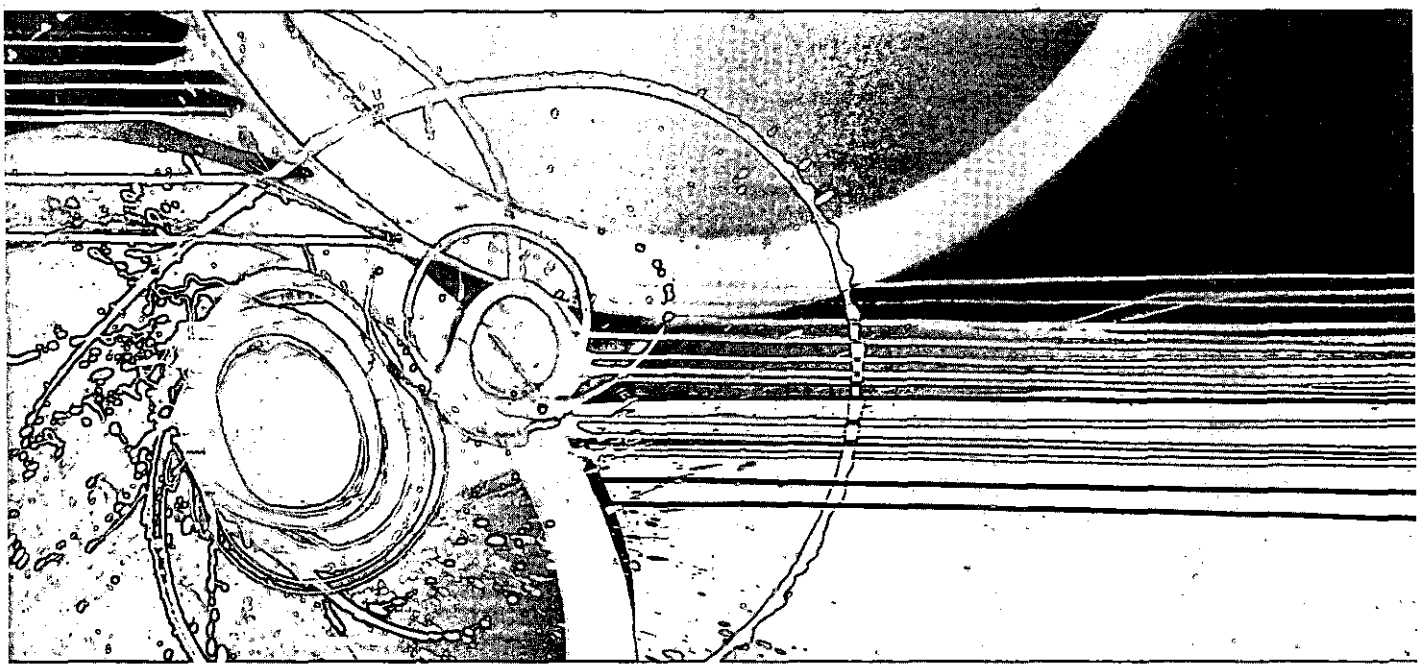
DISCOVERING STATISTICS USING R



ANDY FIELD | JEREMY MILES | ZOË FIELD

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

CONTENTS

Preface xxi
How to use this book xxv
Acknowledgements xxix
Dedication xxxi
Symbols used in this book xxxii
Some maths revision xxxiv

1 Why is my evil lecturer forcing me to learn statistics? 1

1.1. What will this chapter tell me? ① 1
1.2. What the hell am I doing here? I don't belong here ① 2
1.3. Initial observation: finding something that needs explaining ① 4
1.4. Generating theories and testing them ① 4
1.5. Data collection 1: what to measure ① 7
1.5.1. Variables ① 7
1.5.2. Measurement error ① 11
1.5.3. Validity and reliability ① 12
1.6. Data collection 2: how to measure ① 13
1.6.1. Correlational research methods ① 13
1.6.2. Experimental research methods ① 13
1.6.3. Randomization ① 17
1.7. Analysing data ① 19
1.7.1. Frequency distributions ① 19
1.7.2. The centre of a distribution ① 21
1.7.3. The dispersion in a distribution ① 24
1.7.4. Using a frequency distribution to go beyond the data ① 25
1.7.5. Fitting statistical models to the data ① 28
What have I discovered about statistics? ① 29
Key terms that I've discovered 29
Smart Alex's tasks 30
Further reading 31
Interesting real research 31

