

Modern Statistics
for the Social and
Behavioral Sciences
A Practical Introduction

Rand Wilcox

University of Southern California
Los Angeles, USA



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group an Informa business
A CHAPMAN & HALL BOOK

Contents

Preface	xix
1 INTRODUCTION	1
1.1 Samples versus Populations	2
1.2 Software	3
1.3 R Basics	4
1.3.1 Entering Data	4
1.3.2 R Functions and Packages	11
1.3.3 Data Sets	13
1.3.4 Arithmetic Operations	14
2 NUMERICAL AND GRAPHICAL SUMMARIES OF DATA	17
2.1 Basic Summation Notation	18
2.2 Measures of Location	19
2.2.1 The Sample Mean	19
2.2.2 R Function Mean	21
2.2.3 The Sample Median	22
2.2.4 R Function for the Median	24
2.2.5 A Criticism of the Median: It Might Trim Too Many Values	24
2.2.6 R Function for the Trimmed Mean	26
2.2.7 A Winsorized Mean	26
2.2.8 R Function winmean	27
2.2.9 What Is a Measure of Location?	28
2.3 Measures of Variation or Scale	28
2.3.1 Sample Variance and Standard Deviation	29
2.3.2 R Functions for the Variance and Standard Deviation	30
2.3.3 The Interquartile Range	30
2.3.4 R Function idealf	32
2.3.5 Winsorized Variance	32
2.3.6 R Function winvar	32
2.3.7 Median Absolute Deviation	32
2.3.8 R Function mad	33
2.3.9 Average Absolute Distance from the Median	33
2.3.10 Other Robust Measures of Variation	34

2.3.11	R Functions bivar, pbvar, tauvar, and tbs	35
2.4	Detecting Outliers	36
2.4.1	A Method Based on the Mean and Variance	37
2.4.2	A Better Outlier Detection Rule: The MAD-Median Rule	38
2.4.3	R Function out	39
2.4.4	The Boxplot	39
2.4.5	R Function boxplot	40
2.4.6	Modifications of the Boxplot Rule for Detecting Outliers	41
2.4.7	R Function outbox	42
2.4.8	Other Measures of Location	43
2.4.9	R Functions mom and onestep	45
2.5	Histograms	46
2.5.1	R Functions hist and splot	54
2.6	Kernel Density Estimators	54
2.6.1	R Functions kdplot and akerd	55
2.7	Stem-and-Leaf Displays	57
2.7.1	R Function stem	58
2.8	Skewness	58
2.8.1	Transforming Data	59
2.9	Choosing a Measure of Location	61
2.10	Covariance and Pearson's Correlation	62
2.11	Exercises	63
3	PROBABILITY AND RELATED CONCEPTS	67
3.1	Basic Probability	68
3.2	Expected Values	69
3.3	Conditional Probability and Independence	71
3.4	Population Variance	77
3.5	The Binomial Probability Function	79
3.6	Continuous Variables and the Normal Curve	84
3.6.1	Computing Probabilities Associated with Normal Distributions	87
3.6.2	R Function pnorm	89
3.7	Understanding the Effects of Non-Normality	94
3.7.1	Skewness	99
3.8	Pearson's Correlation and the Population Covariance	101
3.8.1	Computing the Population Covariance and Pearson's Correlation	102
3.9	Some Rules about Expected Values	104
3.10	Chi-Squared Distributions	105
3.11	Exercises	106

