Angewandte Statistik mit R

Reiner Hellbrück

Angewandte Statistik mit R

Eine Einführung für Ökonomen und Sozialwissenschaftler

3. Auflage

Springer Gabler

Reiner Hellbrück
Fakultät Wirtschaftswissenschaften
Hochschule für angewandte Wissenschaften
Würzburg-Schweinfurt, Deutschland

ISBN 978-3-658-12861-6 ISBN 978-3-658-12862-3 (eBook)
DOI 10.1007/978-3-658-12862-3

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

Springer Gabler

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer Gabler ist Teil von Springer Nature
Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH

Vorwort zur dritten Auflage

Die Einleitung und Kapitel 2 wurden überarbeitet; bei letzterem wurde der Abschnitt über LimeSurvey und Moodle entfernt. Wegen seiner Bedeutung im Wirtschaftsleben ist die Berechnung des geometrischen Mittels aufgenommen worden. Da das Paket 'QRMlib' für R , Version 3.2.3, nicht verfügbar ist, wurde in Kapitel 3 eine entsprechende Anpassung vorgenommen.

In Kapitel 4 wurde der Abschnitt 'Einfache und zusammengesetzte Hypothesen' eingefügt. Die einseitigen Tests wurden auf den Standardfall, dass das Gleichheitszeichen in der Nullhypothese steht, umgestellt. Leicht geändert wurde die Gliederung des Kapitels 6 und zusätzliche Erläuterungen in Abschnitt 6.5 wurden eingefügt. Erläuterungen zur Kovarianz wurden in Kapitel 7 vorgenommen. Die Einleitung des Kapitels 8 beginnt nun mit einem Praxisbeispiel, in der Einleitung des Kapitels 12 wurde ebenfalls die praktische Bedeutung der Regressionsanalyse etwas hervorgehoben.

In Anhang C wird nun zusätzlich erläutert, wie in einzelnen Paketen vorhandene Daten eingelesen werden können und wie man Updates einzelner Pakete erhalten kann. Um Wechselkursdaten herunterzuladen gibt es Oanda, hierzu wird nun auch eine Alternative, der Service 'TrueFX', vorgestellt. Die Befehle dieses Kapitels wurden zudem auf Funktion geprüft.

In der jüngsten Zeit hat Herr Florian Schuberth, M.Sc., die Veranstaltung in parallelen Gruppen zu mir gehalten. Diskussionen, die sich hierbei mit ihm ergaben, sind in die neue Auflage eingeflossen. Vielen Dank, das machte Laune. Und auch Frau Hasenbalg, Lektorin beim Verlag Springer Gabler, möchte ich gerne für die gute Zusammenarbeit danken. Alle verbliebenen Fehler gehen, wie immer, zu meinen Lasten.

Würzburg, im März 2016

Reiner Hellbrück

Vorwort zur ersten Auflage

Dieses Buch entstand im Zuge der Neustrukturierung meiner Statistikveranstaltungen an der FH Würzburg-Schweinfurt. Die fortschreitende Digitalisierung macht auch vor der Statistik nicht halt und so entstand der Wunsch, die Veranstaltungen neu auszurichten. Wie an Fachhochschulen üblich, liegt der Schwerpunkt auf der Anwendung. Aus diesem Grund sind im allgemeinen nach einer kurzen Darstellung des nötigen Hintergrundwissens Beispiele angefügt. Hierbei kommt die Statistiksoftware R zum Einsatz.

R wird sehr selektiv eingesetzt, allen Anwendungen ist zuvor ein Grundlagenkapitel vorgeschaltet, in dem Maßzahlen, Teststatistiken, Hypothesen und dergleichen vorgestellt werden. Leser, die einen schnellen Überblick über die Software wünschen, seien auf den Anhang C verwiesen, wo die wichtigsten Befehle dargestellt werden. Um Mißverständnissen vorzubeugen, sei ausdrücklich darauf hingewiesen, dass es sich hier um keine Einführung in das Programmpaket R handelt: Methoden und Anwendung, gestützt mit Software, stehen gleichberechtigt nebeneinander. Das Erlernen der Software ergibt sich als nützlicher Nebeneffekt.

