



## Anleitung zur Datenerfassung in Excel/SPSS

### 1 Ziel und Zweck

Diese Anleitung beschreibt die Vorgehensweise bei der Erfassung von Daten mittels geeigneter Programme, um einen einheitlichen Datenaufbau zu gewährleisten und die anschließende Auswertung zu erleichtern.

### 2 Anwendung

ÄrztInnen, DoktorandInnen etc., die Daten zu Auswertungszwecken erfassen.

### 3 Beschreibung des Vorgehens

#### 3.1 Datenerfassungsprogramme

Für die Datenerfassung eignen sich

- Tabellenkalkulationsprogramme (z.B. MS Excel, OpenOffice Calc, Apple Numbers),
- Datenbanken (z.B. MS Access, MySQL)
- Statistikprogramme (z.B. IBM SPSS)

Die im Folgenden aufgestellten Regeln beziehen sich primär auf die Nutzung von Excel zur Erfassung und SPSS zur Auswertung der Daten, sind aber auch auf andere Programme übertragbar.

#### 3.2 Datenaufbau

Bei der Anordnung der Daten ist folgendes zu beachten:

- pro Patient bzw. Untersuchungsobjekt (= Datensatz) eine Zeile, pro Merkmal/Variable eine Spalte
- jeder Datensatz erhält eine eindeutige Identifikationsnummer
- Patientennamen werden aus Datenschutzgründen nicht erfasst. Die Zuordnung von Patientennamen und zugehöriger ID erfolgt separat
- für jede Spalte wird in der ersten Zeile ein eindeutiger Variablenname vergeben
- die Tabelle enthält keine Leerzeilen, keine Leerspalten, keine Formatierungen (z.B. Fettdruck, farbig hinterlegte Zellen etc.) und keine zusätzlichen, spaltenweise berechneten Werte
- jedes Arbeitsblatt enthält nur eine Tabelle

#### 3.3 Variablennamen

Für die Benennung der Variablen gelten folgende Regeln:

- jeder Variablenname muss eindeutig sein, Groß- und Kleinschreibung bleibt dabei unberücksichtigt (gebdat = GebDat)
- erlaubte Zeichen sind Buchstaben, Ziffern, Unterstrich, Punkt, At-Zeichen (@) und Nummernzeichen (#)
- das erste Zeichen muss ein Buchstabe sein, das letzte Zeichen sollte ein Buchstabe oder eine Ziffer sein
- nicht erlaubt sind Leerzeichen und sämtliche Sonderzeichen außer den oben genannten
- In SPSS sind Variablennamen mit bis zu 64 Zeichen möglich, es ist aber ratsam diese trotzdem kurz und sprechend zu halten (<= 15 Zeichen)



### 3.4 Werte, Kodierung

Es ist darauf zu achten, alle Merkmale nach Möglichkeit mit numerischen Werten zu erfassen. Angaben in Textform sind entsprechend zu kodieren (z.B. Geschlecht: 0 = männlich, 1 = weiblich / Diabetes: 0 = nein, 1 = ja). Die Kodierung der Variablen wird nicht im gleichen Arbeitsblatt, sondern separat dokumentiert.

Falls es notwendig sein sollte Textvariablen mit vielen unterschiedlichen Ausprägungen zu erfassen (z.B. Diagnosen), so sollten diese bei der nachfolgenden Kodierung in inhaltlich sinnvollen Gruppen zusammengefasst werden, um zu viele kleine Häufigkeiten zu vermeiden.

Bei fehlenden Werten wird die entsprechende Zelle leer gelassen, oder es wird ein Wert eingetragen, der außerhalb des Wertebereichs der Variablen liegt (z.B. -9 oder 999). Dieser Wert kann später in SPSS als fehlender Wert deklariert werden.



