

**Michael J. Crawley**

# **StatistikmitR**

Aus dem Englischen von  
Silvia Kinkel

**WILEY**

WILEY-VCH Verlag GmbH & Co. KGaA

# Inhaltsverzeichnis

## **Vorwort** *n*

RundS-PLUS 12

## **1 Die Grundlagen** 15

Alles variiert 16

Signifikanz 17

Gute und schlechte Hypothesen 18

Nullhypothesen 18

p-Werte 19

Interpretation 19

Statistisches Modellieren 20

Maximale Wahrscheinlichkeit 21

Versuchsordnung 23

Das Sparsamkeitsprinzip (Ockhams Rasiermesser) 24

Beobachtung, Theorie und Versuch 25

Kontrollen 25

Wiederholungen: Die *ns* rechtfertigen die Mittelwerte 25

Wie viele Wiederholungen? 26

Teststärke 27

Randomisierung 28

Starke Inferenz 31

Schwache Inferenz 31

Wann ist es genug? 32

Pseudowiederholung 33

Ausgangsbedingungen 34

Orthogonales Design und nichtorthogonale Beobachtungsdaten 35

## **z Dataframes (Datentabellen)** 37

Teile eines Dataframes auswählen: Indizes 42

5

Sortieren 44  
Ergebnisse speichern 46  
Aufraumen 46

## **Zentrale Tendenz 49**

Hilfefunktion in R 59

## **Varianz 61**

Freiheitsgrade 64  
Varianz 66  
Ein Arbeitsbeispiel 68  
Varianz und Stichprobengröße 72  
Die Anwendung von Varianz 74  
Ein Maß für Unzuverlässigkeit 74  
Konfidenzintervalle 76  
Bootstrapping 78

## **Einzelne Stichproben 83**

Zusammenfassung der Daten im Fall von Einstichproben 83  
Die Normalverteilung 89  
Berechnungen unter Verwendung von  $z$  der Normalverteilung 96  
Grafische Darstellungen zum Testen auf Normalverteilung einzelner Stichproben 100  
Schlussfolgerung im Fall von Einstichproben 102  
Bootstrapping beim Testen von Hypothesen mit nur einer Stichprobe 102  
T-Verteilung nach Student 104  
Hoherwertige Momente einer Verteilung 106  
Schiefe 106  
Kurtosis 109

## **Zwei Stichproben 111**

Der Vergleich von zwei Varianzen 111  
Der Vergleich von zwei Mittelwerten 114  
Der t-Test nach Student 115  
Der Wilcoxon-Rangsummentest 120  
Tests mit gekoppelten Stichproben 122  
Der Vorzeichentest 124  
Binomialtests zum Vergleich von zwei Verhältnissen 126

Chi-Quadrat-Kontingenztest	127
Exakter Test nach Fisher	133
Korrelation und Kovarianz	137
Data Dredging	140
Partielle Korrelation	142
Korrelation und die Varianz von Unterschieden zwischen Variablen	142
Skalenabh'ngige Korrelationen	145
Kolmogorov-Smirnov-Test	147

## **Statistisches Modellieren 151**

Die Schritte bei der Modellvereinfachung	153
Vorsichtsmafnahmen	134
Reihenfolge beim Weglassen	155
Modellformeln in R	155
Wechselwirkungen zwischen erklärenden Variablen	157
Terme mit multiplen Fehlern	159
Der Achsenabschnitt als Parameter 1	160
Die Funktion Update bei der Modellvereinfachung	161
Beispiele für Modellformeln in R	161
Modellformeln für Regression	162
GLMs: Generalisierte Lineare Modelle	165
Die Fehlerstruktur	166
Der Linear-Prädiktor	167
Angepasste Werte	168
Die Link-Funktion	168
Kanonische Link-Funktionen	169
Verhältnisdaten und binomial-verteilte Fehler	170
Zahldaten und Poisson-verteilte Fehler	172
GAMs: Generalisierte Additive Modelle	172
Modellkritik	173
Zusammenfassung der statistischen Modelle in R	173
Modellüberprüfung	173
Nichtkonstante Varianz: Heteroskedastizität	176
Nicht normalverteilte Fehler	176
Beeinflussung	178
Leverage	179
Fehlspezifiziertes Modell	179

