

<p style="text-align: center;">Wichtige Informationen zur Nutzung der Setups für das Mikrozensus Scientific Use File 1978</p>
--

Der vorliegende Text soll Ihnen dabei helfen, den Rohdatensatz des Statistischen Bundesamtes, den Sie im ASCII-Format erhalten haben, korrekt einzulesen. Dazu werden vom German Microdata Lab (GML) der GESIS-ZUMA die Setups der Statistikprogramme SPSS, SAS und STATA bereitgestellt. Sie enthalten Programmanweisungen zur Definition von fehlenden Werten sowie zum Labeln der Variablen und ihrer Ausprägungen.

Das Rohdatenfile des Statistischen Bundesamtes, das Sie erhalten, enthält nur die von Ihnen bestellten Variablen des Mikrozensus SUF1978. Variablen, die Sie nicht bestellt haben, sind in Ihrem Rohdatenfile auf 0 oder BLANK gesetzt. Damit das Setup für jede mögliche Auswahl an Variablen verwendbar ist, wurden jedoch darin alle Variablen erfasst. Um ein Systemfile mit den von Ihnen bestellten Variablen zu erzeugen, nennen Sie diese Variablen an der Stelle des Setups, an der die eingelesenen Daten abgespeichert werden, mittels des „KEEP“-Befehls.

Spezifika der Statistikprogramme

In SPSS werden fehlende Werte als benutzerdefinierte Missings deklariert, denen ein Wert und ein Label zugewiesen werden kann. Sie zählen bei Analysen nicht als gültige Werte.

In SAS werden zunächst 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. Beim Einlesen des Rohdatenfiles ist in der Option LRECL der INFILE-Anweisung angegeben, wie viele Stellen eine Zeile (d.h. eine Beobachtung) im Rohdatenfile umfasst. 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.

Damit das Stata-Setup durchgeführt werden kann, ist es nötig im Do-File (*setup78.do*) noch die genauen Angaben über den Pfad zum Rohdatenfile und zum zu erstellenden Systemfile (DTA-File) in die entsprechenden Stellen am Anfang des Setups (*setup78.do*) einzufügen. 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.

Um eine reibungsfreie Aufbereitung der Daten zu gewährleisten, empfiehlt es sich, die in den jeweiligen Setups vorgesehenen Voreinstellungen zu berücksichtigen und das Setup – je nach Leistung des verarbeitenden Systems – schrittweise durchführen zu lassen. Ist das entsprechende

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

Fallzahl Mikrozensus 1978 (ohne Gewichtung, ohne Selektion) = 441.563

Des Weiteren können die Verteilungen ausgewählter Variablen des Mikrozensus SUF 1978 zur Prüfung des fehlerfreien Ablaufs des Setups herangezogen werden (vgl.: http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_1978/randverteilungen_mz78.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 1978 sind dies 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 1978 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 1978 vgl.: http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_1978/hochrechnungsfaktoren_mz78.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 (zwischen 0 bis 5 Prozent) 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 1978 auf ihre Plausibilität hin überprüft. Dabei zeigten sich keine Inkonsistenzen.

GESIS-ZUMA
German Microdata Lab
B2,1
68159 Mannheim
Tel: 0621-1246-268
<http://www.gesis.org/Dauerbeobachtung/GML/index.htm>
E-Mail: paul.luettinger@gesis.org

Weitere Informationen zum Mikrozensus

Was ist der Mikrozensus?

Der Mikrozensus ist mit einem Auswahlsatz von 1 Prozent der Bevölkerung die größte jährliche Haushaltsbefragung in Europa. Sie wird seit 1957 in Westdeutschland und seit 1991 in den neuen Bundesländern durchgeführt. Grundlage des Mikrozensus ist die gesamte wohnberechtigte Bevölkerung in Deutschland.