4	SAMPLING DISTRIBUTIONS AND CONFIDENCE INTERVALS	111
4.1	Random Sampling	112
4.2	Sampling Distributions	112
4.2.1	Sampling Distribution of the Sample Mean	114
4.2.2	Computing Probabilities Associated with the Sample Mean	117
4.3	A Confidence Interval for the Population Mean	121
4.3.1	Known Variance	121
4.3.2	Confidence Intervals When σ Is Not Known	124
4.3.3	R Functions pt and qt	126
4.3.4	Confidence Interval for the Population Mean Using Student's T	127
4.3.5	R Function t.test	127
4.4	Judging Location Estimators Based on Their Sampling Distribution	129
4.4.1	Trimming and Accuracy: Another Perspective	134
4.5	An Approach to Non-Normality: The Central Limit Theorem	135
4.6	Student's T and Non-Normality	138
4.7	Confidence Intervals for the Trimmed Mean	147
4.7.1	Estimating the Standard Error of a Trimmed Mean	147
4.7.2	R Function trimse	153
4.8	A Confidence Interval for the Population Trimmed Mean	153
4.8.1	R Function trimci	157
4.9	Transforming Data	159
4.10	Confidence Interval for the Population Median	159
4.10.1	R Function sint	161
4.10.2	Estimating the Standard Error of the Sample Median	162
4.10.3	R Function msmedse	162
4.10.4	More Concerns about Tied Values	162
4.11	A Remark About MOM and M-Estimators	164
4.12	Confidence Intervals for the Probability of Success	164
4.12.1	R Functions binomci and acbinomci	167
4.13	Exercises	168
5	HYPOTHESIS TESTING	173
5.1	The Basics of Hypothesis Testing	174
5.1.1	P-Value or Significance Level	180
5.1.2	R Function t.test	182
5.1.3	Criticisms of Two-Sided Hypothesis Testing and P-Values	182
5.1.4	Summary and Generalization	184
5.2	Power and Type II Errors	185
5.2.1	Understanding How n , α , and σ Are Related to Power	190

5.3	Testing Hypotheses about the Mean When σ Is Not Known	193
5.4	Controlling Power and Determining n	195
5.4.1	Choosing n Prior to Collecting Data	195
5.4.2	R Function <code>power.t.test</code>	196
5.4.3	Stein's Method: Judging the Sample Size When Data Are Available	196
5.4.4	R Functions <code>stein1</code> and <code>stein2</code>	198
5.5	Practical Problems with Student's T Test	199
5.6	Hypothesis Testing Based on a Trimmed Mean	205
5.6.1	R Function <code>trimci</code>	206
5.6.2	R Functions <code>stein1.tr</code> and <code>stein2.tr</code>	208
5.7	Testing Hypotheses about the Population Median	208
5.7.1	R Function <code>sintv2</code>	208
5.8	Making Decisions about Which Measure of Location to Use	209
5.9	Exercises	209
6	REGRESSION AND CORRELATION	213
6.1	The Least Squares Principle	214
6.2	Confidence Intervals and Hypothesis Testing	217
6.2.1	Classic Inferential Techniques	221
6.2.2	Multiple Regression	223
6.2.3	R Functions <code>ols</code> , <code>lm</code> , and <code>olsplot</code>	225
6.3	Standardized Regression	228
6.4	Practical Concerns about Least Squares Regression and How They Might Be Addressed	230
6.4.1	The Effect of Outliers on Least Squares Regression	231
6.4.2	Beware of Bad Leverage Points	234
6.4.3	Beware of Discarding Outliers among the Y Values	236
6.4.4	Do Not Assume Homoscedasticity or That the Regression Line Is Straight	237
6.4.5	Violating Assumptions When Testing Hypotheses	239
6.4.6	Dealing with Heteroscedasticity: The HC4 Method	241
6.4.7	R Functions <code>olshc4</code> and <code>hc4test</code>	242
6.5	Pearson's Correlation and the Coefficient of Determination	243
6.5.1	A Closer Look at Interpreting r	246
6.6	Testing $H_0: \rho = 0$	250
6.6.1	R Functions <code>cor.test</code> and <code>pwr.t.test</code>	251
6.6.2	R Function <code>pwr.r.test</code>	253
6.6.3	Testing $H_0: \rho = 0$ When There is Heteroscedasticity	254
6.6.4	R Function <code>pcorhc4</code>	254
6.6.5	When Is It Safe to Conclude That Two Variables Are Independent?	254
6.7	A Regression Method for Estimating the Median of Y and Other Quantiles	255
6.7.1	R Function <code>rqfit</code>	257

