LiMeLDaS: Linguistische Methoden: Linguistic Data Science

Prof. Dr. Eva Belke (Fakultät für Philologie, Sprachwissenschaftliches Institut)

Was zeichnet das Lehrmuster aus?

Das Modul vermittelt Studierenden sprachwissenschaftlicher Fächer studiengangsübergreifend Grundlagen der Statistik sowie Fertigkeiten um Umgang mit der Analyseumgebung R, in der mit den Verfahren gerechnet werden kann. Das Modul exemplifiziert den Nutzen verschiedener Verfahren an Forschungsthemen der antragstellenden Lehrenden, die diese als Beispielsitzungen in die Kurse des Moduls einspeisen.

Fakten im Überblick:

In welcher Form existiert eine Präsenzphase?
Einzeltermine, Sonstiges (Modul im Inverted Classroom Setting (selbstständige Erarbeitung der Inhalte durch Online-Lektionen, Lehrvideos und Übungsaufgaben, wöchentliche Übungssitzung))

In welchen Zeitraum wird das Lehrmuster durchgeführt?
Während Vorlesungszeit

Wird das Lehrmuster über einen Zeitraum von mehreren Semester durchgeführt?
Ja (2 Semester)

Welchen Umfang hat das Lehrmuster?
Creditpoints: 5
Teilnehmerzahl: 60

Das Modul wird als Grundlagenmodul im neu eingerichteten Studienfach VAMoS im 2-Fach BA eingespeist und darüber hinaus im Master im Rahmen eines Lehrauftrags für Studierende anderer sprachwissenschaftlicher Fächer angeboten.
Credit Points: variierend je nach Fach (3 – 5 CP pro Kurs/Semester)

In welchem Studienabschnitt ist das Lehrmuster angesiedelt?
Master

In welcher Art ist das Lehrmuster curricular verankert?
Pflichtmodul, Wahlmodul

Worum geht es in dem Lehrmuster insbesondere?
Eine Verbesserung der Studieneingangsphase, Sonstiges (Andere Fächer: Eine Ausbildung in Grundlagen der Statistik und in R, basierend auf Beispielen aus der Linguistik, gerade auch für Philolog:innen, von denen zunehmend datenbasiertes/- getriebenes Arbeiten erwartet wird.)

Welche Zielsetzung hat das Lehrmuster?

Mit dem empirical turn in der Sprachwissenschaft gewinnen statistische Verfahren besonders im Kontext des forschenden Lernens in allen Studiengängen an Bedeutung, können aber insbesondere in den philologischen Fächern bislang nur sehr punktuell vermittelt werden. Aufgrund der besonderen Eigenschaften von Sprachdaten (z.B. Zipfsches Gesetz) kann die Ausbildung in statistischen Verfahren nicht ohne Weiteres aus anderen Fächern (z. B. Soziologie, Psychologie) übernommen werden, sondern sollte auf sprachwissenschaftliche Fragestellungen und Konstrukte zugeschnitten sein. Das

Modul wird diesen Anforderungen gerecht, indem es
(1) eine Ausbildung in statistischen Grundlagen speziell für sprachwissenschaftliche Fragestellungen gewährleistet,
(2) die Teilnehmer_innen schrittweise an die Arbeit mit R heranführt und veranstaltungsbegleitend den Einstieg in die Programmierung von Skripten zur Datenanalyse in der Programmierumgebung RStudio ermöglicht,
(3) auf dem Prinzip des Inverted Classroom (IC) als aktuellem didaktischem Ansatz basiert, der u.a. Gelegenheit zum Einüben der erworbenen Verfahren in erheblichem Umfang bietet.

Indem die Studierenden vielfältige Praxisbeispiele aus verschiedenen

Fachbereichen bearbeiten, erleben sie statistische Verfahren als methodische Schnittstelle der Fächer, die sie studieren.

Was sind wesentliche Inhalte des Lehrmusters?

Das im Projekt entwickelte Modul umfasst zwei Lehrveranstaltungen à 2 SWS. In der ersten werden Grundlagen der deskriptiven Statistik und der Inferenzstatistik anhand von Unterschiedstests (z.B. t-Tests) eingeführt und angewendet. In der zweiten werden die erlernten Grundlagen auf varianzanalytische und regressionsbasierte Verfahren erweitert und ihre Verwendung geübt. Darüber hinaus wird die Berücksichtigung zufälliger Effekte der Versuchspersonen und der sprachlichen Einheiten bzw. Stimuli bei sprachwissenschaftlichen Fragestellungen thematisiert und ihre Berücksichtigung in gemischten Regressionsmodellen vermittelt. Sämtliche Lektionen basieren auf Beispieldatensätzen zu linguistischen bzw. psycholinguistischen Fragestellungen. Dazu gehören insbesondere Sitzungen, in denen die fachlich breit aufgestellte Runde der Antragsteller/innen Anwendungen quantitativer Methoden in den beteiligten Fächern aufzeigt.

Wie ist das Lehrmuster strukturiert?

