



Statistik und Data Science – Schlüsselkompetenzen für unsere Zukunft

Helmut Küchenhoff
Campus Tag

2025



Was ist Statistik?

Fragen wir chatGPT:

Was ist Statistik ? Bitte 3 kurze Sätze

“Statistik ist die Wissenschaft der Datensammlung, -analyse und -interpretation, um Muster und Trends zu erkennen. Sie ermöglicht es, aus großen Datenmengen sinnvolle Schlüsse zu ziehen und fundierte Entscheidungen zu treffen. Statistische Verfahren finden Anwendung in verschiedenen Bereichen, um Vorhersagen zu machen und Unsicherheiten zu bewerten.“



- Daten und Wissen sind die großen Themen des 21. Jahrhunderts
- In vielen Lebensbereichen stehen immer mehr Daten zur Verfügung
- Daten-gesteuerte Politik und Wirtschaft
- Adäquate Interpretation und Nutzung von Daten zentrale gesellschaftliche Aufgabe



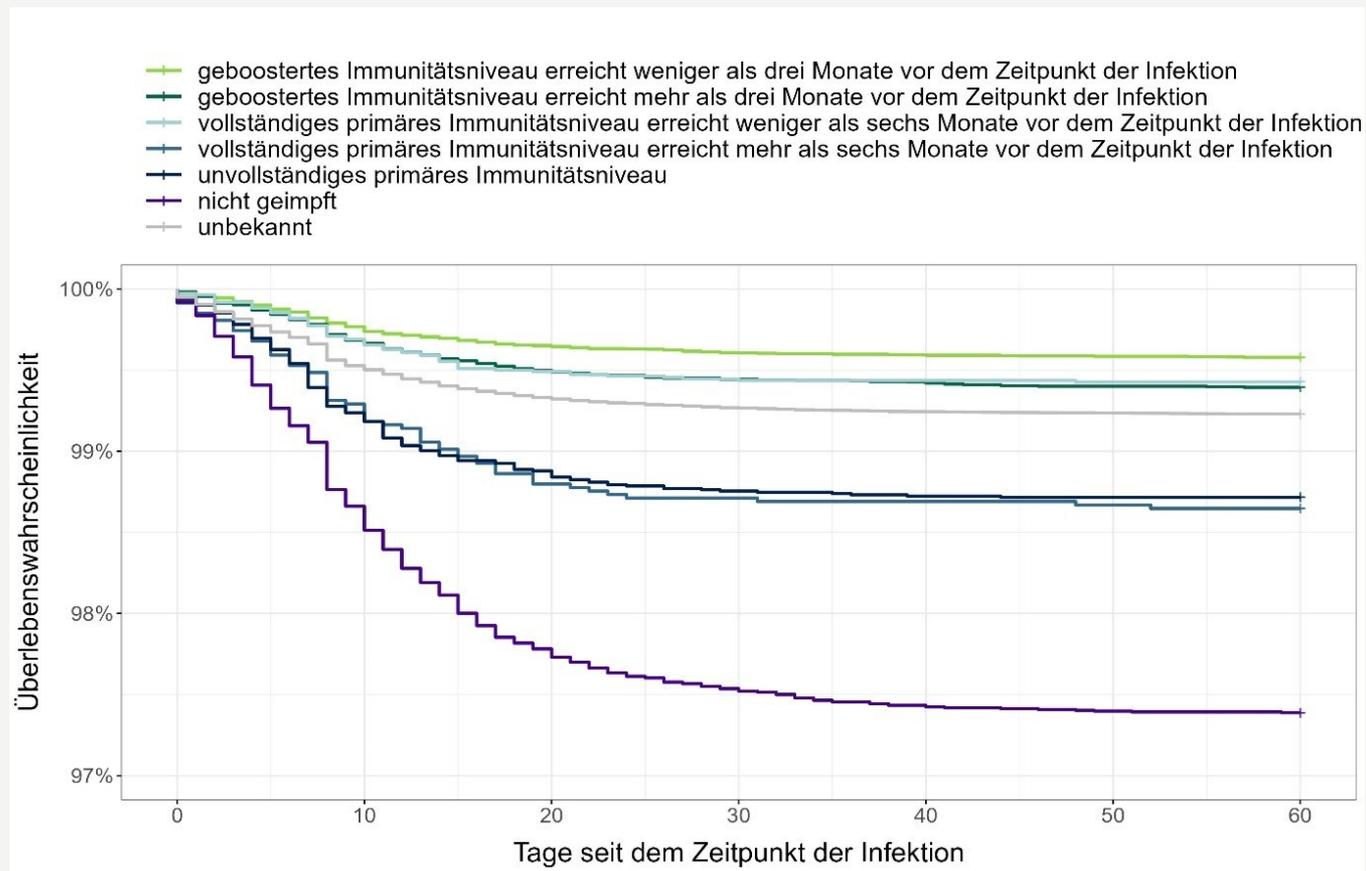
- Grundlagen für Lernalgorithmen: Statistik bildet die Basis für Algorithmen im maschinellen Lernen, einem Hauptbereich der KI.
- Entscheidungsfindung: Statistische Analyse ermöglicht KI-Systemen, Wahrscheinlichkeiten für Entscheidungen zu berechnen.
- Datenanalyse: Statistische Methoden unterstützen bei der Datenaufbereitung, essentiell für die Effizienz von KI-Modellen
- Leistungsbewertung: Statistik hilft, die Performance von KI-Modellen zu messen und zu optimieren.



