Universität Zürich[UZH]

UniversitätsSpital Zürich

**Horten Zentrum
für praxisorientierte Forschung
und Wissenstransfer**

Universität Zürich / UniversitätsSpital
Pestalozzistrasse 24
CH-8091 Zürich
Telefon +41 44 255 86 62
Telefax +41 44 255 97 20
www.evimed.ch

Anforderungen an Datensätze für die statistische Analyse

1. Vorbereiten eines Datensatzes zur Analyse

In dieser Anleitung erklären wir den korrekten Aufbau Ihres Datensatzes, damit er anschliessend mit gängigen Statistik-Programmen ausgewertet werden kann. Zudem weisen wir auch auf die häufigsten Fehler hin. Wir halten uns vor, im Aufbau mangelnde Datensätze zur korrekten Aufbereitung zurückzuschicken.

1.1. Datenschutz

Die Einhaltung des Datenschutzes ist von besonderer Wichtigkeit. Die Datenfiles sollten unter keinen Umständen Informationen, die das Datenschutzgesetz (siehe http://www.edoeb.admin.ch/) verletzen, enthalten. Insbesondere dürfen die Patienten in medizinischen Studien nicht identifizierbar sein, d.h. **Patientennamen müssen durch eindeutige Identifikationsnummern ersetzt werden** (siehe untenstehende Tabellen).
Achtung: Die unerwünschte Übermittlung von Patientennamen geschieht häufig durch „nicht gebrauchten“ resp. „vergessen gegangenen" Tabellenblättern (sheets) in einem Excel-Dokument.

2. Datenformat

Die gängigen statistischen Auswertungsprogramme setzen voraus, dass die zu verarbeitenden Rohdaten „rechteckig" angeordnet sind. Die ersten Felder jeder Zeile sind üblicherweise solchen Variablen zugeordnet, mit denen sich die jeweilige Beobachtungseinheit identifizieren lässt. Häufig ist dies eine einzige Variable, etwa eine Identifikationsnummer für die Patienten einer Stichprobe. Daran schliessen sich die Felder an, in denen die Werte weiterer Merkmale erfasst werden.

Wichtig:
Das Datenfile wird von uns als **final** angesehen, ständige Änderungen im Laufe der statistischen Analyse können **nicht** akzeptiert werden. Falls ausnahmsweise Änderungen am Datenfile erforderlich sind, **muss** das **korrigierte Datenfile erneut** geschickt werden.

Universität Zürich[UZH]
UniversitätsSpital Zürich
Horten Zentrum für praxisorientierte Forschung und Wissenstransfer

Universität Zürich / UniversitätsSpital
Pestalozzistrasse 24
CH-8091 Zürich
Telefon +41 44 255 86 62
Telefax +41 44 255 97 20
www.evimed.ch

2.1. Optische Formatierung Tabellenblatt

Das Datenblatt resp. Tabellenblatt sollte nicht optisch formatiert werden (keine Farben, verschiedene Schriftarten und Grössen etc.). Zudem dürfen keine Grafiken oder Formeln enthalten sein. In der ersten Zeile (Nr. 1) stehen immer die Spaltenbeschriftungen (Abbildung 1). Diese dürfen **KEINE** Umlaute (ä, ö, ü) und **KEINE** Sonderzeichen (z.B. *, #, @) ausser „%“ enthalten.
In der ersten Spalte (Spalte A oder 1) muss immer die eindeutige Fall-Kennung enthalten (z.B. Pat-ID, Paper-Nr, etc.).
Es dürfen keine Zeilen irgendwo zwischendurch mit neuen Überschriften resp. Kommentaren versehen werden (ausgenommen Kommentarspalte, welche immer die letzte Spalte sein muss. Dort darf man beliebig Text reinschreiben).
Falls man nicht alle Zeilen in die Berechnungen einbeziehen möchte (z.B. einzelne Patienten oder Untersuchungenzeitpunkte), sollen diese durch die Extraspalte „In/Out“ mit den Werten in = 1 und out = 0 gekennzeichnet werden (Abbildung 2). Somit werden diese Zeilen (und somit die Inhalte aller Zellen der entsprechenden Zeile) für die Berechnung ausgelassen.

