



# R

## Einführung durch angewandte Statistik

2., aktualisierte Auflage

Reinhold Hatzinger  
Kurt Hornik  
Herbert Nagel  
Marco J. Maier

**EX**TRAS  
ONLINE

**R**

**Einführung durch angewandte Statistik**

# R

## Inhaltsverzeichnis

R - Einführung durch angewandte Statistik

Vorwort

Vorwort zur zweiten Auflage

Fragestellungen und Methoden

Kapitel 1 Einführung

1.1 Konzeption des Buchs

1.2 Aufbau des Buchs

1.3 Programmversionen von R

1.3.1 Das REdaS Package

1.4 Wie kann dieses Buch verwendet werden?

1.5 Typografische und andere Konventionen in diesem Buch

Kapitel 2 Statistische Grundbegriffe

2.1 Einige Beispiele

2.1.1 Hochrechnung (statistisches Schätzen)

2.1.2 Prüfen von Fragestellungen (Testen von Hypothesen)

2.1.3 Erstellen von Modellen (statistisches Modellieren)

2.2 Grundlegende Konzepte

2.3 Messung und Typen von Daten

2.4 Arten von Fragestellungen und Variablen

2.5 Zusammenfassung der Konzepte

2.6 Übungen

Teil I Einführung in R

Kapitel 3 Erste Schritte

3.1 Download und Installation von R

# Inhaltsverzeichnis

3.1.1 Download

3.1.2 Installation

3.1.3 Starten und Beenden von R

3.1.4 Anpassen von R

3.1.5 Installation von Erweiterungen (Contributed Packages)

3.2 Aller Anfang ist leicht

3.3 Übungen

3.4 R-Befehle im Überblick

## Kapitel 4 Daten in R vom Fragebogen zum fertigen Datensatz

4.1 Fragebogen und Kodierung

4.2 Erfassen der kodierten Daten

4.2.1 Eingabe der Daten

4.2.2 Abspeichern und Einlesen der Daten

4.3 Organisation eines Datensatzes Data Frames

4.3.1 Manipulation einzelner Variablen eines Data Frame

4.3.2 Listen

4.3.3 Faktoren

4.3.4 Auswählen von Beobachtungseinheiten (Fällen) mit logischen Operationen

4.4 Transformieren der Daten bzw. Erzeugen von neuen Variablen

4.4.1 Berechnen neuer Variablen

4.4.2 Unendliche und undefinierte Werte

4.4.3 Umkodieren von Variablen

4.4.4 Modifikation eines Data Frame

4.4.5 Datenkontrolle

4.5 Übungen

4.6 R-Befehle im Überblick

## Kapitel 5 Mehr R

5.1 Die R-Arbeitsumgebung

5.1.1 Der Workspace

5.1.2 Working Directory das Arbeitsverzeichnis

5.2 R-Grafik

5.2.1 High-level Plotting Functions

5.2.2 Hinzufügen von Grafikelementen (low-level plotting functions)

# Inhaltsverzeichnis

5.2.3 Spezielle Einstellungen (Graphical Parameters)

5.2.4 Weitere Grafikfunktionen und Packages

## 5.3 Weiterverwenden des R-Outputs

## 5.4 Einlesen von R-Befehlen

5.4.1 Der R Editor

5.4.2 Einlesen von R-Skripts (Quellcode)

5.4.3 Direktes Kopieren von R-Code Einfügen über die Zwischenablage

## 5.5 Einlesen und Schreiben externer Dateien

5.5.1 Daten aus Microsoft Excel bzw. LibreOffice Calc

5.5.2 Dateien aus anderen Statistikpaketen (z.B. SPSS)

5.5.3 Direktes Kopieren Einfügen über die Zwischenablage

5.5.4 Schreiben von Dateien