6.8	Detecting Heteroscedasticity	257
6.8.1	R Function <code>khomreg</code>	258
6.9	Concluding Remarks	259
6.10	Exercises	259
7	BOOTSTRAP METHODS	265
7.1	Bootstrap-t Method	265
7.1.1	Symmetric Confidence Intervals	271
7.1.2	Exact Nonparametric Confidence Intervals for Means Are Impossible	273
7.2	The Percentile Bootstrap Method	274
7.3	Inferences about Robust Measures of Location	277
7.3.1	Using the Percentile Method	277
7.3.2	R Functions <code>onesampb</code> , <code>momci</code> , and <code>trimpb</code>	278
7.3.3	The Bootstrap-t Method Based on Trimmed Means	279
7.3.4	R Function <code>trimcibt</code>	281
7.4	Estimating Power When Testing Hypotheses about a Trimmed Mean	282
7.4.1	R Functions <code>powt1est</code> and <code>powt1an</code>	283
7.5	A Bootstrap Estimate of Standard Errors	286
7.5.1	R Function <code>bootse</code>	287
7.6	Inferences about Pearson's Correlation: <i>Dealing with</i> <i>Heteroscedasticity</i>	287
7.6.1	R Function <code>pcorb</code>	289
7.7	Bootstrap Methods for Least Squares Regression	290
7.7.1	R Functions <code>hc4wtest</code> , <code>olswbtest</code> , <code>lsfitci</code>	292
7.8	Detecting Associations Even When There Is Curvature	294
7.8.1	R Functions <code>indt</code> and <code>medind</code>	296
7.9	Quantile Regression	297
7.9.1	R Functions <code>qregci</code> and <code>rqtest</code>	298
7.9.2	A Test for Homoscedasticity Using a Quantile Regression Approach	298
7.9.3	R Function <code>qhomt</code>	300
7.10	Regression: Which Predictors Are Best?	300
7.10.1	R Function <code>regpre</code>	303
7.10.2	Least Angle Regression	305
7.10.3	R Function <code>larsR</code>	306
7.11	Comparing Correlations	306
7.11.1	R Functions <code>TWOpov</code> and <code>TWOpNOV</code>	308
7.12	Empirical Likelihood	309
7.13	Exercises	309

8	COMPARING TWO INDEPENDENT GROUPS	313
8.1	Student's T Test	314
8.1.1	Choosing the Sample Sizes	317
8.1.2	R Function <code>power.t.test</code>	317
8.2	Relative Merits of Student's T	318
8.3	Welch's Heteroscedastic Method for Means	322
8.3.1	R Function <code>t.test</code>	324
8.3.2	Tukey's Three-Decision Rule	326
8.3.3	Non-Normality and Welch's Method	327
8.3.4	Three Modern Insights Regarding Methods for Comparing Means	328
8.4	Methods for Comparing Medians and Trimmed Means	329
8.4.1	Yuen's Method for Trimmed Means	329
8.4.2	R Functions <code>yuen</code> and <code>fac2list</code>	330
8.4.3	Comparing Medians	331
8.4.4	R Function <code>msmed</code>	332
8.5	Percentile Bootstrap Methods for Comparing Measures of Location	332
8.5.1	Using Other Measures of Location	335
8.5.2	Comparing Medians	335
8.5.3	R Function <code>medpb2</code>	335
8.5.4	Some Guidelines on When to Use the Percentile Boot- strap Method	336
8.5.5	R Functions <code>trimpb2</code> and <code>pb2gen</code>	336
8.6	Bootstrap-t Methods for Comparing Measures of Location	338
8.6.1	Comparing Means	339
8.6.2	Bootstrap-t Method When Comparing Trimmed Means	339
8.6.3	R Functions <code>yuenbt</code> and <code>yhbt</code>	341
8.6.4	Estimating Power and Judging the Sample Sizes	344
8.6.5	R Functions <code>powest</code> and <code>pow2an</code>	344
8.7	Permutation Tests	345
8.7.1	R Function <code>permg</code>	348
8.8	Rank-Based and Nonparametric Methods	348
8.8.1	Wilcoxon-Mann-Whitney Test	349
8.8.2	R Functions <code>wmw</code> and <code>wilcox.test</code>	351
8.8.3	Handling Tied Values and Heteroscedasticity	352
8.8.4	Cliff's Method	352
8.8.5	R functions <code>cid</code> and <code>cidv2</code>	354
8.8.6	The Brunner-Munzel Method	355
8.8.7	R function <code>bmp</code>	358
8.8.8	The Kolmogorov-Smirnov Test	359
8.8.9	R Function <code>ks</code>	360
8.8.10	Comparing All Quantiles Simultaneously: An Extension of the Kolmogorov-Smirnov Test	361
8.8.11	R Function <code>sband</code>	361