Die anfänglichen Rechnungen erfolgten mit Version R-2.5, dann mit neueren. Die Software ist auf vielen verschiedenen Betriebssystemen lauffähig. Im vorliegenden Fall wurde Windows XP und Suse-Linux genutzt. Bei Linux wurden teilweise Rechnungen mit Hilfe einer Shell, (auch Konsole oder Befehlsfenster genannt) größtenteils aber mit Emacs-ess durchgeführt. Emacs ist ein Text-Editor, der üblicherweise mit jeder Linux-Distribution ausgeliefert wird. Das Kürzel 'ess' steht für 'emacs speaks statistics' und will heißen, dass das Zusatzwerkzeug 'Emacs-ess' als Benutzeroberfläche (als 'frontend') für Statistiksoftware eingesetzt werden kann. Hierüber ist es möglich, mit einer einheitlichen Benutzeroberfläche verschiedene Statistikprogramme, darunter auch 'SPSS' und 'Stata', anzusprechen. Eigene Versuche in dieser Richtung wurden von dem Autor bislang nicht unternommen.

Der Einstieg ist sehr einfach gehalten, um dem Studenten während der ersten Wochen genügend Zeit zu lassen, die neue Software auf seinem eigenen Rechner zu installieren und kennenzulernen. Erfahrungsgemäß stellen sich bereits bei dem Einlesen der Daten die ersten Probleme ein. Dies rührt aus der Verwendung unterschiedlicher Parameter, die zur Trennung von Zeichen bei Textdateien verwendet werden. Desweiteren gibt es üblicherweise Probleme durch die Verwendung unterschiedlicher Betriebssysteme. R ist primär für Linuxsysteme geschrieben. Hier gelten jedoch etwas andere Konventionen bei der Angabe von Pfaden: statt des '\', wie in Windowssystemen üblich, wird das Zeichen '/' verwendet.

Werden die Befehle nicht direkt in dem Befehlsfenster von R geschrieben, sondern in einem Textverarbeitungsprogramm, so kann es nach Kopieren der Befehle in das Befehlsfenster leicht zu Fehlermeldungen kommen. Ursache ist dann häufig die automatische Ersetzung der Anführungszeichen in typographische Anführungszeichen innerhalb des Textverarbeitungsprogramms. Deshalb

wird empfohlen, zum Schreiben oder Bearbeiten von Befehlen eine Software zu verwenden, die solche automatischen Ersetzungen nicht vornimmt, oder dass solche Funktionen ausgeschaltet werden.

Der deskriptiven Statistik ist vergleichsweise wenig Raum gewidmet, der Schwerpunkt liegt auf der schließenden Statistik und multivariaten Verfahren, bei denen seitens Ökonomen (speziell meiner Kolleginnen und Kollegen) Nachfrage besteht. Der Text kann, je nach Belieben, unterschiedlich verwendet werden. Einerseits besteht die Möglichkeit, die Theorie weitestgehend in den Hintergrund zu drängen, um sich ausschließlich auf die Anwendung zu konzentrieren: die Kapitel 6 und 7 zur Wahrscheinlichkeitstheorie können dann übersprungen werden. Dies bietet sich an, wenn eine Veranstaltung zur Wahrscheinlichkeitstheorie vorgeschaltet ist.

Andererseits ist es möglich, Inhalte anwendungsnah zu präsentieren, und bei Bedarf nötiges Wissen in Wahrscheinlichkeitstheorie einzuflechten. Dann bietet es sich an, die Kapitel in der angegebenen Folge zu besprechen. Da die Kapitel 10 und 11, ohne statistische Tests auskommen, können sie auch zur Veranschaulichung multivariater Verfahren vorgezogen werden.

