

Auf einen Blick

Einführung	31
Teil I: Eine Einführung in R	35
Kapitel 1: R – Was es macht und wie es dies macht.....	37
Kapitel 2: Mit Packages arbeiten, Importieren und Exportieren.....	65
Teil II: Daten beschreiben	79
Kapitel 3: Grafik.....	81
Kapitel 4: Finden Sie Ihre Mitte!.....	119
Kapitel 5: Weg vom Durchschnitt.....	129
Kapitel 6: Standards und Ränge.....	137
Kapitel 7: Eine Zusammenfassung des Ganzen.....	147
Kapitel 8: Was ist normal?.....	165
Teil III: Daten analysieren	183
Kapitel 9: Ein Spiel mit dem Vertrauen: Schätzung.....	185
Kapitel 10: Hypothesentests mit einer Stichprobe.....	201
Kapitel 11: Hypothesentests mit zwei Stichproben.....	227
Kapitel 12: Tests von mehr als zwei Stichproben.....	253
Kapitel 13: Kompliziertere Tests.....	277
Kapitel 14: Regression: Lineare und multiple und das allgemeine lineare Modell.....	297
Kapitel 15: Korrelation: Aufstieg und Fall von Beziehungen.....	331
Kapitel 16: Kurvilineare Regression: Wenn Beziehungen kompliziert werden.....	351
Kapitel 17: Zu gegebener Zeit.....	375
Kapitel 18: Nichtparametrische Statistik.....	387
Kapitel 19: Einführung in die Wahrscheinlichkeitsrechnung.....	407
Kapitel 20: Wahrscheinlichkeit trifft auf Regression: Logistische Regression.....	427
Teil IV: Aus Daten lernen	435
Kapitel 21: Tools und Daten für Projekte zum maschinellen Lernen.....	437
Kapitel 22: Entscheidungen, Entscheidungen!.....	463
Kapitel 23: In den Wald – per Zufall.....	481
Kapitel 24: Unterstützen Sie Ihren lokalen Vektor.....	497
Kapitel 25: k-Means-Clustering.....	517
Kapitel 26: Neuronale Netze.....	533
Kapitel 27: Datenerforschung im Marketing.....	549
Kapitel 28: Aus der Stadt, die niemals schläft.....	567

Teil V: R in der Praxis: Einige Projekte, die Sie auf Trab halten.....	583
Kapitel 29: Mit einem Browser arbeiten.....	585
Kapitel 30: Dashboards verwenden	613
Abbildungsverzeichnis	647
Stichwortverzeichnis	657

Inhaltsverzeichnis

Einführung	31
Über dieses Buch	31
Teil 1: Einführung in R	31
Teil 2: Daten beschreiben	31
Teil 3: Daten auswerten	32
Teil 4: Aus Daten lernen	32
Teil 5: R nutzen – Einige Projekte, die Sie auf die Praxis vorbereiten	32
Was Sie getrost überspringen können	33
Törichte Annahmen über den Leser	33
In diesem Buch verwendete Symbole	33
Wie es jetzt weitergeht	34
 TEIL I	
EINE EINFÜHRUNG IN R	35
 Kapitel 1	
R – Was es macht und wie es dies macht	37
Statistische (und verwandte) Konzepte, die man einfach kennen muss	37
Stichproben und Grundgesamtheiten	38
Variablen: abhängige und unabhängige	38
Verschiedene Arten von Daten	40
Eine kleine Wahrscheinlichkeit	40
Inferenzstatistik: Hypothesen testen	42
Null- und Alternativhypothesen	42
Zwei Arten von Fehlern	43
R beschaffen	44
RStudio beschaffen	45
Eine Sitzung mit R	48
Das Arbeitsverzeichnis	48
Erste Schritte	49
Funktionen in R	52
Benutzerdefinierte Funktionen	53
Kommentare	54
Strukturen in R	54
Vektoren	55
Numerische Vektoren	55
Matrizen	57
Listen	59
Datenrahmen	60
for-Schleifen und if-Anweisungen	63

