



Joseph Schmuller

Statistik mit R

dummies[®]

Übersetzung aus dem Amerikanischen von
Rainer G. Haselier

WILEY

WILEY-VCH Verlag GmbH & Co. KGaA

Auf einen Blick

Über den Autor	9
Einführung	21
Teil I: Erste Schritte bei der statistischen Analyse mit R	25
Kapitel 1: Daten, Statistiken und Entscheidungen	27
Kapitel 2: R: Was R kann und wie R das macht	37
Teil II: Daten beschreiben	69
Kapitel 3: Daten grafisch darstellen.....	71
Kapitel 4: Suchen Sie Ihre Mitte	111
Kapitel 5: Abweichungen vom Durchschnitt	123
Kapitel 6: Standards und Wertungen kennenlernen	131
Kapitel 7: Alles zusammenfassen.....	143
Kapitel 8: Was ist normal?.....	163
Teil III: Rückschlüsse aus Daten ziehen	183
Kapitel 9: Die Sache mit dem Vertrauen: Schätzung	185
Kapitel 10: Ein-Stichproben-Hypothesentest	201
Kapitel 11: Zwei-Stichproben-Hypothesentest	227
Kapitel 12: Mehr als zwei Stichproben testen	253
Kapitel 13: Komplexere Tests	277
Kapitel 14: Lineare, multiple und allgemeine lineare Regression	297
Kapitel 15: Korrelation: Aufstieg und Fall von Zusammenhängen.....	331
Kapitel 16: Kurvenförmige Regression: Wenn Beziehungen kompliziert werden.....	351
Teil IV: Umgang mit der Wahrscheinlichkeit	375
Kapitel 17: Einführung in die Wahrscheinlichkeit	377
Kapitel 18: Einführung in die statistische Modellierung.....	399
Teil V: Der Top-Ten-Teil	419
Kapitel 19: Zehn nützliche R-Onlineressourcen.....	421
Stichwortverzeichnis	425

Inhaltsverzeichnis

Über den Autor	9
Einführung	21
Über dieses Buch	21
Ähnlichkeiten mit diesem anderen »Für Dummies«-Buch	22
Was Sie nicht lesen müssen	22
Törichte Annahmen über den Leser	22
Wie dieses Buch aufgebaut ist	22
Symbole, die in diesem Buch verwendet werden	23
Wie es weitergeht	24
TEIL I	
ERSTE SCHRITTE BEI DER STATISTISCHEN ANALYSE MIT R	25
Kapitel 1	
Daten, Statistiken und Entscheidungen	27
Die statistischen (und damit verwandten) Begriffe, die Sie einfach kennen müssen	28
Stichproben und Grundgesamtheiten	28
Abhängige und unabhängige Variablen	29
Arten von Daten	30
Ein bisschen Wahrscheinlichkeit	31
Inferenzstatistik: Testen von Hypothesen	32
Nullhypothese und Alternativhypothese	33
Zwei Arten von Fehlern	34
Kapitel 2	
R: Was R kann und wie R das macht	37
R und RStudio herunterladen	37
Eine Session mit R	41
Das Arbeitsverzeichnis	41
Jetzt geht es richtig los	42
Fehlende Daten	45
R-Funktionen	46
Benutzerdefinierte Funktionen	47
Kommentare	48
R-Strukturen	48
Vektoren	49
Numerische Vektoren	49
Matrizen	50
Faktoren	52

