

## **Campus File zur Sozialhilfestatistik, Stichtag 31.12.1998**

Das Campus File zur Sozialhilfestatistik ist ein 5%iger Auszug aus dem Public Use File. Das Datenmaterial wurde bereits für den Public Use File so verändert, dass nach menschlichem Ermessen keine Rückschlüsse auf einzelnen Personen aus der Grundgesamtheit mehr möglich sind, deshalb ist auch der Campus File absolut anonym.

Das Campus File für das Jahr 1998 enthält Angaben zu insgesamt 35 551 Personen (Hochrechnungsfaktor = 80) aus 18 478 Haushalten.

Seine Zielrichtung ist das Üben der Analyse von amtlichen Einzeldaten mit Hilfe gängiger Statistikprogramme, wie SPSS oder SAS. Es unterscheidet sich von gängigen standardisierten Public Use File (PUF) dadurch, dass die Anonymisierung nicht darauf abzielt den Informationsgehalt bei gegebenem Datenschutz möglichst hoch zu halten, um damit eine möglichst repräsentative Forschung zu ermöglichen. Es soll vielmehr ein ausreichend großes Übungsmaterial darstellen, das zeitnah für Universitäten zur Methodenausbildung verfügbar ist.

Ausgangspunkt für dieses Campus-File ist das Public Use File über die Empfänger laufender Hilfe zum Lebensunterhalt außerhalb von Einrichtungen zum 31.12.1998, das im Folgenden näher charakterisiert wird.

## **Informationen zur Stichprobe der Sozialhilfeempfängerstatistik**

### **Beschreibung des Datensatzes**

Die Sozialhilfeempfängerstatistik ist eine Bestandsstatistik, die im zugrunde liegenden Datensatz die Anzahl der Empfänger von laufender Hilfe zum Lebensunterhalt (HLU) außerhalb von Einrichtungen zum Stichtag 31.12.1998 ausweist. Diese werden auch als "Sozialhilfeempfänger im engeren Sinne" bezeichnet. Gemäß § 132 Abs. 2 des Bundessozialhilfegesetzes stellen die Statistischen Ämter der Länder dem Statistischen Bundesamt jedes Jahr u.a. Einzelangaben aus einer Zufallsstichprobe mit einem Auswahlatz von 25 vom Hundert der Empfänger von laufender Hilfe zum Lebensunterhalt zur Verfügung. Der Datensatz für das Jahr 1998, der zur Erstellung des Campus Files verwendet wurde, umfasst insgesamt 719.641 Fälle (Hochrechnungsfaktor = 4). In der Datei sind sowohl personen- als auch haushaltsbezogene Daten enthalten. Insgesamt 61 Variablen, wie z.B. Geburtsjahr, Geschlecht, Staatsangehörigkeit, Erwerbsstatus, höchster Schul-

bzw. Berufsausbildungsabschluss, Haushaltstyp, Anspruch und Bruttobedarf je Monat, vorhandene Einkommensarten etc. ermöglichen eine Vielzahl von Auswertungsmöglichkeiten. Die vollständige Variablenliste kann der Datensatzbeschreibung entnommen werden, die als gesonderte Datei zum Download angeboten wird.

Aus datenschutzrechtlichen Gründen musste eine Anonymisierung der Daten vorgenommen werden, wodurch zum Teil eine Modifizierung der Variablen erforderlich wurde. Die Auswirkungen dieser Anpassungen werden im Folgenden näher erläutert:

Im anonymisierten Material sind lediglich drei **Gebietsstände** ausgewiesen. Das frühere Bundesgebiet Nord inkl. Berlin-Ost umfasst die drei Stadtstaaten sowie die Länder Schleswig-Holstein, Niedersachsen und Nordrhein-Westfalen. Das frühere Bundesgebiet Süd ist definiert durch die fünf verbleibenden Bundesländer Hessen, Rheinland-Pfalz, Baden-Württemberg, Bayern und das Saarland. Die neuen Länder ohne Berlin-Ost werden durch die fünf verbleibenden ostdeutschen Bundesländer abgedeckt.

