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

Bitte beachten: Das Modul läuft über die zweite Hälfte des Wintersemesters (ab dem 11. Dezember). Genauere Informationen zur Kursorganisation finden Sie ab Mitte November auf StudOn. Für aktuelle Informationen bitte die Kommunikation über StudOn beachten!

Vorlesung: Dienstag, 11:30 – 13:00, LG H4 (erste Vorlesung am 12.12.)
Donnerstag, 11:30 – 13:00, LG H4 (erste Vorlesung am 14.12.)

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 4.12. um 12:00 möglich. Die Übungen starten ab dem 18. Dezember.
Übungsgruppe 1: Dienstag, 16:45 – 18:15, LG H6
Übungsgruppe 2: Donnerstag, 8:00 – 9:30, LG H1
Übungsgruppe 3: Donnerstag, 15:00 – 16:30, LG H3

R-Übung: Mittwoch, 13:15 – 14:45, LG H4 (erste R-Übung am 20. Dezember)

Tutorium: Bitte beachten Sie, dass Sie nur den Tutoriumstermin besuchen können, für den Sie sich angemeldet haben. Die Anmeldung erfolgt über StudOn/Campo und ist ab dem 4.12. um 12:00 möglich. Auf StudOn/Campo finden Sie auch die genauen Veranstaltungszeiten. Die Tutorien starten ab dem 18.12.. (Im Gegensatz zum Modul „Data Science: Datenauswertung“ finden die Tutorien in Präsenz statt und es gibt keine Lösungsvideos.)

R-Tutorium: Montag, 15:00 – 16:30, LG 0.215 (startet am 11.12.)
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. 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?

Dieser Kurs gibt eine Einführung in statistische Methoden, die im Bereich Data Science dazu gebraucht werden, um i) Zufallsvorgänge zu analysieren, ii) Datenerhebungen zu planen, iii) von Stichproben auf ganze Populationen zu schließen und iv) statistische Hypothesentests durchzuführen.

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 Methoden mittels der Statistiksoftware R.

Studierende, die den Kurs erfolgreich abschließen, können i) grundlegende Stichprobenverfahren erläutern, ii) Parameter von Verteilungen schätzen, iii) verschiedene Arten von Hypothesentests anwenden, iv) induktive Methoden als Grundlage des Arbeitens in empirischen Wissenschaften einsetzen und statistische Ergebnisse kritisch überprüfen und v) die behandelten Methoden praktisch in der Statistiksoftware R implementieren.

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

Klausurtermin: Wird später vom Prüfungsamt bekanntgegeben. Klausur am Ende des Wintersemesters. Nachholtermin am Ende des Sommersemesters.

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 Kurses

Stichprobenverfahren:

Wie erhebt man eigentlich eine Datenstichprobe?

Wahrscheinlichkeitsrechnung:

Was ist eine Wahrscheinlichkeit?

Wie kann man bedingte Wahrscheinlichkeiten berechnen?

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?

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.]
- 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 Wahrscheinlichkeitsrechnung und schließenden Statistik zu lernen. Wöchentlich sollten Sie die beiden Vorlesungen, eine Übung und die R-Übung besuchen. Darüber hinaus können Sie in einem Tutorium die Anwendung der Kursinhalte vertiefen und in den R-Sprechstunden Unterstützung bei 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 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 am Ende jeder Woche ein Online-Quiz, das Sie dazu nutzen können Ihren Lernfortschritt zu überprüfen. Die Quizze sind jeweils für 2 Wochen zugänglich.

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, und es wird Hilfe bei der Vor- und Nachbereitung der Aufgaben aus der R-Übung angeboten.

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“).

Wir gehen davon aus, dass Sie mit den Grundlagen des Programms nach der Teilnahme am Modul „Data Science: Datenauswertung“ vertraut sind. Sollte dies nicht der Fall sein, bearbeiten Sie bitte vor Modulbeginn den von uns entworfenen [Onlinekurs „Basiskurs R/RStudio“](#) auf StudOn! 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!