Jürgen Groß

Grundlegende Statistik mit R

Jürgen Groß

Grundlegende Statistik mit R

Eine anwendungsorientierte Einführung
in die Verwendung der Statistik Software R

STUDIUM

VIEWEG+
TEUBNER

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<http://dnb.d-nb.de> abrufbar.

Dr. Jürgen Groß
E-Mail: gross@statistik.uni-dortmund.de

1. Auflage 2010

Lektorat: Ulrike Schmickler-Hirzebruch | Nastassja Vanselow

Vieweg+Teubner Verlag ist eine Marke von Springer Fachmedien.
Springer Fachmedien ist Teil der Fachverlagsgruppe Springer Science+Business Media.
www.viewegteubner.de

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg
Druck und buchbinderische Verarbeitung: MercedesDruck, Berlin
Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.
Printed in Germany

ISBN 978-3-8348-1039-7

Vorwort

Das Buch zeigt, wie die statistische Aufbereitung und Auswertung von Daten mit Hilfe des unter der GPL frei verfügbaren Paketes R vorgenommen werden kann. Auf der Basis von aufeinander aufbauenden Lerneinheiten wird das notwendige Rüstzeug vermittelt, um auch ohne vorherige Programmierkenntnisse statistische Auswertungen durchführen zu können. Dabei werden eine Reihe statistischer Methoden beispielhaft angewendet.

Zielgruppe sind Studierende unterschiedlicher Fachrichtungen, die sich im Rahmen ihres Studiums mit der statistischen Aufbereitung und Auswertung von Daten beschäftigen.

- Es soll gezeigt werden, wie statistische Auswertungen auf möglichst unkomplizierte Art und Weise durchgeführt werden können. Daher werden nicht alle Möglichkeiten, die R bietet, in diesem Buch Erwähnung finden. Insbesondere der Aspekt der Programmierung, der beispielsweise von Ligges (2008) abgehandelt wird, wird auf das notwendige Maß eingeschränkt. Stattdessen ist es das Ziel, aufzuzeigen, wie bestimmte statistische Methoden praktisch angewendet werden können.
- Die Verfahren werden anhand von Beispiel-Datensätzen erläutert, die von R selbst zur Verfügung gestellt werden. Die Beispiele sind in den meisten Fällen in Form einer Aufgabe gestellt, deren mögliche Lösung unter Verwendung von R dann jeweils diskutiert wird.
- Das Buch hat den Charakter eines Kurses. Es kann als ein Begleiter und sinnvolle Ergänzung für Vorlesungen verwendet werden, die grundlegende statistische Methoden behandeln, wie sie z.B. in Fischer (2005), Schlittgen (2003) oder Fahrmeier et al. (2003) abgehandelt werden.

Die ersten drei Kapitel (Teil 1) des Buches bieten eine Einführung in den grundlegenden Umgang mit R. Kapitel 4 bis 18 (Teil 2) beschäftigen sich mit dem Umgang mit Daten, der Datenanalyse, sowie grundlegenden statistischen Verteilungsmodellen. Kapitel 19 bis 24 (Teil 3) beinhalten weiterführende Methoden der Varianz- und Regressionsanalyse, sowie den grundlegenden Umgang mit Methoden der Zeitreihenanalyse.

- Der größte Vorteil von R liegt natürlich in der freien Verfügbarkeit unter GPL. Nachdem man sich an die Bedienung etwas gewöhnt hat, ist es recht unkompliziert eigene oder fremde Datensätze zu laden und aufzubereiten.
- Ein weiterer Vorteil liegt darin, dass es eine große Gemeinschaft von Beitragenden gibt. Fehlerhafte Prozeduren werden schnell als solche erkannt und korrigiert. Zu-

dem gibt es zu nahezu jedem statistischen Themengebiet frei verfügbare Pakete, welche entsprechende Methoden bereits zur Verfügung stellen.

- Schließlich ist R auch eine vollständige komplexe Programmiersprache. Der fortgeschrittene Anwender kann damit eigene Methoden implementieren, oder auch bestehende Funktion für eigene Zwecke geeignet ändern.

Einsteigern oder Wiedereinsteigern fällt es zunächst meist schwer, herauszufinden, welche Möglichkeiten von R zur Bearbeitung bestimmter statistischer Fragestellungen zur Verfügung gestellt werden. Hier ist es die Absicht und Hoffnung des Buches, sinnvolle Hilfestellungen zu geben und als nützliches Nachschlagewerk dienen zu können.

Mein Dank gilt den Autoren von R, siehe R Development Core Team (2009), sowie den zahlreichen Beitragenden. Zudem möchte ich mich bei all denjenigen Studierenden bedanken, deren Lernbereitschaft und motivierendes Interesse für die Anwendung statistischer Methoden mit Hilfe von R für mich eine Anregung dargestellt haben, dieses Buch zu schreiben. Schließlich danke ich Annette Möller, die durch intensives Korrekturlesen zu zahlreichen Verbesserungen beigetragen hat. Sämtliche verbleibenden Fehler und Irrtümer gehen zu Lasten des Autors.

Dortmund, im Februar 2010 Jürgen Groß

Inhalt