

Inhaltsverzeichnis

Vorwort	11
Fragestellungen und Methoden	13
Kapitel 1 Einführung	17
1.1 Konzeption des Buchs	18
1.2 Aufbau des Buchs	19
1.3 Programmversionen von R	20
1.4 Wie kann dieses Buch verwendet werden?	20
1.5 Typografische und andere Konventionen in diesem Buch	21
Kapitel 2 Statistische Grundbegriffe	25
2.1 Einige Beispiele	27
2.1.1 Hochrechnung (statistisches Schätzen)	27
2.1.2 Prüfen von Fragestellungen (Testen von Hypothesen)	28
2.1.3 Erstellen von Modellen (statistisches Modellieren)	30
2.2 Grundlegende Konzepte	31
2.3 Messung und Typen von Daten	33
2.4 Arten von Fragestellungen und Variablen	36
2.5 Zusammenfassung der Konzepte	40
2.6 Übungen	40
Teil I Einführung in R	43
Kapitel 3 Erste Schritte	45
3.1 Download und Installation von R	46
3.1.1 Download	46
3.1.2 Installation	48
3.1.3 Aufrufen und Beenden von R	49
3.1.4 Installation von Ergänzungen (Contributed Packages)	49
3.2 Aller Anfang ist leicht	52
3.3 R-Befehle im Überblick	68
3.4 Übungen	69
Kapitel 4 Daten in R – vom Fragebogen zum fertigen Datensatz	71
4.1 Fragebogen und Kodierung	72
4.2 Erfassen der kodierten Daten	75

4.2.1	Eingabe der Daten	75
4.2.2	Abspeichern und Wiedereinlesen der Daten	79
4.3	Organisation eines Datensatzes – Data Frames	80
4.3.1	Das Ansprechen einzelner Variablen eines Data Frame	80
4.3.2	Faktoren	82
4.3.3	Auswählen von Beobachtungseinheiten (Fällen)	85
4.3.4	Transformieren der Daten bzw. Erzeugen von neuen Variablen	90
4.3.5	Berechnen neuer Variablen	90
4.3.6	Umkodieren von Variablen	93
4.3.7	Modifikation eines Data Frame	99
4.3.8	Datenkontrolle	102
4.4	R-Befehle im Überblick	105
4.5	Übungen	107
Kapitel 5	Mehr R	109
5.1	Die R-Arbeitsumgebung	110
5.1.1	Die R-Benutzeroberfläche	110
5.1.2	Der Workspace	112
5.1.3	Working Directory – das Arbeitsverzeichnis	113
5.2	R-Grafik	114
5.2.1	High-level Plotting Functions	115
5.2.2	Hinzufügen von Grafikelementen (Low-level Plotting Functions)	118
5.2.3	Spezielle Einstellungen (Graphical Parameters)	121
5.3	Weiterverwenden des R-Outputs	124
5.4	Einlesen von R-Befehlen	125
5.4.1	Der R-Editor	125
5.4.2	Einlesen von R-Skripts	127
5.4.3	Direktes Kopieren von R-Code – Einfügen über die Zwischenablage	129
5.5	Einlesen und Schreiben externer Dateien	129
5.5.1	Daten aus Excel bzw. OpenOffice.org Calc	129
5.5.2	Dateien aus anderen Statistikpaketen (z. B. SPSS)	132
5.5.3	Direktes Kopieren – Einfügen über die Zwischenablage	133
5.5.4	Schreiben von Dateien	134
5.6	Das R-Hilfesystem und weiterführende Information	135
5.6.1	Hilfe zu einzelnen Funktionen und Packages	136
5.6.2	Dokumente, Webseiten und weiterführende Information	140
5.7	R-Befehle im Überblick	142
5.8	Übungen	144

