Florian Wickelmaier
Forschungsmethoden und
Mathematische Psychologie

Fachbereich Psychologie
Universität Tübingen

Hausaufgaben
Data structures in R (2)

Erzeugen Sie ein lauffähiges und kommentiertes R-Skript (Dateiendung `.R`), in dem Sie die folgenden Aufgaben bearbeiten.

a) (5 Pkt.) Laden Sie von der Website
`https://www.john-fox.ca/AppliedRegression/`
den Datensatz `Vocabulary.txt` herunter und speichern Sie ihn in einem Ordner. Setzen Sie diesen Ordner als Arbeitsverzeichnis in R mit `setwd()` (im Skript auskommentieren!). Lesen Sie die Daten in den Arbeitsspeicher mit `read.table()`.

Welche R-Funktionen eignen sich, um folgende Informationen aus den Daten zu erhalten:

- Anzahl der Beobachtungen,
- Anzahl der Variablen,
- Variablennamen,
- deskriptive Statistik der Variablen?

b) (4 Pkt.) Verwenden Sie den `$`-Operator, um einzelne Variablen aus dem Dataframe zu extrahieren: Berechnen Sie für die Variable `vocabulary` den Mittelwert und die Standardabweichung.

Betrachten Sie die Zeile 2511 des Datensatzes und finden Sie folgende Informationen heraus: Wann wurde diese Person getestet? Wie lange war diese Person in Ausbildung? Wie viele Punkte hatte sie im Wortschatztest?

Indizieren Sie den Dataframe so, dass Sie ausschließlich die gesamten Informationen aller Männer mit weniger als sieben Bildungsjahren erhalten, die 2004 getestet wurden.

c) (2 Pkt.) Sortieren Sie den Datensatz nach "years of education" und innerhalb der "years of education" nach "vocabulary test score". Benutzen Sie dafür die Funktion `order()`.

Wenden Sie auf den sortierten Datensatz die Funktionen `head()` und `tail()` an, um die ersten und die letzten 20 Beobachtungen anzuzeigen. Welche Hypothese würden Sie aufgrund dieser Ergebnisse über den Zusammenhang von "years of education" und "vocabulary test score" aufstellen?

d) (2 Pkt.) Gibt es Geschlechterunterschiede bezüglich "years of education" und "vocabulary test score"? Bestimmen Sie dazu die Mittelwerte dieser beiden Variablen getrennt für Männer und Frauen. Benutzen Sie dafür die Funktion `aggregate()`.

Prüfen Sie, ob die Anzahl der "years of education" im Laufe der letzten Jahre zugenommen hat. Bestimmen Sie dazu deren Mittelwert für jedes "year of testing".

e) (1 Pkt.) Wandeln Sie den in R vorhandenen Datensatz `sleep` (siehe `?sleep`) mit der Funktion `reshape()` um in einen Dataframe im "breiten" Format.