	A	B	C	D	E	F
1	Pat-ID	Sex	Age	initial date	blood pressure baseline systolic	blood pressure baseline diastolic
2	1	f	58	10.10.2010	142	94
3	2	f	85	10.10.2010	137	85
4	3	m	53	10.10.2010	130	87
5	4	f	64	14.10.2010	143	93
6	5	m	75	14.10.2010	134	88
7	6	m	79	14.10.2010	158	105
8	7	m	59	14.10.2010	131	90
9	8	m	61	14.10.2010	143	103
10	9	f	82	18.10.2010	141	85
11	10	f	69	18.10.2010	152	102

Abbildung 1

	A	B	C	D	E	F	G
1	Pat-ID	Sex	Age	initial date	blood pressure baseline systolic	blood pressure baseline diastolic	In / Out
2	1	f	58	10.10.2010	142	94	1
3	2	f	85	10.10.2010	137	85	1
4	3	m	53	10.10.2010	130	87	1
5	4	f	64	14.10.2010	143	93	1
6	5	m	75	14.10.2010	134	88	0
7	6	m	79	14.10.2010	158	105	1
8	7	m	59	14.10.2010	131	90	1
9	8	m	61	14.10.2010	143	103	1
10	9	f	82	18.10.2010	141	85	0
11	10	f	69	18.10.2010	152	102	0

Abbildung 2

2.2. Leere Felder

Bei fehlenden Werten müssen die Felder zwingend **LEER** gelassen werden! Es sind **KEINE** Platzhalter erlaubt (wie z.B. „999“, „–“ (minus), „NA“, „KA“ etc.)

2.3. Formatierung Zahlenfelder

Zellen mit numerischen Variablen dürfen nur Ziffern, das Vorzeichen „–“ (minus) sowie Dezimalpunkt enthalten. Es dürfen **KEIN** Dezimalkomma (z.B. 4,58) oder Sonderzeichen wie z.B. „+“ oder „±“ verwendet werden.
Zahlen müssen korrekt formatiert werden (Zellen formatieren → Reiter „Zahlen“ → Zahl auswählen (mit entsprechenden Dezimalstellen): als Zahl formatierte Zahlen werden dabei rechtsbündig dargestellt, als Text formatierte dagegen links (Abbildung 3).

	A	B	C
1	1234		als **Zahl** formatiert
2	1234		als **Text** formatiert

Abbildung 3

Universität Zürich[UZH]
UniversitätsSpital Zürich
Horten Zentrum für praxisorientierte Forschung und Wissenstransfer

Universität Zürich / UniversitätsSpital
Pestalozzistrasse 24
CH-8091 Zürich
Telefon +41 44 255 86 62
Telefax +41 44 255 97 20
www.evimed.ch

2.4. Formatierung alphanumerische Felder

Variabelwerte, die nicht nur Ziffern sondern auch alphanumerische Zeichen enthalten (z.B. T0 oder CO2), dürfen **KEINE** Umlaute oder Sonderzeichen enthalten (z.B. Ü1a oder CO_2).

2.5. Formatierung Datumszellen

Kalenderdaten müssen zwingend wie folgt formatiert und dargestellt werden: rechte Maustaste auf die entsprechende Zelle (oder ganze Spalte markieren und rechte Maustaste) → Zellen formatieren (Abbildung 4) → Reiter „Zahlen“ → Datum auswählen (wie in Abbildung 5 dargestellt)

