
Inhaltsverzeichnis

1	Erste Schritte	1
1.1	Vorstellung	1
1.1.1	Pro und Contra R	3
1.1.2	Typografische Konventionen	5
1.1.3	R installieren	5
1.1.4	Grafische Benutzeroberflächen	6
1.1.5	Weiterführende Informationsquellen und Literatur	8
1.2	Grundlegende Elemente	8
1.2.1	R Starten, beenden und die Konsole verwenden	8
1.2.2	Befehlssequenzen im Editor bearbeiten	10
1.2.3	Einstellungen vornehmen	10
1.2.4	Umgang mit dem Workspace	11
1.2.5	Einfache Arithmetik	12
1.2.6	Funktionen mit Argumenten aufrufen	13
1.2.7	Hilfe-Funktionen	14
1.2.8	Zusatzpakete verwenden	15
1.2.9	Empfehlungen und typische Fehlerquellen	16
1.3	Datenstrukturen: Klassen, Objekte, Datentypen	17
1.3.1	Objekte benennen	18
1.3.2	Zuweisungen an Objekte	18
1.3.3	Objekte ausgeben	19
1.3.4	Objekte anzeigen lassen und entfernen	19
1.3.5	Datentypen	20
1.3.6	Logische Werte, Operatoren und Verknüpfungen	21
2	Daten importieren und exportieren	23
2.1	Daten im Textformat lesen und schreiben	23
2.2	R-Objekte speichern und laden	25
2.3	Daten im Editor bearbeiten	26

2.4	Daten mit anderen Programmen austauschen	26
2.4.1	Programme zur Tabellenkalkulation	27
2.4.2	SPSS, Stata und SAS	28
2.4.3	Datenbanken	29
3	Elementare Datenverarbeitung	33
3.1	Vektoren	33
3.1.1	Vektoren erzeugen	33
3.1.2	Elemente auswählen und verändern	35
3.1.3	Datentypen in Vektoren	37
3.1.4	Elemente benennen	37
3.2	Logische Operatoren	38
3.2.1	Logischer Vergleich von Vektoren	38
3.2.2	Logische Indexvektoren	40
3.3	Systematische und zufällige Wertefolgen erzeugen	41
3.3.1	Numerische Sequenzen erstellen	42
3.3.2	Wertefolgen wiederholen	43
3.3.3	Zufällig aus einer Urne ziehen	44
3.3.4	Zufallszahlen aus bestimmten Verteilungen erzeugen	44
3.4	Daten transformieren	45
3.4.1	Werte sortieren	46
3.4.2	Werte in zufällige Reihenfolge bringen	47
3.4.3	Teilmengen von Daten auswählen	47
3.4.4	Daten umrechnen	49
3.4.5	Neue aus bestehenden Variablen bilden	51
3.4.6	Werte ersetzen oder recodieren	52
3.4.7	Kontinuierliche Variablen in Kategorien einteilen	54
3.5	Gruppierungsfaktoren	55
3.5.1	Ungeordnete Faktoren	55
3.5.2	Faktorstufen nachträglich ändern	58
3.5.3	Geordnete Faktoren	59
3.5.4	Reihenfolge von Faktorstufen	59
3.5.5	Faktoren nach Muster erstellen	60
3.5.6	Quantitative in kategoriale Variablen umwandeln	61
3.6	Deskriptive Kennwerte numerischer Daten	62
3.6.1	Summen, Differenzen und Produkte	63
3.6.2	Extremwerte	64
3.6.3	Mittelwert, Median und Modalwert	64
3.6.4	Prozentrang, Quartile und Quantile	65
3.6.5	Varianz, Streuung, Schiefe und Wölbung	66
3.6.6	Kovarianz und Korrelation	67
3.6.7	Kennwerte getrennt nach Gruppen berechnen	68