Das Wichtigste über den Mikrozensus im Überblick

Grundgesamtheit:	gesamte wohnberechtigte Bevölkerung
Stichprobe:	Flächenstichprobe (Klumpenstichprobe)
Auswahleinheiten:	Auswahlbezirke (künstlich abgegrenzte Flächen bestehend aus zusammenliegenden Gebäuden oder Gebäudeteile); ein Auswahlbezirk verbleibt vier Jahre in der Stichprobe, in jedem Jahr wird ein Viertel der Auswahlbezirke durch neue Auswahlbezirke ersetzt (rotierendes Panel)
Auswahlgrundlage:	Volkszählung unter Berücksichtigung der Neubautätigkeit
Auswahlverfahren:	einstufig geschichtete Stichprobe
Schichtungsmerkmale:	Bundesland, Regierungsbezirk, Anpassungsschicht, Regionalschicht, Gebäudeschicht
Auswahlsatz:	1 Prozent
Auswahltechnik:	systematische Ziehung mit Zufallsstart
Erhebungseinheiten:	Haushalte, alle vier Jahre Wohnungen
Analyseeinheiten:	Personen, Familien, Haushalte, Wohnungen
Zuordnung Person/Haushalt:	Personen mit mehreren Haushalten werden dem Haushalt am Ort ihrer Hauptwohnung zugerechnet, gleichzeitig können Personen am Ort ihrer Nebenwohnung erfasst werden, so dass Mehrfachzählungen möglich sind
Erhebungsverfahren:	i.d.R. mündliche Befragung, zum Teil auch schriftlich
Scientific Use File:	faktisch-anonymisierte 70%-Substichprobe der Haushalte (bei Mikrozensus mit Fragen zur Wohnsituation eine 70%-Stichprobe der Wohnungen) des Mikrozensus
Gewichte in der Stichprobe:	das Scientific Use File enthält die Anpassungsvariable des MZ Originalfiles
Regionalinformationen:	das Scientific Use File des MZ78 enthält an Regionalinformationen Angaben zum Bundesland und zur Gemeindegrößenklasse

Datenerhebung

Die Angaben im Mikrozensus beziehen sich jeweils auf die letzte feiertagsfreie Woche im April bzw. auf die erste feiertagsfreie Woche im Mai (Berichtswochenkonzept). Für den Mikrozensus 1978 ist die Berichtswoche der 24. bis 30. April 1978. Die Daten des Mikrozensus werden durch die Statistischen Landesämter erhoben. Die Befragungen finden i.d.R. mündlich statt, ein kleiner Teil erfolgt mittels schriftlicher Auskunft der Befragten. Für alle Beteiligten besteht gemäß der

gesetzlichen Regelung Auskunftspflicht. Nur wenige Fragen sind freiwillig zu beantworten. Der Mikrozensus ist eine Haushaltsstichprobe, d.h. alle Personen in einem Haushalt werden entweder direkt befragt oder eine erwachsene Person gibt Auskunft über die weiteren (alle minderjährigen) Haushaltsmitglieder. Der Mikrozensus umfasst auch ausländische Haushalte. Zur Erhebung ihrer Angaben wird der gleiche Fragebogen verwendet wie bei deutschen Haushalten. Die Daten werden von den Statistischen Landesämtern aufbereitet und beim Statistischen Bundesamt zu einem Datensatz zusammengespült.

Stichprobe

Für den Mikrozensus 1978 galt der Stichprobenplan von 1972 (Flächenstichprobe auf der Basis der Volkszählung von 1970). Das stichprobenmethodische Grundkonzept des Mikrozensus 1978 ist eine einstufige Klumpenstichprobe. Die bebaute Fläche der Bundesrepublik wurde auf der Grundlage der Wohnungszählung von 1970, die im Rahmen der Volkszählung stattfand, in etwa gleich große Flächen (Auswahlbezirke bzw. Segmente) aufgeteilt. Die Neubautätigkeit wurde dabei berücksichtigt. Ein Prozent dieser Auswahlbezirke gelangen in die Stichprobe. Innerhalb eines Auswahlbezirkes wurden alle Haushalte befragt. Grundlage für die Auswahlbezirke sind ganze Gebäude oder bei größeren Gebäuden Gebäudeteile. Die Gebäude der Auswahlbezirke wurden dabei nach der Zahl ihrer Wohnung in drei Größenklassen bzw. Schichten eingeteilt, zusätzlich wurden Gebäude mit vermuteten Gemeinschaftsunterkünften einer Sonderschicht zugeordnet. Ein Auswahlbezirk umfasste durchschnittlich 23 Wohnungen.