- Datensammler, Erbsenzähler, Lügner ??
- Datendetektive, Erzeuger von Wissen aus dem Rohstoff Daten !!
- Interdisziplinäre Forscher*innen, die auf vielen Gebieten in Wissenschaft und Wirtschaft arbeiten

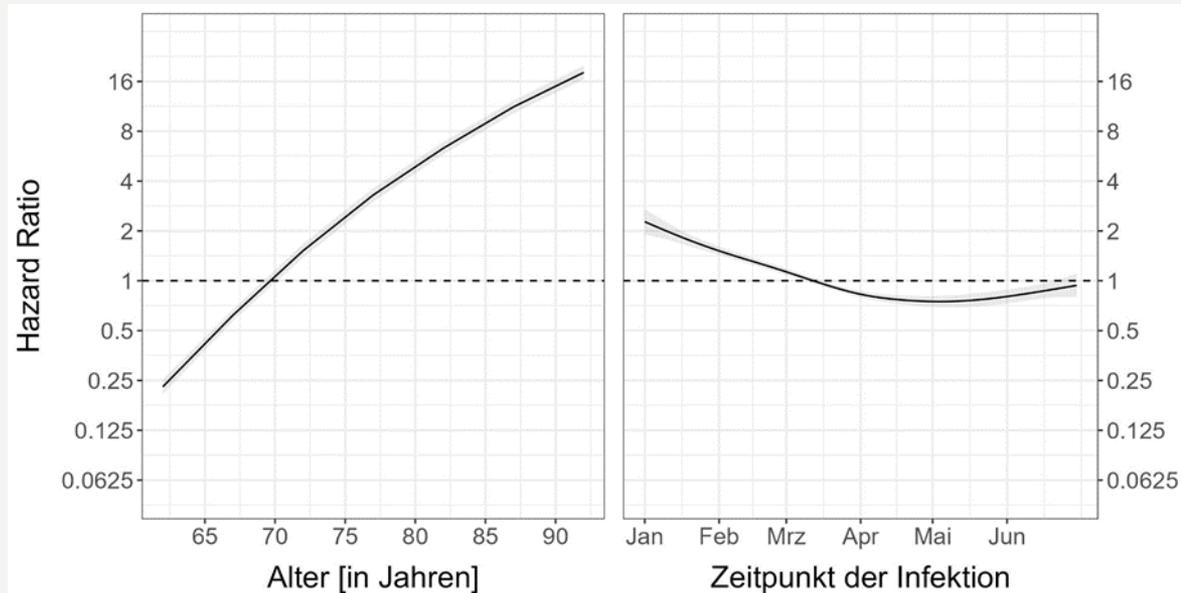


- Infizierte Ü60 in Bayern
- Zusammenhang Impfstatus/Vorinfektionen und Sterberisiko