Das Modul ist im Inverted Classroom Setting implementiert: Grundlagen zu den pro Sitzung thematisierten statistischen Verfahren werden auf Basis von YouTube-Videos erarbeitet und anhand von Übungsaufgaben in Moodle, die u. a. unter Hinzuziehung von R bearbeitet werden müssen, eingeübt. Beide Teile der Vorbereitung werden in Präsenzsitzungen gemeinsam besprochen, Fragen und Schwierigkeiten in der anwe4ndung in R thematisiert. Die spezifischen Inhalte sind:

- Statistische Grundbegriffe und Installation von R/RStudio
- Deskriptive Statistik (Skalenniveaus, Häufigkeiten, Kennwerte der zentralen Tendenz und Dispersionsmaße)
- Population/Grundgesamtheit vs. Stichprobe
- Darstellung deskriptiver Daten und Plots, Datenorganisation
- z-Transformation und (Standard-)Normalverteilung, Dichte- und Verteilungsfunktion
- Einführung in die Inferenzstatistik: Null- und Alternativhypothese, alpha-

und beta-Fehler, Power und Stichprobengröße
- Varianten des t-Tests
- Nicht-parametrische Tests (Chi-Quadrat, Mann-Whitney-U und Wilcoxon Tests)
- Chi2-Test
- Varianzanalyse (Haupteffekte und Interaktionen)
- Einfache und multiple lineare Regression (Haupteffekte und Interaktionen)
- Gemischte lineare Regression
- Logistische Regression
- Gemischte logistische Regression
- Gemischte ordinale Regression

Die folgenden Gastsitzungen wurden erarbeitet:

- Prof. Dr. Eric Fuß (Historische Linguistik) zum Plotten in R: Wortstellungsvariation in frühmittelalterlichen (Übersetzungs-)Texten
- Prof. Dr. Tanja Anstatt (Slavistische Linguistik) zur z-Transformation: Subjektive Frequenz und russischer Verbalaspekt
- Prof. Dr. Lena Heine (Sprachbildung und Mehrsprachigkeit) zu Korrelationen: Einflüsse auf das Passivverstehen bei Schüler:innen der Sek. I
- Prof. Dr. Eva Belke (Psycholinguistik) zur Varianzanalyse und gemischter linearer Regression: Erwerbsalterseffekte auf die Wortfindung beim Benennen
- Prof. Dr. Stefanie Dipper (Computerlinguistik) zur logistischen Regression: Relativsatz-Extraposition
- Prof. Dr. Tatjana Scheffler (Digitale forensische Linguistik) zur gemischten logistischen Regression: Pronominale it-clefts im Englischen
- Prof. Dr. Agata Renans (Experimentelle Linguistik) zur gemischten ordinalen Regression: Verarbeitung von Pluralmarkierungen bei zählbaren und nicht zählbaren Nomen im Griechischen

Welches Prüfungsform ist in dem Lehrmuster vorgesehen?

Klausur (erste Implementierung im Wintersemester 20/21 und Sommersemester 21);
Mündliche Prüfung (zweite Implementierung im Sommersemester 22 und Wintersemester 22/23). Die mündliche Prüfung hat den signifikanten Vorteil, dass Fertigkeiten im Umgang mit den Verfahren in R und die darauf bezogenen

Lernziele besser individuell überprüft werden können.

Welche E-Learning-Elemente werden eingesetzt?

Das Modul ist im Inverted Classroom Setting implementiert.

Tipps für die Umsetzung:

"Das implementierte Verfahren (Lehrvideos mit Moodle-Übungen zum Einüben der vorgestellten Verfahren) hat sich insgesamt als sehr erfolgreich für die Vermittlung der Lehrinhalte erwiesen; die Lernziele wurden erreicht und die Studierenden fanden das Modul insgesamt trotz des schwierigen Themas ansprechend." Prof. Dr. Eva Belke

Veröffentlichungen zum Lehrmuster:

Das Projekt wurde in zwei Vorträgen vorgestellt; dabei wurde es mit einem inhaltlich ähnlich gelagerten Projekt zu computerlinguistischen Verfahren in den den Digital Humanities inhaltlich verzahnt, das ein Jahr nach LIMELDAS implementiert wurde:

- 09/2021:
  Vortrag (online, E. Belke & S. Heimgartner) bei der Tagung „Zukunftskompetenzen und ihre Strukturen in der Hochschullehre" der Netzwerke Curriculum 4.0.nrw und Data Literacy Education.nrw: Data Literacy in der Philologie: Erfahrungen aus der Implementation interdisziplinärer Module
- 07/2022:
  Poster (online, E. Belke, koordiniert mit einem Partnerposter von Stephanie Heimgartner, beide Fakultät für Philologie) beim 4. DH-Day der RUB: Statistische Grundlagen für Linguist:innen aller Fächer.
- 06/2023 (Tag der Philologie; geplanter Beitrag):
  LiMeLDaS: Gemeinsame Lehre über Fächergrenzen hinweg

Konzipierung:

Kontaktperson: Prof. Dr. Eva Belke (eva.belke@rub.de) , Fakultät für Philologie, Sprachwissenschaftliches Institut
Internetseite zum Lehrmuster:
https://linguistics.rub.de/institut/personen/belke/teaching/forschendes-lernen/

Weitere Informationen:

Veröffentlichungsdatum: 06.06.2023, 11:28 Uhr

Schlagwörter: Data Literacy, Data Science, Programmiersprache, Sprache, Statistik

Fächergruppen: Geisteswissenschaften, Interdisziplinär

Das Lehrmuster ist online abrufbar unter: https://lehrmuster.ruhr-uni-bochum.de/?p=1183

Die PDF-Datei wurde generiert am: 16.11.2026, 17:08 Uhr