

## Hinweise zur Dateneingabe in Excel

Die folgenden Hinweise zur Dateneingabe in ein Excel-Tabellenblatt sollten von Anfang an berücksichtigt werden, um später eine Auswertung mit einem geeigneten Statistik-Programm zu ermöglichen (z.B. SPSS), ohne die Daten noch einmal komplett neu eingeben oder strukturieren zu müssen.

### Allgemein

- Anonymisierung der Daten: Vergabe von Nummern statt Patientennamen. Eine separate Tabelle enthält die Zuordnung der Nummern zu den Namen.
- Je einfacher die Struktur der Tabelle, umso besser. Kompliziert angelegte Tabellen sind i.A. nicht exportierbar.
- Farbmarkierungen sind nicht exportierbar.
- Keine Auswertungen wie Mittelwerte oder Häufigkeiten in der Tabelle.
- Kommentare zu den Daten dürfen nicht im Tabellenblatt stehen.
- Löschung nicht genutzter Zellen durch den Befehl "Zellen löschen" mittels rechter Maustaste.

### Daten-Struktur

- Pro Zeile: eine Beobachtung (z.B. Patient), auch bei Mehrfachmessungen.
- Pro Spalte: eine Variable.
- 1. Zeile: Variablenname nach bestimmten Vorgaben (s.u.), damit diese exportierbar sind.
- Bei mehreren Gruppen o.ä.: Eine einzige Tabelle anlegen mit einer Spalte, in der die Gruppenzugehörigkeit steht und nicht pro Gruppe eine Tabelle.

### Variablennamen

- Variablennamen sollten nicht zu lang sein.
- Buchstaben, Ziffern und einige Sonderzeichen sind erlaubt, z.B. `_`. Keine Leerzeichen!
- Variablennamen sollten mit einem Buchstaben anfangen.
- Eine gewisse Verwandtschaft zum Inhalt sollte erkennbar sein, also nicht X1-X345 (s. Beispieltabellen unten).
- Insbesondere bei Mehrfachmessungen ist eine Benennung in der Art MESS1-MESS10 aber sinnvoll.
- Keine „Kapitelüberschriften“ (wie z.B. „Aufnahme“, „Entlassung“), unter denen dann die gleichen Variablennamen vergeben werden. Falls nötig, z.B. Aufnahme\_Gewicht, Entlassung\_Gewicht o.ä. kodieren.
- Grundsätzlich dürfen Variablennamen nicht doppelt vergeben werden.

### Variablen

Die Spalten müssen einheitlich formatiert sein und einheitliche Einträge haben, d.h.:

- Ein Format (d.h. Datentyp) pro Spalte, also nur Zahlen, nur Datumsangaben oder nur Text.
- Datumsangaben als Datum formatieren, am besten in dieser Form: 01.01.2000.
- Keine Kommentare zwischen die Werte schreiben, z.B. beim Todesdatum die Todesursache dazuschreiben. Die Todesursache sollte in einer separaten Spalte kodiert werden. Ggf. eine eigene Spalte für Kommentare/Besonderheiten anlegen. Freitext kann allerdings nicht statistisch ausgewertet werden.
- Keine Sonderzeichen verwenden (auch in Text-Spalten nur wenn unvermeidbar). Insbesondere kein Semikolon verwenden, da dieser beim Einlesen als Trennzeichen verstanden werden kann.
- Spalten komplett ausfüllen, soweit Daten vorhanden sind. In freigebliebene Felder nicht „nein“ eintragen, da sonst keine Unterscheidung zu fehlenden Werten (= Daten nicht erhoben) möglich.
- Bei fehlenden Werten Feld leer lassen oder z.B. -1 oder 99 als Kodierung für Fehlwert eintragen.
- Spalten, die sich durch Berechnungen aus anderen Spalten ergeben, sind i.A. überflüssig, da man diese mit der Statistiksoftware einfach berechnen kann.
- Wenn in Excel Berechnungen durchgeführt werden: Ergebnisse als Werte eintragen und nicht als Formeln, die bei jedem Öffnen der Tabelle neu durchgeführt werden.

## Codes

- Für die verwendeten Codes am besten ein neues Tabellenblatt der Excel-Datei mit folgendem Inhalt erstellen:
  1. Spalte: Variablenname
  2. Spalte: Erklärung
  3. Spalte: Codes
- Ja/Nein-Fragen mit 0/1 kodieren (1=ja, 0=nein)
- Qualitative Merkmale müssen kodiert werden (z.B. 1=Therapie A, 2=Therapie B, 3=Therapie C)
- Wenn Mehrfachnennungen möglich sind, z.B. Therapie A + Therapie C, dann besser für jedes Element eine eigene Spalte anlegen und mit 0/1 kodieren, ob dies vorliegt (nein/ja).

## Beispieltabelle Daten

PATNR	GESCHLECHT	ALTER	THERARM	RANDODAT	DIAGNOSE	BLOODSYS1	BLOODSYS2
1	1	40	1	15.11.2014	1	160	152
2	2	37	2	18.11.2014	1	152	145
3	2	68	2	21.11.2014	3	165	147
4	1	43	1	28.11.2014	2	143	142
5	1	70	1	04.12.2014	3	165	154
6	2	75	2	12.12.2014	3	163	151

## Beispieltabelle Codes/Variablenbeschreibung

Variablenname	Label	Kodierung
PATNR	Patientennummer	
GESCHLECHT	Geschlecht	1=männlich, 2=weiblich
ALTER	Alter in Jahren	
THERARM	Therapiearm	1=CHOP, 2=CHOEP
RANDODAT	Datum der Randomisation	
DIAGNOSE	Pathologische Diagnose	1=centroblastisch, 2=immunoblastisch, 3=andere diffus großzellige
BLOODSYS1	Syst. Blutdruck in mmHg (vor Therapie)	
BLOODSYS2	Syst. Blutdruck in mmHg (nach Therapie)	