

Auswertungsbeispiele: Generierung einer fortlaufenden Haushaltsnummer

Zur Identifikation von Privathaushalten sind im Mikrozensus häufig mehrere Variablen notwendig. In der Regel sind dies das Bundesland, die Nummer des Auswahlbezirks sowie die Nummer des Haushalts im Auswahlbezirk. In anderen Mikrozensusen ist die Abgrenzung anhand einer Variable ausreichend.

Auf der Grundlage dieser Angaben lässt sich eine fortlaufende Haushaltsnummer generieren. Dies hat den Vorteil, dass sich Haushalte über eine einzige Variable identifizieren und voneinander abgrenzen lassen. Außerdem ist das Maximum einer solchen Variable gleichbedeutend mit der Gesamtzahl der Haushalte, die im Datensatz enthalten sind.

Wie eine fortlaufende Haushaltsnummer gebildet werden kann, wird im Folgenden für ausgewählte Erhebungsjahre des Mikrozensus gezeigt:

Statistik-Programm-Routinen

Mikrozensus 2005

SPSS	Stata
<pre> SORT CASES BY EF1 EF3 EF4. IF \$CASENUM=1 hhnr=1. IF EF1 NE LAG(EF1) OR EF3 NE LAG(EF3) OR EF4 NE LAG(EF4) hhnr=hhnr+1. LEAVE hhnr. EXECUTE.</pre>	<pre> egen hhnr=group (EF1 EF3 EF4)</pre>

Mikrozensus 1991

SPSS	Stata
<pre> SORT CASES BY O402. IF \$CASENUM=1 hhnr=1. IF O402 NE LAG(O402) hhnr=hhnr+1. LEAVE hhnr. EXECUTE.</pre>	<pre> egen hhnr=group (o402)</pre>

Mikrozensus 1976

SPSS	Stata
<pre> SORT CASES BY EF1 EF5 EF203. IF \$CASENUM=1 hhnr=1. IF (EF1 NE LAG(EF1) OR EF5 NE LAG(EF5) OR EF203 NE LAG(EF203)) hhnr=hhnr+1. LEAVE hhnr. EXECUTE.</pre>	<pre> egen hhnr=group (ef1 ef5 ef203)</pre>

Erläuterungen

SPSS	Stata
<p>Zunächst wird der Datensatz nach den Variablen sortiert, die zur Abgrenzung von Haushalten notwendig sind. Über den zeilenweisen Abgleich dieser Variablen wird sodann eine fortlaufende Nummer generiert. Zunächst wird die neu zu bildende Variable hhnr in der ersten Zeile des Datensatzes auf den Wert 1 gesetzt. Ein Wert von 1 wird hinzuaddiert, wenn sich die Ausprägung einer der zur Abgrenzung von Haushalten notwendigen Variablen ändert. Ist dies nicht der Fall, bewirkt der LEAVE-Befehl, dass die Ausprägung der Variable hhnr in die jeweils nächste Zeile des Datensatzes unverändert übernommen wird.</p>	<p>Zunächst müssen die Variablen identifiziert werden, die zur Abgrenzung der Haushalte notwendig sind. Daraufhin wird eine Variable (hhnr) erstellt, die für die eindeutig mit diesen Variablen identifizierten Gruppen eine laufende Nummer erzeugt. Jede Gruppe (Haushalt) hat eine einmalige fortlaufende Nummer, jede Beobachtung eines Haushaltes hat dieselbe Haushaltsnummer.</p> <p>Das Maximum der Variable hhnr gibt nun die Gesamtzahl der Haushalte wieder, die im Datensatz enthalten sind.</p> <ul style="list-style-type: none"> - Im Mikrozensus 2005 sind dies 226.116 Haushalte. - Im Mikrozensus 1991 sind dies 224.735 Haushalte.

Das Maximum der Variable hhnr gibt nun die Gesamtzahl der Haushalte wieder, die im Datensatz enthalten sind. – Im Mikrozensus 1976 sind dies 161.020 Haushalte.

- Im Mikrozensus 2005 sind dies 226.116 Haushalte.
- Im Mikrozensus 1991 sind dies 224.735 Haushalte.
- Im Mikrozensus 1976 sind dies 161.020 Haushalte.

Ermitteln lässt sich die Gesamtzahl der Haushalte über:

SPSS

FREQUENCIES hhnr / FORMAT=NOTABLE / STAT=MAX.

Stata

summarize hhnr