CONTENTS

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

1.1 What will this chapter tell me? 1
1.2 What the hell am I doing here? I don’t belong here 2
1.3 Initial observation: finding something that needs explaining 4
1.4 Generating theories and testing them 4
1.5 Data collection 1: what to measure 7
   1.5.1 Variables 7
   1.5.2 Measurement error 11
   1.5.3 Validity and reliability 12
1.6 Data collection 2: how to measure 13
   1.6.1 Correlational research methods 13
   1.6.2 Experimental research methods 13
   1.6.3 Randomization 17
1.7 Analysing data 19
   1.7.1 Frequency distributions 19
   1.7.2 The centre of a distribution 21
   1.7.3 The dispersion in a distribution 24
   1.7.4 Using a frequency distribution to go beyond the data 25
   1.7.5 Fitting statistical models to the data 28
What have I discovered about statistics? 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? 32
2.2 Building statistical models 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.</td>
<td>Populations and samples</td>
<td>36</td>
</tr>
<tr>
<td>2.4.</td>
<td>Simple statistical models</td>
<td>36</td>
</tr>
<tr>
<td>2.4.1.</td>
<td>The mean: a very simple statistical model</td>
<td>36</td>
</tr>
<tr>
<td>2.4.2.</td>
<td>Assessing the fit of the mean: sums of squares, variance and standard deviations</td>
<td>37</td>
</tr>
<tr>
<td>2.4.3.</td>
<td>Expressing the mean as a model</td>
<td>40</td>
</tr>
<tr>
<td>2.5.</td>
<td>Going beyond the data</td>
<td>41</td>
</tr>
<tr>
<td>2.5.1.</td>
<td>The standard error</td>
<td>42</td>
</tr>
<tr>
<td>2.5.2.</td>
<td>Confidence intervals</td>
<td>43</td>
</tr>
<tr>
<td>2.6.</td>
<td>Using statistical models to test research questions</td>
<td>49</td>
</tr>
<tr>
<td>2.6.1.</td>
<td>Test statistics</td>
<td>53</td>
</tr>
<tr>
<td>2.6.2.</td>
<td>One- and two-tailed tests</td>
<td>55</td>
</tr>
<tr>
<td>2.6.3.</td>
<td>Type I and Type II errors</td>
<td>56</td>
</tr>
<tr>
<td>2.6.4.</td>
<td>Effect sizes</td>
<td>57</td>
</tr>
<tr>
<td>2.6.5.</td>
<td>Statistical power</td>
<td>58</td>
</tr>
<tr>
<td>3. The R environment</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>3.1.</td>
<td>What will this chapter tell me?</td>
<td>62</td>
</tr>
<tr>
<td>3.2.</td>
<td>Before you start</td>
<td>63</td>
</tr>
<tr>
<td>3.2.1.</td>
<td>The R-chitecture</td>
<td>63</td>
</tr>
<tr>
<td>3.2.2.</td>
<td>Pros and cons of R</td>
<td>64</td>
</tr>
<tr>
<td>3.2.3.</td>
<td>Downloading and installing R</td>
<td>65</td>
</tr>
<tr>
<td>3.2.4.</td>
<td>Versions of R</td>
<td>66</td>
</tr>
<tr>
<td>3.3.</td>
<td>Getting started</td>
<td>66</td>
</tr>
<tr>
<td>3.3.1.</td>
<td>The main windows in R</td>
<td>67</td>
</tr>
<tr>
<td>3.3.2.</td>
<td>Menus in R</td>
<td>67</td>
</tr>
<tr>
<td>3.4.</td>
<td>Using R</td>
<td>71</td>
</tr>
<tr>
<td>3.4.1.</td>
<td>Commands, objects and functions</td>
<td>71</td>
</tr>
<tr>
<td>3.4.2.</td>
<td>Using scripts</td>
<td>75</td>
</tr>
<tr>
<td>3.4.3.</td>
<td>The R workspace</td>
<td>76</td>
</tr>
<tr>
<td>3.4.4.</td>
<td>Setting a working directory</td>
<td>77</td>
</tr>
<tr>
<td>3.4.5.</td>
<td>Installing packages</td>
<td>78</td>
</tr>
<tr>
<td>3.4.6.</td>
<td>Getting help</td>
<td>80</td>
</tr>
<tr>
<td>3.5.</td>
<td>Getting data into R</td>
<td>81</td>
</tr>
<tr>
<td>3.5.1.</td>
<td>Creating variables</td>
<td>81</td>
</tr>
<tr>
<td>3.5.2.</td>
<td>Creating dataframes</td>
<td>81</td>
</tr>
<tr>
<td>3.5.3.</td>
<td>Calculating new variables from existing ones</td>
<td>83</td>
</tr>
<tr>
<td>3.5.4.</td>
<td>Organizing your data</td>
<td>85</td>
</tr>
<tr>
<td>3.5.5.</td>
<td>Missing values</td>
<td>92</td>
</tr>
<tr>
<td>3.6.</td>
<td>Entering data with R Commander</td>
<td>92</td>
</tr>
<tr>
<td>3.6.1.</td>
<td>Creating variables and entering data with R Commander</td>
<td>94</td>
</tr>
<tr>
<td>3.6.2.</td>
<td>Creating coding variables with R Commander</td>
<td>95</td>
</tr>
<tr>
<td>3.7.</td>
<td>Using other software to enter and edit data</td>
<td>95</td>
</tr>
<tr>
<td>3.7.1.</td>
<td>Importing data</td>
<td>97</td>
</tr>
<tr>
<td>3.7.2.</td>
<td>Importing SPSS data files directly</td>
<td>99</td>
</tr>
</tbody>
</table>
### 3.7.3 Importing data with R Commander

