



Joseph Schmuller

# Statistik mit R

für  
**dummies**<sup>®</sup>

2. Auflage

**WILEY**  
WILEY-VCH GmbH

# Inhaltsverzeichnis

<b>Über den Autor</b> .....	<b>9</b>
<b>Einführung</b> .....	<b>21</b>
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
Teil I: Erste Schritte bei der statistischen Analyse mit R .....	23
Teil II: Daten beschreiben .....	23
Teil III: Rückschlüsse aus Daten ziehen .....	23
Teil IV: Umgang mit der Wahrscheinlichkeit .....	23
Teil V: Der Top-Ten-Teil .....	23
Symbole, die in diesem Buch verwendet werden .....	24
Wie es weitergeht .....	24
<b>TEIL I</b>	
<b>ERSTE SCHRITTE BEI DER STATISTISCHEN ANALYSE MIT R</b> .....	<b>25</b>
<b>Kapitel 1</b>	
<b>Daten, Statistiken und Entscheidungen</b> .....	<b>27</b>
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
<b>Kapitel 2</b>	
<b>R: Was R kann und wie R das macht</b> .....	<b>37</b>
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 .....	48
Kommentare .....	48
R-Strukturen .....	49
Vektoren .....	49
Numerische Vektoren .....	50

## 14 Inhaltsverzeichnis

Matrizen.....	51
Faktoren.....	52
Listen.....	53
Listen und Statistik.....	54
Datensätze (Data Frames).....	55
Daten aus einem Datensatz extrahieren.....	57
Packages.....	58
Weitere Packages.....	60
Die R-Formelschnittstelle.....	62
Lesen und Schreiben.....	63
Tabellenkalkulation.....	63
CSV-Dateien.....	65
Textdateien.....	66
<b>TEIL II</b>	
<b>DATEN BESCHREIBEN.....</b>	<b>69</b>
<b>Kapitel 3</b>	
<b>Daten grafisch darstellen.....</b>	<b>71</b>
Muster erkennen.....	71
Verteilung grafisch darstellen.....	72
Säulensprünge.....	72
Die Torte schneiden.....	74
Das verstreute Diagramm.....	75
Kastengrafik: Kästchen und Antennen.....	76
Diagramme mit dem R-Basispaket erstellen.....	77
Histogramme.....	77
Diagrammfeatures hinzufügen.....	79
Säulendiagramme.....	80
Kreisdiagramme.....	82
Punktdiagramme.....	82
Noch einmal Säulendiagramme.....	83
Streudiagramme.....	86
Boxplots.....	90
Zu ggplot2 aufsteigen.....	90
Histogramme.....	91
Säulendiagramme.....	94
Punktdiagramme.....	95
Noch einmal Säulendiagramme, die Zweite.....	98
Streudiagramme.....	101
Matrix von Streudiagrammen.....	103
Boxplots.....	105
Zusammenfassung und Ausblick.....	108
<b>Kapitel 4</b>	
<b>Suchen Sie Ihre Mitte.....</b>	<b>109</b>
Mittelwert: Die Lehre vom Durchschnitt.....	109
Der Mittelwert in R: mean().....	111

Wie lauten Ihre Bedingungen?.....	111
Mit with() die Dollarzeichen weglassen.....	112
Die Daten erforschen.....	112
Ausreißer: Schönheitsfehler der Mittelwerte.....	114
Und schließlich noch ein paar andere Mittel.....	114
Mediane: Auf halber Strecke erwischt.....	116
Der Median in R: median().....	117

## Kapitel 5

### Abweichungen vom Durchschnitt ..... 119

Die Streuung berechnen.....	119
Mittelwert von quadratischen Abweichungen: Varianz, und wie sie berechnet wird.....	120
Varianz einer Stichprobe.....	122
Varianz in R.....	123
Zurück zu den Wurzeln: Standardabweichung.....	124
Standardabweichung einer Grundgesamtheit.....	124
Standardabweichung einer Stichprobe.....	124
Standardabweichung in R.....	125
Bedingungen, Bedingungen, Bedingungen.....	125

