

CONTENTS

List of Figures	xiii
List of Tables	xxi
Foreword	xxiii
Preface	xxv
Acknowledgments	xxxix

1	An Introduction to R	1
1.1	Getting Started	2
1.1.1	Windows OS	2
1.1.2	Mac OS	2
1.1.3	Add-On Packages	2
1.2	Arithmetic: R as a Calculator	4
1.3	Computations in R: Functions	4
1.4	Connecting Computations	7
1.4.1	Naming Conventions	8
1.5	Data Structures: Vectors	9
1.5.1	Creating Vectors in R	9
1.5.2	Computation with Vectors	11
1.5.3	Character and Logical Vectors	12

1.6	Getting Help	13
1.7	Alternative Ways to Run R	14
1.8	Extension: Matrices and Matrix Operations	14
1.8.1	Computation with Matrices	15
1.9	Further Reading	18
	Problems	19
2	Data Representation and Preparation	21
2.1	Tabular Data	23
2.1.1	External Formats for Storing Tabular Data	23
2.2	Data Entry	24
2.2.1	Data Codebooks	25
2.3	Reading Delimited Data into R	25
2.3.1	Identifying the Location of a File	26
2.3.2	Examining the Data in a Text Editor	28
2.3.3	Reading Delimited Separated Data: An Example	28
2.4	Data Structure: Data Frames	29
2.4.1	Examining the Data Read into R	29
2.5	Recording Syntax using Script Files	33
2.5.1	Documentation File	34
2.6	Simple Graphing in R	34
2.6.1	Saving Graphics to Insert into a Word-Processing File	35
2.7	Extension: Logical Expressions and Graphs for Categorical Variables	37
2.7.1	Logical Operators	38
2.7.2	Measurement Level and Analysis	40
2.7.3	Categorical Data	42
2.7.4	Plotting Categorical Data	44
2.8	Further Reading	45
	Problems	46
3	Data Exploration: One Variable	49
3.1	Reading In the Data	50
3.2	Nonparametric Density Estimation	52
3.2.1	Graphically Summarizing the Distribution	52
3.2.2	Histograms	52
3.2.3	Kernel Density Estimators	53
3.2.4	Controlling the Density Estimation	53

	3.2.5	Plotting the Estimated Density	55
3.3		Summarizing the Findings	58
	3.3.1	Creating a Plot for Publication	59
	3.3.2	Writing Up the Results for Publication	61
3.4		Extension: Variability Bands for Kernel Densities	62
3.5		Further Reading	62
		Problems	63
4		Exploration of Multivariate Data: Comparing Two Groups	65
4.1		Graphically Summarizing the Marginal Distribution	66
4.2		Graphically Summarizing Conditional Distributions	66
	4.2.1	Indexing: Accessing Individuals or Subsets	68
	4.2.2	Indexing Using a Logical Expression	69
	4.2.3	Density Plots of the Conditional Distributions	70
	4.2.4	Side-by-Side Box-and-Whiskers Plots	70
4.3		Numerical Summaries of Data: Estimates of the Population Parameters	72
	4.3.1	Measuring Central Tendency	73
	4.3.2	Measuring Variation	74
	4.3.3	Measuring Skewness	76
	4.3.4	Kurtosis	78
4.4		Summarizing the Findings	80
	4.4.1	Creating a Plot for Publication	80
	4.4.2	Using Color	81
	4.4.3	Selecting a Color Palette	85
4.5		Extension: Robust Estimation	87
	4.5.1	Robust Estimate of Location: The Trimmed Mean	87
	4.5.2	Robust Estimate of Variation: The Winsorized Variance	89
4.6		Further Reading	91
		Problems	91
5		Exploration of Multivariate Data: Comparing Many Groups	93
5.1		Graphing Many Conditional Distributions	94
	5.1.1	Panel Plots	96
	5.1.2	Side-by-Side Box-and-Whiskers Plots	97
5.2		Numerically Summarizing the Data	100
5.3		Summarizing the Findings	101
	5.3.1	Writing Up the Results for Publication	102

5.3.2	Enhancing a Plot with a Line	102
5.4	Examining Distributions Conditional on Multiple Variables	103
5.5	Extension: Conditioning on Continuous Variables	107
5.5.1	Scatterplots of the Conditional Distributions	110
5.6	Further Reading	112
	Problems	113
6	Randomization and Permutation Tests	115
6.1	Randomized Experimental Research	118
6.2	Introduction to the Randomization Test	119
6.3	Randomization Tests with Large Samples: Monte Carlo Simulation	122
6.3.1	Rerandomization of the Data	124
6.3.2	Repeating the Randomization Process	125
6.3.3	Generalizing Processes: Functions	126
6.3.4	Repeated Operations on Matrix Rows or Columns	127
6.3.5	Examining the Monte Carlo Distribution and Obtaining the p -Value	127
6.4	Validity of the Inferences and Conclusions Drawn from a Randomization Test	130
6.4.1	Exchangeability	130
6.4.2	Nonexperimental Research: Permutation Tests	131
6.4.3	Nonexperimental, Nongeneralizable Research	131
6.5	Generalization from the Randomization Results	132
6.6	Summarizing the Results for Publication	133
6.7	Extension: Tests of the Variance	133
6.8	Further Reading	134
	Problems	135
7	Bootstrap Tests	137
7.1	Educational Achievement of Latino Immigrants	138
7.2	Probability Models: An Interlude	140
7.3	Theoretical Probability Models in R	141
7.4	Parametric Bootstrap Tests	143
7.4.1	Choosing a Probability Model	144
7.4.2	Standardizing the Distribution of Achievement Scores	144
7.5	The Parametric Bootstrap	146