Teil II	Kategoriale Daten	145
Kapitel 6	Eine kategoriale Variable	147
6.1	Einleitung	148
6.2	Kommen alle Kategorien gleich häufig vor?	152
6.2.1	Numerische Beschreibung	154
6.2.2	Grafische Beschreibung	156
6.2.3	Statistische Analyse der Problemstellung	159
6.3	Entsprechen Häufigkeiten bestimmten Vorgaben?	166
6.3.1	Numerische und grafische Beschreibung	167
6.3.2	Statistische Analyse der Problemstellung	170
6.4	Hat ein Prozentsatz (Anteil) einen bestimmten Wert?	172
6.4.1	Statistische Analyse der Problemstellung	175
6.5	In welchem Bereich kann man einen Prozentsatz (Anteil) erwarten?	180
6.6	R-Befehle im Überblick	186
6.7	Zusammenfassung der Konzepte	188
6.8	Übungen	188
6.9	Vertiefung: Die Chi-Quadrat-Verteilung oder wie entsteht ein p-Wert? ...	190
Kapitel 7	Mehrere kategoriale Variablen	193
7.1	Beschreibung mehrerer kategorialer Variablen	194
7.1.1	Numerische Beschreibung	195
7.1.2	Grafische Beschreibung	198
7.2	Ist die Verteilung von Häufigkeiten in verschiedenen Gruppen gleich?...	202
7.3	Unterscheiden sich Anteile in zwei oder mehreren Gruppen?	206
7.3.1	Eingabe einer Tabelle	207
7.3.2	Vergleich der Anteile	208
7.3.3	Exakter Test nach Fisher	209
7.4	Sind zwei kategoriale Variablen unabhängig?	211
7.4.1	Datenaufbereitung	211
7.4.2	Unabhängigkeitstest	212
7.5	Unterscheidet sich das Risiko in zwei Gruppen?	215
7.5.1	Odds-Ratio	216
7.5.2	Odds-Ratio-Test	218
7.6	Wie kann man Veränderungen von Anteilen testen?	220
7.6.1	Unabhängige und abhängige Stichproben	220
7.6.2	McNemar-Test	221
7.7	R-Befehle im Überblick	223
7.8	Zusammenfassung der Konzepte	224
7.9	Übungen	224

Teil III	Metrische Daten	227
Kapitel 8	Eine metrische Variable	229
8.1	Wie kann man die Verteilung einer metrischen Variablen beschreiben? ..	230
8.1.1	Klassifizieren, Tabellen und Histogramme	231
8.1.2	Maßzahlen zur Beschreibung der Verteilung	236
8.1.3	Boxplot	243
8.1.4	Ausreißer	246
8.1.5	Weitere grafische Beschreibungsmethoden	247
8.2	Ist der Mittelwert der Grundgesamtheit anders als eine bestimmte Vorgabe?	251
8.3	Bereich für den Mittelwert in einer Grundgesamtheit	256
8.4	Folgt eine metrische Variable einer bestimmten Verteilung?	257
8.4.1	Q-Q-Plot	259
8.4.2	Kolmogorov-Smirnov-Test und Shapiro-Wilk-Test	262
8.4.3	Anpassungstest mit der χ^2 -Verteilung	265
8.5	R-Befehle im Überblick	268
8.6	Zusammenfassung der Konzepte	270
8.7	Übungen	270
Kapitel 9	Mehrere metrische Variablen	273
9.1	Wie stark ist der Zusammenhang zwischen zwei metrischen Variablen? ..	275
9.1.1	Grafische Beschreibung	276
9.1.2	Korrelationskoeffizient nach Pearson	276
9.1.3	Korrelationskoeffizient nach Spearman	280
9.2	Welche Form hat der Zusammenhang zwischen zwei Variablen?	281
9.2.1	Lineares Regressionsmodell	281
9.2.2	Rechenergebnisse	284
9.3	Vorhersage von Werten	287
9.3.1	Punktprognosen	287
9.3.2	Intervallprognosen	289
9.4	Zusammenhang einer mit mehreren Variablen	291
9.4.1	Multiple lineare Regression	292
9.4.2	Kategoriale als erklärende Variablen	295
9.4.3	Modellselektion	298
9.4.4	Modelldiagnostik	300
9.4.5	Prognose	304
9.5	Unterscheiden sich Mittelwerte zu zwei oder mehreren Zeitpunkten? ...	306
9.5.1	Grafische Beschreibung	307
9.5.2	Analyse der Fragestellung	308
9.6	Wie kann man den zeitlichen Verlauf einer Variablen beschreiben?	309
9.6.1	Zeitreihen	309
9.6.2	Zeitreihenzerlegung	311

