

## Informationen zur Nutzung des Setups des Mikrozensus Scientific Use Files 2002

Der vorliegende Text soll Ihnen dabei helfen, den Rohdatensatz des Statistischen Bundesamtes, den Sie im ASCII-Format zusammen mit dem SPSS-Setup (*setup02.sps*) erhalten haben, korrekt einzulesen. Zudem wird auf „Besonderheiten“ des Mikrozensus Scientific Use File (SUF) 2002 hingewiesen.

Ausführliche Informationen über den Mikrozensus und die Mikrozensus Scientific Use Files werden auf den Websites von GESIS bereitgestellt:

<http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/grundfile/>

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 German Microdata Lab (GML) in GESIS bietet für den Mikrozensus 2002 zahlreiche Informationen - den Erhebungsbogen, das Schlüsselverzeichnis, eine thematische Übersicht, relevante Klassifikationen u.v.m. - an. Auf der folgenden Website finden Sie zudem einen ZUMA-Methodenbericht zur Dokumentation und Aufbereitung des Mikrozensus SUF 2002:

<http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/rundfile/mz2002/>.

Das vom GML bereitgestellte SPSS-Setup für das Mikrozensus Scientific Use File 2002 (*setup02.sps*) dient zum Einlesen des Rohdatenmaterials und zum Erstellen eines SPSS-Systemfiles. Es beinhaltet Programmanweisungen zum Umwandeln von nicht-numerischen Variablen in numerische Variablen, zur Definition von fehlenden Werten sowie zum Versehen der Variablen und ihrer Ausprägungen mit entsprechenden Labeln.

Um eine reibungsfreie Aufbereitung der Daten zu gewährleisten, empfiehlt es sich, die in dem Setup vorgesehenen Voreinstellungen zu berücksichtigen und je nach Leistung des verarbeitenden Systems das Setup schrittweise durchführen zu lassen. 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 2002 (ohne Gewichtung, ohne Selektion) = **503.075***

Des Weiteren können die Verteilungen ausgewählter Variablen des Mikrozensus SUF 2002 zur Prüfung des fehlerfreien Ablaufs des Setups herangezogen werden (vgl.:

<http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/grundfile/mz2002/randverteilungen/>).

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 2002 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 2002 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 2002 siehe: <http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/grundfile/mz2002/abgrenzungen-hochrechnung/>).

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 Prozent 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 2002 auf ihre Plausibilität hin überprüft. Dabei zeigten sich einige Unplausibilitäten. So ergeben sich im Mikrozensus SUF 2002 bei der Gegenüberstellung der Variablen EF538 (Anzahl der Familien im Haushalt) und EF509 (Stellung innerhalb der Familie) in drei Haushalten mehr Familienbezugspersonen als vorhandene Familien. Es empfiehlt sich, diese Fälle bei Bedarf von den Analysen auszuschließen.

Im Mikrozensus SUF 2002 sind darüber hinaus Familienbezugspersonen enthalten, die unter 15 Jahre alt sind. Auch diese sollten bei Analysen mit entsprechenden Bandsatzergänzungen gegebenenfalls ausgeschlossen werden.

---

GESIS - Leibniz-Institut für Sozialwissenschaften  
German Microdata Lab  
B2,1  
68159 Mannheim  
Tel: 0621-1246-265  
Fax: 0621-1246-100

<http://www.gesis.org/das-institut/wissenschaftliche-arbeitsbereiche/dauerbeobachtung-der-gesellschaft/german-microdata-lab/>

Kontakt: Andrea Lengerer, Julia H. Schroedter, Hossein Shahla (GML, Mikrozensus Grundfiles)  
Email: [gml@gesis.org](mailto:gml@gesis.org)

---