- Things that can go wrong

### 3.8 Saving data

### 3.9 Manipulating data

- Selecting parts of a dataframe
- Selecting data with the subset() function
- Dataframes and matrices
- Reshaping data

### What have I discovered about statistics?

### R packages used in this chapter

### R functions used in this chapter

### Key terms that I've discovered

### Smart Alex's tasks

### Further reading

---

### 4 Exploring data with graphs

#### 4.1 What will this chapter tell me?

#### 4.2 The art of presenting data

- Why do we need graphs
- What makes a good graph?
- Lies, damned lies, and ... erm ... graphs

#### 4.3 Packages used in this chapter

#### 4.4 Introducing ggplot2

- The anatomy of a plot
- Geometric objects (geoms)
- Aesthetics
- The anatomy of the ggplot() function
- Stats and geoms
- Avoiding overplotting
- Saving graphs
- Putting it all together: a quick tutorial

#### 4.5 Graphing relationships: the scatterplot

#### 4.6 Histograms: a good way to spot obvious problems

#### 4.7 Boxplots (box-whisker diagrams)

#### 4.8 Density plots

#### 4.9 Graphing means

- Bar charts and error bars
- Line graphs

#### 4.10 Themes and options

### What have I discovered about statistics?

### R packages used in this chapter

### R functions used in this chapter

### Key terms that I've discovered

### Smart Alex's tasks

### Further reading
5 Exploring assumptions

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

What have I discovered about statistics? 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

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

6.6. Partial correlation ©
   6.6.1. The theory behind part and partial correlation © 234
   6.6.2. Partial correlation using R © 235
   6.6.3. Semi-partial (or part) correlations © 237

6.7. Comparing correlations ©
   6.7.1. Comparing independent rs © 238
   6.7.2. Comparing dependent rs © 239

6.8. Calculating the effect size © 240

6.9. How to report correlation coefficients © 240

What have I discovered about statistics? © - 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 © 243
Further reading 244
Interesting real research 244

7 Regression © 245

7.1. What will this chapter tell me? © 245

7.2. An introduction to regression ©
   7.2.1. Some important information about straight lines © 247
   7.2.2. The method of least squares © 248
   7.2.3. Assessing the goodness of fit: sums of squares, $R$ and $R^2$ © 249
   7.2.4. Assessing individual predictors © 252

