

# Statistik & Data Science aus Daten lernen und Zukunft gestalten

Helmut Küchenhoff  
Campus Tag

2023



## Was ist Statistik?

Fragen wir Wiki:

*“Statistics is the study of the collection, organization, analysis, interpretation and presentation of **data**. It deals with all aspects of this, including the planning of **data** collection in terms of the design of surveys and experiments.”*



- 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

- Datenmengen erhöhen sich drastisch
- Verfügbarkeit von Daten bietet viele Möglichkeiten
- IT Trend Nr 1
- “Data Science” “Big Data analytics” “Machine learning” “Algorithmen bestimmen unsere Zukunft”
- Zusammenarbeit von Informatik und Statistik der LMU in Forschung und Lehre



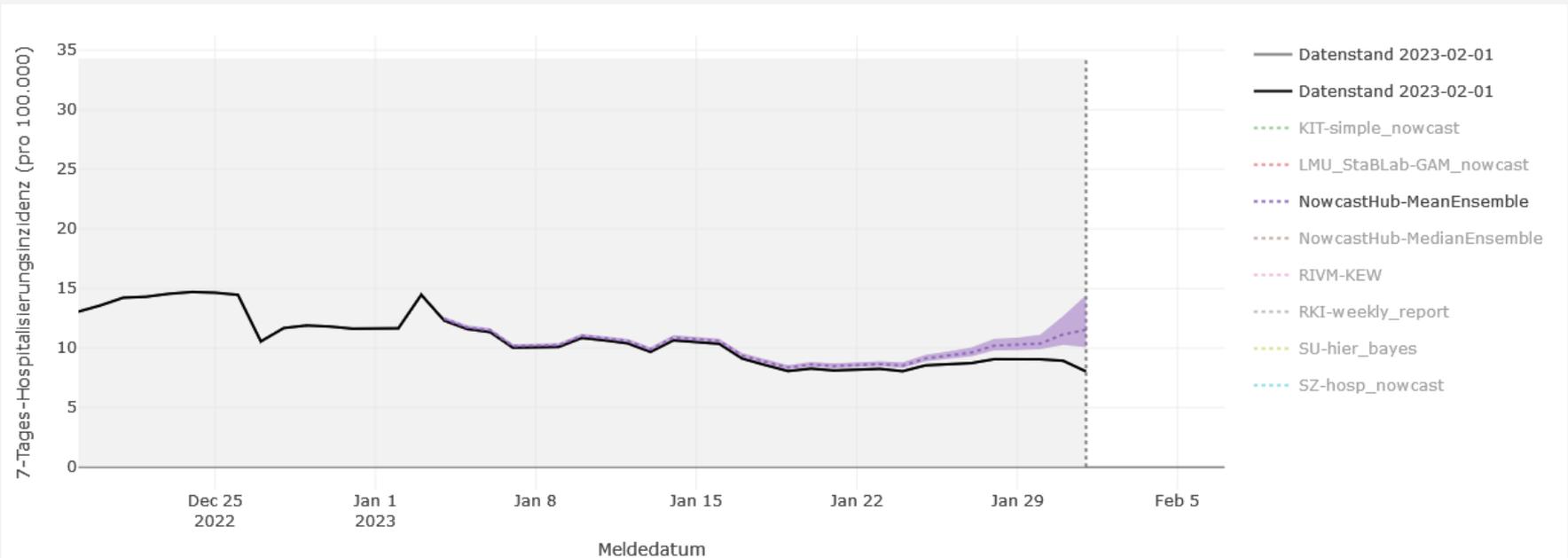
- Datensammler, Erbsenzähler, Lügner ??
- Datendetektive, Erzeuger von Wissen aus dem Rohstoff Daten !!
- Interdisziplinäre Forscher(innen), die auf vielen Gebieten in Wissenschaft und Wirtschaft arbeiten



