

Projekt 1 - Deskriptive Auswertung demographischer Daten

Die internationale Datenbank (*International Data Base, IDB*) des *U.S. Census Bureau* umfasst verschiedene demographische Daten (zur Zeit von 1950 bis 2100) zu allen Staaten und Regionen unserer Welt, die vom Auswärtigen Amt der USA anerkannt werden und mindestens 5000 Einwohner zählen. Die Quellen der Daten der IDB sind Angaben staatlicher Institutionen aus z.B. Volkszählungen, Umfragen oder Einwohnermelderegistern sowie Schätzungen und Projektionen des U.S. Census Bureau selbst.

Der Datensatz in der Datei `census_2022_2002.csv` beinhaltet einen kleinen Auszug aus der IDB. Er enthält Informationen zur Lebenserwartung und Fertilitätsrate für 227 Länder und Territorien aus den Jahren 2002 und 2022. Die genauen Definitionen dieser Variablen entnehmen Sie <https://www.census.gov/programs-surveys/international-programs/about/glossary.html>. Die Lebenserwartung ist aufgeschlüsselt nach Geschlecht. Die Länder sind geographisch in 5 Regionen und 21 Subregionen eingeteilt. Für weitere Details bezüglich der Datenerhebung siehe <https://www.census.gov/programs-surveys/international-programs/about/idb.html>.

Aufgabenstellung

1. Beschreiben Sie die Häufigkeitsverteilungen der Merkmale. Betrachten Sie auch die Unterschiede zwischen den Geschlechtern.
2. Gibt es bivariate Zusammenhänge zwischen den Merkmalen?
3. Wie unterscheiden sich die Merkmale zwischen den Subregionen?
4. Wie haben sich die Merkmale in den letzten 20 Jahren, d.h. im Vergleich von 2002 zu 2022, entwickelt?

Betrachten Sie für die Aufgaben 1–3 nur das Jahr 2022. Dieses Projekt dient der Übung explorativer und deskriptiver Verfahren. Verwenden Sie daher bei allen Aufgabenteilen adäquate statistische Kennzahlen und graphische Methoden zur Auswertung.

Literatur

- Fahrmeir, L., Pigeot, I. und Tutz, Gerhard (2023): Statistik: Der Weg zur Datenanalyse, 9. Auflage, Springer, Berlin.
- Hedderich, J., Sachs, L. (2016): Angewandte Statistik: Methodensammlung mit R, 15. Auflage, Springer-Verlag, Berlin Heidelberg.
- Groß, J. (2010): Grundlegende Statistik mit R : Eine anwendungsorientierte Einführung in die Verwendung der Statistik Software R, Vieweg+Teubner Verlag / GWV Fachverlage GmbH, Wiesbaden.

Abgabe

Abgabe des Berichts und des zugehörigen (lauffähigen und kommentierten) Programmcodes bis Montag, den 28.10.2024, 10:00 Uhr, im Moodle.