

R-Vorkurs

Kursinformationen

Hochschule für Technik und Wirtschaft Berlin

Finanzmathematik, Aktuarwissenschaften und Risikomanagement (M)

Sommersemester 2025

Tutor: Patrick Spohr

Termine

1. Einführung in R Programmierung
2. April 2025 um 13:00-19:00 Uhr, TGS 311
2. Statistik und Wahrscheinlichkeitstheorie in R
11. April 2025 um 9:45-15:30 Uhr, WH C 241

Was ist R?

R ist eine freie Programmiersprache und Softwareumgebung, die speziell für statistische Berechnungen, Datenanalyse und grafische Darstellungen entwickelt wurde. R eignet sich sowohl für den Einsteiger, der grundlegende Datenanalysen durchführen möchte, als auch für erfahrene Anwender, die komplexe statistische Verfahren implementieren wollen.

Hier einige wichtige Punkte zu R:

- **Statistische Analyse:** R bietet eine umfangreiche Sammlung an statistischen Funktionen und Modellen, die es ermöglichen, komplexe Analysen einfach und effizient durchzuführen.
- **Datenvisualisierung:** Mit R lassen sich hochwertige Grafiken und Diagramme erstellen, über Pakete wie `ggplot2` wird die Visualisierung besonders mächtig und flexibel.
- **Open Source und Erweiterbarkeit:** R ist unter der GNU General Public License veröffentlicht und kann frei genutzt und erweitert werden. Tausende von Zusatzpaketen (z.B. mittels CRAN) ermöglichen es, nahezu jeden Bedarf im Bereich Data Science abzudecken.
- **Anwendung in der Wissenschaft und Industrie:** R wird weltweit von Statistikern, Data Scientists und Forschern in den Natur-, Sozial- und Wirtschaftswissenschaften verwendet, um Daten zu analysieren und Modelle zu erstellen.

RStudio

RStudio ist eine beliebte integrierte Entwicklungsumgebung (IDE) für die Programmiersprache R. Es bietet eine benutzerfreundliche Oberfläche, die das Arbeiten mit R erleichtert, insbesondere für statistische Analysen, Datenvisualisierung und maschinelles Lernen.

Hauptmerkmale von RStudio:

- **Code-Editor:** Ein leistungsstarker Editor, der speziell für R-Code entwickelt wurde und Syntaxhervorhebung bietet.
- **Konsole:** Eine Schnittstelle zur Ausführung von R-Befehlen direkt in Echtzeit.
- **Umgebungspanel:** Ermöglicht die Überwachung und Verwaltung von Variablen, Daten und Objekten im Arbeitsbereich.
- **Plot-Panel:** Zeigt erstellte Grafiken an und bietet Exportoptionen.
- **Pakete:** Einfaches Installieren und Verwalten von R-Paketen.
- **Markdown-Unterstützung:** Ideal für die Erstellung von Berichten und Dokumentationen.

Installation

Für eine ausführliche Installation-Guide bitte lesen Sie [RStudio Guide](#).

1. R installieren

- **Besuche die CRAN-Website:** Gehe zu <https://cran.r-project.org/>.
- **Wähle dein Betriebssystem:** Klicke auf das passende Betriebssystem (Windows, macOS oder Linux).
- **Download und Installation:** Lade die Installationsdatei herunter und folge den Anweisungen, um R zu installieren.

2. RStudio installieren

- **Besuche die Posit-Website:** Gehe zu <https://posit.co/download/rstudio/>.
- **Lade RStudio herunter:** Wähle die RStudio-Desktop-Version für dein Betriebssystem aus und lade sie herunter.
- **Installiere RStudio:** Führe die Installationsdatei aus und folge den Anweisungen.

3. R und RStudio starten

- **RStudio öffnen:** Nach der Installation kannst du RStudio öffnen. Es sollte deine R-Installation automatisch erkennen.
- **Beginnen:** Du kannst direkt mit der Arbeit an deinem R-Code starten!

Alternativ für online Anwendung

Wenn Sie keine Installation brauchen, können Sie RStudio auf die Cloud benutzen mit [Posit Cloud](#).

R Hilfe-Dokumentation

```
# suchen Sie für ein Thema
help('matrix')

# suchen Sie direkt mit ??
??matrix

# suchen Sie mit einem Stichwort
help.search('linear regression')
help.search('optimization', package = 'stats')
```

Wichtige Basispakete

R enthält mehrere Basispakete, die standardmäßig installiert und geladen werden. Diese bieten grundlegende Funktionen und Tools für die Arbeit mit Daten, Berechnungen und mehr. Hier ist eine Liste der wichtigsten Basispakete:

