



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

Prof. Dr. Dirk Ostwald (dirk.ostwald@ovgu.de)



Seit 2021	W2 Professur Methodenlehre I
2014 - 2020	W1 Professur Freie Universität Berlin
2010 - 2014	Postdoc BCCN & MPIB Berlin
2007 - 2010	PhD Psychologie Birmingham
2004 - 2006	MSc Neurowissenschaften Tübingen
2005 - 2012	BSc Mathematik Hagen
2000 - 2003	BSc Medizin Hamburg

Forschung	Komputationale Kognitive Neurowissenschaften
Lehre	Datenwissenschaft

(1) Einführung

Einführung

- Datenwissenschaft
- Statistik
- Modul B2 Inferenzstatistik
- Selbstkontrollfragen

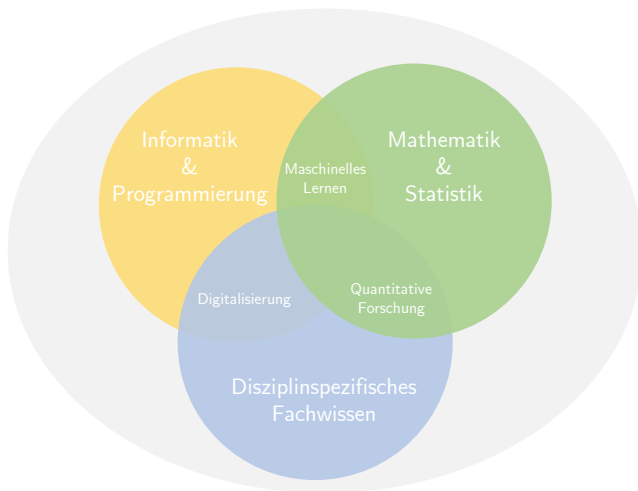
Einführung

- **Datenwissenschaft**
- Statistik
- Modul B2 Inferenzstatistik
- Selbstkontrollfragen

