

Angewandte Statistik I

INF3223 im Wintersemester 2018/19

Statistik

“Data science is the study of the generalizable extraction of knowledge from data” (Wikipedia)

Stochastik umschließt Wahrscheinlichkeitstheorie und Statistik

- beschreibende Statistik
- Wahrscheinlichkeitstheorie
- Schlußfolgernde Statistik

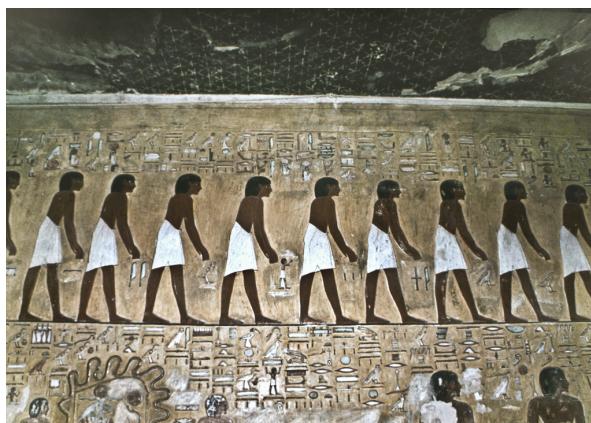
Geschichte

Frühzeit



Quelle: Clemens Schmillen, Wikipedia, CC BY-SA 3.0

Abgaben und Volkszählung in Ägypten



Tal der Könige. Quelle: Manfred Lentz, CC BY-NC-SA 2.0

Griechen, Babylonier

Musiklehre: arithmetisches, harmonisches, geometrisches Mittel



Sappho mit achtsaitiger Barbitos auf Vase. Quelle: Wikipedia, Public Domain.

Mittelalter, Aufklärung, Moderne

Stochastik [griech. *stochastikē* (*téchnē*) = zum Erraten gehörend(e Kunst)
(Begriff von Jakob Bernoulli)



- Jakob Bernoulli (Schweiz, 16. Jh)
- Thomas Bayes (England, 18. Jh)
- Carl Friedrich Gauß (D, 19. Jh)
- Gregor Mendel (D, 19. Jh)

Heute

I keep saying the sexy job in the next ten years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s? The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it [...]

(Hal Varian, Chief Economist at Google. 2008)

Schlagworte

**big data, Kostenexplosion, Volkszählung, Datenkrake
Wirtschaftswissenschaftliche Spieltheorie, Aktienkurse, ...
Wissenschaft, machine learning, ...**

Beispiele

Chancen im Lotto, T€uro? Globale Erderwärmung, Kausalität, Lerneffekt in Wahrnehmung, ...

Ziele

Statistik-Werkzeuge mit denen der wissenschaftliche Alltag bearbeitet wird.

- Kennenlernen
- Verstehen
 - dazu Theorie
- Anwendungsbereiche
- Anwenden
- Programmieren
- Ergebnis
 - interpretieren und darstellen

Über mich

Physiker, Biologe, Elektroingenieur, Administrator.

Kein Statistiker

Alternativ besuchen Sie bitte eine spezielle Vorlesung z.B. Statistik I+II in den Kognitionswissenschaften oder Mathematik.

Ich werde versuchen, Ihnen die Verbindung von all dem, die *praktische* Statistik nahezubringen.

Über Sie

Was sind Ihre Vorkenntnisse in

1. Mathematik
 - Lineare Algebra
 - Analysis
2. Statistik
 - Kombinatorik
 - Wahrscheinlichkeitsrechnung
3. Programmieren
4. Python
5. Notebooks

Was sind Ihre Erwartungen?

Spezielle Problemstellung?

Formalien

Links

Homepage *Neuronale Informationsverarbeitung*

<http://www.nip.uni-tuebingen.de/teaching/lectures-seminars/winter-term-201819/zeitplan-angewandte-statistik-i.html>
<http://www.nip.uni-tuebingen.de/teaching/lectures-seminars/winter-term-201819/zeitplan-angewandte-statistik-i.html>

Informatik

<http://www.wsi.uni-tuebingen.de/studium> (<http://www.wsi.uni-tuebingen.de/studium>)

ilias

- Veranstaltungen
 - Wintersemester 2018-2019
 - 7 Mathematisch-Naturwissenschaftliche Fakultät
 - Informatik
 - Neuronale Informationsverarbeitung (Prof. Wichmann)
 - Angewandte Statistik I (WiSe 2018/19)