Kapitel 2

Mit Packages arbeiten, Importieren und

Exportieren	65
Packages installieren	65
Daten prüfen	67
Kopf und Zahl	67
Fehlende Daten	68
Untergruppen	68
Formeln in R	69
Weitere Packages	70
Das Tidyversum	71
Importieren und Exportieren	75
Tabellenkalkulationen	75
CSV-Dateien	76
Textdateien	76

TEIL II

DATEN BESCHREIBEN 79

Kapitel 3

Grafik	81
Muster finden	81
Grafische Darstellung einer Verteilung	82
Bar-Hopping	82
Tortenstücke	84
Die Darstellung der Streuung	85
Von Boxen und Schnurrhaaren	85
Grundlagen beherrschen: Basisgrafiken in R	87
Histogramme	87
Diagrammeigenschaften	89
Balkendiagramme	90
Tortendiagramme	92
Punktdiagramme	92
Balkendiagramme neu betrachtet	93
Streudiagramme	96
Anders ausgedrückt	97
Streudiagramm-Matrix	97
Box-Plots	99
Einen Zahn zulegen – mit ggplot2	100
Histogramme	101
Balkendiagramme	103
Punktdiagramme	104
Balkendiagramme neu aufgelegt	107
Streudiagramme	111

Wie es zu diesem Wandel kommt	111
Streudiagramm-Matrix	113
Box-Diagramme	115
Dranbleiben!	117

Kapitel 4

Finden Sie Ihre Mittel!..... 119

Mittelwerte: Die Verlockung des Durchschnitts	119
Berechnung des Mittelwerts	120
Der Durchschnitt in R: <code>mean()</code>	121
Wie lautet die Bedingung?	121
Die \$-Zeichen loswerden	122
Daten genauer erkunden	122
Ausreißer: Der Fehler der Durchschnittswerte	124
Mediane: In der Mitte gefangen	124
Der Median in R: <code>median()</code>	125
Statistik à la Mode	126
Der Modus in R	127

Kapitel 5

Weg vom Durchschnitt..... 129

Messung der Varianz	129
Mittelwertbildung für die quadratischen Abweichungen:	
Varianz und wie man sie berechnet	130
Varianz der Stichprobe	132
Varianz in R	133
Zurück zu den Ursprüngen: Standardabweichung	134
Standardabweichung der Grundgesamtheit	134
Standardabweichung der Stichprobe	134
Standardabweichung in R	135
Bedingungen, Bedingungen, Bedingungen	135

Kapitel 6

Standards und Ränge..... 137

Ich kaufe ein Z!	137
Merkmale der z-Werte	138
Bonds gegen Bambino	138
Werte genauer untersuchen	139
Standardwerte in R	140
Wo stehen Sie?	142
Ranglisten in R	142
Punktgliche Ergebnisse	142
N-t kleinste, N-t größte	143
Perzentile	143
Prozentuale Ränge	145
Zusammenfassung	146

Kapitel 7

Eine Zusammenfassung des Ganzen	147
Wie viele?	147
Das Hoch und das Tief	148
In den Momenten leben.....	149
Ein lehrreicher Moment.....	149
Zurück zu den Deskriptiven	150
Schiefe	150
Wölbung (Kurtosis)	153
Bestimmung der Häufigkeit.....	155
Nominale Variablen: <code>table()</code> et al.	155
Numerische Variablen: <code>hist()</code>	155
Numerische Variablen: <code>stem()</code>	161
Zusammenfassung eines Datenrahmens	162

Kapitel 8

Was ist normal?	165
Gut auf der Kurve liegen	165
Genauer betrachtet.....	166
Parameter einer Normalverteilung.....	167
Mit Normalverteilungen arbeiten	168
Verteilungen in R.....	169
Dichtefunktion der Normalverteilung.....	169
Eine Normalkurve zeichnen	169
Kumulative Dichtefunktion	174
Ein Diagramm für die CDF.....	175
Quantile von Normalverteilungen.....	176
Ausgabe der CDF mit Quartilen	177
Zufallsstichproben	178
Treffen mit einem angesehenen Mitglied der Familie	179
Die Standardnormalverteilung in R.....	180
Ausgabe der Standardnormalverteilung	181