7.5.1	The Parametric Bootstrap: Approximating the Distribution of the Mean Difference	146
7.6	Implementing the Parametric Bootstrap in R	148
7.6.1	Writing a Function to Randomly Generate Data for the boot() Function	148
7.6.2	Writing a Function to Compute a Test Statistic Using the Randomly Generated Data	150
7.6.3	The Bootstrap Distribution of the Mean Difference	151
7.7	Summarizing the Results of the Parametric Bootstrap Test	154
7.8	Nonparametric Bootstrap Tests	154
7.8.1	Using the Nonparametric Bootstrap to Approximate the Distribution of the Mean Difference	157
7.8.2	Implementing the Nonparametric Bootstrap in R	158
7.9	Summarizing the Results for the Nonparametric Bootstrap Test	160
7.10	Bootstrapping Using a Pivot Statistic	161
7.10.1	Student's t -Statistic	161
7.11	Independence Assumption for the Bootstrap Methods	164
7.12	Extension: Testing Functions	166
7.12.1	Ordering a Data Frame	166
7.13	Further Reading	168
	Problems	168
8	Philosophical Considerations	171
8.1	The Randomization Test vs. the Bootstrap Test	172
8.2	Philosophical Frameworks of Classical Inference	173
8.2.1	Fisher's Significance Testing	174
8.2.2	Neyman–Pearson Hypothesis Testing	175
8.2.3	p -Values	176
9	Bootstrap Intervals and Effect Sizes	179
9.1	Educational Achievement Among Latino Immigrants: Example Revisited	180
9.2	Plausible Models to Reproduce the Observed Result	180
9.2.1	Computing the Likelihood of Reproducing the Observed Result	181
9.3	Bootstrapping Using an Alternative Model	185
9.3.1	Using R to Bootstrap under the Alternative Model	187

9.3.2	Using the Bootstrap Distribution to Compute the Interval Limits	190
9.3.3	Historical Interlude: Student's Approximation for the Interval Estimate	190
9.3.4	Studentized Bootstrap Interval	191
9.4	Interpretation of the Interval Estimate	191
9.5	Adjusted Bootstrap Intervals	192
9.6	Standardized Effect Size: Quantifying the Group Differences in a Common Metric	192
9.6.1	Effect Size as Distance—Cohen's δ	193
9.6.2	Robust Distance Measure of Effect	195
9.7	Summarizing the Results	197
9.8	Extension: Bootstrapping the Confidence Envelope for a Q-Q Plot	197
9.9	Confidence Envelopes	198
9.10	Further Reading	202
	Problems	204
10	Dependent Samples	205
10.1	Matching: Reducing the Likelihood of Nonequivalent Groups	206
10.2	Mathematics Achievement Study Design	206
10.2.1	Exploratory Analysis	209
10.3	Randomization/Permutation Test for Dependent Samples	211
10.3.1	Reshaping the Data	212
10.3.2	Randomization Test Using the Reshaped Data	214
10.4	Effect Size	216
10.5	Summarizing the Results of a Dependent Samples Test for Publication	217
10.6	To Match or Not to Match . . . That is the Question	218
10.7	Extension: Block Bootstrap	220
10.8	Further Reading	223
	Problems	224
11	Planned Contrasts	227
11.1	Planned Comparisons	228
11.2	Examination of Weight Loss Conditioned on Diet	228
11.2.1	Exploration of Research Question 1	229
11.2.2	Exploration of Research Question 2	230

11.2.3	Exploration of Research Question 3	231
11.3	From Research Questions to Hypotheses	232
11.4	Statistical Contrasts	233
11.4.1	Complex Contrasts	236
11.5	Computing the Estimated Contrasts Using the Observed Data	237
11.6	Testing Contrasts: Randomization Test	239
11.7	Strength of Association: A Measure of Effect	240
11.7.1	Total Sum of Squares	241
11.8	Contrast Sum of Squares	243
11.9	Eta-Squared for Contrasts	243
11.10	Bootstrap Interval for Eta-Squared	244
11.11	Summarizing the Results of a Planned Contrast Test Analysis	245
11.12	Extension: Orthogonal Contrasts	245
11.13	Further Reading	251
	Problems	251
12	Unplanned Contrasts	253
12.1	Unplanned Comparisons	254
12.2	Examination of Weight Loss Conditioned on Diet	254
12.3	Omnibus Test	257
12.3.1	Statistical Models	257
12.3.2	Postulating a Statistical Model to Fit the Data	258
12.3.3	Fitting a Statistical Model to the Data	260
12.3.4	Partitioning Variation in the Observed Scores	262
12.3.5	Randomization Test for the Omnibus Hypothesis	268
12.4	Group Comparisons After the Omnibus Test	269
12.5	Ensemble-Adjusted p -values	270
12.5.1	False Discovery Rate	272
12.6	Strengths and Limitations of the Four Approaches	273
12.6.1	Planned Comparisons	273
12.6.2	Omnibus Test Followed by Unadjusted Group Comparisons	274
12.6.3	Omnibus Test Followed by Adjusted Group Comparisons	274
12.6.4	Adjusted Group Comparisons without the Omnibus Test	275
12.6.5	Final Thoughts	276
12.7	Summarizing the Results of Unplanned Contrast Tests for Publication	276

12.8	Extension: Plots of the Unplanned Contrasts	276
12.8.1	Simultaneous Intervals	280
12.9	Further Reading	282
	Problems	283

References	285
-------------------	------------