## Contents

**List of Figures**  
*page xi*

**List of Tables**  
*xiv*

**Acknowledgements**  
*xv*

### Getting Started

1. **Why We Use Statistics**  

2. **How Statistics Support Science**
   - 1.1 The Problem of Induction
   - 1.2 Severe Testing
   - 1.3 Evidence in HCI
   - 1.4 New Experimentalism in HCI
   - 1.5 Big Data
   - 1.6 Conclusions

3. **Testing the Null**
   - 2.1 The Basics of NHST
   - 2.2 Going beyond *p*-Values
   - 2.3 NHST and Severe Testing
   - 2.4 Honesty in Statistics

4. **Constraining Bayes**
   - 3.1 Defining Probability
   - 3.2 Plausibility
   - 3.3 Unconstrained Bayes
   - 3.4 The Bayesian Critique of Frequentism
   - 3.5 Being Careful: A Response to the Critique
   - 3.6 So, Frequentist or Bayesian?

5. **Effects: What Tests Test**
   - 4.1 Location
   - 4.2 Dominance
   - 4.3 Variation
   - 4.4 Estimation and Significance
Contents

4.5 Big, Small and Zero Effects 62
4.6 Choosing Tests to See Effects 64

PART II HOW TO USE STATISTICS 69

5 Planning Your Statistical Analysis 71
  5.1 Principle 1: Articulation 73
  5.2 Principle 2: Simplicity 75
  5.3 Principle 3: Honesty 77
  5.4 Conclusions 79

6 A Cautionary Tail: Why You Should Not Do a One-Tailed Test 80
  6.1 A Tale of Two Tails 81
  6.2 One-Tail Bad, Two-Tails Better 82

7 Is This Normal? 86
  7.1 What Makes Data Normal? 86
  7.2 The Problems of Non-normal Data 90
  7.3 Testing for Normality 91
  7.4 Implications 93

8 Sorting Out Outliers 95
  8.1 Detecting Outliers 96
  8.2 Sources and Remedies for Outliers 98
    8.2.1 Errors in Data 99
    8.2.2 Mischievous Participants 99
    8.2.3 Faulty Study Design 100
    8.2.4 Natural Variation 101
  8.3 Conclusions 102

9 Power and Two Types of Error 104
  9.1 Type I and Type II Errors 105
  9.2 Defining Power 106
  9.3 Power and Sample Sizes 108
  9.4 Power and the Quality of Tests 110
  9.5 Summary 112

10 Using Non-Parametric Tests 114
  10.1 The Mechanics of Ranks 115
  10.2 Analysing Errors 117
    10.2.1 Type I Errors 117
    10.2.2 Type II Errors 119
  10.3 Practical Use 120
  10.4 Reporting Non-Parametric Tests 122
  10.5 Summary 123
Contents

11 A Robust t-Test 125
   11.1 A Traditional t-Test 126
   11.2 Simple Solutions? 130
   11.3 Location, Location, Location 131
   11.4 Trimmed and Winsorized Means 132
   11.5 M-Estimators 134
   11.6 Back to t-Tests 135
   11.7 Overall Advice 136

12 The ANOVA Family and Friends 139
   12.1 What ANOVA Does 140
   12.2 Is ANOVA Robust? 145
   12.3 Robust Alternatives to ANOVA 147
      12.3.1 Non-Parametric Alternatives 147
      12.3.2 Changes of Location 149
      12.3.3 Do Something Else 150
   12.4 Summary 151

13 Exploring, Over-Testing and Fishing 155
   13.1 Exploring After a Severe Test 156
   13.2 Exploratory Studies 157
   13.3 Over-Testing 159
      13.3.1 ANOVA Can (Sometimes) Help 160
      13.3.2 Planned Comparisons 161
      13.3.3 The Bonferroni Correction 161
      13.3.4 Bayesian Methods Can Over-Test Too 162
   13.4 Fishing 162
   13.5 Some Rules of Exploration 164

14 When Is a Correlation Not a Correlation? 167
   14.1 Defining Correlation 168
   14.2 Outlying Points 170
   14.3 Clusters 171
   14.4 Avoiding Problems 173
   14.5 A Final Warning 174

15 What Makes a Good Likert Item? 175
   15.1 Some Important Context 176
   15.2 Should Items Have a Midpoint? 178
   15.3 How Many Options? 180
   15.4 Label All Options or Just End-Points? 181
   15.5 The Final Story? 182

16 The Meaning of Factors 186
   16.1 From Concepts to Items 188
   16.2 From Items to Factors 190
## Contents

16.2.1 The Methods of Factor Analysis 192
16.2.2 Finding Factors 194
16.3 From Factors to Concepts? 198
16.4 What Does It Mean? 200

17 Unreliable Reliability: The Problem of Cronbach’s Alpha 204
  17.1 Reliability and Validity 205
  17.2 A Simple Model 206
  17.3 When $\alpha$ Is Low 208
  17.4 When $\alpha$ Is Too High 209
  17.5 Beyond $\alpha$ 211

18 Tests for Questionnaires 214
  18.1 Testing Likert Items 215
    18.1.1 Type I Analysis 216
    18.1.2 Power Analysis 218
    18.1.3 Which Test for Likert Items? 222
  18.2 Questionnaire Data 222
    18.2.1 Type I Analysis 224
    18.2.2 Power Analysis 225
    18.2.3 Which Test for Questionnaires? 229
  18.3 One Final Observation 229

Index 231