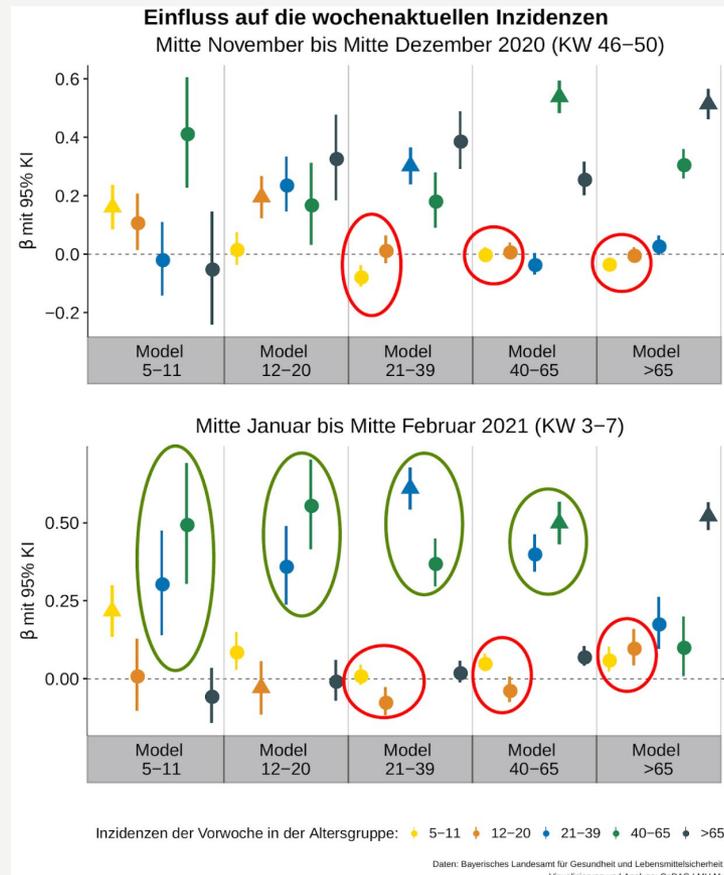


- Infizierte Ü60 in Bayern
- Zusammenhang Impfstatus/Vorinfektionen und Sterberisiko
- Berücksichtigung von Alter, Zeitpunkt der Infektion
- Statistisches Modell ermöglicht gleichzeitige Betrachtung verschiedener Einflüsse