Die Variable "**Stellung zum Haushaltsvorstand**" liefert wichtige Informationen zur Bestimmung des Haushaltstyps. Sie hat fünf Ausprägungen: Haushaltsvorstand (1), Ehegatte(in) (2), Kind (3), Verwandte(r) oder Verschwägerte(r) (4) und sonstige Person (5). Der Haushaltsvorstand ist die Person, die die Generalkosten des Haushalts trägt. Aus diesem Grund steht dem Haushaltsvorstand ein erhöhter Eckregelsatz zu. Eine nicht ganz eindeutige Abgrenzung verbirgt sich hinter der Klassifizierung "Kind". Der Oberbegriff "Stellung zum Haushaltsvorstand" lässt vermuten, dass es sich hierbei um die leiblichen Kinder des Haushaltsvorstandes handelt. Der Begriff ist aber in dem hier verwendeten Zusammenhang weiter gefasst. Tatsächlich werden als Kinder neben ehelichen und nichtehelichen auch Stief-, Adoptiv- und Pflegekinder mit einbezogen, die zusammen mit ihren Eltern oder einem Elternteil in einem Haushalt wohnen.

Aus der Variable "**Geburtsjahr**" lässt sich das Alter der HLU-Empfänger berechnen. Im Rahmen der Anonymisierung wurde das Höchstalter auf 90 Jahre festgesetzt. Im Datensatz ist somit Personen, die vor 1908 geboren wurden, das Geburtsjahr 1908 zugewiesen.

Als ein Teil der laufenden Hilfe zum Lebensunterhalt werden bestimmten Personengruppen, beispielsweise behinderten Menschen und Pflegebedürftigen, zusätzlich "**Mehrbedarfszuschläge**" gewährt, welche im einzelnen der Datensatzbeschreibung entnommen werden können. Ein Sozialhilfeempfänger kann von insgesamt sieben unterschiedlichen Mehrbedarfszuschlägen maximal

vier gleichzeitig erhalten. Dadurch sind für dieses Merkmal im Datensatz vier Positionen vorgesehen, welche jeweils die Ausprägungen 1-7 annehmen können.

Ähnlich wie bei der Höchstaltersbegrenzung auf 90 Jahre (siehe oben) war es in dieser Stichprobe aus datenschutzrechtlichen Gründen erforderlich, bei bestimmten Variablen eine "Abschneidengrenze" festzulegen. So musste bei den Merkmalen **"Nummer der Person"**, **"Anzahl der Personen in der Bedarfsgemeinschaft"**, **"Zahl der Haushaltsmitglieder"** und **"Zahl aller HLU-Empfänger im Haushalt"** eine Beschränkung auf 10 Personen vorgenommen werden. Das bedeutet, dass beispielsweise eine Bedarfsgemeinschaft, die real aus 12 Personen besteht, im vorliegenden Datensatz lediglich mit 10 Personen erfasst wird. Da dies nur für relativ wenige Fälle zutrifft, ist die zu erwartende Untererfassung als gering zu beurteilen.

Die Zahlbeträge **"Bruttobedarf"**, **"Anerkannte Bruttokaltmiete"**, **"Anspruch der Bedarfsgemeinschaft"** und die aus diesen Angaben ermittelte Größe **"Angerechnetes Einkommen"** wurden auf Zehnerstellen gerundet (z.B.: 1452,-DM → 1450,-DM). Um Rückschlüsse auf einzelne Empfänger zu vermeiden, wurden bei diesen Beträgen Höchstgrenzen eingeführt. Der maximal ausgewiesene *Bruttobedarf* liegt bei 5.000,-DM. Der größte ausgewiesene Wert für die *anerkannte Bruttokaltmiete* beläuft sich auf 1.800,-DM; beim *Nettoanspruch* beträgt er 3.500,-DM. Dies ist auch der Höchstwert für das *angerechnete Einkommen* (Hinweis zur Umrechnung: 1 Euro = 1,95583 DM).

Die **"Längste bisherige Dauer der Hilfestellung an die Bedarfsgemeinschaft in der aktuellen Zusammensetzung"** und die **"Bisherige Dauer der ununterbrochenen Hilfestellung für mindestens ein Mitglied der Bedarfsgemeinschaft"** ist auf 120 Monate beschränkt. Einer Bezugsdauer, die länger als 120 Monate anhält, ist der Wert "120" zugeordnet. Diese Kappungsgrenzen führen dazu, dass die durchschnittliche Bezugsdauer gegenüber dem nicht-anonymisierten Material und gegenüber den Daten der Totalerhebung geringer ausfällt.

