

## **Informationen zur Nutzung des SAS-Setups für das Mikrozensus Scientific Use File 1996**

Der vorliegende Text soll Ihnen dabei helfen, den Rohdatensatz des Statistischen Bundesamtes, den Sie im ASCII-Format erhalten haben, korrekt in SAS einzulesen. Dazu wird vom German Microdata Lab (GML) des Zentrums für Umfragen, Methoden und Analysen (ZUMA) ein Setup (*setup96.sas*) bereitgestellt.

Ausführliche Informationen über den Mikrozensus und die Mikrozensus Scientific Use Files finden Sie auf den Websites des Zentrums für Umfragen, Methoden und Analysen (ZUMA) unter:

<http://www.gesis.org/Dauerbeobachtung/GML/index.htm>

Weitere Informationen sind auf den Seiten des Statistischen Bundesamtes abrufbar:

[http://www.destatis.de/themen/d/thm\\_mikrozen.php](http://www.destatis.de/themen/d/thm_mikrozen.php)

Das vom GML bereitgestellte SAS-Setup für das Mikrozensus Scientific Use File 1996 (*setup96.sas*) dient zum Einlesen des Rohdatenmaterials und zum Erstellen eines SAS-Systemfiles. Es beinhaltet Programmanweisungen zum Ersetzen fehlender Werte sowie zum Versehen der Variablen und ihrer Ausprägungen mit entsprechenden Labels.

Das Setup gliedert sich in verschiedene Bereiche, wobei Folgendes zu beachten ist:

Zunächst werden Formate definiert, um diese später den Ausprägungen der Variablen zuzuweisen. Sollen diese Formate dauerhaft verwendet werden, müssen sie in einer separaten Datei gespeichert werden. Das entsprechende Verzeichnis ist dann mit dem speziell dafür vorgesehenen Bibliotheksnamen LIBRARY zu referenzieren. Um in späteren SAS-Sitzungen auf die Formate zugreifen zu können, muss dem Verzeichnis, in dem die Formate abgespeichert sind, wiederum der Name LIBRARY zugewiesen werden.

Die fehlenden Werte, d.h. die Leerstellen im Rohdatenfile, werden über IF-Anweisungen durch gültige Werte ersetzt und später über die Format-Anweisung mit Labels versehen. Weil SAS keine benutzerdefinierten Missings kennt, müssen die Werte für spätere Auswertungen gegebenenfalls manuell auf Missing zurückgesetzt werden. Optional können die IF-Anweisungen im Setup auch auskommentiert werden. Dann werden die fehlenden Werte allerdings auch nicht gelabelt.

Das Rohdatenfile des StBa enthält alle Variablen des Mikrozensus 1996. Damit das SAS-Setup für jede mögliche Variablen-Auswahl verwendbar ist, wurden alle Variablen erfasst. Variablen, die Sie nicht bestellt haben, sind in Ihrem Rohdatenfile auf 0 oder Leerzeichen gesetzt. Um ein Systemfile zu erzeugen, das nur die von Ihnen bestellten Variablen enthält, fügen Sie im Setup am Ende des DATA-step die Anweisung KEEP ein und nennen dort Ihre bestellten Variablen.

Um eine reibungsfreie Aufbereitung der Daten zu gewährleisten, empfiehlt es sich, die in dem Setup vorgesehenen Voreinstellungen zu berücksichtigen. Insbesondere die Anweisung COMPRESS=YES im DATA-step bewirkt, dass die erzeugte Datei deutlich reduzierten Speicherplatz benötigt.

Sind die Daten eingelesen und ist das entsprechende Systemfile erstellt, kann die folgende Fallzahl (ohne Gewichtung, ohne Selektion) zur Kontrolle, ob der Rohdatensatz fehlerfrei eingelesen wurde, mit der des eingelesenen Datensatzes verglichen werden. Unterscheiden sich die Fallzahlen, weist dies auf einen Fehler beim Einlesen hin.

*Fallzahl Mikrozensus 1996 (ohne Gewichtung, ohne Selektion) = 509.243*

Des Weiteren können die Verteilungen ausgewählter Variablen des Mikrozensus SUF 1996 zur Prüfung des fehlerfreien Ablaufs des Setups herangezogen werden (vgl.: [http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz\\_1996/randverteilungen\\_mz96.htm](http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_1996/randverteilungen_mz96.htm)).