Datenwissenschaft

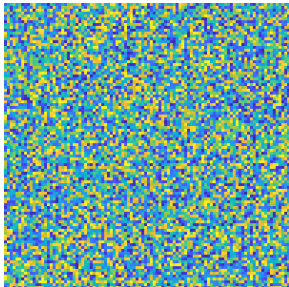
Die Kunst, aus Daten Sinn zu generieren

Datenwissenschaft



Datenanalyse ist Datenreduktion

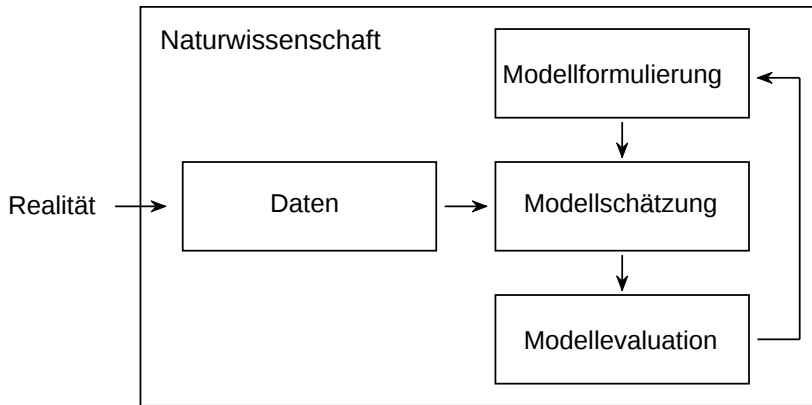
Rohdaten



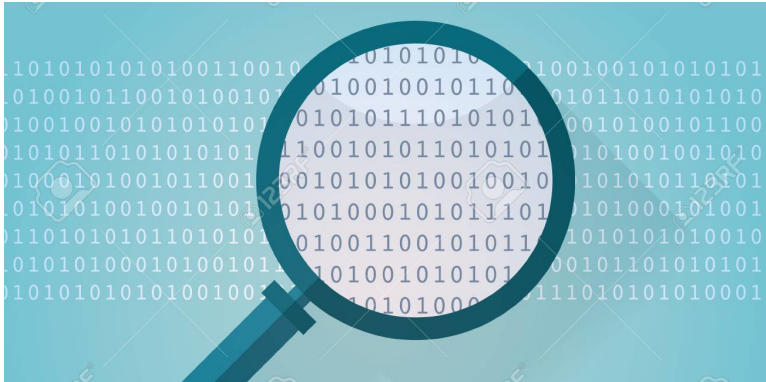
Reduzierte Daten



Datenanalyse ist modellbasiert



Datenanalyse ist Dateninterpretation



Datenwissenschaft

Statistik = Maschinelles Lernen = Künstliche Intelligenz

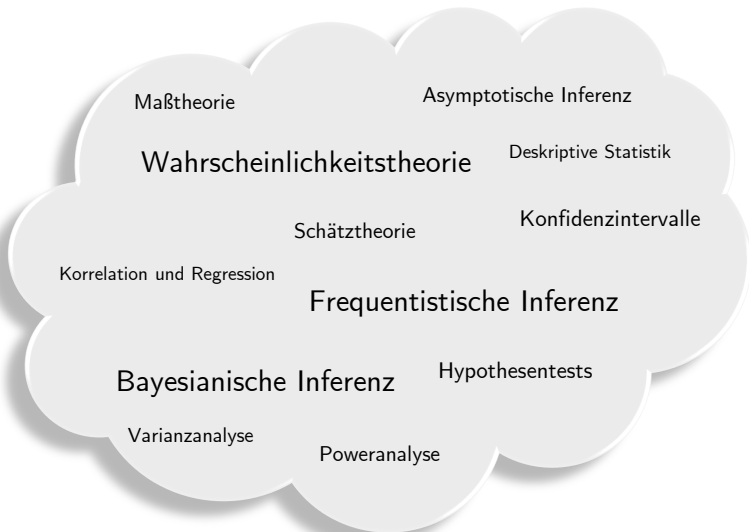
Statistik	Maschinelles Lernen	Künstliche Intelligenz
Probabilistische Modelle	Deterministische Modelle	Agenten-basierte Modelle
Theoretische Analyse	Klassifikation	Reinforcement learning
Optimalitätstheorie	Bayesianische Modelle	Symbolik
Asymptotische Theorie	Anwendung	Anwendung
Wissenschaftsphilosophie	Benchmarking	Hype

Einführung

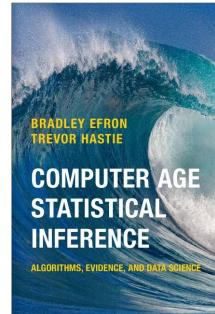
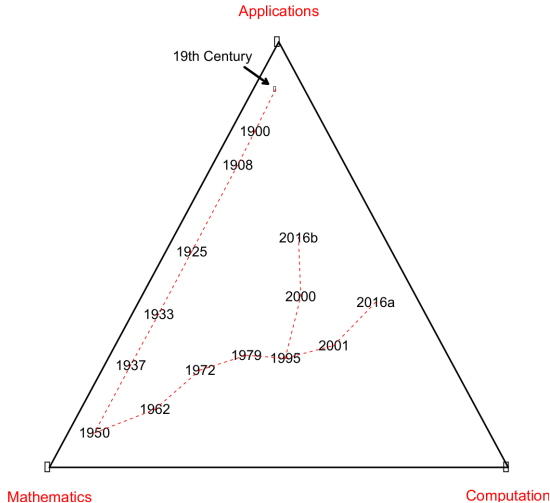
- Datenwissenschaft
- **Statistik**
- Modul B2 Inferenzstatistik
- Selbstkontrollfragen

Statistik

Die Kunst, aus Daten Sinn zu generieren
und seine Unsicherheit zu quantifizieren



Geschichte



Geschichte

1900	Pearsons Chi-Quadrat Test
1908	Studentsche t-Verteilung
1925	Fishers "Statistical Methods for Research Workers"
1933	Neyman und Pearsons Hypothesentesttheorie
1937	Neymans Konfidenzintervalle
1950	Walds Statistische Entscheidungstheorie
1950	Savages & de Finettis Bayesianische Entscheidungstheorie
1961	Raiffa & Schlaifers Statistische Entscheidungstheorie
1962	Tukeys "The future of data analysis"
1971	Lindleys Bayesianische Statistik
1979	Efrons Bootstrap
1995	Benjamini und Hochbergs Falscherkennungsrate
1996	Vapniks Statistische Lerntheorie
2000	Multiples Testen bei Microarray and Neuroimaginganalysen
2003	Beals Variational inference
2010	Renaissance neuronaler Netze als Deep Learning
2015	Datenwissenschaft

