

# Statistik für Digital Humanities

## Organisatorisches

Dr. Jochen Tiepmar

Institut für Informatik  
Abteilung Computational Humanities  
**Universität Leipzig**

14. Oktober 2019

[Letzte Aktualisierung: 14/10/2019, 11:11]

# Modulinformationen

- Akademischer Grad: Bachelor Digital Humanities
- Modulnummer:10-207-0003
- Empfohlen für 3. Semester
- Arbeitsaufwand: 5 LP = 150 Arbeitsstunden
- Verwendbarkeit: Fakultätsinterne Schlüsselqualifikation im B.Sc. Digital Humanities und B.Sc. Informatik
- keine Teilnahmevoraussetzungen
- Anmeldung per Moodlekurs und/oder Mail an [jtiepmar@informatik.uni-leipzig.de](mailto:jtiepmar@informatik.uni-leipzig.de)
- Vorlesung: montags 11:15
- Übung: dienstags 11:15

# Moodle Kurs

- Kursname: Statistik für Digital Humanities
- Kurskennzeichen: StatDH
- Vorlesungen, Übungsserien
- Weiterführende Informationen (Literatur etc.)
- Anmeldung = Prüfungsanmeldung

# Lernziele

Nach der aktiven Teilnahme am Modul *Statistik in den Digital Humanities* sind die Studierenden in der Lage

- statistische Grundbegriffe und Verfahren zu benennen und zu erklären
- ausgewählte Verfahren zu analysieren, zu beurteilen und diese selbstständig auf Problemstellungen im Kontext der Digital Humanities anzuwenden

# Inhalt

- Grundlagen der Statistik
- Explorative Datenanalyse (EDA)
- Datenvisualisierung
- Wahrscheinlichkeitsverteilungen und Sampling
- Statistische Experimente und Signifikanztests
- Regression, Prädiktion, Korrelation
- Statistische Anwendungen in den Digital Humanities

# Literatur

Andy Field:  
Discovering Statistics Using R  
(English Edition)

# Weitere Literaturempfehlungen

Peter Bruce:  
Practical Statistics for Data  
Scientists: 50 Essential Concepts

Eileen Magnello, Borin Van Loon:  
Statistics: A Graphical Guide

# The R Project for Statistical Computing

- Frei verfügbare Softwareumgebung für statistische Analysen
- <https://www.r-project.org/>
- Selbstständige frühzeitige Installation und Einarbeitung empfehlenswert
- Beispiele werden soweit passend mit R Skripten veranschaulicht
- R selbst wird nicht gelehrt
- Zur Prüfung steht R nicht zur Verfügung



# Prüfungsleistung

- Modulprüfung/Klausur (60 Minuten)
- Übungsschein = Prüfungsvorleistung
- **Unbedingt rechtzeitige Anmeldung / Abmeldung beachten**, auch bei Beendigung der Teilnahme (jtiepmar@informatik.uni-leipzig.de)
- Versäumtes Abmelden kann zu Fehlversuch der Klausur führen

# Übungen

- 5 Übungsserien a 20 Punkten
- 50% der Punkte = Übungsschein/Klausurzulassung
- Übungsschein ist im Semester nicht wiederholbar
- Ausgabe nach der Vorlesung
- Abgabe vor der Vorlesung
- Rückgabe 1 Woche später im Seminar
- Handschriftlich oder ausgedruckt
- Gruppenarbeit (2-3 Studierende) möglich, **ist aber als solche zu markieren**
- 1 Abgabe pro Student\_In
- Als Kopien interpretierte Abgaben werden mit 0 Punkten bewertet
- Briefkastenabgabe möglich, werden montags früh eingesammelt

# Vorlesung

- Raumzeit: montags 11:15 P801 (Paulinum)

# Seminare

- Raumzeit: dienstags 11:15 P801 (Paulinum)
- Je 1 Termin pro Übungsblatt
  - Termine: 12.11 26.11 17.12 21.01 04.02
- Weitere Termine spontan oder auf Zuruf möglich, werden zur Vorlesung angekündigt
- Die aktuell zu lösenden Aufgaben werden nicht im Seminar besprochen
- Jeder sollte die Aufgaben im Seminar lösen und erklären können

# Syllabus

#	Datum	Vorlesung	Übungsaufgaben	Seminartermin
01	14.10	Organisatorisches und Überblick	–	–
02	21.10	Statistische Forschung	–	–
03	28.10	Statistische Modelle	Ausgabe Blatt 1	–
04	04.11	Visualisierung	Einsammeln Blatt 1	–
05	11.11	Statistische Annahmen, Korrelation	Ausgabe Blatt 2	Blatt 1
06	18.11	Regression	Einsammeln Blatt 2	–
07	25.11	Mittelwertvergleiche	Ausgabe Blatt 3	Blatt 2
–	02.12	–	–	–
08	09.12	Kovarianz	Einsammeln Blatt 3	–
09	16.12	Messwiederholungsdesigns & Mixed Designs	Ausgabe Blatt 4	Blatt 3
10	13.01	Multivariate Varianzanalyse	Einsammeln Blatt 4	–
11	20.01	Faktoranalyse	Ausgabe Blatt 5	Blatt 4
12	27.01	Kategorische Daten	Einsammeln Blatt 5	–
13	03.02	Mehrstufige Lineare Modelle	–	Blatt 5