

# Contents

<i>Preface</i>	x
<b>1 An introduction to R</b>	<b>1</b>
1.1 R as a calculator	2
1.2 Getting data into and out of R	4
1.3 Accessing information in data frames	6
1.4 Operations on data frames	10
1.4.1 Sorting a data frame by one or more columns	10
1.4.2 Changing information in a data frame	12
1.4.3 Extracting contingency tables from data frames	13
1.4.4 Calculations on data frames	15
1.5 Session management	18
<b>2 Graphical data exploration</b>	<b>20</b>
2.1 Random variables	20
2.2 Visualizing single random variables	21
2.3 Visualizing two or more variables	32
2.4 Trellis graphics	37
<b>3 Probability distributions</b>	<b>44</b>
3.1 Distributions	44
3.2 Discrete distributions	44
3.3 Continuous distributions	57
3.3.1 The normal distribution	58
3.3.2 The $t$ , $F$ , and $\chi^2$ distributions	63
<b>4 Basic statistical methods</b>	<b>68</b>
4.1 Tests for single vectors	71
4.1.1 Distribution tests	71
4.1.2 Tests for the mean	75
4.2 Tests for two independent vectors	77
4.2.1 Are the distributions the same?	78
4.2.2 Are the means the same?	79
4.2.3 Are the variances the same?	81
4.3 Paired vectors	82
4.3.1 Are the means or medians the same?	82
4.3.2 Functional relations: linear regression	84

4.3.3	What does the joint density look like?	97
4.4	A numerical vector and a factor: analysis of variance	101
4.4.1	Two numerical vectors and a factor: analysis of covariance	108
4.5	Two vectors with counts	111
4.6	A note on statistical significance	114
<b>5</b>	<b>Clustering and classification</b>	<b>118</b>
5.1	Clustering	118
5.1.1	Tables with measurements: principal components analysis	118
5.1.2	Tables with measurements: factor analysis	126
5.1.3	Tables with counts: correspondence analysis	128
5.1.4	Tables with distances: multidimensional scaling	136
5.1.5	Tables with distances: hierarchical cluster analysis	138
5.2	Classification	148
5.2.1	Classification trees	148
5.2.2	Discriminant analysis	154
5.2.3	Support vector machines	160
<b>6</b>	<b>Regression modeling</b>	<b>165</b>
6.1	Introduction	165
6.2	Ordinary least squares regression	169
6.2.1	Nonlinearities	174
6.2.2	Collinearity	181
6.2.3	Model criticism	188
6.2.4	Validation	193
6.3	Generalized linear models	195
6.3.1	Logistic regression	195
6.3.2	Ordinal logistic regression	208
6.4	Regression with breakpoints	214
6.5	Models for lexical richness	222
6.6	General considerations	236
<b>7</b>	<b>Mixed models</b>	<b>241</b>
7.1	Modeling data with fixed and random effects	242
7.2	A comparison with traditional analyses	259
7.2.1	Mixed-effects models and quasi- $F$	260
7.2.2	Mixed-effects models and Latin Square designs	266
7.2.3	Regression with subjects and items	269
7.3	Shrinkage in mixed-effects models	275
7.4	Generalized linear mixed models	278
7.5	Case studies	284
7.5.1	Primed lexical decision latencies for Dutch neologisms	284
7.5.2	Self-paced reading latencies for Dutch neologisms	287
7.5.3	Visual lexical decision latencies of Dutch eight-year-olds	289
7.5.4	Mixed-effects models in corpus linguistics	295

<b>Appendix A</b>	<b>Solutions to the exercises</b>	<b>303</b>
<b>Appendix B</b>	<b>Overview of R functions</b>	<b>335</b>
<i>References</i>		342
<i>Index</i>		347
	Index of data sets	347
	Index of R	347
	Index of topics	349
	Index of authors	352