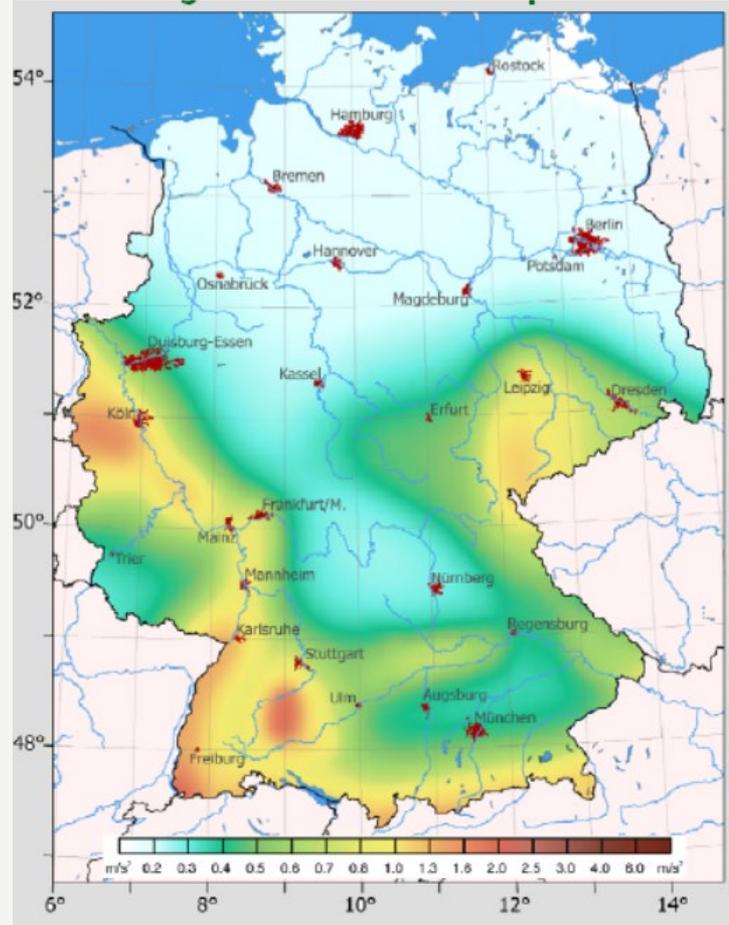




CODAG Bericht Nr. 10 vom 05.03.2021
 Infektionen unter Schulkindern zeigen keinen Zusammenhang
 mit den
 Infektionszahlen in anderen Altersgruppen







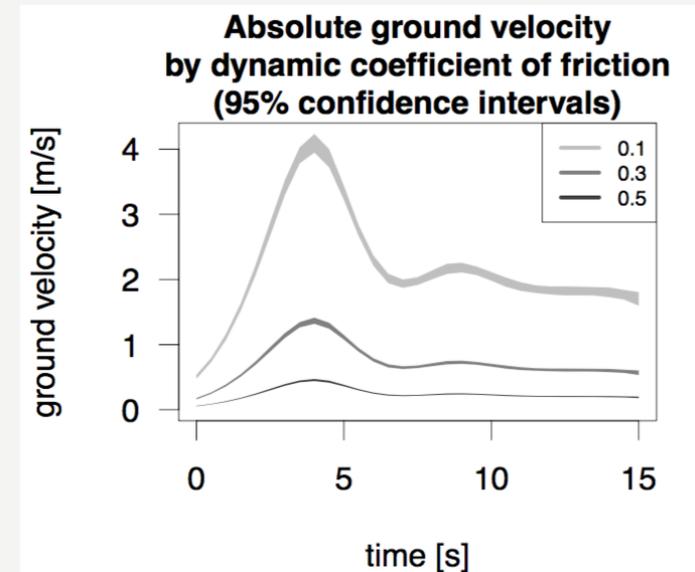
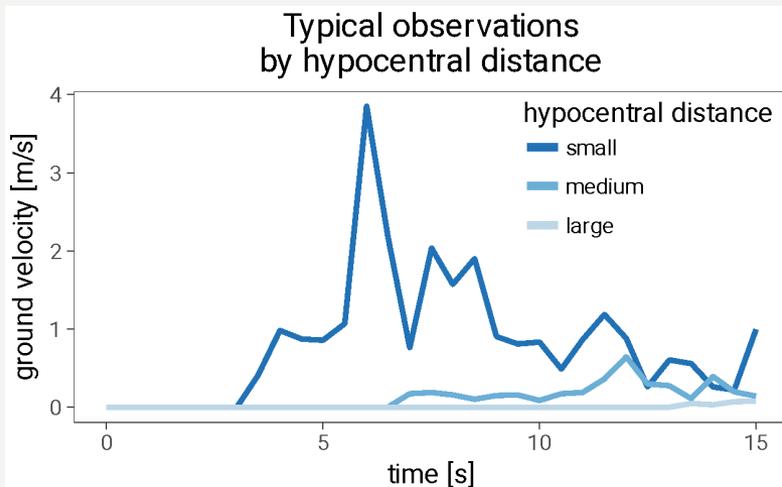


Wie erhält man die Karte?

