

FAU – Fachbereich Wirtschafts- und Sozialwissenschaften  
Lehrstuhl für Statistik und Ökonometrie  
**Data Science: Datenauswertung**  
Wintersemester 2020/21 – Syllabus

**Bitte beachten:** Wegen der Corona-Pandemie findet dieses Modul im Wintersemester 2020/21 teilweise im Onlineformat statt. Für aktuelle Informationen bitte die Kommunikation über StudOn beachten!

Das Modul läuft über die erste Hälfte des Wintersemesters. Genauere Informationen zur Kursorganisation finden Sie auf StudOn.

**Vorlesung:** Aufzeichnungen der Vorlesung stehen in Form von kurzen Videoclips auf StudOn zur Verfügung.

**Übung:** Findet alle zwei Wochen im Präsenzformat statt. Außerdem stehen Videos mit Erklärungen zu den Lösungswegen auf StudOn zur Verfügung. Mehr Informationen zur Organisation finden Sie auf StudOn. Bitte beachten Sie, dass Sie nur die Übungsgruppe besuchen können, für die Sie sich angemeldet haben. Die Anmeldung erfolgt über StudOn und ist ab dem 19. Oktober 2020 um 9:00 möglich.

Übungsgruppe 1: Montag, 8:00 – 9:30, LG H4, Start in erster Woche

Übungsgruppe 2: Montag, 8:00 – 9:30, LG H4, Start in zweiter Woche

Übungsgruppe 3: Montag, 13:15 – 14:45, LG H4, Start in erster Woche

Übungsgruppe 4: Montag, 13:15 – 14:45, LG H4, Start in zweiter Woche

Übungsgruppe 5: Donnerstag, 8:00 – 9:30, LG H4, Start in erster Woche

Übungsgruppe 6: Donnerstag, 8:00 – 9:30, LG H4, Start in zweiter Woche

**Tutorium:** Bitte beachten Sie, dass Sie nur den Tutoriumstermin besuchen können, für den Sie sich angemeldet haben. Die Anmeldung erfolgt über StudOn und ist ab dem 19. Oktober 2020 um 9:00 möglich. Dort finden Sie auch die genauen Veranstaltungszeiten. Die Tutorien starten in der ersten Vorlesungswoche.

**R-Übung:** Findet als Onlineveranstaltung über Zoom statt. Zugangsdaten erfahren Sie über StudOn.

Dienstag, 11:30 – 13:00. Die R-Übung startet in der ersten Vorlesungswoche.

Es gibt darüber hinaus noch drei von studentischen Hilfskräften betreute (Online-)Sprechstunden zu Fragen rund um R. Termine und Zugangsdaten finden sie auf StudOn.

**Dozent:**

Jonas Dovern

Raum: LG 4.169

[jonas.dovern@fau.de](mailto:jonas.dovern@fau.de)

Sprechstunde: n. V.

**Koordinator Übungen/Tutorien**

Alexander Glas

Raum: LG 4.173

[alexander.glas@fau.de](mailto:alexander.glas@fau.de)

Online-Sprechstunde: Donnerstag, 13:00 – 14:00

**Koordinator R-Übung**

Alexander Glas

Raum: LG 4.173

[alexander.glas@fau.de](mailto:alexander.glas@fau.de)

**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. 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 im Bereich Data Science dazu gebraucht werden, um i) Daten zu beschreiben, und v) große Datensätze mittels Maschinellen Lernens zu analysieren.

Da mehr und mehr Berufsbilder in Wirtschaft und Forschung das praktische Arbeiten mit Datensätzen beinhalten, umfasst der Kurs auch eine Einführung in die praktische Umsetzung der Datenauswertung mittels der Statistiksoftware R.

Studierende, die den Kurs erfolgreich abschließen, können i) Datensätze mittels geeigneter Maßzahlen beschreiben, ii) geeignete Formen zur Visualisierung von Daten auswählen, iii) die Funktionsweise ausgewählter Verfahren des Maschinellen Lernens erklären und iv) die behandelten Methoden praktisch in der Statistiksoftware R implementieren.

**Prüfung:** Das Modul wird mit einer Multiple-Choice-Klausur (60 Minuten) in der Mitte des Semesters geprüft. Dabei beinhaltet die Klausur auch Fragen zur praktischen Umsetzung von statistischen Analysen in R. Termine zum Nachschreiben der Klausur werden jeweils am Ende des Winter- und des Sommersemesters angeboten.

**Klausurtermin:** Wird später vom Prüfungsamt bekanntgegeben.

**Kommunikation im Kurs:** Wir kommunizieren alles Wichtige zum Kurs über die StudOn-Seite des Kurses. Bitte nutzen Sie für Fragen die jeweiligen Foren auf StudOn (zu organisatorischen Fragen, zur Übung, zu den Tutorien, zu R) und sehen Sie – außer bei persönlichen Fragen – von Anfragen per E-Mail ab.

## Themen des Kurses

### Deskriptive 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?

### 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?

## 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.]

**Aufbau des Kurses:** Die Veranstaltung bietet Ihnen vielfältige Möglichkeiten die Grundlagen der Datenauswertung zu lernen. Wöchentlich können Sie zum Lernen die Vorlesungsaufzeichnungen zu 2 Themenbereichen, 1 Übung, 1 Tutorium, und 1 R-Übung (online) nutzen. 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 in die Lösung zu schauen) sowie 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 Veranstaltungen gibt es auf StudOn zu jedem Themenkomplex ein Online-Quiz, das Sie dazu nutzen können Ihren Lernfortschritt zu überprüfen. Machen Sie diese Quizze kontinuierlich während des Semesters, um von diesem Angebot maximal zu profitieren.

## Informationen zur R-Übung

In der R-Übung wird gezeigt, wie mit Hilfe des Statistikprogramms R der aus Vorlesung und Übung bekannte Stoff auf reale Datensätze angewendet werden kann. Zusätzlich zu den in den R-Übungen besprochenen Aufgaben gibt es Übungsblätter, die der eigenständigen Kontrolle Ihres Lernfortschritts dienen. Bei Problemen mit diesen Aufgaben können Sie Hilfe im Rahmen der R-Sprechstunden bekommen, die online von studentischen Hilfskräften angeboten werden. 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.

## 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“).

Um die Grundlagen des Programms zu lernen, müssen Sie in den ersten beiden Semesterwochen den von uns entworfenen [Onlinekurs „Basiskurs R/RStudio“](#) auf StudOn bearbeiten. Dort wird auch erklärt, wie Sie R und R-Studio auf Ihrem Computer installieren können.

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!