Modern Statistics for the Social and Behavioral Sciences

A Practical Introduction

SECOND EDITION

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

CHAPTER 1 ▪ INTRODUCTION 1

1.1 SAMPLES VERSUS POPULATIONS 2
1.2 SOFTWARE 2
1.3 R BASICS 3
1.3.1 Entering Data 3
1.3.2 R Functions and Packages 8
1.3.3 Data Sets 10
1.3.4 Arithmetic Operations 11

CHAPTER 2 ▪ NUMERICAL AND GRAPHICAL SUMMARIES OF DATA 15

2.1 BASIC SUMMATION NOTATION 16
2.2 MEASURES OF LOCATION 17
2.2.1 The Sample Mean 18
2.2.2 R Function Mean 19
2.2.3 The Sample Median 20
2.2.4 R Function for the Median 22
2.3 A CRITICISM OF THE MEDIAN: IT MIGHT TRIM TOO MANY VALUES 22
2.3.1 R Function for the Trimmed Mean 23
2.3.2 A Winsorized Mean 23
2.3.3 R Function winmean 24
2.3.4 What is a Measure of Location? 25
2.4 MEASURES OF VARIATION OR SCALE 25
2.4.1 Sample Variance and Standard Deviation 25
2.4.2 R Functions var and sd 26
2.4.3 The Interquartile Range 27
2.4.4 R Functions idealf and ideafIQR 28
2.4.5 Winsorized Variance 28
2.4.6 R Function winvar 28
2.4.7 Median Absolute Deviation 28
2.4.8 R Function mad 29
2.4.9 Average Absolute Distance from the Median 29

2.4.10 Other Robust Measures of Variation 29
2.4.11 R Functions bivar, pbvar, tauvar, and tbs 30
2.5 DETECTING OUTLIERS 32
2.5.1 A Method Based on the Mean and Variance 32
2.5.2 A Better Outlier Detection Rule: The MAD-Median Rule 33
2.5.3 R Function out 33
2.5.4 The Boxplot 34
2.5.5 R Function boxplot 35
2.5.6 Modifications of the Boxplot Rule for Detecting Outliers 36
2.5.7 R Function outbox 36
2.5.8 Other Measures of Location 37
2.5.9 R Functions mom and onestep 39
2.6 HISTOGRAMS 39
2.6.1 R Functions hist and splot 42
2.7 KERNEL DENSITY ESTIMATORS 43
2.7.1 R Functions kdplot and akerd 45
2.8 STEM-AND-LEAF DISPLAYS 46
2.8.1 R Function stem 48
2.9 SKEWNESS 49
2.9.1 Transforming Data 49
2.10 CHOOSING A MEASURE OF LOCATION 49
2.11 EXERCISES 51

CHAPTER 3 ▪ PROBABILITY AND RELATED CONCEPTS 55

3.1 BASIC PROBABILITY 56
3.2 EXPECTED VALUES 57
3.3 CONDITIONAL PROBABILITY AND INDEPENDENCE 58
3.4 POPULATION VARIANCE 63
3.5 THE BINOMIAL PROBABILITY FUNCTION 65
3.5.1 R Functions dbinom and pbinom 69
3.6 CONTINUOUS VARIABLES AND THE NORMAL CURVE 69
3.6.1 Computing Probabilities Associated with Normal Curves 71
3.6.2 R Function pnorm 71
3.6.3 R Function qnorm 77
3.7 UNDERSTANDING THE EFFECTS OF NONNORMALITY 77
3.7.1 Skewness 80
3.8 PEARSON'S CORRELATION AND THE POPULATION COVARIANCE (OPTIONAL) 82
3.8.1 Computing the Population Covariance and Pearson's Correlation 84