## Kapitel 6

### Standards und Wertungen kennenlernen ..... 127

z-Werte einfangen.....	127
Eigenschaften von z-Werten.....	128
Bonds und Ruth.....	128
Prüfungsergebnisse.....	129
Standardwerte in R.....	130
Wo stehen Sie?.....	132
Rangermittlung in R.....	132
Gleiche Werte.....	132
k-kleinsten und k-größten Wert.....	133
Quantile.....	133
Prozentrang.....	134
Zusammenfassen.....	136

## Kapitel 7

### Alles zusammenfassen ..... 137

Wie viele?.....	137
Groß und klein.....	139
Im Moment leben.....	139
Ein lehrreicher Moment.....	140
Zurück zu den Beschreibungen.....	140
Schiefe.....	141
Kurtosis - Wölbung.....	144
Nun kommt die Häufigkeit ins Spiel.....	145
Nominalskalierte Variablen: table() et al.....	145

Numerische Variable: hist() .....	146
Kumulierte Häufigkeit .....	147
Schritt für Schritt: Die empirische kumulative Verteilungsfunktion .....	147
Numerische Variable: stem() .....	151
Einen Datensatz zusammenfassen .....	153

## Kapitel 8

### Was ist normal? ..... 157

So kratzen Sie die Kurve .....	157
Tiefer graben .....	158
Parameter einer Normalverteilung .....	159
Mit Normalverteilungen arbeiten .....	161
Verteilungen in R .....	161
Dichtefunktion .....	161
Normalverteilung als Kurve darstellen .....	162
Kumulierte Dichtefunktion .....	165
Die Verteilungsfunktion zeichnen .....	167
Quantile der Normalverteilungen .....	168
Die Verteilungsfunktion mit Quartilen zeichnen .....	170
Zufällige Stichproben .....	171
Eine ganz besondere Verteilung .....	171
Die Standardnormalverteilung in R .....	173
Die Standardnormalverteilung als Graphen darstellen .....	174

## TEIL III

### RÜCKSCHLÜSSE AUS DATEN ZIEHEN ..... 175

## Kapitel 9

### Die Sache mit dem Vertrauen: Schätzung ..... 177

Stichprobenverteilungen verstehen .....	177
Ein BESONDERS wichtiges Konzept: der zentrale Grenzwertsatz .....	179
(Näherungsweise) den zentralen Grenzwertsatz simulieren .....	180
Vorhersagen des zentralen Grenzwertsatzes .....	185
Vertrauen: Es gibt Grenzen .....	186
So ermitteln Sie die Vertrauensgrenzen für einen Mittelwert .....	186
Passend für ein t .....	188

## Kapitel 10

### Ein-Stichproben-Hypothesentest ..... 191

Hypothesen, Tests und Fehler .....	191
Hypothesentests und Stichprobenverteilungen .....	193
Noch einmal z-Werte .....	195
z-Test in R .....	197
t-Test für eine Stichprobe .....	199
t-Tests in R .....	200
Mit t-Verteilungen arbeiten .....	200

t-Verteilungen visualisieren .....	201
t mit den R-Basisfunktionen darstellen.....	202
Diagramm mit ggplot2 erstellen .....	204
Eine Sache noch zu ggplot2.....	208
Testen einer Varianz .....	209
In R testen .....	210
Mit Chi-Quadrat-Verteilungen arbeiten.....	212
Chi-Quadrat-Verteilungen visualisieren.....	212
Chi-Quadrat mit den R-Basisfunktionen darstellen.....	212
Chi-Quadrat mit ggplot2 darstellen.....	214

## Kapitel 11

### Zwei-Stichproben-Hypothesentest..... 217

Hypothesen für zwei.....	217
Noch einmal Stichprobenverteilungen.....	218
Den zentralen Grenzwertsatz anwenden.....	219
Noch einmal z-Werte.....	220
Z-Test für zwei Stichproben in R.....	222
t-Test für zwei Stichproben .....	223
Wie ein Ei dem anderen: gleiche Varianzen .....	224
t-Test in R.....	225
Zwei Vektoren verwenden .....	226
Einen Datensatz und eine Formel verwenden.....	226
Die Ergebnisse visualisieren.....	227
Wie Äpfel und Birnen: unterschiedliche Varianzen.....	230
Ein passendes Paar: Hypothesentest für abhängige Stichproben.....	231
t-Test für abhängige Stichproben in R.....	233
Zwei Varianzen testen.....	233
F-Test in R.....	235
F zusammen mit t.....	236
Mit F-Verteilungen arbeiten.....	236
F-Verteilungen visualisieren.....	237

