Contents

List of Boxes *ix*
Preface *xi*
About the Companion Site *xv*

1 A Primer on Data Summarization and Visualization *1*
1.1 Numerical Summary of Data *1*
1.2 Data Visualization *5*

2 The *p*-Value and Concept of Statistical Significance *13*
2.1 Beware of Misconceptions *17*

3 Statistical Tests: How to Get the *p*-Value *25*
3.1 *t*-Tests *29*
3.1.1 Unpaired *t*-Test *29*
3.1.2 Welch's Unpaired *t*-Test *37*
3.1.3 Paired *t*-Test *38*
3.1.4 Equivalence Testing *40*
3.2 Non-parametric Alternatives to *t*-Tests for Comparison of Two Groups *42*
3.2.1 Mann–Whitney U Test (Wilcoxon Rank-Sum) *42*
3.2.2 Brunner–Munzel Test *48*
3.2.3 Other Non-parametric Unpaired Tests *48*
3.2.4 Wilcoxon Signed-Rank Test *49*
3.2.5 Sign Test *49*
3.3 Tests for Frequency of Occurrence *50*
3.3.1 Fisher's Exact Test and Some of Its Alternatives *50*
3.3.2 Chi-Square Test *52*

3.4 Correlation and Regression Analyses *52*
3.4.1 Correlation *52*
3.4.1.1 Pearson Correlation *54*
3.4.1.2 Spearman Correlation *56*
3.4.1.3 Kendall Correlation *56*
3.4.1.4 Final Comments on Correlation *58*
3.4.2 Linear Regression *60*
3.4.3 Non-linear Regression and Comparing Curve Data *67*
3.5 ANOVA Methods for Comparison of Multiple Groups *70*
3.5.1 One-Way ANOVA *70*
3.5.2 Repeated Measures ANOVA *75*
3.5.3 Two-Way ANOVA *76*
3.6 Permutation Tests *77*

4 Common Pitfalls Associated with the Use of *p*-Values and Statistical Significance *83*
4.1 Multiple Testing Problem (and How to Correct for It) *83*
4.1.1 Multiple Hypothesis Testing Corrections *85*
4.2 Underpowered Study Design (and Power Calculations) *88*
4.2.1 Power Calculations *89*
4.3 Pseudoreplication (and Hierarchical Statistics) *94*
4.4 N-Hacking (and Interim and Sequential Analysis) *99*

5 Problems of the Concept of Statistical Significance; Alternative Approaches *105*

6 Experimental Design *111*
6.1 Starting the Project *111*
6.1.1 Pre-registration *111*
6.2 Stopping the Project *112*
6.3 Control Experiments *113*
6.4 Discarding Data *113*
6.5 Blinding *114*
6.6 Randomization and Blocking *115*
6.7 Uncontrolled Factors *117*
6.8 Which Experimental Model Should Be Used? *118*

7 Concluding Remarks *121*

Further Reading *123*

Appendix A A Brief Overview of Tools for Statistical Analysis *125*

Appendix B Additional Points on Tests for Contingency Tables *129*
B.1 Fisher's Test Conservatism *129*
B.2 Additional Details on Useful Alternatives to Fisher's Test *131*

Appendix C On the *p*-Value as a Measure of Evidence *133*

Code Samples *135*
Samples for Section 1.1 *136*
Samples for Section 1.2 *136*
Samples for Section 3.1 *143*
Samples for Section 3.2 *147*
Samples for Section 3.3 *150*
Samples for Section 3.4 *151*
Samples for Section 3.5 *159*
Samples for Section 3.6 *161*
Samples for Section 4.1 *165*
Samples for Section 4.2 *166*
Samples for Section 4.3 *170*

Bibliography *173*
Index *181*