8.9	Graphical Methods for Comparing Groups	363
8.9.1	Error Bars	363
8.9.2	R Function ebarplot	366
8.9.3	Plotting the Shift Function	367
8.9.4	Plotting the Distributions	371
8.9.5	R Function sumplot2g	371
8.9.6	Other Approaches	371
8.10	Comparing Measures of Scale	373
8.11	Methods for Comparing Measures of Variation	374
8.11.1	R Function comvar2	375
8.11.2	Brown–Forsythe Method	375
8.11.3	Comparing Robust Measures of Variation	376
8.12	Measuring Effect Size	377
8.12.1	R Functions yuenv2 and akp.effect	385
8.13	Comparing Correlations and Regression Slopes	385
8.13.1	R Functions twopcor, twolsreg, and tworegwb	387
8.14	Comparing Two Binomials	388
8.14.1	Storer–Kim Method	388
8.14.2	Beal’s Method	389
8.14.3	R Functions twobinom, twobici, and power.prop.test	390
8.15	Making Decisions about Which Method to Use	391
8.16	Exercises	393
9	COMPARING TWO DEPENDENT GROUPS	399
9.1	The Paired T Test	401
9.1.1	When Does the Paired T Test Perform Well?	402
9.1.2	R Function t.test	403
9.2	Comparing Robust Measures of Location	404
9.2.1	R Functions yuend, ydbt, and dmedpb	406
9.2.2	Comparing Marginal M-Estimators	411
9.2.3	R Function rmrest	412
9.3	Handling Missing Values	412
9.3.1	R Functions rm2miss and rmmismcp	415
9.4	A Different Perspective When Using Robust Measures of Location	415
9.4.1	R Functions loc2dif and l2drnci	416
9.5	The Sign Test	417
9.5.1	R Function signt	417
9.6	Wilcoxon Signed Rank Test	418
9.6.1	R Function wilcox.test	420
9.7	Comparing Variances	420
9.8	Comparing Robust Measures of Scale	421
9.8.1	R Function rnrvar	422
9.9	Comparing All Quantiles	423
9.9.1	R Function lband	423

9.10	Plots for Dependent Groups	423
9.10.1	R Function <code>g2plotdifxy</code>	425
9.11	Exercises	425
10	ONE-WAY ANOVA	427
10.1	Analysis of Variance for Independent Groups	428
10.1.1	A Conceptual Overview	429
10.1.2	ANOVA via Least Squares Regression and Dummy Coding	434
10.1.3	R Functions <code>anova</code> , <code>anova1</code> , <code>aov</code> , and <code>fac2list</code>	435
10.1.4	Controlling Power and Choosing the Sample Sizes	439
10.1.5	R Functions <code>power.anova.test</code> and <code>anova.power</code>	440
10.2	Dealing with Unequal Variances	441
10.2.1	Welch's Test	442
10.3	Judging Sample Sizes and Controlling Power When Data Are Available	446
10.3.1	R Functions <code>bdanova1</code> and <code>bdanova2</code>	449
10.4	Trimmed Means	450
10.4.1	R Functions <code>t1way</code> , <code>t1wayv2</code> , and <code>t1wayF</code>	452
10.4.2	Comparing Groups Based on Medians	454
10.4.3	R Function <code>med1way</code>	455
10.5	Bootstrap Methods	455
10.5.1	A Bootstrap-t Method	455
10.5.2	R Function <code>t1waybt</code>	456
10.5.3	Two Percentile Bootstrap Methods	457
10.5.4	R Functions <code>b1way</code> and <code>pbadepth</code>	461
10.5.5	Choosing a Method	462
10.6	Random Effects Model	462
10.6.1	A Measure of Effect Size	466
10.6.2	A Heteroscedastic Method	468
10.6.3	A Method Based on Trimmed Means	470
10.6.4	R Function <code>rananova</code>	471
10.7	Rank-Based Methods	472
10.7.1	The Kruskal-Wallis Test	473
10.8	R Function <code>kruskal.test</code>	474
10.8.1	Method BDM	475
10.8.2	R Function <code>bdm</code>	476
10.9	Exercises	477
11	TWO-WAY AND THREE-WAY DESIGNS	483
11.1	Basics of a Two-Way ANOVA Design	483
11.1.1	Interactions	487
11.1.2	R Functions <code>interaction.plot</code> and <code>interplot</code>	494
11.1.3	Interactions When There Are More than Two Levels	495
11.2	Testing Hypotheses about Main Effects and Interactions	497

