

# Statistik für Digital Humanities

## Organisatorisches / Vorläufiger Pandemieplan

Dr. Jochen Tiepmar

Institut für Informatik  
Computational Humanities  
**Universität Leipzig**

06. April 2020

[Letzte Aktualisierung: 01/04/2020, 10:46]

## „Statistik für die Digital Humanities“



Die Statistik gibt Forscherinnen und Forschern wertvolle datenanalytische Werkzeuge an die Hand, die wesentlich den grundlegenden Charakter von Wissenschaft und Erkenntnisgewinn beeinflussen. Statistik bietet die Möglichkeit, mathematische, theoretische und abstrakte Modelle oder Annahmen mittels empirisch nachgewiesener realer Effekte zu validieren.

Das als fakultätsinterne Schlüsselqualifikation angelegte Modul *Statistik für die Digital Humanities* lehrt Studierende des Fachs Digital Humanities diese grundsätzlichen Methoden mit einem Fokus auf die typischen Aufgaben und Probleme, die sich in dem sehr heterogenen geisteswissenschaftlich orientierten Fachbereich ergeben. Dadurch werden Studierenden bereits frühzeitig die Fähigkeiten vermittelt, einerseits Datenanalysen kritisch zu hinterfragen und andererseits die eigene Forschungsarbeit kompetent und mit der nötigen Objektivität zu strukturieren.

Neben einem grundlegenden Überblick über das Fach werden Methoden der Datenexploration und Visualisierung gelehrt und geübt. Weiterhin werden die Prinzipien empirischer, stichprobenbasierter Experimente sowie statistischer Testverfahren vermittelt und Verfahren zur Datenanalyse, bspw. Prädiktion oder Berechnung von Regression und Korrelation, beigebracht.

---

**Ansprechpartner:** Jochen Tiepmar (jtiepmar@informatik.uni-leipzig.de),  
Computational Humanities Group (<https://ch.uni-leipzig.de/>)

- Bachelor Digital Humanities
- Empfohlen ab 3. Semester
- Modulnummer:10-207-0003
- Arbeitsaufwand: 5 LP = 150 Arbeitsstunden
- Verwendbarkeit: Fakultätsinterne Schlüsselqualifikation im B.Sc. Digital Humanities und B.Sc. Informatik
- In Absprache mit Studiengangverantwortlichen auch andere Verwendbarkeit möglich (Bspw. Kernmodul DH)
- keine Teilnahmevoraussetzungen
- Anmeldung für Moodlekurs per Mail an [jtiepmar@informatik.uni-leipzig.de](mailto:jtiepmar@informatik.uni-leipzig.de)
- Vorlesung: montags 11:15
- Übung: mittwochs 9:15

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

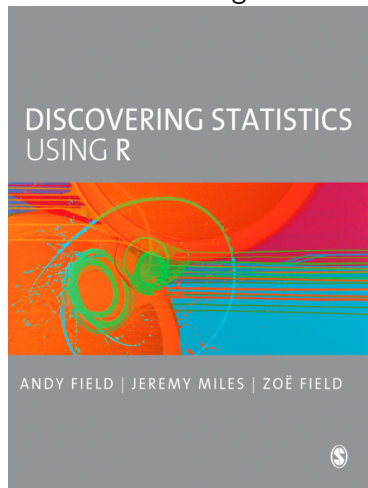
- statistische Grundbegriffe und Verfahren zu benennen und zu erklären
- ausgewählte Verfahren zu analysieren, zu beurteilen und diese selbstständig auf Problemstellungen im Kontext der Digital Humanities anzuwenden

- Kursname: Statistik für Digital Humanities
- Kurskennzeichen: StatDH\_SS20
- Vorlesungen, Übungsserien
- Weiterführende Informationen (Literatur etc.)
- Ergänzungsmaterialien neben der Vorlesung und den Übungsaufgaben sind nicht prüfungsrelevant

Andy Field:

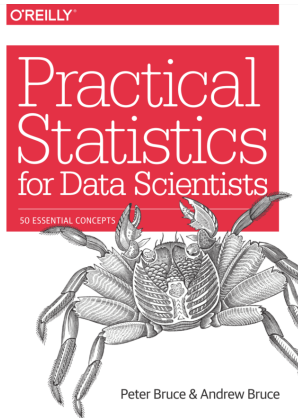
Discovering Statistics Using R (English Edition)

In der UBL vorrätig

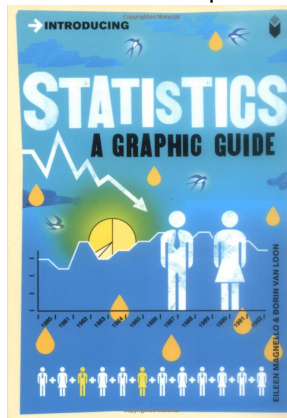


# Weitere Literaturempfehlungen

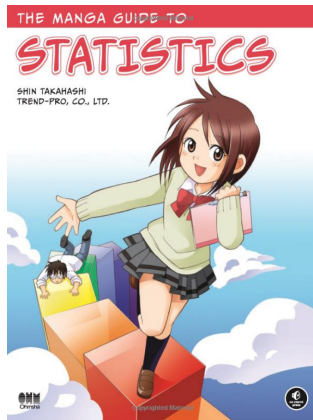
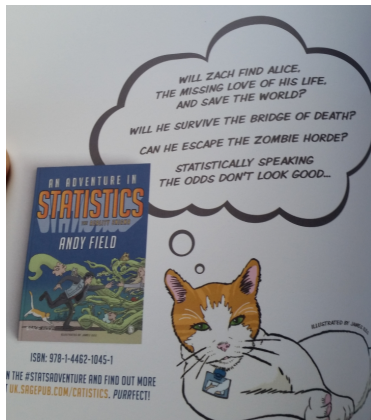
Peter Bruce:  
Practical Statistics for Data  
Scientists: 50 Essential Concepts



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



# Weitere Literaturempfehlungen





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

- Details situationsbedingt unsicher
- 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

- 5 Übungsserien a 20 Punkten
- 50% der Punkte = Übungsschein/Klausurzulassung
- Übungsschein ist im Semester nicht wiederholbar
- Ausgabe nach der Vorlesung (Moodle)
- Abgabe per Email spätestens am Abgabetag (montags) 24 Uhr  
jtiepmar@informatik.uni-leipzig.de
- Abgabeformat PDF (wichtig)
- Eindeutigen (geheimen) Namen für die PDF wählen (Nicht die Matrikelnummer!)
- Korrigierte PDF per Link mit Angabe des Dateinamens abholbar
- 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

- Raumzeit: montags 11:15 SG 3-12
- Aufgezeichnete Vorlesung auf Youtube
- Backup der Videos auf Uniserver
- Fragen usw. per Moodle-Forum oder Email
- Individuelle lokale Kopien sinnvoll zur Reduzierung der Serverlast

- Raumzeit: mittwochs 09:15 P801 (Paulinum)
- Regulär bis auf weiteres keine Termine
- Lösungen der Aufgaben werden online gestellt
- Situationsabhängig Skype-Termine, Forumdiskussionen, Erklärvideos o.ä. (auch auf Zuruf von Studierendenseite!)

| Vorlesung | Inhalt                                                |
|-----------|-------------------------------------------------------|
| 01        | Organisation                                          |
| 02        | Statistische Forschung                                |
| 03        | Statistische Modelle                                  |
| 04        | Visualisierung                                        |
| 05        | Annahmen Parametrischer Tests                         |
| 06        | Kovarianz und Korrelation                             |
| 07        | Regression                                            |
| 08        | Logistische Regression                                |
| 09        | Vergleich zweier Mittelwerte (t-Test)                 |
| 10        | Vergleich mehrerer Mittelwerte (ANOVA & ANCOVA)       |
| 11        | ANOVA II (Mehrfaktoriell, Messwiederholung, Gemischt) |
| ... *     | ...                                                   |

\*Es werden noch Themen dazukommen