TEIL III

DATEN ANALYSIEREN	183
--------------------------------	------------

Kapitel 9

Ein Spiel mit dem Vertrauen: Schätzung	185
Stichprobenverteilungen verstehen.....	185
Ein AUSSERORDENTLICH wichtiges Konzept: Der zentrale Grenzwertsatz.....	187
(Annähernde) Simulation des zentralen Grenzwertsatzes.....	188
Vorhersagen des zentralen Grenzwertsatzes	192
Vertrauen: Alles hat seine Grenzen!.....	194
Ermittlung von Konfidenzgrenzen für einen Mittelwert.....	195
Mit R die Konfidenzgrenzen für einen Mittelwert ermitteln.....	196
Annähern an ein t	197

Kapitel 10

Hypothesentests mit einer Stichprobe 201

Hypothesen, Tests und Fehler	201
Hypothesentests und Stichprobenverteilungen	203
Wir kaufen noch ein Z	205
Z-Tests in R	207
t for One	209
t -Test in R	210
Die Arbeit mit t -Verteilungen	210
Visualisierung von t -Verteilungen	211
t in grundlegenden R-Diagrammen ausgeben	212
Ausgabe von t in ggplot2	213
Ein Wort zu ggplot2	218
Eine Varianz testen	218
Ein Beispiel anlegen	219
Testen in R	220
Mit Chi-Quadrat-Verteilungen arbeiten	221
Visualisierung von Chi-Quadrat-Verteilungen	222
Darstellung von Chi-Quadrat in grundlegender R-Grafik	222
Chi-Quadrat-Darstellung in ggplot2	223

Kapitel 11

Hypothesentests mit zwei Stichproben 227

Hypothesen für zwei	227
Stichprobenverteilungen unter der Lupe	228
Anwendung des zentralen Grenzwertsatzes	229
Und nochmal Z	230
Z-Test für zwei Stichproben in R	231
t for Two	233
Wie die Erbsen in der Schote: gleiche Varianzen	234
t -Tests in R	235
Mit zwei Vektoren arbeiten	235
Mit einem Datenrahmen und einer Formel arbeiten	236
Visualisierung der Ergebnisse	237
Wie p s und q s: ungleiche Varianzen	240
Perfekt kombiniert: Hypothesentests für paarweise Stichproben	241
t -Test für paarweise Stichproben in R	243
Zwei Varianzen testen	243
F-Tests in R	245
F in Verbindung mit t	246
Mit F -Verteilungen arbeiten	246
Visualisierung von F -Verteilungen	247

Kapitel 12**Tests von mehr als zwei Stichproben..... 253**

Mehr als zwei testen.....	253
Ein heikles Problem.....	254
Eine Lösung.....	255
Bedeutsame Beziehungen.....	259
ANOVA in R.....	259
Einen Boxplot zur Visualisierung der Daten ausgeben.....	261
Nach der ANOVA.....	261
Kontraste in R.....	264
Ungeplante Vergleiche.....	265
Eine andere Art von Hypothese, eine andere Art von Test.....	266
Mit ANOVA mit wiederholten Messungen arbeiten.....	266
ANOVA mit wiederholten Messungen in R.....	268
Visualisierung der Ergebnisse.....	270
Dem Trend folgen.....	271
Trendanalyse in R.....	274

Kapitel 13**Kompliziertere Tests..... 277**

Die Kombinationen knacken.....	277
Wechselwirkungen.....	278
Die Analyse.....	279
Zwei-Faktoren-ANOVA in R.....	281
Visualisierung der wechselseitigen Ergebnisse.....	282
Zwei Arten von Variablen ... auf einmal.....	284
Gemischte ANOVA in R.....	287
Visualisierung der gemischten ANOVA-Ergebnisse.....	288
Nach der Analyse.....	290
Multivariate Varianzanalyse.....	290
MANOVA in R.....	292
Visualisierung der MANOVA-Ergebnisse.....	293
Nach der MANOVA.....	295