<https://corona.stat.uni-muenchen.de/maps/>

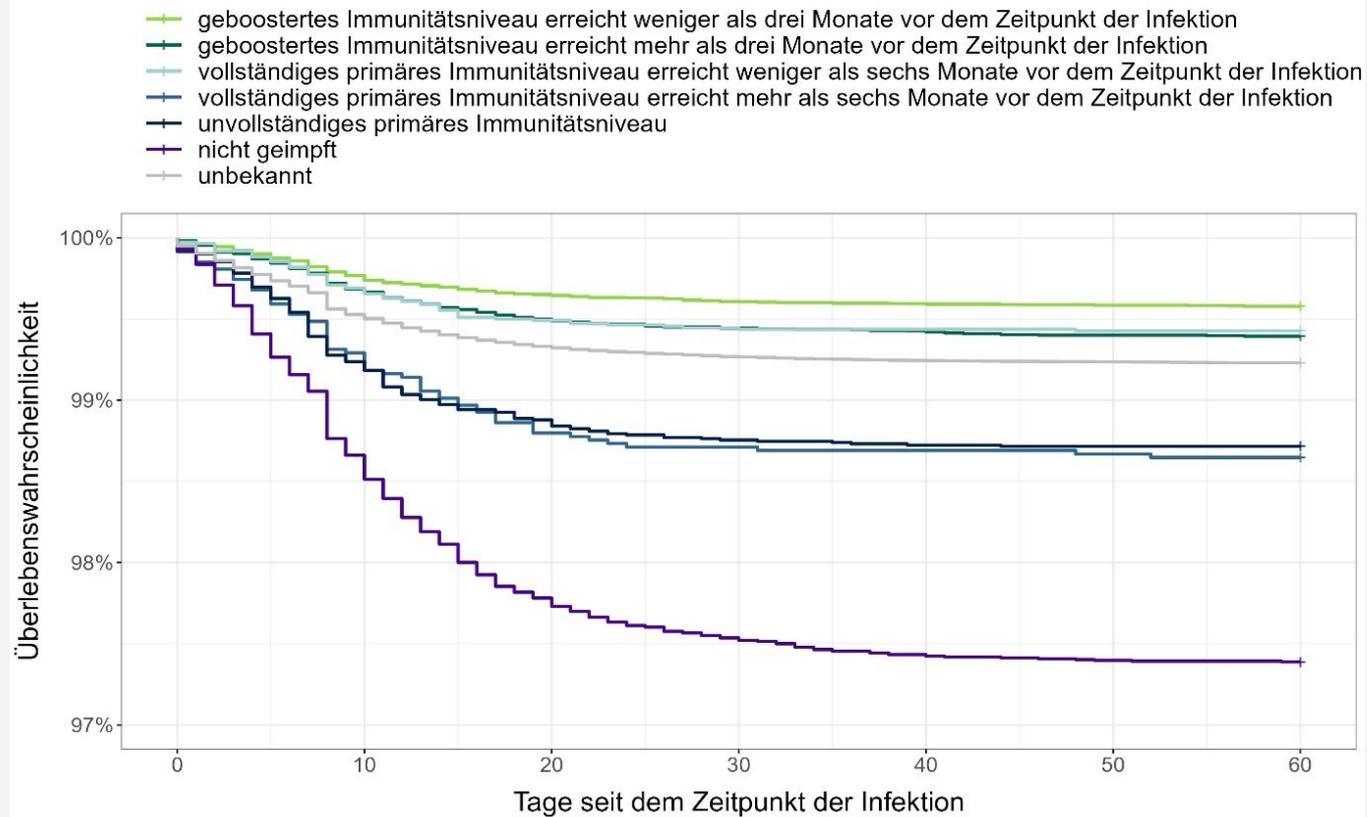
<https://covid19nowcasthub.de/>

## Lage in Bayern am 2.2.2023



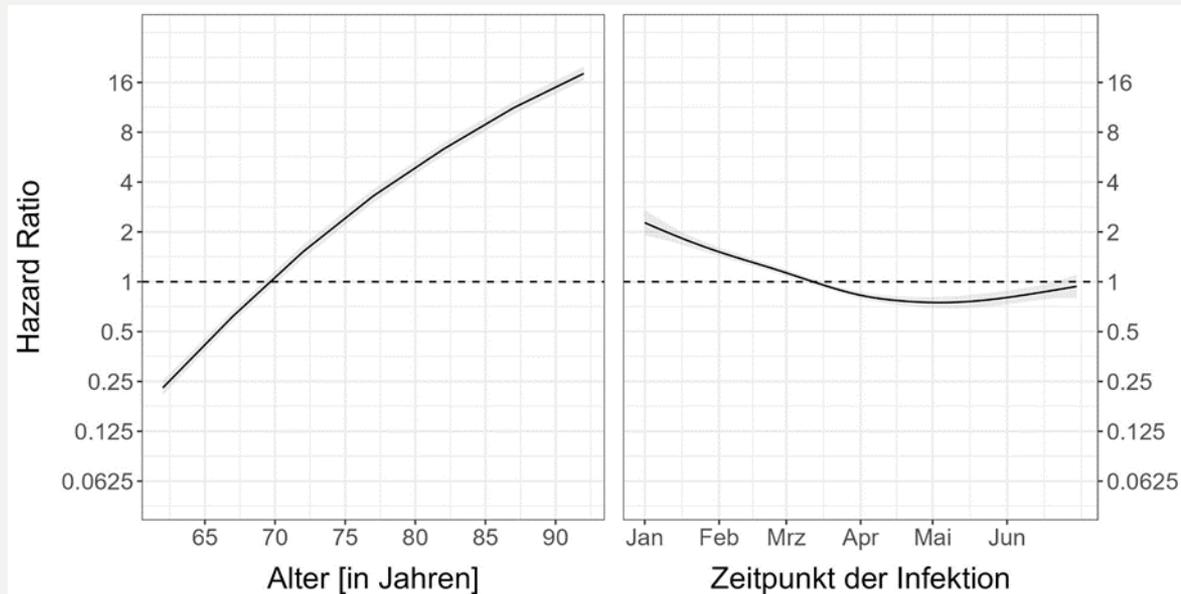


- Infizierte Ü60 in Bayern
- Zusammenhang Impfstatus/Vorinfektionen und Sterberisiko