Fundamentale Annahmen der Wahrscheinlichkeitstheorie

- Zufallsprozesse können mathematisch modelliert werden.
- Mathematik kann zur Vorhersage zufälliger Ereignisse genutzt werden.
- Die Wahrscheinlichkeitstheorie ist mengentheoretisch begründet.

Fundamentale Annahmen der Frequentistischen Statistik

- Wahrscheinlichkeiten spiegeln die relative Frequenz des Auftretens eines zufälligen Ereignisses und beschreiben objektive Eigenschaften der realen Welt.
- Die Parameter probabilistischer Modelle sind feste, unbekannte Konstanten, die als *wahre, aber unbekannte, Parameterwerte* bezeichnet werden. Über Parameterwerte und Modelle werden keine probabilistischen Aussagen getroffen.
- Statistische Methoden werden so gestaltet, dass sie gute langfristige relative Frequenzeigenschaften besitzen und werden typischerweise anhand ihrer Stichprobenverteilungen bewertet.

Im BSc Psychologie lernen Sie nur Methoden der Frequentistischen Statistik kennen.

Fundamentale Annahmen der Bayesianischen Statistik

- Wahrscheinlichkeiten werden als Grade der Sicherheit, nicht als langfristige relative Häufigkeiten interpretiert. Aussagen der Form “Die Wahrscheinlichkeit, dass das Wintersemester 2021/22 in Präsenzlehre stattfindet, ist 0.7” haben eine Bedeutung.
- Die Parameter probabilistischer Modelle sind feste, unbekannte Konstanten, die als *wahre, aber unbekannte, Parameterwerte* bezeichnet werden. Über Parameterwerte und Modelle werden probabilistische Aussagen getroffen, die unseren Grad an Sicherheit hinsichtlich ihrer quantitativen Ausprägung und Validität widerspiegeln.
- Probabilistische Aussagen über Parameter werden mithilfe von Wahrscheinlichkeitsverteilungen getroffen, auf deren Grundlage optimale Entscheidungen im Sinne von Kosten-Nutzenfunktionen getroffen werden können.

Im MSc Psychologie lernen Sie auch Methoden der Bayesianischen Statistik kennen.

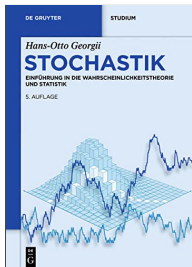
Einführung

- Datenwissenschaft
- Statistik
- **Modul B2 Inferenzstatistik**
- Selbstkontrollfragen

Modul B2 Inferenzstatistik Sommersemester 2021

Einheit	Datum	Thema
(0) Mathematische Grundlagen	Ad libitum	Allgemeine Grundlagen
(1) Einführung	07.04.2021	Allgemeine Grundlagen
(2) Wahrscheinlichkeitsräume und Zufallsvariablen	14.04.2021	Wahrscheinlichkeitstheorie
(3) Multivariate Verteilungen und Stichproben	21.04.2021	Wahrscheinlichkeitstheorie
(4) Erwartungswert und Kovarianz	28.04.2021	Wahrscheinlichkeitstheorie
(5) Ungleichungen und Grenzwerte	05.05.2021	Wahrscheinlichkeitstheorie
(6) Transformationen der Normalverteilung	12.05.2021	Wahrscheinlichkeitstheorie
(7) Parameterschätzung	19.05.2021	Wahrscheinlichkeitstheorie
(8) Schätzeigenschaften	26.05.2021	Grundlagen der Statistik
(9) Konfidenzintervalle	02.06.2021	Grundlagen der Statistik
(10) Hypothesentests	09.06.2021	Grundlagen der Statistik
(11) T-Tests	16.06.2021	Spezielle Verfahren
(12) Poweranalyse	23.06.2021	Spezielle Verfahren
(13) Einfaktorielle Varianzanalyse	30.06.2021	Spezielle Verfahren
(14) Zweifaktorielle Varianzanalyse	07.07.2021	Spezielle Verfahren
Klausur	Juli	
Wiederholungstermin	September	

