

Rezension

**Alexandrowicz, Rainer W.
(2013). *R in 10 Schritten.
Einführung in die statistische
Programmierungsumgebung***

Wien: facultas.wuv. 304 S., 27,99 EUR,
ISBN 978-3-8252-8484-8.

Das Buch „R in 10 Schritten. Einführung in die statistische Programmierungsumgebung“ von Rainer W. Alexandrowicz ist eine Einführung in die freie statistische Programmierungsumgebung R.

Alexandrowicz ist Assistenzprofessor an der Alpen-Adria-Universität Klagenfurt und lehrt dort Statistik und Testtheorie am Institut für Psychologie, wo er auch IT-Sprecher ist.

Das Buch von Alexandrowicz stellt die Statistiksoftware R vor. Ein solches Werk ist insofern interessant, als dass die Software, die zwar bereits vor 20 Jahren entwickelt wurde, erst in den letzten Jahren vermehrt Einzug in sozial- und geisteswissenschaftliche Forschungsbereiche erhalten hat. Insbesondere für die universitäre Lehre macht die *open source*-Idee, die dem Programm zugrunde liegt, den Einsatz attraktiv.

Im Gegensatz zu anderen Statistikprogrammen ist R eine Programmiersprache, die für rechnerische und grafische Statistikauswertungen eingesetzt wird und die sich vor allem dadurch auszeichnet, dass sich der Anwender ergänzend zur gegebenen Grundfunktionalität

individuell Erweiterungen installieren kann, die zusätzliche (z. B. fachspezifische, wissenschaftlich aktuelle) Anwendungen möglich machen. Die Anzahl dieser Erweiterungen erhöht sich kontinuierlich, so dass R in immer mehr Anwendungsfeldern die gleiche oder sogar darüber hinausgehende Funktionalität aufweist wie kommerzielle Software. Der Aufbau und die Flexibilität von R setzen beim Anwender ein gewisses Statistikverständnis voraus, da statistische Prozeduren nicht durch reines Klicken auf Buttons erstellt werden können.

Weil der Umgang mit R als Statistiksoftware selten Teil der akademischen Ausbildung in sozial- und geisteswissenschaftlichen Fachrichtungen ist, muss diese Lücke mit Anleitungsliteratur zum Selbstlernen geschlossen werden. Ein solches Werk stellt auch das hier diskutierte Buch von Alexandrowicz dar. Der Autor verfolgt aber nicht die Absicht, im Stile eines „Kochbuches“ einzelne Verfahren zu fokussieren und ihre Umsetzung vorzuführen, sondern möchte „dem Anfänger das Funktionsprinzip von R nahebringen und ihn so in die Lage versetzen, die Möglichkeiten von R uneingeschränkt für die eigene Fragestellung auszunutzen“ (S. 6). Dadurch soll der Leser befähigt werden, das Werkzeug R soweit zu verstehen, dass Informationen über spezifische statistische Verfahren

eigenständig eingeordnet und angewendet werden können.

Wie der Autor im Vorwort hervorhebt, wendet er sich nicht explizit an eine bestimmte Zielgruppe im Sinne einer Fachrichtung, sondern allgemein an Anfänger, bei denen er „keine technischen Detailkenntnisse“ und „keine Programmierkenntnisse“ voraussetzt. Es ist vielmehr sein Ziel, „die Grundlagen für solche verständlich zu machen“ (S. 6). Die Kapitel des Buches, konsistent zum Buchtitel „Schritte“ genannt, widmen sich dann auch überwiegend eben dieser Vermittlung grundlegender Funktionsweisen von R; lediglich in einem Kapitel geht es um konkrete „grundlegende statistische Prozeduren“. Viele Kapitel enthalten gekennzeichnete Abschnitte mit vertiefenden Informationen. Abgerundet werden alle Kapitel mit Übungsaufgaben, mit denen die gelesenen Beispiele aktiv nachvollzogen und die neuen Kenntnisse zur Aufgabenlösung eingesetzt werden können.

Im ersten Kapitel „Kennenlernen von R“ wird der Leser dort abgeholt, wo er als Anfänger vermutlich steht: Der Autor erklärt zunächst, wie das Programm heruntergeladen und installiert wird, wobei Screenshots der einzelnen Installationsschritte die Ausführungen unterstützen. Nach der Installation des Programms lernt der Leser die Oberfläche von R kennen und kann mit den Übungsaufgaben bereits die ersten Befehle umsetzen.

