



Leseprobe aus: Luhmann, R für Einsteiger, ISBN 978-3-621-28249-9

© 2015 Beltz Verlag, Weinheim Basel

<http://www.beltz.de/de/nc/verlagsgruppe-beltz/gesamtprogramm.html?isbn=978-3-621-28249-9>

I Einleitung

Seit einiger Zeit macht die Statistiksoftware R auch außerhalb eingeweihter Statistiker-Kreise von sich reden. Aussagen wie „R macht so tolle Graphiken“ und „R kann sogar Strukturgleichungsmodelle“ hört man immer häufiger. Allerdings halten sich einige ungerechtfertigte Gerüchte hartnäckig, zum Beispiel „In R kann man seine Daten nicht sehen“ oder „R ist nur was für Programmierer“.

Diese und andere Vorurteile sollen mit diesem Buch aus der Welt geschafft werden. R ist viel leichter und schneller zu erlernen, als häufig geglaubt wird. Das spricht sich mittlerweile herum. R wird zunehmend auch von ganz normalen Anwendern benutzt. Jedes Jahr gibt es zudem mehr Fachbereiche, an denen nicht mehr kommerzielle Statistikprogramme gelehrt werden, sondern R. Mit Ihrer Entscheidung, R eine Chance zu geben, liegen Sie daher voll im Trend. Ich wünsche Ihnen dabei viel Erfolg und hoffentlich auch etwas Spaß!

Ach ja: In R kann man seine Daten sehen. Und Programmieren muss man nicht unbedingt können.

1.1 Warum R?

Es gibt viele Gründe, die für R sprechen. Hier sind die wichtigsten:

- ▶ **R kann mehr.** Die Software enthält viele Funktionen, die man bei kommerziellen Statistik-Programmen wie zum Beispiel SPSS vergeblich sucht.
- ▶ **R ist schnell.** Das Programm wird ständig weiterentwickelt, sodass viele neue statistische Methoden zuerst in R und erst Jahre später in anderen Programmen implementiert sind.
- ▶ **R ist ansprechbar.** Die Programmierer der einzelnen Pakete sind über E-Mail erreichbar und reagieren meiner Erfahrung nach innerhalb weniger Tage. So können Fragen schnell geklärt und eventuelle Programmierfehler behoben werden.
- ▶ **R schläft nicht.** R-Nutzer gibt es auf der ganzen Welt, und viele von ihnen sind in der Mailingliste eingetragen. So kann man seine Fragen rund um die Uhr klären.
- ▶ **R ist kostenlos!**

1.2 Für wen ist dieses Buch?

Dieses Buch richtet sich sowohl an Einsteiger, die zum ersten Mal mit einer Statistiksoftware arbeiten, als auch an Umsteiger, die R als eine Alternative zu anderen Statistikprogrammen ausprobieren möchten. Besondere Computer- oder gar Programmierkenntnisse brauchen Sie nicht. Da dieses Buch aber keine Statistik-Einführung ist, sollten Sie die Grundlagen der hier behandelten statistischen Tests kennen. Dafür empfehle ich das Lehrbuch von Eid, Gollwitzer und Schmitt (2015), an dem auch die Gliederung dieses Buchs orientiert ist.

Ich bin selbst Psychologin und habe daher Datenbeispiele aus psychologischen Studien gewählt. Darüber hinaus behandle ich einige Funktionen und statistische Verfahren, die besonders für Psychologinnen und Psychologen von Interesse sind, z. B. die Berechnung von Skalenwerten (Kap. 7) oder Verfahren für die Testkonstruktion wie Itemanalysen und Faktorenanalysen (Kap. 19). Grundsätzlich richtet sich das Buch jedoch an Interessierte aus allen Fachrichtungen der Sozialwissenschaften.

Leider ist in diesem Einführungsbuch kein Platz, um komplexere Graphiken oder spezielle statistische Verfahren wie Modelle der Item Response Theorie detailliert zu besprechen. Sie werden jedoch in der Lage sein, sich in die Funktionen und Pakete, die Sie für solche Verfahren brauchen, selbst einzuarbeiten. Weitere Hinweise dazu finden Sie im Anhang A: Datensätze.

1.3 Wie benutzt man dieses Buch?

Den Umgang mit einer Statistiksoftware lernt man am besten durch Ausprobieren. Daher empfehle ich, bei der Lektüre dieses Buchs immer einen Computer mit R dabei zu haben. So können Sie die beschriebenen Funktionen direkt ausprobieren. Die Datensätze dafür können Sie von der Webseite zu diesem Buch herunterladen: <http://www.beltz.de/r-fuer-einsteiger>. Dort finden Sie auch die Skripte für jedes Kapitel, die oftmals noch weitere Beispiele enthalten, sowie zahlreiche Zusatzmaterialien.