3.9 SOME RULES ABOUT EXPECTED VALUES 85
3.10 CHI-SQUARED DISTRIBUTIONS 86
3.11 EXERCISES 86

CHAPTER 4 ▪ SAMPLING DISTRIBUTIONS AND CONFIDENCE INTERVALS 91

4.1 RANDOM SAMPLING 92
4.2 SAMPLING DISTRIBUTIONS 92
4.2.1 Sampling Distribution of the Sample Mean 93
4.2.2 Computing Probabilities Associated with the Sample Mean 96
4.3 A CONFIDENCE INTERVAL FOR THE POPULATION MEAN 99
4.3.1 Known Variance 100
4.3.2 Confidence Intervals When σ Is Not Known 103
4.3.3 R Functions pt and qt 104
4.3.4 Confidence Interval for the Population Mean Using Student's T 105
4.3.5 R Function t.test 105
4.4 JUDGING LOCATION ESTIMATORS BASED ON THEIR SAMPLING DISTRIBUTION 106
4.4.1 Trimming and Accuracy: Another Perspective 109
4.5 AN APPROACH TO NONNORMALITY: THE CENTRAL LIMIT THEOREM 110
4.6 STUDENT'S T AND NONNORMALITY 111
4.7 CONFIDENCE INTERVALS FOR THE TRIMMED MEAN 118
4.7.1 Estimating the Standard Error of a Trimmed Mean 119
4.7.2 R Function trimse 124
4.7.3 A Confidence Interval for the Population Trimmed Mean 124
4.7.4 R Function trimci 126
4.8 TRANSFORMING DATA 128
4.9 CONFIDENCE INTERVAL FOR THE POPULATION MEDIAN 129
4.9.1 R Function sint 130
4.9.2 Estimating the Standard Error of the Sample Median 130
4.9.3 R Function msmedse 131
4.9.4 More Concerns About Tied Values 131
4.10 A REMARK ABOUT MOM AND M-ESTIMATORS 131
4.11 CONFIDENCE INTERVALS FOR THE PROBABILITY OF SUCCESS 132
4.11.1 R Functions binomci, acbinomci and and binomLCO 134
4.12 BAYESIAN METHODS 135
4.13 EXERCISES 136

CHAPTER 5 ▪ HYPOTHESIS TESTING 141

5.1 THE BASICS OF HYPOTHESIS TESTING 142

5.1.1 p-Value or Significance Level 147
5.1.2 Criticisms of Two-Sided Hypothesis Testing and p-Values 149
5.1.3 Summary and Generalization 150
5.2 POWER AND TYPE II ERRORS 151
5.2.1 Understanding How n, α, and σ Are Related to Power 154
5.3 TESTING HYPOTHESES ABOUT THE MEAN WHEN σ IS NOT KNOWN 157
5.3.1 R Function t.test 158
5.4 CONTROLLING POWER AND DETERMINING THE SAMPLE SIZE 159
5.4.1 Choosing n Prior to Collecting Data 159
5.4.2 R Function power.t.test 159
5.4.3 Stein's Method: Judging the Sample Size When Data Are Available 160
5.4.4 R Functions stein1 and stein2 162
5.5 PRACTICAL PROBLEMS WITH STUDENT'S T TEST 162
5.6 HYPOTHESIS TESTING BASED ON A TRIMMED MEAN 166
5.6.1 R Function trimci 167
5.6.2 R Functions stein1.tr and stein2.tr 168
5.7 TESTING HYPOTHESES ABOUT THE POPULATION MEDIAN 169
5.7.1 R Function sintv2 169
5.8 MAKING DECISIONS ABOUT WHICH MEASURE OF LOCATION TO USE 169
5.9 BOOTSTRAP METHODS 170
5.10 BOOTSTRAP-T METHOD 170
5.10.1 Symmetric Confidence Intervals 175
5.10.2 Exact Nonparametric Confidence Intervals for Means Are Impossible 176
5.11 THE PERCENTILE BOOTSTRAP METHOD 176
5.12 INFERENCES ABOUT ROBUST MEASURES OF LOCATION 179
5.12.1 Using the Percentile Method 179
5.12.2 R Functions onesampb, momci, and trimpb 180
5.12.3 The Bootstrap-t Method Based on Trimmed Means 181
5.12.4 R Function trimcibt 182
5.13 ESTIMATING POWER WHEN TESTING HYPOTHESES ABOUT A TRIMMED MEAN 183
5.13.1 R Functions powt1est and powt1an 184
5.14 A BOOTSTRAP ESTIMATE OF STANDARD ERRORS 185
5.14.1 R Function bootse 185
5.15 EXERCISES 185

