

FAU – Fachbereich Wirtschafts- und Sozialwissenschaften
Lehrstuhl für Statistik und Ökonometrie
Statistik (RUW-2172/2170)
Wintersemester 2019/20 – Syllabus

Vorlesung: Dienstag, 11:30-13:00/Donnerstag, 11:30-13:00, LG H4 (in den ersten vier Wochen zusätzlich auch LG H1).
Die erste Vorlesung findet am 22. Oktober statt. Eine Anmeldung ist nicht erforderlich.

Übung: Montag, 16:45-18:15, LG H2
Dienstag, 08:00-9:30, LG H3
Dienstag, 15:00-16:30, LG H3
Mittwoch, 18:30-20:00, LG H1
Donnerstag, 08:00-9:30, LG H6
Donnerstag, 16:45-18:15, LG H1
Bitte beachten Sie, dass Sie nur den einen Übungstermin besuchen können, für den Sie sich angemeldet haben. Die Anmeldung erfolgt über StudOn und ist ab dem 4. Okt. 2019 (18:00 h) möglich. Die Übungen starten ab dem Nachmittag des 22. Oktobers. Studierende, die für die Übungen am Montag oder Dienstagmorgen angemeldet sind, besuchen in der ersten Woche bitte ausnahmsweise eine der anderen Übungen.

Tutorium: Montag, 08:00–9:30, LG H3
Montag, 11:30-13:00, LG 0.222/3
Montag, 13:15-14:45, LG 0.222/3
Dienstag, 9:45-11:15 LG 0.142
Dienstag, 13:15-14:45 LG 0.424
Mittwoch, 8:00-9:30, LG 0.144
Mittwoch, 9:45-11:15, LG H2
Donnerstag, 9:45-11:15, LG 0.142
Donnerstag, 13:15-14:45, LG H2
Freitag, 11:30-13:00, LG 0.222/3
Die ersten Tutorien finden in der Woche ab dem 28. Oktober statt. Bitte beachten Sie, dass Sie nur den einen Tutoriumstermin besuchen können, für den Sie sich angemeldet haben. Die Anmeldung erfolgt über StudOn und ist ab dem 5. Oktober 2019 (18:00 h) möglich.

R-Übung: Mittwoch, 13:15-14:45, LG H4.
Die erste R-Übung findet am 23. Oktober statt. Eine Anmeldung ist nicht erforderlich.

R-Tutorium: Dienstag, 9:45-11:15, LG 0.422
Mittwoch, 8:00-9:30, LG 0.422
Die ersten R-Tutorien finden in der Woche ab dem 28. Oktober statt. Eine Anmeldung ist nicht erforderlich

Dozent:

Jonas Dovern

Raum: LG 4.169

E-Mail: jonas.dovern@fau.de

Sprechstunde: nach Vereinbarung

Koordinator für Übungen/Tutorien

Matthias Schnaubelt

Raum: LG 4.174

E-Mail: Matthias.schnaubelt@fau.de

Sprechstunde: nach Vereinbarung

Koordinator für R-Übung/R-Tutorien

Alexander Glas

Raum: LG 4.173

E-Mail: alexander.glas@fau.de

Sprechstunde: nach Vereinbarung

Kurzbeschreibung: Statistik und Wahrscheinlichkeitsrechnung verfolgen uns tagtäglich. Hilft ein bestimmtes Medikament gegen Kopfschmerzen? Wie lange warte ich im Schnitt auf die U-Bahn? Können Frauen besser Autofahren als Männer? Und natürlich sind statistische Verfahren auch nicht aus der Welt der BWL, VWL oder Soziologie wegzudenken. Zieht ein neues Webseiten-Design mehr Käufer an als das alte? Wie viele Kunden sollte eine Firma befragen, um belastbare Aussagen über die Interessen ihrer gesamten Kundschaft treffen zu können? Wie hoch ist das durchschnittliche Haushaltseinkommen in Deutschland? In welchem Land sind die Einkommen besonders ungleich verteilt? Dauert es im Schnitt weniger als X Monate bis ein(e) Arbeitslose(r) wieder einen Job gefunden hat?

Dieser Kurs gibt eine Einführung in statistische Methoden, die dazu genutzt werden können, um i) Daten zu beschreiben, ii) Datenerhebungen zu planen, iii) von Stichproben auf ganze Populationen zu schließen, iv) Zufallsvorgänge zu analysieren und v) große Datensätze mittels Maschinellen Lernens zu analysieren.

Da mehr und mehr Berufsbilder in Wirtschaft und Forschung das praktische Arbeiten mit Datensätzen beinhalten, bietet der Kurs auch eine Einführung in die praktische Datenanalyse auf Basis der Statistiksoftware R.

