

# Inhaltsverzeichnis

<b>1</b>	<b>Erste Schritte</b>	<b>1</b>
1.1	Vorstellung	1
1.1.1	Pro und Contra R	2
1.1.2	Typografische Konventionen	4
1.1.3	R installieren	4
1.1.4	Grafische Benutzeroberflächen	5
1.1.5	Weiterführende Informationsquellen und Literatur	6
1.2	Grundlegende Elemente	7
1.2.1	R Starten, beenden und die Konsole verwenden	7
1.2.2	Befehlssequenzen im Editor bearbeiten	8
1.2.3	Einstellungen vornehmen	9
1.2.4	Umgang mit dem Workspace	10
1.2.5	Einfache Arithmetik	10
1.2.6	Funktionen mit Argumenten aufrufen	11
1.2.7	Hilfe-Funktionen	12
1.2.8	Zusatzpakete verwenden	13
1.2.9	Empfehlungen und typische Fehlerquellen	14
1.3	Datenstrukturen: Klassen, Objekte, Datentypen	15
1.3.1	Objekte benennen	15
1.3.2	Zuweisungen an Objekte	15
1.3.3	Objekte ausgeben	16
1.3.4	Objekte anzeigen lassen und entfernen	16
1.3.5	Datentypen	17
1.3.6	Logische Werte, Operatoren und Verknüpfungen	18
<b>2</b>	<b>Daten importieren und exportieren</b>	<b>20</b>
2.1	Daten im Textformat lesen und schreiben	20
2.2	R-Objekte speichern und laden	22
2.3	Daten im Editor bearbeiten	23
2.4	Daten mit anderen Programmen austauschen	23
2.4.1	Programme zur Tabellenkalkulation	23
2.4.2	SPSS, Stata und SAS	24
2.4.3	Datenbanken	25
<b>3</b>	<b>Elementare Datenverarbeitung</b>	<b>27</b>
3.1	Vektoren	27
3.1.1	Vektoren erzeugen	27
3.1.2	Elemente auswählen und verändern	28
3.1.3	Datentypen in Vektoren	30
3.1.4	Elemente benennen	30

## Inhaltsverzeichnis

3.2	Logische Operatoren . . . . .	31
3.2.1	Logischer Vergleich von Vektoren . . . . .	31
3.2.2	Logische Indexvektoren . . . . .	33
3.3	Systematische und zufällige Wertefolgen erzeugen . . . . .	34
3.3.1	Numerische Sequenzen erstellen . . . . .	34
3.3.2	Wertefolgen wiederholen . . . . .	35
3.3.3	Zufällig aus einer Urne ziehen . . . . .	36
3.3.4	Zufallszahlen aus bestimmten Verteilungen erzeugen . . . . .	36
3.4	Daten transformieren . . . . .	37
3.4.1	Werte sortieren . . . . .	37
3.4.2	Werte in zufällige Reihenfolge bringen . . . . .	38
3.4.3	Teilmengen von Daten auswählen . . . . .	39
3.4.4	Daten umrechnen . . . . .	40
3.4.5	Neue aus bestehenden Variablen bilden . . . . .	42
3.4.6	Werte ersetzen oder recodieren . . . . .	43
3.4.7	Kontinuierliche Variablen in Kategorien einteilen . . . . .	44
3.5	Gruppierungsfaktoren . . . . .	45
3.5.1	Ungeordnete Faktoren . . . . .	45
3.5.2	Faktorstufen nachträglich ändern . . . . .	47
3.5.3	Geordnete Faktoren . . . . .	48
3.5.4	Reihenfolge von Faktorstufen . . . . .	48
3.5.5	Faktoren nach Muster erstellen . . . . .	49
3.5.6	Quantitative in kategoriale Variablen umwandeln . . . . .	50
3.6	Deskriptive Kennwerte numerischer Daten . . . . .	51
3.6.1	Summen, Differenzen und Produkte . . . . .	51
3.6.2	Extremwerte . . . . .	52
3.6.3	Mittelwert, Median und Modalwert . . . . .	52
3.6.4	Prozentrang, Quartile und Quantile . . . . .	53
3.6.5	Varianz, Streuung, Schiefe und Wölbung . . . . .	54
3.6.6	Kovarianz und Korrelation . . . . .	55
3.6.7	Kennwerte getrennt nach Gruppen berechnen . . . . .	55
3.7	Matrizen . . . . .	56
3.7.1	Dimensionierung, Zeilen und Spalten . . . . .	57
3.7.2	Elemente auswählen und verändern . . . . .	58
3.7.3	Matrizen verbinden . . . . .	59
3.7.4	Matrizen sortieren . . . . .	60
3.7.5	Randkennwerte berechnen . . . . .	61
3.7.6	Matrix zeilen- oder spaltenweise mit Kennwerten verrechnen . . . . .	62
3.7.7	Kovarianz- und Korrelationsmatrizen . . . . .	62
3.8	Häufigkeitsauszählungen . . . . .	64
3.8.1	Einfache Tabellen absoluter und relativer Häufigkeiten . . . . .	64
3.8.2	Absolute, relative und bedingte relative Häufigkeiten in Kreuztabellen . . . . .	65
3.8.3	Randkennwerte von Kreuztabellen . . . . .	67
3.9	Codierung, Identifikation und Behandlung fehlender Werte . . . . .	68
3.9.1	Fehlende Werte codieren und ihr Vorhandensein prüfen . . . . .	69
3.9.2	Fehlende Werte ersetzen und umcodieren . . . . .	70
3.9.3	Behandlung fehlender Werte bei der Berechnung einfacher Kennwerte . . . . .	70