Hinweise für Einsteiger

Ich habe mich bemüht, dieses Buch so zu schreiben, dass auch Einsteiger ohne jegliche Erfahrung mit Statistik-Software mit R zurechtkommen können. In diesem Buch stelle ich die statistischen Verfahren in der Reihenfolge vor, in der sie typischerweise gelehrt werden. Wenn Sie dieses Buch also in Begleitung zu einer Statistik-Vorlesung in Psychologie oder anderen sozialwissenschaftlichen Fächern lesen, können Sie die Kapitel einfach von vorne nach hinten durcharbeiten.

Hinweise für Umsteiger

Als Umsteiger haben Sie bereits Erfahrung mit anderen kommerziellen Statistik-Programmen und interessieren sich möglicherweise nur für bestimmte Funktionen in R.

In diesem Fall brauchen Sie natürlich nicht das Buch von vorne bis hinten durchzuarbeiten. Beginnen Sie stattdessen mit Kapitel 23. Dieses Kapitel ist ein Crash-Kurs für Umsteiger und dient als Wegweiser durch das Buch. In diesem Kapitel werden auch die wesentlichen Unterschiede zwischen R und SPSS vorgestellt.

1.4 Weiterentwicklungen und Aktualität des Buchs

R wird ständig weiterentwickelt. Sowohl die Basisversion als auch die zusätzlichen Pakete werden regelmäßig aktualisiert, und fast wöchentlich werden neue Pakete zur Verfügung gestellt. Wenn Sie regelmäßig mit R arbeiten, werden Sie vermutlich immer wieder neue Pakete entdecken, die Ihnen die statistischen Analysen noch leichter machen. Für die Erstellung dieses Buchs habe ich mich bemüht, immer die aktuellsten Versionen zu verwenden (Stand März 2015). Dieses Buch verwendet die R Version 3.1.3. Die hier vorgestellten Funktionen und Pakete sollten jedoch auch mit neueren Versionen weiterhin einwandfrei funktionieren. Wenn Sie trotzdem veraltete Funktionen oder gar Fehler finden, lassen Sie mich dies bitte wissen. Aktualisierungen und Fehlerkorrekturen werden regelmäßig unter <http://www.beltz.de/r-fuer-einsteiger> veröffentlicht. Und noch ein Tipp: Die aktuellsten Hinweise zu den Funktionen und den entsprechenden Argumenten finden Sie in der dazugehörigen Hilfedatei (s. Abschn. 3.3).

1.5 Verwendete Schriftarten

Um die Übersichtlichkeit des Texts zu erhöhen, werden in diesem Buch verschiedene Schriftarten verwendet (s. Tab. 1.1).

Tabelle 1.1 Verwendung von Schriftarten im Buch

Text	Beispiel
Menübefehle	DATEI → SPEICHERN
Namen von Paketen	psych-Paket
Namen von Funktionen	mean-Funktion
Befehle	> c(1, 2, 3)

2 Installation

R ist als kostenlose Software im Internet verfügbar. Alles was wir brauchen, um loszulegen, ist also ein Internetzugang. Um R zu installieren, muss man Schreibrechte für das Laufwerk haben, auf dem R installiert werden soll. Auch für die Arbeit mit R sollte man Schreibrechte für das Laufwerk haben, da man so weitere Pakete herunterladen und installieren kann (zur Rolle von Paketen s. Abschn. 2.3).

2.1 Download

Die Internetseite cran.r-project.org ist der Ausgangspunkt für die Arbeit mit R. Die Abkürzung CRAN steht für Comprehensive R Archive Network. Hier werden sowohl die Software als auch verschiedene Zusatzdateien (so genannte Pakete, s. Abschn. 2.3) und Dokumente zur Verfügung gestellt. Direkt auf der Startseite dieser Webseite befindet sich die Option *Download and Install R*. Hier kann man R für die Betriebssysteme Linux, MacOS X und Windows herunterladen.

Zum Download der **Windows-Version** gelangt man über den Link *Download R for Windows*. Auf der neuen Seite wählen wir den Link *base*. Auf der folgenden Seite klicken wir auf den Link *Download R X.X.X for Windows*. Damit laden wir die Setup-Datei herunter. (Das X steht für die aktuelle Versionsnummer. Auf der CRAN-Seite wird immer nur die neueste R-Version angeboten.)

Zum Download der **Mac-Version** gelangt man über den Link *Download R for MacOS X*. Wir gelangen nun auf die Seite *R for MacOS X*, wo wir die Datei *R-X.X.X.pkg* auswählen und herunterladen.

