Auswertungsbeispiele: Generierung einer fortlaufenden Nummer der Person im Haushalt

In den verschiedenen Erhebungen des Mikrozensus ist bereits eine Nummer der Person im Haushalt enthalten. Allerdings ist diese Nummer nicht immer fortlaufend. Sofern gedoppelte Fälle in den Daten vorhanden sind (Mikrozensen vor 1989), haben diese häufig eine eigene Nummer. Mit dem Ausschluss gedoppelter Fälle entstehen Lücken, so dass auch dann die Nummer der Person im Haushalt nicht mehr fortlaufend ist. Nicht zuletzt gibt es einige wenige Fälle mit unplausiblen Angaben, d.h. zum selben Haushalt gehörende Personen mit identischer Nummer.

Eine fortlaufende Nummer der Person im Haushalt kann folgendermaßen gebildet werden (hier am Beispiel des Mikrozensus 1995):

Statistik-Programm-Routinen

SPSS	Stata
SORT CASES BY EF2. COMPUTE pnr=1. DO IF EF2=LAG(EF2). COMPUTE pnr=LAG(pnr)+1. END IF. EXECUTE.	by ef2, sort: generate pnr=_n

Erläuterungen

SPSS	Stata
Zunächst wird der Datensatz nach der Variable sortiert, die zur Abgrenzung von Haushalten notwendig ist (EF2). Eine Variable pnr wird gebildet und auf den Wert 1 gesetzt. Über den zeilenweisen Abgleich der zur Abgrenzung von Haushalten notwendigen Variable wird sodann eine fortlaufende Nummer generiert. Zur Variable tnr wird der Wert 1 addiert, wenn sich die Ausprägung der zur Abgrenzung von Haushalten notwendigen Variable nicht verändert.	Zunächst wird der Datensatz nach der Variable sortiert, die zur Abgrenzung von Haushalten notwendig ist (ef2). Eine Variable pnr wird gebildet, die der Systemvariablen *_n* entspricht, welche zusammen mit *by* die Position einer Beobachtung innerhalb der *by*-Gruppe angibt (d.h. hier innerhalb des Haushalts).