## Inhaltsverzeichnis

3.9.4	Behandlung fehlender Werte in Matrizen . . . . .	71
3.9.5	Behandlung fehlender Werte beim Sortieren von Daten . . . . .	73
3.9.6	Behandlung fehlender Werte in inferenzstatistischen Tests . . . . .	73
3.10	Zeichenketten verarbeiten . . . . .	73
3.10.1	Zeichenketten erstellen und ausgeben . . . . .	74
3.10.2	Zeichenketten manipulieren . . . . .	74
3.10.3	Zeichenfolgen finden . . . . .	75
3.10.4	Zeichenfolgen ersetzen . . . . .	77
3.11	Datum und Uhrzeit . . . . .	78
3.11.1	Datumsangaben erstellen und formatieren . . . . .	78
3.11.2	Uhrzeit . . . . .	78
3.11.3	Berechnungen mit Datum und Uhrzeit . . . . .	79
<b>4</b>	<b>Datensätze</b>	<b>81</b>
4.1	Mit Datensätzen arbeiten . . . . .	81
4.1.1	Datentypen in Datensätzen . . . . .	82
4.1.2	Elemente auswählen und verändern . . . . .	83
4.1.3	Namen von Variablen und Beobachtungen . . . . .	84
4.1.4	Datensätze in den Suchpfad einfügen . . . . .	85
4.2	Datensätze transformieren . . . . .	86
4.2.1	Variablen hinzufügen und entfernen . . . . .	86
4.2.2	Datensätze sortieren . . . . .	87
4.2.3	Teilmengen von Daten auswählen . . . . .	88
4.2.4	Doppelte und fehlende Werte behandeln . . . . .	90
4.2.5	Datensätze teilen . . . . .	91
4.2.6	Datensätze zusammenfügen . . . . .	92
4.2.7	Organisationsform von Datensätzen ändern . . . . .	92
4.3	Daten aggregieren . . . . .	95
4.3.1	Funktionen auf Variablen anwenden . . . . .	95
4.3.2	Funktionen getrennt nach Gruppen anwenden . . . . .	97
4.4	Listen . . . . .	98
4.4.1	Komponenten auswählen und verändern . . . . .	98
4.4.2	Komponenten hinzufügen und entfernen . . . . .	99
<b>5</b>	<b>Hilfsmittel für die Inferenzstatistik</b>	<b>101</b>
5.1	Lineare Modelle formulieren . . . . .	101
5.2	Funktionen von Zufallsvariablen . . . . .	103
5.2.1	Dichtefunktionen . . . . .	103
5.2.2	Verteilungsfunktionen . . . . .	104
5.2.3	Quantilfunktionen . . . . .	104
5.3	Gemeinsamer Datensatz für alle Auswertungen . . . . .	105
<b>6</b>	<b>Lineare Regression</b>	<b>107</b>
6.1	Test auf Korrelation . . . . .	107
6.2	Einfache lineare Regression . . . . .	108
6.2.1	Deskriptive Modellanpassung . . . . .	108
6.2.2	Regressionsanalyse . . . . .	109