- 8 Regression 181**
- Lineare Regression 184
  - Lineare Regression in R 186
  - Fehlervarianz bei der Regression:  $SSY = SSR + SSE$  195
  - Die Anpassungsgute,  $r^2$ , messen 204
  - Modellüberprüfung 203
  - Polynomiale Regression 208
  - Nichtlineare Regression 213
  - Testen auf gekrummte Beziehungen 217
  - Generalisierte Additive Modelle (GAMs) 218
- 9 Varianzanalyse 221**
- Einfaktorielle Varianzanalyse 221
  - Kurzformeln 230
  - Effektgroßen 232
  - Diagramme zur Interpretation einer einfachen Varianzanalyse 237
  - Faktorielle Versuchspläne 243
  - Pseudowiederholung: Geschachtelte Versuchsdesigns und Split-Plot-Designs 249
  - Split-Plot-Versuchspläne 249
  - Zufällige Effekte und geschachtelte Designs 233
  - Fixed oder Random Effects? 234
  - Entfernen der Pseudowiederholung 255
  - Analyse von Langsschnittdaten 255
  - Abgeleitete Variablenanalyse 256
  - Varianzkomponentenanalyse (VCA) 257
  - Was ist der Unterschied zwischen Split-Plot- und hierarchischen Proben? 261
- 10 Ko varianzanalyse 263**
- 11 Multiple Regression 273**
- Ein einfaches Beispiel 273
  - Ein etwas komplexeres Beispiel 281
  - Automatisierung des Modellvereinfachungsprozesses durch die Verwendung von step 288
  - AIC (Akaikes Informationskriterium) 289

- 12 Kontraste 291**
  - Kontrastkoeffizienten 293
  - Ein Beispiel mit Kontrasten in R 294
  - A-priori-Kontraste 293
  - Modellvereinfachung durch schrittweises Entfernen 297
  - Quadratsummen manuell kontrastieren 301
  - Vergleich der drei Kontrastarten 303
  - Aliasing 308
  - Kontraste und die Parameter des Ancova-Modells 309
  - Multiple Vergleiche 314
  
- 13 Zahldaten 315**
  - Eine Regression mit Poisson-verteilten Fehlern 316
  - Abweichungsanalyse bei Zahldaten 317
  - Die Gefahr bei Kontingenztabellen 326"
  - Kovarianzanalyse mit Zahldaten 330
  - Häufigkeitsverteilungen 333
  
- 14 Verhältnisdaten 341**
  - Analysen von Daten zu einem und zwei Verhältnissen 343
  - Zahldaten'bei Verhältnissen 343
  - Odds 345
  - Oberdispersion und Hypothesentest 347
  - Anwendungen 348
  - Logistische Regression mit binomial-verteilten Fehlern 348
  - Verhältnisdaten mit kategorischen erklärenden Variablen 332
  - Kovarianzanalyse mit Binomialdaten 358
  
- 15 Todes- und Ausfalldaten 363**
  - Überlebensanalyse mit Zensieren 365
  
- 16 Bina>e Zielvariable 371**
  - Ereignisfunktionen 373
  - Ancova mit binarer Zielvariable 378
  
- Anhang 1:**
  - Grundlagen der Programmiersprache R 387**
    - Rechnen mit R 387
    - Wertzuweisung an Variablen 388

Wiederholungen erzeugen	389
Faktorausprägungen erzeugen	390
Das Erscheinungsbild von Diagrammen verändern	391
Daten aus einer Datei lesen	394
Vektorfunktionen in R	395
Indexierung: Teile von Vektoren erhalten	397
Indexierung mit logischen Ausdrücken	397
Indexierung mehrdimensionaler Tabellen	398
Indexierung von Listen	400
Das Schreiben von Funktionen in R	401
Sortieren und Anordnen	402
Elemente innerhalb von Tabellen zählen	404
Tabellen für zusammenfassende Statistiken	404
Umwandlung von kontinuierlichen in kategoriale Variablen durch die Verwendung von cut	403
Die split-Funktion	406
Trellis-Diagramme	408
Die xyplot-Funktion	409
3 D-Diagramme	412
• Matrixarithmetik	413
Lösungssysteme für lineare Gleichungen	416

**Quellen und weiterführende Literatur** 419.

**Index** 422