Angewandte Datenanalyse

Applied Statistical Learning

Modulnummer	Workload	Credits	Häufigkeit des Angebots	Dauer
32491	300 h	10	jedes Semester	1 Semester

1 Lehrveranstaltungen

Einheit	Titel	Workload
1	Angewandte Datenanalyse	300 h

2 Lernergebnisse (learning outcomes) / Kompetenzen

Mit diesem Modul werden vier Qualifikationsziele verfolgt:

1. Die Studierenden entwickeln umfassende Kenntnisse moderner Verfahren des statistischen Lernens und beherrschen deren Anwendung für die Modellierung und Prognose relevanter Variablen aus verschiedenen Bereichen der Wirtschaftswissenschaften.
2. Die Studierenden erlernen, wie sie ökonomische Fragestellungen sachgerecht in ein Modell überführen, die geeigneten Daten auszuwählen und die statistischen Schätz- sowie Testverfahren hinsichtlich ihrer Angemessenheit für die jeweilige Fragestellung beurteilen. Sie lernen zudem empirische Befunde kritisch zu reflektieren.
3. Die Studierenden erlernen eigenständig mit Hilfe der open-Source Software R praktische Probleme zu lösen und die Analysen, sowie Ergebnisse beispielsweise mit Abbildungen und Tabellen adressatengerecht aufzubereiten.
4. Den Studierenden wird ein allgemeines Grundverständnis für die praktische Anwendung (auch als Vorbereitung auf eigene empirische Untersuchungen im Rahmen von Seminar- und Abschlussarbeiten) vermittelt.

3 Inhalte

Das Modul basiert auf dem etablierten Lehrtext „An Introduction to Statistical Learning: with Applications in R“ (Second Edition, Springer Texts in Statistics) von Gareth James, Daniela Witten, Trevor Hastie und Robert Tibshirani. Dieser ist für Studierende der FernUniversität als E-Book frei verfügbar. Zusätzlich werden auf Basis von Vorlesungsfolien Lehrvideos angeboten. In synchronen virtuellen Übungsveranstaltungen werden die Inhalte vertieft und anhand realer Datensätze erprobt. Es werden zu den verschiedenen Kapiteln R-Codes zur Verfügung gestellt, in denen Methoden anhand von Beispieldaten in der open-source Software R angewendet und verglichen werden.

Die Darstellung der Methoden ist so formal wie nötig. Im Vordergrund stehen die empirische Modellierung und das intuitive Verständnis. Die Methoden werden durchweg anhand wirtschaftswissenschaftlicher Fragestellungen illustriert. Größtenteils werden Methoden des überwachten Lernens behandelt, bei denen verschiedenartige Zielvariablen durch andere erklärende Variablen modelliert und prognostiziert werden. Abschließend werden Methoden des unüberwachten Lernens betrachtet, bei denen es um Dimensionsreduktion und die Aufdeckung von Clustern geht.

Kapitel 1 (Kapitel 2 im Lehrtext) umfasst eine Einführung in das „Statistische Lernen“, um stetige Variablen, wie z. B. den Umsatz eines Unternehmens, mit Hilfe von Regressionsmodellen und qualitativen Variablen, wie etwa die Kaufentscheidung für ein bestimmtes Produkt, mit Klassifikationsmodellen zu prognostizieren. Darüber hinaus wird erklärt, wie die Prognosen evaluiert werden können und wichtige Konzepte wie etwa der Bias-Varianz-Kompromiss vermittelt. Verschiedene Modelle werden zudem hinsichtlich ihrer Prognosefähigkeit und Interpretierbarkeit verglichen.

Kapitel 2 (Kapitel 3 im Lehrtext) widmet sich linearen Regressionsmodellen. Neben der Schätzung der Modelle mit Hilfe der Kleinst-Quadrate-Methode und der Prognose der abhängigen Variablen wird erläutert, wie ökonomische Hypothesen empirisch überprüft werden können.