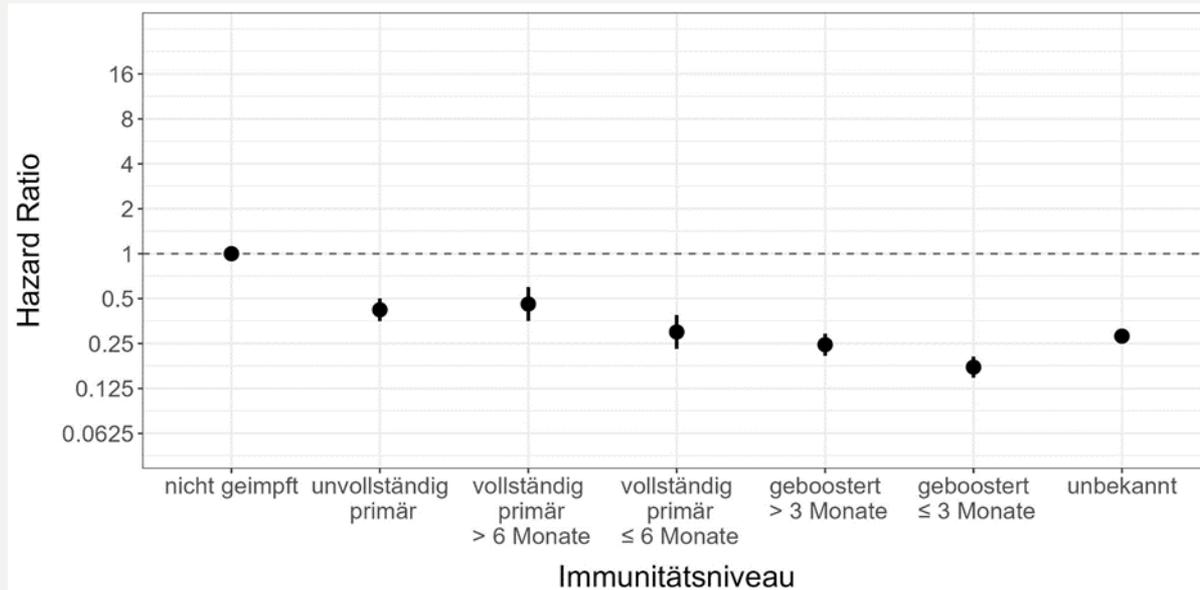
- Infizierte Ü60 in Bayern
- Zusammenhang Impfstatus/Vorinfektionen und Sterberisiko
- Berücksichtigung von Alter, Zeitpunkt der Infektion
- Statistisches Modell ermöglicht gleichzeitige Betrachtung verschiedener Einflüsse