7.3. Packages used in this chapter © 253

7.4. General procedure for regression in R © 254
   7.4.1. Doing simple regression using R Commander © 254
   7.4.2. Regression in R © 255

7.5. Interpreting a simple regression © 257
   7.5.1. Overall fit of the object model © 258
   7.5.2. Model parameters © 259
   7.5.3. Using the model © 260

7.6. Multiple regression: the basics ©
   7.6.1. An example of a multiple regression model © 261
   7.6.2. Sums of squares, $R$ and $R^2$ © 262
   7.6.3. Parsimony-adjusted measures of fit © 263
   7.6.4. Methods of regression © 263

7.7. How accurate is my regression model? © 266
   7.7.1. Assessing the regression model I: diagnostics © 266
   7.7.2. Assessing the regression model II: generalization © 271

7.8. How to do multiple regression using R Commander and R © 276
   7.8.1. Some things to think about before the analysis © 276
   7.8.2. Multiple regression: running the basic model © 277
   7.8.3. Interpreting the basic multiple regression © 280
   7.8.4. Comparing models © 284

7.9. Testing the accuracy of your regression model © 287
   7.9.1. Diagnostic tests using R Commander © 287
   7.9.2. Outliers and influential cases © 288
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.9.3.</td>
<td>Assessing the assumption of independence</td>
<td>291</td>
</tr>
<tr>
<td>7.9.4.</td>
<td>Assessing the assumption of no multicollinearity</td>
<td>292</td>
</tr>
<tr>
<td>7.9.5.</td>
<td>Checking assumptions about the residuals</td>
<td>294</td>
</tr>
<tr>
<td>7.9.6.</td>
<td>What if I violate an assumption?</td>
<td>298</td>
</tr>
<tr>
<td>7.10.</td>
<td>Robust regression: bootstrapping</td>
<td>298</td>
</tr>
<tr>
<td>7.11.</td>
<td>How to report multiple regression</td>
<td>301</td>
</tr>
<tr>
<td>7.12.</td>
<td>Categorical predictors and multiple regression</td>
<td>302</td>
</tr>
<tr>
<td>7.12.1.</td>
<td>Dummy coding</td>
<td>302</td>
</tr>
<tr>
<td>7.12.2.</td>
<td>Regression with dummy variables</td>
<td>305</td>
</tr>
<tr>
<td>7.12.3.</td>
<td>What have I discovered about statistics?</td>
<td>308</td>
</tr>
<tr>
<td></td>
<td>R packages used in this chapter</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>R functions used in this chapter</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>Key terms that I've discovered</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>Smart Alex's tasks</td>
<td>310</td>
</tr>
<tr>
<td>7.13.</td>
<td>Further reading</td>
<td>311</td>
</tr>
<tr>
<td>7.14.</td>
<td>Interesting real research</td>
<td>311</td>
</tr>
</tbody>
</table>

