Auswertungsbeispiele: Zahl der Kinder unter 18 Jahren in der Familie

Oft benötigt man Angaben zu einer Einheit von Personen, etwa zum Haushalt, zur Familie oder zur Lebensgemeinschaft. Ein Beispiel ist die Zahl der Kinder unter 18 Jahren in der Familie.
In Form so genannter Bandsatzerweiterungen sind solche Variablen häufig bereits in den Daten des Mikrozensus enthalten. Sie werden vom Statistischen Bundesamt generiert. Hier wird nun gezeigt, wie solche Variablen gebildet werden, um entweder die im Datensatz bereits enthaltenen Variablen zu rekonstruieren und so eventuell vorhandene unplausible Angaben zu korrigieren, oder um eigene Variablen zu generieren.

Die Zahl der Kinder unter 18 Jahren in der Familie kann folgendermaßen ermittelt und als Variable allen Personen der jeweiligen Familie zugespielt werden (hier am Beispiel des Mikrozensus 2003):

Statistik-Programm-Routinen

SPSS	Stata

SELECT IF EF508=1.
 	
COMPUTE x=0.
IF (EF509=3 AND EF30<18) x=1.
 	
SORT CASES BY EF3 EF4 EF28.
AGGREGATE OUTFILE=* MODE=ADDVARIABLES
  / PRESORTED
  / BREAK EF3 EF4 EF28
  / V585=SUM(x).
EXECUTE.
	/* Es gibt in STATA mehrere Möglichkeiten die Konstruktion von Bandsatzerweiterungen umzusetzen.*/
	Option 1: egen
	egen v585 = total(ef508==1 & ef509==3 & ef30<18), by(ef3 ef4 ef28)
	Option 2: merge
	sort ef3 ef4 ef28
	save Laufwerk:\<Padangabe>\filename1.dta, replace
	
	sort ef3 ef4 ef28
	  generate v585_merge1 = ef508==1 & ef509==3 & ef30<18
	  collapse(sum)v585_merge1, by(ef3 ef4 ef28)
	  sort ef3 ef4 ef28
	  save Laufwerk:\<Padangabe>\filename2.dta, replace
	use Laufwerk:\<Padangabe>\filename1.dta, clear
	merge ef3 ef4 ef28 using ///
	Laufwerk:\<Padangabe>\filename2.dta, uniqusing

Erläuterungen

SPSS	Stata
Zunächst erfolgt eine Einschränkung der interessierenden Population auf die Bevölkerung am Familienwohnsitz (EF508=1). Eine Hilfsvariable x wird gebildet, die für ledige Kinder in der Familie (EF509=3) unter 18 Jahren (EF30<18) den Wert 1 erhält, ansonsten den Wert 0. Dann wird diese Hilfsvariable x auf Familienebene (EF3, EF4, EF28) aggregiert und als neue Variable (V585) allen Personen der jeweiligen Familie zugewiesen.	**Option 1: egen** Mit *egen* wird die Variable v585 (Anzahl Kinder unter 18 Jahre in der Familie) gebildet, welche die bereits vorhandene Bandsatzerweiterung ef585 replizieren soll. Die Funktion *total* bewirkt, dass die Summe der ledigen Kinder in der Familie (EF509==3) unter 18 Jahren (ef30<18) am Familienwohnsitz (ef508==1) gebildet wird. Die Sortierung des Datensatzes nach den zur Abgrenzung von Familien notwendigen Variablen (ef3, ef4, ef28) erfolgt in den Optionen nach dem Komma. **Option 2: merge** Zunächst wird der Datensatz nach den zur Abgrenzung von Haushalten notwendigen Variablen (ef3, ef4, ef28) sortiert und anschließend gespeichert. Daraufhin wird die Variable v585 (Anzahl Kinder unter 18 Jahre in der Familie) generiert, welche die bereits vorhandene Bandsatzerweiterung ef585 replizieren soll. Über diese Variable erfolgt die Abgrenzung auf Individualebene, d.h. auf die Bevölkerung am Familienwohnsitz (nach dem

	traditionellen Familienkonzept)(ef508==1) und ledige Kinder in der Familie (ef509==3) unter 18 Jahren (ef30<18). Die Aggregation der Beobachtungen auf Haushaltsebene erfolgt mittels *collapse*. Anschließend wird erneut sortiert und gespeichert. Das Individualdatenfile (filename1.dta) wird nun eingelesen und mittels *merge* mit dem Aggregatdatenfile (filename2.dta) zusammengeführt.

Zur Kontrolle kann nun die neu gebildete Variable V585 mit der im Datensatz bereits enthaltenen Variable EF585 abgeglichen werden:

SPSS	Stata
CROSSTABS / TABLES=V585 BY EF585.	tabulate v585 ef585 if ef508==1

Weitere Erläuterungen finden sich in:

John, Kristina (2008): Rekonstruktion von abgeleiteten Variablen im Mikrozensus 2005 im Vergleich zu den Mikrozensen bis 2004. GESIS Methodenbericht 2008/7.
John, Kristina (2007): Rekonstruktion von Bandsatzerweiterungen mittels zeilen- bzw. satzübergreifender Operation in STATA im Mikrozensus. Gesis-Methodenberichte 3/2007.
John, Kristina (2006): Rekonstruktion von Bandsatzerweiterungen mittels zeilen- bzw. satzübergreifender Operationen in STATA.
Lengerer, Andrea; Boehle, Mara (2006): Rekonstruktion von Bandsatzerweiterungen zu Haushalt, Familie und Lebensformen im Mikrozensus. ZUMA-Methodenbericht 2006/05.