Kapitel 14**Regression: Lineare und multiple und das allgemeine lineare Modell..... 297**

Das Streudiagramm.....	297
Linien grafisch darstellen.....	299
Regression: Was für eine Linie!.....	301
Verwendung der Regression für Prognosen.....	302
Variation um die Regressionslinie.....	303
Hypothesen über die Regression testen.....	304
Prüfung der Passform.....	305
Test der Steigung.....	308
Test des Schnittpunkts.....	309

Lineare Regression in R.....	309
Merkmale des linearen Modells.....	311
Prognosen treffen	311
Visualisierung des Streudiagramms und der Regressionslinie	311
Ausgabe der Residuen	312
Mit vielen Beziehungen gleichzeitig jonglieren: Mehrfache Regression.....	314
Mehrfache Regression in R.....	315
Prognosen machen.....	316
Visualisierung des 3D-Streudiagramms und der Regressionsebene.....	317
Das Package scatterplot3d.....	317
car und rgl: ein Paketangebot.....	318
ANOVA: Eine andere Perspektive.....	319
Analyse der Kovarianz: Die letzte Komponente des allgemeinen linearen Modells	323
Doch halt – das war noch nicht alles!.....	328

Kapitel 15

Korrelation: Aufstieg und Fall von Beziehungen..... 331

Korrelation verstehen	331
Korrelation und Regression	334
Hypothesen über die Korrelation testen.....	337
Ist ein Korrelationskoeffizient größer als Null?	337
Unterscheiden sich zwei Korrelationskoeffizienten?.....	338
Korrelation in R.....	339
Berechnung eines Korrelationskoeffizienten.....	340
Einen Korrelationskoeffizienten testen	340
Die Differenz zwischen zwei Korrelationskoeffizienten testen.....	340
Eine Korrelationsmatrix berechnen.....	341
Korrelationsmatrizen visualisieren.....	341
Multiple Korrelation.....	344
Multiple Korrelation in R	345
Anpassung des R-Quadrats.....	346
Partielle Korrelation.....	346
Partielle Korrelation in R	347
Semipartielle Korrelation.....	348
Semipartielle Korrelation in R.....	349

Kapitel 16

Kurvilineare Regression: Wenn Beziehungen

kompliziert werden..... 351

Was ist ein Logarithmus?.....	352
Was ist e?	354
Potenzregression	357
Exponentielle Regression.....	362

24 Inhaltsverzeichnis

Logarithmische Regression.....	366
Polynomielle Regression: Eine höhere Potenz.....	369
Welches Modell sollten Sie verwenden?.....	372
Kapitel 17	
Zu gegebener Zeit.....	375
Eine Zeitreihe und ihre Komponenten	375
Prognose: Eine bewegliche Erfahrung.....	379
Prognose: Ein anderer Weg.....	382
Mit realen Daten arbeiten	383
Kapitel 18	
Nichtparametrische Statistik.....	387
Unabhängige Stichproben.....	387
Zwei Stichproben: Wilcoxon Rangsummentest.....	388
Mehr als zwei Stichproben: Kruskal-Wallis Einweg-ANOVA.....	391
Paarweise Stichproben	394
Zwei paarweise Stichproben: Wilcoxon-Vorzeichenrang-Test.....	394
Mehr als zwei Stichproben: Friedman-ANOVA.....	396
Mehr als zwei Stichproben: Cochrans Q.....	398
Korrelation: Spearmans r_s	401
Korrelation: Kendalls Tau.....	402
Eine Vorwarnung.....	405
Kapitel 19	
Einführung in die Wahrscheinlichkeitsrechnung.....	407
Was ist Wahrscheinlichkeit?.....	407
Experimente, Versuche, Ereignisse und Ergebnismengen.....	408
Ergebnismengen und Wahrscheinlichkeiten	408
Zusammengesetzte Ereignisse.....	409
Vereinigung und Schnitt.....	409
Schnitt, Wiederholung.....	410
Bedingte Wahrscheinlichkeit.....	411
Mit den Wahrscheinlichkeiten arbeiten	412
Die Grundlage der Hypothesentests.....	412
Große Ergebnismengen	412
Permutationen.....	413
Kombinationen.....	414
R-Funktionen für Zählregeln.....	415
Diskrete und stetige Zufallsvariablen.....	416
Wahrscheinlichkeitsverteilungen und Dichtefunktionen	417
Die Binomialverteilung.....	419