Erhebungsprogramm

Das Fragenprogramm des Mikrozensus besteht aus einem gleich bleibenden Grundprogramm, das sozioökonomische Grundinformationen liefert und die laufende Beobachtung des Arbeitsmarktes ermöglicht. Im Grundprogramm, das sich an alle Haushalte richtet, sind im Mikrozensus folgende Bereiche enthalten:

- Angaben zu allen Personen im Haushalt und zur Haushaltszusammensetzung
- zu Erwerbstätigkeit, Arbeitslosigkeit, Arbeitssuche, Nichterwerbspersonen
- zum Besuch von Kindergarten, Schule, Hochschule
- zu allgemeiner und beruflicher Ausbildung
- zu Quellen des Lebensunterhalts, zur Höhe des Einkommens
- zur gesetzlichen Rentenversicherung
- zur gesetzlichen und privaten Krankenversicherung
- zur Pflegeversicherung
-

Der Scientific Use File des MZ78 enthält nicht die im Grundprogramm erhobenen Fragen zur Gesundheit.

Neben den jährlich zu erhebenden Merkmalen des Grundprogramms (Auswahlsatz 1 Prozent) werden in Ergänzungsprogrammen mit einem Auswahlsatz von 0,45% weitere Angaben zu den Themen Aus- und Fortbildung, aktuelle und frühere Erwerbstätigkeit und Erwerbsbeteiligung erhoben. Seit 1996 werden diese Ergänzungsprogramme ebenfalls jährlich durchgeführt. Darüber hinaus gibt es Zusatzprogramme mit Merkmalen unterschiedlicher Themenschwerpunkte, die nicht jährlich erhoben werden. Im Rahmen der Zusatzprogramme enthält der Mikrozensus 1978 jeweils mit dem Auswahlsatz 1 Prozent:

- Zusatzangaben zur beruflichen Ausbildung
- Angaben zur Pendlereigenschaft und zu Pendlermerkmalen
- Vertriebenen-/Flüchtlingseigenschaft

Der faktisch anonymisierte Mikrozensus – Das Scientific Use File

Das Mikrozensus Scientific Use File (SUF) ist eine faktisch anonymisierte 70-Prozent-Substichprobe des Mikrozensus. Für die faktisch anonymisierte Substichprobe gilt, dass sie als systematische Zufallsauswahl aus dem Originalmaterial des Statistischen Bundesamtes gezogen wird. Das Material wird dazu nach Bundesland, Regierungsbezirk, Gemeindegrößenklasse, Anzahl der Personen in Privathaushalten, Auswahlbezirksnummer und Nummer des Haushalts (bei Mikrozensus mit dem Zusatzprogramm zur Wohnsituation: Nummer der Wohnung) sortiert. Durch diese Anordnung ist gewährleistet, dass die Substichprobe hinsichtlich dieser Merkmale nur geringe zufallsbedingte Abweichungen aufweist. Bei Gemeinschaftsunterkünften erhält jede Person eine eigene fortlaufende ("Haushalts"- bzw. "Wohnungs"-)Nummer. Zur Stichprobenziehung dient die letzte (Einer-)Stelle der Haushalts- bzw. der Wohnungsnummer. Alle Haushalte (Wohnungen), deren Einerstelle einer von drei zufällig gezogenen einstelligen Zahlen entspricht, gelangen nicht in die Stichprobe (30 Prozent der Haushalte bzw. Wohnungen).

Weiterhin unterscheidet sich das Mikrozensus SUF vom Originalfile des Mikrozensus dadurch, dass bestimmte Variablen im Scientific Use File, bedingt durch die Anonymisierung, nur in klassifizierter und vergrößerter Form verfügbar sind.

