

# Formatting Requirements

## Hagenbuch Statistical Consulting

1. In your data set, include only the variables needed for the statistical analysis. Variables that are not of interest should be removed (no “data spamming”).
2. Use a table form. Rows represent observational units, columns the measured variables. A single row comprises all information of the respective subject of the study.
3. Variable names are in the first row. Use short (max. 8 characters) and unique names that you will understand later. They are used throughout the report. The first character has to be a letter. Do not use spaces, punctuation (exception: period) or other special characters (ä, ö, ü, ...).

**Allowed:** ID, Weight.pre, pCO2, T1a, Groesse, ...

**Not allowed:** Gebärmutterhalsumfang, Pat.ID, Alk%, dead/alive, #Infarcts, 5-Pt.Fixation, Bennett’s Scale, ...

4. The above applies for the labelling of levels in factors as well.
5. Numeric variables only contain figures, the negative sign (–) and the period.

**Note:** Excel displays numbers right-justified; when formatted as text by mistake, they appear left-justified.

6. In case you have values based on time calculations like age difference or a duration, report the final value. For dates, use a text string, preferably in the format YYYY-MM-DD (be careful with the American representation month-day-year).

**Recommended:** 2014-11-23

**Not recommended:** 2014-November-23, 2-August-2011, 07.05.2012, 10/06/11.

7. The first column is usually the identifier (ID).
8. Do not use any real names, not even abbreviations. Remove addresses. (Special rules apply for spatial analyses.)

**Note:** Names and confidential information are sometimes listed on additional tabs in Excel. Check for hidden rows or columns!

**Note:** Compare the number of rows with the number of observations, and the number or columns with the number of variables.

9. Few repeated measurements are stored in separated variables/columns (so-called wide format).

ID	Gender	Age	Weight.pre	Weight.post
1	m	27	82	79
2	f	39	71	66
3	f	31	77	72
⋮	⋮	⋮	⋮	⋮

10. Numerous repeated measurements or measurements taken at irregular intervals are stored in separate rows (so-called long format). In addition to the measured variable, another variable indicating the time or the sequence of the observation is required. Variables that do not change are reiterated.

ID	Gender	Age	Weight	Group	Time	Conc.	BP	Pulse
1	m	27	79	A	0	124	138	109
1	m	27	79	A	10	102	133	97
1	m	27	79	A	30	88	129	82
2	f	39	66	B	0	119	142	110
2	f	39	66	B	20	96	128	101
2	f	39	66	B	35	89	121	92
2	f	39	66	B	60	79	118	83
3	f	31	72	A	2	128	143	115
3	f	31	72	A	30	101	127	99
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

11. A separate text file with explanations is highly appreciated:
- List all variables with their units, and give a short description.
  - Indicate the lower and upper potential limit of *continuous* and *count* data.  
**Example:** – Visual Analog Scale (VAS): 0–10 cm.
  - For *nominal* variables, indicate all categories  
**Examples:** – Pregnant: 0 (no), 1 (yes)  
– State: 0 (alive), 1 (dead)  
– Type: Lim (limousine), MV (minivan), Sed (sedan), St (station wagon).  
**Note:** Be consistent! (E.g., in the variable **Gender**, do not mix **m, male, M, f, female, F.**)
  - For *ordinal* variables, indicate all categories and their order.  
**Examples:** – NYHA Class: 1 (no symptoms), 2 (mild symptoms), 3 (marked limitations), 4 (severe limitations)  
– Opioids (in increasing analgetic potency): Mo (morphine), Oxy (oxycodone), Fen (fentanyl), Rem (remifentanil), Suf (sufentanil)
12. In case of missing values, leave the cells empty. Do not implicitly use spaces, 0, 99, 9999, or any other value. Declare if such values were used.
13. Do not code information using special formatting in Excel, like green background = “control group”. Use a separate variable.
14. The preferred file format is `.csv` or `.txt` (all common software allows to choose this save or export type).