8 Logistic regression

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1.</td>
<td>What will this chapter tell me?</td>
<td>312</td>
</tr>
<tr>
<td>8.2.</td>
<td>Background to logistic regression</td>
<td>313</td>
</tr>
<tr>
<td>8.3.</td>
<td>What are the principles behind logistic regression?</td>
<td>313</td>
</tr>
<tr>
<td>8.3.1.</td>
<td>Assessing the model: the log-likelihood statistic</td>
<td>315</td>
</tr>
<tr>
<td>8.3.2.</td>
<td>Assessing the model: the deviance statistic</td>
<td>316</td>
</tr>
<tr>
<td>8.3.3.</td>
<td>Assessing the model: R and $R^2$</td>
<td>316</td>
</tr>
<tr>
<td>8.3.4.</td>
<td>Assessing the model: information criteria</td>
<td>318</td>
</tr>
<tr>
<td>8.3.5.</td>
<td>Assessing the contribution of predictors: the z-statistic</td>
<td>318</td>
</tr>
<tr>
<td>8.3.6.</td>
<td>The odds ratio</td>
<td>319</td>
</tr>
<tr>
<td>8.3.7.</td>
<td>Methods of logistic regression</td>
<td>320</td>
</tr>
<tr>
<td>8.4.</td>
<td>Assumptions and things that can go wrong</td>
<td>321</td>
</tr>
<tr>
<td>8.4.1.</td>
<td>Assumptions</td>
<td>321</td>
</tr>
<tr>
<td>8.4.2.</td>
<td>Incomplete information from the predictors</td>
<td>322</td>
</tr>
<tr>
<td>8.4.3.</td>
<td>Complete separation</td>
<td>323</td>
</tr>
<tr>
<td>8.5.</td>
<td>Packages used in this chapter</td>
<td>325</td>
</tr>
<tr>
<td>8.6.</td>
<td>Binary logistic regression: an example that will make you feel eel</td>
<td>325</td>
</tr>
<tr>
<td>8.6.1.</td>
<td>Preparing the data</td>
<td>326</td>
</tr>
<tr>
<td>8.6.2.</td>
<td>The main logistic regression analysis</td>
<td>327</td>
</tr>
<tr>
<td>8.6.3.</td>
<td>Basic logistic regression analysis using R</td>
<td>329</td>
</tr>
<tr>
<td>8.6.4.</td>
<td>Interpreting a basic logistic regression</td>
<td>330</td>
</tr>
<tr>
<td>8.6.5.</td>
<td>Model 1: Intervention only</td>
<td>330</td>
</tr>
<tr>
<td>8.6.6.</td>
<td>Model 2: Intervention and Duration as predictors</td>
<td>336</td>
</tr>
<tr>
<td>8.6.7.</td>
<td>Casewise diagnostics in logistic regression</td>
<td>338</td>
</tr>
<tr>
<td>8.6.8.</td>
<td>Calculating the effect size</td>
<td>341</td>
</tr>
<tr>
<td>8.7.</td>
<td>How to report logistic regression</td>
<td>341</td>
</tr>
<tr>
<td>8.8.</td>
<td>Testing assumptions: another example</td>
<td>342</td>
</tr>
<tr>
<td>8.8.1.</td>
<td>Testing for multicollinearity</td>
<td>343</td>
</tr>
<tr>
<td>8.8.2.</td>
<td>Testing for linearity of the logit</td>
<td>344</td>
</tr>
<tr>
<td>8.9.</td>
<td>Predicting several categories: multinomial logistic regression</td>
<td>346</td>
</tr>
<tr>
<td>8.9.1.</td>
<td>Running multinomial logistic regression in R</td>
<td>347</td>
</tr>
<tr>
<td>8.9.2.</td>
<td>Interpreting the multinomial logistic regression output</td>
<td>350</td>
</tr>
</tbody>
</table>
9 Comparing two means 359

9.1. What will this chapter tell me? 359
9.2. Packages used in this chapter 360
9.3. Looking at differences 360
  9.3.1. A problem with error bar graphs of repeated-measures designs 361
  9.3.2. Step 1: calculate the mean for each participant 364
  9.3.3. Step 2: calculate the grand mean 364
  9.3.4. Step 3: calculate the adjustment factor 364
  9.3.5. Step 4: create adjusted values for each variable 365
9.4. The t-test 368
  9.4.1. Rationale for the t-test 369
  9.4.2. The t-test as a general linear model 370
  9.4.3. Assumptions of the t-test 372
9.5. The independent t-test 372
  9.5.1. The independent t-test equation explained 372
  9.5.2. Doing the independent t-test 375
9.6. The dependent t-test 386
  9.6.1. Sampling distributions and the standard error 386
  9.6.2. The dependent t-test equation explained 387
  9.6.3. Dependent t-tests using R 388
9.7. Between groups or repeated measures? 394