## Kapitel 12

### Mehr als zwei Stichproben testen..... 243

Mehr als zwei Stichproben testen.....	243
Eine harte Nuss.....	244
Eine Lösung.....	245
Wichtige Zusammenhänge .....	249
ANOVA in R.....	249
Die Ergebnisse visualisieren.....	250
Nach der ANOVA.....	251
Kontraste in R.....	254
Nicht geplante Vergleiche .....	255
Eine andere Art Hypothese, eine andere Art Test .....	256
Mit wiederholten Messungen bei der Varianzanalyse arbeiten.....	256

Varianzanalyse für wiederholte Messungen in R.....	258
Die Ergebnisse visualisieren.....	260
Jetzt wird es trendy.....	261
Trendanalyse in R.....	264

## Kapitel 13

<b>Komplexere Tests .....</b>	<b>267</b>
Die Kombinationen knacken.....	267
Interaktionen.....	269
Die Analyse.....	269
Zweifaktorielle Varianzanalyse in R.....	271
Ergebnisse der zweifaktoriellen Varianzanalyse visualisieren.....	272
Zwei Arten von Variablen, und zwar gleichzeitig.....	275
Gemischte ANOVA in R.....	277
Ergebnisse der gemischten ANOVA visualisieren.....	279
Nach der Analyse.....	280
Multivariate Varianzanalyse.....	280
MANOVA in R.....	282
MANOVA-Ergebnisse visualisieren.....	283
Nach der Analyse.....	285

## Kapitel 14

<b>Lineare, multiple und allgemeine lineare Regression.....</b>	<b>287</b>
Das Streudiagramm.....	287
Geraden zeichnen.....	289
Regression: Was für eine Gerade!.....	291
Die Regression für Schätzungen verwenden.....	293
Streuung um die Regressionsgerade.....	293
Hypothesen über die Regression testen.....	295
Lineare Regression in R.....	300
Features des linearen Modells.....	301
Vorhersagen treffen.....	301
Das Streudiagramm und die Regressionsgerade visualisieren.....	302
Residuendiagramm erstellen.....	303
Irrsinnig viele Zusammenhänge auf einmal: multiple Regression.....	304
Multiple Regression in R.....	305
Vorhersagen treffen.....	306
Das 3-D-Streudiagramm und die Regressionsebene visualisieren.....	307
ANOVA: Eine andere Perspektive.....	310
Kovarianzanalyse: Die letzte Komponente des allgemeinen linearen Modells ...	313
Moment bitte – da gibt’s noch mehr.....	319

## Kapitel 15

<b>Korrelation: Aufstieg und Fall von Zusammenhängen.....</b>	<b>321</b>
Noch einmal Streudiagramme.....	321
Grundlegendes zur Korrelation.....	322

Korrelation und Regression .....	324
Hypothesen über Korrelationen testen.....	327
Ist ein Korrelationskoeffizient größer als null?.....	327
Unterscheiden sich zwei Korrelationskoeffizienten voneinander? .....	328
Korrelation in R.....	329
Korrelationskoeffizient berechnen.....	329
Korrelationskoeffizient testen .....	330
Die Differenz zwischen zwei Korrelationskoeffizienten testen.....	330
Eine Korrelationsmatrix berechnen.....	331
Korrelationsmatrizen visualisieren.....	331
Multiple Korrelation.....	334
Multiple Korrelation in R.....	334
Das Bestimmtheitsmaß korrigieren .....	335
Partialkorrelation .....	336
Partialkorrelation in R.....	337
Semipartialkorrelation .....	338
Semipartialkorrelation in R.....	338