2.2 Installation

Die Installationsschritte unterscheiden sich etwas zwischen Windows und MacOS. Daher besprechen wir die Installation für die beiden Betriebssysteme getrennt.

2.2.1 Installation unter Windows

Wir starten die Installation von R, indem wir die heruntergeladene Setup-Datei öffnen. Als erstes müssen wir die Sprache für den Setup auswählen. Danach öffnet sich der Setup-Assistent. Dieser Assistent führt uns Schritt für Schritt durch die Installation. Wenn wir einen Schritt abgeschlossen haben, klicken wir auf WEITER. Wenn wir einen Schritt rückgängig machen möchten, klicken wir auf ZURÜCK und ändern die

Einstellungen. Die Installation kann jederzeit durch die Option **ABBRECHEN** beendet werden. Folgende Schritte werden bei der Installation durchlaufen:

- (1) Der Lizenztext wird angezeigt und sollte durchgelesen werden.
- (2) **ZIELORDNER AUSWÄHLEN**. Wir wählen den Ordner, in dem das Programm installiert wird. Standardmäßig wird der Ordner `C:\Programme\R\R-X.X.X` vorgeschlagen.
- (3) **AUSWAHL VON KOMPONENTEN**. Hier wählen wir die Dateien aus, die installiert werden sollen. Für die meisten Leser wird die Option **BENUTZERINSTALLATION** geeignet sein.
- (4) **STARTOPTIONEN ANPASSEN**. Wählen wir die Option **JA**, können wir im Folgenden bestimmte Einstellungen von R bei der Installation festlegen. Dazu zählen der Anzeigemodus (Darstellung der einzelnen Fenster), der Hilfestil (Hilfdateien als PDF oder im Windows-Format) sowie der Internetzugang. Wählen wir **NEIN**, werden die einzelnen Schritte übersprungen und die Standardeinstellungen für R übernommen. Wir wählen die letztere Option.
- (5) **STARTMENÜORDNER AUSWÄHLEN**. Hier kann man entscheiden, ob ein Ordner im Startmenü erstellt werden soll und den Namen im Startmenü ändern.
- (6) **ZUSÄTZLICHE AUFGABEN AUSWÄHLEN**. Wir entscheiden nun, ob zusätzliche Symbole auf dem Desktop oder in der Schnellstartleiste angelegt werden sollen. Außerdem legen wir hier die Art der Verknüpfung von R mit Windows fest (Speicherung der Versionsnummer, R mit `.RData`-Dateien verknüpfen). Für die meisten Leser sind die Standardeinstellungen ausreichend.
- (7) Wir haben es geschafft: Das Programm wird installiert!

2.2.2 Installation auf MacOS

Um R auf dem Mac zu installieren, öffnen wir die Setup-Datei, die wir von der CRAN-Seite heruntergeladen haben (s. Abschn. 2.1). Dadurch öffnet sich der Setup-Assistent. Dieser Assistent führt uns Schritt für Schritt durch die Installation. Bei der Installation werden die folgenden Schritte durchlaufen:

- (1) Wichtige Informationen lesen und zustimmen.
- (2) Software-Lizenzvereinbarungen lesen und akzeptieren.
- (3) Zielvolumen wählen: Soll die Software für alle Benutzer oder nur auf einem bestimmten Laufwerk installiert werden?
- (4) Installationstyp: Hier kann der Ort der Installation verändert werden.
- (5) Durchführung der Installation.
- (6) Fertig!

An dieser Stelle noch ein allgemeiner Hinweis: Dieses Lehrbuch wurde mit Windows erstellt. Die Benutzeroberfläche sieht auf dem Mac manchmal etwas anders aus. Grundsätzlich lassen sich aber alle Funktionen an denselben Stellen wiederfinden. Alle Pakete lassen sich auch auf MacOS verwenden.

2.3 Zusätzliche Pakete

Bisher haben wir nur die Basisversion von R installiert. Mit dieser Version können wir schon ganz nette Graphiken erstellen und die wichtigsten statistischen Tests durchführen. Für komplexere statistische Analysen ist die Basisversion jedoch nur bedingt geeignet. Hier kommen die zusätzlichen Pakete ins Spiel: Pakete sind Dateien, die man gesondert installieren und laden muss. Sie enthalten Zusatzfunktionen, die die Basisfunktionen von R ergänzen.