Studierende, die den Kurs erfolgreich abschließen, können i) Datensätze mittels geeigneter Maßzahlen beschreiben und geeignete Formen zur Visualisierung von Daten auswählen, ii) grundlegende Stichprobenverfahren erläutern, iii) Parameter von Verteilungen schätzen und verschiedene Arten von Hypothesentests anwenden, iv) induktive Methoden als Grundlage des Arbeitens in empirischen Wissenschaften einsetzen und statistische Ergebnisse kritisch überprüfen, v) die Funktionsweise ausgewählter Verfahren des Maschinellen Lernens erklären und vi) die behandelten Methoden praktisch in der Statistiksoftware R implementieren.

Prüfung: Das Modul wird mit einer schriftlichen Klausur (120 Minuten) am Ende des Semesters geprüft. Die Klausur besteht zu rund 2/3 aus einem Multiple-Choice-Teil und zu rund 1/3 aus (Teil-)Aufgaben, die frei beantwortet werden müssen. Für Studierende, die die Kursversion RUW-2172 belegen (im Prinzip alle außer Studierende des Wirtschaftsingenieurwesens), beinhaltet die Klausur auch Fragen zur praktischen Umsetzung statistischer Analysen in R.

Klausurtermin: Wird später vom Prüfungsamt bekanntgegeben.

Themen des Kurses

Deskriptive Statistik:

- Wofür braucht man Statistik?
- Welche Arten von Daten gibt es?
- Wie kann man die Verteilung von Daten darstellen?
- Wie misst man den „Durchschnitt“ einer Variablen?
- Welche Maße geben sonst noch Auskunft über die Verteilung einer Variablen?
- Wie misst man Konzentration in Märkten bzw. Ungleichheit (z.B. der Einkommen)?
- Wie kann man Datensätze mit mehreren Variablen auswerten?
- Wie misst man einen (linearen) Zusammenhang zwischen zwei Variablen?
- Wie visualisiert man Daten am besten?

Stichprobenverfahren:

- Wie erhebt man eigentlich eine Datenstichprobe?

Wahrscheinlichkeitsrechnung:

- Was ist eine Wahrscheinlichkeit?
- Wie kann man bedingte Wahrscheinlichkeiten berechnen?
- Wie viele Möglichkeiten gibt es bestimmte Elemente zu kombinieren?
- Was ist eine Zufallsvariable und wie beschreibt man sie mathematisch?
- Was gibt es für wichtige Verteilungsfunktionen für diskrete Daten?
- Was gibt es für wichtige Verteilungsfunktionen für stetige Daten?

Schließende Statistik:

- Was kann man aus einer zufälligen Stichprobe schließen?
- Wie kann man einen unbekanntem Verteilungsparameter schätzen?
- Wie kann man Schätzunsicherheit darstellen?
- Wie kann man testen, ob ein Parameter einen bestimmten Wert hat?
- Wie kann man testen, ob zwei Stichproben ähnlich sind?
- Wie kann man testen, welche Verteilung am besten zu den Daten passt?
- Wie testet man, ob es einen Zusammenhang zwischen zwei Variablen gibt?

Grundlagen des Maschinellen Lernens:

- Was ist Maschinelles Lernen?
- Wie kann man Beobachtungseinheiten datengetrieben in Gruppen aufteilen?
- Wie kann man Zufallsvariablen auf Basis anderer Information voraussagen?

Die Geschichte der Statistik/Wahrscheinlichkeitsrechnung

Lehrbücher:

- Schlittgen, Rainer (2012), Einführung in die Statistik – Analyse und Modellierung von Daten (12. Auflage), Oldenbourg Wissenschaftsverlag GmbH, München. [Hauptlehrbuch; als elektronische Version in der FAU-Bibliothek verfügbar.]
- James, Gareth, Daniela Witten, Trevor Hastie und Robert Tibshirani (2013), An Introduction to Statistical Learning, Springer, Heidelberg. [Einige Exemplare in der FAU-Bibliothek verfügbar; als elektronische Version unter <http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf> verfügbar.]
- Kauermann, Göran und Helmut Küchenhoff (2010), Stichproben – Methoden und praktische Umsetzung in R, Springer, Heidelberg. [Als elektronische Version in der FAU-Bibliothek verfügbar.]

Aufbau des Kurses: Die Veranstaltung bietet Ihnen vielfältige Möglichkeiten die Grundlagen der Statistik zu lernen. Wöchentlich können Sie 2 Vorlesungen, 1 Übung, 1 Tutorium, 1 R-Übung und 1 R-Tutorium besuchen. Um einen optimalen Lernerfolg zu erzielen, empfehlen wir Ihnen jeweils i) die relevante Literatur vor den Vorlesungen anzuschauen, ii) die Übungsaufgaben eigenständig vor den Übungen und Tutorien zu rechnen und, falls Sie Probleme mit einer Aufgabe hatten, nach den Übungen/Tutorien nochmal zu probieren (ohne die Lösung daneben liegen zu haben) und iii) die Implementation des Stoffs in R eigenständig zu üben. Wichtig ist, dass Sie ab der ersten Woche mitarbeiten, weil viele Elemente des Kurses aufeinander aufbauen und Sie sonst schnell den Anschluss verlieren.