## 4 Anhang

### Anhang 1: Beispiel für korrekte Datenerfassung

	A	B	C	D	E	F	G	H	I	J	K	L
1	patnr	gruppe	gebdat	studegdat	sex	größe	gewicht	rrsys_tag1	rrsys_tag7	cholesterin	hdl	
2	1	1	02.04.1914	30.08.1998	0	162	70	182	159	338,0	26,0	
3	2	0	07.01.1950	26.12.1997	1	164	90	180	128	200,0	57,0	
4	3	0	04.03.1926	18.05.1998	0	184	95	180	166	162,0	23,0	
5	4	0	13.11.1924	06.08.1998	0	150	56	260	180	238,0	45,0	
6	5	0	15.11.1937	12.11.1998	1	165	86	250	160	295,3	41,8	
7	6	1	04.05.1926	09.02.1998	0	157	83	240	195	143,0	23,2	
8	7	0	18.12.1944	29.10.1998	1	166	85	240	190	183,0	37,0	
9	8	1	10.03.1918	11.05.1998	0	152	70	240	160	281,0	39,0	
10	9	0	05.06.1941	15.09.1998	1	178	90	240	160	234,0	35,0	
11	10	1	31.03.1923	22.03.1998	9	162	72	235	160	217,0	38,0	
12	11	0	06.06.1945	17.01.1998	1	187	99	230	190	174,0	73,0	
13	12	0	15.12.1936	28.02.1998	1	165	54	230	180	291,0	29,0	
14	13	1	25.08.1954	17.12.1997	1	178	126	220	160	247,0	24,0	
15	14	0	26.08.1914	11.02.1998	0	160	74	220	160	193,0	39,0	
16	15	1	02.05.1943	03.09.1998	1	175	85	220	140	-9,0	-9,0	
17	16	0	19.11.1939	21.07.1998	0	161	80	220	185	223,0	53,0	
18	17	0	04.10.1922	02.02.1998	1	176	80	220	190	178,0	29,0	
19	18	0	27.06.1930	17.02.1998	1	170	70	220	170	264,0	63,0	
20	19	0	18.08.1928	14.12.1997	1	160	90	220	210	85,5	15,9	
21	20	0	27.12.1929	23.06.1998	1	168	93	210	220	155,0	37,0	
22												
23												



Anhang 2: Beispiel für fehlerhafte Datenerfassung

	A	B	C	D	E	F	G	H	I	J	K	L
1	patnr	gruppe	gebdat	studbegdat	sex 0=weiblich 1=männlich	größe in cm	gewicht	rrsys-tag1	rrsys_tag7	cholesterin	hdl	
2	1	1	02.04.1914	30.08.1998	0	162	70	182	159 mmHg	338,0	26,0	
3	2	0	07.01.1950	26.12.1997	1	164	90	180	128	200,0	57,0	
4	3	0	04.03.1926	18.05.1998	0	184	95	180	166	162,0	23,0	
5	4	0	13.11.1924	06.08.1998	0	150	56	260	180	238,0	45,0	
6	5	0	15.11.1937	12.11.1998	1	165	86	250	160	295,3	41,8	
7	6	1	04.05.1926	09.02.1998	0	157	83	240	195	143,0	23,2	
8	7	0	18.12.1944	29.10.1998	1	166	85	240	190	183,0	37,0	
9	8	1	10.03.1918	11.05.1998	0	152	70	240	160	281,0	39,0	
10	9	0	05.06.1941	15.09.1998	1	178	90	240	160	234,0	35,0	
11	10	1	31.03.1923	22.03.1998	9	162	72	235	160	217,0	38,0	
12												
13	11	0	06.06.1945	17.01.1998	1	187	99	230	190	174,0	73,0	
14	12	0	15.12.1936	28.02.1998	1	165	54	230	180	291,0	29,0	
15	13	1	25.08.1954	17.12.1997	1	178	126	220	160	247,0	24,0	
16	14	0	26.08.1914	11.02.1998	0	160	74	220	160	193,0	39,0	
17	15	1	02.05.1943	03.09.1998	1	175	85	220	140 ?		-9,0	
18	16	0	19.11.1939	21.07.1998	0	161	80	220	185	223,0	53,0	
19	17	0	04.10.1922	02.02.1998	1	176	80	220	190	178,0	29,0	
20	18	0	27.06.1930	17.02.1998	1	170	70	220	170	264,0	63,0	
21	19	0	18.08.1928	14.12.1997	1	160	90	220	210	85,5	15,9	
22	20	0	27.12.1929	23.06.1998	1	168	93	210	220	155,0	37,0	
23												MW: 35,94
24												
25												
26												
27												
28												
29												

Leerzeile

Formatierung

zusätzliche Info zum Variablennamen

ungültige Variablennamen

keine numerischen Angaben

zusätzlich berechneter Wert