CHAPTER 6 ▪ REGRESSION AND CORRELATION 189

6.1 THE LEAST SQUARES PRINCIPLE 190
6.2 CONFIDENCE INTERVALS AND HYPOTHESIS TESTING 192

6.2.1 Classic Inferential Techniques 196
6.2.2 Multiple Regression 198
6.2.3 R Functions ols and lm 199
6.3 STANDARDIZED REGRESSION 201
6.4 PRACTICAL CONCERNS ABOUT LEAST SQUARES REGRESSION AND HOW THEY MIGHT BE ADDRESSED 203
6.4.1 The Effect of Outliers on Least Squares Regression 204
6.4.2 Beware of Bad Leverage Points 205
6.4.3 Beware of Discarding Outliers Among the Y Values 208
6.4.4 Do Not Assume Homoscedasticity or that the Regression Line is Straight 209
6.4.5 Violating Assumptions When Testing Hypotheses 211
6.4.6 Dealing with Heteroscedasticity: The HC4 Method 211
6.4.7 R Functions olshc4 and hc4test 212
6.4.8 Interval Estimation of the Mean Response 213
6.4.9 R Function olshc4band 213
6.5 PEARSON'S CORRELATION AND THE COEFFICIENT OF DETERMINATION 214
6.5.1 A Closer Look at Interpreting r 216
6.6 TESTING H_0: $\rho = 0$ 220
6.6.1 R Function cor.test 220
6.6.2 R Function pwr.r.test 222
6.6.3 Testing H_0: $\rho = 0$ When There is Heteroscedasticity 223
6.6.4 R Function pcorhc4 223
6.6.5 When Is It Safe to Conclude that Two Variables Are Independent? 224
6.7 A REGRESSION METHOD FOR ESTIMATING THE MEDIAN OF Y AND OTHER QUANTILES 224
6.7.1 R Function rqfit 225
6.8 DETECTING HETEROSCEDASTICITY 226
6.8.1 R Function khomreg 227
6.9 INFERENCES ABOUT PEARSON'S CORRELATION: DEALING WITH HETEROSCEDASTICITY 227
6.9.1 R Function pcorb 229
6.10 BOOTSTRAP METHODS FOR LEAST SQUARES REGRESSION 229
6.10.1 R Functions hc4wtest, olswbtest, and lsfitci 231
6.11 DETECTING ASSOCIATIONS EVEN WHEN THERE IS CURVATURE 232
6.11.1 R Functions indt and medind 232
6.12 QUANTILE REGRESSION 233
6.12.1 R Functions qregci and rqtest 234
6.12.2 A Test for Homoscedasticity Using a Quantile Regression Approach 235
6.12.3 R Function qhomt 236

6.13 REGRESSION: WHICH PREDICTORS ARE BEST? 236
6.13.1 The 0.632 Bootstrap Method 238
6.13.2 R function regpre 238
6.13.3 Least Angle Regression 240
6.13.4 R Function larsR 240
6.14 COMPARING CORRELATIONS 241
6.14.1 R Functions TWOpov and TWOpNOV 242
6.15 CONCLUDING REMARKS 243
6.16 EXERCISES 243

CHAPTER 7 ▪ COMPARING TWO INDEPENDENT GROUPS 249

7.1 STUDENT'S T TEST 251
7.1.1 Choosing the Sample Sizes 252
7.1.2 R Function power.t.test 252
7.2 RELATIVE MERITS OF STUDENT'S T 254
7.3 WELCH'S HETEROSCEDASTIC METHOD FOR MEANS 257
7.3.1 R function t.test 259
7.3.2 Tukey's Three-Decision Rule 260
7.3.3 Nonnormality and Welch's Method 261
7.3.4 Three Modern Insights Regarding Methods for Comparing Means 262
7.4 METHODS FOR COMPARING MEDIANS AND TRIMMED MEANS 262
7.4.1 Yuen's Method for Trimmed Means 262
7.4.2 R Functions yuen and fac2list 263
7.4.3 Comparing Medians 264
7.4.4 R Function msmed 265
7.5 PERCENTILE BOOTSTRAP METHODS FOR COMPARING MEASURES OF LOCATION 265
7.5.1 Using Other Measures of Location 267
7.5.2 Comparing Medians 267
7.5.3 R Function medpb2 268
7.5.4 Some Guidelines on When to Use the Percentile Bootstrap Method 268
7.5.5 R Functions trimpb2, med2g, and pb2gen 268
7.6 BOOTSTRAP-T METHODS FOR COMPARING MEASURES OF LOCATION 270
7.6.1 Comparing Means 270
7.6.2 Bootstrap-t Method When Comparing Trimmed Means 271
7.6.3 R Functions yuenbt and yhbt 272
7.6.4 Estimating Power and Judging the Sample Sizes 274
7.6.5 R Functions powest and pow2an 275
7.7 PERMUTATION TESTS 275