11.2.1	R Function <code>anova</code>	501
11.2.2	Inferences about Disordinal Interactions	502
11.2.3	The Two-Way ANOVA Model	504
11.3	Heteroscedastic Methods for Trimmed Means, Including Means	505
11.3.1	R Function <code>t2way</code>	507
11.4	Bootstrap Methods	509
11.4.1	R Functions <code>pbad2way</code> and <code>t2waybt</code>	512
11.5	Testing Hypotheses Based on Medians	514
11.5.1	R Function <code>m2way</code>	515
11.6	A Rank-Based Method for a Two-Way Design	516
11.6.1	R Function <code>bdm2way</code>	517
11.6.2	The Patel–Hoel Approach to Interactions	517
11.7	Three-Way ANOVA	519
11.7.1	R Functions <code>anova</code> and <code>t3way</code>	521
11.8	Exercises	522
12 COMPARING MORE THAN TWO DEPENDENT GROUPS		527
12.1	Comparing Means in a One-Way Design	527
12.1.1	R Function <code>aox</code>	530
12.2	Comparing Trimmed Means When Dealing with a One-Way Design	531
12.2.1	R Functions <code>rmanova</code> and <code>rmdat2mat</code>	533
12.2.2	A Bootstrap-t Method for Trimmed Means	535
12.2.3	R Function <code>rmanovab</code>	536
12.3	Percentile Bootstrap Methods for a One-Way Design	536
12.3.1	Method Based on Marginal Measures of Location	537
12.3.2	R Function <code>bd1way</code>	537
12.3.3	Inferences Based on Difference Scores	538
12.3.4	R Function <code>rmzero</code>	541
12.4	Rank-Based Methods for a One-Way Design	541
12.4.1	Friedman’s Test	541
12.4.2	R Function <code>friedman.test</code>	542
12.4.3	Method BPRM	543
12.4.4	R Function <code>bprm</code>	544
12.5	Comments on Which Method to Use	544
12.6	Between-by-Within Designs	545
12.6.1	Method for Trimmed Means	546
12.6.2	R Function <code>bwtrim</code> and <code>bw2list</code>	549
12.6.3	A Bootstrap-t Method	550
12.6.4	R Function <code>tsplitbt</code>	551
12.6.5	Inferences Based on M-estimators and Other Robust Measures of Location	552
12.6.6	R Functions <code>sppba</code> , <code>sppbb</code> , and <code>sppbi</code>	553

12.6.7	A Rank-Based Test	554
12.6.8	R Function <code>bwrank</code>	559
12.7	Within-by-Within Design	560
12.7.1	R Function <code>wwtrim</code>	561
12.8	Three-Way Designs	561
12.8.1	R Functions <code>bbwtrim</code> , <code>bwtrim</code> , and <code>wwwtrim</code>	561
12.8.2	Data Management: R Functions <code>bw2list</code> and <code>bbw2list</code>	561
12.9	Exercises	562
13	MULTIPLE COMPARISONS	565
13.1	One-Way ANOVA, Independent Groups	566
13.1.1	Fisher's Least Significant Difference Method	566
13.1.2	The Tukey-Kramer Method	568
13.1.3	R Function <code>TukeyHSD</code>	569
13.1.4	Tukey-Kramer and the ANOVA F Test	571
13.1.5	A Step-Down Method	571
13.1.6	Dunnett's T3	575
13.1.7	Games-Howell Method	576
13.1.8	Comparing Trimmed Means	577
13.1.9	R Function <code>lincon</code>	577
13.1.10	Alternative Methods for Controlling FWE	578
13.1.11	Percentile Bootstrap Methods for Comparing Trimmed Means, Medians, and M-estimators	581
13.1.12	R Functions <code>medpb</code> , <code>mcppb</code> , <code>pbmcp</code> , and <code>mcppb20</code>	582
13.1.13	A Bootstrap-t Method	583
13.1.14	R Function <code>linconb</code>	584
13.1.15	Rank-Based Methods	584
13.1.16	R Functions <code>cidmul</code> , <code>cidmulv2</code> , and <code>bmpmul</code>	585
13.2	Two-Way, between-by-between Design	585
13.2.1	Scheffé's Homoscedastic Method	588
13.2.2	Heteroscedastic Methods	589
13.2.3	Extension of Welch-Šidák and Kaiser-Bowden Methods to Trimmed Means	592
13.2.4	R Function <code>kbcon</code>	594
13.2.5	R Function <code>con2way</code>	594
13.2.6	Linear Contrasts Based on Medians	597
13.2.7	R Functions <code>mmed</code> and <code>mcp2med</code>	598
13.2.8	Bootstrap Methods	598
13.2.9	R Functions <code>linconb</code> , <code>mcp2a</code> , and <code>bbmcpb</code>	599
13.2.10	The Patel-Hoel Rank-Based Interaction Method	600
13.2.11	R Function <code>rimul</code>	600
13.3	Judging Sample Sizes	600
13.3.1	Tamhane's Procedure	601
13.3.2	R Function <code>tamhane</code>	602
13.3.3	Hochberg's Procedure	603