2 Everything you ever wanted to know about statistics (well, sort of) 32

2.1. What will this chapter tell me? ① 32
2.2. Building statistical models ① 33

2.3. Populations and samples ① 36
2.4. Simple statistical models ① 36
2.4.1. The mean: a very simple statistical model ① 36
2.4.2. Assessing the fit of the mean: sums of squares, variance and standard deviations ① 37
2.4.3. Expressing the mean as a model ② 40
2.5. Going beyond the data ① 41
2.5.1. The standard error ① 42
2.5.2. Confidence intervals ② 43
2.6. Using statistical models to test research questions ① 49
2.6.1. Test statistics ① 53
2.6.2. One- and two-tailed tests ① 55
2.6.3. Type I and Type II errors ① 56
2.6.4. Effect sizes ② 57
2.6.5. Statistical power ② 58
What have I discovered about statistics? ① 59
Key terms that I've discovered 60
Smart Alex's tasks 60
Further reading 60
Interesting real research 61

3 The R environment 62

3.1. What will this chapter tell me? ① 62
3.2. Before you start ① 63
3.2.1. The R-chitecture ① 63
3.2.2. Pros and cons of R ① 64
3.2.3. Downloading and installing R ① 65
3.2.4. Versions of R ① 66
3.3. Getting started ① 66
3.3.1. The main windows in R ① 67
3.3.2. Menus in R ① 67
3.4. Using R ① 71
3.4.1. Commands, objects and functions ① 71
3.4.2. Using scripts ① 75
3.4.3. The R workspace ① 76
3.4.4. Setting a working directory ② 77
3.4.5. Installing packages ① 78
3.4.6. Getting help ① 80
3.5. Getting data into R ① 81
3.5.1. Creating variables ① 81
3.5.2. Creating dataframes ① 81
3.5.3. Calculating new variables from exisiting ones ① 83
3.5.4. Organizing your data ① 85
3.5.5. Missing values ① 92
3.6. Entering data with R Commander ① 92
3.6.1. Creating variables and entering data with R Commander ① 94
3.6.2. Creating coding variables with R Commander ① 95
3.7. Using other software to enter and edit data ① 95
3.7.1. Importing data ① 97
3.7.2. Importing SPSS data files directly ① 99

3.7.3. Importing data with R Commander ① 101
3.7.4. Things that can go wrong ① 102
3.8. Saving data ① 103
3.9. Manipulating data ③ 103
3.9.1. Selecting parts of a dataframe ② 103
3.9.2. Selecting data with the subset() function ② 105
3.9.3. Dataframes and matrices ② 106
3.9.4. Reshaping data ③ 107
What have I discovered about statistics? ① 113
R packages used in this chapter 113
R functions used in this chapter 113
Key terms that I've discovered 114
Smart Alex's tasks 114
Further reading 115

4 Exploring data with graphs 116

4.1. What will this chapter tell me? ① 116
4.2. The art of presenting data ① 117
4.2.1. Why do we need graphs ① 117
4.2.2. What makes a good graph? ① 117
4.2.3. Lies, damned lies, and ... erm ... graphs ① 120
4.3. Packages used in this chapter ① 121
4.4. Introducing ggplot2 ① 121
4.4.1. The anatomy of a plot ① 121
4.3.2. Geometric objects (geoms) ① 123
4.4.3. Aesthetics ① 125
4.4.4. The anatomy of the ggplot() function ① 127
4.4.5. Stats and geoms ③ 128
4.4.6. Avoiding overplotting ② 130
4.4.7. Saving graphs ① 131
4.4.8. Putting it all together: a quick tutorial ② 132
4.5. Graphing relationships: the scatterplot ① 136
4.5.1. Simple scatterplot ① 136
4.5.2. Adding a funky line ① 138
4.5.3. Grouped scatterplot ① 140
4.6. Histograms: a good way to spot obvious problems ① 142
4.7. Boxplots (box–whisker diagrams) ① 144
4.8. Density plots ① 148
4.9. Graphing means ③ 149
4.9.1. Bar charts and error bars ② 149
4.9.2. Line graphs ② 155
4.10. Themes and options ① 161
What have I discovered about statistics? ① 163
R packages used in this chapter 163
R functions used in this chapter 164
Key terms that I've discovered 164
Smart Alex's tasks 164
Further reading 164
Interesting real research 165