7.8 RANK-BASED AND NONPARAMETRIC METHODS 276
7.8.1 Wilcoxon–Mann–Whitney Test 277
7.8.2 R Functions wmw and wilcox.test 279
7.8.3 Handling Tied Values and Heteroscedasticity 279
7.8.4 Cliff's Method 279
7.8.5 R Functions cid and cidv2 281
7.8.6 The Brunner–Munzel Method 281
7.8.7 R Functions bmp and loc2dif.ci 284
7.8.8 The Kolmogorov–Smirnov Test 284
7.8.9 R Function ks 285
7.8.10 Comparing All Quantiles Simultaneously: An Extension of the Kolmogorov–Smirnov Test 285
7.8.11 R Function sband 286
7.9 GRAPHICAL METHODS FOR COMPARING GROUPS 287
7.9.1 Error Bars 288
7.9.2 R Functions ebarplot and ebarplot.med 290
7.9.3 Plotting the Shift Function 291
7.9.4 Plotting the Distributions 293
7.9.5 R Function sumplot2g 294
7.9.6 Other Approaches 294
7.10 COMPARING MEASURES OF VARIATION 295
7.10.1 R Function comvar2 296
7.10.2 Brown–Forsythe Method 297
7.10.3 Comparing Robust Measures of Variation 298
7.11 MEASURING EFFECT SIZE 299
7.11.1 R Functions yuenv2 and akp.effect 304
7.12 COMPARING CORRELATIONS AND REGRESSION SLOPES 305
7.12.1 R Functions twopcor, twolsreg, and tworegwb 306
7.13 COMPARING TWO BINOMIALS 307
7.13.1 Storer–Kim Method 307
7.13.2 Beal's Method 308
7.13.3 R Functions twobinom, twobici, bi2KMSv2, and power.prop.test 308
7.13.4 Comparing Two Discrete Distributions 309
7.13.5 R Function disc2com 310
7.14 MAKING DECISIONS ABOUT WHICH METHOD TO USE 310
7.15 EXERCISES 311

CHAPTER 8 ▪ COMPARING TWO DEPENDENT GROUPS 317

8.1 THE PAIRED T TEST 319

8.1.1 When Does the Paired T Test Perform Well? 320
8.1.2 R Function t.test 321
8.2 COMPARING ROBUST MEASURES OF LOCATION 322
8.2.1 R Functions yuend, ydbt, and dmedpb 323
8.2.2 Comparing Marginal M-Estimators 327
8.2.3 R Function rmmest 328
8.2.4 Measuring Effect Size 328
8.2.5 R Function D.akp.effect 328
8.3 HANDLING MISSING VALUES 329
8.3.1 R Functions rm2miss and rmmismcp 330
8.4 A DIFFERENT PERSPECTIVE WHEN USING ROBUST MEASURES OF LOCATION 331
8.4.1 R Functions loc2dif and l2drmci 331
8.5 THE SIGN TEST 332
8.5.1 R Function signt 332
8.6 WILCOXON SIGNED RANK TEST 333
8.6.1 R Function wilcox.test 334
8.7 COMPARING VARIANCES 335
8.7.1 R Function comdvar 335
8.8 COMPARING ROBUST MEASURES OF SCALE 336
8.8.1 R Function rmrvar 336
8.9 COMPARING ALL QUANTILES 336
8.9.1 R Functions lband 337
8.10 PLOTS FOR DEPENDENT GROUPS 337
8.10.1 R Function g2plotdifxy 338
8.11 EXERCISES 338