Binom und negatives Binom in R	420
Binomialverteilung.....	420
Negative Binomialverteilung.....	422
Hypothesentests mit der Binomialverteilung.....	423
Mehr über Hypothesentests: R vs. Tradition	424

Kapitel 20

Wahrscheinlichkeit trifft auf Regression: Logistische

Regression	427
Beschaffung der Daten	430
Durchführung der Analyse	430
Ergebnisse visualisieren	433

TEIL IV

AUS DATEN LERNEN

435

Kapitel 21

Tools und Daten für Projekte zum maschinellen Lernen

437

Das ML-Repository der UCI (University of California-Irvine).....	438
Mit einem UCI-Datensatz arbeiten.....	438
Daten bereinigen.....	441
Erkundung der Daten.....	442
Kurzer Projektvorschlag: Dichtediagramme	444
Untersuchung von Beziehungen in den Daten.....	444
Grafiken im grundlegenden R.....	445
Die ggplot-Version	447
Das Rattle-Package – eine Einführung	450
Verwendung von Rattle für iris.....	454
Beschaffung und (weitere) Erkundung der Daten.....	454
Cluster in Daten finden.....	458

Kapitel 22

Entscheidungen, Entscheidungen!.....

463

Komponenten des Entscheidungsbaums.....	463
Wurzeln und Blätter	464
Baumkonstruktion	465
Entscheidungsbäume in R.....	465
Lassen Sie den Baum in R gedeihen.....	466
Den Baum in R zeichnen.....	467
Entscheidungsbäume in Rattle.....	469
Den Baum erstellen	470
Die Ausgabe des Baums	471
Bewertung des Baums	472
Projekt: Ein komplexerer Entscheidungsbaum	473
Die Daten: Fahrzeugbewertung	474
Daten erkunden	475

Den Baum erstellen und ausgeben.....	476
Bewertung des Baums	477
Schneller Projektvorschlag: Das Verständnis des Komplexitätsparameters	478
Vorgeschlagenes Projekt: Titanic.....	479

Kapitel 23

In den Wald – per Zufall 481

Das Wachsen eines Zufallswaldes	481
Zufallswälder in R.....	483
Aufbau des Waldes	483
Bewertung des Waldes	484
Ein genauerer Blick	485
Fehler bei der Ausgabe.....	486
Den Einfluss darstellen.....	488
Projekt: Glas identifizieren.....	489
Die Daten	489
Eingabe der Daten in Rattle.....	490
Die Daten erkunden.....	491
Der Zufallswald wächst.....	492
Visualisierung der Ergebnisse.....	494
Vorgeschlagenes Projekt: Pilze identifizieren	495

Kapitel 24

Unterstützen Sie Ihren lokalen Vektor 497

Einige Daten, mit denen Sie arbeiten können	497
Eine Teilmenge verwenden	498
Eine Grenze festlegen.....	498
Stützvektoren verstehen.....	499
Trennbarkeit: Normalerweise nichtlinear	500
Stützvektor-Maschinen in R.....	503
Mit e1071 arbeiten.....	503
Training der SVM.....	504
Mit kernlab arbeiten	508
Projekt: Und wer sind Sie?.....	510
Einlesen der Daten	511
Erkundung der Daten	512
Erstellen der SVM	514
Auswertung der SVM.....	515

Kapitel 25

k-Means-Clustering 517

Wie das Ganze funktioniert	517
k-Means-Clustering in R.....	519
Daten einrichten und analysieren.....	519
Die Ausgabe verstehen.....	519
Die Cluster visualisieren	521