- Daten aus der Vergangenheit
- Fachwissen (Geophysik)
- Modellbildung mit Unsicherheit
(Probabilistic hazard analysis)



Wie stark sind die Bodenbewegungen, die man unter bestimmten physikalischen Bedingungen erwarten kann?



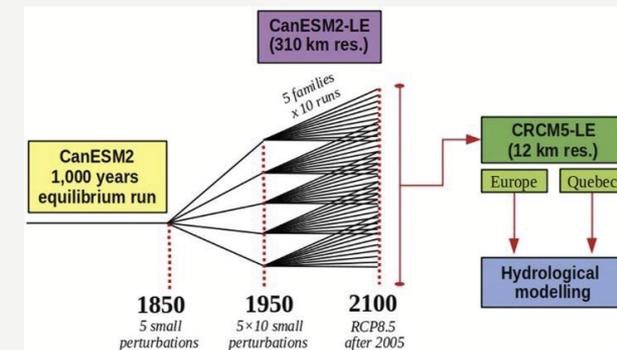


- erstaunliche Fortschritte in KI
- Wie vertrauenswürdig sind Algorithmen?
 - Sicherheit: keinen Schaden verursachen
 - Gesichert gegen Angriffe oder Manipulation
 - Privatsphäre der Daten
 - Verantwortung: ethische Prinzipien (Fairness)
- grundlegendes Verständnis der Methoden nötig





- Kooperation mit CLIMEX-Projekt der LMU
- Prognose von Dürren in Bayern unter verschiedenen Szenarien des Klimawandels
- Interdisziplinäres Seminar Geowissenschaften Statistik
- Bewertung der Gesundheitsrisiken durch Hitze
- Zusammenhang zwischen Lufstaschadstoffwirkungen und Hitze
- Stand der wissenschaftlichen Evidenz





Gibt es eine Midlife Crisis?

Analysen von Panel-Daten zur subjektiven Lebenszufriedenheit mit Regressionsmodellen.

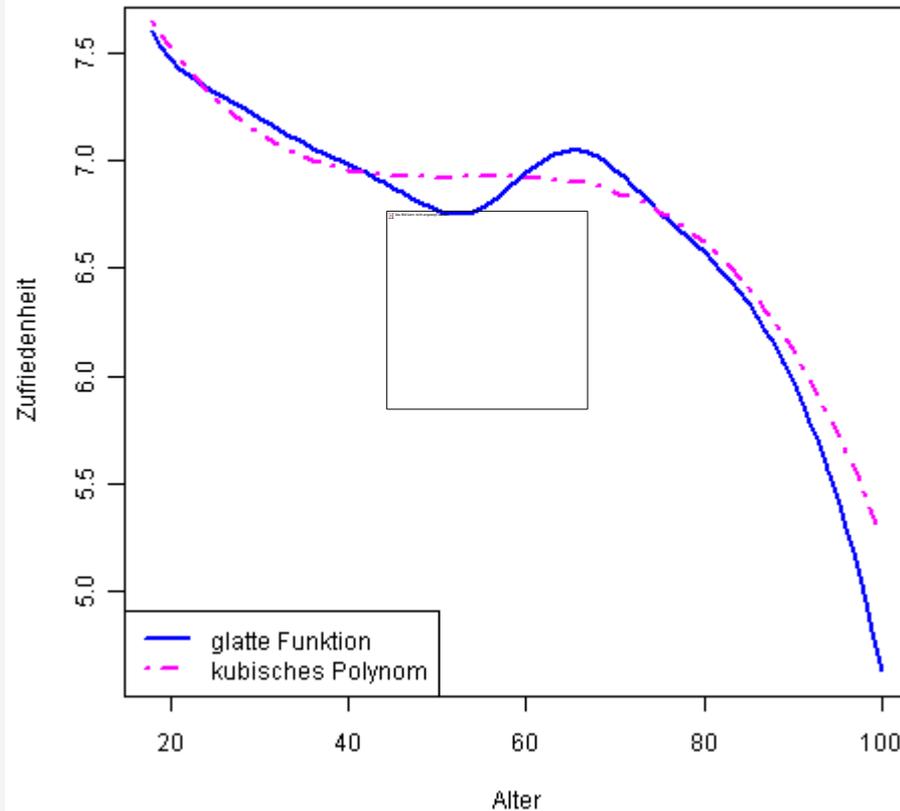
In Zusammenarbeit mit Sonja Greven, Andrea Wiencierz, Christoph Wunder