14 Inhaltsverzeichnis

Listen.....	53
Listen und Statistik	54
Datensätze (Data Frames)	55
Daten aus einem Datensatz extrahieren	57
Packages.....	58
Weitere Packages	61
Die R-Formelschnittstelle.....	62
Lesen und Schreiben	63
Tabellenkalkulation	63
CSV-Dateien	65
Textdateien	66
TEIL II	
DATEN BESCHREIBEN.....	69
Kapitel 3	
Daten grafisch darstellen	71
Muster erkennen	71
Verteilung grafisch darstellen.....	72
Säulensprünge	73
Die Torte schneiden	74
Das verstreute Diagramm	75
Kastengrafik: Kästchen und Antennen	76
Diagramme mit dem R-Basispaket erstellen.....	77
Histogramme.....	78
Diagrammfeatures hinzufügen	79
Säulendiagramme	81
Kreisdiagramme	82
Punktdiagramme	83
Noch einmal Säulendiagramme.....	84
Streudiagramme	87
Boxplots	91
Zu ggplot2 aufsteigen.....	92
Histogramme.....	92
Säulendiagramme	96
Punktdiagramme	96
Noch einmal Säulendiagramme, die Zweite	100
Streudiagramme	103
Matrix von Streudiagrammen	106
Boxplots	108
Zusammenfassung und Ausblick	110
Kapitel 4	
Suchen Sie Ihre Mitte	111
Mittelwert: Die Lehre vom Durchschnitt.....	111
Der Mittelwert in R: mean().....	113

Wie lauten Ihre Bedingungen?	113
Mit <code>with()</code> die Dollarzeichen weglassen.....	114
Die Daten erforschen.....	114
Ausreißer: Schönheitsfehler der Mittelwerte.....	116
Und schließlich noch ein paar andere Mittel.....	116
Mediane: Auf halber Strecke erwischt.....	118
Der Median in R: <code>median()</code>	119
Der Modalwert	120
Der Modalwert in R.....	120
Kapitel 5 Abweichungen vom Durchschnitt	123
Die Streuung berechnen	123
Mittelwert von quadratischen Abweichungen: Varianz, und wie sie berechnet wird.....	124
Varianz einer Stichprobe	126
Varianz in R	127
Zurück zu den Wurzeln: Standardabweichung	127
Standardabweichung einer Grundgesamtheit	128
Standardabweichung einer Stichprobe	128
Standardabweichung in R	129
Bedingungen, Bedingungen, Bedingungen	129
Kapitel 6 Standards und Wertungen kennenlernen	131
z-Werte einfangen	131
Eigenschaften von z-Werten	132
Bonds und Ruth	133
Prüfungsergebnisse	133
Standardwerte in R	134
Wo stehen Sie?	136
Rangermittlung in R	136
Gleiche Werte	137
k-kleinster und k-größter Wert.....	137
Quantile	138
Prozentrang	139
Zusammenfassen	140
Kapitel 7 Alles zusammenfassen	143
Wie viele?	143
Groß und klein.....	145
Im Moment leben	145
Ein lehrreicher Moment	145
Zurück zu den Beschreibungen	146
Schiefe.....	146
Kurtosis – Wölbung.....	149

16 Inhaltsverzeichnis

Nun kommt die Häufigkeit ins Spiel	151
Nominalskalierte Variablen: <code>table()</code> et al.	151
Numerische Variablen: <code>hist()</code>	152
Kumulierte Häufigkeit	153
Schritt für Schritt: Die empirische kumulative Verteilungsfunktion.....	154
Numerische Variablen: <code>stem()</code>	158
Einen Datensatz zusammenfassen	159
Kapitel 8	
Was ist normal?	163
So kratzen Sie die Kurve	163
Tiefer graben	164
Parameter einer Normalverteilung	165
Mit Normalverteilungen arbeiten	167
Verteilungen in R.....	167
Dichtefunktion.....	167
Normalverteilung als Kurve darstellen.....	168
Kumulierte Dichtefunktion	172
Die Verteilungsfunktion zeichnen	174
Quantile der Normalverteilungen	175
Die Verteilungsfunktion mit Quartilen zeichnen.....	176
Zufällige Stichproben	177
Eine ganz besondere Verteilung.....	178
Die Standardnormalverteilung in R.....	179
Die Standardnormalverteilung als Graphen darstellen.....	180
TEIL III	
RÜCKSCHLÜSSE AUS DATEN ZIEHEN	183
Kapitel 9	
Die Sache mit dem Vertrauen: Schätzung	185
Stichprobenverteilungen verstehen	186
Ein BESONDERS wichtiges Konzept: der zentrale Grenzwertsatz.....	187
(Näherungsweise) den zentralen Grenzwertsatz simulieren	189
Vorhersagen des zentralen Grenzwertsatzes	193
Vertrauen: Es gibt Grenzen	195
So ermitteln Sie die Vertrauengrenzen für einen Mittelwert.....	195
Passend für ein t	197
Kapitel 10	
Ein-Stichproben-Hypothesentest.....	201
Hypothesen, Tests und Fehler	201
Hypothesentests und Stichprobenverteilungen	203
Noch einmal z-Werte	205
z-Test in R.....	207
t-Test für eine Stichprobe	209