## 5.6 Statistische Funktionen

5.6.1 Die drei Funktionen statistischer Verteilungen

5.6.2 Erzeugen von (Pseudo-)Zufallszahlen

5.6.3 Verfügbare Verteilungsfunktionen

## 5.7 Rechnen mit Matrizen

5.7.1 Transponieren

5.7.2 Matrixmultiplikation

5.7.3 Matrixinversion

5.7.4 Lösen von Gleichungssystemen

5.7.5 Spezielle Funktionen für quadratische Matrizen

5.7.6 Hilfreiche Funktionen für Matrizen

5.7.7 Ein Anwendungsbeispiel: OLS-Schätzer einer Regression

## 5.8 Erzeugung, Verkettung und Ausgabe von Textelementen

5.8.1 Escapesequenzen

5.8.2 Textelemente verketteten

5.8.3 Textausgabe

## 5.9 Entwicklungsumgebungen (IDEs) für R

## 5.10 Das R-Hilfesystem und weiterführende Information

5.10.1 Hilfe zu einzelnen Funktionen und Packages

5.10.2 Dokumente, Webseiten und weiterführende Information

## 5.11 R für Fortgeschrittene

5.11.1 Definition kleiner Funktionen

# Inhaltsverzeichnis

5.11.2 Konstanten in R

5.11.3 Kontrollstrukturen (Control Flow)

5.11.4 Objektorientiertes Programmieren (OOP), generische Funktionen und Methoden

5.11.5 Funktionsdefinitionen und Objektorientierung

5.11.6 Initialisieren von Objekten

5.11.7 Testen und Konvertieren von Typen und Klassen

5.11.8 Optimierung

5.12 Übungen

5.13 R-Befehle im Überblick

## Teil II Kategoriale Daten

### Kapitel 6 Eine kategoriale Variable

6.1 Einleitung

6.2 Kommen alle Kategorien gleich häufig vor?

6.2.1 Numerische Beschreibung

6.2.2 Grafische Beschreibung

6.2.3 Statistische Analyse der Problemstellung

6.3 Entsprechen Häufigkeiten bestimmten Vorgaben?

6.3.1 Numerische und grafische Beschreibung

6.3.2 Statistische Analyse der Problemstellung

6.4 Hat ein Prozentsatz (Anteil) einen bestimmten Wert?

6.4.1 Statistische Analyse der Problemstellung

6.5 In welchem Bereich kann man einen Prozentsatz (Anteil) erwarten?

6.6 Zusammenfassung der Konzepte

6.7 Übungen

6.8 Vertiefung: Die Chi-Quadrat-Verteilung oder wie entsteht ein p-Wert?

6.9 R-Befehle im Überblick

### Kapitel 7 Mehrere kategoriale Variablen

7.1 Beschreibung mehrerer kategorialer Variablen

7.1.1 Numerische Beschreibung

7.1.2 Grafische Beschreibung

7.2 Ist die Verteilung von Häufigkeiten in verschiedenen Gruppen gleich?

7.3 Unterscheiden sich Anteile in zwei oder mehreren Gruppen?

# Inhaltsverzeichnis

- 7.3.1 Eingabe einer Tabelle
- 7.3.2 Vergleich der Anteile
- 7.3.3 Exakter Test nach Fisher
- 7.4 Sind zwei kategoriale Variablen unabhängig?
  - 7.4.1 Datenaufbereitung
  - 7.4.2 Unabhängigkeitstest
- 7.5 Unterscheidet sich das Risiko in zwei Gruppen?
  - 7.5.1 Odds-Ratio
  - 7.5.2 Odds-Ratio-Test
- 7.6 Wie kann man Veränderungen von Anteilen testen?
  - 7.6.1 Unabhängige und abhängige Stichproben
  - 7.6.2 McNemar-Test
- 7.7 Zusammenfassung der Konzepte
- 7.8 Übungen
- 7.9 R-Befehle im Überblick