## Kapitel 16

### Kurvenförmige Regression: Wenn Beziehungen

<b>kompliziert werden .....</b>	<b>341</b>
Was ist ein Logarithmus?.....	342
Was ist e? .....	344
Potenzregression .....	346
Exponentielle Regression.....	352
Logarithmische Regression.....	356
Polynomische Regression: Eine größere Potenz.....	359
Welches Modell sollten Sie verwenden?.....	363

## TEIL IV

### UMGANG MIT DER WAHRSCHEINLICHKEIT..... 365

## Kapitel 17

### Einführung in die Wahrscheinlichkeit..... 367

Was ist Wahrscheinlichkeit? .....	367
Experimente, Versuche, Ereignisse und Stichprobenräume .....	368
Wahrscheinlichkeitsräume und Wahrscheinlichkeit .....	368
Zusammengesetzte Ereignisse.....	369
Vereinigung und Schnitt.....	369
Noch mehr zum Schnitt.....	370
Bedingte Wahrscheinlichkeit.....	371
Mit Wahrscheinlichkeiten arbeiten.....	372
Die Grundlage des Testens von Hypothesen .....	372
Große Wahrscheinlichkeitsräume.....	372
Permutationen .....	373
Kombinationen.....	374
R-Funktionen für Zählregeln.....	375

Zufallsvariablen: diskret und stetig.....	376
Wahrscheinlichkeitsverteilungen und Dichtefunktionen .....	377
Die Binomialverteilung.....	379
Binomial- und negative Binomialverteilung in R.....	380
Binomialverteilung.....	380
Negative Binomialverteilung.....	382
Hypothesen mit der Binomialverteilung testen.....	383
Weitere Informationen zum Testen von Hypothesen: R vs. Tradition .....	385

## Kapitel 18

### **Einführung in die statistische Modellierung..... 387**

Die Modellierung einer Verteilung.....	387
Näheres zur Poissonverteilung.....	388
Modellierung mit der Poissonverteilung.....	389
Prüfen, ob das Modell passt.....	393
Ein kurzer Hinweis zu <code>chisq.test()</code> .....	394
Modelle für Baseballstatistiken .....	396
Simulationen .....	399
Es darauf ankommen lassen: die Monte-Carlo-Methode.....	399
Den Würfel »zinken« .....	399
Simulation des zentralen Grenzwertsatzes .....	403

## TEIL V

### **DER TOP-TEN-TEIL..... 407**

## Kapitel 19

### **(Fast) zehn nützliche R-Onlineressourcen ..... 409**

Websites für R-Anwender.....	409
R-bloggers .....	409
Microsoft R Application Network.....	409
Quick-R.....	410
Stack Overflow .....	410
Online-Bücher und weitere Dokumentation.....	410
R-Handbuch.....	410
R-Dokumentation .....	411
RDocumentation .....	411
The R Journal.....	411