Zusätzlich zu den Präsenzveranstaltungen gibt es auf StudOn noch wöchentliche Online-Quizzes, die Sie dazu nutzen können Ihren Lernerfolg zu überprüfen. Die einzelnen Quizzes sind jeweils nur für eine Woche freigeschaltet und beziehen sich auf die Inhalte der vorherigen Vorlesungen. Auch hier müssen Sie also kontinuierlich am Ball bleiben, wenn Sie von diesem Angebot profitieren möchten.

Informationen zu R-Übung und R-Tutorium

In der R-Übung wird gezeigt, wie mit Hilfe des Statistikprogrammes R der aus Vorlesung und Übung bekannte Stoff auf reale Datensätze angewendet werden kann. Zusätzlich zu den in den Übungen besprochenen Aufgaben gibt es Übungsblätter, die der eigenständigen Kontrolle Ihres Lernfortschritts dienen. Bei Problemen mit diesen Aufgaben können Sie Hilfe im Rahmen des R-Tutoriums bekommen. In diesen Tutorien können auch Fragen zu R im Allgemeinen geklärt werden und es wird Hilfe bei der Vor- und Nachbereitung der Aufgaben aus der R-Übung angeboten. Eine Vermittlung von zusätzlichen Inhalten in Form von Frontalunterricht findet in den R-Tutorien nicht statt. (Bitte beachten Sie, dass diese Veranstaltungen für Studierende des Wirtschaftsingenieurwesens nicht verpflichtend sind; die Studierenden können aber natürlich gerne freiwillig an dem Lehrangebot teilnehmen.)

Informationen zu R/R-Studio

R ist eine freie Programmiersprache, die vor allem für statistische Analysen und zur Visualisierung von Daten genutzt wird. R-Studio ist eine integrierte Entwicklungsumgebung und grafische Benutzeroberfläche für R (ein „Editor“).

Für diesen Kurs werden wir auf das Angebot *R Studio Cloud* zurückgreifen, das es Ihnen erlaubt, die R-Aufgaben über einen Browser zu bearbeiten ohne R/R-Studio auf Ihrem Computer zu installieren. Um sich dort für den Kurs zu registrieren, müssen Sie die beiden folgenden Schritte befolgen:

1. Sie müssen sich auf <https://rstudio.cloud/> registrieren.
2. Über [diesen Link](#) können Sie dem Bereich für diesen Kurs beitreten.

Dieses Anmeldeverfahren wird auch in der ersten R-Übung nochmal erklärt. Alle Übungsblätter, Datensätze und sonstige Dateien, die Sie zur Bearbeitung der R-Aufgaben im Kurs benötigen, werden wir über *R Studio Cloud* bereitstellen.

Wie Sie R und R-Studio zusätzlich auf Ihrem Computer installieren können (um z.B. für eine Seminar- oder Abschlussarbeit damit zu arbeiten oder auch offline Zugriff auf R zu haben), wird in einer der Übungen erklärt. Außerdem können Sie eine Anleitung dazu in diesem [separaten Dokument](#) nachlesen, das auf der Lehrstuhlwebseite bereitgestellt wird. Für den Statistik-Kurs ist eine Installation aber wegen der Nutzung von *R Studio Cloud* nicht unbedingt notwendig!

Falls Sie mehr R lernen möchten und mehr Lernmaterialien brauchen, als wir im Rahmen dieses Statistikurses bereitstellen, bieten sich viele exzellente Online-Ressourcen an. Eine Auswahl:

- Freier Kurs auf [Coursera](#), angeboten von der Johns Hopkins University;
- Das Buch „R Programming for Data Science“ von einem der Autoren des Coursera-Kurses;
- Viele über das [swirl project](#) angebotene Tutorials zu verschiedenen R-Themen;
- Der Kurs „[Topics in R Statistical Language](#)“, angeboten von der PennState University;
- Der Kurs „[Learning R](#)“, angeboten von der NYU.

Das Wichtigste, um die Datenanalyse mit R (oder jedem anderen Software-Paket) zu lernen, ist: Üben, üben, üben!!! Haben Sie keine Angst vor Fehlermeldungen! Googeln Sie nach Lösungen (es hatten immer schon andere das gleiche Problem)! Probieren Sie sich an den Aufgaben der Übungsblätter ohne direkt in die Lösung zu schauen! Programmieren Sie Beispiele (ggf. leicht modifiziert) nach!