- Daten stammen aus den Haushaltsstichproben A (Westdeutsche) und C (Ostdeutsche) des Sozio-Ökonomischen Panels (SOEP)
- für die ausgewählten Modellvariablen liegen Beobachtungen aus den Jahren 1992, 1994 bis 2006 vor
- durchschnittliche Anzahl von Beobachtungen pro Person: 7.77
- in die Modellberechnungen ging 102 708 vollständige Beobachtungen von 13224 Individuen ein
- Anzahl Beobachtungen pro Jahr:



geschätzte Funktion inkl. AR(1) für *Durchschnittsmensch*



Midlife-Crisis nur bei glatter Funktion erkennbar.



Consulting Project Alexander Bauer and Matthias Nist:
Früherkennung von Risiken im Lebensmittelhandel
"Project ISAR LGL"

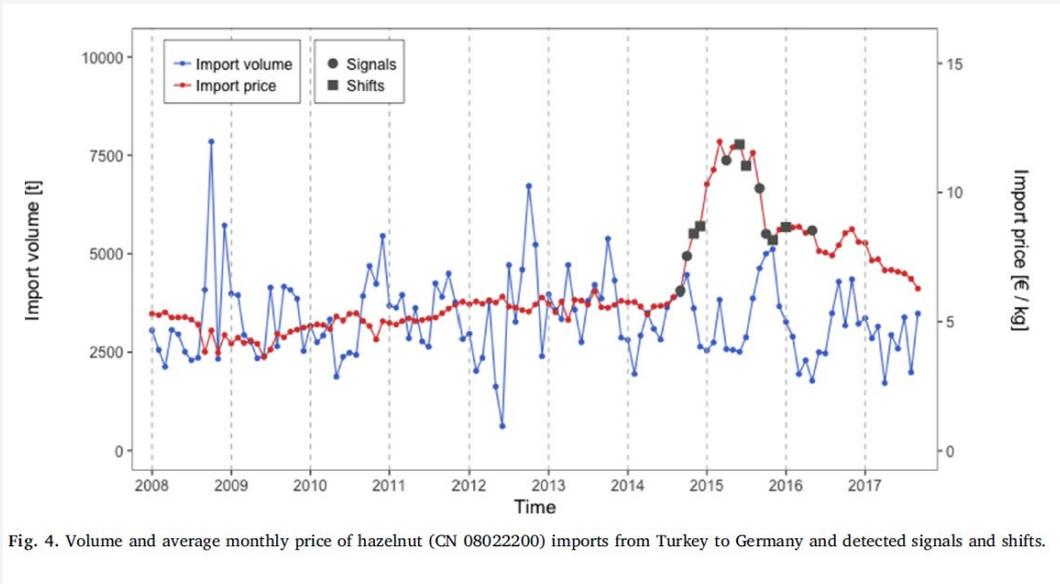


Fig. 4. Volume and average monthly price of hazelnut (CN 08022200) imports from Turkey to Germany and detected signals and shifts.



