

## Infobogen

### Informationsbogen - Datencheckliste

Prinzipiell werden Daten, die im Rahmen der von Mitarbeitern des Instituts für Med. Informatik unterstützten Projekte anfallen, extern erfasst, bevor sie statistisch ausgewertet werden können. Diese Datenerfassung ist im Sinne des Beratenen grundsätzlich vorher mit dem zuständigen Berater abzuklären, um Fehler und dadurch häufig anfallende Mehrarbeit zu vermeiden. Im folgenden sind die wesentlichen Punkte zusammengefasst, die bei der Datenerfassung unbedingt beachtet werden müssen:

#### 1. Datenstruktur

Zur Auswertung müssen die Daten in Rechteckform als Datenmatrix (2-dim. Tabelle) vorliegen. Dabei sind die Zeilen der Tabelle die Untersuchungsobjekte bzw. Fälle (in der Praxis meist Patienten, Probanden, Versuchstiere usw.), die Spalten entsprechen den erhobenen Merkmalen bzw. Variablen. Mehrstufig, d. h. wiederholt gemessene Variablen stehen in entsprechend vielen Spalten hintereinander. Alle Daten eines Projektes müssen in einer Datenmatrix, d.h. in einer Tabelle stehen. Die Daten sollten nicht in mehreren Arbeitsblättern oder gar Dateien abgelegt werden. Die Datenmatrix d.h. die Tabelle darf nicht durch Leerzeichen, Leerspalten oder andere Markierungen (Trennlinien, Unterlegen von Farben usw.) strukturiert werden.

Beispiel für einen Auszug aus einer formal korrekten Datei:

Patnr	Geschl	Alter	Diagnose	Therapie	RRsys_1	RRsys_2	HF_1	HF_2
1	1	38	3	0	145	130	80	96
2	2	42	2	0	150	140	80	76
3	2	26	3	1	155	155	84	84

#### 2. Variablennamen

In der ersten Zeile der Datentabelle stehen grundsätzlich die Bezeichnungen der erfassten Merkmale, die sogenannten Variablennamen, die einigen Formalien genügen müssen:

- maximale Länge: sollte 10 Zeichen nicht überschreiten,
- 
- das erste Zeichen muß ein Buchstabe sein,
- 
- die weiteren maximal 9 Stellen können Buchstaben oder Zahlen oder ein "\_" (=
- Unterstrich) sein.
- Sonderzeichen, Umlaute, Punkte, "ß", Leerzeichen, etc. sind nicht erlaubt.
-

- Variablennamen müssen eindeutig sein, d.h. sie dürfen nicht mehrfach vorkommen

Die Bedeutung der 10 Zeichen muß in einer gesondert angelegten Liste, der sogenannten Schlüsselliste, erläutert werden (nicht in der gleichen Tabelle; für jede Variable gesondert).

### 3. Datenbereich

Ab der zweiten Zeile stehen in der Datentabelle nur noch Zahlen. Messungen oder Zählungen werden direkt numerisch realisiert, während qualitative Merkmale eindeutig numerisch codiert werden müssen. Diese Codes müssen, bei 0 oder 1 beginnend, ganzzahlig so gewählt werden, wie sie für die Auswertung später praktikabel sind. Ansonsten ist die Wahl der Codes beliebig. Die Bedeutung muß der Anwender in der Schlüsselliste bei der jeweiligen Variablen dokumentieren. In der auszuwertenden Datenmatrix steht also letzten Endes an jeder Stelle (i-te Zeile, j-te Spalte) genau eine eindeutige Zahl. Ausnahme: bei fehlenden Werten soll die entsprechende Stelle leer gelassen werden.

### 4. Hardware/Betriebssystem

Die Daten müssen anonymisiert in einem IBM-kompatiblen WINDOWS-Format vorliegen und können auf entsprechenden Datenträgern (USB-Stick, CD, Diskette) geliefert oder auch per E-Mail geschickt werden.

### 5. Dateiformat

Die Daten können zur statistischen Auswertung in verschiedenen Formaten geliefert werden, d.h. mit verschiedenen kommerziellen Softwareprodukte (Programmen) erstellt werden:

- EXCEL

Insbesondere das Programm EXCEL hat sich als integriertes System nicht nur zur Datenerfassung, sondern auch mit Möglichkeiten zur Erstellung von Graphiken und Tabellen kleineren deskriptiven Auswertungen, Berechnungen neuer Variablen und Datenmodifikation bewährt.

Falls in einer Datenmatrix neue Variablen berechnet werden müssen, sollte durch Kopieren der gesamten Datenmatrix in eine andere Tabelle dafür gesorgt werden, dass in dieser neuen Tabelle nur die Werte und nicht die Formeln enthalten sind.

Der Beratene sollte sich in das Programm einarbeiten.

- ACCESS

Auch mit dem Datenbanksystem ACCESS können auswertbare Dateien erstellt werden (Dateinamenerweiterung: \*.mdb). Dabei ist zu beachten, dass nur einfache Tabellen, nicht jedoch Relationen importiert werden können. Bei Fehlwerten sollte das Feld leer bleiben; es darf kein Standardwert gesetzt werden.

ACCESS-Daten müssen für Auswertungen mit SPSS in EXCEL-Dateien konvertiert werden.

- Natürlich können Daten auch direkt mit Statistik-Software (SPSS, SAS) erfasst werden.

### 6. Bemerkungen

- Die Durchführung der Datenerfassung ist allein Sache des Beratenen, der also mit

dem benutzen Programm und Betriebssystem vertraut sein muss. Er ist auch für die Richtigkeit und Vollständigkeit seiner Daten und Schlüsselliste verantwortlich. Exportierte bzw. importierte Daten können nur mit erheblichem Aufwand nachträglich verändert werden.

- Der Beratene ist ebenso zuständig für die Ausstattung seiner Arbeit mit Graphiken, Tabellen und Text, wobei der Berater Hinweise geben kann.
  - Von der Arbeitsgruppe Med. Statistik werden zur Zeit die kommerziellen Programmpakete SAS und SPSS unterstützt. Dabei können für SAS die Daten im EXCEL- oder ACCESS-Format geliefert werden.
  - Datensicherung ist Aufgabe des Beratenen.
  - Falls mit anderer als der oben erwähnten Software Daten erfasst werden, sollte sichergestellt sein, dass die erhobenen Daten in Standardsoftware wie EXCEL exportiert werden können.
  - Den Datenschutzrichtlinien entsprechend müssen die Daten in anonymisierter Form zur Verfügung gestellt werden.
-