

Tips zur Dateneingabe

Zusammengestellt von Erik Hölzl

1) *Fragebögen numerieren*. Jeder ausgefüllte Fragebogen wird mit einer Nummer versehen, damit man spätere Unklarheiten überprüfen kann (z.B. hat die Person mit der Nummer 97 wirklich angegeben, 50 Semester zu studieren, oder ist das doch ein Tippfehler?).

2) *Codeplan erstellen*. Auf einem Fragebogen werden Variablennamen und verwendete Codes/Ziffern notiert, um bei Unklarheiten eine Hilfe zu haben (etwa wenn man zwei Wochen Pause macht und sich nicht mehr erinnern kann, ob man bei Frage 4 "ja" als 1 oder 2 eingegeben hat).

3) *Variablennamen festlegen*. In SPSS kann ein Variablenname maximal 8 Zeichen umfassen. Allerdings ist es oft sinnvoll, am Anfang nur maximal 6 Zeichen auszunützen, da manchmal später Transformationen nötig sind, die den Namen verlängern. Beispielsweise besteht im Befehl Descriptives die Möglichkeit, standardisierte Werte als neue Variablen speichern zu lassen - dabei wird dem alten Variablennamen ein "z" vorangestellt. Eine Variable namens F_EINFL1 ergibt dann ZF_EINFL, und die Eins am Ende verschwindet. Was die Namen angeht, gibt es 2 Schulen. Die einen meinen, man sollte am Namen bereits den Inhalt der Variablen erkennen; z.B. ALTER, BERUF, etc. Dies hat den Nachteil, daß bei längeren Fragebatterien oft Phantasie gefordert ist, um doppelte Namen zu vermeiden. Die anderen numerieren einfach zusammengehörige Fragen durch - z.B. die Items des 16PA heißen dann PA_1 bis PA_32. Damit weiß man zwar nicht sofort, was PA_15 ist - aber dafür ist es wesentlich schneller und meist werden die Einzelitems ohnehin zu Skalen zusammengefaßt- z.B. NORM. Letztlich ist die Namensgebung Geschmackssache.

Üblicherweise entspricht jedem Item eine Variable. Wenn allerdings mehrere Antworten bei einem Item möglich sind, muß für jede Antwortkategorie eine eigene Variable vorgesehen werden (z.B. Hobby_1 bis Hobby_5, siehe Anhang).

Falls verschiedene Bedingungen vorliegen (z.B. eine Gruppe füllt den Fragebogen auf der Straße aus, die andere daheim; oder verschiedene Reihenfolgen in der Fragebogenbatterie), muß auch dafür eine Variable vorgesehen werden. Weiters ist es oft sinnvoll, eine Textvariable für Anmerkungen vorzusehen, oder eine Variable, in der "auffällige" Fragebögen gekennzeichnet werden.

4) *Codierungen festlegen*. Wird eine Ankreuzung "ja" als "1", "2", oder anders eingegeben? Auch hier ist die Wahl der Zahlen relativ willkürlich - sinnvoll ist jedoch, sie so zu wählen, daß innerhalb des Fragebogens Konsistenz herrscht (also nicht "ja" einmal als 1 und einmal als 2 eingegeben wird), und in Leserichtung des Fragebogens gearbeitet wird (die erste Alternative bekommt "1", die zweite "2", ...). Dies soll unnötige Eingabefehler vermeiden. Fehlende Angaben müssen in den meisten Fällen nicht extra gekennzeichnet werden - die entsprechende Stelle wird einfach frei gelassen. Bis auf freie Antworten, die eine Textvariable erfordern, sind alle Variablen "numerisch", d.h. sie enthalten Zahlen (nicht "J" für "ja!").

5) *Datenfile erstellen*. Dies kann entweder Variable für Variable im SPSS-Datenfenster gemacht werden, oder über ein Syntaxfile (siehe Anhang). Hierbei ist auch sinnvoll, "VARIABLE LABELS" und "VALUE LABELS" zu verwenden, da sie später die Lesbarkeit des Outputs wesentlich erhöhen.

