Contents

Preface ix

Introduction xi

Chapter 1. Working with the R Software 1

1.1. Working with the R software 1
1.1.1. Why and how to work with the R software? 1
1.1.2. Recommended working method 2
1.1.3. Finding help with R 3
1.2. Basic operations for statistics in R 6
1.2.1. Manipulating the data set in R 6
1.2.2. Format the data set 12
1.2.3. Arranging the data set 13
1.2.4. Pivot tables. 14
1.3. A few graphs to summarize the data set 16
1.3.1. Display of graphs 16
1.3.2. Simple graphs (plots). 17
1.3.3. Histograms. 19
1.3.4. Boxplots 20
1.3.5. Barplots. 23
1.3.6. Pair representations. 25
1.3.7. Graphical representation of contingency diagrams 27
1.3.8. Study of the dispersion and interaction of data on a balanced and crossed model 29

Chapter 2. Fundamental Concepts in Statistics 31

2.1. Basic statistical vocabulary 31
2.1.1. Element, population and sample 31
2.1.2. Variables 33

2.2. Summarizing a sample 35
2.2.1. Main trend parameters 35
2.2.2. Dispersion parameters 38
2.3. The laws of probability 39
2.3.1. Normal distribution or Laplace Gauss's law 42
2.3.2. Other useful laws of probability 48

Chapter 3. Developing a Sampling or Experimental Plan 51

3.1. Sampling plans 51
3.1.1. Principles and criteria to be respected 52
3.1.2. Generalization and reliability of the plan 55
3.1.3. Variants of random sampling 56
3.2. Experimental plans 60
3.2.1. Principles and criteria to be respected 61
3.2.2. Generalization and sensitivity of the plan 63
3.2.3. Different experimental plans 64
3.2.4. Experimental plans and reality in the field: crucial choices for statistical analysis 67

Chapter 4. Principle of a Statistical Test 73

4.1. The usefulness of statistics 73
4.2. Decision theory 74
4.3. The statistical approach 76
4.3.1. Concepts of error risk, power and robustness 77
4.3.2. Independent and paired samples 82
4.4. Example of the application of a statistical test 82
4.4.1. Case of a Student's *t*-test of two independent bilateral samples . . . 84
4.4.2. Case of a Student's *t*-test of two independent unilateral samples . . 87
4.4.3. Case of a Student's *t*-test for comparing a sample to a standard . . 88
4.4.4. Case of a Student's *t*-test of two paired samples 90
4.4.5. Power test and research for the minimum number of elements to be considered per sample for an imposed power 94

Chapter 5. Key Choices in Statistical Tests 95

5.1. How are keys chosen? 95
5.1.1. Key 1 95
5.1.2. Key 2 96
5.1.3. Key 3 96
5.1.4. Key 4 96
5.2. Verification tests of application conditions 103
5.2.1. Cochran's rule 103

5.2.2. Kolmogorov–Smirnov test 104
5.2.3. Shapiro–Wilk test. 106
5.2.4. Fisher–Snedecor test 107
5.2.5. Bartlett's test. 109
5.2.6. Levene's test. 110
5.2.7. Goldfeld–Quandt test 112
5.2.8. Durbin–Watson test 113
5.2.9. Homogeneity test of betadisp variances. 114

Chapter 6. Comparison Tests of Unilateral and Bilateral Parameters 115

6.1. Comparisons of numbers and proportions. 115
6.1.1. Fisher's exact test. 115
6.1.2. The chi-squared conformity test 118
6.1.3. Chi-squared homogeneity test 120
6.1.4. McNemar's test 123
6.1.5. Binomial test. 124
6.1.6. Cochran–Mantel–Hantzel test 125
6.2. Comparisons of means 128
6.2.1. Student's *t*-test 128
6.2.2. Wilcoxon–Mann–Whitney U test 139
6.3. Correlation test of two quantitative variables 146
6.3.1. Linear correlation: the Bravais–Pearson coefficient 146
6.3.2. Monotonous correlation: Spearman or Kendall coefficients. 149

Chapter 7. Classical and Generalized Linear Models. 155

7.1. Principle of linear models 157
7.1.1. Classical linear models: the least squares method. 157
7.1.2. Generalized linear models: the principle of maximum likelihood . 166
7.2. Conditions of application of the model 170
7.2.1. Classic linear models. 170
7.2.2. Generalized linear models. 173
7.3. Other useful analyses 174
7.3.1. Transformations, linearization 174
7.3.2. Post hoc tests for ANOVA or equivalent GLM 177
7.3.3. The variance partition 179
7.3.4. Selection criteria for linear models: AIC or BIC criteria. 181
7.3.5. From model to prediction 182
7.4. Example of the application of different linear models. 183
7.4.1. Simple linear regression 183
7.4.2. Multiple linear regression 191

7.4.3. One-factor ANOVA 199
7.4.4. Multifactorial ANOVA 206
7.4.5. Covariance analysis (ANCOVA) 226
7.5. Examples of the application of GLMs 230
7.5.1. GLM for proportion or binary data under binomial distribution . . 230
7.5.2. GLM for discrete data under Poisson distribution 236
7.5.3. GLM for true positive data under gamma distribution 243
7.5.4. GLM for true data under Gaussian law 249

Chapter 8. Non-parametric Alternatives to Linear Models 253

8.1. Principle of non-parametric tests 253
8.2. ANOVA alternatives 254
8.2.1. Rank-based ANOVAs 254
8.2.2. Non-parametric alternative to a one-factor ANOVA: the Kruskal–Wallis test and associated *post hoc* 254
8.2.3. Non-parametric alternative to a two-factor ANOVA: Scheirer–Ray–Hare test 257
8.2.4. Non-parametric alternative to a matching ANOVA: the Friedman test 259
8.3. Non-parametric ANOVAs based on permutation tests (PERMANOVA) 262
8.3.1. Principle 262
8.3.2. Example of application in R 264
8.4. Nonlinear models 271

Conclusion 273

References 275

Index 277