

---

# Contents

---

<b>Introduction</b> . . . . .	ix
<b>Chapter 1. Elements of the Language</b> . . . . .	1
1.1. Before proceeding . . . . .	1
1.1.1. Installing R . . . . .	1
1.1.2. RStudio . . . . .	1
1.1.3. List of useful packages . . . . .	2
1.1.4. Find help . . . . .	2
1.1.5. R scripts . . . . .	2
1.2. Data representation in R . . . . .	3
1.2.1. Management of numerical variables . . . . .	3
1.2.2. Operations with a numerical variable . . . . .	4
1.2.3. Management of categorical variables . . . . .	6
1.2.4. Manipulation of categorical variables . . . . .	7
1.3. Selection of observations . . . . .	9
1.3.1. Index-based selection . . . . .	9
1.3.2. Criterion-based selection . . . . .	9
1.4. Representation and processing of missing values . . . . .	10
1.5. Importing and storing data . . . . .	11
1.5.1. Univariate data . . . . .	11
1.5.2. Multivariate data . . . . .	12
1.5.3. Storing the data in an external file . . . . .	13
1.6. Multidimensional data management . . . . .	14
1.6.1. Construction of a structured data table . . . . .	14
1.6.2. Birthwt data . . . . .	15
1.7. Key points . . . . .	18
1.8. Going further . . . . .	18
1.9. Applications . . . . .	18

---

<b>Chapter 2. Descriptive Statistics and Estimation</b> . . . . .	23
2.1. Summarizing a numerical variable . . . . .	23
2.1.1. Central tendency and shape of the distribution . . . . .	23
2.1.2. Distribution indicators . . . . .	24
2.2. Summarizing a categorical variable . . . . .	25
2.3. Graphically representing the distribution of a variable . . . . .	27
2.3.1. The case of numerical variables . . . . .	28
2.3.2. The case of categorical variables . . . . .	31
2.4. Interval estimation for a mean or a proportion . . . . .	33
2.4.1. Confidence interval for a mean . . . . .	33
2.4.2. Confidence interval for a proportion . . . . .	34
2.5. Key points . . . . .	35
2.6. Applications . . . . .	36
<b>Chapter 3. Measures and Tests of Association Between Two Variables</b> . . . . .	41
3.1. Bivariate descriptive statistics . . . . .	41
3.1.1. Describing a numeric variable according to the modalities of a qualitative variable . . . . .	41
3.1.2. Describing two qualitative variables . . . . .	44
3.2. Comparisons of two group means . . . . .	46
3.2.1. Independent samples . . . . .	46
3.2.2. Non-independent samples . . . . .	49
3.3. Comparisons of proportions . . . . .	52
3.3.1. Case of two proportions . . . . .	52
3.3.2. Chi-squared test . . . . .	53
3.3.3. The case of non-independent samples . . . . .	54
3.4. Risk and odds ratio measures . . . . .	55
3.5. Non-parametric approaches and exact tests . . . . .	56
3.6. Key points . . . . .	58
3.7. Going further . . . . .	58
3.8. Applications . . . . .	58
<b>Chapter 4. Analysis of Variance and Experimental Design</b> . . . . .	65
4.1. Data representation and descriptive statistics . . . . .	65
4.1.1. Data representation format . . . . .	65
4.1.2. Descriptive statistics and data structuring . . . . .	66
4.2. One-way ANOVA . . . . .	68
4.2.1. The one-way ANOVA model . . . . .	68
4.2.2. Comparisons using pairs of treatments . . . . .	71
4.2.3. Linear trend test . . . . .	72

---

4.3. Non-parametric one-way ANOVA . . . . .	76
4.4. Two-way ANOVA . . . . .	76
4.4.1. Construction of an ANOVA table . . . . .	77
4.4.2. Diagnostic model . . . . .	79
4.5. Key points . . . . .	80
4.6. Applications . . . . .	80
<b>Chapter 5. Correlation and Linear Regression . . . . .</b>	<b>89</b>
5.1. Descriptive statistics . . . . .	89
5.1.1. Scatterplot and Loess curve . . . . .	91
5.1.2. Parametric and non-parametric association measures . . . . .	91
5.1.3. Interval estimation and inference test . . . . .	92
5.2. Simple linear regression . . . . .	93
5.2.1. Regression line . . . . .	93
5.2.2. Interval estimation and variance analysis table . . . . .	96
5.2.3. Regression model predictions . . . . .	97
5.2.4. Diagnostic and residual analysis of the model . . . . .	98
5.2.5. Connection with ANOVA . . . . .	99
5.3. Multiple linear regression . . . . .	101
5.4. Key points . . . . .	102
5.5. Going further . . . . .	102
5.6. Applications . . . . .	102
<b>Chapter 6. Measures of Association in Epidemiology and Logistic Regression . . . . .</b>	<b>111</b>
6.1. Contingency tables and measures of association . . . . .	111
6.1.1. Contingency tables . . . . .	111
6.1.2. Measures and association tests . . . . .	112
6.1.3. Odds ratio and stratification . . . . .	112
6.2. Diagnostic studies . . . . .	115
6.2.1. Sensibility and specificity of a diagnostic test . . . . .	115
6.2.2. Positive and negative predictive values . . . . .	116
6.2.3. Synthesis table of the diagnostic properties of a test . . . . .	117
6.3. Logistic regression . . . . .	118
6.3.1. Estimation of the model's parameters . . . . .	118
6.3.2. Predictions with confidence intervals . . . . .	121
6.3.3. The case of grouped data . . . . .	123
6.4. ROC curve . . . . .	124
6.5. Key points . . . . .	125
6.6. Applications . . . . .	126

<b>Chapter 7. Survival Data Analysis</b> . . . . .	137
7.1. Data representation and descriptive statistics . . . . .	138
7.1.1. Data representation format . . . . .	138
7.2. Survival function and Kaplan–Meier curve . . . . .	138
7.2.1. Survival table . . . . .	138
7.2.2. Kaplan–Meier curve . . . . .	140
7.2.3. Cumulative hazard function . . . . .	142
7.2.4. Survival function equality test . . . . .	143
7.3. Cox regression . . . . .	144
7.4. Key points . . . . .	146
7.5. Going further . . . . .	147
7.6. Applications . . . . .	147
<b>Appendices</b> . . . . .	155
<b>Appendix 1. Introduction to RStudio</b> . . . . .	157
<b>Appendix 2. Graphs with the Lattice Package</b> . . . . .	161
<b>Appendix 3. The Hmisc and rms Packages</b> . . . . .	173
<b>Bibliography</b> . . . . .	187
<b>Index</b> . . . . .	191