6) *Daten eingeben*. Alles - bis auf offensichtliche Fehler - genau so eingeben, wie es am Fragebogen ist. Insbesondere nicht: a) schon vor der Eingabe umpolen, b) nur die Summenwerte einzelner Items eingeben, c) scheinbar überflüssige Daten nicht eingeben. Falls etwas anders eingegeben wird als am Fragebogen, dies am Fragebogen vermerken! Wesentlich schneller und

fehlerfreier ist man bei der Eingabe, wenn man sich Blindschreiben am Ziffernblock angewöhnt, oder sich die Daten ansagen läßt.

Wenn einem bei der Dateneingabe etwas auffällt (z.B. sehr viele fehlende Werte, offensichtliche Instruktionsverfehlung, nicht ernsthaft ausgefüllt), sollte der entsprechende Fragebogen nur mit äußerster Vorsicht verwendet werden, besser aber gar nicht. In einer Variable für "Anmerkungen" oder "Auffälligkeiten" kann man diese Fälle auch kennzeichnen, um sie später wiederzufinden.

Wenn die Dateneingabe abgeschlossen ist, wird das File unter dem Namen "ROHDATEN.SAV" gespeichert.

7. Datencheck. Auch wenn man am liebsten gleich losrechnen möchte - vorher sollte man einen gründlichen Datencheck durchführen. Nichts ärgerlicher, als wenn man am Ende der Auswertungen bemerkt, daß bei einer einzigen Person statt "9" "999" steht, und sich dadurch alle Mittelwerte ändern. Der Datencheck zielt vor allem darauf ab, Eingabefehler zu entdecken. Mit dem Befehl DESCRIPTIVES oder FREQUENCIES wird für alle Variablen überprüft, ob Minimum und Maximum innerhalb der zulässigen Grenzen liegen, ob "auffällige" Mittelwerte bestehen, oder etwa gar die Hälfte aller Fälle fehlende Werte enthält. Eine weitere Möglichkeit besteht darin, logische "Unmöglichkeiten" zu überprüfen (z.B. jemand hat bei Alter einen Wert von 12 und bei Bildung "Hochschule"). Dies ist am einfachsten über Kreuztabellen (CROSSTABS) möglich. Eingabefehler werden nach Vergleich mit den Fragebögen im Rohdatenfile korrigiert. Am Ende dieser Prozedur steht ein File mit den kontrollierten, NICHT umkodierten Rohdaten mit dem Namen "ROHDATEN.SAV". Dieses File kommt in der Diplomarbeit in den Anhang (Kleiner Tip: Dafür im SPSS-Datenfenster als Excel-File abspeichern, dann wird das Layoutieren leichter).

8) *Umkodieren, Transformationen, Skalenbildung*. Ausgehend vom Rohdatenfile werden jetzt Items umgepolt, neue Variablen gebildet, usw. und das entstehende File *unter einem neuen Namen* abgespeichert! (z.B. "DATEN_1.sav"). Dadurch läßt sich der Prozeß jederzeit nachvollziehen, und man kommt nicht in Verlegenheit, ob ein bestimmtes Items eigentlich jetzt schon umgepolt ist oder nicht. Weiters empfiehlt sich eine Protokollierung aller Transformationsschritte - erstens über Syntax, zweitens in Form eines Auswertungstagebuchs.

Umpolungen sollten so erfolgen, daß ein hoher Wert die stärkere Ausprägung des gemessenen Konstruktes anzeigt. Bei einem Zufriedenheitsitem sollte eine Antwort, die Zufriedenheit signalisiert, auch einen hohen Zahlenwert bekommen - das erleichtert später die Lesbarkeit. (SPSS-Befehl RECODE)

Bevor man einzelne Items zu einer Skala zusammenfaßt, sollte man - sofern es sich nicht um altbewährte, standardisierte Fragebögen handelt - die Reliabilität überprüfen und eine Faktorenanalyse durchführen. Bei der Skalenbildung empfiehlt sich, den *Mittelwert*, nicht die *Summe* der Items zu bilden - damit bleibt der Skalenwert innerhalb der Grenzen der Items und wird leichter interpretierbar (z.B. 1 "unzufrieden" bis 5 "zufrieden"; SPSS-Befehl MEANS).