3.7	Matrizen	68
3.7.1	Dimensionierung, Zeilen und Spalten	69
3.7.2	Elemente auswählen und verändern	70
3.7.3	Matrizen verbinden	72
3.7.4	Matrizen sortieren	73
3.7.5	Randkennwerte berechnen	74
3.7.6	Matrix zeilen- oder spaltenweise mit Kennwerten verrechnen . . .	75
3.7.7	Kovarianz- und Korrelationsmatrizen	76
3.8	Häufigkeitsauszählungen	78
3.8.1	Einfache Tabellen absoluter und relativer Häufigkeiten	78
3.8.2	Absolute, relative und bedingte relative Häufigkeiten in Kreuztabellen	79
3.8.3	Randkennwerte von Kreuztabellen	82
3.9	Codierung, Identifikation und Behandlung fehlender Werte	83
3.9.1	Fehlende Werte codieren und ihr Vorhandensein prüfen	83
3.9.2	Fehlende Werte ersetzen und umcodieren	85
3.9.3	Behandlung fehlender Werte bei der Berechnung einfacher Kennwerte	85
3.9.4	Behandlung fehlender Werte in Matrizen	86
3.9.5	Behandlung fehlender Werte beim Sortieren von Daten	88
3.9.6	Behandlung fehlender Werte in inferenzstatistischen Tests	89
3.10	Zeichenketten verarbeiten	89
3.10.1	Zeichenketten erstellen und ausgeben	89
3.10.2	Zeichenketten manipulieren	90
3.10.3	Zeichenfolgen finden	91
3.10.4	Zeichenfolgen ersetzen	93
3.11	Datum und Uhrzeit	94
3.11.1	Datumsangaben erstellen und formatieren	94
3.11.2	Uhrzeit	95
3.11.3	Berechnungen mit Datum und Uhrzeit	96
4	Datensätze	99
4.1	Mit Datensätzen arbeiten	99
4.1.1	Datentypen in Datensätzen	101
4.1.2	Elemente auswählen und verändern	102
4.1.3	Namen von Variablen und Beobachtungen	103
4.1.4	Datensätze in den Suchpfad einfügen	104
4.2	Datensätze transformieren	106
4.2.1	Variablen hinzufügen und entfernen	106
4.2.2	Datensätze sortieren	107
4.2.3	Teilmengen von Daten auswählen	108
4.2.4	Doppelte und fehlende Werte behandeln	110

4.2.5	Datensätze teilen	111
4.2.6	Datensätze zusammenfügen	112
4.2.7	Organisationsform von Datensätzen ändern	113
4.3	Daten aggregieren	116
4.3.1	Funktionen auf Variablen anwenden	117
4.3.2	Funktionen getrennt nach Gruppen anwenden	118
4.4	Listen	119
4.4.1	Komponenten auswählen und verändern	120
4.4.2	Komponenten hinzufügen und entfernen	121
5	Hilfsmittel für die Inferenzstatistik	123
5.1	Lineare Modelle formulieren	123
5.2	Funktionen von Zufallsvariablen	125
5.2.1	Dichtefunktionen	125
5.2.2	Verteilungsfunktionen	126
5.2.3	Quantilfunktionen	127
5.3	Gemeinsamer Datensatz für alle Auswertungen	128
6	Lineare Regression	129
6.1	Test des Korrelationskoeffizienten	129
6.2	Einfache lineare Regression	130
6.2.1	Deskriptive Modellanpassung	130
6.2.2	Regressionsanalyse	131
6.3	Multiple lineare Regression	133
6.3.1	Deskriptive Modellanpassung und Regressionsanalyse	134
6.3.2	Modelle vergleichen	134
6.4	Erweiterungen der linearen Regression	136
6.5	Regressionsmodelle auf andere Daten anwenden	137
6.6	Kreuzvalidierung von Regressionsmodellen	138
6.7	Regressionsdiagnostik	139
6.7.1	Extremwerte, Ausreißer und Einfluss	139
6.7.2	Verteilungseigenschaften der Residuen	142
6.7.3	Multikollinearität	142
6.8	Partialkorrelation und Semipartialkorrelation	144
6.9	Logistische Regression	145
6.9.1	Modellanpassung für dichotome Daten	145
6.9.2	Anpassungsgüte	146
6.9.3	Vorhersage, Klassifikation und Anwendung auf neue Daten	147
6.9.4	Signifikanztests für Parameter und Modell	148
7	t-Tests und Varianzanalysen	151
7.1	Tests auf Normalverteilung	151
7.2	Levene-Test auf Varianzhomogenität	152