6.3	Multiple lineare Regression . . . . .	110
6.3.1	Deskriptive Modellanpassung und Regressionsanalyse . . . . .	111
6.3.2	Modelle vergleichen . . . . .	111
6.4	Regressionsmodelle auf andere Daten anwenden . . . . .	113
6.5	Kreuzvalidierung von Regressionsmodellen . . . . .	113
6.6	Regressionsdiagnostik . . . . .	114
6.6.1	Extremwerte, Ausreißer und Einfluss . . . . .	115
6.6.2	Verteilungseigenschaften der Residuen . . . . .	116
6.6.3	Multikollinearität . . . . .	117
6.7	Partialkorrelation und Semipartialkorrelation . . . . .	118
6.8	Logistische Regression . . . . .	119
6.8.1	Modellanpassung für dichotome Daten . . . . .	119
6.8.2	Anpassungsgüte . . . . .	120
6.8.3	Vorhersage, Klassifikation und Anwendung auf neue Daten . . . . .	121
6.8.4	Signifikanztests für Parameter und Modell . . . . .	122
<b>7</b>	<b>t-Tests und Varianzanalysen</b> . . . . .	<b>123</b>
7.1	Tests auf Normalverteilung . . . . .	123
7.2	Levene-Test auf Varianzhomogenität . . . . .	124
7.3	t-Tests . . . . .	125
7.3.1	t-Test für eine Stichprobe . . . . .	125
7.3.2	t-Test für zwei unabhängige Stichproben . . . . .	126
7.3.3	t-Test für zwei abhängige Stichproben . . . . .	127
7.4	Einfaktorielle Varianzanalyse . . . . .	127
7.4.1	Auswertung mit <code>av()</code> . . . . .	128
7.4.2	Auswertung mit <code>anova()</code> . . . . .	129
7.4.3	Grafische Prüfung der Voraussetzungen . . . . .	129
7.4.4	Paarvergleiche mit t-Tests und $\alpha$ -Adjustierung . . . . .	130
7.4.5	Simultane Konfidenzintervalle nach Tukey . . . . .	131
7.5	Einfaktorielle Varianzanalyse mit abhängigen Gruppen . . . . .	132
7.5.1	Univariat formulierte Auswertung . . . . .	132
7.5.2	Multivariat formulierte Auswertung . . . . .	133
7.6	Zweifaktorielle Varianzanalyse . . . . .	134
7.6.1	Auswertung . . . . .	135
7.6.2	Quadratsummen vom Typ I, II und III . . . . .	136
7.7	Zweifaktorielle Varianzanalyse mit zwei Intra-Gruppen Faktoren . . . . .	137
7.7.1	Univariat formulierte Auswertung . . . . .	137
7.7.2	Multivariat formulierte Auswertung . . . . .	138
7.8	Zweifaktorielle Varianzanalyse mit Split-Plot-Design . . . . .	139
7.8.1	Univariat formulierte Auswertung . . . . .	139
7.8.2	Multivariat formulierte Auswertung . . . . .	140
7.9	Kovarianzanalyse . . . . .	141
<b>8</b>	<b>Nonparametrische Methoden</b> . . . . .	<b>144</b>
8.1	Häufigkeiten kategorialer Variablen analysieren . . . . .	144
8.1.1	Binomialtest . . . . .	144
8.1.2	$\chi^2$ -Test auf eine feste Verteilung . . . . .	145