Weitere Hinweise findet man auch in Kapitel 8: Syntax bei Bühl, Achim & Zöfel, Peter. (1996). Professionelle Datenanalyse mit SPSS für Windows. Bonn: Addison-Wesley.

Beispiel Codeplan

Für mich persönlich ist das Institut für Psychologie:

wichtig	1	2	3	4	5	6	7	unwichtig
irrelevant	1	2	3	4	5	6	7	relevant
wertvoll	1	2	3	4	5	6	7	wertlos
interessant	1	2	3	4	5	6	7	uninteressant
unbedeutend	1	2	3	4	5	6	7	wesentlich

wie angekreuzt

	trifft gar	trifft wenig	teils	trifft ziemlich	trifft völlig
	nicht	zu	zu	zu	zu

Das Institut für Psychologie ist dienstleistungsorientiert

Das Lehrpersonal ist kompetent.....

Der organisatorische Ablauf am Institut ist chaotisch.....

Das Lehrpersonal engagiert sich wenig für Anliegen der Studierenden.....

Alter: _____ Jahre 1 2 1 2 3 4 5

Geschlecht: weiblich männlich

Hobbies: Sport
 Reisen
 Musik
 Bücher
 Film
 Anderes: _____

1, wenn angekreuzt

NR

I_1 *
I_2
I_3 *
I_4 *
I_5

Eingabemuster

O_1
P_1
O_2 *
P_2 *

Alter
Geschl
Hobby_1
Hobby_2
Hobby_3
Hobby_4
Hobby_5
Hobby_6
Hobby_A

Variablen-
Bezeichnung

* kennzeichnet jene Variablen, die später umcodiert werden

```

*Syntax zur Erstellung des Beispiel-Rohdatenfiles.
*Zeilen mit einem Stern am Anfang sind Anmerksungszeilen, keine Befehlszeilen.
*Jede Anmerksungszeile und jeder Befehl endet mit einem Punkt.

* Neues Datenfile erstellen, erste Zeile aktivieren.
new file.
data list/NR (F3).
Variable Label NR "Fragebogennummer".
begin data.
1
end data.

*Ab hier kommen die Definitionen der Variablen laut Codeplan.
Numeric I_1 to I_5 (F1).
Variable Label I_1 "unwichtig (-)" /
I_2 "relevant" /
I_3 "wertlos (-)" /
I_4 "uninteressant (-)" /
I_5 "wesentlich".
*Anmerkung: später umcodiert werden I_1, I_3, I_4.
Numeric O_1 P_1 O_2 P_2 (F1).
Variable Label O_1 "Organisation: dienstleistungsorientiert" /
P_1 "Personal: kompetent" /
O_2 " Organisation: chaotisch (-)" /
P_2 " Personal: wenig engagiert (-)".
*Anmerkung: später umcodiert werden O_2, P_2.
Numeric ALTER (F2).
Variable Label Alter "Alter in Jahren ".
Numeric GESCHL (F1).
Variable Label GESCHL "Geschlecht".
Value Label GESCHL 1 "weiblich" 2 "männlich".
Numeric HOBBY_1 to HOBBY_6 (F1).
Variable Label HOBBY_1 "Sport" /
HOBBY_2 "Reisen" /
HOBBY_3 "Musik" /
HOBBY_4 "Bücher" /
HOBBY_5 "Film" /
HOBBY_6 "Anderes Hobby".
Value Label HOBBY_1 to HOBBY_6 1 "angekreuzt".
String HOBBY_A (A25).
Variable Label HOBBY_A "Anderes Hobby: Was".
execute.

*Datenfile speichern.
SAVE OUTFILE = "C:\ROHDATEN.SAV".
    
```

Beispiel Syntax