1. **base**: Enthält grundlegende Funktionen wie Rechenoperationen, Kontrollstrukturen und Arbeiten mit Datentypen.
2. **stats**: Stellt statistische Funktionen bereit, z. B. für lineare Modelle, Clusteranalysen und Tests.
3. **graphics**: Funktionen für die Erstellung von Basisgrafiken und Visualisierungen.
4. **utils**: Bietet Werkzeuge wie das Lesen und Schreiben von Daten sowie Dateiverwaltung.

```
# Beschreibung
packageDescription('stats')

# Liste von Funktionen
library(help = 'stats')

# Liste von allen verfügbaren Packages
library()

# Liste von allen aktuellen Bibliotheken
search()
```

Übungen

1. Taschenrechner

Berechnen Sie die folgenden Aufgaben:

Ü1A1

$$203 + 43 - 12$$

Ü1A2

$$5 \cdot 43 + 100$$

Ü1A3

$$\sqrt{10^3}$$

Ü1A4

$$\frac{\sin(3)}{4.23 - 3.2}$$

Ü1A5

$$\left(\frac{e^3}{5!}\right)^{0.5}$$

Ü1A6

$$\ln(e)$$

Ü1A7

$$\frac{1}{3}|10 - 100|$$

Ü1A8

$$\sin^{-1}(1)$$

Ü1A9

$$\frac{100}{\pi} - 10$$

Ü1A10

$$\sin\left(e^{\frac{1}{2}} \cdot e^{-4}\right)$$

2. Standardfunktionen

Erzeugen Sie die folgenden Ausgabe nur mit Standardfunktionen und `vect` :

```
vect <- c(1, -1, 5, 0, -8, 2, 4, 0, 15, 20, -30, 6)
```

Ü2A1

```
[1] 20
```

Ü2A2

```
[1] -30
```

Ü2A3

```
[1] 14
```

Ü2A4

```
[1] 1 0 5 5 -3 -1 3 3 18 38 8 14
```

Ü2A5

```
[1] -2 6 -5 -8 10 2 -4 15 5 -50 36
```

Ü2A6

```
[1] 1 1 5 0 8 2 4 0 15 20 30 6
```

Ü2A7

```
[1] 1.166667
```

Ü2A8

```
[1] 12.26846
```

Ü2A9

```
[1] "1"  "-1"  "5"   "0"   "-8"  "2"   "4"   "0"   "15"  "20"  "-30"
"6"
```

Ü2A10

```
[1] 1 -1 5 0 -8 2 4 0 15 20 -30 6 Inf
```

Ü2A11

```
[1] 12
```

Ü2A12

	[,1]	[,2]	[,3]	[,4]
[1,]	1	-1	5	0
[2,]	-8	2	4	0
[3,]	15	20	-30	6

3. Datenstrukturen

Erledigen Sie die folgenden Aufgaben:

Ü3A1

- Erstellen Sie einen Vektor mit 5 Namen, die `namen` heißt.
-

Ü3A2

- Fügen Sie dem Vektor `namen` 5 weitere Namen hinzu.
-

Ü3A3

- Drucken Sie jeden Name von `namen` aus.
-

Ü3A4

- Drucken Sie den ersten Buchstabe jedes Names von `namen` aus.
-

Ü3A5

- Drucken Sie nur Namen von `namen` mit einem 'a' oder 'e' drin.
-

Ü3A6

- Erstellen Sie `nachnamen` mit `namen`, indem Sie den ersten und mittleren Buchstaben entfernen.
-

Ü3A7

- Speichern Sie `nachnamen` als nur große Buchstaben.
-

Ü3A8

- Erstellen Sie einen Dataframe `df_mitarbeiter` mit `namen` und `nachnamen` als Spalten.
-

Ü3A9

- Erstellen Sie einen Vektor `gehalt`, der beliebige ganze Zahlen enthält.
-

Ü3A10

- Fügen Sie `gehalt` dem Datafram `df_mitarbeiter` hinzu.
-

Ü3A11

- Ändern Sie die Namen der Spalten von `df_mitarbeiter` zu 'Vorname', 'Nachname', und 'Jahresgehalt'

4. Benutzerdefinierte Funktionen

Bestimmen Sie die folgenden Funktion, sodass für beliebigen Vektor `vect` nur TRUE ergibt. Benutzen Sie keine Standardfunktionen.

```
set.seed(123)
N <- 100 # beliebig
MIN <- 100 # beliebig
MAX <- 100 # beliebig
vect <- runif(N, min = -MIN, max = MAX)
# all() prüft, ob alle Elemente wahr sind
```

Ü4A1

```
meine_summe(vect) == sum(vect)
```

Ü4A2

```
all(meine_kumsumme(vect) == cumsum(vect))
```

Ü4A3

```
mein_max(vect) == max(vect)
```

Ü4A4

```
mein_min(vect) == min(vect)
```

Ü4A5

```
all(mein_abs(vect) == abs(vect))
```

Ü4A6

```
mein_mittelwert(vect) == mean(vect)
```

Ü4A7

```
meine_sd(vect) == sd(vect)
```

Ü4A8 - Herausforderung

```
all(meine_diff(vect) == diff(vect))
```

Ü4A9 - Herausforderung

```
all(meine_kov(vect, vect2) == cov(vect, vect2))
```