Im zweiten Schritt „Arbeiten mit R“ präsentiert der Autor, wie die wichtigsten Rechenoperationen in R umgesetzt werden. So lernt der Leser R als Taschenrechner kennen, was vor allem ein erstes Gefühl für Eingabe

(Aktion des Benutzers) und Ausgabe (Reaktion/Antwort von R) erzeugen soll. Außerdem zeigt der Autor verschiedene Möglichkeiten auf, wie und wo Hilfe zu finden ist – innerhalb des Programms und auf ergänzenden Internetseiten.

In Kapitel 3 „Variablen, Objekte und der *workspace*“ werden die „Regeln zur Handhabung von Objekten“ eingeführt. Da in R letztlich alle Strukturen (das können z. B. Daten, Ergebnisse oder Auswertungsroutinen sein) Objekte verschiedenen Typs sind, erklärt Alexandrowicz dem Leser nun wie diese Objekte angelegt, ausgegeben, benannt und gespeichert werden können. Ein kurzer Absatz „für Umsteiger“ ergänzt an dieser Stelle, welche Parallelen und Unterschiede zu den Statistikprogrammen SPSS und SAS bestehen.

Die darauffolgenden beiden Kapitel „Objekttypen“ und „Datenstrukturen in R“ können als Fortsetzung von Kapitel drei verstanden werden. Nachdem der Leser mit der ersten groben Handhabung von R-Objekten vertraut gemacht wurde, wird jetzt in die Unterscheidung einfacher Objekte nach Inhalt eingeführt, um diese Kenntnisse weiter auf die Unterscheidung komplexerer Objekte nach ihrer Form zu erweitern. Die meisten empirisch erhobenen Daten liegen in Form von Matrizen (wenn nur Zahlen enthalten sind) oder sogenannter Dataframes (Zahlen und z. B. Text) vor. Alexandrowicz vermittelt dem Leser wie diese Datenstrukturen in R gehandhabt werden können.

Nachdem diese Grundlagen erklärt wurden, kommen die Kapitel 6 „Datenimport und -export“ und 7 „Datenaufbereitung“ der konkreten Anwendung des Lesers näher. Der Autor

beschreibt, wie spezifische Datenformate eingelesen werden können. Hier wird explizit auf den Import von Daten aus SPSS und EXCEL eingegangen. Im Folgenden geht es um fehlende Werte, das Rekodieren von Variablen und das Sortieren und Zusammenführen von Datensätzen.

Im Kapitel „Erstellung von Grafiken“ widmet sich der Autor dann der Durchführung deskriptiver, grafischer Auswertungsmöglichkeiten. Hier bietet R große Flexibilität, was die Erstellung von Standardgrafiken aber eben auch von individuellen, anwendungsspezifischen Abbildungen ermöglicht. Jeder Grafikparameter kann bei Bedarf einzeln verändert werden. So kann in R sowohl die Datenauswertung als auch die grafische Aufbereitung der Ergebnisse nach allen didaktischen, fachlichen und kreativen Ansprüchen umgesetzt werden.

In Kapitel 9 führt Alexandrowicz dann „Grundlegende statistische Prozeduren“ vor, also einfache Verfahren deskriptiver und schließender Statistik. Damit bekommt der Leser einen Einblick in die konkrete statistische Auswertung mit R. Mit den bis zu diesem Kapitel erworbenen Kenntnissen soll der Leser nun in der Lage sein, zusätzliche Literatur und Hilfeseiten mit Erklärungen ohne große Mühe zu verstehen und neue Funktionen anzuwenden, um so das Werkzeug R für komplexere Anwendungen zu verwenden.

Abgerundet wird das Buch mit dem Kapitel „Grundzüge des Programmierens in R“, das einige hilfreiche Details enthält, wie individuelle Funktionen geschrieben und bestimmte Routinen wohldurchdacht formuliert werden können, um beispielsweise die

Auswertungsgeschwindigkeit zu erhöhen.

