

Seminar
Rechnergestützte Statistik

Uwe Ligges

TU Dortmund, SoSe 2026

1 Seminar Rechnergestützte Statistik

Modul anrechenbar für Studiengang

B.Sc. Statistik	BS 12
M.Sc. Statistik	MS 4
B.Sc. Data Science	BD 14
M.Sc. Data Science	MD 4

1 Seminar Rechnergestützte Statistik

Allgemeine Informationen

Dozent	Dr. Uwe Ligges
Teilnahmevoraussetzungen	siehe Studienordnung ideal: Numerik / Computergest. Stat.
Ort und Zeit	tba
Sprache	Deutsch
Erster Termin	1. Vorlesungswoche
Prüfungszulassung	siehe Studienordnung
Prüfungsform	Vortrag, Bericht
LP	4

Anmeldung

Verbindliche Anmeldung bis zum **25. Februar 2026** über den Link in der Mail vom Dekanat.

2 Themen

1. Big Data und R(-Pakete)

- ➊ Wie groß dürfen Daten sein? Was macht man sonst? Datenbanken und geschickter Umgang.
- ➋ Einsatz von Hauptspeicher und/oder Parallelisierung
- ➌ Lineare Modelle und GLMs bei großen Datenmengen: Updating Algorithmen
- ➍ Clusterverfahren für große Datenmengen (k-means, clara, ...)
- ➎ Klassifikationsverfahren für große Datenmengen (Baum, Random Forest, Diskriminanzanalyse, SVM, ...)
- ➏ ...

Es wird jeweils ein Vergleich von herkömmlichen Methoden mit speziellen Methoden für große Daten anhand von geeigneten Testdatensätzen hinsichtlich Genauigkeit und Geschwindigkeit erwartet.

2 Themen

2. Berechnung empirischer Varianzen und Kovarianzen

- ① verwendete Algorithmen in gängiger Statistiksoftware
- ② Konstruktion von Stichproben mit exakt bekannter, einstellbarer Varianz
- ③ Überprüfung der Rechengenauigkeit und Vergleich von Excel, R, SAS und SPSS

3. Sortierverfahren

- ① verwendete Algorithmen von (partiellen) Sortierern in gängiger Statistiksoftware
- ② Konstruktion geeigneter Testdatensätze für Geschwindigkeitsvergleich von (partiellen) Sortierern in Excel, R, SAS und SPSS.

2 Themen

4. Lösung des linearen KQ Problems

- ④ Vergleich von Excel, R, SAS und SPSS hinsichtlich der Rechengenauigkeit (Stichwort: „Zielke Matrizen“)

5. Nichtlineare Optimierung

- ④ Optimierungsverfahren für Optimierung unter Nebenbedingungen (box constraints)

6. GLM: Logistische Regression, Cox-Modell, ...

- ④ verwendete (Optimierungs-)Algorithmen in gängiger Statistiksoftware
- ④ Konvergenz in numerisch schwierigen Situationen
- ④ Verhalten bei Verletzung der Modellannahmen
- ④ Vergleich von R, SAS und SPSS hinsichtlich der Lösungsgüte in numerisch schwierigen Situationen (Stichwort: Separierbarkeit)

2 Themen

7. EM Algorithmus:

- ① (Ab)arten des EM Algorithmus
- ② Identifizierbarkeitsprobleme und deren Auswirkungen

8. Pseudozufallszahlen (gemäß einer Normalverteilung)

- ① verwendete Algorithmen in gängiger Statistiksoftware
- ② Vergleich der Güte solcher Zufallszahlen aus Excel, R, SAS und SPSS anhand statistischer Tests

9. Boostingverfahren

3 Literatur

- Benutzerhandbücher von Excel, R, SAS und SPSS
- diverse Papiere zu den einzelnen Algorithmen
- Lange, Kenneth (1999): Numerical Analysis for Statisticians. Springer-Verlag, New York.
- Weihs, Mersmann, Ligges (2013): Foundations of Statistical Algorithms. Taylor & Francis / CRC Press, Boca Raton.