Statistik & Data Science – DAS Berufsfeld im Zeitalter der Digitalisierung

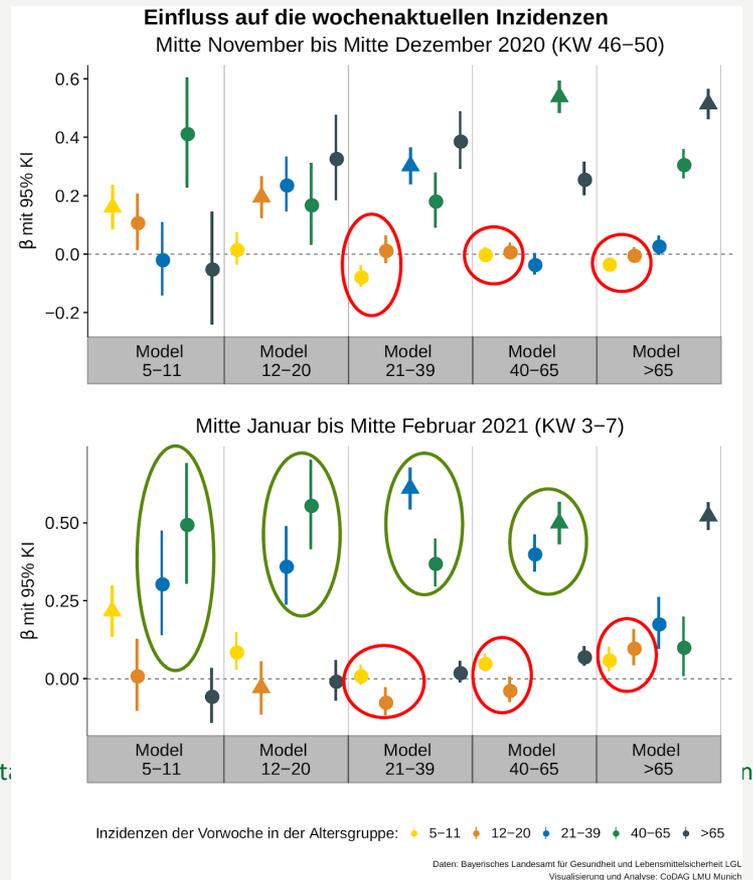


- Infizierte Ü60 in Bayern
- Zusammenhang Impfstatus/Vorinfektionen und Sterberisiko
- Berücksichtigung von Alter, Zeitpunkt der Infektion
- Statistisches Modell ermöglicht gleichzeitige Betrachtung verschiedener Einflüsse

