

## Vorbereitung

Bitte führen Sie zur Vorbereitung folgende Schritte aus:

1. Starten Sie RStudio.
2. Löschen Sie den Workspace.
3. Setzen Sie das Arbeitsverzeichnis: `Session >> Set Working Directory >> Choose Directory`.
4. Öffnen Sie ein R-Skript.
5. Nachdem Sie die Aufgaben bearbeitet haben, speichern Sie das Skript unter einem geeigneten Namen ab.

Für die folgenden Aufgaben benutzen wir den `erstis_neu.RData` Datensatz.

## Aufgabe 1 - Modus & Häufigkeitstabelle

Erstellen Sie eine Häufigkeitstabelle für die Variable der Kursgruppe (`gruppe`) und bestimmen Sie den Modus, (a) indem Sie ihn aus der Tabelle ablesen und (b) indem Sie eine angemessene Grafik erstellen und den Modus daran ablesen.

**Zusatzaufgabe:** Benutzen Sie einen anderen Befehl, um sich den Modus direkt ausgeben zu lassen. (Hinweis: Installieren Sie dafür ggf. neue Pakete.)

## Aufgabe 2 - Median

- (i) Erstellen Sie eine Tabelle, die die prozentuale und die kumulierte prozentuale Häufigkeitsverteilung des Items `1z17` („Wenn ich mein Leben noch einmal leben könnte, würde ich kaum etwas ändern.“) gemeinsam abbildet (auf zwei Nachkommastellen gerundet). Bestimmen Sie den Median für diese Variable (a) anhand der erstellten Tabelle, (b) anhand des Befehls für Quantile, und (c) anhand des Befehls für den Median.
- (ii) Eine Person sagt „Mindestens 10 % der Personen haben einen Wert angekreuzt, der kleiner oder gleich meinem Wert ist.“ Welchen Wert hat die Person?
- (iii) Eine Person sagt „Mindestens 70 % der Personen haben einen Wert angekreuzt, der kleiner oder gleich meinem Wert ist.“ Welchen Wert hat die Person?

## Aufgabe 3 - Median II

- (i) Schauen Sie sich die fünf Persönlichkeitsdimensionen (`neuro`, `extra`, `gewiss`, `vertraeg`, `intell`) an. Beantworten Sie die Frage, für welche Skala der Median am höchsten ist. (a) graphisch, (b) mithilfe eines Befehls.
- (ii) Erstellen Sie ein Balkendiagramm für die Variable mit dem höchsten Median. Färben Sie alle Balken in unterschiedlichen Farben, dabei die Mediankategorie rot. Schreiben Sie den Namen der Variablen an die X-Achse.

## Zusatzaufgabe: Relativer Informationsgehalt

Suchen Sie aus dem Datensatz die Stimmungs-Variable (`stim1` bis `stim12`) mit den meisten fehlenden Werten. Bestimmen Sie für diese Variable den relativen Informationsgehalt  $H$ . Ist Ihrer Meinung nach der relative Informationsgehalt  $H$  bestmöglich geeignet, um die Streuung der Merkmale darzustellen?