Bei einigen Lehrbüchern hat sich zwischenzeitlich die Unart eingeschlichen, während des laufenden Textes nicht zu zitieren. Es scheint, als habe ein sehr bekanntes Lehrbuch der Mikroökonomie, diese Entwicklung eingeleitet. Dem Autor des Lehrbuches verbrannte das Manuskript mitsamt der Zitate. Aus den verbliebenen Resten wurde es fast gänzlich ohne Zitate fertiggestellt. Hierdurch wird dem Studenten der Eindruck vermittelt, als brauche man nicht zu zitieren. Diesem Zeitgeist wird hier nicht gefolgt. Es wird angegeben, woher der Autor seine Weisheiten hat.

Dank schulde ich vielen, insbesondere meinem akademischen Lehrer Prof. Dr. Volker Steinmetz, der es außerordentlich gut verstand, theoretische Statistik und Ökonometrie zu vermitteln. Herr Prof. Dr. Rudolf Richter bot bereits in den 80-er Jahren PC-gestützte ökonometrische Auswertungen an, damals ein Novum. Beide Ansätze werden hier miteinander verknüpft. Danken möchte ich an dieser Stelle auch meinem wissenschaftlichen Mitarbeiter Manuel Hertel, für die gute Zusammenarbeit und die Entlastung durch seine Übungsstunden, die er mit großer Umsicht anbietet. Schließlich möchte ich bei meinen Söhnen, David und Simon um Nachsicht bitten, für die Zeit, die ich in meinem Arbeitszimmer den PC blockiert habe. Meine Frau genoß die Zeit, während ich 'aufgeräumt' war, ebenso wie ich.

Würzburg, im Juni 2009: Reiner Hellbrück

Vorwort zur zweiten Auflage

Kleinere Veränderungen sind vorgenommen worden. So wurde die Bedeutung der Messbarkeit besser herausgearbeitet. In Kapitel 3 ist die logarithmische Skala hinzugefügt worden und bei der Regression wird die Thematik der Kointegration angesprochen. In Kapitel C wurde eine weitere Möglichkeit zur Installation zusätzlicher R -Pakete in Unixsystemen eingefügt. Zudem wurden einige Internetadressen und Befehle aktualisiert sowie Schreibfehler der 1. Auflage korrigiert. Ein herzliches Dankeschön geht an zwei meiner Studenten, Herrn Daniel Back und Herrn Felix Kreß, die mich freundlicherweise auf Druckfehler hingewiesen haben. Alle verbliebenen Fehler gehen selbstverständlich zu meinen Lasten.

Freundlichst wird darauf hingewiesen, dass die verwendeten Daten von der Homepage des Verlages heruntergeladen werden können. Hierbei handelt es sich um eine *.zip-Datei. Die enthaltenen Dateien müssen zuerst entpackt werden, damit R darauf zugreifen kann. Insbesondere für Dozenten finden sich zusätzliche Materialen; so werden beispielsweise alle Abbildungen zur Verfügung gestellt.

Gerne bin ich bereit, eine Befragung mit LimeSurvey zu ermöglichen. Das Programm ist zwar kostenfrei, doch seine Installation auf einem Server, die Nutzung und die Einrichtung von Nutzungsrechten verursachen Kosten. Aktuelle Konditionen erhalten Sie auf Anfrage. Senden Sie bei Interesse eine E-Mail an

reiner.hellbrueck@fhws.de.

Um alle Funktionen des Programms verfügbar zu haben, ist es notwendig, zumindest R 2.11 zu installieren. Zudem kann es notwendig sein, Pakete zu aktualisieren; ansonsten kann es zu Fehlermeldungen kommen. So ist die logarithmische Skalierung beispielsweise in älteren Distributionen nicht enthalten.

Herr Christian Schuld hat mich freundlicherweise bei der Beschaffung von Literatur unterstützt. Text, Layout, Stichwortverzeichnis, Glossar und Literaturverzeichnis wurden, wie an Fachhochschulen meist der Fall, selbst erstellt. LaTeXhat hier wertvolle Dienste geleistet.

Würzburg, im Oktober 2010: Reiner Hellbrück

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis