

Projekt 2

Namenshäufigkeiten von ProfessorInnen

Oft werden bei wissenschaftlichen Publikationen die Autoren in der Reihenfolge der Anfangsbuchstaben ihrer Nachnamen aufgeführt. Das ist insbesondere in der Mathematik häufig. In vielen Fächern spielt aber der sogenannte Erstautor eine große Rolle. Daher stellt sich die Frage, ob Personen, die aufgrund ihres Nachnamens eher Erstautor sind, eine größere Chance haben, eine Professur zu bekommen. Das soll in diesem Projekt untersucht werden, indem die Verteilung der Anfangsbuchstaben der Nachnamen von ProfessorInnen mit der in der Gesamtbevölkerung verglichen wird. Die Verteilung der Anfangsbuchstaben in der Bevölkerung in Deutschland ist in der Datei

`buchstabenhaeufigkeit_dt_namen.xlsx`

von der Seite https://www.datendieter.de/item/Buchstabenhaeufigkeit_von_deutschen_Nachnamen enthalten. Die Datei

`Liste der Universitäten in Deutschland - Uni Übersicht
deutsche Unis.pdf`

von der Seite <http://www.fernhochschule.net/universitaeten-deutschland.php> enthält die wesentlichen deutschen Universitäten. Anhand der in dieser Datei aufgeführten Universitäten sollen die Anfangsbuchstaben der in diesen Universitäten beschäftigten ProfessorInnen nach folgenden Fächern erfasst werden:

- Ges: Gesellschaftswissenschaften inkl. Geschichtswissenschaften, Sprachwissenschaften, Philosophie
- WiWi: Wirtschaftswissenschaften
- Psy: Psychologie
- Med: Medizin inkl. Zahn- und Veterinärmedizin
- Nat: Naturwissenschaften inkl. Agrarwissenschaften
- Math: Mathematik inkl. Statistik
- Info: Informatik
- Ing: Ingenieurwissenschaften
- Sonst: Sonstige Fächer

Dabei sollen Juniorprofessoren nicht erfasst werden. Die Ergebnisse sollen in die Datei

`Daten_Vorlage.xlsx`

eingetragen werden. Diese Datei hat die in Tabelle 1 gezeigte Struktur. Insbesondere sollen die Abkürzungen für die Fächer beibehalten werden und in den Spalten A bis Z sollen die absoluten Anzahlen der Anfangsbuchstaben der ProfessorInnen-Namen in den jeweiligen Fächern der Universität eingetragen werden. Alle Dateien finden Sie im Moodle-Raum.

Universität	Fach	A	B	C	D	E	...	Q	R	S	T	U	V	X	Y	Z
	Ges															
	WiWi															
	Psy															
	Med															
	Nat															
	Math															
	Info															
	Ing															
	Sonst															

Tabelle 1: Vorlage für Datenerfassung in Exel-Tabelle `Daten_Vorlage.xlsx`

Aufgaben

- 1) Bestimmen Sie für die Ihnen zugewiesenen Universitäten die Anzahlen der Anfangsbuchstaben der ProfessorInnen gemäß Tabelle `Daten_Vorlage.xlsx`, tragen Sie diese Ergebnisse in diese Tabelle ein und laden Sie diese in Moodle hoch.
- 2) Kommen Buchstaben am Anfang des Alphabets bei den ProfessorInnen häufiger als bei der Gesamtbevölkerung vor? Gibt es diesbezüglich Unterschiede zwischen den Fächern?
- 3) Gibt es Unterschiede bei den Häufigkeitsverteilungen der Anfangsbuchstaben in der Professorenschaft und in der Gesamtbevölkerung? Gibt es Unterschiede zwischen den Fächern?
- 4) Erstellen Sie neben dem Bericht auch eine Presseerklärung.

Literatur

- 1) Christensen, R. (1997). *Log-Linear Models and Logistic Regression*. Springer, New York.
- 2) Faraway, J.J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Second Edition. Taylor & Francis, London.

- 3) Filzmoser, P., Hron, K. und Templ, M. (2018) Applied Compositional Data Analysis. With Worked Examples in R. Springer, New York.
- 4) Hastie, T., Tibshirani, R. und Friedman, J. (2002). *The Elements of Statistical Learning. Data Mining, Inference and Prediction*. Springer, New York.
- 5) James, G., Witten, D., Hastie, T. und Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer, New York.
- 6) https://www.datendieter.de/item/Buchstabenhaeufigkeit_von_deutschen_Nachnamen (23.9.2019)
- 7) <http://www.fernhochschule.net/universitaeten-deutschland.php>
(23.9.2019)

Abgabetermin

Abgabe am **Mittwoch, dem 19.2.2020**, in der Veranstaltung.