What have I discovered about statistics? 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? 398
10.2. The theory behind ANOVA 399
  10.2.1. Inflated error rates 399
  10.2.2. Interpreting F 400
  10.2.3. ANOVA as regression 400
  10.2.4. Logic of the F-ratio 405
  10.2.5. Total sum of squares (SS_T) 407
  10.2.6. Model sum of squares (SS_M) 409
  10.2.7. Residual sum of squares (SS_R) 410
  10.2.8. Mean squares 411
10.2.9. The F-ratio © 411
10.3. Assumptions of ANOVA © 412
  10.3.1. Homogeneity of variance © 412
  10.3.2. Is ANOVA robust? © 412
10.4. Planned contrasts © 414
  10.4.1. Choosing which contrasts to do © 415
  10.4.2. Defining contrasts using weights © 419
  10.4.3. Non-orthogonal comparisons © 425
  10.4.4. Standard contrasts © 426
  10.4.5. Polynomial contrasts: trend analysis © 427
10.5. Post hoc procedures © 428
  10.5.1. Post hoc procedures and Type I (α) and Type II error rates © 431
  10.5.2. Post hoc procedures and violations of test assumptions © 431
  10.5.3. Summary of post hoc procedures © 432
10.6. One-way ANOVA using R © 432
  10.6.1. Packages for one-way ANOVA in R © 433
  10.6.2. General procedure for one-way ANOVA © 433
  10.6.3. Entering data © 433
  10.6.4. One-way ANOVA using R Commander © 434
  10.6.5. Exploring the data © 436
  10.6.6. The main analysis © 438
  10.6.7. Planned contrasts using R © 443
  10.6.8. Post hoc tests using R © 447
10.7. Calculating the effect size © 454
10.8. Reporting results from one-way independent ANOVA © 457
  What have I discovered about statistics? © 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? © 462
  11.2. What is ANCOVA? © 463
  11.3. Assumptions and issues in ANCOVA © 464
    11.3.1. Independence of the covariate and treatment effect © 464
    11.3.2. Homogeneity of regression slopes © 466
  11.4. ANCOVA using R © 467
    11.4.1. Packages for ANCOVA in R © 467
    11.4.2. General procedure for ANCOVA © 468
    11.4.3. Entering data © 468
    11.4.4. ANCOVA using R Commander © 471
    11.4.5. Exploring the data © 471
    11.4.6. Are the predictor variable and covariate independent? © 473
    11.4.7. Fitting an ANCOVA model © 473
    11.4.8. Interpreting the main ANCOVA model © 477
12 Factorial ANOVA (GLM 3)  498

  12.1. What will this chapter tell me?  498
  12.2. Theory of factorial ANOVA (independent design)  499
    12.2.1. Factorial designs  499
  12.3. Factorial ANOVA as regression  501
    12.3.1. An example with two independent variables  501
    12.3.2. Extending the regression model  501
  12.4. Two-way ANOVA: behind the scenes  505
    12.4.1. Total sums of squares (SS\textsubscript{T})  506
    12.4.2. The model sum of squares (SS\textsubscript{M})  507
    12.4.3. The residual sum of squares (SS\textsubscript{R})  510
    12.4.4. The F-ratios  511
  12.5. Factorial ANOVA using R  511
    12.5.1. Packages for factorial ANOVA in R  511
    12.5.2. General procedure for factorial ANOVA  512
    12.5.3. Factorial ANOVA using R Commander  512
    12.5.4. Entering the data  513
    12.5.5. Exploring the data  516
    12.5.6. Choosing contrasts  518
    12.5.7. Fitting a factorial ANOVA model  520
    12.5.8. Interpreting factorial ANOVA  520
    12.5.9. Interpreting contrasts  524
    12.5.10. Simple effects analysis  525
    12.5.11. Post hoc analysis  528
    12.5.12. Overall conclusions  530
    12.5.13. Plots in factorial ANOVA  530
  12.6. Interpreting interaction graphs  530
  12.7. Robust factorial ANOVA  534
  12.8. Calculating effect sizes  542
  12.9. Reporting the results of two-way ANOVA  544
    What have I discovered about statistics?  546
13 Repeated-measures designs (GLM 4)

