

Konzept zur faktischen Anonymisierung statistischer Einzelangaben aus der Volkszählung 1970 zur Erstellung eines standardisierten Scientific-Use-File für die Off-Site-Auswertung

Einleitung

Die mit Erhebungsstichtag 27. Mai 1970 durchgeführte Volks- und Berufszählung wurde organisatorisch unterteilt in eine Totalerhebung und in einen Repräsentativteil, der nur 10% der Bevölkerung erfasst und mit der Berufszählung identisch ist. Während die Totalerhebung zum Ziel hatte, regional tief gegliederte Ergebnisse u. a. für die Stadt- und Regionalplanung zu liefern, sollte der Repräsentativteil einer detaillierteren und spezifischeren Erfassung von Merkmalen über die Erwerbstätigkeit und die Ausbildung der Bevölkerung dienen. Die 10%-Stichprobe wurde seinerzeit von allen Bundesländern mit Ausnahme des Saarlands gezogen. Diese Länderfiles einschließlich des Totalbestandes des Saarlands liegen dem seit 1.7.2007 für die Zensen fachlich zuständigen FDZ-Standort Bad Ems vor. Derselbe Datenbestand war Mitte der 1979er-Jahre bereits in anonymisierter Form der Universität Mannheim für das Projekt "Vergleichende Analysen der Sozialstruktur mit Massendaten (VASMA)"¹⁾ von den Statistischen Landesämtern zur Verfügung gestellt worden.

Die Einzeldaten der 10%-Stichprobe aller 11 Bundesländer einschl. Westberlin sollen der Wissenschaft als standardisierter Scientific-Use-File (SUF) angeboten werden. Dazu muss der Mikrodatensatz den Anforderungen der faktischen Anonymität genügen, die sicherstellt, dass die Einzelangaben den Merkmalsträgern nur mit unverhältnismäßig großem Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können. Durch die genannten Maßnahmen ist der Versuch einer Deanonymisierung für einen potentiellen Datenangreifer mit einem erheblichen Aufwand im Sinne des § 16 Abs. 6 BStatG verbunden. Die Datennutzung wird weiterhin durch einen Nutzungsvertrag abgesichert, in dem ein Reidentifikationsverbot sowie eine Strafe bei Vertragsbruch festgeschrieben ist.

Anonymisierungsmaßnahmen

Grundsätzlich ist zu sehen, dass die VZ-1970 mehr als 37 Jahre zurückliegt. In Bezug auf die heutige Bevölkerung bedeutet dies:

- Nahezu alle damals über 55-Jährigen dürften inzwischen verstorben sein
- Die heute 0- bis 36-Jährigen sind im Datenbestand nicht enthalten
- Für einen Großteil der heute 37- bis 55-Jährigen gibt es zu vielen Merkmalen keine Angaben (z.B. zu schulischer, beruflicher Ausbildung, Erwerbstätigkeit, Familie etc.)

Des Weiteren waren die Ende der 70er Jahre von den Statistischen Landesämtern an das VASMA-Projekt übermittelten Volkszählungsdaten bereits absolut anonymisiert. Zu diesem Zweck hatten die Landesämter mit den Angaben über Regierungsbezirke, Kreise und Gemeinden eine Reihe detaillierter regionaler Informationen und mit der Zählerlistennummer und den Informationen zu den Zählbezirken relevante Kennzeichen des Stichprobenplans gelöscht bzw. anonymisiert. Die ebenfalls gelöschte Haushaltsnummer im Erfassungsfeld EF198 bedeutet nur eine Nummerierung der Haushalte innerhalb eines Zählbezirks. Da aber weder die Identifikation der

¹⁾ <http://www.gesis.org/Dauerbeobachtung/GML/Daten/VZ1970/doc70.htm>

Zählbezirke, der Gemeinden, der Kreise oder Regierungsbezirke möglich ist, ergibt die Haushaltsnummer keinerlei Rückschluss für die Identifikation des Haushalts. Dies auch deshalb nicht, weil nicht bekannt ist, nach welchen Prinzipien die Vergabe des Haushaltsnummern in den einzelnen Zählbezirken erfolgte.