t-Tests in R	210
Mit t-Verteilungen arbeiten	210
t-Verteilungen visualisieren	211
t mit den R-Basisfunktionen darstellen	212
Diagramm mit ggplot2 erstellen	214
Eine Sache noch zu ggplot2	218
Testen einer Varianz	219
In R testen	220
Mit Chi-Quadrat-Verteilungen arbeiten	222
Chi-Quadrat-Verteilungen visualisieren	222
Chi-Quadrat mit den R-Basisfunktionen darstellen	223
Chi-Quadrat mit ggplot2 darstellen	224
Kapitel 11	
Zwei-Stichproben-Hypothesentest	227
Hypothesen für zwei	227
Noch einmal Stichprobenverteilungen	228
Den zentralen Grenzwertsatz anwenden	229
Noch einmal z-Werte	230
Z-Test für zwei Stichproben in R	232
t-Test für zwei Stichproben	234
Wie ein Ei dem anderen: gleiche Varianzen	234
t-Test in R	235
Zwei Vektoren verwenden	236
Einen Datensatz und eine Formel verwenden	236
Die Ergebnisse visualisieren	237
Wie Äpfel und Birnen: unterschiedliche Varianzen	241
Ein passendes Paar: Hypothesentest für abhängige Stichproben	242
t-Test für abhängige Stichproben in R	243
Zwei Varianzen testen	244
F-Test in R	245
F zusammen mit t	246
Mit F-Verteilungen arbeiten	247
F-Verteilungen visualisieren	247
Kapitel 12	
Mehr als zwei Stichproben testen	253
Mehr als zwei Stichproben testen	253
Eine harte Nuss	254
Eine Lösung	255
Wichtige Zusammenhänge	258
ANOVA in R	259
Die Ergebnisse visualisieren	260
Nach der ANOVA	261

18 Inhaltsverzeichnis

Kontraste in R	264
Nicht geplante Vergleiche	265
Eine andere Art Hypothese, eine andere Art Test.....	266
Mit wiederholten Messungen bei der Varianzanalyse arbeiten	266
Varianzanalyse für wiederholte Messungen in R	268
Die Ergebnisse visualisieren	270
Jetzt wird es trendy.....	271
Trendanalyse in R	275

Kapitel 13

Komplexere Tests.....

277

Die Kombinationen knacken.....	277
Interaktionen.....	278
Die Analyse.....	279
Zweifaktorielle Varianzanalyse in R	281
Ergebnisse der zweifaktoriellen Varianzanalyse visualisieren	282
Zwei Arten von Variablen, und zwar gleichzeitig.....	285
Gemischte ANOVA in R	287
Ergebnisse der gemischten ANOVA visualisieren	289
Nach der Analyse	290
Multivariate Varianzanalyse	291
MANOVA in R	292
MANOVA-Ergebnisse visualisieren.....	293
Nach der Analyse	295

Kapitel 14

Lineare, multiple und allgemeine lineare Regression

297

Das Streudiagramm	297
Geraden zeichnen.....	299
Regression: Was für eine Gerade!	301
Die Regression für Schätzungen verwenden	302
Streuung um die Regressionsgerade	303
Hypothesen über die Regression testen.....	304
Lineare Regression in R	309
Features des linearen Modells	311
Vorhersagen treffen	311
Das Streudiagramm und die Regressionsgerade visualisieren	311
Residuendiagramm erstellen	312
Irrsinnig viele Zusammenhänge auf einmal: multiple Regression	314
Multiple Regression in R	315
Vorhersagen treffen	316
Das 3D-Streudiagramm und die Regressionsebene visualisieren.....	317
ANOVA: Eine andere Perspektive.....	320
Kovarianzanalyse: Die letzte Komponente des allgemeinen linearen Modells	323
Moment bitte – da gibt's noch mehr	328