An dieser Stelle sei deutlich darauf hingewiesen, dass Alexandrowicz' Buch sich in seinen Erklärungen auf die Vorführung sehr einfacher Prozeduren beschränkt, mehrschrittige Auswertungsroutinen werden nicht erörtert. Dadurch wirken die erwähnten Kniffe, die vor allem für diese mehrschrittigen Auswertungen sinnvoll und hilfreich sind, im abschließenden Kapitel etwas verloren.

Zusammenfassend kann das Buch als Einführungsbuch für Anfänger beschrieben werden, die insbesondere an einem Einblick in die Grundlagen von R-Prozessen interessiert sind und sich über das losgelöste Durchführen einzelner statistischer Prozeduren hinaus ein Verständnis von R aneignen wollen. Das Studium des Buches soll den Leser befähigen selbstständig weiterzulernen und seine Kenntnisse auszubauen, wozu Alexandrowicz alle nötigen Werkzeuge präsentieren möchte. Es sei aber darauf hingewiesen, dass es in dem Buch ausschließlich um die Software R geht. Auch wenn die meisten Anwender einen Editor zum Schreiben von R-Code benutzen, belässt es Alexandrowicz dabei, einige übliche R-Editoren tabellarisch aufzuführen. Das ist schade, da die Fähigkeit zur Bedienung eines Editors viele Vorteile bringt. Zum Beispiel trägt die automatische Formatierung (*syntax highlighting*) in einem Editor wesentlich dazu bei, Fehler zu vermeiden oder einfacher zu finden.

Durch den Aufbau der Kapitel als aufeinander aufbauende Schritte eignet sich das Buch vor allem dazu, es in der vorgegebenen Reihenfolge der Kapitel zu studieren und mit Hilfe der

Übungen kontinuierlich mitzulernen. Dabei sollte sich der Leser nicht von den teilweise sehr in die Tiefe gehenden Technikdetails abschrecken lassen. Nicht immer wird deutlich, nach welcher Systematik einzelne Kapitelabschnitte im Gegensatz zu anderen als vertiefend markiert wurden. Dem Leser sei empfohlen, sich nach eigenem Interesse mit den Zusatzinformationen, etwa zur internen Funktionalität eines Computers, zu beschäftigen oder diese zu überspringen.

Positiv hervorzuheben ist, dass der Autor ausdrücklich (wenn auch etwas versteckt) darauf hinweist, dass „in diesem Buch der *open source*-Gedanke gefördert werden soll“ (S. 167). So finden sich neben der Installationsanleitung unter Windows auch kurze Hinweise zur Installation unter Linux und alle im Buch erwähnten Editoren sind freie Produkte.

Äußerlich besticht das vom *facultas.wuv* Universitätsverlag herausgegebene Buch durch ein durchdachtes Layout, das sich in abgesetzten Kapiteleinleitungen, Einschüben, Schriftarten und -farben sowie den Abbildungen zeigt. Besonders die Schemata zur Veranschaulichung von

R-Strukturen verdienen eine Erwähnung, tragen sie doch maßgeblich zu einem schnellen Verständnis und „Aha“-Effekten bei. Lediglich bei den Übungen wäre noch eine Strukturierung wünschenswert gewesen.

Vereinzelt weisen die Beispiele Flüchtigkeiten auf, die die Nachvollziehbarkeit für R-Lernende erschweren. In einer Neuauflage des Buches sollte dies korrigiert werden.

Insgesamt kann das Buch „R in 10 Schritten“ von Rainer W. Alexandrowicz denjenigen empfohlen werden, die als technikaffine Nutzer R nicht nur anwenden, sondern auch in seiner Funktionsweise verstehen wollen.

Die Zielgruppe ist also nicht fachspezifisch. Der individuelle Nutzen ist sicherlich dann gegeben, wenn der Leser Zeit investieren möchte, um die zugrundeliegenden Prinzipien von R-Prozessen und technische Hintergrundinformationen zu studieren. Für Erklärungen und Beispiele zu einzelnen statistischen Verfahren muss hingegen ergänzende Literatur zu Rate gezogen werden.

Katharina Borg

Katharina Borg, Institut für Schulentwicklungsforschung, Technische Universität
Dortmund, Vogelpothsweg 78, 44227 Dortmund, Deutschland
E-Mail: borg@ifs.tu-dortmund.de