### **Abbildungsverzeichnis..... 413**

### **Stichwortverzeichnis..... 421**

# Abbildungsverzeichnis

- Abbildung 1.1: Die Beziehung zwischen Grundgesamtheit, Stichprobe, Parametern und Statistiken 29
- Abbildung 2.1: RStudio, wie es beim ersten Start nach der Installation aussieht 39
- Abbildung 2.2: Die Registerkarte PACKAGES in RStudio 40
- Abbildung 2.3: Die Registerkarte HELP in RStudio 40
- Abbildung 2.4: RStudio, nachdem Sie an der rechten Seite der Titelleiste der Konsole das große Symbol angeklickt haben 41
- Abbildung 2.5: Die Registerkarte ENVIRONMENT nach der Erzeugung des Vektors  $x$  43
- Abbildung 2.6: RStudio, nachdem ein Vektor erstellt und mit ihm gearbeitet wurde 44
- Abbildung 2.7: Das Dialogfeld QUIT R SESSION 45
- Abbildung 2.8: Die Funktion `edit()` öffnet eine tabellenblattähnliche Ansicht eines Datensatzes. 56
- Abbildung 2.9: Die Registerkarte HELP zeigt Informationen über das MASS-Package an. 58
- Abbildung 2.10: Der Datensatz `anorexia` aus dem Package MASS 59
- Abbildung 2.11: Das Dialogfeld INSTALL PACKAGES 61
- Abbildung 2.12: Die Registerkarte PACKAGES, nachdem `ggplot2` installiert und der Bibliothek hinzugefügt wurde 62
- Abbildung 2.13: Der Datensatz `anorexia` wurde in eine Excel-Datei exportiert. 65
- Abbildung 2.14: Der Datensatz `anorexia` als Textdatei, in der der Tabulator als Trennzeichen verwendet wurde 67
- Abbildung 3.1: Auszug aus dem Datensatz `Cars93` 72
- Abbildung 3.2: Histogramm der Anschaffungspreise der Autos im Datensatz `Cars93` 73
- Abbildung 3.3: Tabelle 3.1 als Säulendiagramm 74
- Abbildung 3.4: Tabelle 3.1 als Kreisdiagramm 75
- Abbildung 3.5: Verbrauch im Stadtverkehr und PS für die Daten in `Cars93` 75
- Abbildung 3.6: Kastengrafik der Beziehung PS vs. Anzahl der Zylinder des Datensatzes `Cars93` 76
- Abbildung 3.7: Das anfängliche Histogramm mit der Verteilung der Preise in `Cars93` 78
- Abbildung 3.8: Dichteplot der Verteilung der Preise in `Car93` 79
- Abbildung 3.9: Dichteplot mit ergänzter Kurve 80
- Abbildung 3.10: Das anfängliche Säulendiagramm für `table(Cars93$Type)` 81
- Abbildung 3.11: Dotplot der Daten aus Tabelle 3.1 82
- Abbildung 3.12: Die Daten aus Tabelle 3.2 in einem Diagramm 84
- Abbildung 3.13: Anfängliches Säulendiagramm der Daten aus Tabelle 3.2 86
- Abbildung 3.14: Verbrauch im Stadtverkehr und PS mit gefüllten Kreisen (`pch=16`) 87
- Abbildung 3.15: Verbrauch vs. PS mit Datenpunkten, die die Anzahl der Zylinder angeben 88
- Abbildung 3.16: Mehrere Streudiagramme für die Beziehungen zwischen Verbrauch, Preis und PS 89
- Abbildung 3.17: `ggplot()` fast ohne Argumente verwendet 92
- Abbildung 3.18: Das anfängliche Histogramm für Price in `Cars93` 93