7.3	<i>t</i> -Tests	153
7.3.1		153
7.3.2	<i>t</i> -Test für zwei unabhängige Stichproben	154
7.3.3	<i>t</i> -Test für zwei abhängige Stichproben	155
7.4	Einfaktorielle Varianzanalyse	156
7.4.1	Auswertung mit <code>aov()</code>	156
7.4.2	Auswertung mit <code>anova()</code>	158
7.4.3	Grafische Prüfung der Voraussetzungen	158
7.4.4	Paarvergleiche mit <i>t</i> -Tests und α -Adjustierung	159
7.4.5	Simultane Konfidenzintervalle nach Tukey	160
7.5	Einfaktorielle Varianzanalyse mit abhängigen Gruppen	161
7.5.1	Univariat formulierte Auswertung	161
7.5.2	Multivariat formulierte Auswertung	163
7.6	Zweifaktorielle Varianzanalyse	164
7.6.1	Auswertung	164
7.6.2	Quadratsummen vom Typ I, II und III	166
7.7	Zweifaktorielle Varianzanalyse mit zwei Intra-Gruppen Faktoren	167
7.7.1	Univariat formulierte Auswertung	168
7.7.2	Multivariat formulierte Auswertung	169
7.8	Zweifaktorielle Varianzanalyse mit Split-Plot-Design	170
7.8.1	Univariat formulierte Auswertung	170
7.8.2	Multivariat formulierte Auswertung	171
7.9	Kovarianzanalyse	172
8	Nonparametrische Methoden	175
8.1	Häufigkeiten kategorialer Variablen analysieren	175
8.1.1	Binomialtest	175
8.1.2	χ^2 -Test auf eine feste Verteilung	176
8.1.3	χ^2 -Test auf Unabhängigkeit	177
8.1.4	χ^2 -Test für mehrere Auftretenswahrscheinlichkeiten	178
8.1.5	Fishers exakter Test auf Unabhängigkeit	179
8.1.6	Kennwerte von (2×2) -Konfusionsmatrizen	180
8.1.7	ROC-Kurve und AUC	183
8.2	Maße für Zusammenhang und Übereinstimmung	184
8.2.1	Spearman's ρ und Kendall's τ	184
8.2.2	Zusammenhang kategorialer Variablen	185
8.2.3	Inter-Rater-Übereinstimmung	186
8.3	Tests auf Übereinstimmung von Verteilungen	188
8.3.1	Wilcoxon-Vorzeichen-Rang-Test	188
8.3.2	Wilcoxon-Rangsummen-Test / Mann-Whitney- <i>U</i> -Test	189
8.3.3	Kruskal-Wallis- <i>H</i> -Test	190
8.3.4	Friedman-Test	191

9	Multivariate Verfahren	193
9.1	Hauptkomponentenanalyse	193
9.2	Faktorenanalyse	195
9.3	Multivariate multiple Regression	197
9.4	Hotellings T^2	198
9.4.1	Test für eine Stichprobe	198
9.4.2	Test für zwei Stichproben	199
9.5	Multivariate Varianzanalyse	200
10	Diagramme erstellen	203
10.1	Grafik-Devices	203
10.1.1	Aufbau und Verwaltung von Grafik-Devices	203
10.1.2	Grafiken speichern	204
10.2	Streu- und Liniendiagramme	205
10.2.1	Streudiagramme mit <code>plot()</code>	205
10.2.2	Streudiagramme mit <code>matplot()</code>	207
10.3	Diagramme formatieren	208
10.3.1	Grafikelemente formatieren	209
10.3.2	Farben spezifizieren	210
10.3.3	Achsen formatieren	210
10.4	Säulen- und Punktdiagramme	211
10.4.1	Einfache Säulendiagramme	211
10.4.2	Gruppierte und gestapelte Säulendiagramme	212
10.4.3	Dotchart	214
10.5	Elemente einem bestehenden Diagramm hinzufügen	215
10.5.1	Punkte	215
10.5.2	Linien	216
10.5.3	Polygone	218
10.5.4	Funktionsgraphen	219
10.5.5	Text	220
10.5.6	Achsen	221
10.5.7	Fehlerbalken	221
10.6	Verteilungsdiagramme	223
10.6.1	Histogramm und Schätzung der Dichtefunktion	223
10.6.2	Boxplot	225
10.6.3	Stripchart	226
10.6.4	Quantil-Quantil-Diagramm	228
10.6.5	Kreisdiagramm	229
10.6.6	Gemeinsame Verteilung zweier Variablen	230
10.7	Multivariate Daten visualisieren	232
10.7.1	Höhenlinien	233
10.7.2	Dreidimensionale Gitter und Streudiagramme	234

10.7.3 Bedingte Diagramme für mehrere Gruppen mit ggplot2	236
10.7.4 Matrix aus Streudiagrammen	242
10.8 Mehrere Diagramme in einem Grafik-Device darstellen	244
11 R als Programmiersprache	247
11.1 Kontrollstrukturen	247
11.1.1 Fallunterscheidungen	247
11.1.2 Schleifen	248
11.2 Eigene Funktionen erstellen	250
11.2.1 Funktionskopf und Funktionsrumpf	251
11.2.2 Rückgabewert	251
11.2.3 Generische Funktionen	252
11.2.4 Funktionen analysieren	252
11.2.5 Effizienz von Auswertungen steigern	254
Literatur	257
Sachverzeichnis	263
R-Funktionen, Klassen und Schlüsselwörter	271
Zusatzpakete	277