CHAPTER 9 ▪ ONE-WAY ANOVA 341

9.1 ANALYSIS OF VARIANCE FOR INDEPENDENT GROUPS 343
9.1.1 A Conceptual Overview 343
9.1.2 ANOVA via Least Squares Regression and Dummy Coding 348
9.1.3 R Functions anova, anova1, aov, and fac2list 348
9.1.4 Controlling Power and Choosing the Sample Sizes 352
9.1.5 R Functions power.anova.test and anova.power 352
9.2 DEALING WITH UNEQUAL VARIANCES 353
9.2.1 Welch's Test 354
9.3 JUDGING SAMPLE SIZES AND CONTROLLING POWER WHEN DATA ARE AVAILABLE 357
9.3.1 R Functions bdanova1 and bdanova2 360
9.4 TRIMMED MEANS 361

9.4.1 R Functions t1way, t1wayv2, t1wayF, and g5plot 362
9.4.2 Comparing Groups Based on Medians 364
9.4.3 R Function med1way 365
9.5 BOOTSTRAP METHODS 365
9.5.1 A Bootstrap-t Method 365
9.5.2 R Functions t1waybt and BFBANOVA 366
9.5.3 Two Percentile Bootstrap Methods 367
9.5.4 R Functions b1way, pbadepth, and Qanova 370
9.5.5 Choosing a Method 370
9.6 RANDOM EFFECTS MODEL 371
9.6.1 A Measure of Effect Size 373
9.6.2 A Heteroscedastic Method 376
9.6.3 A Method Based on Trimmed Means 377
9.6.4 R Function rananova 378
9.7 RANK-BASED METHODS 379
9.7.1 The Kruskall–Wallis Test 379
9.7.2 R Function kruskal.test 380
9.7.3 Method BDM 381
9.7.4 R Functions bdm and bdmP 382
9.8 EXERCISES 383

CHAPTER 10 ▪ TWO-WAY AND THREE-WAY DESIGNS 387

10.1 BASICS OF A TWO-WAY ANOVA DESIGN 387
10.1.1 Interactions 391
10.1.2 R Functions interaction.plot and interplot 395
10.1.3 Interactions When There Are More Than Two Levels 396
10.2 TESTING HYPOTHESES ABOUT MAIN EFFECTS AND INTERACTIONS 398
10.2.1 R function anova 401
10.2.2 Inferences About Disordinal Interactions 402
10.2.3 The Two-Way ANOVA Model 403
10.3 HETEROSCEDASTIC METHODS FOR TRIMMED MEANS, INCLUDING MEANS 404
10.3.1 R Function t2way 405
10.4 BOOTSTRAP METHODS 408
10.4.1 R Functions pbad2way and t2waybt 410
10.5 TESTING HYPOTHESES BASED ON MEDIANS 412
10.5.1 R Function m2way 413
10.6 A RANK-BASED METHOD FOR A TWO-WAY DESIGN 413
10.6.1 R Function bdm2way 414

10.6.2 The Patel–Hoel Approach to Interactions 414
10.7 THREE-WAY ANOVA 415
10.7.1 R Functions anova and t3way 417
10.8 EXERCISES 418

CHAPTER 11 ▪ COMPARING MORE THAN TWO DEPENDENT GROUPS 421

11.1 COMPARING MEANS IN A ONE-WAY DESIGN 422
11.1.1 R Function aov 424
11.2 COMPARING TRIMMED MEANS WHEN DEALING WITH A ONE-WAY DESIGN 425
11.2.1 R Functions rmanova and rmdat2mat 426
11.2.2 A Bootstrap-t Method for Trimmed Means 428
11.2.3 R Function rmanovab 429
11.3 PERCENTILE BOOTSTRAP METHODS FOR A ONE-WAY DESIGN 429
11.3.1 Method Based on Marginal Measures of Location 430
11.3.2 R Function bd1way 430
11.3.3 Inferences Based on Difference Scores 431
11.3.4 R Function rmdzero 432
11.4 RANK-BASED METHODS FOR A ONE-WAY DESIGN 432
11.4.1 Friedman's Test 432
11.4.2 R Function friedman.test 434
11.4.3 Method BPRM 434
11.4.4 R Function bprm 435
11.5 COMMENTS ON WHICH METHOD TO USE 436
11.6 BETWEEN-BY-WITHIN DESIGNS 436
11.6.1 Method for Trimmed Means 437
11.6.2 R Function bwtrim and bw2list 439
11.6.3 A Bootstrap-t Method 441
11.6.4 R Function tsplitbt 441
11.6.5 Inferences Based on M-estimators and Other Robust Measures of Location 442
11.6.6 R Functions sppba, sppbb, and sppbi 443
11.6.7 A Rank-Based Test 444
11.6.8 R Function bwrank 448
11.7 WITHIN-BY-WITHIN DESIGN 449
11.7.1 R Function wwtrim 449
11.8 THREE-WAY DESIGNS 449
11.8.1 R Functions bbwtrim, bwwtrim, and wwwtrim 449
11.8.2 Data Management: R Functions bw2list and bbw2list 450
11.9 EXERCISES 451