That's Cool. Computational Sociolinguistic Methods for Investigating Individual Lexico-grammatical Variation

Hans-Jörg Schmid^{1*}, Quirin Würschinger¹, Sebastian Fischer² and Helmut Küchenhoff²

¹Department of English and American Studies, LMU, Munich, Germany, ²Statistical Consulting Unit StaBLab, Department of Statistics, LMU, Munich, Germany

Linguistik: Nutzung von Konstruktionen



That's Cool. Computational Sociolinguistic Methods for Investigating Individual Lexico-grammatical Variation

Hans-Jörg Schmid^{1*}, Quirin Würschinger¹, Sebastian Fischer² and Helmut Küchenhoff²

¹Department of English and American Studies, LMU, Munich, Germany, ²Statistical Consulting Unit StaBLab, Department of Statistics, LMU, Munich, Germany



- Sammeln von komplexen Daten
- Beschreiben von komplexen Daten
- Auffinden von Strukturen in Daten
- Bildung von Modellen mit zufälligen Komponenten

*Statistics is a body of wise methods
for making wise decisions
in the face of uncertainty.*

- W.A. Wallis, H.V. Roberts -



- 20 HochschullehrerInnen + x
- über 60 MitarbeiterInnen
- Forschungsprojekte
- Kooperationen mit div. Partnern
- Statistisches Beratungslabor StaBLab
- Munich Center of Machine Learning
- Konrad Zuse School for Reliable AI
- AI Hub@LMU



www.stat.uni-muenchen.de
„Statistik lmu“



1. (Beratung bei der) **Planung** von Untersuchungen, Studien und Projekten
2. **Organisation** der Durchführung
3. **Analyse** der Daten
4. **Beratung** bei der Interpretation der Ergebnisse



- Fachwissen auf dem Gebiet der Fragestellung
- Kenntnisse der mathematisch-statistischen Theorie
- Fähigkeiten im Umgang mit Computern



- **Bachelor „Statistik und Data Science“**

6 Semester Grundlage mit Abschluss
- **Master „Statistics and Data Science“**

4 Semester



Pflichtveranstaltungen im 1. Semester:

Statistik I: Deskriptive Statistik &
W'keitstheorie

Einführung in die statistische Software

Analysis I

Nebenfach

Pflichtveranstaltungen im 2. Semester:

Statistik II: Deskriptive Statistik & W'keitstheorie

Einführung in die Wahrscheinlichkeitstheorie (V+Ü)[6]

Analysis II

Methoden der linearen Algebra



- Betriebswirtschaftslehre.
- Biologie
- Experimentalphysik
- Geographie
- Informatik
- Insurance and Risk-Management
- Philosophie
- Pädagogik / Bildungswissenschaft
- Politikwissenschaft
- Soziologie
- Theoretische Physik
- Volkswirtschaftslehre



Ein zentraler englischsprachiger Masterstudiengang, mit Spezialisierung auf einen der folgenden Aspekte:

- Machine Learning / Data Science
- Methodology and Modelling
- Social Statistics & Social Data Science
- Econometrics
- Biostatistics

Umfang: 4 Semester, 120 ECTS



- Einführung in die praktische Statistik („Anfängerpraktikum“)
- Kurse Statistische Software
- Statistisches Praktikum
- Consulting (Master)
- Interdisziplinäre Seminare
- Bachelor, Master Thesis
- Arbeiten als studentische Hilfskraft
- Jobs außerhalb der Universität
- Anwendungsgebiete



- Forschungseinrichtungen
- Medizin/Pharmazeutische Industrie
- Umweltforschung
- Biotechnologie
- Informationstechnologie / Softwareindustrie
- Banken
- Versicherungen
- Unternehmensberatungen
- Markt- und Meinungsforschung



<https://www.statistik.uni-muenchen.de/stellen/extern/index.html>

Suche unter Data Scientist Stellen



Vorteile

- Anwendungsbezug und Interdisziplinarität
- Vielfalt der Anwendungsgebiete
- Familiäre Atmosphäre
- Möglichkeiten zu HIWI-Jobs
- Auslandsaufenthalte
- Bezug zu aktuellen Forschungsthemen

Herausforderungen

- Hohe Anforderungen
(insbesondere **math. Grundlagen**)
- Verschiedene Ansprüche
- ?



Webseite:

<https://www.statistik.uni-muenchen.de>

<http://www.statistik.uni-muenchen.de/studieninteressierte/index.html>

„Under Cover“ Besuch