



Empfehlungen zur Struktur und der Erfassung der Daten

Die folgenden Anmerkungen sollen als Empfehlung verstanden werden. Eine Nichtbeachtung kann jedoch bei der EDV-gestützten Dateneingabe sowohl zu merklichem Mehraufwand führen als auch die spätere Auswertung der Daten erheblich komplizieren.

Datenstruktur

Die gängigen statistischen Auswertungsprogramme (u. a. auch die am IMSB verfügbaren Programmpakete IBM SPSS® Statistics und SAS®) setzen voraus, dass die zu verarbeitenden Rohdaten in einer "rechteckigen" Datenstruktur angeordnet sind.

Ein Datensatz besteht aus Zeilen, jede Zeile wird in Spalten eingeteilt. Jede Spalte kann ein oder mehrere Zeichen (Ziffer, Buchstabe) aufnehmen.

Jede *Zeile* repräsentiert eine Untersuchungseinheit (z. B. Patient, Proband, Zahnspange, Hand, Wirbelkörper usw.).

Jede *Spalte* repräsentiert ein Merkmal (Variable) einer Untersuchungseinheit (z. B. Geschlecht, Alter, Größe, Material, Blutdruck, Kinderkrankheit usw.)

	Patient	Geschlecht	Geburtsdatum	Alter	Raucher	Sport	sysbldr1	sysbldr2	var
1	263	1	29.07.1962	18	1	2	124	110	
2	294	0	01.06.1963	17	0	9	140	142	
3	398	0	12.09.1962	18	1	0	134	124	
4	455	0	03.07.1962	18	1	1	142	144	
5	472	1	30.01.1960	21	1	3	132	126	
6	513	0	07.12.1962	18	1	0	144	134	

Die Länge der Felder kann von Merkmal zu Merkmal unterschiedlich sein, muss aber so gewählt werden, dass jede denkbare Messung des Merkmals erfasst werden kann. So ließe sich etwa das Merkmal "Körpergröße in [cm]" für Patienten in einem Feld mit drei Stellen für alle denkbaren Messergebnisse erfassen, nicht jedoch in einem Feld mit nur zwei Stellen.

Werden einzelne Merkmale für jede Beobachtungseinheit zu verschiedenen Zeitpunkten wiederholt erhoben, wie etwa bei Messung des systolischen Blutdrucks unmittelbar vor und zwei Stunden nach Gabe eines Medikamentes, so muss für dieses Merkmal für jeden Messzeitpunkt ein "eigenes" Feld und somit eine gesonderte Variable zugeordnet werden (z. B. "Systolischer Blutdruck 1. Messung" und "Systolischer Blutdruck 2. Messung").



Datenerfassung

Für die Auswertung mit dem Programmpaket IBM SPSS® besteht die Möglichkeit, die Rohdaten mit dem Dateneditor von IBM SPSS® zu erfassen. Sollten die Rohdaten jedoch mit anderen Programmen (z. B. Microsoft Excel®) erfasst werden, sind folgende Hinweise zu beachten:

- Variablenamen müssen eindeutig sein.
- Sie dürfen maximal 64 Zeichen lang sein und beginnen mit einem der Zeichen „A-Z“ oder „a-z“. Leerzeichen sind nicht gestattet.
- Dem ersten Zeichen eines Namens dürfen beliebige Kombinationen aus Buchstaben, Ziffern, die Zeichen „@“, „#“, „\$“, „.“, und „_“ folgen.
- Variablenamen dürfen nicht mit „.“ oder „_“ enden.
- Felder mit numerischen Variablen dürfen nur Ziffern, das Vorzeichen „-“ sowie Dezimalpunkt oder -komma enthalten.
- Fehlende Werte ("missing values") sind durch einen speziellen Code zu kennzeichnen oder die Felder bleiben leer. Auf keinen Fall sollten Sonderzeichen wie „?“ oder Buchstaben „k. A.“ benutzt werden.
- Klartext ist bei der Auswertung problematisch und sollte, wenn möglich, sinnvoll kodiert werden (z. B. Kinderkrankheiten: 1 = Masern, 2 = Röteln, ...).

Die Verwendung eines Tabellenkalkulationsprogramms wie z. B. Microsoft Excel® zur *Datenauswertung* kann nicht empfohlen werden, da für diese Zwecke leistungsfähige Statistikpakete (s. o.) zur Verfügung stehen.

Wichtige Hinweise

In den Amtlichen Mitteilungen 24/2011 ist die "Ordnung der Universität zu Köln zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten vom 22. Juli 2011" (zuletzt geändert: 20.07.2016) veröffentlicht worden.

Abfragedatum Downloadadresse: 07.02.2017

http://strategy.uni-koeln.de/e142759/e240/e136815/Ordnung_gute_wiss_Praxis_ger.pdf

In § 4 heißt es u. a., dass die oder der für ein Forschungsprojekt Verantwortliche sicherzustellen hat, dass Primärdaten als Grundlage für Veröffentlichungen auf haltbaren und gesicherten Trägern zehn Jahre in der Institution, in der sie entstanden sind, aufbewahrt werden.

Zudem wird auf die ergänzte und aktualisierte Denkschrift "Sicherung guter wissenschaftlicher Praxis" der Deutschen Forschungsgemeinschaft (DFG) ausdrücklich hingewiesen (zur Datenhaltung: Empfehlung 7; Stand: 07/2013), Download unter

http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/index.html

ACHTUNG: Personendaten müssen grundsätzlich anonymisiert werden (auf keinen Fall Namen erfassen)! Unbedingt sind in diesem Zusammenhang die rechtlichen Vorgaben zu beachten - Hinweise dazu z. B. auf den Seiten der Datenschutzbeauftragten (NRW, Bund).

Quelle bzw. Download unter

<http://imsiweb.uni-koeln.de/beratung/daten/>