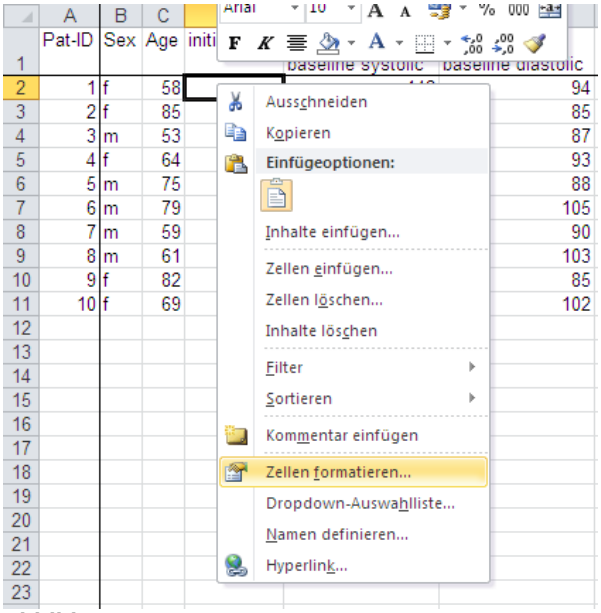

Abbildung 4

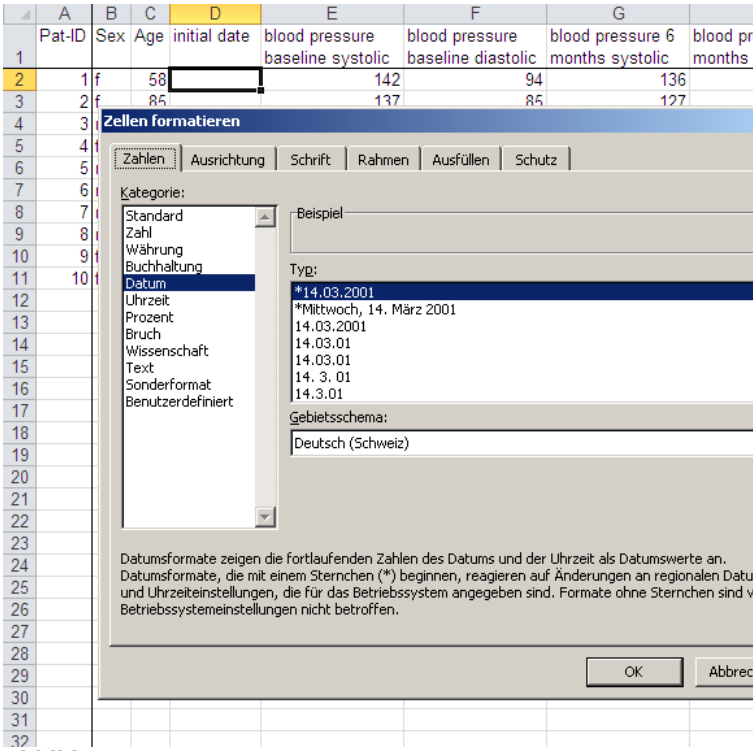

Abbildung 5

Universität Zürich[UZH]
UniversitätsSpital Zürich
Horten Zentrum für praxisorientierte Forschung und Wissenstransfer

Universität Zürich / UniversitätsSpital
Pestalozzistrasse 24
CH-8091 Zürich
Telefon +41 44 255 86 62
Telefax +41 44 255 97 20
www.evimed.ch

2.6. Text

Text ist nicht unmittelbar auswertbar und muss deshalb kodiert werden. Anstatt „gelb“, „grün“ und „blau“ zu schreiben (Abbildung 7), sollte man diese numerisch kodieren: 1 = gelb, 2 = grün, 3 = blau (Abbildung 6). Wichtig dabei ist, dass man ein Codebook anlegt (eigenes Tabellenblatt, siehe „Beispiel Codebook“), wo man die Kodierung aufführt. Sie sollte nicht in der Spaltenbeschriftung mitaufgeführt werden (Abbildung 8).

✓

	A	B	C
1	Pat-ID	Alter	Farbe Gruppe
2	1	45	1
3	2	31	2
4	3	65	3

Abbildung 6

✗

	A	B	C
1	Pat-ID	Alter	Farbe Gruppe
2	1	45	gelb
3	2	31	grün
4	3	65	blau

Abbildung 7

	A	B	C
1	Pat-ID	Alter	Farbe Gruppe (1 = gelb, 2 = grün, 3 = blau)
2	1	45	1
3	2	31	2
4	3	65	3