5 Exploring assumptions 166

5.1. What will this chapter tell me? ① 166
5.2. What are assumptions? ① 167
5.3. Assumptions of parametric data ① 167
5.4. Packages used in this chapter ① 169
5.5. The assumption of normality ① 169
5.5.1. Oh no, it's that pesky frequency distribution again: checking normality visually ① 169
5.5.2. Quantifying normality with numbers ① 173
5.5.3. *Exploring groups of data* ① 177
5.6. Testing whether a distribution is normal ① 182
5.6.1. Doing the Shapiro–Wilk test in R ① 182
5.6.2. Reporting the Shapiro–Wilk test ① 185
5.7. Testing for homogeneity of variance ① 185
5.7.1. Levene's test ① 186
5.7.2. Reporting Levene's test ① 188
5.7.3. Hartley's F_{max}: the variance ratio ① 189
5.8. Correcting problems in the data ② 190
5.8.1. Dealing with outliers ② 190
5.8.2. Dealing with non-normality and unequal variances ② 191
5.8.3. Transforming the data using R ② 194
5.8.4. When it all goes horribly wrong ③ 201
What have I discovered about statistics? ① 203
R packages used in this chapter 204
R functions used in this chapter 204
Key terms that I've discovered 204
Smart Alex's tasks 204
Further reading 204

6 Correlation 205

6.1. What will this chapter tell me? ① 205
6.2. Looking at relationships ① 206
6.3. How do we measure relationships? ① 206
6.3.1. A detour into the murky world of covariance ① 206
6.3.2. Standardization and the correlation coefficient ① 208
6.3.3. The significance of the correlation coefficient ③ 210
6.3.4. Confidence intervals for r ③ 211
6.3.5. A word of warning about interpretation: causality ① 212
6.4. Data entry for correlation analysis ① 213
6.5. Bivariate correlation ① 213
6.5.1. Packages for correlation analysis in R ① 214
6.5.2. General procedure for correlations using R Commander ① 214
6.5.3. General procedure for correlations using R ① 216
6.5.4. Pearson's correlation coefficient ① 219
6.5.5. Spearman's correlation coefficient ① 223
6.5.6. Kendall's tau (non-parametric) ① 225
6.5.7. Bootstrapping correlations ③ 226
6.5.8. Biserial and point-biserial correlations ③ 229

6.6. Partial correlation ② 234
6.6.1. The theory behind part and partial correlation ② 234
6.6.2. Partial correlation using R ② 235
6.6.3 Semi-partial (or part) correlations ② 237
6.7. Comparing correlations ③ 238
6.7.1. Comparing independent *r*s ③ 238
6.7.2. Comparing dependent *r*s ③ 239
6.8. Calculating the effect size ① 240
6.9. How to report correlation coefficents ① 240
What have I discovered about statistics? ① 242
R packages used in this chapter 243
R functions used in this chapter 243
Key terms that I've discovered 243
Smart Alex's tasks ① 243
Further reading 244
Interesting real research 244

7 Regression 245

7.1. What will this chapter tell me? ① 245
7.2. An introduction to regression ① 246
7.2.1. Some important information about straight lines ① 247
7.2.2. The method of least squares ① 248
7.2.3. Assessing the goodness of fit: sums of squares, R and R^2 ① 249
7.2.4. Assessing individual predictors ① 252
7.3. Packages used in this chapter ① 253
7.4. General procedure for regression in R ① 254
7.4.1. Doing simple regression using R Commander ① 254
7.4.2. Regression in R ① 255
7.5. Interpreting a simple regression ① 257
7.5.1. Overall fit of the object model ① 258
7.5.2. Model parameters ① 259
7.5.3. Using the model ① 260
7.6. Multiple regression: the basics ② 261
7.6.1. An example of a multiple regression model ② 261
7.6.2. Sums of squares, R and R^2 ② 262
7.6.3. Parsimony-adjusted measures of fit ② 263
7.6.4. Methods of regression ② 263
7.7. How accurate is my regression model? ② 266
7.7.1. Assessing the regression model I: diagnostics ② 266
7.7.2. Assessing the regression model II: generalization ② 271
7.8. How to do multiple regression using R Commander and R ② 276
7.8.1. Some things to think about before the analysis ② 276
7.8.2. Multiple regression: running the basic model ② 277
7.8.3. Interpreting the basic multiple regression ② 280
7.8.4. Comparing models ② 284
7.9. Testing the accuracy of your regression model ② 287
7.9.1. Diagnostic tests using R Commander ② 287
7.9.2. Outliers and influential cases ② 288

7.9.3. Assessing the assumption of independence ② 291
7.9.4. Assessing the assumption of no multicollinearity ② 292
7.9.5. Checking assumptions about the residuals ② 294
7.9.6. What if I violate an assumption? ② 298
7.10. Robust regression: bootstrapping ③ 298
7.11. How to report multiple regression ② 301
7.12. Categorical predictors and multiple regression ③ 302
7.12.1. Dummy coding ③ 302
7.12.2. Regression with dummy variables ③ 305
What have I discovered about statistics? ① 308
R packages used in this chapter 309
R functions used in this chapter 309
Key terms that I've discovered 309
Smart Alex's tasks 310
Further reading 311
Interesting real research 311

