

Tools in Data Science und Machine Learning

rapp.fhnw.ch/RAPP/MDS/

Ihr Dozent

Dr. Fabian Sander

[\(fabian.sander@fhnw.ch\)](mailto:fabian.sander@fhnw.ch)

- Dozent im Bereich empirische Wirtschafts- & Sozialforschung, Business Analytics & Applied Data Science:
- Forschungsschwerpunkte: (höhere) Berufsbildung, Bildungsökonomie/Soziologie
- Erfahrung in Praxis- und Lehre in den Bereichen: empirische Umfragemethoden, angewandte Statistik (Paneldatenanalyse, statistische Dekompositionsmethoden, Strukturgleichungsmodelle), wissenschaftliches Arbeiten

Ziele des Moduls

Erlernen der wichtigsten Prozeduren zur Be- und Verarbeitung von Datensätzen mit «R» und «Excel».

Ablauf

Tools in DS und ML – Teil I: Arbeiten mit der Programmiersprache «R»

1) Einführung in R (support mit leaRn)

- Installation (R, R-Studio, Pakete), Anlegen R-Projekt, Datenimport
- R-Basics: Objekte, Funktionen, Datentypen
- Vektoren (erstellen, ändern, sortieren, indexieren)

2) Datenmanagement (data wrangling I)

- Erstellen von Datentabellen (data frames)
- Bearbeitung von Datentabellen mit dem dplyr-package (sortieren, filtern, auswählen, «pipe-operator», gruppieren, modifizieren, zusammenfassen)

Ablauf

Tools in DS und ML – Teil I: Arbeiten mit der Programmiersprache «R»

3) Datenmanagement II (data wrangling II)

- Datumstypen - & formate, extrahieren von Datumsangaben mit dem lubridate-package
- Kombinieren ganzer Datentabellen oder Zeilen/Spalten; Kombinieren von Vektoren (verbinden) mit dem dplyr-package

4) Datenmanagement III ((data wrangling III))

- Umstrukturierung ganzer Datentabellen oder Zeilen/Spalten (wide, tidy, separieren, vereinigen) mit dem tidyr-package

Ablauf

Tools in DS und ML – Teil I: Arbeiten mit der Programmiersprache «R»

5) Grafische Darstellung

- Grundlagen
- Detaillierte grafische Darstellung von Daten (gg-plot)

6) Programmierung in R

- Grundlagen
- Bedingungen (if-else, any, all)
- Funktionen (function)
- Schleifen (for-loops)

Ablauf

Tools in DS und ML – Teil II: Arbeiten mit dem Tabellenkalkulationsprogramm «Excel»

7) Einführung in Excel, Datenbe- & Verarbeitung, Datenaggregation- & Auswertung

- Funktionen; Namensmanager
- Zusammenfassen und Analysieren von Daten mit Pivot-Tabellen; Pivot-Grafiken

8) Makros I – Automatisierung und Programmierung

- Makros aufzeichnen
- Makros mit VBA programmieren

9) Makros II – Automatisierung und Programmierung

- Makros mit VBA programmieren

10) SQL (mit R) & Prüfungsvorbereitung

Literatur - R

- Heimsch, F., Niederer, R. & Zöfel, P. (2018). Statistik im Klartext. Halbergmoos: Pearson Education Deutschland GmbH.
- Irizarry, R. A. (2022). Introduction to Data Science. Data Analysis and Prediction Algorithms with R.
<https://rafalab.github.io/dsbook/>
- <https://rapp.fhnw.ch/RAPP>
- <https://rstudio-education.github.io/hopr/>
- Wickham, H., & Grolemund, G. (2017). R for Data Science. Beijing: O'Reilly.
<https://r4ds.had.co.nz/index.html>
- Wickham, H., & Raton, B. (2019). Advanced R. Second Edition. London/New York: CRC Press.
<https://adv-r.hadley.nz/>

Literatur - Excel

- Benker, H. (2014). Excel in der Wirtschaftsmathematik. Wiesbaden. Springer Vieweg Verlag.
- Etheridge, D. (2010). Excel. Data Analysis. Indianapolis. Wiley Publishing, Inc.
- Fuchs, E. (2021). SQL Grundlagen und Datenbankdesign. Bodenheim, Herdt-Verlag.
- García, R. H. (2016). Excel 365 Automatisierung, Programmierung. Bodenheim, Herdt Verlag.
- García, R.H. & Wies, P. (2021). Excel 365 Grundlagen. Bodenheim, Herdt Verlag.
- García, R.H. & Wies, P. (2021). Excel 365 Fortgeschrittene Techniken. Bodenheim, Herdt Verlag.
- García, R. H. & Wies, P. (2021). Excel 365 Pivot-Tabellen und Filter. Daten professionell auswerten. Bodenheim, Herdt Verlag.
- Kellner, F. & Brabänder, Ch. (2019). VBA mit Excel. Berlin. Springer Gabler.
- Schels, I. (2014). Excel Formeln und Funktionen. München: Carl Hanser Verlag.
- Schels, I. (2020). Business Intelligence with Excel. München: Carl Hanser Verlag.

Übersicht der Modulsoftware

Software	Excel	R
Vorteil	<ul style="list-style-type: none"> • Einfache(re) Bedienung als mit R • Programmierung mittels Macros möglich 	<ul style="list-style-type: none"> • Breite Anwendungsfelder (ökonometrische Modellierung, Stichprobenziehung, Automatisierungsmöglichkeiten [bspw. Serienbriefe/Versand, Web-scraping, ...]) • Kostenlos
Nachteil	<ul style="list-style-type: none"> • Weniger Optionen als mit R • Kostenpflichtig 	<ul style="list-style-type: none"> • Vergleichsweise komplizierte Bedienung (meist Schulung notwendig)
Download	https://www.microsoft.com/de-ch/microsoft-365	Tutorial für <ul style="list-style-type: none"> • Microsoft Windows: http://learn.fhnw.ch/install_en_win.html • Apple Mac OS: http://learn.fhnw.ch/install_en_mac.html