13.1. What will this chapter tell me?

13.2. Introduction to repeated-measures designs

13.2.1. The assumption of sphericity

13.2.2. How is sphericity measured?

13.2.3. Assessing the severity of departures from sphericity

13.2.4. What is the effect of violating the assumption of sphericity?

13.2.5. What do you do if you violate sphericity?

13.3. Theory of one-way repeated-measures ANOVA

13.3.1. The total sum of squares (SS_T)

13.3.2. The within-participant sum of squares (SS_W)

13.3.3. The model sum of squares (SS_M)

13.3.4. The residual sum of squares (SS_R)

13.3.5. The mean squares

13.3.6. The F-ratio

13.3.7. The between-participant sum of squares

13.4. One-way repeated-measures designs using R

13.4.1. Packages for repeated measures designs in R

13.4.2. General procedure for repeated-measures designs

13.4.3. Repeated-measures ANOVA using R Commander

13.4.4. Entering the data

13.4.5. Exploring the data

13.4.6. Choosing contrasts

13.4.7. Analysing repeated measures: two ways to skin a .dat

13.4.8. Robust one-way repeated-measures ANOVA

13.5. Effect sizes for repeated-measures designs

13.6. Reporting one-way repeated-measures designs

13.7. Factorial repeated-measures designs

13.7.1. Entering the data

13.7.2. Exploring the data

13.7.3. Setting contrasts

13.7.4. Factorial repeated-measures ANOVA

13.7.5. Factorial repeated-measures designs as a GLM

13.7.6. Robust factorial repeated-measures ANOVA

13.8. Effect sizes for factorial repeated-measures designs

13.9. Reporting the results from factorial repeated-measures designs

What have I discovered about statistics?
14 Mixed designs (GLM 5)

14.1. What will this chapter tell me? 604
14.2. Mixed designs 605
14.3. What do men and women look for in a partner? 606
14.4. Entering and exploring your data 606
  14.4.1. Packages for mixed designs in R 606
  14.4.2. General procedure for mixed designs 608
  14.4.3. Entering the data 608
  14.4.4. Exploring the data 610
14.5. Mixed ANOVA 613
14.6. Mixed designs as a GLM 617
  14.6.1. Setting contrasts 617
  14.6.2. Building the model 619
  14.6.3. The main effect of gender 622
  14.6.4. The main effect of looks 623
  14.6.5. The main effect of personality 624
  14.6.6. The interaction between gender and looks 625
  14.6.7. The interaction between gender and personality 628
  14.6.8. The interaction between looks and personality 630
  14.6.9. The interaction between looks, personality and gender 635
  14.6.10. Conclusions 639
14.7. Calculating effect sizes 640
14.8. Reporting the results of mixed ANOVA 641
14.9. Robust analysis for mixed designs 643

