

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	7
1.2	Grundlegende Elemente	8
1.2.1	R Starten, beenden und die Konsole verwenden	8
1.2.2	Befehlssequenzen im Editor bearbeiten	9
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	Hinweise zum Arbeiten mit Zusatzpaketen	16
1.2.10	Empfehlungen und typische Fehlerquellen	16
1.3	Datenstrukturen: Klassen, Objekte, Datentypen	18
1.3.1	Objekte benennen	18
1.3.2	Zuweisungen an Objekte	19
1.3.3	Objekte ausgeben	19
1.3.4	Objekte anzeigen lassen und entfernen	20
1.3.5	Datentypen	20
1.3.6	Logische Werte, Operatoren und Verknüpfungen	21
2	Daten importieren und exportieren	25
2.1	Daten im Textformat lesen und schreiben	25
2.2	R-Objekte speichern und laden	27
2.3	Daten mit anderen Programmen austauschen	28

2.3.1	Programme zur Tabellenkalkulation	28
2.3.2	SPSS, Stata und SAS	29
2.3.3	Datenbanken	31
3	Elementare Datenverarbeitung	35
3.1	Vektoren	35
3.1.1	Vektoren erzeugen	35
3.1.2	Elemente auswählen und verändern	36
3.1.3	Datentypen in Vektoren	38
3.1.4	Elemente benennen	39
3.2	Logische Operatoren	39
3.2.1	Logischer Vergleich von Vektoren	39
3.2.2	Logische Indexvektoren	41
3.3	Systematische und zufällige Wertefolgen erzeugen	42
3.3.1	Numerische Sequenzen erstellen	43
3.3.2	Wertefolgen wiederholen	44
3.3.3	Zufällig aus einer Urne ziehen	44
3.3.4	Zufallszahlen aus bestimmten Verteilungen erzeugen	45
3.4	Daten transformieren	46
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	53
3.5	Gruppierungsfaktoren	54
3.5.1	Ungeordnete Faktoren	54
3.5.2	Faktorstufen nachträglich ändern	56
3.5.3	Geordnete Faktoren	58
3.5.4	Reihenfolge von Faktorstufen	58
3.5.5	Faktoren nach Muster erstellen	59
3.5.6	Quantitative in kategoriale Variablen umwandeln	60
3.6	Deskriptive Kennwerte numerischer Daten	61
3.6.1	Summen, Differenzen und Produkte	61
3.6.2	Extremwerte	62
3.6.3	Mittelwert, Median und Modalwert	63
3.6.4	Prozentrang, Quartile und Quantile	63
3.6.5	Varianz, Streuung, Schiefe und Wölbung	64
3.6.6	Kovarianz und Korrelation	65
3.6.7	Kennwerte getrennt nach Gruppen berechnen	65

3.7	Matrizen	66
3.7.1	Dimensionierung, Zeilen und Spalten	67
3.7.2	Elemente auswählen und verändern	68
3.7.3	Matrizen verbinden	69
3.7.4	Matrizen sortieren	70
3.7.5	Randkennwerte berechnen	71
3.7.6	Kovarianz- und Korrelationsmatrizen	72
3.8	Datensätze	73
3.8.1	Datentypen in Datensätzen	75
3.8.2	Elemente auswählen und verändern	75
3.8.3	Namen von Variablen und Beobachtungen	76
3.8.4	Datensätze in den Suchpfad einfügen	77
3.9	Listen	78
3.9.1	Komponenten auswählen und verändern	79
3.9.2	Komponenten hinzufügen und entfernen	80
3.10	Häufigkeiten bestimmen	81
3.10.1	Einfache Tabellen absoluter und relativer Häufigkeiten	81
3.10.2	Absolute, relative und bedingte relative Häufigkeiten in Kreuztabellen	82
3.10.3	Randkennwerte von Kreuztabellen	84
3.11	Codierung, Identifikation und Behandlung fehlender Werte	85
3.11.1	Fehlende Werte codieren und ihr Vorhandensein prüfen	85
3.11.2	Fehlende Werte ersetzen und umcodieren	86
3.11.3	Behandlung fehlender Werte bei der Berechnung einfacher Kennwerte	87
3.11.4	Behandlung fehlender Werte in Matrizen	88
3.11.5	Behandlung fehlender Werte beim Sortieren von Daten	90
3.11.6	Behandlung fehlender Werte in inferenzstatistischen Tests	90
3.12	Zeichenketten verarbeiten	91
3.12.1	Zeichenketten erstellen und ausgeben	91
3.12.2	Zeichenketten manipulieren	92
3.12.3	Zeichenfolgen finden	93
3.12.4	Zeichenfolgen ersetzen	94
3.13	Datum und Uhrzeit	95
3.13.1	Datumsangaben erstellen und formatieren	95
3.13.2	Uhrzeit	96
3.13.3	Berechnungen mit Datum und Uhrzeit	97