CHAPTER 12 ▪ MULTIPLE COMPARISONS 453

12.1 ONE-WAY ANOVA AND RELATED SITUATIONS, INDEPENDENT GROUPS 455
12.1.1 Fisher's Least Significant Difference Method 455
12.1.2 The Tukey–Kramer Method 457
12.1.3 R Function TukeyHSD 458
12.1.4 Tukey–Kramer and the ANOVA F Test 459
12.1.5 Step-Down Methods 459
12.1.6 Dunnett's T3 462
12.1.7 Games–Howell Method 464
12.1.8 Comparing Trimmed Means 464
12.1.9 R Functions lincon, stepmcp and twoKlin 464
12.1.10 Alternative Methods for Controlling FWE 465
12.1.11 Percentile Bootstrap Methods for Comparing Trimmed Means, Medians, and M-estimators 468
12.1.12 R Functions medpb, linconpb, pbmcp, and p.adjust 469
12.1.13 A Bootstrap-t Method 470
12.1.14 R Function linconbt 470
12.1.15 Rank-Based Methods 471
12.1.16 R Functions cidmul, cidmulv2, and bmpmul 471
12.1.17 Comparing the Individual Probabilities of Two Discrete Distributions 471
12.1.18 R Functions binband, splotg2, cumrelf, and cumrelfT 472
12.1.19 Comparing the Quantliles of Two Independent Groups 473
12.1.20 R Functions qcomhd and qcomhdMC 474
12.1.21 Multiple Comparisons for Binomial and Categorical Data 475
12.1.22 R Functions skmcp and discmcp 475
12.2 TWO-WAY, BETWEEN-BY-BETWEEN DESIGN 475
12.2.1 Scheffé's Homoscedastic Method 477
12.2.2 Heteroscedastic Methods 479
12.2.3 Extension of Welch–Šidák and Kaiser–Bowden Methods to Trimmed Means 481
12.2.4 R Function kbcon 482
12.2.5 R Functions con2way and conCON 482
12.2.6 Linear Contrasts Based on Medians 485
12.2.7 R Functions msmed and mcp2med 485
12.2.8 Bootstrap Methods 486
12.2.9 R Functions mcp2a, bbmcppb, bbmcp 487
12.2.10 The Patel–Hoel Rank-Based Interaction Method 487
12.2.11 R Function rimul 487
12.3 JUDGING SAMPLE SIZES 487

12.3.1 Tamhane's Procedure 488
12.3.2 R Function tamhane 489
12.3.3 Hochberg's Procedure 489
12.3.4 R Function hochberg 491
12.4 METHODS FOR DEPENDENT GROUPS 491
12.4.1 Linear Contrasts Based on Trimmed Means 492
12.4.2 R Function rmmcp 493
12.4.3 Comparing M-estimators 493
12.4.4 R Functions rmmcppb, dmedpb, dtrimpb, and boxdif 494
12.4.5 Bootstrap-t Method 494
12.4.6 R Function bptd 495
12.4.7 Comparing the Quantiles of the Marginal Distributions 495
12.4.8 R Function Dqcomhd 495
12.5 BETWEEN-BY-WITHIN DESIGNS 495
12.5.1 R Functions bwmcp, bwamcp, bwbmcp, bwimcp, spmcpa, spmcpb, spmcpi, and bwmcppb 497
12.6 WITHIN-BY-WITHIN DESIGNS 499
12.6.1 Three-Way Designs 500
12.6.2 R Functions con3way, mcp3atm, and rm3mcp 500
12.6.3 Bootstrap Methods for Three-Way Designs 502
12.6.4 R Functions bbwmcp, bwwmcp, bwwmcppb, bbbmcppb, bbwmcppb, bwwmcppb, and wwwmcppb 502
12.7 EXERCISES 503