## 414 **Abbildungsverzeichnis**

- Abbildung 3.19: Das fertige Price-Histogramm 94
- Abbildung 3.20: Säulendiagramm für Fahrzeugtypen 95
- Abbildung 3.21: Das anfängliche Punktdiagramm für die Fahrzeugtypen 97
- Abbildung 3.22: Das bearbeitete Punktdiagramm für die Fahrzeugtypen 98
- Abbildung 3.23: Säulendiagramm der Daten aus Tabelle 3.2, mit ggplot() erstellt 99
- Abbildung 3.24: PS und Benzinverbrauch in Cars93 102
- Abbildung 3.25: Das anfängliche ggplot2-Streudiagramm Benzinverbrauch vs. PS mit der Zylinderanzahl als »Symbol« der Datenpunkte 103
- Abbildung 3.26: Das bearbeitete Streudiagramm Benzinverbrauch vs. PS mit der Zylinderanzahl als »Symbol« der Datenpunkte 104
- Abbildung 3.27: Streudiagrammmatrix für Verbrauch, Preis und PS 105
- Abbildung 3.28: Das Hinzufügen von Cylinders erzeugt diese Diagrammmatrix. 106
- Abbildung 3.29: Boxplot für PS vs. Zylinder 107
- Abbildung 3.30: Boxplot mit Datenpunkten 107
- Abbildung 3.31: Boxplot mit verlagerten Datenpunkten 108
- Abbildung 4.1: PS-Histogramme für die in den USA und anderorts produzierten Autos im Datensatz Cars93 113
- Abbildung 7.1: Drei Diagramme, die die unterschiedlichen Schiefen zeigen 141
- Abbildung 7.2: PS-Histogramme von in den USA und anderswo hergestellten Autos 142
- Abbildung 7.3: PS-DichtepLOTS von in den USA und anderswo hergestellten Autos 143
- Abbildung 7.4: Zwei Histogramme mit unterschiedlicher Kurtosis 144
- Abbildung 7.5: Kumuliertes Häufigkeitshistogramm der Preisangaben aus Cars93 148
- Abbildung 7.6: Empirische kumulative Verteilungsfunktion für die Preisdaten in Cars93 149
- Abbildung 7.7: Die empirische kumulative Verteilungsfunktion für die Preisdaten in Cars93, mit ggplot() erstellt 150
- Abbildung 7.8: Die empirische kumulative Verteilungsfunktion für die Preisdaten mit gestrichelten Linien an den Quartilen 151
- Abbildung 7.9: Die empirische kumulative Verteilungsfunktion für die Preisdaten mit Quartilswerten an der x-Achse 152
- Abbildung 7.10: Diagramm, das mit datadensity(autos) erstellt wurde 155
- Abbildung 8.1: Die Glockenkurve 158
- Abbildung 8.2: Die Normalverteilung des Intelligenzquotienten, unterteilt in Standardabweichungen 160
- Abbildung 8.3: Das Ausgangsdiagramm der Dichtefunktion für die Normalverteilung des IQ 163
- Abbildung 8.4: Die Dichtefunktion für IQ, bei der die Standardabweichungen auf der x-Achse dargestellt werden 164
- Abbildung 8.5: Das IQ-Diagramm mit vertikalen, gestrichelten Liniensegmenten an den Standardabweichungen 165
- Abbildung 8.6: Das fertige Ergebnis: das IQ-Diagramm ohne Leerraum zwischen den x-Werten und der x-Achse 166
- Abbildung 8.7: Visualisierung der Wahrscheinlichkeit eines IQ-Werts zwischen 85 und 100, mit dem Package tigerstats erstellt 167
- Abbildung 8.8: Kumulierte Dichtefunktion der IQ-Verteilung 168
- Abbildung 8.9: Das Diagramm, das mit qnormGC() gezeichnet wurde 169

Abbildung 8.10: Die normale kumulative Verteilungsfunktion mit Quartilswerten 170

Abbildung 8.11: Die Standardnormalverteilung, aufgeteilt in Standardabweichungen 173

Abbildung 8.12: Die Standardnormalverteilung nach Standardabweichungen eingeteilt und mit ggplot() gezeichnet 174

Abbildung 9.1: Erstellen einer Stichprobenverteilung des Mittelwerts 178

Abbildung 9.2: Die Stichprobenverteilung des Mittelwerts, in Standardfehlereinheiten unterteilt 180

Abbildung 9.3: Stichprobenverteilung des Mittelwerts basierend auf 600 Stichproben mit drei Elementen aus einer Grundgesamtheit der gleichermaßen wahrscheinlichen Werte 1, 2 und 3 183

Abbildung 9.4: Die Stichprobenverteilung des Mittelwerts mit einer neu skalierten x-Achse und coolen Achsenbeschriftungen 184

Abbildung 9.5: Die Stichprobenverteilung des Mittelwerts für die FarBlonJet-Batterie 187

Abbildung 9.6: Die 95 igen Vertrauensgrenzen auf der FarBlonJet-Stichprobenverteilungskurve 188

Abbildung 9.7: Einige t-Verteilungskurven 189

Abbildung 10.1:  $H_0$  und  $H_1$  entsprechen jeweils einer Stichprobenverteilung. 194

Abbildung 10.2: Der zweiseitige kritische Bereich für  $\alpha = 0,05$  197

Abbildung 10.3: t-Verteilung mit drei Freiheitsgraden, R-Basisfunktionen 202

Abbildung 10.4: Drei Verteilungen suchen nach einer Legende. 203

Abbildung 10.5: Das fertige Diagramm, einschließlich der Legende 204

Abbildung 10.6: Die Graphen dreier t-Verteilungen, mit ggplot2 erstellt 206

Abbildung 10.7: Drei t-Verteilungen mit anders zugeordneten Linientypen 207

Abbildung 10.8: Das fertige Ergebnis mit neu sortierter Legende 208

Abbildung 10.9: Zwei Chi-Quadrat-Verteilungskurven 210

Abbildung 10.10: Zwei Mitglieder der Chi-Quadrat-Familie, mit den R-Basisfunktionen gezeichnet 214