## Teil III Metrische Daten

### Kapitel 8 Eine metrische Variable

- 8.1 Wie kann man die Verteilung einer metrischen Variablen beschreiben?
  - 8.1.1 Klassifizieren, Tabellen und Histogramme
  - 8.1.2 Maßzahlen zur Beschreibung der Verteilung
  - 8.1.3 Boxplot
  - 8.1.4 Ausreißer
  - 8.1.5 Weitere grafische Beschreibungsmethoden
- 8.2 Ist der Mittelwert der Grundgesamtheit anders als eine bestimmte Vorgabe?
- 8.3 Bereich für den Mittelwert in einer Grundgesamtheit
- 8.4 Folgt eine metrische Variable einer bestimmten Verteilung?
  - 8.4.1 Q-Q-Plot
  - 8.4.2 Kolmogorov-Smirnov-Test und Shapiro-Wilk-Test
  - 8.4.3 Anpassungstest mit der  $\chi^2$ -Verteilung
- 8.5 Zusammenfassung der Konzepte
- 8.6 Übungen

# Inhaltsverzeichnis

## 8.7 R-Befehle im Überblick

## Kapitel 9 Mehrere metrische Variablen

### 9.1 Wie stark ist der Zusammenhang zwischen zwei metrischen Variablen?

#### 9.1.1 Grafische Beschreibung

#### 9.1.2 Korrelationskoeffizient nach Pearson

#### 9.1.3 Korrelationskoeffizient nach Spearman

### 9.2 Welche Form hat der Zusammenhang zwischen zwei Variablen?

#### 9.2.1 Lineares Regressionsmodell

#### 9.2.2 Rechenergebnisse

### 9.3 Vorhersage von Werten

#### 9.3.1 Punktprognosen

#### 9.3.2 Intervallprognosen

### 9.4 Zusammenhang einer mit mehreren Variablen

#### 9.4.1 Multiple lineare Regression

#### 9.4.2 Kategoriale als erklärende Variablen

#### 9.4.3 Modellselektion

#### 9.4.4 Modelldiagnostik

#### 9.4.5 Prognose

### 9.5 Unterscheiden sich Mittelwerte zu zwei oder mehreren Zeitpunkten?

#### 9.5.1 Grafische Beschreibung

#### 9.5.2 Analyse der Fragestellung

### 9.6 Wie kann man den zeitlichen Verlauf einer Variablen beschreiben?

#### 9.6.1 Zeitreihen

#### 9.6.2 Zeitreihenzerlegung

#### 9.6.3 Trend

#### 9.6.4 Saison

#### 9.6.5 Zusammenfassung der Zeitreihenzerlegung

#### 9.6.6 Prognose

#### 9.6.7 Autokorrelation

### 9.7 Zusammenfassung der Konzepte

### 9.8 Übungen

### 9.9 R-Befehle im Überblick

# Inhaltsverzeichnis

## Teil IV Metrische und kategoriale Daten

### Kapitel 10 Metrische und kategoriale Variablen

- 10.1 Unterscheiden sich die Mittelwerte in zwei Gruppen?
  - 10.1.1 Grafische und numerische Beschreibung
  - 10.1.2 Analyse der Fragestellung
- 10.2 Unterscheidet sich die Lage einer Variablen zwischen zwei Gruppen?
  - 10.2.1 Beschreibung der Stichprobe
  - 10.2.2 Analyse der Fragestellung
- 10.3 Unterscheiden sich die Mittelwerte mehrerer Gruppen?
  - 10.3.1 Grafische und numerische Beschreibung
  - 10.3.2 Analyse der Fragestellung
  - 10.3.3 Post-hoc-Tests
- 10.4 Unterscheidet sich die Lage einer Variablen zwischen mehreren Gruppen?
- 10.5 Wie wirken zwei kategoriale Variablen kombiniert auf Mittelwerte?
  - 10.5.1 Numerische und grafische Beschreibung
  - 10.5.2 Analyse der Fragestellung
  - 10.5.3 Modellselektion
- 10.6 Hängen Chancen von einer oder mehreren Variablen ab?
  - 10.6.1 Logistische Regression
  - 10.6.2 Logistische Regression mit mehreren erklärenden Variablen
- 10.7 Unterscheiden sich Chancen und Odds-Ratios zwischen Gruppen?
  - 10.7.1 Vergleich von Odds in mehreren Gruppen?
  - 10.7.2 Vergleich von Odds-Ratios in mehreren Gruppen?
- 10.8 Zusammenfassung der Konzepte
- 10.9 Übungen
- 10.10 R-Befehle im Überblick