CHAPTER 13 ▪ SOME MULTIVARIATE METHODS 507

13.1 LOCATION, SCATTER, AND DETECTING OUTLIERS 508
13.1.1 Detecting Outliers Via Robust Measures of Location and Scatter 511
13.1.2 R Functions cov.mve and cov.mcd 512
13.1.3 More Measures of Location and Covariance 513
13.1.4 R Functions rmba, tbs, and ogk 514
13.1.5 R Function out 514
13.1.6 A Projection-Type Outlier Detection Method 514
13.1.7 R Functions outpro, outproMC, outproad, outproadMC, and out3d 516
13.1.8 Skipped Estimators of Location 517
13.1.9 R Function smean 517
13.2 ONE-SAMPLE HYPOTHESIS TESTING 518
13.2.1 Comparing Dependent Groups 521
13.2.2 R Functions smeancrv2, hotel1, and rmdzeroOP 522
13.3 TWO-SAMPLE CASE 523
13.3.1 R Functions smean2, mat2grp, matsplit, and mat2list 524

13.3.2 R functions matsplit, mat2grp, and mat2list 524
13.4 MANOVA 525
13.4.1 R Function manova 526
13.4.2 Robust MANOVA Based on Trimmed Means 527
13.4.3 R Functions MULtr.anova and MULAOVp 528
13.5 A MULTIVARIATE EXTENSION OF THE WILCOXON–MANN–WHITNEY TEST 529
13.5.1 Explanatory Measure of Effect Size: A Projection-Type Generalization 530
13.5.2 R Function mulwmwv2 531
13.6 RANK-BASED MULTIVARIATE METHODS 531
13.6.1 The Munzel–Brunner Method 532
13.6.2 R Function mulrank 533
13.6.3 The Choi–Marden Multivariate Rank Test 534
13.6.4 R Function cmanova 536
13.7 MULTIVARIATE REGRESSION 536
13.7.1 Multivariate Regression Using R 537
13.7.2 Robust Multivariate Regression 538
13.7.3 R Function mlrreg and mopreg 539
13.8 PRINCIPAL COMPONENTS 539
13.8.1 R Functions prcomp and regpca 541
13.8.2 Robust Principal Components 544
13.8.3 R Functions outpca, robpca, robpcaS, Ppca, and Ppca.summary 544
13.9 EXERCISES 547

CHAPTER 14 ▪ ROBUST REGRESSION AND MEASURES OF ASSOCIATION 549

14.1 ROBUST REGRESSION ESTIMATORS 550
14.1.1 The Theil–Sen Estimator 551
14.1.2 R Functions tsreg, tshdreg, and regplot 552
14.1.3 Least Median of Squares 553
14.1.4 Least Trimmed Squares and Least Trimmed Absolute Value Estimators 553
14.1.5 R Functions lmsreg, ltsreg, and ltareg 554
14.1.6 M-estimators 554
14.1.7 R Function chreg 554
14.1.8 Deepest Regression Line 555
14.1.9 R Function mdepreg 555
14.1.10 Skipped Estimators 555
14.1.11 R Functions opreg and opregMC 555
14.1.12 S-estimators and an E-type Estimator 555
14.1.13 R Function tstsreg 556