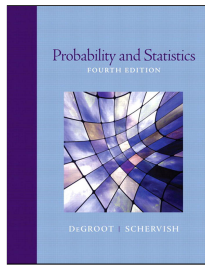
Weiterführende Literatur



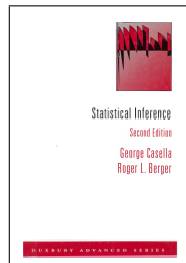
Georgii (2009)



Held and Sabanés Bové
(2014)



DeGroot and Schervish
(2012)



Casella and Berger (2012)

Ziele des Moduls

- Erwerben von grundlegenden Kenntnisse der Inferenzstatistik
- Tieferes Verständnis statistisch gesicherter wissenschaftlicher Erkenntnis
- Förderung des logischen, quantitativen, und methodenkritischen Denkens
- Theoretischer Hintergrund zu Modul C2 Computergestützte Datenanalyse

Keine Ziele des Moduls

- Wecken des Interesses an Wissenschaft oder psychologische Datenanalyse
⇒ Beides bringen Sie schon mit, sonst wären Sie nicht hier
- Extrinsische Lernmotivation
⇒ Wir setzen intrinsische Motivation voraus
- Durch Mathematikverzicht unverständliche “Statistik für Psycholog:innen”

Organisation des Moduls

- Synchrone Onlinevorlesung mittwochs 9 - 12 Uhr (7.4.2021 - 7.7.2021)
- Vorlesungsfolien und Vorlesungsvideolinks auf Abteilungswebseite
- Interaktive Q&A Session donnerstags 11 - 13 Uhr (8.4.2021 - 8.7.2021)
- Selbstkontrollfragen zur selbständigen Bearbeitung

Leistungsnachweis

- Klausur am Ende des Semesters
- Prüfungsnote entspricht der Klausurnote
- Klausurfragen angelehnt an Selbstkontrollfragen

Einige Ratschläge

- Bearbeiten Sie die Inhalte anhand der Selbstkontrollfragen wöchentlich aktiv nach.
⇒ Schauen Sie sich das Vorlesungsvideo ruhig nochmal an.
- Lernen Sie ihre Antworten auf die Selbstkontrollfragen auswendig.
⇒ Der Kurs ist sehr hierarchisch, Wissen aus Woche k wird in Woche $k + 1$ benötigt.
⇒ Mathematik ist auch nur Detailwissen!
- Lesen Sie alternative Darstellungen aus den Literaturhinweisen oder auf Wikipedia.
⇒ Prüfungsrelevant ist allerdings allein die Darstellung im Kurs.
- Bilden Sie Lerngruppen und/oder besuchen Sie das Tutorium.
⇒ Üben Sie, das Gelernte aktiv und laut wiederzugeben.
- Bereiten Sie Fragen für die Q&A Session am Donnerstagvormittag vor.

Fragen?

Einführung

- Datenwissenschaft
- Statistik
- Modul B2 Inferenzstatistik
- **Selbstkontrollfragen**

1. Diskutieren Sie den Begriff "Datenwissenschaft".
2. Nennen und erläutern Sie drei zentrale Charakteristika der Datenanalyse.
3. Diskutieren Sie Gemeinsamkeiten und Unterschiede von Statistik, Maschinellem Lernen und künstlicher Intelligenz.
4. Nennen und erläutern Sie drei fundamentale Annahmen der Wahrscheinlichkeitstheorie.
5. Nennen und erläutern Sie drei fundamentale Annahmen der Frequentistischen Statistik.
6. Nennen und erläutern Sie drei fundamentale Annahmen der Bayesianischen Statistik.

Literatur

Casella, G. and Berger, R. (2012). *Statistical Inference*. Duxbury.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.

Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. Aufl edition.

Held, L. and Sabanés Bové, D. (2014). *Applied Statistical Inference*. Springer Berlin Heidelberg, Berlin, Heidelberg.