Als weitere Anonymisierungsmaßnahme wurde vom FDZ-Standort Bad Ems aus dem originären Totalbestand des Saarlands – 1.148.384 Personendatensätze – eine 10%-Stichprobe gezogen. Methodisch geschah dies in Anlehnung an das in den 1970er Jahren vom VASMA-Projekt praktizierte fachliche Konzept, nach dem dort damals aus der originären 10%-Stichprobe der Länder aus rechentechnischen Gründen nochmals eine 10%-Auswahl gezogen worden war.

Die Auswahlgrundlage für die Ziehung der Stichprobe bilden die Privathaushalte – dazu gehören auch die Privathaushalte in den Anstalten – und die zusammengefassten Anstaltspersonen, die keinem Haushalt zugehörig sind. Beide zusammen, Haushalte und Anstaltsinsassen, ergeben die Stichprobendatei. Als Auswahlmethode wurde die einfache Zufallsstichprobe gewählt.

Im ersten Schritt wurden die Anstaltsinsassen – Ausprägung des Merkmals Haushaltsgröße (EF101) = 00 – aus dem saarländischen Gesamtbestand selektiert und in einer Anstaltsdatei zusammengefasst. Aus dieser Anstaltsdatei wurde per einfache Zufallsstichprobe (Zufallsgenerator: Mersenne Twister als Standardfunktion von SPSS)) ein 10%-Datenbestand ausgewählt.

Zur Erstellung der Haushaltsdatei erfolgte die Zusammenfassung der Personensätze zu Haushaltsdatensätzen mit Hilfe der in den Originaldatensätzen enthaltenen Haushaltsidentifikatoren

- Gemeindegrößenklasse des Wohnorts (EF78)
- Haushaltsgröße: Wohnberechtigte Bevölkerung (W1) (EF101)
- Haushaltsgröße: Wohnbevölkerung (W2) (EF102)
- Haushaltsgröße: Bevölkerung in Privathaushalten und Anstalten (W7) (EF103)
- Korrekturnummer (EF192)
- Bündelnummer (EF193)
- Haushaltsnummer (EF198)

Ob eine Person noch zum gleichen Haushalt gehört wie die vorherige Person, wurde durch Vergleich der o. a. Identifikatoren des aktuellen Datensatzes mit denen des vorangehenden ermittelt. Ein Haushaltssatz ist dann erstellt, wenn die Anzahl der entsprechend deckungsgleichen Personendatensätze mit der jeweiligen Haushaltsgröße übereinstimmt. Nach dieser Methodik wurden 391.752 Privathaushalte gefunden. Eine Zahl, die mit den Veröffentlichungen des Saarlands zur Volkszählung 1970 korrespondiert.

Aus der so gebildeten Haushaltsdatei wurde per einfache Zufallsstichprobe analog der Vorgehensweise bei der Anstaltsdatei eine 10%-Stichprobe gezogen.

Anschließend wurden den so ausgewählten Haushalten die zugehörigen Personendatensätze wieder zugeordnet und diese Datei mit der 10%-Stichprobendatei der Anstaltsinsassen zusammengefasst.

Nachdem nun von allen Bundesländern einschl. Westberlin eine 10%-Stichprobe vorlag, wurden noch die Variablen EF2 (Tag des Geburtsdatums) sowie EF186 bis EF205 (signifikante Kennzeichen des Stichprobenplans wie Zählerlistennummer, Zählbezirke etc., gelöscht.

Aufgrund dieses durchgeführten Maßnahmenbündels erfüllen alle 10%-Länderfiles den Anspruch der faktischen Anonymität, wie er für einen standardisierten Scientific-Use-File im Angebot des FDZ gestellt wird.

Merkmale des standardisierten Scientific-Use-File der VZ-1970 auf Basis 10%-Stichprobe