8 Logistic regression 312

8.1. What will this chapter tell me? ① 312
8.2. Background to logistic regression ① 313
8.3. What are the principles behind logistic regression? ③ 313
8.3.1. Assessing the model: the log-likelihood statistic ③ 315
8.3.2. Assessing the model: the deviance statistic ③ 316
8.3.3. Assessing the model: R and R^2 ③ 316
8.3.4. Assessing the model: information criteria ③ 318
8.3.5. Assessing the contribution of predictors: the z-statistic ② 318
8.3.6. The odds ratio ③ 319
8.3.7. Methods of logistic regression ② 320
8.4. Assumptions and things that can go wrong ④ 321
8.4.1. Assumptions ② 321
8.4.2. Incomplete information from the predictors ④ 322
8.4.3. Complete separation ④ 323
8.5. Packages used in this chapter ① 325
8.6. Binary logistic regression: an example that will make you feel eel ② 325
8.6.1. Preparing the data 326
8.6.2. The main logistic regression analysis ② 327
8.6.3. Basic logistic regression analysis using R ② 329
8.6.4. Interpreting a basic logistic regression ② 330
8.6.5. Model 1: Intervention only ② 330
8.6.6. Model 2: Intervention and Duration as predictors ② 336
8.6.7. Casewise diagnostics in logistic regression ② 338
8.6.8. *Calculating the effect size* ② 341
8.7. How to report logistic regression ② 341
8.8. Testing assumptions: another example ② 342
8.8.1. Testing for multicollinearity ③ 343
8.8.2. Testing for linearity of the logit ③ 344
8.9. Predicting several categories: multinomial logistic regression ③ 346
8.9.1. Running multinomial logistic regression in R ③ 347
8.9.2. Interpreting the multinomial logistic regression output ③ 350

8.9.3. Reporting the results 355
What have I discovered about statistics? ① 355
R packages used in this chapter 356
R functions used in this chapter 356
Key terms that I've discovered 356
Smart Alex's tasks 357
Further reading 358
Interesting real research 358

9 Comparing two means 359

9.1. What will this chapter tell me? ① 359
9.2. Packages used in this chapter ① 360
9.3. Looking at differences ① 360
9.3.1. A problem with error bar graphs of repeated-measures designs ① 361
9.3.2. Step 1: calculate the mean for each participant ② 364
9.3.3. Step 2: calculate the grand mean ② 364
9.3.4. Step 3: calculate the adjustment factor ② 364
9.3.5. Step 4: create adjusted values for each variable ② 365
9.4. The *t*-test ① 368
9.4.1. Rationale for the *t*-test ① 369
9.4.2. The *t*-test as a general linear model ② 370
9.4.3. Assumptions of the *t*-test ① 372
9.5. The independent *t*-test ① 372
9.5.1. The independent *t*-test equation explained ① 372
9.5.2. Doing the independent *t*-test ① 375
9.6. The dependent *t*-test ① 386
9.6.1. Sampling distributions and the standard error ① 386
9.6.2. The dependent *t*-test equation explained ① 387
9.6.3. Dependent *t*-tests using *R* ① 388
9.7. Between groups or repeated measures? ① 394
What have I discovered about statistics? ① 395
R packages used in this chapter 396
R functions used in this chapter 396
Key terms that I've discovered 396
Smart Alex's tasks 396
Further reading 397
Interesting real research 397

10 Comparing several means: ANOVA (GLM 1) 398

10.1. What will this chapter tell me? ① 398
10.2. The theory behind ANOVA ② 399
10.2.1 Inflated error rates ② 399
10.2.2. Interpreting *F* ② 400
10.2.3. ANOVA as regression ② 400
10.2.4. Logic of the *F*-ratio ② 405
10.2.5. Total sum of squares (SS_T) ② 407
10.2.6. Model sum of squares (SS_M) ② 409
10.2.7. Residual sum of squares (SS_R) ② 410
10.2.8. Mean squares ② 411