13.3.4	R Function hochberg	605
13.4	Methods for Dependent Groups	605
13.4.1	Linear Contrasts Based on Trimmed Means	606
13.4.2	R Function rmmcp	607
13.4.3	Comparing M-estimators	608
13.4.4	R Functions rmmcplb, dmedpb, and dtrimpb	608
13.4.5	Bootstrap-t Method	609
13.4.6	R Function bptd	609
13.5	Between-by-within Designs	609
13.5.1	R Functions bwmcp, bwamcp, bwbmcp, bwimcp, spmcpa, spmcplb, spmcpi, and bwmcplb	612
13.6	Within-by-within Designs	615
13.6.1	Three-Way Designs	616
13.6.2	R Functions con3way, mcp3atm, and rm3mcp	616
13.6.3	Bootstrap Methods for Three-Way Designs	618
13.6.4	R Functions bbwmcplb, bwwmcp, bbbmcplb, bbwmcplb, bwwmcplb, and wwwmcplb	618
13.7	Exercises	620
14	SOME MULTIVARIATE METHODS	625
14.1	Location, Scatter, and Detecting Outliers	626
14.1.1	Detecting Outliers via Robust Measures of Location and Scatter	630
14.1.2	R Functions cov.mve and com.mcd	631
14.1.3	More Measures of Location and Covariance	633
14.1.4	R Functions rmba, tbs, and ogk	633
14.1.5	R Function out	633
14.1.6	A Projection-Type Outlier Detection Method	634
14.1.7	R Functions outpro, outproMC, outproad, outproadMC, and out3d	636
14.1.8	Skipped Estimators of Location	638
14.1.9	R Functions smean	638
14.2	One-Sample Hypothesis Testing	638
14.2.1	Comparing Dependent Groups	642
14.2.2	R Functions smeancrv2, hotell, and rmdzeroOP	643
14.3	Two-Sample Case	645
14.3.1	R Functions smean2, mat2grp, and matsplit	645
14.4	MANOVA	648
14.4.1	R Function manova	649
14.4.2	Robust MANOVA Based on Trimmed Means	650
14.4.3	R Functions MULtr.anova and MULAOPv	651
14.4.4	A Multivariate Extension of the Wilcoxon–Mann–Whitney Test	652
14.4.5	Explanatory Measure of Effect Size: A Projection-Type Generalization	652

14.4.6	R Function <code>mulwmwv2</code>	653
14.5	Rank-Based Multivariate Methods	656
14.5.1	The Munzel–Brunner Method	656
14.5.2	R Function <code>mulrank</code>	657
14.5.3	The Choi–Marden Multivariate Rank Test	658
14.5.4	R Function <code>cmanova</code>	660
14.6	Multivariate Regression	661
14.6.1	Multivariate Regression Using R	662
14.6.2	Robust Multivariate Regression	662
14.6.3	R Function <code>mlrreg</code> and <code>mopreg</code>	663
14.7	Principal Components	664
14.7.1	R Functions <code>prcomp</code> and <code>regpca</code>	666
14.7.2	Robust Principal Components	671
14.7.3	R Functions <code>outpca</code> , <code>robpca</code> , <code>robpcaS</code> , <code>Ppca</code> , and <code>Ppca.summary</code>	671
14.8	Exercises	674