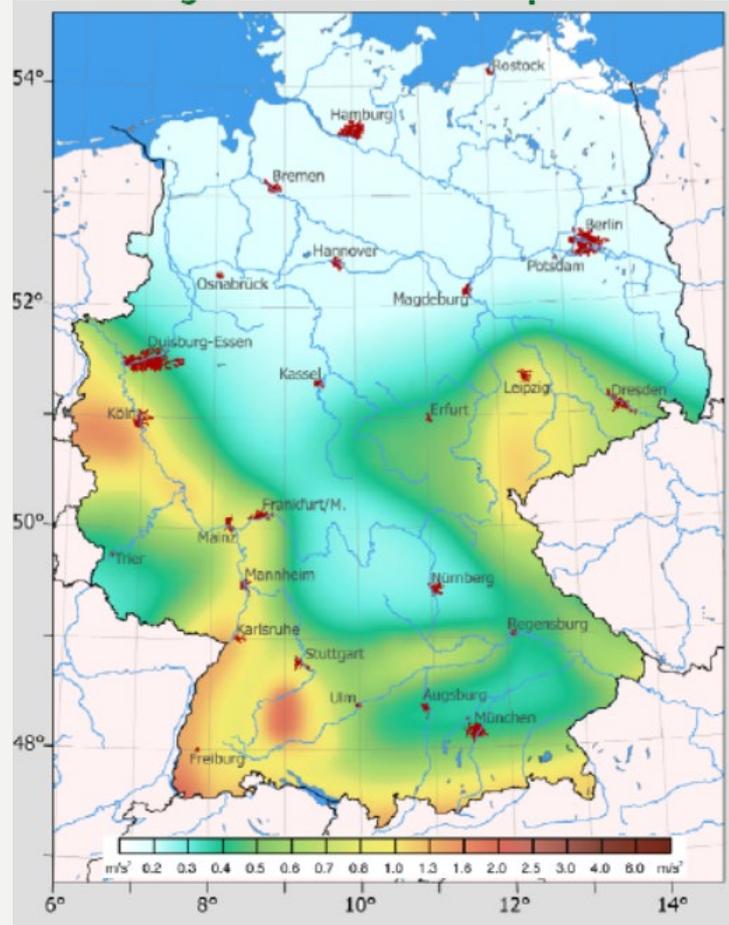




CODAG Bericht Nr. 10 vom 05.03.2021  
 Infektionen unter Schulkindern zeigen keinen Zusammenhang mit den  
 Infektionszahlen in anderen Altersgruppen









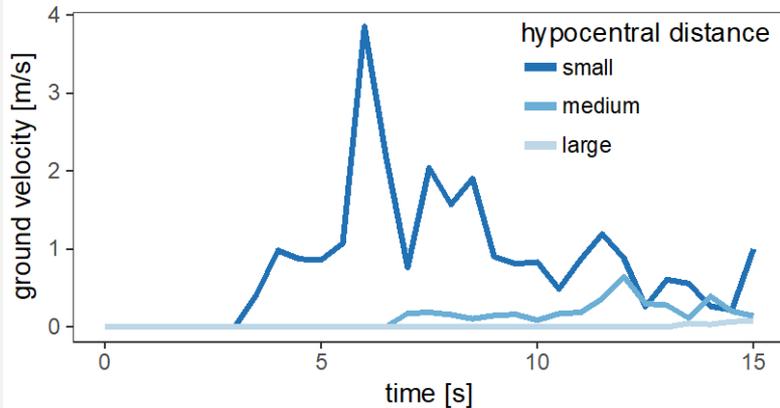
## Wie erhält man die Karte?

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

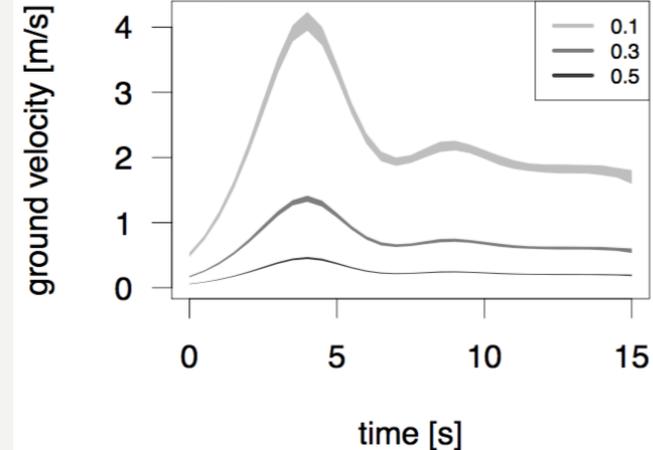


Wie stark sind die Bodenbewegungen, die man unter bestimmten physikalischen Bedingungen erwarten kann?

Typical observations  
by hypocentral distance



Absolute ground velocity  
by dynamic coefficient of friction  
(95% confidence intervals)



## Projekt in Kooperation mit Politikwissenschaft (Prof. Thurner)

- Wählerwanderung zwischen verschiedenen Wahlen
- Studentische Befragung
- Statistische Methodenentwicklung
- Non response, Erinnerungsvermögen der Befragten...



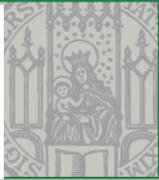
## Gibt es eine Midlife Crisis?

Analysen von Panel-Daten zur subjektiven  
Lebenszufriedenheit mit semiparametrischen  
Regressionsmodellen.