10.2.9. The *F*-ratio ② 411
10.3. Assumptions of ANOVA ③ 412
10.3.1. Homogeneity of variance ② 412
10.3.2. Is ANOVA robust? ③ 412
10.4. Planned contrasts ② 414
10.4.1. Choosing which contrasts to do ② 415
10.4.2. Defining contrasts using weights ② 419
10.4.3. Non-orthogonal comparisons ② 425
10.4.4. Standard contrasts ② 426
10.4.5. Polynomial contrasts: trend analysis ② 427
10.5. *Post hoc* procedures ② 428
10.5.1. *Post hoc* procedures and Type I (α) and Type II error rates ② 431
10.5.2. *Post hoc* procedures and violations of test assumptions ② 431
10.5.3. Summary of *post hoc* procedures ② 432
10.6. One-way ANOVA using R ② 432
10.6.1. Packages for one-way ANOVA in R ① 433
10.6.2. General procedure for one-way ANOVA ① 433
10.6.3. Entering data ① 433
10.6.4. One-way ANOVA using R Commander ② 434
10.6.5. Exploring the data ② 436
10.6.6. The main analysis ② 438
10.6.7. Planned contrasts using R ② 443
10.6.8. *Post hoc* tests using R ② 447
10.7. Calculating the effect size ② 454
10.8. Reporting results from one-way independent ANOVA ② 457
What have I discovered about statistics? ① 458
R packages used in this chapter 459
R functions used in this chapter 459
Key terms that I've discovered 459
Smart Alex's tasks 459
Further reading 461
Interesting real research 461

11 Analysis of covariance, ANCOVA (GLM 2) 462

11.1. What will this chapter tell me? ② 462
11.2. What is ANCOVA? ② 463
11.3. Assumptions and issues in ANCOVA ③ 464
11.3.1. Independence of the covariate and treatment effect ③ 464
11.3.2. Homogeneity of regression slopes ③ 466
11.4. ANCOVA using R ② 467
11.4.1. Packages for ANCOVA in R ① 467
11.4.2. General procedure for ANCOVA ① 468
11.4.3. Entering data ① 468
11.4.4. ANCOVA using R Commander ② 471
11.4.5. Exploring the data ② 471
11.4.6. Are the predictor variable and covariate independent? ② 473
11.4.7. Fitting an ANCOVA model ② 473
11.4.8. Interpreting the main ANCOVA model ② 477

11.4.9. Planned contrasts in ANCOVA ② 479
11.4.10. Interpreting the covariate ② 480
11.4.11. *Post hoc* tests in ANCOVA ② 481
11.4.12. Plots in ANCOVA ② 482
11.4.13. Some final remarks ② 482
11.4.14. Testing for homogeneity of regression slopes ③ 483
11.5. Robust ANCOVA ③ 484
11.6. Calculating the effect size ② 491
11.7. Reporting results ② 494
What have I discovered about statistics? ① 495
R packages used in this chapter 495
R functions used in this chapter 496
Key terms that I've discovered 496
Smart Alex's tasks 496
Further reading 497
Interesting real research 497

12 Factorial ANOVA (GLM 3) 498

12.1. What will this chapter tell me? ② 498
12.2. Theory of factorial ANOVA (independent design) ② 499
12.2.1. Factorial designs ② 499
12.3. Factorial ANOVA as regression ③ 501
12.3.1. An example with two independent variables ② 501
12.3.2. Extending the regression model ③ 501
12.4. Two-way ANOVA: behind the scenes ② 505
12.4.1. Total sums of squares (SS_T) ② 506
12.4.2. The model sum of squares (SS_M) ② 507
12.4.3. The residual sum of squares (SS_R) ② 510
12.4.4. The *F*-ratios ② 511
12.5. Factorial ANOVA using R ② 511
12.5.1. Packages for factorial ANOVA in R ① 511
12.5.2. General procedure for factorial ANOVA ① 512
12.5.3. Factorial ANOVA using R Commander ② 512
12.5.4. Entering the data ② 513
12.5.5. Exploring the data ② 516
12.5.6. Choosing contrasts ② 518
12.5.7. Fitting a factorial ANOVA model ② 520
12.5.8. Interpreting factorial ANOVA ② 520
12.5.9. Interpreting contrasts ② 524
12.5.10. Simple effects analysis ③ 525
12.5.11. *Post hoc* analysis ② 528
12.5.12. Overall conclusions 530
12.5.13. Plots in factorial ANOVA ② 530
12.6. Interpreting interaction graphs ② 530
12.7. Robust factorial ANOVA ③ 534
12.8. Calculating effect sizes ③ 542
12.9. Reporting the results of two-way ANOVA ② 544
What have I discovered about statistics? ① 546

R packages used in this chapter 546
R functions used in this chapter 546
Key terms that I've discovered 547
Smart Alex's tasks 547
Further reading 548
Interesting real research 548