Anmeldung mit Passwort

- XXXXX
- Bei Problemen bitte Email an mich (Siehe Homepage)

Folien der Vorlesung, Skript

Email an Mitglieder

- hier findet alle Kommunikation statt
- Terminänderungen falls nötig

Forum

- Für Fragen
- ... und Antworten

Übungsaufgaben

- **Angewandte** Statistik
- Zu Ihrer eigenen Verständnisvertiefung
- jede / jede zweite Woche
- Besprechung in der Übungsstunde
- Voraussetzung für Teilnahme an Prüfung: 50% richtig bearbeitet
 - Bonusregelung?
- Nicht Teil der Benotung

Klausur

- Schriftliche Klausur
- Termin: 20. Februar 2019 14:15 (=Vorlesung)
- Raum: F119 Sand 6
- Dauer: 90 Minuten
- Voraussetzung: 50% der Punkte aus den Übungsaufgaben

Anmeldung zur Klausur

Siehe Prüfungsordnung bzw. neue Richtlinien.

Prüfungsordnungen 2010, 2015 und 2016

Anmeldeverfahren

Studierende der X-INF- Studiengänge müssen alle Prüfungen (Haupt- oder Nachprüfung) anmelden, für die sie eine Note verbucht haben wollen. Bei Nichtbestehen einer Klausur gibt es keine automatische Anmeldung zu Nachklausuren. Auch Master-Studierende nach der Prüfungsordnung 2016, d.h. alle Masterstudierenden, die seit WS 2016/17 ihr Studium begonnen haben, können sich jetzt über Campus anmelden und ihre Noten über Campus einsehen.

Kognitionswissenschaft: Wenn die Prüfungsanmeldung in Campus nicht möglich ist, melden sich Studierende der Kognitionswissenschaft beim jeweiligen Dozenten an und nicht schriftlich im Prüfungssekretariat.

Anmeldefristen

Die Anmeldung zu Prüfungen ist ab 1.12.2018 möglich. Die Anmeldeperiode endet in Campus einen Werktag vor der jeweiligen Prüfung.

Rücktritt von Prüfungen

Falls Studierende aus gesundheitlichen Gründen an einer angemeldeten Prüfung nicht teilnehmen können, müssen sie innerhalb von 3 Werktagen nach dem Prüfungstermin dem Prüfungssekretariat ein ärztliches Attest vorlegen. Andernfalls wird die Prüfung als nicht erschienen verzeichnet und mit 5,0 bewertet.

auch abmelden, falls nicht aufgrund der Kriterien innerhalb des Kurses zugelassen!

Veranstaltungsspezifische Regeln für die Teilnahme an Nachklausuren

Die Zulassung zu einer eventuellen Nachklausur in Veranstaltungen des FB Informatik ist veranstaltungsspezifisch geregelt, wird zu Beginn des Semesters in der Veranstaltung bekanntgegeben

Eventuelle Nachprüfung mit Hauptprüfungspflicht

Es gibt eine Hauptprüfung (schriftlich) und es liegt im Ermessen des Dozenten, ob es eine Nachprüfung (hier mündlich) gibt, was auch erst vom Ausgang der Hauptprüfung abhängig gemacht werden kann. An der Nachprüfung kann nur teilnehmen, wer an der Hauptprüfung teilgenommen hat.

Terminübersicht

(vorläufig)

Datum	Titel
17.10.2018	Einführung: Statistik, Python
24.10.2018	Beschreibende Statistik
31.10.2018	(--Allerheiligen--)
07.11.2018	Korrelation und Regression
14.11.2018	Wahrscheinlichkeitstheorie
21.11.2018	Diskrete Verteilungen
28.11.2018	Kontinuierliche Verteilungen
05.12.2018	Sätze der Statistik
12.12.2018	Schätzer, Chi-Quadrat
19.12.2018	Tests
26.12.2018	(--Weihnachtsferien--)
02.01.2019	(--Weihnachtsferien--)
09.01.2019	t-Test, Vertrauensintervalle
16.01.2019	Bootstrapping, Varianzanalyse
23.01.2019	Lineares Modell, OLS, Pandas, Patsy
30.01.2019	OLS, Versuchsplanung
06.02.2019	Repetitorium
20.02.2019	Klausur

Übungsstunden

- Ergänzung zur Vorlesung
- Besprechen der Übungsaufgaben
- Fragen

Fragen

- Verständnisproblem? Sofort fragen.
- Am Kapitelende
- Nach den Vorlesungen
- In den Übungsstunden
- Per ilias-Mail

Rückmeldungen

- gerne offen für Korrekturen und Verbesserungsvorschläge

Fragen?

