

Leitfaden zur Anonymisierung der Einkommens- und Verbrauchsstichprobe 1962/63 zur Verwendung als Public Use File

1. Vorbemerkung

Die Einkommens- und Verbrauchsstichprobe 1962/63 (EVS62/63) wurde als freiwillige Erhebung durchgeführt. Die damaligen Rechtsgrundlagen sahen eine Weitergabe der Daten an Einrichtungen oder Personen außerhalb der amtlichen Statistik nicht vor. Nach der vorherrschenden Rechtsauffassung ist eine Weitergabe deshalb nur in der Form eines so genannten Public Use Files (PUF) zulässig. Die Veröffentlichung als PUF setzt voraus, dass eine De-Anonymisierung von Befragten ausgeschlossen ist und die Daten mithin absolut anonymisiert sind.

2. Basismaterial

Die in den Daten enthaltenen Merkmale geben insbesondere Informationen über Haushaltsstruktur, detaillierte Angaben zur Erwerbstätigkeit des Haushaltsvorstandes, Einkommensentstehung, Besitz ausgewählter langlebiger Gebrauchs- und Konsumgüter. Die Mikrodatenfiles zur Einkommensverwendung (insb. die Angaben aus den Haushaltsbüchern und die Feinschreibungen) liegen nicht vor.

Bei der Rückerschließung der Daten konnte man nicht mehr feststellen, aus welchem Schritt der Datenproduktion die vorliegenden Daten stammen. Aus diesem Grunde ist eine Reproduktion der in den entsprechenden Fachserien veröffentlichten Ergebnisse mit dem vorliegenden Material nicht uneingeschränkt möglich.

3. Anonymisierungsmaßnahmen

Folgende Anonymisierungsmaßnahmen und Eigenschaften des vorliegenden Materials führen zur absoluten Anonymisierung der Daten:

(a) Alter der Daten

Da die Erhebung der EVS62 mittlerweile über vierzig Jahre zurückliegt, kann davon ausgegangen werden, dass Zusatzinformationen nur in eingeschränktem Umfang verfügbar sind und vorliegende Informationen von nur geringer Verlässlichkeit sind. Ins-

besondere kann davon ausgegangen werden, dass viele der 1962 befragten Haushalte in ihrer Gesamtheit oder zumindest in ihrer damaligen Struktur und regionalen Zuordnung nicht mehr existieren. Gleiches gilt für Informationen über Haushaltsmitglieder (etwa zur Berufstätigkeit des Haushaltsvorstandes).

(b) Stichprobenziehung

In der Rechtsgrundlage der EVS62 ist der Auswahlatz mit höchstens 0,3% aller bundesdeutschen Haushalte festgelegt. Bei damals rund 20 Mill. Haushalten in der Bundesrepublik konnten höchstens 60.000 in die Erhebung einbezogen werden. Das vorliegende Material umfasst etwa 34.000 Haushalte. Es ist nicht mehr nachvollziehbar, ob es sich bei diesen Haushalten um alle in der EVS62 erhobenen Einheiten handelt, und wenn ja, welcher Auswahlmechanismus eingesetzt wurde. Als zusätzliche Anonymisierungsmaßnahme vor der Stichprobenziehung werden aus dem Datenmaterial diejenigen 841 Haushalte entfernt, deren Größe 6 Personen übersteigt. Anschließend wird der PUF als eine 98%-Zufallsstichprobe aus dem Datenmaterial gezogen und stellt insgesamt eine rund 95%-ige Stichprobe des vorliegenden Originalmaterials dar.

(c) Datenqualität

Das vorliegende Material weist eine Reihe von Fällen mit inkonsistenten Angaben auf (insb. mit Codierfehlern). Von einer nachträglichen Plausibilisierung des Materials wird aus methodischen Gründen abgesehen. Diese Inkonsistenzen erlauben zwar Analysen bezüglich der Verteilung von Variablen, erschweren aber die Identifikation einzelner Merkmalsträger und erhöhen mithin den Anonymisierungsgrad des PUF.

(d) Vergrößerung von Regionalinformationen

In der EVS62 wurde neben dem Bundesland auch erfragt, ob und wenn ja in welcher Agglomeration (Ballungsraum) ein Haushalt liegt. Die zugrunde liegende Variable wurde durch Zusammenfassungen vergrößert.

Die Variablen EAH031, EAH038, EAH049-EAH051, EAH053-EAH060, EAH177 und EAH180 werden aus dem Material gelöscht. Es handelt sich dabei um Variablen, die entweder leer, Dupletten anderer Variablen oder aus anderen Variablen abgeleitet sind. Für die Variablen EAH045, EAH046, EAH047 werden neue Variablen EAH045_, EAH046_ und EAH047_ hergeleitet. Diese weisen differenziertere Merkmalsausprägungen als die Originalvariablen auf.

Fazit

Die unter 3. beschriebenen Anonymisierungsmaßnahmen führen in Verbindung mit der Qualität und dem Alter der Daten zu einem Mikrodatenfile, bei dem eine De-Anonymisierung einzelner Merkmalsträger ausgeschlossen ist. Der Datensatz ist damit absolut anonym und kann in dieser Form als Public Use File veröffentlicht werden.