>	139
8.4.	<i>Continuous * Continuous Interactions</i>	146
8.5.	<i>Nonlinear Effects</i>	150
8.6.	<i>Higher-Order Interactions</i>	155
8.7.	<i>Chapter Conclusions</i>	156
9	Inferential Statistics 1: Significance Testing	157
9.1.	<i>Introduction</i>	157
9.2.	<i>Effect Size: Cohen's d</i>	159
9.3.	<i>Cohen's d in R</i>	161
9.4.	<i>Standard Errors and Confidence Intervals</i>	162
9.5.	<i>Null Hypotheses</i>	165
9.6.	<i>Using t to Measure the Incompatibility with the Null Hypothesis</i>	166
9.7.	<i>Using the t-Distribution to Compute p-Values</i>	167
9.8.	<i>Chapter Conclusions</i>	169
10	Inferential Statistics 2: Issues in Significance Testing	171
10.1.	<i>Common Misinterpretations of p-Values</i>	171
10.2.	<i>Statistical Power and Type I, II, M, and S Errors</i>	171
10.3.	<i>Multiple Testing</i>	175
10.4.	<i>Stopping rules</i>	177
10.5.	<i>Chapter Conclusions</i>	178
11	Inferential Statistics 3: Significance Testing in a Regression Context	180
11.1.	<i>Introduction</i>	180
11.2.	<i>Standard Errors and Confidence Intervals for Regression Coefficients</i>	180
11.3.	<i>Significance Tests with Multilevel Categorical Predictors</i>	184

11.4.	<i>Another Example: The Absolute Valence of Taste and Smell Words</i>	188
11.5.	<i>Communicating Uncertainty for Categorical Predictors</i>	190
11.6.	<i>Communicating Uncertainty for Continuous Predictors</i>	194
11.7.	<i>Chapter Conclusions</i>	197
12	Generalized Linear Models 1: Logistic Regression	198
12.1.	<i>Motivating Generalized Linear Models</i>	198
12.2.	<i>Theoretical Background: Data-Generating Processes</i>	198
12.3.	<i>The Log Odds Function and Interpreting Logits</i>	202
12.4.	<i>Speech Errors and Blood Alcohol Concentration</i>	204
12.5.	<i>Predicting the Dative Alternation</i>	207
12.6.	<i>Analyzing Gesture Perception</i>	210
12.7.	<i>Chapter Conclusions</i>	216
13	Generalized Linear Models 2: Poisson Regression	218
13.1.	<i>Motivating Poisson Regression</i>	218
13.2.	<i>The Poisson Distribution</i>	218
13.3.	<i>Analyzing Linguistic Diversity Using Poisson Regression</i>	220
13.4.	<i>Adding Exposure Variables</i>	225
13.5.	<i>Negative Binomial Regression for Overdispersed Count Data</i>	227
13.6.	<i>Overview and Summary of the Generalized Linear Model Framework</i>	229
13.7.	<i>Chapter Conclusions</i>	230
14	Mixed Models 1: Conceptual Introduction	232
14.1.	<i>Introduction</i>	232
14.2.	<i>The Independence Assumption</i>	232
14.3.	<i>Dealing with Non-independence via Experimental Design and Averaging</i>	233
14.4.	<i>Mixed Models: Varying Intercepts and Varying Slopes</i>	234
14.5.	<i>More on Varying Intercepts and Varying Slopes</i>	237
14.6.	<i>Interpreting Random Effects and Random Effect Correlations</i>	238
14.7.	<i>Specifying Mixed Effects Models: <code>lme4</code> syntax</i>	240
14.8.	<i>Reasoning About Your Mixed Model: The Importance of Varying Slopes</i>	241
14.9.	<i>Chapter Conclusions</i>	244
15	Mixed Models 2: Extended Example, Significance Testing, Convergence Issues	245
15.1.	<i>Introduction</i>	245
15.2.	<i>Simulating Vowel Durations for a Mixed Model Analysis</i>	245

15.3. <i>Analyzing the Simulated Vowel Durations with Mixed Models</i>	253
15.4. <i>Extracting Information out of lme4 Objects</i>	255
15.5. <i>Messing up the Model</i>	257
15.6. <i>Likelihood Ratio Tests</i>	260
15.7. <i>Remaining Issues</i>	264
15.8. <i>Mixed Logistic Regression: Ugly Selfies</i>	267
15.9. <i>Shrinkage and Individual Differences</i>	271
15.10. <i>Chapter Conclusions</i>	272
16 Outlook and Strategies for Model Building	274
16.1. <i>What You Have Learned So Far</i>	274
16.2. <i>Model Choice</i>	275
16.3. <i>The Cookbook Approach</i>	275
16.4. <i>Stepwise Regression</i>	276
16.5. <i>A Plea for Subjective and Theory-Driven Statistical Modeling</i>	277
16.6. <i>Reproducible Research</i>	279
16.7. <i>Closing Words</i>	280
<i>References</i>	281
<i>Appendix A. Correspondences Between Significance Tests and Linear Models</i>	290
A1. <i>t-Tests</i>	290
A2. <i>Tests for Categorical Data</i>	295
A3. <i>Other Tests</i>	299
<i>Appendix B. Reading Recommendations</i>	301
B1. <i>Book Recommendations</i>	301
B2. <i>Article Recommendations</i>	302
B3. <i>Staying Up-to-Date</i>	303
<i>Index</i>	304
<i>Index of R Functions</i>	308