Contents

Introduction ix

Chapter 1. Language Elements 1

1.1. Data representation in Stata 1
1.1.1. The Stata language 1
1.1.2. Creating and manipulating variables 3
1.1.3. Indexed or criteria-based selection of observations 5
1.1.4. Processing the missing values 6
1.1.5. Data management 7
1.1.6. Importing external data 7
1.1.7. Variable management 9
1.1.8. Converting a numerical variable into a categorical variable 11
1.2. Descriptive univariate statistics and estimation 12
1.2.1. Summarizing a numerical variable 12
1.2.2. Summarizing a categorical variable 13
1.3. Bivariate descriptive statistics 14
1.3.1. Describing a numeric variable by the levels of a categorical variable 14
1.3.2. Describing two qualitative variables 18
1.4. Key points 18
1.5. Further reading 19
1.6. Applications 19

Chapter 2. Measures of Association, Comparisons of Means and Proportions for Two Samples or More 25

2.1. Comparisons of two group means 25
2.1.1. Independent samples 25
2.1.2. Non-independent samples 29

2.1.3. Non-parametric approach . . . 29
2.2. Comparaisons of two proportions . . . 30
2.2.1. Independent samples . . . 30
2.2.2. Non-independent samples . . . 34
2.3. Risk measures and OR . . . 34
2.4. Analysis of variance . . . 37
2.4.1. One-way ANOVA . . . 37
2.4.2. Pairwise comparisons of means . . . 39
2.4.3. Linear trend test . . . 40
2.4.4. Computing specific contrasts . . . 42
2.4.5. Non-parametric approach . . . 43
2.4.6. Two-factor ANOVA . . . 44
2.5. Key points . . . 47
2.6. Further reading . . . 47
2.7. Applications . . . 47

Chapter 3. Linear Regression . . . 59

3.1. Measures of association between two numeric variables . . . 59
3.1.1. Bivariate descriptive statistics . . . 59
3.1.2. Pearson's correlation . . . 61
3.1.3. Non-parametric correlation . . . 62
3.2. Linear regression . . . 62
3.2.1. Estimation of the model parameters . . . 62
3.2.2. Pointwise and interval-based prediction . . . 64
3.2.3. Model diagnostic . . . 65
3.3. Multiple linear regression . . . 68
3.4. Key points . . . 69
3.5. Further reading . . . 70
3.6. Applications . . . 70

Chapter 4. Logistic Regression and Epidemiological Analyses . . . 79

4.1. Measures of association in epidemiology . . . 79
4.1.1. Prognostic studies and risk measures . . . 79
4.1.2. Diagnostic studies . . . 84
4.2. Logistic regression . . . 86
4.2.1. Estimation of the model parameters . . . 86
4.2.2. Logistic regression and diagnostic studies . . . 87
4.2.3. Point and interval prediction . . . 88
4.2.4. Case of grouped data . . . 90
4.3. Key points . . . 92
4.4. Further reading . . . 92
4.5. Applications . . . 92

Chapter 5. Survival Data Analysis 101

5.1. Data representation and descriptive statistics 101
5.1.1. Survival data representation format 101
5.2. Descriptive statistics 102
5.3. Survival function and Kaplan–Meier curve 103
5.3.1. Mortality table 103
5.3.2. Kaplan–Meier curve 105
5.3.3. Cumulative hazard function 106
5.3.4. Survival functions equality test 106
5.4. Cox regression 108
5.5. Key points 110
5.6. Further reading 111
5.7. Applications 111

Bibliography 117

Index 119