

ZUSAMMENFASSUNG

VOLKSZÄHLUNG UND PUS

Die Volkszählung ist eine der grössten und wichtigsten statistischen Datensammlungen der amtlichen Statistik in der Schweiz. Dieser im wahren Sinne des Wortes umfassende Datensatz enthält Daten zur gesamten Bevölkerung der Schweiz aus so unterschiedlichen Bereichen wie Demographie, berufliche Stellung und soziale Situation. Ausserdem stehen Informationen über Haushalte und Wohnverhältnisse zur Verfügung. Die Volkszählung wird alle 10 Jahre durchgeführt und ermöglicht dadurch einen vertieften Einblick in die Entwicklung des Landes. Es ist deshalb sehr wichtig, dass Forschende und weitere interessierte Kreise leicht Zugang zu diesen Informationen haben.

Diese Vorzüge der Volkszählung stellen für die Analyse gleichzeitig erhebliche Schwierigkeiten dar, u.a. wegen des enormen Umfangs dieser umfassenden Datensammlung. Eine Datensammlung über mehrere Millionen Einzelpersonen ist selbst mit modernen Informatiklösungen schwierig zu verwalten. Die Hauptschwierigkeit liegt jedoch nicht in diesem Punkt, sondern bei der Sicherung der Vertraulichkeit der Daten: Da eine Volkszählung grundsätzlich die ganze Bevölkerung umfasst, wäre es durch Kombination verschiedener Kriterien möglich, eine bestimmte Person zu identifizieren, was sowohl aus ethischer als auch aus rechtlicher Sicht inakzeptabel ist. Damit VZ-Daten den Interessierten trotzdem ohne Einschränkungen zur Verfügung gestellt werden können, beschloss das Bundesamt für Statistik (BFS), unter Sicherstellung der Anonymität eine zufällig ausgewählte Stichprobe («Public Use Sample» oder PUS) bereit zu stellen.¹

Die beschriebenen Daten stammen nicht allein aus der Volkszählung 2000, sondern auch aus den Vorläufererhebungen der Jahre 1970, 1980 und 1990. Dies ermöglicht Analysen über berufliche, regionale und andere Entwicklungen in der Schweiz in diesem Zeitraum. Der Vergleich mehrerer Volkszählungen erfordert einen spezifischen Arbeitsgang zur Harmonisierung der Daten. Wie bei jeder Erhebung sind die Daten der Volkszählung ein Spiegelbild ihrer Zeit: So sind beispielsweise die Informatikberufe von 1970 nicht die gleichen wie heute, was eine Überarbeitung der Nomenklaturen erfordert. Es gilt somit Überlegungen zu den Vergleichsmodalitäten der Daten aus verschiedenen Epochen anzustellen, die über rein technische Aspekte hinausgehen.

Diese letzte Feststellung zeigt auch deutlich die Bedeutung der Dokumentation. Diese muss sämtliche relevanten Aspekte der Datenverarbeitung abdecken, gleichzeitig aber auch den Zielen der Datenaufbereitung und den Arbeitsweisen der potenziellen Benutzenden Rechnung tragen. Die vorliegende Beschreibung der Daten stützt sich im Wesentlichen auf eine umfangreiche, vom BFS erarbeitete technische Dokumentation sowie auf die ursprünglichen Informationen über die Volkszählungen, die in zahlreichen Publikationen des BFS zu finden sind. Die vorliegende

¹ Zu erwähnen sind zwei Anonymisierungsstufen: In einem ersten Schritt, der bereits bei der Aufbereitung der Daten vorgenommen wird, werden in Einklang mit dem Datenschutzgesetz Namen und Adressen eliminiert, während in einem zweiten Schritt, der hier besprochen wird, eine Identifizierung durch Rückschlüsse verunmöglicht wird.

Unter gewissen Voraussetzungen können auch die Originaldaten verwendet werden. In diesem Fall wird in einem spezifischen Vertrag festgehalten, dass die Einhaltung des Berufsgeheimnisses und eine vertrauliche Behandlung zu gewährleisten sind. Die Kontaktpersonen beim BFS sind am Schluss des Textes aufgeführt.

Beschreibung soll einfach und verständlich sein und gleichzeitig Informationen aus zahlreichen Publikationen zusammenfassen.

WOZU UND WEM DIENEN “PUBLIC USE SAMPLES” (PUS)?