13 Repeated-measures designs (GLM 4) 549

13.1. What will this chapter tell me? ② 549
13.2. Introduction to repeated-measures designs ② 550
13.2.1. The assumption of sphericity ② 551
13.2.2. How is sphericity measured? ② 551
13.2.3. Assessing the severity of departures from sphericity ② 552
13.2.4. What is the effect of violating the assumption of sphericity? ③ 552
13.2.5. What do you do if you violate sphericity? ② 554
13.3. Theory of one-way repeated-measures ANOVA ② 554
13.3.1. The total sum of squares (SS_T) ② 557
13.3.2. The within-participant sum of squares (SS_W) ② 558
13.3.3. The model sum of squares (SS_M) ② 559
13.3.4. The residual sum of squares (SS_R) ② 560
13.3.5. The mean squares ② 560
13.3.6. The *F*-ratio ② 560
13.3.7. The between-participant sum of squares ② 561
13.4. One-way repeated-measures designs using R ② 561
13.4.1. Packages for repeated measures designs in R ① 561
13.4.2. General procedure for repeated-measures designs ① 562
13.4.3. Repeated-measures ANOVA using R Commander ② 563
13.4.4. Entering the data ② 563
13.4.5. Exploring the data ② 565
13.4.6. Choosing contrasts ② 568
13.4.7. Analysing repeated measures: two ways to skin a .dat ② 569
13.4.8. Robust one-way repeated-measures ANOVA ③ 576
13.5. Effect sizes for repeated-measures designs ③ 580
13.6. Reporting one-way repeated-measures designs ② 581
13.7. Factorial repeated-measures designs ② 583
13.7.1. Entering the data ② 584
13.7.2. Exploring the data ② 586
13.7.3. Setting contrasts ② 588
13.7.4. Factorial repeated-measures ANOVA ② 589
13.7.5. Factorial repeated-measures designs as a GLM ③ 594
13.7.6. Robust factorial repeated-measures ANOVA ③ 599
13.8. Effect sizes for factorial repeated-measures designs ③ 599
13.9. Reporting the results from factorial repeated-measures designs ② 600
What have I discovered about statistics? ② 601
R packages used in this chapter 602
R functions used in this chapter 602
Key terms that I've discovered 602
Smart Alex's tasks 602

Further reading 603
Interesting real research 603

14 Mixed designs (GLM 5) 604

14.1. What will this chapter tell me? ① 604
14.2. Mixed designs ② 605
14.3. What do men and women look for in a partner? ② 606
14.4. Entering and exploring your data ② 606
14.4.1. Packages for mixed designs in R ① 606
14.4.2. General procedure for mixed designs ① 608
14.4.3. Entering the data ② 608
14.4.4. Exploring the data ② 610
14.5. Mixed ANOVA ② 613
14.6. Mixed designs as a GLM ③ 617
14.6.1. Setting contrasts ② 617
14.6.2. Building the model ② 619
14.6.3. The main effect of gender ② 622
14.6.4. The main effect of looks ② 623
14.6.5. The main effect of **personality** ② 624
14.6.6. The interaction between **gender** and **looks** ② 625
14.6.7. The interaction between **gender** and **personality** ② 628
14.6.8. The interaction between **looks** and **personality** ② 630
14.6.9. The interaction between **looks, personality** and **gender** ③ 635
14.6.10. Conclusions ③ 639
14.7. Calculating effect sizes ③ 640
14.8. Reporting the results of mixed ANOVA ② 641
14.9. Robust analysis for mixed designs ③ 643
What have I discovered about statistics? ② 650
R packages used in this chapter 650
R functions used in this chapter 651
Key terms that I've discovered 651
Smart Alex's tasks 651
Further reading 652
Interesting real research 652

15 Non-parametric tests 653

15.1. What will this chapter tell me? ① 653
15.2. When to use non-parametric tests ① 654
15.3. Packages used in this chapter ① 655
15.4. Comparing two independent conditions: the Wilcoxon rank-sum test ① 655
15.4.1. Theory of the Wilcoxon rank-sum test ② 655
15.4.2. Inputting data and provisional analysis ① 659
15.4.3. Running the analysis using R Commander ① 661
15.4.4. Running the analysis using R ① 662
15.4.5. Output from the Wilcoxon rank-sum test ① 664
15.4.6. Calculating an effect size ② 664
15.4.7. Writing the results ① 666