14.2 COMMENTS ON CHOOSING A REGRESSION ESTIMATOR 556
14.3 INFERENCES BASED ON ROBUST REGRESSION ESTIMATORS 557
14.3.1 Testing Hypotheses About the Slopes 558
14.3.2 Inferences About the Typical Value of Y Given X 558
14.3.3 R Functions regtest, regtestMC, regci, regciMC, regYci, and regYband 559
14.3.4 Comparing Measures of Location via Dummy Coding 563
14.4 DEALING WITH CURVATURE: SMOOTHERS 563
14.4.1 Cleveland's Smoother 564
14.4.2 R Functions lowess, lplot, lplot.pred, and lplotCI 565
14.4.3 Smoothers Based on Robust Measures of Location 566
14.4.4 R Functions rplot, rplotCIS, rplotCI, rplotCIv2, rplotCIM, rplot.pred, qhdsm, and qhdsm.pred 568
14.4.5 Prediction When X Is Discrete: The R Function rundis 570
14.4.6 Seeing Curvature with More Than Two Predictors 571
14.4.7 R Function prplot 571
14.4.8 Some Alternative Methods 571
14.4.9 Detecting Heteroscedasticity Using a Smoother 573
14.4.10 R Function rhom 573
14.5 SOME ROBUST CORRELATIONS AND TESTS OF INDEPENDENCE 573
14.5.1 Kendall's tau 573
14.5.2 Spearman's rho 574
14.5.3 Winsorized Correlation 575
14.5.4 R Function wincor 575
14.5.5 OP or Skipped Correlation 575
14.5.6 R Function scor 576
14.5.7 Inferences about Robust Correlations: Dealing with Heteroscedasticity 577
14.5.8 R Functions corb and scorci 577
14.6 MEASURING THE STRENGTH OF AN ASSOCIATION BASED ON A ROBUST FIT 577
14.7 COMPARING THE SLOPES OF TWO INDEPENDENT GROUPS 578
14.7.1 R Function reg2ci 579
14.8 TESTS FOR LINEARITY 579
14.8.1 R Functions lintest, lintestMC, and linchk 581
14.9 IDENTIFYING THE BEST PREDICTORS 582
14.9.1 Inferences Based on Independent Variables Taken in Isolation 582
14.9.2 R Functions regpord, ts2str, and sm2strv7 584
14.9.3 Inferences When Independent Variables Are Taken Together 585
14.9.4 R Function regIVcom 587
14.10 INTERACTIONS AND MODERATOR ANALYSES 588

14.10.1 R Functions olshc4.inter, ols.plot.inter, regci.inter, reg.plot.inter and adtest 590
14.10.2 Graphical Methods for Assessing Interactions 591
14.10.3 R Functions kercon, runsm2g, regi 592
14.11 ANCOVA 593
14.11.1 Classic ANCOVA 594
14.11.2 Robust ANCOVA Methods Based on a Parametric Regression Model 596
14.11.3 R Functions ancJN, ancJNmp, anclin, reg2plot, and reg2g.p2plot 597
14.11.4 ANCOVA Based on the Running-interval Smoother 599
14.11.5 R Functions ancsm, Qancsm, ancova, ancovaWMW, ancpb, ancovaUB, ancboot, ancdet, runmean2g, qhdsm2g, and l2plot 600
14.11.6 R Functions Dancts, Dancols, Dancova, Dancovapb, DancovaUB, and Dancdet 604
14.12 EXERCISES 605

CHAPTER 15 ▪ BASIC METHODS FOR ANALYZING CATEGORICAL DATA 609

15.1 GOODNESS OF FIT 610
15.1.1 R Functions chisq.test and pwr.chisq.test 611
15.2 A TEST OF INDEPENDENCE 612
15.2.1 R Function chi.test.ind 614
15.3 DETECTING DIFFERENCES IN THE MARGINAL PROBABILITIES 615
15.3.1 R Functions contab and mcnemar.test 617
15.4 MEASURES OF ASSOCIATION 618
15.4.1 The Proportion of Agreement 620
15.4.2 Kappa 621
15.4.3 Weighted Kappa 622
15.4.4 R Function Ckappa 622
15.5 LOGISTIC REGRESSION 623
15.5.1 R Functions glm and logreg 623
15.5.2 A Confidence Interval for the Odds Ratio 625
15.5.3 R Function ODDSR.CI 626
15.5.4 Smoothers for Logistic Regression 626
15.5.5 R Functions logrsm, rplot.bin, and logSM 627
15.6 EXERCISES 627

APPENDIX A ▪ ANSWERS TO SELECTED EXERCISES 631

APPENDIX B ▪ TABLES 635

APPENDIX C ▪ BASIC MATRIX ALGEBRA 663

Index 697