15 ROBUST REGRESSION AND MEASURES OF ASSOCIATION

		677
15.1	Robust Regression Estimators	677
15.1.1	The Theil–Sen Estimator	678
15.1.2	R Functions <code>tsreg</code> and <code>regplot</code>	679
15.1.3	Least Median of Squares	680
15.1.4	Least Trimmed Squares and Least Trimmed Absolute Value Estimators	680
15.1.5	R Functions <code>lmsreg</code> , <code>ltsreg</code> , and <code>ltareg</code>	681
15.1.6	M-Estimators	681
15.1.7	R Function <code>chreg</code>	682
15.1.8	Deepest Regression Line	682
15.1.9	R Function <code>mdepreq</code>	682
15.1.10	Skipped Estimators	683
15.1.11	R Functions <code>opreg</code> and <code>opregMC</code>	683
15.1.12	S-estimators and an E-Type Estimator	683
15.1.13	R Function <code>tsts</code>	683
15.2	Comments on Choosing a Regression Estimator	684
15.3	Testing Hypotheses When Using Robust Regression Estimators	686
15.3.1	R Functions <code>regtest</code> , <code>regtestMC</code> , <code>regci</code> , and <code>regciMC</code>	687
15.3.2	Comparing Measures of Location via Dummy Coding	689
15.4	Dealing with Curvature: Smoothers	691
15.4.1	Cleveland’s Smoother	691
15.4.2	R Functions <code>lowess</code> and <code>lplot</code>	693
15.4.3	Smoothers Based on Robust Measures of Location	696
15.4.4	R Functions <code>rplot</code> and <code>rplotsm</code>	697
15.4.5	More Smoothers	697

15.4.6	R Functions <code>kerreg</code> , <code>runpd</code> , and <code>qsmcobs</code>	697
15.4.7	Prediction When X Is Discrete: The R Function <code>rundis</code>	698
15.4.8	Seeing Curvature with More than Two Predictors	698
15.4.9	R Function <code>prplot</code>	699
15.4.10	Some Alternative Methods	701
15.5	Some Robust Correlations and Tests of Independence	701
15.5.1	Kendall's tau	702
15.5.2	Spearman's rho	703
15.5.3	Winsorized Correlation	703
15.5.4	R Function <code>wincor</code>	704
15.5.5	OP Correlation	704
15.5.6	R Function <code>scor</code>	705
15.5.7	Inferences about Robust Correlations: Dealing with Heteroscedasticity	706
15.5.8	R Function <code>corb</code>	706
15.6	Measuring the Strength of an Association Based on a Robust Fit	706
15.7	Comparing the Slopes of Two Independent Groups	708
15.7.1	R Functions <code>reg2ci</code> , <code>runmean2g</code> , and <code>l2plot</code>	708
15.8	Tests for Linearity	709
15.8.1	R Functions <code>lintest</code> , <code>lintestMC</code> , and <code>linchk</code>	712
15.9	Identifying the Best Predictors	713
15.9.1	R Functions <code>regpod</code> , <code>ts2str</code> , and <code>sm2strv7</code>	716
15.10	Detecting Interactions and Moderator Analysis	717
15.10.1	R Functions <code>adtest</code>	720
15.10.2	Graphical Methods for Assessing Interactions	720
15.10.3	R Functions <code>kercon</code> , <code>runsm2g</code> , <code>regi</code> , <code>ols.plot.inter</code> , and <code>reg.plot.inter</code>	722
15.11	ANCOVA	725
15.11.1	Classic ANCOVA	725
15.11.2	Some Modern ANCOVA Methods	727
15.11.3	R Functions <code>ancsm</code> , <code>Qancsm</code> , <code>ancova</code> , <code>ancpb</code> , <code>ancbbpb</code> , and <code>ancboot</code>	729
15.12	Exercises	732

16 BASIC METHODS FOR ANALYZING CATEGORICAL DATA

16.1	Goodness of Fit	735
16.1.1	R Functions <code>chisq.test</code> and <code>pwr.chisq.test</code>	737
16.2	A Test of Independence	739
16.2.1	R Function <code>chi.test.ind</code>	741
16.3	Detecting Differences in the Marginal Probabilities	742
16.3.1	R Functions <code>contab</code> and <code>mcnemar.test</code>	745
16.4	Measures of Association	746

16.4.1	The Proportion of Agreement	749
16.4.2	Kappa	750
16.4.3	Weighted Kappa	751
16.4.4	R Function Ckappa	751
16.5	Logistic Regression	752
16.5.1	R Functions glm and logreg	753
16.5.2	A Confidence Interval for the Odds Ratio	755
16.5.3	R Function ODDSR.CI	755
16.5.4	Smoothers for Logistic Regression	756
16.5.5	R Functions logrsm, rplot.bin, and logSM	757
16.6	Exercises	758
Appendix A ANSWERS TO SELECTED EXERCISES		761
Appendix B TABLES		767
Appendix C BASIC MATRIX ALGEBRA		795
Appendix D REFERENCES		803
Index		831