Kapitel 15	
Korrelation: Aufstieg und Fall von Zusammenhängen	331
Noch einmal Streudiagramme	331
Grundlegendes zur Korrelation	332
Korrelation und Regression	334
Hypothesen über Korrelationen testen	337
Ist ein Korrelationskoeffizient größer als null?	337
Unterscheiden sich zwei Korrelationskoeffizienten voneinander?	338
Korrelation in R	339
Korrelationskoeffizient berechnen	339
Korrelationskoeffizient testen	340
Die Differenz zwischen zwei Korrelationskoeffizienten testen	340
Eine Korrelationsmatrix berechnen	341
Korrelationsmatrizen visualisieren	341
Multiple Korrelation	344
Multiple Korrelation in R	344
Das Bestimmtheitsmaß korrigieren	345
Partialkorrelation	346
Partialkorrelation in R	347
Semipartialkorrelation	348
Semipartialkorrelation in R	348
Kapitel 16	
Kurvenförmige Regression: Wenn Beziehungen kompliziert werden	351
Was ist ein Logarithmus?	352
Was ist e?	354
Potenzregression	356
Exponentielle Regression	362
Logarithmische Regression	366
Polynomische Regression: Eine größere Potenz	369
Welches Modell sollten Sie verwenden?	373
TEIL IV	
UMGANG MIT DER WAHRSCHEINLICHKEIT	375
Kapitel 17	
Einführung in die Wahrscheinlichkeit	377
Was ist Wahrscheinlichkeit?	377
Experimente, Versuche, Ereignisse und Stichprobenräume	378
Wahrscheinlichkeitsräume und Wahrscheinlichkeit	378
Zusammengesetzte Ereignisse	379
Vereinigung und Schnitt	379
Noch mehr zum Schnitt	380
Bedingte Wahrscheinlichkeit	381
Mit Wahrscheinlichkeiten arbeiten	382
Die Grundlage des Testens von Hypothesen	382

20 Inhaltsverzeichnis

Große Wahrscheinlichkeitsräume	383
Permutationen	384
Kombinationen	384
R-Funktionen für Zählregeln	385
Zufallsvariablen: diskret und stetig	387
Wahrscheinlichkeitsverteilungen und Dichtefunktionen	387
Die Binomialverteilung	389
Binomial- und negative Binomialverteilung in R	390
Binomialverteilung	391
Negative Binomialverteilung	393
Hypothesen mit der Binomialverteilung testen	394
Weitere Informationen zum Testen von Hypothesen: R versus Tradition	395
Kapitel 18 Einführung in die statistische Modellierung	399
Die Modellierung einer Verteilung	399
Näheres zur Poissonverteilung	400
Modellierung mit der Poissonverteilung	401
Prüfen, ob das Modell passt	405
Ein kurzer Hinweis zu chisq.test()	406
Modelle für Baseballstatistiken	408
Simulationen	411
Es darauf ankommen lassen: die Monte-Carlo-Methode	411
Den Würfel »zinken«	411
Simulation des zentralen Grenzwertsatzes	415
TEIL V DER TOP-TEN-TEIL	419
Kapitel 19 Zehn nützliche R-Onlineressourcen	421
Websites für R-Anwender	421
R-bloggers	421
Microsoft R Application Network	421
Quick-R	422
RStudio Online Learning	422
Stack Overflow	422
Online-Bücher und weitere Dokumentation	422
R-Handbuch	423
R-Dokumentation	423
RDocumentation	423
YOU CANalytics	423
The R Journal	423
Stichwortverzeichnis	425