Die Pakete werden von hunderten Personen auf der ganzen Welt programmiert und zur Verfügung gestellt. Mittlerweile sind schon über 6000 Pakete verfügbar (Stand März 2015)! Durch dieses Prinzip sind neu entwickelte statistische Methoden meistens als erstes in R verfügbar, Jahre bevor sie in kommerziellen Statistikprogrammen implementiert werden. Auf der Webseite cran.r-project.org sind alle verfügbaren Pakete aufgelistet und beschrieben. Um Pakete nutzen zu können, müssen zwei Schritte durchgeführt werden: Die Pakete müssen einmalig installiert werden (Abschn. 2.3.1) und die Pakete müssen in jeder Sitzung neu geladen werden (Abschn. 2.3.2).

Tip

Eine Liste der beliebtesten Pakete sowie die neuesten Aktualisierungen erhält man unter <http://www.crantastic.org/>.

2.3.1 Pakete installieren

Die Installation von Paketen unterscheidet sich zwischen den Betriebssystemen Windows und MacOS, daher zeigen wir hier die Installation für beide Betriebssysteme. Zur Veranschaulichung installieren wir auf beiden Betriebssystemen das Paket `psych` (Revelle, 2015). Dieses Paket wurde von dem Persönlichkeitspsychologen William Revelle programmiert und enthält eine Reihe nützlicher Funktionen, die den Umgang mit R erleichtern, zum Beispiel die Funktion `corr.test`, die die Erstellung von Korrelationsmatrizen vereinfacht (s. Kap. 10). Zur Erinnerung: Um Pakete installieren zu können, müssen wir Schreibrechte auf dem entsprechenden Laufwerk haben.

Windows

Um Pakete zu installieren, öffnen wir R und wählen in der Menüleiste die Option `PAKETE → INSTALLIERE PAKET(E)` aus. Es öffnet sich ein Fenster, in dem wir einen CRAN-Mirror auswählen. Dabei handelt es sich um einen Server, auf dem die Dateien gespeichert sind. Im deutschsprachigen Bereich stehen zurzeit sechs Mirrors zur Verfügung: Austria, Germany (Berlin), Germany (Göttingen), Germany (Frankfurt),

Germany (Münster) und Switzerland. Allgemein wird empfohlen, einen Mirror in der Nähe zu wählen.

Nachdem wir einen Mirror gewählt haben, erscheint ein weiteres Fenster mit einer alphabetisch sortierten Liste aller Pakete. Hier wählen wir das gewünschte Paket aus und klicken auf OK. Das Paket wird nun heruntergeladen und installiert.

Wir installieren nun das Paket `psych`. Sobald wir das Paket in der Liste ausgewählt und auf OK geklickt haben, erscheint der folgende Text in der R Konsole:

```
--- Bitte einen CRAN Spiegel für diese Sitzung auswählen ---  
versuche URL 'http://cran.uni-muenster.de/bin/windows/  
contrib/3.1/psych_1.5.1.zip'  
Content type 'application/zip' length 3198677 bytes (3.1 MB)  
URL geöffnet  
downloaded 3.1 MB
```

```
Paket 'psych' erfolgreich ausgepackt und MD5 Summen  
abgeglichen
```

```
Die heruntergeladenen Binärpakete sind in  
C:\Ordner1\Ordner2\downloaded_packages
```

Tipp

Wenn das gesuchte Paket nicht in der Liste aufgeführt wird, kann man es sich auch direkt von der CRAN-Homepage herunterladen. Auf dieser Seite gibt es links den Link `PACKAGES`. Für jedes Paket ist hier eine eigene Seite angelegt, auf der man u. a. die Installationsdateien und das Manual herunterladen kann. Windows-Nutzer laden die Zip-Datei herunter und gehen anschließend in R auf die Option `PAKETE → INSTALLIERE PAKETE AUS LOKALEN ZIP-DATEIEN`. Jetzt nur noch die Zip-Datei auswählen und die Installation wird durchgeführt.

MacOS

Wir öffnen die R Konsole und gehen auf `PAKETE UND DATENSÄTZE → PAKETINSTALLATION`. In dem Menüfenster sollte oben die Option `CRAN (BINARIES)` eingestellt sein. Wir geben nun den Namen des gesuchten Pakets ein und klicken dann auf `LISTE HOLEN`. Jetzt wird uns eine Liste mit allen Paketen angezeigt, die diesen Namen enthalten. Wir wählen das gewünschte Paket aus und klicken auf `INSTALLIEREN`.

Zur Veranschaulichung installieren wir nun das Paket `psych`. Wir geben den Namen `psych` oben rechts ein und erhalten dann eine Liste mit mehreren Paketen (Abb. 2.1). Wir wählen das Paket `psych` aus und klicken auf `INSTALLIEREN`. Das Fenster `R PAKETINSTALLATION` bleibt auch nach der Installation des Pakets geöffnet. Wir können jetzt noch weitere Pakete suchen und installieren. Wenn alle gewünschten Pakete installiert sind, schließen wir das Fenster.