

Einführung in R

Übungsblatt

Magnus Bender
Institut für Informationssysteme
Universität zu Lübeck

Übung: 18.01.2022
Abgabe: 24.01.2022

- Die Aufgaben sollen in Gruppenarbeit gemeinsam gelöst werden.
- Eine Gruppe darf nicht mehr als drei Studierende umfassen.
- Die abzugebende Lösung muss die Befehle umfassen, die für das Lösen der einzelnen Aufgaben erforderlich sind.
- Jede*r Studierende muss die erarbeitete Lösung eigenständig im Moodle hochladen, am besten als eine R-Quelldatei `.R`.
- Generierte Grafiken sollen als PDF oder PNG ausgegeben werden und auch im Moodle hochgeladen werden.
- Auf der Abgabe sind die Gruppenmitglieder in der ersten Zeile anzugeben, alle Gruppenmitglieder müssen die gleiche Abgabe hochladen.

-
1. Laden Sie sich den Datensatz `berufstaetig.txt` aus dem Moodle-Kurs herunter. Lesen Sie den heruntergeladenen Datensatz in R ein, sodass die Daten in *einer* Variable gespeichert werden. Achten Sie darauf, dass Sie einen sinnvollen Datentyp für die Variable verwenden, sodass Sie die nachfolgenden Aufgabe möglichst einfach umsetzen können.

Hinweis: Trennzeichen, wie Komma und Leerzeichen, sind ggf. vor dem Einlesen des Datensatzes anzupassen. Beachten Sie hierbei die Wahl des Symbols für die Separierung der Eingabeparameter.

2. Fügen Sie zwei neue Spalten (`berufstaetig`, `erwerbspersonen`) zur zuvor erstellten Variable hinzu. Die erste neue Spalte (`berufstaetig`) soll die Anzahl an Berufstätigen (Arbeitnehmer + Selbstständige) pro Jahr enthalten. Die zweite Spalte (`erwerbspersonen`) soll die Anzahl an Erwerbspersonen (Berufstätige + Erwerbslose) pro Jahr enthalten.

3. Fügen Sie nun in einer weiteren Spalte (`nichtErwerbspersonen`) den Anteil der Nicht-Erwerbspersonen im Bezug auf die Gesamtbevölkerung hinzu.

Lassen Sie sich anschließend eine statistische Kurzzusammenfassung ausgeben (`summary`) und bestimmen Sie die Jahre, in denen der Anteil der Nicht-Erwerbspersonen am höchsten und am niedrigsten war. Speichern Sie die Jahre jeweils in eine Variable ab.

4. Visualisieren Sie

a) die empirische Verteilungsfunktion des Arbeitnehmeranteils an der Bevölkerung und

b) den zeitlichen Verlauf des Arbeitnehmeranteils über die Jahre.

Hinweis: Verwenden Sie zur Visualisierung die Funktion `plot` und `ecdf` mit passenden Parametern.

5. Geben Sie die jeweilige Änderung der Anteile an der Gesamtbevölkerung von Jahr zu Jahr an. Erstellen Sie hierzu die drei folgenden neuen Variablen, die nicht als Teil des Datensatzes `berufstaetig.txt` abgespeichert werden sollen:

- `diff_arbeitnehmer` für die Änderung des Arbeitnehmeranteil,
- `diff_selbststaendige` für die Änderung des Selbstständigenanteils und
- `diff_erwerbslose` für die Änderung des Erwerbslosenanteils.

6. Fassen Sie die in der letzten Aufgabe gebildeten Variablen zu einem neuen Datensatz `aenderung` zusammen. In diesem neuen Datensatz sollen den einzelnen Werten auch die jeweiligen Jahre zugeordnet werden, dabei wird für jede Änderung das angegeben Jahr mit dem Vorjahr verglichen.

Hinweis: Verwenden Sie eine passende Struktur des Datensatzes für die Visualisierung mit `GGPlot2` in der nächsten Aufgabe.

7. Visualisieren Sie die Veränderungen für alle drei Variablen in einem Plot. Wählen Sie hierbei eine geeignete Form der Visualisierung.

Hinweis: Verwenden Sie zur Visualisierung das Paket `GGPlot2`.