Als Referenz zur Plausibilitätsprüfung der Mikrozensus Scientific Use Files dienen die in den Fachserien des Statistischen Bundesamtes veröffentlichten Ergebnisse des jeweiligen Mikrozensus. Für den Mikrozensus 1996 sind dies die Reihe 3 (Haushalte und Familien), die Reihe 4.1.1 (Stand und Entwicklung der Erwerbstätigkeit) und die Reihe 4.1.2 (Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen) der Fachserie 1.

Die Plausibilitätsprüfung des Mikrozensus SUF 1996 erfolgte durch den Vergleich des aufbereiteten Datensatzes mit den vom Statistischen Bundesamt veröffentlichten Tabellen in den entsprechenden Reihen der Fachserie 1. Dabei wurde nach den vom Statistischen Bundesamt verwendeten Bevölkerungs- und Erwerbskonzepten gewichtet und selektiert. (zur Abgrenzung und Hochrechnung der Bevölkerungsbegriffe im Mikrozensus 1996 vgl.: [http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz\\_1996/hochrechnungsfaktoren\\_mz96.htm](http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_1996/hochrechnungsfaktoren_mz96.htm)).

Bedingt durch die Ziehung der 70%-Substichprobe ergeben sich Abweichungen zwischen den Häufigkeiten der Variablen des Mikrozensus Scientific Use Files und den in den Fachserien des Statistischen Bundesamtes veröffentlichten Zahlen (beruhend auf den Original-Mikrozensusdaten). Die meisten Variablen weichen nur in geringem Maße (meist 0% bis 1%, max. 5%) von den veröffentlichten Daten ab. Abweichungen über 5% wurden bei Variablen festgestellt, deren Merkmalsausprägungen mit sehr geringen Fallzahlen besetzt sind (z.B. Staatsangehörigkeit und Wirtschaftsbereiche).

Weiterhin wurden verschiedene generierte Merkmale (sog. Bandsatzergänzungen) des Mikrozensus 1996 auf ihre Plausibilität hin überprüft. In der Variable EF569 "Bevölkerungsgruppe der Bezugsperson" tritt im Rohdatenfile die Kategorie "0" mit 102 Fällen auf. Diese 102 Fälle sind Personen in Gemeinschaftsunterkünften, die der Missingkategorie "leer" zugehören. Das gleiche Problem tritt bei den Variablen EF538 und EF558 auf. Mit dem Setup werden die 102 Fälle jeweils automatisch mit der Kategorie "Entfällt (Personen in Gemeinschaftsunterkünften)" zusammengefasst. Problematisch

wirkt sich die Fehlcodierung in der Variable EF569 aus. Sie ist Grundlage für die Erstellung weiterer Bandsatzerweiterungen. In den Variablen EF528-EF531 "Anzahl der Kinder von ... bis ... Jahren im Haushalt" und EF576-581 "Anzahl der Kinder von ... bis ... Jahren in der Familie" wurden die 102 Fälle, die als Personen in Gemeinschaftsunterkünften zu identifizieren sind, jeweils der Kategorie "0 = Keine Kinder im Alter von ... bis ... Jahren" zugeordnet. Am Ende des Setups befindet sich eine Befehlssyntax, die dieses Problem behebt und die gegebenenfalls aktiviert werden kann.

Bei der Kontrolle der Variable EF688 (Beruf frühere Erwerbstätigkeit nach ISCO88 (COM)) fiel auf, dass sich in der Missingkategorie 11146 Fälle befinden bei denen nicht klar ersichtlich ist, warum ihnen für die ISCO88 (COM) keine Ausprägung zugeschrieben wurde. Denn diese Fälle haben sowohl bei der Angabe zur früheren Erwerbstätigkeit sowie des Berufs der früheren Erwerbstätigkeit nach KIdB (EF114) gültige Werte.

---

Zentrum für Umfragen, Methoden und Analysen (ZUMA)  
German Microdata Lab  
B2,1  
68159 Mannheim  
Tel: 0621-1246-265  
Fax: 0621-1246-100  
<http://www.gesis.org/Dauerbeobachtung/GML/index.htm>  
Kontakt: [mikrodaten@zuma-mannheim.de](mailto:mikrodaten@zuma-mannheim.de)

---