

Evaluation Randomisiert-Kontrollierter Studien und Experimente mit R

Datenvorbereitung bei RCTs: Dos & Don'ts

Prof. Dr. David Ebert & Mathias Harrer

Graduiertenseminar TUM-FGZ

Psychology & Digital Mental Health Care, Technische Universität München

Vorbereitung der Daten: Dos & Don'ts

Die Vorbereitung der Daten kann mühselig sein, ist aber das **Grundgerüst für alle späteren Schritte**.

→ Die Daten vor Import in R in das korrekte Format zu bringen, ist also lohnenswert.

Dateiformat

Da der Import von *Microsoft Excel* in R sehr einfach ist, ist es empfehlenswert, Daten dort teilweise vorzubereiten. Excel Sheets können unkompliziert als *comma-separated values*-Datei (.csv) exportiert werden. Diese Dateiarart kann besonders einfach und ohne großes Fehlerrisiko in R importiert werden (via `read.csv`).

Dos and Dont's

- Eine **korrekte Benennung der Spalten** bereits in Excel spart oft viel Zeit, da keine Umbenennung mehr in R stattfinden muss.
- Die **Bezeichnung der Variablen** sollte **möglichst kurz sein** und einem **einheitlichen Schema** folgen.
- Für die Bezeichnung der **Zeitpunkte** empfiehlt sich eine **Zahlenkodierung**, die an das Ende des Variablennames angehängt wird. (z.B. `variable.0` für Baseline, `variable.1` für Postmessung etc.)
- Spaltennamen sollten **keine Leerzeichen** beinhalten. Zur Trennung zweier Worte sollten Punkte oder `camelCase` genutzt werden.
- Auch **Sonderzeichen** in Variablennamen sollten vermieden werden, da sie beim Import ggf. verändert werden.

Dos and Dont's

- Wenn der Name der Spalte in der **ersten Zeile der Spalte** eingetragen ist, wird R den Namen **automatisch als Spaltennamen** erkennen. Weitere Formatierung der Spalten ist dann nicht mehr notwendig.
- Die **Reihenfolge der Spalten** in der Excel-Datei ist **unerheblich**.
- Die Excel-Datei sollte nur **ein Tabellenblatt** haben.
- Sollte es mehrere **leere Zeilen und Spalten** geben, ist es ratsam, diese zu **löschen**. Beim Import könnten diese sonst als Missings interpretiert werden.
- Es sollte ein **Codebook** angelegt werden, in dem vermerkt ist, welcher Variable welches Kürzel zugeordnet ist und wie Antworten/Beobachtungen kodiert wurden.

siehe [Beispiel](#) für korrekt vorstrukturiertes Excel-Sheet.