

FAU – Fachbereich Wirtschafts- und Sozialwissenschaften
Lehrstuhl für Statistik und Ökonometrie
Data Science: Datenauswertung
Wintersemester 2023/24 – Syllabus

Bitte beachten: Das Modul läuft über die erste Hälfte des Wintersemesters. Aufgrund des Planspiels beginnen Veranstaltungen dieses Moduls frühestens erst in der zweiten Semesterwoche. Genauere Informationen zur Kursorganisation finden Sie auf StudOn. Für aktuelle Informationen bitte die Kommunikation über StudOn beachten!

Vorlesung: Dienstag, 11:30 – 13:00, LG H4 (erster Termin am 24.10.)
Donnerstag, 11:30 – 13:00, LG H4 (erster Termin am 26.10.)

Übung: 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 9. Oktober 2023 um 12:00 möglich. Die Übungen starten wie die Vorlesung in der zweiten Semesterwoche.

Übungsgruppe 1: Dienstag, 16:45 – 18:15, LG H6

Übungsgruppe 2: Mittwoch, 8:00 – 9:30, LG H2

Übungsgruppe 3: Donnerstag, 8:00 – 9:30, LG H1

Übungsgruppe 4: Donnerstag, 16:45 – 18:15, LG H5 (30.11. – LG H4)

R-Übung: Mittwoch, 13:15 – 14:45, LG H4

Die R-Übung startet erst in der zweiten Semesterwoche. Um die Grundlagen des Programms zu lernen, müssen Sie in den ersten drei Semesterwochen den von uns entworfenen [Onlinekurs „Basiskurs R/RStudio“](#) bearbeiten und das Zertifikat im StudOn-Kurs für dieses Modul einreichen.

Tutorium: Lösungen für die Arbeitsblätter des Tutoriums bieten wir dieses Jahr in Form von Videos an (Zugriff via StudOn). Es finden keine Tutorien in Präsenz statt. Sie können (und sollten!!!) bei Fragen aber die Sprechstunden mit studentischen Hilfskräften in Anspruch nehmen (nächster Punkt).

Sprechstunde: Eine Anmeldung ist nicht notwendig. Besuchen Sie die Sprechstunden, wenn Sie nach dem Schauen der Lösungsvideos noch konkrete Fragen zu Aufgabenblättern des Tutoriums haben. Die Sprechstunden starten in der zweiten Semesterwoche.

Sprechstunde 1: Mittwoch, 11:30 – 13:00, LG 0.424

Sprechstunde 2: Donnerstag, 16:45 – 18:15, LG 0.142

Sprechstunde 3: Donnerstag, 18:30 – 20:00, LG 0.144

R-Tutorium: Montag, 15:00 – 16:30, LG 0.215 (für einzelne Raumänderungen beachten Sie bitte Informationen auf Campo)

Startet erst in der dritten Semesterwoche (ab dem 30.10.). Eine Anmeldung ist nicht notwendig; es gibt allerdings nur begrenzte Kapazitäten. Ein Besuch lohnt sich nur mit konkreten Fragen zu den R-Aufgabenblättern.

Dozent:

Jonas Dovern
 Raum: LG 4.169
jonas.dovern@fau.de
 Sprechstunde: n.V.

Koordinator Übungen/Tutorien

Johannes Frank
 Raum: LG 4.174
johannes.jf.frank@fau.de
 Sprechstunde: n.V.

Koordinator R-Übung

Johannes Frank
 Raum: LG 4.174
johannes.jf.frank@fau.de
 Sprechstunde: n.V.

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 6 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 mit Methoden des 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 Umsetzung der Datenauswertung mit Hilfe 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 Single-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. (Bitte beachten Sie, dass Studierende des ersten Fachsemesters des BA Wirtschaftswissenschaften über eine sogenannte Serviceanmeldung automatisch für den ersten Klausurtermin registriert werden. Diese Studierenden müssen sich aktiv abmelden, wenn Sie die Klausur an diesem Termin nicht schreiben möchten! Alle anderen Studierenden müssen sich aktiv für den Klausurtermin anmelden und dabei die frühe Deadline beachten.)

Klausurtermine: Voraussichtlich am 9. Dezember. Weitere Klausurtermine werden jeweils 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 und sehen Sie – außer bei persönlichen Fragen – von Anfragen per E-Mail ab.

Themen des KursesDeskriptive 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 in der FAU-Bibliothek verfügbar.]

Aufbau des Kurses: Die Veranstaltung bietet Ihnen vielfältige Möglichkeiten die Grundlagen der Datenauswertung zu lernen. Wöchentlich sollten Sie die beiden Vorlesungen, eine Übung und die R-Übung besuchen. Darüber hinaus können Sie in einem Tutorium, das wir in Form von Videos mit Musterlösungen anbieten, die Anwendung der Kursinhalte vertiefen. In der Sprechstunde und im R-Tutorium können Sie von studentischen Hilfskräften Unterstützung bei konkreten Fragen zu den Tutoriumsaufgaben bzw. der praktischen Implementierung in R erhalten.

Um einen optimalen Lernerfolg zu erzielen, empfehlen wir Ihnen i) die relevante Literatur vor den Vorlesungen anzuschauen, ii) die Übungsaufgaben eigenständig vor den Übungen zu bearbeiten und, falls Sie Probleme mit einer Aufgabe hatten, *nach* den Übungen 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 am Ende jeder Woche ein Online-Quiz, das Sie dazu nutzen können Ihren Lernfortschritt zu überprüfen. Die Quizze sind jeweils nur für 2 Wochen zugänglich, um einen Anreiz für kontinuierliches Lernen zu setzen.

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 des R-Tutoriums bekommen, das von studentischen Hilfskräften angeboten wird. Dort können auch allgemeine Fragen zu R geklärt werden.

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“](#) bearbeiten und das Kurszertifikat dann im StudOn-Kurs für dieses Modul als Nachweis hochladen. Im R-Kurs 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!

Das Investment zahlt sich im Übrigen für Sie bereits schnell aus, weil auch nachfolgende Module (Data Science – Statistik, Data Science – Datenmanagement und -analyse, Data Science – Ökonometrie und viele Wahlmodule) auf R als Software zurückgreifen.