Abbildung 10.11: Zwei Mitglieder der Chi-Quadrat-Familie, mit ggplot2 gezeichnet 215

Abbildung 11.1: Erstellen einer Stichprobenverteilung der Differenz zwischen Mittelwerten 219

Abbildung 11.2: Die Stichprobenverteilung der Differenz zwischen Mittelwerten nach dem zentralen Grenzwertsatz 220

Abbildung 11.3: Die Stichprobenverteilung der Differenz zwischen Mittelwerten zusammen mit dem kritischen Wert für  $\alpha = 0,05$  und dem berechneten Wert der Testgröße im Beispiel mit dem IQ 221

Abbildung 11.4: Boxplot der FarKlemp-Maschinendaten, der mit Standard-R-Grafikfunktionen erstellt wurde 228

Abbildung 11.5: Boxplot der FarKlemp-Maschinendaten, der mit ggplot erstellt wurde 228

Abbildung 11.6: Die Mittelwerte und Standardfehler der FarKlemp-Maschinendaten 229

Abbildung 11.7: Die Mittelwerte und Standardfehler der FarKlemp-Maschinendaten 234

Abbildung 11.8: Zwei Mitglieder der Familie der F-Verteilungen, mit ggplot2 dargestellt 239

Abbildung 11.9: Zwei mit ggplot2 dargestellte Mitglieder der Familie der F-Verteilungen: das fertige Ergebnis 240

Abbildung 12.1: Die F-Verteilung mit den Freiheitsgraden 2 und 24 248

Abbildung 12.2: Boxplot der Stichprobenergebnisse 251

## 416 **Abbildungsverzeichnis**

Abbildung 12.3: Die Mittelwerte und Standardfehler des Beispiels zur Gewichtsabnahme 260

Abbildung 12.4: Ein quadratischer Trend mit vier Mittelwerten 262

Abbildung 12.5: Ein kubischer Trend mit vier Mittelwerten 262

Abbildung 13.1: Kombination aus den Ausprägungen der Präsentationsmethode mit den Ausprägungen des Präsentationsstils 268

Abbildung 13.2: Mittelwerte und Standardfehler zur Präsentationsstudie 274

Abbildung 13.3: Mittelwerte und Standardfehler für die Studie Buch vs. E-Reader 279

Abbildung 13.4: Drei Boxplots zeigen die Verteilung der Scores für Physik, Chemie und Biologie für die einzelnen Bücher. 284

Abbildung 14.1: Begabung und Leistung bei FarMisht Consulting 288

Abbildung 14.2: Der Graph für  $y = 4 + 2x$  290

Abbildung 14.3: Die Abweichungen in einem Streudiagramm 296

Abbildung 14.4: Streudiagramm und Regressionsgerade für 16 FarMisht-Berater 302

Abbildung 14.5: Residuendiagramm für das FarMisht-Beispiel 303

Abbildung 14.6: Streudiagramm für das FarMisht-Beispiel zur multiplen Regression, das mit `scatterplot3d()` erstellt wurde 307

Abbildung 14.7: Streudiagramm für das FarMisht-Beispiel zur multiplen Regression, komplett mit der Regressionsebene 308

Abbildung 14.8: Streudiagramm für das FarMisht-Beispiel zur multiplen Regression, mit `scatter3d()` erstellt 309

Abbildung 14.9: Gewichtszunahme vs. Behandlungsform im Datensatz anorexia. 315

Abbildung 14.10: Gewichtszunahme vs. Behandlung und vorheriges Gewicht im Datensatz anorexia 317