## Inhaltsverzeichnis

8.1.3	$\chi^2$ -Test auf Unabhängigkeit	146
8.1.4	$\chi^2$ -Test für mehrere Auftretenswahrscheinlichkeiten	147
8.1.5	Fishers exakter Test auf Unabhängigkeit	147
8.1.6	Kennwerte von $(2 \times 2)$ -Konfusionsmatrizen	148
8.1.7	ROC-Kurve und AUC	151
8.2	Maße für Zusammenhang und Übereinstimmung	152
8.2.1	Spearman's $\rho$ und Kendall's $\tau$	152
8.2.2	Zusammenhang kategorialer Variablen	153
8.2.3	Inter-Rater-Übereinstimmung	153
8.3	Tests auf Übereinstimmung von Verteilungen	155
8.3.1	Wilcoxon-Vorzeichen-Rang-Test	155
8.3.2	Wilcoxon-Rangsummen-Test / Mann-Whitney-U-Test	156
8.3.3	Kruskal-Wallis-H-Test	157
8.3.4	Friedman-Test	157
<b>9</b>	<b>Multivariate Verfahren</b>	<b>159</b>
9.1	Hauptkomponentenanalyse	159
9.2	Faktorenanalyse	160
9.3	Multivariate multiple Regression	163
9.4	Hotellings $T^2$	163
9.4.1	Test für eine Stichprobe	163
9.4.2	Test für zwei Stichproben	164
9.5	Multivariate Varianzanalyse	165
9.6	Diskriminanzanalyse	166
<b>10</b>	<b>Diagramme erstellen</b>	<b>168</b>
10.1	Grafik-Devices	168
10.1.1	Aufbau und Verwaltung von Grafik-Devices	168
10.1.2	Grafiken speichern	169
10.2	Streu- und Liniendiagramme	170
10.2.1	Streudiagramme mit <code>plot()</code>	170
10.2.2	Streudiagramme mit <code>matplot()</code>	171
10.3	Diagramme formatieren	172
10.3.1	Grafikelemente formatieren	172
10.3.2	Farben spezifizieren	174
10.3.3	Achsen formatieren	174
10.4	Säulen- und Punktdiagramme	175
10.4.1	Einfache Säulendiagramme	175
10.4.2	Gruppierte und gestapelte Säulendiagramme	176
10.4.3	Dotchart	177
10.5	Elemente einem bestehenden Diagramm hinzufügen	178
10.5.1	Punkte	179
10.5.2	Linien	179
10.5.3	Polygone	181
10.5.4	Funktionsgraphen	181
10.5.5	Text	182
10.5.6	Achsen	183

10.5.7 Fehlerbalken . . . . .	184
10.6 Verteilungsdiagramme . . . . .	185
10.6.1 Histogramm und Schätzung der Dichtefunktion . . . . .	185
10.6.2 Boxplot . . . . .	186
10.6.3 Stripchart . . . . .	187
10.6.4 Quantil-Quantil-Diagramm . . . . .	188
10.6.5 Kreisdiagramm . . . . .	189
10.6.6 Gemeinsame Verteilung zweier Variablen . . . . .	191
10.7 Multivariate Daten visualisieren . . . . .	192
10.7.1 Höhenlinien . . . . .	193
10.7.2 Dreidimensionale Gitter und Streudiagramme . . . . .	194
10.7.3 Bedingte Diagramme für mehrere Gruppen . . . . .	195
10.7.4 Matrix aus Streudiagrammen . . . . .	196
10.8 Mehrere Diagramme in einem Grafik-Device darstellen . . . . .	197
<b>11 R als Programmiersprache</b>	<b>200</b>
11.1 Kontrollstrukturen . . . . .	200
11.1.1 Fallunterscheidungen . . . . .	200
11.1.2 Schleifen . . . . .	201
11.2 Eigene Funktionen erstellen . . . . .	202
11.2.1 Funktionskopf und Funktionsrumpf . . . . .	203
11.2.2 Rückgabewert . . . . .	203
11.2.3 Generische Funktionen . . . . .	204
11.2.4 Funktionen analysieren . . . . .	205
11.2.5 Effizienz von Auswertungen steigern . . . . .	205
<b>Literaturverzeichnis</b>	<b>207</b>