

# Anwendungsorientierte Analyseverfahren

## Einführung

Prof. Dr. Michael Scharkow

Sommersemester 2024

# WARUM (NOCH) EINE VORLESUNG ZUR STATISTIK?

- **Literacy:** Wenn man aktuelle Forschung lesen möchte (oder muss), führt kein Weg an etwas komplexeren Analysen vorbei.
- **Selbstwirksamkeit:** Wer einmal eine Analyse durchgeführt hat, kann in Seminar- und Abschlussarbeiten besser Daten auswerten.
- **Jobaussichten:** Viele AbsolventInnen berichten rückblickend, dass gerade die Methodenskills am besten verwertbar waren bei der Jobsuche und im Beruf.

# ZIELE DER VORLESUNG

- Studierende werden dazu befähigt, die Anwendung ausgewählter Analyseverfahren nachzuvollziehen sowie entsprechende Forschungsergebnisse und Interpretationen zu verstehen.
- Studierende sind in der Lage, für ausgewählte Analyseverfahren anhand vorgegebener Daten Ergebnisse aus der Forschungsliteratur mittels Statistiksoftware zu reproduzieren.
- Studierende verfügen über die die Kompetenz, Angemessenheit und Güte von methodischen Vorgehensweisen zu beurteilen.
- (Studierende finden Statistik weniger schlimm und langweilig).

# WAS DIE VORLESUNG (NICHT) IST

- keine Wiederholung der VL Statistik oder der Datenanalyse-Übungen
- Fokus auf das Verständnis für und die Anwendung von statistischen Verfahren, weniger die Mathematik dahinter
- das Allgemeine Lineare Modell (GLM) als grundlegendes Verfahren
- kein reines Ablesen von p-Werten und Signifikanz-Sternchen
- emanzipierter Umgang mit statistischen Verfahren statt Rezepte abarbeiten

# VORLESUNGSPLAN

Termin	Thema
17.04.24	Einführung
24.04.24	GLM Grundlagen
01.05.24	frei (Feiertag)
08.05.24	Regression - metrische Prädiktoren
15.05.24	Regression - kategorielle Prädiktoren*
22.05.24	Multiple Regression (Felix Dietrich)
29.05.24	Modellannahmen

Termin	Thema
05.06.24	Modellvorhersagen*
12.06.24	Moderationsanalyse - kategorielle Moderatoren
19.06.24	Moderationsanalyse - metrische Moderatoren*
26.06.24	frei (Tag der offenen Uni)
03.07.24	Logistische Regression
10.07.24	Multilevel-Regression
17.07.24	Abschluss

# **ABLAUF DER SITZUNGEN UND ANWESENHEIT**

## **Ablauf**

1. Besprechung der praktischen Übungen/Hausaufgabe (max. 15 min)
2. Vorlesungsteil (max. 60 min)
3. Fragen und Antworten zur Vorlesung und praktischen Übung

## **Anwesenheit**

- keine Anwesenheitspflicht, aber auch keine Nachhilfepflicht meinerseits
- eigenständige Nachbereitung der praktischen Übungen

# E-LEARNING UND STUDIENLEISTUNG

## Material

- Folien und Übungsmaterialien samt Daten und R-Code auf <https://stats.ifp.uni-mainz.de/ba-aa-vl>

## Studienleistung

- während der Vorlesungszeit **3** Teil-Studienleistungen (je ca. 20 min)
- sowohl Interpretations- als auch praktische Analyseaufgaben
- Deadline jeweils 2 Wochen nach Aufgabenstellung, Do 12h
- Benotung jeweils Pass/Fail, min 2x Pass nötig

# PRAKTISCHE ÜBUNGEN

- zu jeder Sitzung eine praktische Übung auf Basis einer publizierten Studie
- kurze Besprechung in der Vorlesung, meist mit einer exemplarischen Analyse
- R-Code zum Replizieren der Analysen zuhause oder während der Vorlesung
- praktische Anwendung als integraler Teil der Vorlesung und der Studienleistung
- Copy & Paste/Anpassung von bestehendem Code ist ok!



# SOFTWARE

- in der VL vorgestellten Analysen lassen sich mit praktisch jeder Statistiksoftware reproduzieren
- jede Statistiksoftware ist nur ein Werkzeug
- Lektürekompetenz heißt auch, man kann sowohl SPSS als auch Stata oder R-Output lesen
- wegen Verfügbarkeit und Zukunftsfähigkeit verwende ich R

**Für die Studienleistung ist irrelevant, welche Software Sie verwenden!**

# WARUM MUSS ICH JETZT AUCH NOCH R LERNEN?

- **Sie müssen nicht!**
- R ist freie Software und durch viele tausend Pakete (packages) erweiterbar, u.a. für
  - Datenerhebung: Web-Scraping, APIs (z.B. für TikTok oder Spotify), Textdaten
  - Auswertung: Statistik, Textanalyse, Audiodaten, Psychophysiologie, etc.
  - Datenpräsentation und -visualisierung: Grafiken, Berichte, Folien (z.B. diese)
- grundlegende Programmierkenntnisse, die auch ohne Statistik nützlich sein können
- das IfP stellt in diesem Semester BA Datenanalyse von SPSS auf R um  
<https://stats.ifp.uni-mainz.de/ba-datenanalyse/>

# KLEINES R-BEISPIEL: BREAKING BAD DEATHS

```
library(tidyverse)
read_csv('https://wegweisr.haim.it/Daten/breaking_bad_deaths.csv') |>
  count(method, sort = TRUE) |>
  head(n = 5)
```

<b>method</b>	<b>n</b>
accidental death	172
shot	59
poisoned	14
stabbed	13
bombed	4

# LITERATUR

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

Miles, J., & Shevlin, M. (2001). Applying regression and correlation: A guide for students and researchers. London: Sage.

Darlington, R. B., & Hayes, A. F. (2016). Regression analysis and linear models: Concepts, applications, and implementation. Guilford Publications.

McElreath, R. (2020). Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press. (für Interessierte)

# AUFGABEN ZUR NÄCHSTEN SITZUNG

1. Lesen Sie die Einführung in R aus dem BA-Datenanalyse Kurs <https://stats.ifp.uni-mainz.de/ba-datenanalyse/r-und-rstudio.html>
2. Lösen Sie dort die Hausaufgaben 1 und 2 (Installieren von R).  
**Achtung:** Bei Aufgabe 2 bitte die Materialien dieser VL <https://stats.ifp.uni-mainz.de/ba-aa-vl/> statt der BA-Datenanalyse herunterladen und öffnen.
3. **Fallback-Option:** RStudio im Browser <https://rstudio.ifp.uni-mainz.de>, dort das Projekt im Ordner `ba-aa-v1` öffnen.  
(nur aus dem JGU-Netzwerk, Username = Passwort = ZDV-Nutzername)
4. Frischen Sie ggf. ihre Statistik-Kenntnisse auf. Alle sollten wissen, was eine (Ko)-Varianz ist oder wie eine Datenmatrix aussieht.

**Fragen?**