Vorlesung Statistik A
Explorative Datenanalyse

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Wintersemester 2026/2027

Statistikausbildung im Überblick

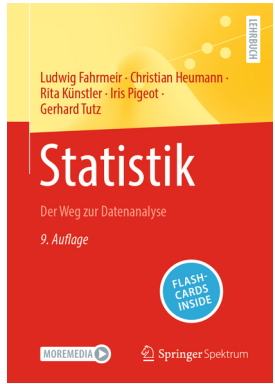
Lehrveranstaltungen

- Modul Statistik 1 (WS)
  VL Statistik A: Explorative Datenanalyse, deskriptive Statistik
  VL Statistik B: Wahrscheinlichkeitstheorie
- Modul Statistik 2 (SS)
  VL Statistik C: Grundlegende Verfahren der Inferenzstatistik
  VL Statistik D: Lineare Modelle
- Digitale Tools (ab SS)
- Forschungsvertiefung Statistik (MSc)
  VL Statistik III: Verfahren für Längsschnittdaten, Multivariate Verfahren

Mehr zur Methodenausbildung in der VL Forschungsmethoden

Literatur

Das Lehrbuch zur Vorlesung



Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I. & Tutz, G. (2023). *Statistik* (9. Aufl.). Berlin: Springer. doi:10.1007/978-3-662-67526-7

Literatur

Weitere Literatur

- Bosch, K. (2011). *Elementare Einführung in die Wahrscheinlichkeitsrechnung* (11. Aufl.). Braunschweig: Vieweg.
- Bosch, K. (2010). *Elementare Einführung in die angewandte Statistik* (9. Aufl.). Braunschweig: Vieweg.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
- Hays, W.L. (1994). *Statistics* (5th ed.). New York: Holt, Rinehart and Winston.
- Meyberg, K. & Vachenauer, P. (2003). *Höhere Mathematik 1*. Berlin: Springer.

Online-Lehrtexte

- Siehe Homepage

(Kommerzielle) Software

Statistik

- SPSS
- SAS

Symbolische Mathematik

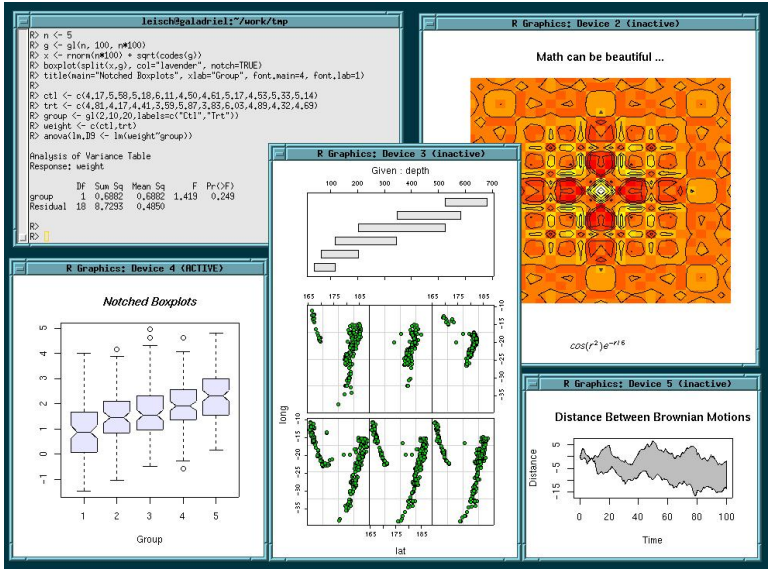
- SageMath (open-source)
  `https://www.sagemath.org/`
- Mathematica
  `https://wolframalpha.com/`

Numerische lineare Algebra

- Matlab
- GNU Octave (open-source)
  `https://www.gnu.org/software/octave/`

Open-Source-Software: R

https://www.r-project.org/



Open-Source-Software: R

Was ist R?