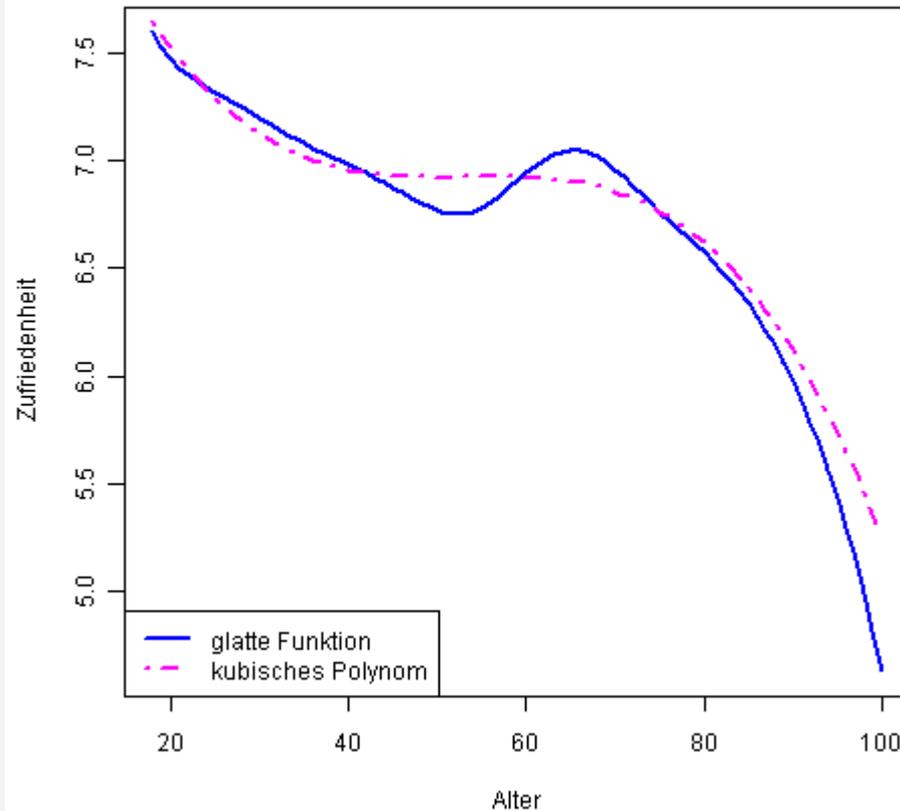
In Zusammenarbeit mit Sonja Greven, Andrea  
Wiencierz, Christoph Wunder



- 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\*



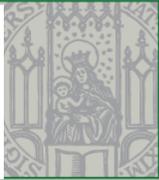
Midlife-Crisis nur bei glatter Funktion erkennbar.



Statistiker ziehen aus Daten Information  
Statistiker **quantifizieren** diese Information

Beispiel:

„Wen haben Sie gerade gewählt ? Was haben Sie  
bei der letzten Wahl gewählt ?“



- 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 -



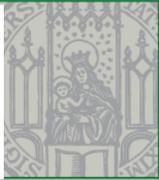
- 9 HochschullehrerInnen + x
- über 25 MitarbeiterInnen
- Forschungsprojekte (DFG, LGL)
- Kooperationen mit div. Partnern
- Statistisches Beratungslabor StaBLab
- Munich Center of Machine Learning
- Kooperation mit Informatik: Data Science



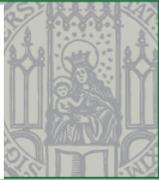
[www.stat.uni-muenchen.de](http://www.stat.uni-muenchen.de)  
„Statistik lmu“



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“**

Wählbarer Fokus:

Machine Learning, Biostatistics, Social Statistics and Data Science, Econometrics, Methodology and Modelling

**Elitestudiengang „Data Science“ (mit Informatik)**



## **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



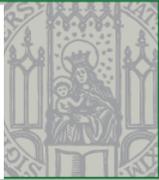
Ein zentraler Masterstudiengang, mit  
Spezialisierung auf einen der folgenden  
Aspekte:

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

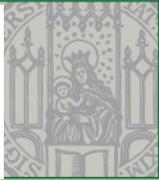
Umfang: 4 Semester, 120 ECTS



- 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
- Behörden und Ämter



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

Web-Portal LinkedIn zur Vermittlung von hochqualifizierten Arbeitskräften genutzt wird:

Liste der auf dem Arbeitsmarkt am meisten gefragten Fähigkeiten („hottest skills“).

## 1. „Statistical Analysis and Data Mining“

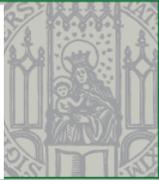


## Vorteile

- 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/index.html>

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

„Under Cover“ Besuch

Videos zu Veranstaltungen :

<https://cast.itunes.uni-muenchen.de/vod/playlists/ySRhdE8WHb.html>