

M.Psy.209 Statistische Methoden I - Einführung in multivariate Verfahren und Data Science

Modulbeschreibung:

Dieses Modul vermittelt theoretischen Grundlagen multivariater Verfahren und Data Science zur Beschreibung und Analyse von Daten und praktiziert deren Anwendung in Übungen unter Verwendung des Statistikprogramms R.

Vorlesung:

Die Studierenden bereiten sich auf den Präsenztermin vor, indem sie jede Woche vier Lernvideos zu einem Thema durcharbeiten. In der anschließenden Präsenzveranstaltung werden Fragen zu den jeweiligen Verfahren besprochen. Folgende Themen werden behandelt:

- 1 Einführung
- 2 Daten und Modelle + Abbildungen
- 3 Lineare Regression
- 4 Multiple Regression
- 5 Allgemeines Lineares Modell
- 6 Logistische Regression
- 7 Zeitreihenanalyse
- 8 Multivariate Analyse
- 9 Principle Component Analysis
- 10 Faktorenanalyse
- 11 Clusteranalyse
- 12 Gruppierte Multivariate Daten
- 13 Statistical Learning
- 14 Resümee

Literatur:

Vehkalahti, K. & Everitt, B. (2018). *Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences*. CRC Press.

Berk, R.A. (2020). *Statistical Learning from a Regression Perspective*. Springer Nature.

Übungen:

Die Übungen finden epochal immer freitags in einem 3-stündigen Block von 8:15 Uhr bis 11:45 Uhr statt im PC-Pool statt. Übung 1 (Einführung in Data Science) umfasst die erste Hälfte des Semesters (Ende Oktober bis Mitte Dezember), Übung 2 (Multivariate Verfahren) umfasst die zweite Hälfte des Semesters (Mitte Dezember bis Anfang Februar).

In der ersten Hälfte des Semesters (Übung 1) werden die Grundlagen zur Datenanalyse in R gelegt mit einem starken Fokus auf allgemeinen skriptbasierten Techniken und Methoden zur Visualisierung, Aufbereitung, Transformation und Säuberung von Daten. Dabei werden sowohl allgemeine Coding-Prinzipien zur selbstständigen Erstellung von R-Skripten eingeübt, sowie die Fähigkeit vorhandenen Code auf Angemessenheit und Korrektheit zu überprüfen. Folgende Themen werden behandelt: Grundlegender Aufbau und Ansatz von R, Coding Basics, tidyverse, publikationsreife Visualisierung von Daten, Data Wrangling (importing, tidying, transforming), Programmierung in R: Schleifen und Funktionen.

In der zweiten Hälfte des Semesters (Übung 2) werden diese erlernten Kompetenzen angewendet, um ausgewählte statistische Verfahren auf vorhandene Datensätze anzuwenden. Folgende Themen werden behandelt: Allgemeines und generalisiertes lineares Modell, Linear-Mixed-Effects Modelle, exploratorische und konfirmatorische Faktorenanalyse, Cluster Analyse sowie Stichprobenplanung mittels Simulationen.

Jeder Termin wird in drei ca. einstündige Abschnitte eingeteilt (jeweils mit einer ca. 15-minütigen Pause zwischen den Abschnitten).

8:15 – 9:15	Kurze Rekapitulation des bisher bearbeiteten Stoffes sowie Einführung des aktuellen Themas
9:30 – 10:30	Hands-On Übungen zum aktuellen Thema mit individueller Unterstützung durch Dozierenden und Tutor*innen.
10:45 – 11:45	Zusammentragen der Ergebnisse der Übungen, Erörterung von Details und Problemen, Zusammenfassen der wichtigsten Inhalte des Termins

Übungsaufgaben zu den einzelnen Terminen ermöglichen es den Studierenden die Inhalte selbstständig zu vertiefen. Zudem werden zusätzlich zu den Freitagsterminen pro Woche zwei Tutorientermine angeboten, in denen die Inhalte der Übungen (und/oder die Übungsaufgaben) noch einmal mit den Tutor*innen vertieft werden können.

Literatur:

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage, London.

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media, Inc.

<https://r4ds.hadley.nz/>