15.5. Comparing two related conditions: the Wilcoxon signed-rank test ① 667
15.5.1. Theory of the Wilcoxon signed-rank test ② 668
15.5.2. Running the analysis with R Commander ① 670
15.5.3. Running the analysis using R ① 671
15.5.4. Wilcoxon signed-rank test output ① 672
15.5.5. Calculating an effect size ② 673
15.5.6. Writing the results ① 673
15.6. Differences between several independent groups: the Kruskal–Wallis test ① 674
15.6.1. Theory of the Kruskal–Wallis test ② 675
15.6.2. Inputting data and provisional analysis ① 677
15.6.3. Doing the Kruskal–Wallis test using R Commander ① 679
15.6.4. Doing the Kruskal–Wallis test using R ① 679
15.6.5. Output from the Kruskal–Wallis test ① 680
15.6.6. *Post hoc* tests for the Kruskal–Wallis test ② 681
15.6.7. Testing for trends: the Jonckheere–Terpstra test ② 684
15.6.8. Calculating an effect size ② 685
15.6.9. Writing and interpreting the results ① 686
15.7. Differences between several related groups: Friedman's ANOVA ① 686
15.7.1. Theory of Friedman's ANOVA ② 688
15.7.2. Inputting data and provisional analysis ① 689
15.7.3. Doing Friedman's ANOVA in R Commander ① 690
15.7.4. Friedman's ANOVA using R ① 690
15.7.5. Output from Friedman's ANOVA ① 691
15.7.6. *Post hoc* tests for Friedman's ANOVA ② 691
15.7.7. Calculating an effect size ② 692
15.7.8. Writing and interpreting the results ① 692
What have I discovered about statistics? ① 693
R packages used in this chapter 693
R functions used in this chapter 693
Key terms that I've discovered 694
Smart Alex's tasks 694
Further reading 695
Interesting real research 695

16 Multivariate analysis of variance (MANOVA) 696

16.1. What will this chapter tell me? ② 696
16.2. When to use MANOVA ② 697
16.3. Introduction: similarities to and differences from ANOVA ② 697
16.3.1. Words of warning ② 699
16.3.2. The example for this chapter ② 699
16.4. Theory of MANOVA ③ 700
16.4.1. Introduction to matrices ③ 700
16.4.2. Some important matrices and their functions ③ 702
16.4.3. Calculating MANOVA by hand: a worked example ③ 703
16.4.4. Principle of the MANOVA test statistic ④ 710
16.5. Practical issues when conducting MANOVA ③ 717
16.5.1. Assumptions and how to check them ③ 717

16.5.2. Choosing a test statistic ③ 718
16.5.3. Follow-up analysis ③ 719
16.6. MANOVA using R ② 719
16.6.1. Packages for factorial ANOVA in R ① 719
16.6.2. General procedure for MANOVA ① 720
16.6.3. MANOVA using R Commander ② 720
16.6.4. Entering the data ② 720
16.6.5. Exploring the data ② 722
16.6.6. Setting contrasts ② 728
16.6.7. The MANOVA model ② 728
16.6.8. Follow-up analysis: univariate test statistics ② 731
16.6.9. Contrasts ③ 732
16.7. Robust MANOVA ③ 733
16.8. Reporting results from MANOVA ② 737
16.9. Following up MANOVA with discriminant analysis ③ 738
16.10. Reporting results from discriminant analysis ② 743
16.11. Some final remarks ④ 743
16.11.1. The final interpretation ④ 743
16.11.2. Univariate ANOVA or discriminant analysis? 745
What have I discovered about statistics? ② 745
R packages used in this chapter 746
R functions used in this chapter 746
Key terms that I've discovered 747
Smart Alex's tasks 747
Further reading 748
Interesting real research 748

17 Exploratory factor analysis 749

17.1. What will this chapter tell me? ① 749
17.2. When to use factor analysis ② 750
17.3. Factors ② 751
17.3.1. Graphical representation of factors ② 752
17.3.2. Mathematical representation of factors ② 753
17.3.3. Factor scores ② 755
17.3.4. Choosing a method ② 758
17.3.5. Communality ② 759
17.3.6. Factor analysis vs. principal components analysis ② 760
17.3.7. Theory behind principal components analysis ③ 761
17.3.8. Factor extraction: eigenvalues and the scree plot ② 762
17.3.9. Improving interpretation: factor rotation ③ 764
17.4. Research example ② 767
17.4.1. Sample size ② 769
17.4.2. Correlations between variables ③ 770
17.4.3. The distribution of data ② 772
17.5. Running the analysis with R Commander 772
17.6. Running the analysis with R 772
17.6.1. Packages used in this chapter ① 772
17.6.2. Initial preparation and analysis 772

17.6.3. Factor extraction using R ② 778
17.6.4. Rotation ② 788
17.6.5. Factor scores ② 793
17.6.6. Summary ② 795
17.7. How to report factor analysis ① 795
17.8. *Reliability analysis* ② 797
17.8.1. Measures of reliability ③ 797
17.8.2. Interpreting Cronbach's α (some cautionary tales ...) ② 799
17.8.3. Reliability analysis with R Commander 800
17.8.4. Reliability analysis using R ② 800
17.8.5. Interpreting the output ② 801
17.9. Reporting reliability analysis ② 806
What have I discovered about statistics? ② 807
R packages used in this chapter 807
R functions used in this chapter 808
Key terms that I've discovered 808
Smart Alex's tasks 808
Further reading 810
Interesting real research 811