15 Non-parametric tests

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

16.5.2. Choosing a test statistic © 718
16.5.3. Follow-up analysis © 719
16.6. MANOVA using R © 719
  16.6.1. Packages for factorial ANOVA in R © 719
  16.6.2. General procedure for MANOVA © 720
  16.6.3. MANOVA using R Commander © 720
  16.6.4. Entering the data © 720
  16.6.5. Exploring the data © 722
  16.6.6. Setting contrasts © 728
  16.6.7. The MANOVA model © 728
  16.6.8. Follow-up analysis: univariate test statistics © 731
  16.6.9. Contrasts © 732
16.7. Robust MANOVA © 733
16.8. Reporting results from MANOVA © 737
16.9. Following up MANOVA with discriminant analysis © 738
16.10. Reporting results from discriminant analysis © 743
16.11. Some final remarks © 743
  16.11.1. The final interpretation © 743
  16.11.2. Univariate ANOVA or discriminant analysis? © 745
What have I discovered about statistics? © 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? © 749
17.2. When to use factor analysis © 750
17.3. Factors © 751
  17.3.1. Graphical representation of factors © 752
  17.3.2. Mathematical representation of factors © 753
  17.3.3. Factor scores © 755
  17.3.4. Choosing a method © 758
  17.3.5. Communality © 759
  17.3.6. Factor analysis vs. principal components analysis © 760
  17.3.7. Theory behind principal components analysis © 761
  17.3.8. Factor extraction: eigenvalues and the scree plot © 762
  17.3.9. Improving interpretation: factor rotation © 764
17.4. Research example © 767
  17.4.1. Sample size © 769
  17.4.2. Correlations between variables © 770
  17.4.3. The distribution of data © 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 © 772
  17.6.2. Initial preparation and analysis
17.6.3. Factor extraction using R © 778
17.6.4. Rotation © 788
17.6.5. Factor scores © 793
17.6.6. Summary © 795
17.7. How to report factor analysis © 795
17.8. Reliability analysis © 797
17.8.1. Measures of reliability © 797
17.8.2. Interpreting Cronbach's α (some cautionary tales ...) © 799
17.8.3. Reliability analysis with R Commander 800
17.8.4. Reliability analysis using R © 800
17.8.5. Interpreting the output © 801
17.9. Reporting reliability analysis © 806
What have I discovered about statistics? © 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? © 812
18.2. Packages used in this chapter © 813
18.3. Analysing categorical data © 813
18.4. Theory of analysing categorical data © 814
18.4.1. Pearson's chi-square test © 814
18.4.2. Fisher's exact test © 816
18.4.3. The likelihood ratio © 816
18.4.4. Yates's correction © 817
18.5. Assumptions of the chi-square test © 818
18.6. Doing the chi-square test using R © 818
18.6.1. Entering data: raw scores © 818
18.6.2. Entering data: the contingency table © 819
18.6.3. Running the analysis with R Commander © 820
18.6.4. Running the analysis using R © 821
18.6.5. Output from the CrossTable() function © 822
18.6.6. Breaking down a significant chi-square test with standardized residuals © 825
18.6.7. Calculating an effect size © 826
18.6.8. Reporting the results of chi-square © 827
18.7. Several categorical variables: loglinear analysis © 829
18.7.1. Chi-square as regression © 829
18.7.2. Loglinear analysis © 835
18.8. Assumptions in loglinear analysis © 837
18.9. Loglinear analysis using R © 838
18.9.1. Initial considerations © 838
18.9.2. Loglinear analysis as a chi-square test © 840
18.9.3. Output from loglinear analysis as a chi-square test © 843
### 18.9.4. Loglinear analysis

18.10. Following up loglinear analysis

18.11. Effect sizes in loglinear analysis

18.12. Reporting the results of loglinear analysis

What have I discovered about statistics?

R packages used in this chapter

R functions used in this chapter

Key terms that I’ve discovered

Smart Alex’s tasks

Further reading

Interesting real research

---

### 19 Multilevel linear models

19.1. What will this chapter tell me?

19.2. Hierarchical data

19.2.1. The intraclass correlation

19.2.2. Benefits of multilevel models

19.3. Theory of multilevel linear models

19.3.1. An example

19.3.2. Fixed and random coefficients

19.4. The multilevel model

19.4.1. Assessing the fit and comparing multilevel models

19.4.2. Types of covariance structures

19.5. Some practical issues

19.5.1. Assumptions

19.5.2. Sample size and power

19.5.3. Centring variables

19.6. Multilevel modelling in R

19.6.1. Packages for multilevel modelling in R

19.6.2. Entering the data

19.6.3. Picturing the data

19.6.4. Ignoring the data structure: ANOVA

19.6.5. Ignoring the data structure: ANCOVA

19.6.6. Assessing the need for a multilevel model

19.6.7. Adding in fixed effects

19.6.8. Introducing random slopes

19.6.9. Adding an interaction term to the model

19.7. Growth models

19.7.1. Growth curves (polynomials)

19.7.2. An example: the honeymoon period

19.7.3. Restructuring the data

19.7.4. Setting up the basic model

19.7.5. Adding in time as a fixed effect

19.7.6. Introducing random slopes

19.7.7. Modelling the covariance structure

19.7.8. Comparing models

19.7.9. Adding higher-order polynomials

19.7.10. Further analysis