

---

## Leitfaden zur Eingabe von Daten

Die Einhaltung der folgenden Hinweise ermöglicht ein effizientes Datenmanagement und die systematische Auswertung durch Dritte. Alle Angaben sind ohne Anspruch auf Vollständigkeit.

### Allgemeines

- Für die effiziente Auswertung stehen die Daten im Vordergrund, nicht das Design, da Daten i.d.R. exportiert und in einer anderen Software analysiert werden
- Inhalt der Zeilen: Untersuchungseinheiten (z.B. Patient\*innen, Zellen, Mäuse)
- Inhalt der Spalten: Merkmale der Untersuchungseinheiten (z.B. Größe, Alter)
- Jede Datei sollte nur eine einzelne Tabelle enthalten
- Zu jedem Datensatz sollte es ein Code-Book geben, in dem für jede Variable die Kodierung und eine kurze Beschreibung festgehalten wird

### Benennung der Variablen

- Erste Spalte: Eindeutige ID der Untersuchungseinheit
- Erste Zeile: Variablennamen
- Aussagekräftige, **kurze** Variablennamen
- Variablennamen sollten keine Sonderzeichen enthalten (z.B. \*, -, /, Leerzeichen). Eine Ausnahme bildet der Unterstrich (“\_”)
- Variablennamen sollten nicht mit einer Zahl beginnen
- Es sollten keine weiteren Textzeilen, wie z.B. Titel, im Datensatz enthalten sein
- Der Datensatz sollte keine leeren Zeilen enthalten

### Inhalt des Datensatzes

- Zahlen sind in der Auswertung einfacher zu handhaben als Wörter und in der Tabelle ebenfalls übersichtlicher. Kategoriale Variablen sollten deshalb besser mit Zahlen kodiert werden als durch Wörter (Beispiel: Biologisches Geschlecht mit 0 und 1, anstatt “weiblich” und “männlich”)
- Bei binären Variablen (nur zwei mögliche Ausprägungen): 0, 1 statt beispielsweise “nein”, “ja”
- Datumsangaben in der Form YYYY-MM-DD eingeben
- Wenn Variablen in Textform unerlässlich sind: Wenn möglich nur ein Wort und ohne Sonderzeichen (auf Groß- und Kleinschreibung bzw. Einheitlichkeit achten)
- Freitextangaben möglichst vermeiden (Gefahr von Rechtschreibfehlern und unklaren Informationen)
- Alle Informationen müssen der ID der Untersuchungseinheit zuzuordnen sein. Bei mehreren Tabellen werden diese ausschließlich über die ID verbunden
- Es sollten keine Berechnungen in den Rohdaten-Dateien durchgeführt werden

- Bei Mehrfachmessungen pro Untersuchungseinheit sollte jeder Messzeitpunkt in einer Zeile dargestellt werden
- Es sollten nicht mehrere Angaben in einer Zelle/Spalte angegeben werden. Beispiel: Blutdruck (systolisch/diastolisch) 120/80 in zwei Spalten statt einer
- Maßeinheiten sollten nicht im Datensatz, sondern im Codebuch angegeben werden
- Jede Spalte darf Werte nur in einem Format enthalten. Zum Beispiel sind numerische (numeric) mit alphanumerischen (character) Werten nicht zu verwechseln.

## Fehlende Werte

- Fehlende Werte wenn möglich vermeiden
- Wenn ein Wert fehlt, die zugehörige Zelle in der Tabelle **leer** lassen. Kein "fehlt", "N/A" oder ähnliches

## Beispiel für eine gute Datenstruktur

ID	sex	time	insulin
321	0	0	
321	0	5	0.205
321	0	10	0.129
322	1	0	0.251
322	1	5	2.228
322	1	15	2.078

- alle über die Zeit unveränderlichen Daten sollten trotzdem in jeder Zeile auftauchen

## Beispiel für eine schlechte Datenstruktur

Mouse ID	SEX	Week 4			Week 6			Week 8		
		date	weight	glucose	date	weight	glucose	date	weight	glucose
3005	M	3/30/2007	19.3	635	4/11/2007	31	460.7	4/27/2007	39.6	530.2
3017	M	10/6/2006	25.9	202.4	10/19/2006	45.1	384.7	11/3/2006	57.2	458.7
3434	M	11/22/2006	26.6	238.9	12/6/2006	45.9	378	4/27/2007	56.2	409.8