Abbildung 8

2.7. Beispiel Codebook

Die Kodierung muss für Variabeln ab zwei Zeichen numerisch sein (Abbildung 9). Einzige Ausnahmen, welche erlaubt sind: y/n (yes/no) resp j/n (ja/nein) und f/m (female/male) resp. w/m (weiblich/männlich). Codes für die gleichen Antwortkategorien müssen für alle Variablen gleich sein (z.B. 0 = nein, 1 = ja)

	A	B	C	D
1	**Codebook**			
2	**Spalte**	**Code**	**Bedeutung**	
3				
4	Farbe Gruppe	1	gelb	
5		2	grün	
6		3	blau	
7				
8	BMI Kategorie	1	BMI <25	
9		2	BMI 25 bis <30	
10		3	BMI 30 bis <35	
11		4	BMI 35 bis <40	
12		5	BMI >40	
13				
14				
45				
46				
47				

Wide | Long | Format | Code | Codebook

Abbildung 9

Universität Zürich[UZH]
UniversitätsSpital Zürich
Horten Zentrum für praxisorientierte Forschung und Wissenstransfer

Universität Zürich / UniversitätsSpital
Pestalozzistrasse 24
CH-8091 Zürich
Telefon +41 44 255 86 62
Telefax +41 44 255 97 20
www.evimed.ch

2.8. Wide Range Tabelle

Bei der Wide Range Tabelle (Abbildung 10) ist jeder Patient nur auf einer Zeile dargestellt. Longitudinale Messungen werden als zusätzliche Spalte hinzugefügt. Falls Zeitpunkte angegeben werden, müssen die in Extraspalten (im Datumsformat) hinzugefügt werden.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Pat-ID	Sex	Age	initial date	blood pressure baseline systolic	blood pressure baseline diastolic	blood pressure 6 months systolic	blood pressure 6 months diastolic	blood pressure 12 months systolic	blood pressure 12 months diastolic	blood pressure 18 months systolic	blood pressure 18 months diastolic
2	1	f	58	10.10.2010	142	94	136	88	133	85	130	82
3	2	f	85	10.10.2010	137	85	127	75	127	75	127	75
4	3	m	53	10.10.2010	130	87	124	81	120	77	119	76
5	4	f	64	14.10.2010	143	93	135	85	130	80	127	77
6	5	m	75	14.10.2010	134	88	128	82	128	82	126	80
7	6	m	79	14.10.2010	158	105	154	101	151	98	148	95
8	7	m	59	14.10.2010	131	90	127	86	125	84	123	82
9	8	m	61	14.10.2010	143	103	141	101	139	99	136	96
10	9	f	82	18.10.2010	141	85	137	81	137	81	137	81
11	10	f	69	18.10.2010	152	102	148	98	146	96	145	95

Abbildung 10

2.9. Long Range Tabelle

Bei der Long Range Tabelle (Abbildung 11) ist für jede Messung eine eigene Zeile vorgesehen (die gleiche Patienten-ID kann mehrfach vorkommen).

	A	B	C	D	E	F	G
1	Pat-ID	Sex	Age	initial date	time point (months)	blood pressure systolic	blood pressure diastolic
2	1	f	58	10.10.2010	0	142	94
3	1	f	58	10.10.2010	6	136	88
4	1	f	58	10.10.2010	12	133	85
5	1	f	58	10.10.2010	18	130	82
6	2	f	85	10.10.2010	0	137	85
7	2	f	85	10.10.2010	6	127	75
8	2	f	85	10.10.2010	12	127	75
9	2	f	85	10.10.2010	18	127	75
10	3	m	53	10.10.2010	0	130	87
11	3	m	53	10.10.2010	6	124	81
12	3	m	53	10.10.2010	12	120	77
13	3	m	53	10.10.2010	18	119	76
14	4	f	64	14.10.2010	0	143	93
15	4	f	64	14.10.2010	6	135	85
16	4	f	64	14.10.2010	12	130	80
17	4	f	64	14.10.2010	18	127	77
18	5	m	75	14.10.2010	0	134	88
19	5	m	75	14.10.2010	6	128	82
20	5	m	75	14.10.2010	12	128	82
21	5	m	75	14.10.2010	18	126	80
22	⋮		⋮	⋮	⋮	⋮	⋮
23							
24							

Abbildung 11