

Informationen zur Nutzung des Stata-Setups für das Mikrozensus Scientific Use File 2002

Der vorliegende Text dient dazu, den Rohdatensatz des Statistischen Bundesamtes, den Sie im ASCII-Format erhalten haben, korrekt in Stata einzulesen.

Dazu wird vom German Microdata Lab (GML) der GESIS ein Setup (*setup02.do*; *setup02.dct*) bereitgestellt.

Ausführliche Informationen über den Mikrozensus und die Mikrozensus Scientific Use Files finden Sie auf den Websites der GESIS unter:

<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

Das vom GML bereitgestellte Stata-Setup für das Mikrozensus Scientific Use File 2002 (*setup02.do*; *setup02.dct*) dient zum Einlesen des Rohdatenmaterials und zum Erstellen eines Stata-Systemfiles. Es beinhaltet Programmanweisungen zum Ersetzen fehlender Werte, zur Umwandlung alphanumerischer Variablen in numerische sowie zum Versehen der Variablen und ihrer Ausprägungen mit entsprechenden Labels. Das Stata-Setup beruht auf dem SPSS-Setup und wurde mit Hilfe des Programms StatTrans automatisch erstellt. Dadurch sind die Variable und Value Labels gekürzt bzw. abgeschnitten. Vollständige Labellisten finden sich auf den GML-Websites.

Das Rohdatenfile des Statistischen Bundesamtes enthält alle Variablen des Mikrozensus 2002. Damit das Stata-Setup für jede mögliche Variablen-Auswahl verwendbar sind, wurden alle Variablen erfasst. Variablen, die Sie nicht bestellt haben, sind in Ihrem Rohdatenfile auf 0 oder BLANK gesetzt.

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

Bevor das Stata-Setup fehlerfrei durchgeführt werden kann, ist es nötig, zunächst den Rohdatensatz über einen Texteditor zu öffnen und dort die Dezimalstellenkommata durch Dezimalpunkte automatisch zu ersetzen.

Danach müssen im Stata-Setup (*setup02.do*) noch die genauen Angaben über den Pfad zum Rohdatenfile, zum zu erstellenden Systemfile (DTA-File) und zum Dictionary-File (*setup02.dct*) in die entsprechenden Stellen am Anfang des Setups (*setup02.do*) eingefügt werden.

Beim Einlesen des Rohdatenfiles mit dem *setup02.do* wird auf den Dictionary-File *setup02.dct* zugegriffen, in dem angegeben ist, wie viele Stellen eine Zeile (d.h. eine Beobachtung) im Rohdatenfile umfasst. Die Variable EF643 ist im Rohdatenfile alphanumerisch und wird in einem nachfolgenden Schritt in eine numerische Variable umgewandelt.

Die fehlenden Werte, d.h. die Leerstellen im Rohdatenfile, werden über Recode-Befehle durch gültige Werte ersetzt und später mit Labels versehen. Weil Stata keine benutzerdefinierten Missings kennt, müssen die Werte für spätere Auswertungen gegebenenfalls manuell auf Systemmissing zurückgesetzt werden.

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 vgl.: <http://www.gesis.org/dienstleistungen/daten/amtliche-mikrodaten/mikrozensus/grundfile/mz2005/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