Die Variable **"Typ der Bedarfsgemeinschaft"** hat insgesamt 24 Ausprägungen und ist aus den Variablen **"Stellung zum Haushaltsvorstand"**, **"Geburtsjahr"**, **"Geschlecht"** und **"Anzahl der Personen in der Bedarfsgemeinschaft"** konstruiert. Die Haushaltstypisierung wurde in der vorliegenden Stichprobe weitgehend an die der Standardtabellen und Veröffentlichungen des Statistischen Bundesamts angepasst. Dem entsprechend wurden auch hier die gleichgeschlechtlichen Paare in der Kategorie **"nichteheliche Lebensgemeinschaft"** erfasst.

## Qualität der Stichprobe

Hinsichtlich der Aussagefähigkeit der Stichprobe ist deren Qualität von besonderer Bedeutung.

. Die vorgenommene Qualitätskontrolle basiert auf zwei Säulen: Auf der Prüfung der Einzeldatensätze und auf dem Vergleich der Stichprobenergebnisse mit denen der Totalerhebung.

Die Prüfung der Einzeldatensätze erfolgt bereits durch die Statistischen Landesämter. Sie prüfen die Plausibilität der Originaldaten. Dabei handelt es sich einerseits um eine *Signierkontrolle* (haben die Variablen auch wirklich nur die vorgesehenen Ausprägungen?) und andererseits um eine *Kombinationskontrolle* (Beispiel: Wenn a gilt, dann muss auch b vorkommen).

Zur Beurteilung der Qualität der Stichprobe wurden für einige ausgewählte Variablen die Abweichungen der Stichprobenergebnisse von denen der Totalerhebung ermittelt. Nach dem Gesetz der großen Zahl nimmt der Stichprobenfehler mit kleiner werdender Fallzahl zu. Je detaillierter die Untergliederung wird und je kleiner die Fallzahlen werden, desto wahrscheinlicher wird die Abweichung der Stichprobenergebnisse von denen der Totalerhebung. Durch die Anonymisierung kommt ein weiterer Aspekt hinzu: Bestimmte Informationen liegen nicht ganz so exakt wie im Originalmaterial vor. Trotz dieser Einschränkungen lässt sich sagen, dass der zugrunde liegende Datensatz von hoher Qualität ist.

- Im Folgenden sind die Ergebnisse des Datenvergleichs von Public Use File und Totalerhebung zusammengefasst:
- Die Stichprobenergebnisse der **über 80 Jahre alten Hilfeempfänger** weichen etwas stärker als die jüngeren Altersgruppen von der Totalerhebung ab.
- Bei der Untergliederung der **Gruppe der Nichtdeutschen nach Alter** können aufgrund geringer Fallzahlen größere Abweichungen von den Daten der Totalerhebung auftreten.
- Auch bei der Unterscheidung nach dem **Typ der Bedarfsgemeinschaft** treten bei geringer Zellenbesetzung größere Abweichungen von der Totalerhebung auf.
- Darüber hinaus können die **Mehrbedarfszuschläge** nach Altersgruppen ebenfalls das Problem geringer Fallzahlen in den jeweiligen Kombinationen widerspiegeln. Bei geringer Zellenbesetzung besteht das Risiko, dass die Stichprobenergebnisse von denen der Totalerhebung stärker abweichen. In einigen Konstellationen mit einer Zellenbesetzung von unter 100 Fällen, können die prozentuellen Abweichung im zweistelligen Bereich liegen.

## Lieferumfang

Die Daten stehen dem Nutzer als kommagetrennter ASCII-File, sowie als SPSS-, SAS und STATA-Datei zur Verfügung. Ergänzend wird eine Datei mitgeliefert, die das Konvertieren des ASCII-Files in eine SPSS-Datei ermöglichen. Die Verwendung der Konvertierungsdateien ("Setup\_cf\_sohi98") empfiehlt sich für alle Anwender, die abweichende Variablennamen oder Kommentierungen vergeben möchten bzw. die nur eine Auswahl der Variablen einlesen wollen. In diesen Fällen sind die beabsichtigten Änderungen in den Konvertierungsdateien vorzunehmen. Die gelieferten Variablen und ihre Spezifizierungen sind in der Datensatzbeschreibung dokumentiert.