18 Categorical data 812

18.1. What will this chapter tell me? ① 812
18.2. Packages used in this chapter ① 813
18.3. Analysing categorical data ① 813
18.4. Theory of analysing categorical data ① 814
18.4.1. Pearson's chi-square test ① 814
18.4.2. Fisher's exact test ① 816
18.4.3. The likelihood ratio ② 816
18.4.4. Yates's correction ② 817
18.5. Assumptions of the chi-square test ① 818
18.6. Doing the chi-square test using R ① 818
18.6.1. Entering data: raw scores ① 818
18.6.2. Entering data: the contingency table ① 819
18.6.3. Running the analysis with R Commander ① 820
18.6.4. Running the analysis using R ① 821
18.6.5. Output from the *CrossTable()* function ① 822
18.6.6. Breaking down a significant chi-square test with standardized residuals ② 825
18.6.7. Calculating an effect size ② 826
18.6.8. Reporting the results of chi-square ① 827
18.7. Several categorical variables: loglinear analysis ③ 829
18.7.1. Chi-square as regression ④ 829
18.7.2. Loglinear analysis ③ 835
18.8. Assumptions in loglinear analysis ② 837
18.9. Loglinear analysis using R ② 838
18.9.1. Initial considerations ② 838
18.9.2. Loglinear analysis as a chi-square test ② 840
18.9.3. Output from loglinear analysis as a chi-square test ② 843

18.9.4. Loglinear analysis ② 845
18.10. Following up loglinear analysis ② 850
18.11. Effect sizes in loglinear analysis ② 851
18.12. Reporting the results of loglinear analysis ② 851
What have I discovered about statistics? ① 852
R packages used in this chapter 853
R functions used in this chapter 853
Key terms that I've discovered 853
Smart Alex's tasks ③ 853
Further reading 854
Interesting real research 854

19 Multilevel linear models 855

19.1. What will this chapter tell me? ① 855
19.2. Hierarchical data ② 856
19.2.1. The intraclass correlation ② 859
19.2.2. Benefits of multilevel models ② 859
19.3. Theory of multilevel linear models ③ 860
19.3.1. An example ② 861
19.3.2. Fixed and random coefficients ③ 862
19.4. The multilevel model ④ 865
19.4.1. Assessing the fit and comparing multilevel models ④ 867
19.4.2. Types of covariance structures ④ 868
19.5. Some practical issues ③ 870
19.5.1. Assumptions ③ 870
19.5.2. Sample size and power ③ 870
19.5.3. Centring variables ④ 871
19.6. Multilevel modelling in R ④ 873
19.6.1. Packages for multilevel modelling in R 873
19.6.2. Entering the data ② 873
19.6.3. Picturing the data ② 874
19.6.4. Ignoring the data structure: ANOVA ② 874
19.6.5. Ignoring the data structure: ANCOVA ② 876
19.6.6. Assessing the need for a multilevel model ③ 878
19.6.7. Adding in fixed effects ③ 881
19.6.8. Introducing random slopes ④ 884
19.6.9. Adding an interaction term to the model ④ 886
19.7. Growth models ④ 892
19.7.1. Growth curves (polynomials) ④ 892
19.7.2. An example: the honeymoon period ② 894
19.7.3. Restructuring the data ③ 895
19.7.4. Setting up the basic model ④ 895
19.7.5. Adding in time as a fixed effect ③ 897
19.7.6. Introducing random slopes ④ 897
19.7.7. Modelling the covariance structure ④ 897
19.7.8. Comparing models ③ 899
19.7.9. Adding higher-order polynomials ③ 901
19.7.10. Further analysis ④ 905

19.8. How to report a multilevel model ③ 906
What have I discovered about statistics? ② 907
R packages used in this chapter 908
R functions used in this chapter 908
Key terms that I've discovered 908
Smart Alex's tasks 908
Further reading 909
Interesting real research 909

Epilogue: life after discovering statistics 910
Troubleshooting R 912
Glossary 913
Appendix 929
A.1. Table of the standard normal distribution 929
A.2. Critical values of the *t*-distribution 935
A.3. Critical values of the *F*-distribution 936
A.4. Critical values of the chi-square distribution 940

References 941
Index 948
Functions in R 956
Packages in R 957