Kurz gesagt sollte es die vorliegende Arbeit allen interessierten Personen ermöglichen, die Daten der Volkszählung im Rahmen eines “reduzierten Modells” zu analysieren. Im Vordergrund stehen dabei Anwendungen:

- im Unterricht, zum Beispiel in Statistik- oder anderen Kursen sowie für praktische Forschungsarbeiten, um den Studierenden die Analyse der Volkszählung näher zu bringen.
- bei Vorarbeiten für komplexere Forschungsvorhaben, z.B. in Form von Sondierungs- bzw. Vorbereitungsanalysen, um Impulse dahingehend zu liefern, Forschungsprojekte mit einer Vorstudie auf der Basis der amtlichen Statistik zu beginnen;
- im Bereich der Vorbereitung von Forschungsarbeiten, die Angaben über bestimmte Merkmale der Bevölkerung – Quoten, Schätzungen zur Grösse bestimmter Gruppen, Vorbereitungen von Klassifikationen, Aggregationsversuche usw. – erfordern;
- im Sinne von Kontextdaten, die Informationen aus anderen Forschungsarbeiten ergänzen und z.B. eine Verknüpfung von Raumgliederungen und in den Unterlagen vorgestellten Aggregationsmöglichkeiten zulassen.

Zum Zielpublikum gehören alle interessierten Personen, ungeachtet ihrer beruflichen Tätigkeit. Dazu zählen namentlich:

- Studierende, die im Rahmen von Kursen und Seminaren betreut werden;
- Doktorierende und wissenschaftliche Mitarbeitende an Universitäten. Diese können sich anhand der PUS mit der amtlichen Statistik vertraut machen und lernen, Fragestellungen präziser zu formulieren und die zur Beantwortung notwendigen Daten zu verwenden;
- Mitarbeitende privater Einrichtungen und öffentlicher Verwaltungen, die sich so bewusst werden, welche Möglichkeiten die Volkszählung bietet und damit leichter Zugang zu Gesamtdaten finden;
- ausländische Forschende, vor allem, wenn sie bereits in ihren Herkunftsländern mit PUS gearbeitet haben.

Ein Hauptinteresse der PUS liegt paradoxerweise nicht in den hier präsentierten Daten. Wie erwähnt wurden die Informationen im Hinblick auf eine Anonymisierung reduziert, was bestimmte Analysen erschwert oder verunmöglicht. Indem jedoch die PUS die Reichhaltigkeit und das Potenzial der vorhandenen Volkszählungsdaten aufzeigen, erleichtern sie den Zugang zu weiteren Daten: Personen, welche die Variablen bereits kennen, können besser abschätzen, inwieweit es angebracht ist, eine Analyse auf die gesamten Daten der Volkszählung auszuweiten.

WELCHE DATEN STEHEN ZUR VERFÜGUNG?

Die Volkszählung umfasst mehrere Datenblöcke – solche über Personen, Haushalte, Wohnungen und Gebäude. Um eine Auswertung zu vereinfachen, hat das BFS entschieden, hier nur die Daten der Einzelpersonen zu berücksichtigen und diese durch Informationen über die Haushalte, Gebäude und Wohnungen zu ergänzen. In der Fachsprache wird deshalb von einem “Public Use Sample” von Einzelpersonen und nicht von einem “Public Use Sample” von Haushalten gesprochen.

Für jede Volkszählung wird durch einfache Zufallsauswahl eine Stichprobe von Personen gebildet, die als repräsentative Stichprobe 5% der Bevölkerung abdeckt. In inhaltlicher Hinsicht war es unser Bestreben, sämtliche verfügbaren Informationen zu verwenden. Einschränkungen wurden

ausschliesslich gemacht, um die Identifikation von Personen auszuschliessen. Deshalb haben wir uns konsequent dafür entschieden, bei Bedarf eher auf Einzelheiten der Antworten zu verzichten, als uns auf eine Untergruppe von Variablen zu beschränken.

Die Grundidee des Projekts besteht also darin, ein "Motivationsprodukt" zur Verfügung zu stellen: Indem wir die ganze Palette der verfügbaren Daten präsentieren, möchten wir die Benutzenden für eine Arbeit mit den Daten der gesamten Volkszählung motivieren. Voraussetzung dafür ist dann allerdings der Abschluss eines Vertrags zur Wahrung der Anonymität.

Konkret umfasst dieses Produkt einen dreifachen Inhalt: Angaben zur Volkszählung in der Schweiz, vor allem aber die dabei gewonnenen Daten; einige Hinweise auf die Anforderungen des Datenschutzes und auf die Mittel zu ihrer Erfüllung; schliesslich eine umfassende Dokumentation der verfügbaren Merkmale.