4	Datensätze aufbereiten und aggregieren mit dplyr	99
4.1	Besonderheiten	100
4.2	Variablen umbenennen	101
4.3	Teilmengen von Daten auswählen	102
4.3.1	Variablen auswählen	102
4.3.2	Beobachtungen auswählen	104
4.4	Variablen entfernen, hinzufügen und transformieren	105
4.5	Doppelte und fehlende Werte behandeln	106
4.6	Datensätze sortieren	107
4.7	Datensätze zeilen- oder spaltenweise verbinden	108
4.8	Datensätze zusammenführen	109
4.9	Organisationsform von Datensätzen ändern	111
4.10	Datensätze getrennt nach Gruppen auswerten und aggregieren	114
4.11	Häufigkeiten bestimmen	117
5	R-Dokumente und Notebooks erstellen	119
5.1	Grundprinzip	119
5.2	Arbeitsschritte	120
5.3	Aufbau eines Quelldokuments	121
5.4	Beispiel	122
6	Hilfsmittel für die Inferenzstatistik	127
6.1	Lineare Modelle formulieren	127
6.2	Funktionen von Zufallsvariablen	129
6.2.1	Dichtefunktionen	130
6.2.2	Verteilungsfunktionen	130
6.2.3	Quantilfunktionen	131
6.3	Gemeinsamer Datensatz für alle Auswertungen	131
7	Lineare Regression	133
7.1	Test des Korrelationskoeffizienten	133
7.2	Einfache lineare Regression	134
7.2.1	Deskriptive Modellanpassung	134
7.2.2	Regressionsanalyse	135
7.3	Multiple lineare Regression	137
7.3.1	Deskriptive Modellanpassung und Regressionsanalyse	137
7.3.2	Modelle vergleichen	138
7.4	Erweiterungen der linearen Regression	140
7.5	Regressionsmodelle auf andere Daten anwenden	141
7.6	Kreuzvalidierung von Regressionsmodellen	142
7.7	Regressionsdiagnostik	143

7.7.1	Extremwerte, Ausreißer und Einfluss.	143
7.7.2	Verteilungseigenschaften der Residuen	144
7.7.3	Multikollinearität	146
7.8	Partialkorrelation und Semipartialkorrelation	147
7.9	Logistische Regression	147
7.9.1	Modellanpassung für dichotome Daten	148
7.9.2	Anpassungsgüte	149
7.9.3	Vorhersage, Klassifikation und Anwendung auf neue Daten	149
7.9.4	Signifikanztests für Parameter und Modell	150
8	t-Tests und Varianzanalysen	153
8.1	Tests auf Normalverteilung	153
8.2	Levene-Test auf Varianzhomogenität.	154
8.3	t-Tests.	155
8.3.1	t-Test für eine Stichprobe.	155
8.3.2	t-Test für zwei unabhängige Stichproben	156
8.3.3	t-Test für zwei abhängige Stichproben	157
8.4	Einfaktorielle Varianzanalyse.	157
8.4.1	Auswertung mit aov ()	158
8.4.2	Auswertung mit anova ()	159
8.4.3	Grafische Prüfung der Voraussetzungen.	159
8.4.4	Paarvergleiche mit t-Tests und α -Adjustierung	160
8.4.5	Simultane Konfidenzintervalle nach Tukey	161
8.5	Einfaktorielle Varianzanalyse mit abhängigen Gruppen	162
8.5.1	Univariat formulierte Auswertung	162
8.5.2	Multivariat formulierte Auswertung.	163
8.6	Zweifaktorielle Varianzanalyse	165
8.6.1	Auswertung	165
8.6.2	Quadratsummen vom Typ I, II und III	166
8.7	Zweifaktorielle Varianzanalyse mit zwei Intra-Gruppen Faktoren	167
8.7.1	Univariat formulierte Auswertung	168
8.7.2	Multivariat formulierte Auswertung.	169
8.8	Zweifaktorielle Varianzanalyse mit Split-Plot-Design.	170
8.8.1	Univariat formulierte Auswertung	170
8.8.2	Multivariat formulierte Auswertung.	171
8.9	Kovarianzanalyse	171
9	Nonparametrische Methoden	175
9.1	Häufigkeiten kategorialer Variablen analysieren	175
9.1.1	Binomialtest	175
9.1.2	χ^2 -Test auf eine feste Verteilung	176