Abbildung 14.11: Gewichtszunahme vs. Behandlung und vorheriges Gewicht im Datensatz anorexia, mit Regressionsgeraden für die Werte der Stufen in `Treat` 320

Abbildung 15.1: Begabung und Leistung bei FarMisht Consulting 322

Abbildung 15.2: Streudiagramm der 16 FarMisht-Berater mit der Regressionsgeraden 324

Abbildung 15.3: Ein Punkt im Streudiagramm und die zugehörigen Abstände 325

Abbildung 15.4: Die Korrelationsmatrix für Begabung, Leistung und Persönlichkeit (mit den Standardgrafikfunktionen von R erstellt) 332

Abbildung 15.5: Die Korrelationsmatrix für Begabung, Leistung und Persönlichkeit, die mit GGally (einem auf ggplot2 basierten Package) erstellt wurde 332

Abbildung 15.6: Die Korrelationsmatrix für Begabung, Leistung und Persönlichkeit (mit dem Package `corrgram` erstellt) 333

Abbildung 16.1: Hypothetischer Plot des Erlernens einer Fertigkeit wie das Schuhebinden 342

Abbildung 16.2: Die Beziehung zwischen Körpergewicht und Hirngewicht für 25 Tierarten 348

Abbildung 16.3: Die Beziehung zwischen dem Logarithmus des Körpergewichts und dem Logarithmus des Gehirngewichts für 25 Tierarten 349

Abbildung 16.4: Die Beziehung zwischen dem Logarithmus des Körpergewichts und dem Logarithmus des Gehirngewichts für 25 Tierarten mit einer Regressionsgeraden 350

Abbildung 16.5: Ursprünglicher Plot zur Beziehung zwischen Körpergewicht und Gehirngewicht für 25 Tierarten mit der Potenzregressionskurve 352

Abbildung 16.6: Wie der Bierschaum (`krone.cm`) im Laufe der Zeit zerfällt 353

Abbildung 16.7: Wie der `log(krone.cm)` im Laufe der Zeit zerfällt, einschließlich der Regressionsgeraden 354

Abbildung 16.8: Der Zerfall der Bierkrone im Verlauf der Zeit, mit der exponentiellen Regressionskurve 356

Abbildung 16.9: Verbrauch (Highway) und PS im Datensatz Cars93 357

Abbildung 16.10: MPG.highway und  $\log(\text{Horsepower})$  in Cars 93, zusammen mit der Regressionsgeraden 358

Abbildung 16.11: Verbrauch (Highway) und PS mit der logarithmischen Regressionskurve 359

Abbildung 16.12: Scatterplot des Medianwerts ( $\text{medv}$ ) gegen Zimmeranzahl ( $\text{rm} = \text{rooms}$ ) im Datensatz Boston, mit Regressionsgerade 360

Abbildung 16.13: Scatterplot des Medianwerts ( $\text{medv}$ ) gegen Zimmeranzahl ( $\text{rm} = \text{rooms}$ ) im Datensatz Boston, mit polynomischer Regressionskurve 363

Abbildung 17.1: Die Wahrscheinlichkeitsverteilung für  $x$ , einer

Zufallsvariablen, die auf den Würfeln eines neutralen, symmetrischen Würfels basiert 377

Abbildung 17.2: Die Standardnormalverteilung: eine Wahrscheinlichkeitsdichtefunktion 378

Abbildung 17.3: Binomialverteilung der Anzahl der Erfolge, wenn mit einem symmetrischen Würfel zehnmal gewürfelt wird 381

Abbildung 17.4: Kumulierte Verteilung der Anzahl der Erfolge, wenn mit einem symmetrischen Würfel zehnmal gewürfelt wird 382

Abbildung 18.1: Die Poissonverteilung mit  $\mu = 3$  390

Abbildung 18.2: Seitenaufrufe der FarBlonJet-Intranet-Homepage; beobachtet und mit Poisson vorhergesagt ( $\mu = 3$ ) 392

Abbildung 18.3: Stichprobenverteilung des Mittelwerts ( $N = 25$ ), basierend auf 10.000 Stichproben aus einer Normalverteilung mit  $\mu = 100$  und  $\sigma = 15$  405