

Einführung in R

Clemens Brunner

18.-19.10.2019

Grundlagen

Inhalt

In dieser Veranstaltung werden Sie das Statistikpaket R kennenlernen. Sie müssen dafür keinerlei Vorwissen mitbringen - die einzige Voraussetzung ist Interesse an den Themen Statistik, Datenanalyse und Programmieren. Folgende Themen werden behandelt:

- Installieren und Warten einer R-Umgebung
- Wichtige Datenstrukturen (Vektoren, Matrizen, Data Frames)
- Importieren von Daten aus verschiedenen Dateiformaten
- Berechnen diverser deskriptiver Statistiken
- Erstellen von statistischen Grafiken
- Berechnen von Korrelationen
- Vergleichen von Mittelwerten

Aufgrund der begrenzten Zeit ist es nicht möglich, alle für Sie vielleicht relevanten statistischen Verfahren abzudecken (wie zum Beispiel lineare Modelle, Varianzanalysen, lineare gemischte Modelle, Strukturgleichungsmodelle, und so weiter). Sie werden aber durch ein umfassendes Basiswissen selbst die Werkzeuge in die Hand bekommen, sich solche spezifischeren Methoden in R selbst anzueignen.

Überblick über R

Was ist R?

R ist eine Programmierumgebung für statistische Datenanalysen, Berechnungen und Grafiken. Wichtige Eigenschaften von R kann man mit folgenden Punkten zusammenfassen:

- Open Source (freie Software)
- Plattformübergreifend (läuft auf Windows, macOS, Linux)
- Populär im Bereich Datenverarbeitung und Statistik
- Relativ einfach zu erlernen
- Einfaches Arbeiten mit Vektoren, Matrizen, Tabellen und Listen
- Riesige Anzahl an Zusatzpaketen
- Umfangreiche Hilfe und Dokumentation
- Große und hilfsbereite Community
- Sehr gut geeignet für reproduzierbare Datenanalysen

Wie sieht R-Code aus?

R bedient man, indem man Befehle entweder in der sogenannten Console eingibt oder ein Script ausführt. Statistische Berechnungen werden daher nicht wie z.B. in SPSS über eine grafische Oberfläche durchgeführt. Dies ist keineswegs ein Nachteil, sondern hat ganz im Gegenteil mehrere Vorteile. Insbesondere wird die Datenanalyse dadurch reproduzierbar und wiederverwendbar. Typische Ein- und Ausgaben in R sehen beispielsweise wie folgt aus:

```
x <- c(1, 18, 12, 3, 8, 22, 12, 4, 13, 21, 27, 1)
mean(x)
```

```
[1] 11.83333
```

```
sd(x)
```

```
[1] 8.768055
```

```
set.seed(1)
```

```
y <- -0.1 * x + rnorm(12, mean=-10, sd=4)
```

```
cor.test(x, y)
```

Pearson's product-moment correlation

data: x and y

t = -1.0611, df = 10, p-value = 0.3136

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

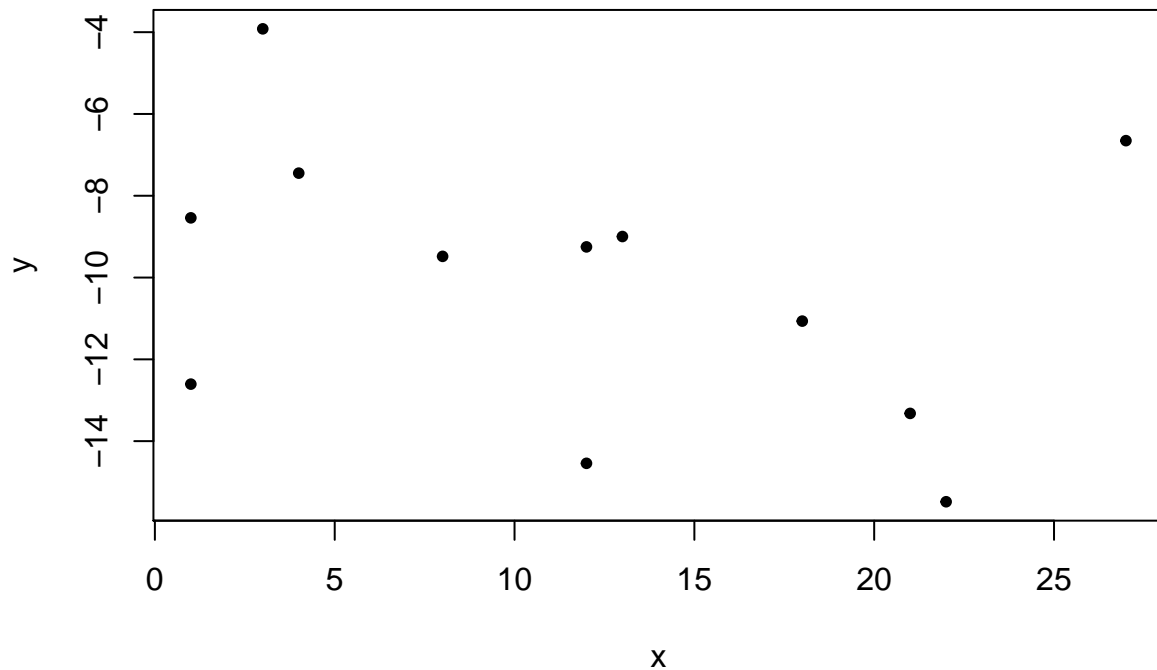
-0.7543040 0.3129176

sample estimates:

cor

-0.3181118

```
plot(x, y, pch=20)
```



Im Laufe dieser Veranstaltung werden wir auf die einzelnen Bestandteile dieses Codes eingehen.

