

Vorbereitung der statistischen Datenauswertung

Nach der Datenerhebung folgen bei einer wissenschaftlichen Studie die *Kodierung* und *statistische Auswertung* der Daten. Die statistische Auswertung wird normalerweise mit einem Computerprogramm durchgeführt - dabei kann ich Ihnen gerne helfen¹. Zur Vorbereitung der Auswertung sollten Sie die Daten kodieren (siehe Abschn. 1 und 2) und die Fragestellungen und Hypothesen (nochmals) präzisieren (siehe Abschn. 3).

1 Kodierung der Daten in eine Rohdaten-Tabelle

Legen Sie in *Excel* (oder falls Sie kein Excel zur Verfügung haben in *Word*) eine **Tabelle** an. Diese Tabelle lesen wir dann mit dem Statistikprogramm ein, so dass wir die Daten auswerten können.

Beispiel einer Rohdaten-Tabelle

nr	geschl	alter	gruppe	nation	f1	f2	f3a	f3b	f3c	iq	rz	...
1	0	10	1	0	3	2			1	104	0,7	...
2	1	12	2	0	0	2	1	1		96	3,2	...
3	1	10	1	2	2					120	1,9	...
4	1	11	2	0	1	3	1	1	1	115	2,1	...
...

In der **ersten Zeile** dieser Tabelle stehen **Bezeichnungen für alle Variablen**, über die Daten vorliegen (z. B. Versuchspersonen-Nummer, Geschlecht, ..., Antworten in Fragebögen, Test- und Messwerte, ...). Diese Bezeichnungen sollten aus einer Zeichenfolge bestehen, die mit einem Buchstaben beginnt.

In den folgenden Zeilen (Zeile 2 ff.) gibt es **für jede Untersuchungseinheit** (d.h. in der Regel für jede Person) **eine Zeile**, in der deren Werte für die jeweiligen Variablen eingetragen werden. Ab hier (der 2. Zeile) dürfen ausschließlich Zahlen eingetragen werden. Dabei können drei Fälle unterschieden werden:

- Die Werte sind Zahlen (z. B. Alter, Anzahl der Fehler, Ergebnis eines Tests). Bei solchen Daten werden direkt die Zahlen eingetragen.
- Die Werte sind keine Zahlen, können aber durch eine Zahl ausgedrückt werden, da pro Person nur eine Antwort möglich ist. Beispiele: Die Variable *Geschlecht* hat die Werte *männlich* und *weiblich*, die in der Tabelle durch 0 und 1 kodiert werden können. Oder in einem Fragebogen wurden Fragen durch Kreuze auf einer Rating-Skala (trifft voll zu - trifft meistens zu - trifft manchmal zu - trifft gar nicht zu) beantwortet; die Antworten der Personen könnten dann durch die Zahlen 3 - 2 - 1 - 0 kodiert werden. Oder die Frage „Was ist deine Muttersprache?“ kann jeweils nur eine Antwort haben, wobei jede vorkommende Sprache durch eine Zahl ausgedrückt wird.
- Die Messungen/Fragen sind so angelegt, dass bei einer Person unterschiedlich viele Antworten vorliegen können (z. B. „Welche Sprachen sprichst du?“ oder Kategorisierungen von freien Antworttexten). Bei solchen Daten muss für jede Antwortmöglichkeit/Kategorie eine eigene Variablenspalte angelegt werden, in denen immer dann eine 1 eingetragen wird, wenn sie zutrifft (vgl. im Beispiel Frage 3, die durch die Variablen f3a, f3b und f3c kodiert wurde).

Bei **fehlenden Werten** wird die entsprechende **Zelle frei gelassen** (im Beispiel hat Person Nr. 3 die Frage 2 nicht beantwortet).

Bitte fügen Sie außer den Variablennamen und den Zahlen **keine Kommentare in die Tabelle** ein. Benutzen Sie in *Excel* keine Berechnungsfunktionen; sie können vom Statistikprogramm nicht eingelesen werden.

1. Wenn Sie sich detaillierter über die statistische Datenanalyse informieren wollen empfehlen wir: Bühl, A. & Zöfel, P. (2005). *SPSS 12. Einführung in die moderne Datenanalyse unter Windows*. München: Pearson Studium.

Eine **Formatierung der Tabelle** (Farben, Spaltenbreite) ist **nicht notwendig**, kann aber zur Vereinfachung der Eingabe für Sie hilfreich sein.

Die Tabelle kann beliebig viele Spalten (Variablen) und Zeilen (Personen) beinhalten. Kodieren Sie daher **alle Daten in einer Tabelle** und legen Sie nicht für bestimmte Variablen- oder Personengruppen separate Tabellen an. Untergruppen von Personen können durch eine entsprechende Variable gekennzeichnet werden.

Tragen Sie in die Tabelle möglichst **direkt die Rohdaten** ein. Alle Berechnungen können effektiv und fehlerfrei mit dem Statistikprogramm vorgenommen werden.

2 Benennung der Variablen und Variablen

Selbstverständlich sollten Sie sich notieren, was die Variablennamen und ggf. die Zahlen inhaltlich bedeuten. Damit die vom Statistikprogramm erzeugten Ausgaben (Tabellen, Graphiken) verständliche Beschriftungen haben, kann man dem Statistikprogramm entsprechende Benennungen „mitteilen“. Dies geschieht durch eine Programmanweisung mit einer festgelegten Grammatik. Hier ist ein Beispiel für die Programmanweisung zur Benennung von Variablen und ein Beispiel für die Benennung von Werten, das Sie auf Ihren Fall übertragen und in einer Text-Datei (z. B. in Word) aufschreiben können.

VARIABLE LABELS

```
nr          'Vp-Nummer'
geschl     'Geschlecht'
nation     'Nationalität'
f1         'Frage 1'
f2         'Frage 2'
f3a       'Frage 3: englisch'
f3b       'Frage 3: französisch'
f3c       'Frage 3: türkisch'
iq         'Intelligenzquotient'
rz         'Reaktionszeit' .
```

VALUE LABELS

```
geschl     0 'männlich'
           1 'weiblich'
/gruppe    1 'Lerngruppe'
           2 'Kontrollgruppe'
/nation    0 'deutsch'
           1 'italienisch'
           2 'türkisch' .
```

3 Fragestellungen und Hypothesen

Erstellen Sie zur Auswertung der Daten eine Liste, auf der möglichst präzise die Fragen oder Hypothesen aufgeschrieben sind, die Sie durch Ihre Studie beantworten bzw. überprüfen wollten. Dabei sollte immer klar sein, anhand welcher Variablen man dies tun kann: Zwischen welchen Variablen wird ein Zusammenhang erwartet? Welche Variablen sollten sich auf andere Variablen auswirken? Kann man aus bestimmten Variablengruppen eine andere (abstraktere) Variable ableiten bzw. berechnen?

4 Checkliste

Bitte bringen Sie zur Auswertung folgendes mit:

- Datentabelle (Excel-Datei auf Memory-Stick, CD oder vorher per Email schicken)
- Benennungs-Anweisungen (als Text-Datei auf Memory-Stick, CD oder vorher per Email schicken)
- Fragestellungen und Hypothesen (als Ausdruck)
- Untersuchungsplan
- Versuchsmaterialien, z. B. Fragebögen