9.1.3	χ^2 -Test auf Unabhängigkeit	177
9.1.4	χ^2 -Test für mehrere Auftretenswahrscheinlichkeiten.	178
9.1.5	Fishers exakter Test auf Unabhängigkeit	179
9.1.6	Kennwerte von (2×2) -Konfusionsmatrizen	180
9.1.7	ROC-Kurve und AUC	182
9.2	Maße für Zusammenhang und Übereinstimmung	184
9.2.1	Zusammenhang stetiger ordinaler Variablen: Spearman's ρ und Kendalls τ	184
9.2.2	Zusammenhang kategorialer Variablen: ϕ , Cramér's V , Kontingenzkoeffizient	185
9.2.3	Inter-Rater-Übereinstimmung	185
9.3	Tests auf Übereinstimmung von Verteilungen	187
9.3.1	Wilcoxon-Vorzeichen-Rang-Test	187
9.3.2	Wilcoxon-Rangsummen-Test / Mann- Whitney- U -Test	188
9.3.3	Kruskal-Wallis- H -Test	189
9.3.4	Friedman-Test	189
10	Multivariate Verfahren	191
10.1	Hauptkomponentenanalyse	191
10.2	Faktorenanalyse	193
10.3	Multivariate multiple Regression	195
10.4	Hotellings T^2	196
10.4.1	Test für eine Stichprobe	196
10.4.2	Test für zwei Stichproben	197
10.5	Multivariate Varianzanalyse	198
11	Diagramme mit ggplot2 erstellen	199
11.1	Grundprinzip	199
11.1.1	Grundsicht	199
11.1.2	Diagramme speichern	201
11.2	Diagrammtypen	201
11.2.1	Punkt-, Streu- und Liniendiagramme	202
11.2.2	Säulendiagramm	202
11.2.3	Histogramm	204
11.2.4	Boxplot	205
11.2.5	Quantil-Quantil-Diagramm	206
11.3	Bedingte Diagramme in Panels darstellen	207
11.4	Diagrammelemente hinzufügen	209
11.5	Diagramme formatieren	211
11.5.1	Elementposition kontrollieren	211
11.5.2	Achsen anpassen	213

11.5.3	Legende ändern	215
11.5.4	Farben, Datenpunktsymbole und Linientypen	217
11.5.5	Aussehen im Detail verändern	219
12	R als Programmiersprache	221
12.1	Kontrollstrukturen	221
12.1.1	Fallunterscheidungen	221
12.1.2	Schleifen.	222
12.2	Eigene Funktionen erstellen.	224
12.2.1	Funktionskopf und Funktionsrumpf.	225
12.2.2	Rückgabewert.	225
12.2.3	Generische Funktionen	225
12.2.4	Funktionen analysieren	226
12.2.5	Effizienz von Auswertungen steigern.	227
	Literatur.	229
	Stichwortverzeichnis.	235
	R-Funktionen, Klassen und Schlüsselwörter	243
	Zusatzpakete	249