So sind im Mikrozensus SUF die Regionalangaben Bundesland und eine vergrößerte Gemeindegrößenklasse enthalten. Auch bei den weiteren Variablen des Scientific Use Files werden je nach Bedarf Merkmalsvergrößerungen vorgenommen, so dass jede Ausprägung in den univariaten Verteilungen mindestens 5.000 Fälle in der Grundgesamtheit aufweist. Die Ausprägungen der Variable Nationalität werden so aggregiert, dass jede Nationalitätengruppe in der Grundgesamtheit wenigstens 50.000 Einwohner umfasst.

Zur Abgrenzung der Bevölkerungsbegriffe und zu Gewichtung und Hochrechnung im Mikrozensus Scientific Use File

Zum Vergleich mit den vom Statistischen Bundesamt veröffentlichten Zahlen sind die jeweiligen Bevölkerungs- und Erwerbskonzepte zu berücksichtigen. Eine gesonderte Aufstellung für den Mikrozensus 1978 kann abgerufen werden unter:

http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_1978/hochrechnungsfaktoren_mz78.htm.

Für den Mikrozensus Scientific Use File 78 stehen keine spezifischen Gewichtungsvariablen zur Verfügung. Bis zum Mikrozensus 1989 erfolgte die Anpassung an die Bevölkerungsfortschreibung durch das zufällige Doppeln und Streichen von Personensätzen, um bei bestimmten Merkmalen auf bestimmte Sollzahlen zu kommen. Die Variable EF76 (Dopplung/Streichung bei Anpassung) zeigt die gedoppelten und gestrichenen Fälle an. Die gestrichenen Fälle sind in den Daten noch enthalten und werden im Normalfall aus einer Analyse ausgeschlossen.

Beim Abgleich der Ergebnisse von Häufigkeitsauszählungen des Scientific Use Files mit den Veröffentlichungen des Statistischen Bundesamtes zum Mikrozensus muss die entsprechende Gewichtungsvariable verwendet werden. Um auf 1 Prozent der Bevölkerung, also auf den vollen Mikrozensus hochzurechnen, wird das Gewicht mit dem Kehrwert der Ziehungswahrscheinlichkeit für das Scientific Use File (100/70) multipliziert.

In den Fachserien des Statistischen Bundesamtes werden die Ergebnisse meist auf die gesamte Bevölkerung hochgerechnet. Zur Herstellung der Vergleichbarkeit müssen in diesem Fall Ergebnisse eigener Berechnungen noch zusätzlich zur Hochrechnung und Gewichtung mit 100 multipliziert werden. Weiterhin sind zum Vergleich mit den vom Statistischen Bundesamt veröffentlichten Zahlen die jeweiligen Bevölkerungs- und Erwerbskonzepte zu berücksichtigen.

http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/mz_1978/hochrechnungsfaktoren_mz78.htm

Zugang zum Mikrozensus

Die Mikrozensus Scientific Use Files können von unabhängigen inländischen Forschungseinrichtungen im Rahmen der Vereinbarung zwischen dem Statistischen Bundesamt und dem BMBF (Bundesministerium für Bildung und Forschung) zu Sonderkonditionen (Bereitstellungsgebühr von jeweils 95 EURO) bei den Forschungsdatenzentren des Bundes und der Länder als Rohdatenfile im ASCII-Format bestellt werden. Weitere Informationen finden Sie auf unserer Web Seite

<http://www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/MZ-Info/SUF.htm>

Service-Angebot des German Microdata Lab bei ZUMA

In Kooperation mit dem Statistischen Bundesamt stellt das German Microdata Lab (GML) dem Nutzer der Mikrozensus Scientific Use Files ausführliche Informationen zur Verfügung und bietet Setups zum Einlesen des Rohdatenmaterials und zur Erstellung von Systemfiles in Datenformaten gängiger Softwarepakete an. Für Interessenten und Nutzer des Mikrozensus besteht weiterhin die Möglichkeit, sich an das GML zu wenden und folgende Serviceleistungen in Anspruch zu nehmen:

- Beratung, inwieweit der Forschungsplan mit Daten des Mikrozensus zu realisieren ist
- Beratung und evtl. Unterstützung beim Verfassen des Antrags (Forschungsplans)
- Beratung zu technischen Fragen (EDV-Voraussetzungen, Datensicherheit etc.)
- Beratung zu Auswertungs- und Analysefragen

Weiterführende Literatur

Informationen zum Mikrozensus und seiner Konzeption

Emmerling, Dieter; Riede, Thomas, 1997: 40 Jahre Mikrozensus. In: *Wirtschaft und Statistik* 3/1997, S. 160-174.

