

Contents

Preface..... v

Chapter 1

Some fundamentals of empirical research 1

1. Introduction..... 1

2. On the relevance of quantitative methods in linguistics 3

3. The design and the logic of quantitative studies 7

3.1. Scouting 8

3.2. Hypotheses and operationalization 10

3.2.1. Scientific hypotheses in text form..... 10

3.2.2. Operationalizing your variables 15

3.2.3. Scientific hypotheses in statistical/mathematical form 18

3.3. Data collection and storage 20

3.4. The decision 26

3.4.1. One-tailed *p*-values from discrete probability distributions..... 29

3.4.2. Two-tailed *p*-values from discrete probability distributions 34

3.4.3. Extension: continuous probability distributions..... 41

4. The design of a factorial experiment: introduction 46

5. The design of a factorial experiment: another example 52

Chapter 2

Fundamentals of R 56

1. Introduction and installation 56

2. Functions and arguments 60

3. Vectors 64

3.1. Generating vectors 64

3.2. Loading and saving vectors..... 69

3.3. Editing vectors 72

4. Factors..... 79

4.1. Generating factors 79

4.2. Loading and saving factors 80

4.3. Editing factors 81

5. Data frames 84

5.1. Generating data frames 84

5.2. Loading and saving data frames..... 86

5.3.	Editing data frames	88
6.	Some programming: conditionals and loops	94
6.1.	Conditional expressions	94
6.2.	Loops.....	95
7.	Writing your own little functions	97

Chapter 3

Descriptive statistics	102
1. Univariate statistics	102
1.1. Frequency data	102
1.1.1. Scatterplots and line plots	104
1.1.2. Pie charts	108
1.1.3. Bar plots	109
1.1.4. Pareto-charts.....	111
1.1.5. Histograms	112
1.1.6. Empirical cumulative distributions	114
1.2. Measures of central tendency	115
1.2.1. The mode.....	115
1.2.2. The median.....	116
1.2.3. The arithmetic mean.....	116
1.2.4. The geometric mean.....	117
1.3. Measures of dispersion.....	119
1.3.1. Relative entropy	120
1.3.2. The range.....	121
1.3.3. Quantiles and quartiles.....	122
1.3.4. The average deviation	123
1.3.5. The standard deviation/variance	124
1.3.6. The variation coefficient	125
1.3.7. Summary functions	126
1.3.8. The standard error	128
1.4. Centering and standardization (z -scores)	130
1.5. Confidence intervals	132
1.5.1. Confidence intervals of arithmetic means.....	133
1.5.2. Confidence intervals of percentages	134
2. Bivariate statistics	136
2.1. Frequencies and crosstabulation	136
2.1.1. Bar plots and mosaic plots	137
2.1.2. Spineplots.....	138
2.1.3. Line plots.....	139

2.2.	Means	140
2.2.1.	Boxplots	141
2.2.2.	Interaction plots.....	143
2.3.	Coefficients of correlation and linear regression	147

Chapter 4

Analytical statistics	157
1. Distributions and frequencies.....	162
1.1. Distribution fitting.....	162
1.1.1. One dep. variable (ratio-scaled).....	162
1.1.2. One dep. variable (nominal/categorical).....	165
1.2. Tests for differences/independence.....	172
1.2.1. One dep. variable (ordinal/interval/ratio scaled) and one indep. variable (nominal) (indep. samples).....	172
1.2.2. One dep. variable (nom./cat.) and one indep. variable (nom./cat.) (indep.samples).....	178
1.2.3. One dep. variable (nom./cat.) (dep. samples).....	192
2. Dispersions.....	195
2.1. Goodness-of-fit test for one dep. variable (ratio-scaled).....	197
2.2. One dep. variable (ratio-scaled) and one indep. variable (nom.).....	199
3. Means	205
3.1. Goodness-of-fit tests	205
3.1.1. One dep. variable (ratio-scaled).....	205
3.1.2. One dep. variable (ordinal)	209
3.2. Tests for differences/independence.....	215
3.2.1. One dep. variable (ratio-scaled) and one indep. variable (nom.) (indep. samples)	215
3.2.2. One dep. variable (ratio-scaled) and one indep. variable (nom.) (dep. samples).....	221
3.2.3. One dep. variable (ordinal) and one indep. variable (nom.) (indep. samples)	227
3.2.4. One dep. variable (ordinal) and one indep. variable (nom.) (dep. samples).....	234
4. Coefficients of correlation and linear regression	238
4.1. The significance of the product-moment correlation	238
4.2. The significance of Kendall's Tau	243
4.3. Correlation and causality	245

Chapter 5

Selected multifactorial and multivariate methods.....	247
1. The notions of interaction and model (selection).....	247
1.1. Interactions.....	247
1.2. Model (selection)	253
1.2.1. Formulating the first model.....	253
1.2.2. Selecting a final model.....	259
2. Linear models.....	261
2.1. A linear model with a binary predictor	264
2.2. A linear model with a categorical predictor	271
2.3. A linear model with a numeric predictor	275
2.4. A linear model with two categorical predictors	276
2.5. A linear model with a categorical and a numeric predictor	280
2.6. A linear model with two numeric predictors.....	282
2.7. A linear model selection process with multiple predictors.....	285
3. Binary logistic regression models	293
3.1. A logistic regression with a binary predictor	296
3.2. A logistic regression with a categorical predictor.....	304
3.3. A logistic regression with a numeric predictor	306
3.4. A logistic regression with two categorical predictors	308
3.5. A logistic regression with a categorical and a numeric predictor	310
3.6. A logistic regression with two numeric predictors.....	311
4. Other regression models.....	316
4.1. An ordinal logistic regression with a categorical and a numeric predictor	317
4.2. A multinomial regression with a categorical and a numeric predictor	322
4.3. A Poisson/count regression with a categorical and a numeric predictor	324
5. Repeated measurements: a primer	327
5.1. One independent variable nested into subjects/items.....	329
5.2. Two independent variables nested into subjects/items	331
5.3. Two independent variables, one between, one within subjects/items	332
5.4. Mixed-effects / multi-level models	333
6. Hierarchical agglomerative cluster analysis.....	336

Chapter 6

Epilog..... 350

References..... 353