9.6.3	Trend	312
9.6.4	Saison	319
9.6.5	Zusammenfassung der Zeitreihenzerlegung	321
9.6.6	Prognose	322
9.6.7	Autokorrelation	324
9.7	R-Befehle im Überblick	328
9.8	Zusammenfassung der Konzepte	330
9.9	Übungen	330

Teil IV Metrische und kategoriale Daten 333

Kapitel 10 Metrische und kategoriale Variablen 335

10.1	Unterscheiden sich die Mittelwerte in zwei Gruppen?	336
10.1.1	Grafische und numerische Beschreibung	337
10.1.2	Analyse der Fragestellung	338
10.2	Unterscheidet sich die Lage einer Variablen zwischen zwei Gruppen? ...	340
10.2.1	Beschreibung der Stichprobe	340
10.2.2	Analyse der Fragestellung	341
10.3	Unterscheiden sich die Mittelwerte mehrerer Gruppen?	343
10.3.1	Grafische und numerische Beschreibung	343
10.3.2	Analyse der Fragestellung	344
10.3.3	Post-hoc-Tests	347
10.4	Unterscheidet sich die Lage einer Variablen zwischen mehreren Gruppen?.....	348
10.5	Wie wirken zwei kategoriale Variablen kombiniert auf Mittelwerte?	351
10.5.1	Numerische und grafische Beschreibung	351
10.5.2	Analyse der Fragestellung	353
10.5.3	Modellselektion	356
10.6	Hängen Chancen von einer oder mehreren Variablen ab?	358
10.6.1	Logistische Regression	359
10.6.2	Logistische Regression mit mehreren erklärenden Variablen ...	362
10.7	Unterscheiden sich Chancen und Odds-Ratios zwischen Gruppen?.....	367
10.7.1	Vergleich von Odds in mehreren Gruppen?.....	367
10.7.2	Vergleich von Odds-Ratios in mehreren Gruppen?	370
10.8	R-Befehle im Überblick	374
10.9	Zusammenfassung der Konzepte	376
10.10	Übungen	376

Teil V	Multivariate Daten	379
Kapitel 11	Dimensionsreduktion	381
11.1	Kann man Komplexität multidimensionaler Daten reduzieren?	382
11.1.1	Grundlagen der Hauptkomponentenanalyse	383
11.1.2	Anwendung der Hauptkomponentenanalyse	393
11.2	Weiterverwenden der Ergebnisse einer Hauptkomponentenanalyse	404
11.3	R-Befehle im Überblick	409
11.4	Zusammenfassung der Konzepte	410
11.5	Übungen	410
11.6	Vertiefung: Extraktion der Hauptkomponenten für zwei Variablen	411
Kapitel 12	Gruppierung von Beobachtungen	415
12.1	Wie entdeckt man Gruppen ähnlicher Beobachtungen?	417
12.1.1	Distanz- und Ähnlichkeitsmaße	418
12.1.2	Hierarchische Clusterverfahren	420
12.1.3	Outputteile	424
12.1.4	Anwendung auf die Demografiekennzahlen	428
12.1.5	Teilungsverfahren	430
12.1.6	Speichern der Clusterzugehörigkeit	431
12.2	Wie findet man Cluster in den Variablen?	432
12.3	Wie findet man Cluster in großen Datensätzen?	434
12.3.1	Centroid-Verfahren	434
12.3.2	Outputteile des Verfahrens	435
12.3.3	Analyse des Outputs	437
12.4	Wie können kategoriale Variablen einbezogen werden?	439
12.4.1	Distanzmatrix	440
12.4.2	Fuzzy-Verfahren	440
12.4.3	Outputteile	441
12.4.4	Analyse des Outputs	442
12.5	R-Befehle im Überblick	445
12.6	Zusammenfassung der Konzepte	446
12.7	Übungen	446
Literaturverzeichnis		449
Index		451