## Teil V Multivariate Daten

### Kapitel 11 Dimensionsreduktion

- 11.1 Kann man Komplexität multidimensionaler Daten reduzieren?
  - 11.1.1 Grundlagen der Hauptkomponentenanalyse

# Inhaltsverzeichnis

11.1.2 Anwendung der Hauptkomponentenanalyse

11.2 Weiterverwenden der Ergebnisse einer Hauptkomponentenanalyse

11.3 Zusammenfassung der Konzepte

11.4 Übungen

11.5 Vertiefung: Extraktion der Hauptkomponenten für zwei Variablen

11.6 R-Befehle im Überblick

## Kapitel 12 Gruppierung von Beobachtungen

12.1 Wie entdeckt man Gruppen ähnlicher Beobachtungen?

12.1.1 Distanz- und Ähnlichkeitsmaße

12.1.2 Hierarchische Clusterverfahren

12.1.3 Outputteile

12.1.4 Anwendung auf die Demografiekennzahlen

12.1.5 Teilungsverfahren

12.1.6 Speichern der Clusterzugehörigkeit

12.2 Wie findet man Cluster in den Variablen?

12.3 Wie findet man Cluster in großen Datensätzen?

12.3.1 Centroid-Verfahren

12.3.2 Outputteile des Verfahrens

12.3.3 Analyse des Outputs

12.4 Wie können kategoriale Variablen einbezogen werden?

12.4.1 Distanzmatrix

12.4.2 Fuzzy-Verfahren

12.4.3 Outputteile

12.4.4 Analyse des Outputs

12.5 Zusammenfassung der Konzepte

12.6 Übungen

12.7 R-Befehle im Überblick

## Kapitel 13 Mehrdimensionale kategoriale Daten

13.1 Modelle für Kreuztabellen

13.1.1 Modelle und erwartete Häufigkeiten

13.1.2 Modellhierarchie

13.1.3 Loglineare Modelle

13.2 Welche Modelle beschreiben die Eintragungen einer Kreuztabelle

# Inhaltsverzeichnis

gut?

13.2.1 Pearson- und Likelihood-Ratio-Statistik

13.2.2 Modellberechnung in R

13.2.3 Modellauswahl

## 13.3 Höherdimensionale loglineare Modelle

13.3.1 Dreidimensionale Kreuztabellen

13.3.2 Loglineare Modelle in drei Dimensionen

13.3.3 Modellauswahl

13.3.4 Interpretation

## 13.4 Weitere Beispiele

13.4.1 Ein Modell mit einer Dreifachwechselwirkung

13.4.2 Ein Modell ohne Dreifachwechselwirkung

13.4.3 Ein Modell mit nur einer Zweifachwechselwirkung

13.4.4 Modelle ohne Wechselwirkungen

## 13.5 Abschließende Bemerkungen

13.5.1 Berechnung von Odds und Odds-Ratios

13.5.2 Modelle mit mehr als drei Variablen

13.5.3 Vergleich mit logistischer Regression

## 13.6 Zusammenfassung der Konzepte

## 13.7 Übungen

## 13.8 R-Befehle im Überblick

## Teil VI Appendix

Literaturverzeichnis

R GUI (Windows)

R-Funktionen

Index

Ins Internet: Weitere Infos zum Buch, Downloads, etc.

Copyright

# Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als **persönliche Einzelplatz-Lizenz** zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschließlich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs und
- der Veröffentlichung

bedarf der **schriftlichen Genehmigung** des Verlags. Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwortschutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: [info@pearson.de](mailto:info@pearson.de)

## Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. **Der Rechtsweg ist ausgeschlossen.**

## Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website herunterladen:

**<http://ebooks.pearson.de>**