- R ist eine freie Programmierumgebung für statistische Auswertungen und zur Erstellung von Abbildungen
- R und S-Plus sind zwei Implementierungen der Sprache S (Becker & Chambers)
- R wurde 1996 von Ross Ihaka und Robert Gentleman entworfen und wird mittlerweile von mehreren Entwicklern (R-Core-Team) betreut
- R ist die lingua franca der Datenanalyse
  - Eine neue R-Version pro Jahr mit nur kleinen Veränderungen
  - Aktive Entwicklung in tausenden Zusatzpaketen
- R ist für Unix-, Mac- und Windows-Rechner verfügbar

Was ist Statistik?

Etymologie

- Das Wort Statistik stammt vom lateinischen *statisticum* ("den Staat betreffend") und dem italienischen *statista* ("Staatsmann" oder "Politiker")
- Die deutsche Statistik, eingeführt von Gottfried Achenwall (1749), bezeichnete ursprünglich die Lehre von den Daten über den Staat, also Staatstheorie
- Im 19. Jahrhundert wurde das Wort durch den Engländer Sir John Sinclair erstmals in seiner heutigen Bedeutung benutzt

Statistik ist ...

- die Gesamtheit der Methoden zur Zusammenfassung, Darstellung und Analyse von Daten
- angewandte Mathematik
- *the art and science of learning from data*

Warum brauchen Psychologen Statistik?

Als empirische Wissenschaft sieht sich die Psychologie mit zufälligem Geschehen konfrontiert

- Zufällig bedeutet hier nicht völlig regellos oder chaotisch, sondern "nicht im Einzelnen vorhersagbar"

Auch in anderen Naturwissenschaften spielt Zufälligkeit eine Rolle

- Physik: Druck (eines Gases) als statistisches Konzept, radioaktiver Zerfall
- Biologie: Genetik, Entwicklung von Populationen einer Art
- Meteorologie: Wettervorhersage

In der Psychologie gibt es aber kaum spezifische Theorien, was wiederum die Zufälligkeit in den Beobachtungen begünstigt

- *Weak theory often goes with noisy data: it is hard to know how to collect relevant data to test a theory that is not well specified*

Teilbereiche der Statistik

Deskriptive Statistik (beschreibende Statistik)
Explorative Datenanalyse

- Untersuchungsergebnisse (Stichprobe) werden in geeigneter Weise beschrieben und zusammengefasst (Tabellen, grafische Darstellungen)
- Berechnung von Kenngrößen (Statistiken), die möglichst viel Information über die vorliegenden Daten liefern

Inferenzstatistik (schließende oder induktive Statistik)

- Aus den Daten einer Stichprobe werden Eigenschaften einer Grundgesamtheit (Population) abgeleitet
- Die Wahrscheinlichkeitstheorie liefert die Grundlagen für die erforderlichen Schätz- und Testverfahren

Was ist Statistik?

"Zwei Welten"

Empirie



Stichprobe

Statistiken
$\bar{x}$, s^2

statistische Inferenz
$\longrightarrow$
Parameterschätzung
Testverfahren

Theorie



Population

Parameter
μ, σ^2

Themen

A: Explorative Datenanalyse, deskriptive Statistik

- Merkmale und Merkmalstypen
- Skalenniveau und Bedeutsamkeit
- (Eindimensionale) Häufigkeitsverteilungen
- Kennwerte von Häufigkeitsverteilungen
- Zweidimensionale Häufigkeitsverteilungen
- Korrelationsrechnung (Teil 1)
- Regressionsrechnung
- Korrelationsrechnung (Teil 2)

Themen

B: Wahrscheinlichkeitstheoretische Grundlagen

- Mengenlehre
- Wahrscheinlichkeitsraum
- Kombinatorik
- Stochastische Unabhängigkeit, bedingte Wahrscheinlichkeit und klassische Wahrscheinlichkeitsverteilungen
- Diskrete und stetige Zufallsvariablen, Verteilungen und Kennwerte, Paare von Zufallsvariablen
- Zentraler Grenzwertsatz
- Testverteilungen