



**RUHR-UNIVERSITÄT BOCHUM**

**STATISTISCHE GRUNDLAGEN FÜR LINGUIST:INNEN ALLER FÄCHER**

Eva Belke, Sprachwissenschaftliches Institut

Vorstellung eines Moduls, das im Rahmen der Förderlinie Forschendes Lernen:  
Schwerpunkt Data Science entwickelt wurde.

## Statistik für Linguist:innen aller Fächer

### Fachübergreifende Koordination der Statistik-Ausbildung

Vermittlung von Kernkompetenzen im Umgang mit quantitativen Daten

- Grundlagenkenntnisse und Umgang mit R
- im Inverted Classroom Format

Gezielter Zuschnitt auf sprachwissenschaftliche Fragestellungen und Konstrukte

Linguist:innen brauchen heutzutage fast immer statistische Kenntnisse und Fertigkeiten für ihre Arbeit!

Gezielter Zuschnitt auf sprachwissenschaftliche Fragestellungen und Konstrukte:  
Viele statistische Verfahren gehen von normalverteilten Daten aus, aber Sprache ist eine Domäne, in denen eine der augenfälligsten Variablen, die Frequenz sprachlicher Zeichen, nicht normalverteilt ist.

## Besondere Eigenschaften von Sprachdaten



erstellt mit: <https://www.wortwolken.com/>

Bei der Arbeit mit Sprachkorpora muss man bspw. zentrale Eigenschaften ihrer Zusammensetzung im Blick haben: So sind z. B. nur wenige Wörter extrem häufig und für viele Untersuchungen uninteressant – das sehen wir an dieser Wortwolke unseres Förderantrags für die Entwicklung des Moduls. Inhaltlich einschlägigere Wörter kommen erst zum Vorschein, wenn man die Funktionswörter (der, die, den, usw.) ausschließt. Welche Daten wie aufbereitet werden, muss fachlich begründbar sein.

## Ziele des Projekts

### Ausbildung in statistischen Grundlagen speziell für sprachwissenschaftliche Fragestellungen

- In Datenstrukturen denken lernen
- schrittweise Heranführung an die Arbeit mit R

### Statistische Methoden für alle linguistisch Forschenden und Studierenden der Fakultät – aber nicht nur!

Der Zuschnitt auf sprachwiss. Fragen hat nicht nur inhaltliche Gründe, sondern soll auch motivieren!

Inverted Classroom Setting optimal geeignet zur Vermittlung von Kenntnissen und Fertigkeiten, die zentrale Ziele des Moduls darstellen:

[...]

## ... und jenseits der Linguistik?

**Grundlagenkenntnisse in Statistik sind nicht nur für Linguist:innen essentiell, um quantitative Daten zu analysieren und zu bewerten**

- Wann handelt es sich um statistisch relevante Unterschiede/Zusammenhänge?
- Wie müssen Untersuchungen gestaltet sein, damit sowohl statistisch relevante als auch statistisch nicht relevante Unterschiede klar erkennbar sind?
- Wie sind statistische Ergebnisse zu bewerten?

### Statistik für den Beruf

- Inhalte im Zusammenhang mit Statistik denken: Quantitative Unterschiede und Zusammenhänge sind u. U. bedeutungslos, wenn sie nicht statistisch signifikant sind – umgekehrt können Unterschiede und Zusammenhänge unter Umständen zwar statistisch signifikant, aber inhaltlich bedeutungslos sein.
- Wie müssen Untersuchungen also gestaltet sein, damit inhaltlich sinnvoll Unterschiede und Zusammenhänge sauber mit Blick auf ihre statistische Bedeutsamkeit untersucht werden können?

## ... und jenseits des Studiums?

**Alle aktuellen Absolvent:innen der RUB werden in ihrem Berufsleben direkt oder indirekt mit Daten arbeiten, z. B. ...**

... im Lehramt:

- Medien- und KI-Kompetenz
- R (und auch) Python sind Programmiersprachen, die nicht nur von ChatGPT beherrscht werden können sollten!

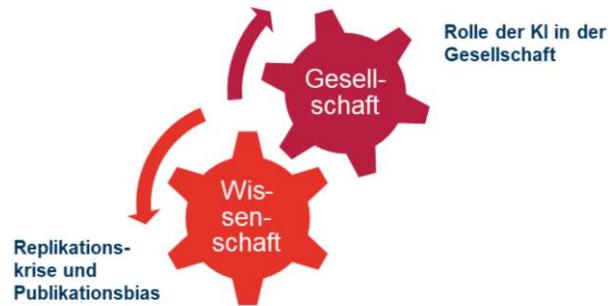
... und außerhalb der Lehramtsausbildung:

- Kenntnisse in R / Python sind (noch) selten bei Philolog:innen
- Die Kenntnis von Datenverarbeitungsverfahren und -tools ist oft Herrschaftswissen

Zu Medienkompetenz: Wie können Daten maximal informativ und transparent visualisiert werden (und wie erkennt man, wenn jemand mit Statistik/Darstellungsformen gezielt in die Irre führen will)?

## ... und jenseits beruflicher Kontexte?

Unsere aktuellen Absolvent:innen werden ihre und unsere Zukunft entscheidend mitgestalten



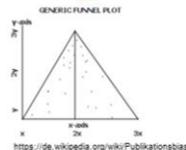
# Datenkompetenz als Ausbildungsziel

Unsere aktuellen Absolvent:innen werden ihre und unsere Zukunft entscheidend mitgestalten



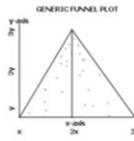
Was ist „big“ in Big Data?  
Was macht insbesondere sprachliche Daten zu so einer mächtigen Ressource für die KI?

Wie sind quantitative Ergebnisse zu bewerten?



# Datenkompetenz als Ausbildungsziel

Unsere aktuellen Absolvent:innen werden ihre und unsere Zukunft entscheidend mitgestalten



## Weitere Informationen

<https://linguistics.rub.de/institut/personen/belke/teaching/forschendes-lernen/>

[https://www.stifterverband.org/sites/default/files/data\\_literacy\\_education\\_duz\\_wissenschaft\\_und\\_management\\_09-2020.pdf](https://www.stifterverband.org/sites/default/files/data_literacy_education_duz_wissenschaft_und_management_09-2020.pdf)