Zeitaufwand

- 6 ECTS
- Vorlesung
 - Nachbereitung
- Übungsaufgaben
 - *Angewandte Statistik = Praxis*
 - Damit arbeiten
 - Verständnis überprüfen (Sie und ich)
- Übungsstunden
 - Besprechen der Aufgaben

Python

- *Universelle* Programmiersprache
- zahlreiche Bibliotheken
- Open Source
- Unterstützte Plattformen:
 - Linux, Windows, Mac, ...

Anaconda / Miniconda

Einfach zu installierende Distribution

Don't panic

Sie werden mit kleinem Aufwand Python *im Vorbeigehen* lernen und hoffentlich großen Nutzen daraus ziehen.

Inhalt Vorlesung

0. Python

- Anaconda Distribution
- numpy
- scipy
- matplotlib

1. Beschreibende Statistik

- absolute und relative Häufigkeit
- kummulierte Häufigkeitsverteilung
- Kennwerte
- Darstellungsformen

1.1 Zweidimensionale Verteilungen

- Kontingenz
- Korrelation
- Regression

2. Wahrscheinlichkeitstheorie

- Wahrscheinlichkeit
- Zufallsvariable
- Axiome von Kolmogoroff
 - Laplace Wahrscheinlichkeit
 - Verbundwahrscheinlichkeit
- bedingte Wahrscheinlichkeit
 - Unabhängigkeit
- Theorem von Bayes
- Kennzahlen
 - Erwartungswert
 - Varianz
 - Schiefe und Kurtosis
- Zufallszahlen
- Gesetz der großen Zahl

2.1 diskrete Verteilungen

- Binomial
- Poisson
- Multinomial
- Summen

2.2 kontinuierliche Verteilungen

- Wahrscheinlichkeitsdichte pdf
- kummulierte Verteilungsfunktion cdf
- Perzentil-Funktion
- Uniform
- Exponential
- Normal
- Gemeinsame und bedingte Verteilung
- Korrelation
- Funktion mehrere Zufallsvariablen

3. Schließende Statistik

3.1 Sätze der Statistik

3.2 Schätzungen

- Maximum Likelihood Methode
- Robuste Schätzungen

3.3 Tests

- Teststatistik
- einzelne und gepaarte Stichproben
- p-Wert
- Sinn und Unsinn

3.4 Vertrauensintervalle

- verschiedene Verteilungen
- Bootstrap Methode
- Vergleich von Schätzung, Test und Vertrauensintervall

4. Datenanalyse

- Nominale Daten
 - Chi-Quadrat-Test
- Varianzanalyse
 - Mehrere Stichproben
 - Multiple Vergleiche und Tests
 - verbundene Stichproben

4.1 Multivariate Statistik

- Mehrdimensionale Zufallsvariable
- Schätzung von Erwartungswert und Kovarianz
- Mehrdimensionale Normalverteilung
- Regression
- multiple Regression

5. Lineare Modelle

- $y = f(x)$
- abhängige und unabhängige Variable
- Parameterschätzung
- Zusammenhang mit
 - t-Test
 - Anova
 - Regression

6. Versuchsplanung

- Stichprobenplanung
- Auswertung von Umfragen

Weiterführende Vorlesungen

- **Angewandte Statistik II** (Fortsetzung: Komplexere Statistik, Bayes-Statistik)
 - Hauptkomponentenanalyse PCA
 - Unabhängigkeitsanalyse ICA
 - general linear models GLM
 - Bayes Statistik
- **Kapitel Statistik** (Informatik im Rahmen der Mathematik IV)
- **Statistik I+II für Kognitionswissenschaftler** (mathematische Behandlung)
- **trees** (weiterführend, spezielle Statistik)
- **Zeitreihen** (weiterführend, spezielle Statistik)
- **Machine Learning / Intelligent Systems** (weiterführend)

Literatur

in der UB mehrfach vorhandene Bücher

- Fahrmeir, Künstler, Pigeot, Tutz: Statistik, Springer-Verlag Berlin Heidelberg, 6. Auflage 2007
- Stahel: Statistische Datenanalyse, Vieweg&Sohn, 5. Auflage 2008

wissenschaftliche Artikel, Quellen im Internet

- zum jeweiligen Kapitel

Weiter

[Python installieren \(002_PythonInstallation.ipynb\)](#)

[Motivation Python \(003_MotivationPython.ipynb\)](#)

Fragen?