Die optimale Anzahl von Clustern ermitteln.....	522
Kurzer Projektvorschlag: Hinzufügen der Kelchblätter	526
Projekt: Glas-Cluster	527
Die Daten	527
Rattle starten und die Daten erkunden.....	528
Vorbereitung auf das Clustering	529
Das Clustering durchführen	529
Über Rattle hinaus.....	530

Kapitel 26

Neuronale Netze..... 533

Netzwerke im Nervensystem.....	533
Künstliche neuronale Netze.....	534
Übersicht.....	534
Eingabeschicht und verborgene Schicht.....	535
Ausgabeschicht.....	536
Wie das Ganze funktioniert.....	536
Neuronale Netze in R.....	537
Ein neuronales Netz für den iris-Datenrahmen aufbauen.....	537
Das Netzwerk zeichnen.....	539
Das Netz bewerten	540
Kurzer Projektvorschlag: Die Kelchblätter.....	540
Projekt: Banknoten.....	541
Die Daten	541
Ein kurzer Blick in die Zukunft.....	542
Rattle einrichten.....	542
Das Netz bewerten	544
Über Rattle hinaus: Visualisierung des Netzwerks.....	545
Projektvorschläge: Mit Rattle experimentieren	547

Kapitel 27

Datenerforschung im Marketing..... 549

Analyse von Einzelhandelsdaten.....	549
Die Daten	550
RFM in R.....	551
Aufbereitung der Daten.....	551
Durchführung der Analyse.....	553
Die Ergebnisse prüfen.....	554
Ein Blick auf die Länder	556
Maschinelles Lernen.....	558
Mit k-Means-Clustering arbeiten	558
Mit Rattle arbeiten.....	559
Tiefer in die Cluster sehen.....	561
Cluster und Klassen	563
Kurzer Projektvorschlag.....	564
Projekt: Ein weiterer Datensatz	565

Kapitel 28

Aus der Stadt, die niemals schläft	567
Der Datensatz	567
Aufwärmen	568
Einblicke und Ausblicke	568
Piping, Filterung und Gruppierung	568
Visualisierung	570
Verknüpfungen	572
Kurzer Projektvorschlag: Namen von Fluggesellschaften	574
Projektvorschlag: Verzögerungen bei der Abreise	574
Eine Variable hinzufügen: Wochentag	575
Projektvorschlag: Analyse der Unterschiede zwischen den Wochentagen	576
Verspätung, Wochentag und Flughafen	576
Verspätung und Flugdauer	579
Projekt: Verspätung und Wetter	580

TEIL V

R IN DER PRAXIS: EINIGE PROJEKTE, DIE SIE AUF TRAB HALTEN	583
--	------------

Kapitel 29

Mit einem Browser arbeiten	585
Auf Hochglanz gebracht	585
Ihr erstes shiny-Projekt	586
Die Benutzeroberfläche	589
Der Server	590
Letzte Schritte	591
Reagieren	592
Mit ggplot arbeiten	595
Den Server ändern	596
Ein paar weitere Änderungen	597
Reaktiv mit ggplot	599
Ein weiteres shiny-Projekt	602
Die Version in grundlegendem R	602
Die ggplot-Version	609
Projektvorschlag	611

Kapitel 30

Dashboards verwenden	613
Das Package shinydashboard	613
Dashboard-Layouts	614
Erste Schritte mit der Benutzeroberfläche	614
Aufbau der Benutzeroberfläche: Boxen, Boxen, Boxen...	615
Aufreihung in Spalten	622
Ein netter Trick: Registerkarten	625

Projektvorschlag: Statistiken hinzufügen	629
Projektvorschlag: valueBoxes in tabPanels platzieren	629
Mit der Seitenleiste arbeiten.....	631
Die Benutzeroberfläche.....	632
Der Server	635
Projektvorschlag: Verlegung des Schiebereglers.....	636
Interaktion mit Grafiken	637
Klicks, Doppelklicks und Markieren – behalten Sie den Überblick!.....	638
Wozu die Mühe?	641
Projektvorschlag: Experiment zur Luftqualität.....	644
Abbildungsverzeichnis	647
Stichwortverzeichnis	657