Hartmann, Peter H., 1989: Der Mikrozensus als Datenquelle für die Sozialwissenschaften. In: *ZUMA-Nachrichten* 24, S. 6-25.

Heidenreich, Hans-Joachim, 1994: Hochrechnung des Mikrozensus ab 1990. In: Siegfried Gabler, Jürgen Hoffmeyer-Zlotnik, Dagmar Krebs (Hg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag. S. 112-123.

Lengerer, Andrea; Bohr, Jeanette; Janßen, Andrea, 2005: Haushalte, Familien und Lebensformen im Mikrozensus - Konzepte und Typisierungen. *ZUMA-Arbeitsbericht* 2005/05.

Lüttinger, Paul; Riede, Thomas, 1997: Der Mikrozensus: amtliche Daten für die Sozialforschung. In: *ZUMA-Nachrichten* 41, S. 19-43.

Meyer, Kurt, 1994: Zum Auswahlplan des Mikrozensus ab 1990. In: Siegfried Gabler, Jürgen Hoffmeyer-Zlotnik, Dagmar Krebs (Hg.): *Gewichtung in der Umfragepraxis*. Opladen: Westdeutscher Verlag. S. 106-111.

Statistisches Bundesamt (Hg.), 1997: *Arbeitsunterlagen zum Mikrozensus. Das Erhebungsprogramm des Mikrozensus seit 1957*. Wiesbaden.

Statistisches Bundesamt (Hg.), 1989: *Inhaltliche Fragen bevölkerungstatistischer Stichproben am Beispiel des Mikrozensus*. Wiesbaden.

Schmidt, Simone, 2000: *Erwerbstätigkeit im Mikrozensus. Konzept, Definition, Umsetzung*. *ZUMA-Arbeitsbericht* 2000/01.

Zur Weitergabe faktisch anonymisierter Einzeldaten

Müller, Walter; Blien, Uwe ; Knoche, Peter; Wirth, Heike, 1991: *Die Faktische Anonymität von Mikrodaten*. Stuttgart: Metzler-Poeschel.

Wirth, Heike, 1992: Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes. In: *ZUMA-Nachrichten* 30, S. 7-65.

Zur Fehlerrechnung im Mikrozensus

Krug, Walter; Nourney, Martin; Schmidt, Jürgen, 1994: *Wirtschafts- und Sozialstatistik. Gewinnung von Daten*. München: Oldenbourg.

Ergebnisse des Mikrozensus

Die Ergebnisse des Mikrozensus werden vom Statistischen Bundesamt in den folgenden Fachserien und Reihen veröffentlicht:

Fachserie 1: Bevölkerung und Erwerbstätigkeit, Reihe 3: Haushalte und Familien (Ergebnisse des Mikrozensus).

Fachserie 1: Bevölkerung und Erwerbstätigkeit, Reihe 4.1.1: Stand und Entwicklung der Erwerbstätigkeit (Ergebnisse des Mikrozensus).

Fachserie 1: Bevölkerung und Erwerbstätigkeit, Reihe 4.1.2: Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen (Ergebnisse des Mikrozensus).

Fachserie 12: Gesundheitswesen, Reihe S. 3: Fragen zur Gesundheit (Ergebnisse des Mikrozensus).

Fachserie 13: Sozialeleistungen, Reihe 1: Altersvorsorge und vermögenswirksame Leistungen (Ergebnisse des Mikrozensus).

Umfassende Informationen zur Aufbereitung und Dokumentation der Mikrozensus Scientific Use Files finden Sie unter

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

Publikationen, die auf den Scientific-Use-Files der Mikrozensus basieren, finden Sie unter

http://www.gesis.org/Dauerbeobachtung/GML/Service/MZ-SUF-Literatur/MZ_Literatur.htm