Installation

R

Alle relevanten Informationen zur Installation von R sind auf der offiziellen Website zu finden. Hier finden sich detaillierte Anleitungen für jede unterstützte Plattform. Der Download-Link befindet sich in der linken Spalte ganz oben unter “Download” - “CRAN”. Es ist sinnvoll, stets mit der aktuellsten Version von R zu arbeiten (bei der Erstellung dieses Dokuments war dies Version 3.6.1).



RStudio

R ist modular aufgebaut und besteht aus einem Kern sowie einer grafischen Oberfläche (GUI¹). Obwohl man mit der mitgelieferten GUI relativ komfortabel arbeiten kann (vor allem unter macOS), hat sich dennoch die alternative Oberfläche RStudio als De-Facto-Standard durchgesetzt. RStudio ist eine reine GUI, d.h. es wird eine installierte R-Umgebung vorausgesetzt. Auch hier ist es ratsam, stets die aktuellste Version zu verwenden (bei der Erstellung dieses Dokuments war dies Version 1.2.5001).



Erste Schritte

Starten von R

Um R zu verwenden, startet man entweder die mitgelieferte grafische Oberfläche oder RStudio - unter Windows z.B. aus dem Startmenü, unter macOS aus den Anwendungen oder per Spotlight. Wir werden in dieser Veranstaltung ausschließlich RStudio verwenden.

Verwendung als Taschenrechner

R kann man als umfangreichen Taschenrechner verwenden. Die Eingabe von Rechenoperationen ist intuitiv, ein sogenannter Prompt (>) in der Console signalisiert, dass R bereit für Eingaben ist. Nach Eingeben eines Befehls und Bestätigen mit der Eingabetaste wird das Ergebnis der Berechnung ausgegeben (ignorieren wir vorerst die Zahlen in eckigen Klammern, die vor jeder Ausgabe stehen). Die Grundrechenarten funktionieren wie erwartet:

```
13 + 7
```

```
[1] 20
```

```
1001 - 93
```

```
[1] 908
```

```
81 * 76
```

```
[1] 6156
```

```
1563 / 43
```

```
[1] 36.34884
```

Für eine Ganzzahl-Division verwendet man den Operator %/%:

¹Graphical User Interface

```
1563 %% 43
```

```
[1] 36
```

Den Rest einer Division erhält man mit %%:

```
1563 %% 43
```

```
[1] 15
```

Potenzieren kann man mit ^ bzw. **:

```
16**2
```

```
[1] 256
```

```
16^2
```

```
[1] 256
```

R kennt selbstverständlich auch die korrekten Vorrangsregeln bei Verkettung von mehreren Operationen:

```
(13 + 6) * 8 - 12**2 / (2.5 + 1.6)
```

```
[1] 116.878
```

Die Quadratwurzel berechnet man mit sqrt:

```
sqrt(144)
```

```
[1] 12
```

Bitte beachten Sie, dass R die *englische* Zahlenschreibweise mit einem Punkt (.) als Dezimaltrennzeichen verwendet und nicht das im deutschen Sprachraum übliche Komma (,). Kommazahlen müssen daher immer mit einem Punkt eingegeben werden, ganz egal welche Sprache im Betriebssystem eingestellt ist.

Übungen

Übung 1

Installieren Sie die neuesten Versionen von R und RStudio auf Ihrem Rechner. Starten Sie dann RStudio und suchen Sie nach einer Möglichkeit, die Versionsnummern sowohl von R als auch von RStudio anzuzeigen (dies ist wichtig, weil die erste Frage bei Problemen normalerweise ist: “Verwenden Sie die aktuellsten Versionen?”).

Übung 2

Berechnen Sie mit R das Ergebnis der Division 619 : 65. Berechnen Sie außerdem das ganzzahlige Ergebnis und den Rest.

Übung 3

Gegeben seien folgende Messwerte: 11, 27, 15, 10, 32, 18, 25, 22, 29, 11. Berechnen Sie deren Mittelwert \bar{x} . Verwenden Sie dafür keine fertigen Funktionen, sondern nur Grundrechenarten.

Übung 4

Berechnen Sie in R das Ergebnis des folgenden Ausdrucks (achten Sie auf die richtige Klammersetzung):

$$\sqrt{2} \cdot \frac{(5^3 - 2) \cdot \sqrt{18}}{\left(\frac{7}{5} + 13.2\right) \cdot 7^{\frac{2}{3}}}$$



Diese Unterlagen sind lizenziert unter einer Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.