Kapitel 3 (Kapitel 4 im Lehrtext) beschäftigt sich mit Klassifikationsproblemen und erläutert Modelle für Prognose der Wahrscheinlichkeit, mit der eine qualitative Variable einen bestimmten Wert annimmt. Ein großer Fokus dieses Kapitels liegt auf der logistischen Regression für binäre Variablen

	und der Maximum-Likelihood-Methode zur Schätzung der Parameter. Es werden zudem die multinomiale/geordnete logistische Regression, die Diskriminanzanalyse und weitere Themen behandelt, welche für Anwendungsfälle mit mehr als zwei Klassen geeignet sind. Darüber hinaus wird auf Besonderheiten der Prognoseevaluation im Rahmen von Klassifikationsproblemen eingegangen. Kapitel 4 (Kapitel 5 im Lehrtext) widmet sich der Kreuzvalidierung im Kontext der Unterteilung von Datensätzen in Trainings- und Testdaten. Dies ist besonders relevant, um eine Einschätzung zur Prognoseleistung zu erhalten. Es werden unterschiedliche Kreuzvalidierungsmethoden für Regressionsmodelle und Klassifikationsprobleme vorgestellt. Kapitel 5 (Kapitel 6 im Lehrtext) beschäftigt sich mit der Modellauswahl im Rahmen der linearen Regression. Als Methoden zur Auswahl der erklärenden Variablen werden die Subset-Selection (Teilmengenauswahl), Informationskriterien, welche auf einem Kompromiss zwischen Reduktion des Schätzfehlers und Modellkomplexität basieren, und Regularisierungsverfahren wie die Ridge- oder Lasso-Regression behandelt. Neben den Methoden des Überwachten Lernens, behandelt Kapitel 6 (Kapitel 12 im Lehrtext) Methoden des Unüberwachten Lernens. Es werden hauptsächlich zwei führende Methoden behandelt: Die Hauptkomponentenanalyse zur Dimensionsreduktion und die Clusteranalyse zur Aufdeckung von Strukturen in multivariaten Datensätzen.
4	**Lehrformen** Fernstudium mit Betreuung, zeitlich und räumlich flexibel, mit folgenden Elementen: – Lehrvideos (asynchron) mit didaktisch aufbereiteten Vorlesungsfolien – Lehrtext – Moodle-Lernumgebung mit zusätzlichen Vorlesungs- und Übungselementen – Freiwillige Online-Übungen in Zoom (synchron)
5	**Teilnahmevoraussetzungen** Formal: Gemäß Prüfungsordnung des jeweiligen Studienganges Inhaltlich: Vorkenntnisse in Mathematik und Statistik sind erforderlich
6	**Prüfungsformen** Zweistündige Abschlussklausur
7	**Voraussetzungen für die Vergabe von Kreditpunkten** Die Leistungspunkte werden vergeben, wenn die Prüfungsklausur bestanden worden ist. Voraussetzung für die Teilnahme an der Prüfungsklausur ist das Bestehen mindestens einer von zwei Einsendearbeiten.
8	**Verwendung des Moduls** Masterstudiengang Wirtschaftswissenschaft Masterstudiengang Volkswirtschaft Masterstudiengang Wirtschaftsinformatik Masterstudiengang Wirtschaftswissenschaft für Ingenieur/-innen und Naturwissenschaftler/-innen Masterstudiengang Wirtschaftspsychologie Akademiestudium
9	**Stellenwert der Note für die Endnote** Gemäß Prüfungsordnung des jeweiligen Studienganges
10	**Modulbeauftragte/r und hauptamtlich Lehrende** Univ.-Prof. Dr. Robinson Kruse-Becher

11	**Sonstige Informationen** Der zugrundliegende Lehrtext „An Introduction to Statistical Learning: with Applications in R“ (Second Edition, Springer Texts in Statistics) von Gareth James, Daniela Witten, Trevor Hastie und Robert Tibshirani ist für Studierende der FernUniversität als E-Book frei verfügbar. Die Lehrvideos werden in Moodle bereitgestellt. Dort befinden sich ebenso die Datensätze und die R-Codes.