

# Einführung in R für SozialwissenschaftlerInnen

## Übungsblatt Woche 6

### Vorgehensweise

Füge, wenn möglich, nur die Befehle in ihre Abgabe ein, die gefordert sind. Alle anderen Befehle können Dein Ergebnis verfälschen. Achte unbedingt darauf, keinen Plot-Befehl einzufügen!

### Aufgabe 1

#### **Ziel: Umgang mit eigenen Funktionen und der Apply-Funktion üben**

Du bist Konsumforscher und untersuchst das Konsumverhalten von Menschen aus anderen Kulturen und Zeiten. Hierbei arbeitest du mit dem in R gespeicherten Datensatz `USPersonalExpenditure`. Der Datensatz umfasst in Milliarden, was die Bewohner der USA in den Jahren 1940-1960 für die Kategorien “Essen und Tabak”, “Haushalt”, “Medizin und Gesundheit”, “Körperpflege” und “Private Bildung” ausgegeben haben.

#### **a) Maximale Punktzahl: 2 Punkte**

Schreibe eine Funktion mit dem Namen “halbieren”, die als Parameter/Argument einen Wert `x` bekommt und diesen durch 2 teilt.

#### **b) Maximale Punktzahl: 2 Punkte**

Schreibe eine Funktion mit dem Namen “spaltensumme”, die als Parameter/Argument einen Dataframe `x` bekommt und für jede Spalte in `x` die Summe der Werte in dieser Spalte berechnet. Nutze hierfür die `apply`-Funktion.

#### **c) Maximale Punktzahl: 2 Punkte**

Schreibe eine Funktion mit dem Namen “zeilendurchschnitt”, die als Parameter/Argument einen Dataframe `x` bekommt und die für jede Zeile in `x` den Durchschnitt der Werte in dieser Zeile berechnet. Nutze hierfür die `apply`-Funktion.

#### **d) Maximale Punktzahl: 1 Punkt**

Berechne mit dem Datensatz `USPersonalExpenditure` für jedes Jahr wie viel Geld die Personen insgesamt in diesem Jahr ausgegeben haben und speichere das Ergebnis in der Variable `d`. Du kannst hierfür die Funktionen benutzen, die du selbst geschrieben hast.

**e) Maximale Punktzahl: 1 Punkt**

Berechne mit dem Datensatz `USPersonalExpenditure` für jede Kategorie den Durchschnitt der Ausgaben über alle Jahre hinweg und speichere das Ergebnis in der Variable `e`. Du kannst hierfür die Funktionen benutzen, die du selbst geschrieben hast.

**f) Maximale Punktzahl: 1 Punkt**

Berechne mit dem Datensatz `USPersonalExpenditure` für jede Kategorie den Durchschnitt der Ausgaben über die Jahre 1940 bis 1945 (eingeschlossen) und speichere das Ergebnis in der Variable `f`. Du kannst hierfür die Funktionen benutzen, die du selbst geschrieben hast.

## Aufgabe 2

**Ziel: Umgang mit der Merge-Funktion üben**

### Hintergrund

Du hast einen Werkstudentenjob bei der Uni und hilfst bei der Kommunikation zwischen Prüfungsamt und Professoren. Der Datensatz `Prfamt.csv` beinhaltet die persönlichen Daten und Noten aller Studenten. Der Datensatz `Statistikurs.csv` beinhaltet die Noten von Studenten in dem Fach Statistik.

### Vorbereitung

Setze folgende Befehle zum Einlesen der Daten an den Anfang deines Skripts:

```
prfamnt<-read.csv2("resources/Prfamnt.csv", encoding="UTF-8")
statistik<-read.csv2("resources/Statistikurs.csv", encoding="UTF-8")
```

Wenn du die Hausaufgaben in R-Studio machst musst du evtl den Pfad ändern und noch zuvor dein Working-Directory setzen. **Bevor du das Skript ins Abgabesystem hochlädst musst du diese Befehle wieder löschen. Im Skript sollen sich nur der Befehl, den du hier siehst und die Befehle für die Hausaufgaben, befinden.**

**a) Maximale Punktzahl: 3 Punkte**

Das Prüfungsamt muss die Noten der Studenten im Fach Statistik eintragen. Füge hierfür die Datensätze `prfamnt` und `statistik` zusammen und speichere das Ergebnis in der Variable `a`. Beachte folgende Dinge:

- Wähle die richtige Spalte zum Zusammenfügen aus, damit keine duplizierten Werte entstehen.
- Das Fach Statistik wird auch von Studenten aus anderen Unis belegt. Füge nur die Studenten ein, die auch im Datensatz des Prüfungsamtes vorkommen.
- Nicht alle Studenten haben Statistik belegt. Diese Studenten sollen statt einer Note ein NA haben.
- Die Zeilen deines Datensatzes sollten nach der Matrikelnummer aufsteigend sortiert sein.
- Verändere nicht die Datentypen der Spalten.

- Entferne alle nicht benötigten Spalten und benenne die Spalten um, sodass dein Datensatz folgendes Format erhält:

```
##      Vorname Nachname Matrikelnummer Fach Englisch Numerik Statistik
## 1      ...      ...      ...      ...      ...      ...      ...
## ...      ...      ...      ...      ...      ...      ...      ...
```

#### b) Maximale Punktzahl: 3 Punkte

Der Professor benötigt die Hauptfächer der Studenten, um den Erfolg der Vorlesung nach Studiengängen zu evaluieren. Füge hierfür die Datensätze prfam und statistik zusammen und speichere das Ergebnis in der Variable b. Beachte folgende Dinge:

- Wähle die richtige Spalte zum Zusammenfügen aus, damit keine duplizierten Werte entstehen.
- Das Fach Statistik wird auch von Studenten aus anderen Unis belegt. Alle Studenten, zu denen keine Informationen im Datensatz des Prüfungsamtes vorliegen, sollen als Fach NA haben.
- Nicht alle Studenten aus dem Prüfungsamt-Datensatz haben Statistik belegt. Füge nur die Studenten ein, die auch Statistik belegt haben.
- Die Zeilen deines Datensatzes sollten nach der Matrikelnummer aufsteigend sortiert sein.
- Verändere nicht die Datentypen der Spalten.
- Entferne alle nicht benötigten Spalten und benenne die Spalten um, sodass dein Datensatz folgendes Format erhält:

```
##      name vorname mtrnr note fach
## 1      ...      ...      ...      ...      ...
## ...      ...      ...      ...      ...      ...
```