Eingabefeld	Merkmal	Anonymisierungsmaßnahme
EF1	Geschlecht	
EF2	Geburtsdatum: Tag	gelöscht
EF3	Geburtsdatum: Monat	
EF4	Geburtsdatum: Jahr	
EF5	Familienstand	
EF6	Stellung zum Haushaltsvorstand	
EF7	Religion	
EF8	Sonstige religiöse Gemeinschaften	
EF9	Deutscher?	
EF10	Staatsangehörigkeit	
EF11	Andere Wohnung?	
EF12	Gang zur Arbeit?	
EF13	Leben in Wohnung?	
EF14	Behördlich gemeldet?	
EF15	Auslandswohnung?	
EF16	Überwiegender Lebensunterhalt	
EF17	Gegenwärtiger Schulbesuch	
EF18	Volksschulabschluss?	
EF19	Berufsschulabschluss?	
EF20	Mittlere Reife?	
EF21	Abitur?	
EF22	Berufsfachschulabschluss?	
EF23	Ingenieurschulabschluss?	
EF24	Hochschulabschluss?	
EF25	Erwerbstätig?	
EF26	Landwirt?	
EF27	Mithelfer im Familienbetrieb?	
EF28	Arbeitslos/-suchend?	
EF29	Nicht erwerbstätig?	
EF30	Hausfrau?	
EF31	Berufssoldat?	
EF32	Verkehrsmittel	
EF33	Zeitaufwand	
EF34	Geschäftszweig	
EF35	Stellung im Beruf	
EF36	Weitere Tätigkeit	
EF37	Arbeitszeit	
EF38	Wohnsitz am 1.9.1939	
EF39	Nach Kriegsende aus Sowjetzone zugezogen	
EF40	Vertriebenenausweis	
EF41	Frühere Erwerbstätigkeit	
EF42	Zeitpunkt Aufgabe frühere Erwerbstätigkeit	
EF43	Beruf	
EF44	Amtsbezeichnung bei Beamten	
EF45	Maschinenbedienung	
EF46	Monatliches Nettoeinkommen	
EF47	Leitende Stellung	
EF48	Bei Selbst: Anzahl Beschäftigter	
EF49	Lohnempfänger unter den Beschäftigten?	
EF50	Landwirtschaftlich genutzte Fläche	
EF51	Praktische Berufsausbildung?	

EF52	Abschluss prakt. Berufsausbildung?	
EF53	Erlerner Beruf (F32)	
EF54	Abschluss Technikerschule?	
EF55	Abschluss Berufsfachschule?	
EF56	Abschluss Ingenieurschule?	
EF57	Abschluss Hochschule?	
EF58	Dauer gesamte Techniker + Hochschulbildung	
EF59	Hauptfachrichtung des Abschlusses	
EF60	Jahr des letzten Abschlusses	
EF61	Eheschließungsjahr	
EF62	Frühere Ehe?	
EF63	Kinderlos?	
EF64	Geburtsjahr des 1. Kindes	
EF65	Geburtsjahr des 2. Kindes	
EF66	Geburtsjahr des 3. Kindes	
EF67	Geburtsjahr des 4. Kindes	
EF68	Geburtsjahr des 5. Kindes	
EF69	Geburtsjahr des 6. Kindes	
EF70	Geburtsjahr des 7. Kindes	
EF71	Geburtsjahr des 8. Kindes	
EF72	Geburtsjahr des 9. Kindes	
EF73	Geburtsjahr des 10. Kindes	
EF74	Geburtsjahr des 11. Kindes	
EF75	Geburtsjahr des 12. Kindes	
EF76	Telefon?	
EF77	Wohnsitzgemeinde / Land	
EF78	Gemeindegrößenklasse des Wohnortes (W2)	
EF79	Anstaltskennzeichen	
EF80	10%-Kennzeichen alle	
EF81	Anstaltskennzeichen	
EF82	10%-Kennzeichen	
EF83	Anstaltsart	
EF84	Zahl der Anstalten je Gemeinde	
EF85	Personentyp (ABCDE)	
EF86	Wohnbevölkerungs-Gruppen	
EF87	Bevölkerung in Privathaushalt/am Wohnsitz (W7,W8)	
EF88	Beteiligung am Erwerbsleben (W1)	
EF89	Landwirtschaftliche Bevölkerung? (W2)	
EF90	Differenzierung Ernährer/Ernährte (W1)	
EF91	Alter (W1)	
EF92	Anzahl der ledigen Personen unter 18 Jahre (W1)	
EF93	Anzahl der lebendgeborenen Kinder	
EF94	Sozio-ökonomische Gliederung	
EF95	Sozio-ökonomische Gliederung (Ausländer)	
EF96	Nettoerwerbseinkommen (erweitert)	
EF97	Vertriebenengruppe	
EF98	Haushaltstyp (W7)	
EF99	Ein-/Mehrfamilienhaushalt	
EF100	Anzahl Familien im Haushalt	
EF101	Haushaltsgröße: Wohnbereichsbevölkerung (W1)	
EF102	Haushaltsgröße: Wohnbevölkerung (W2)	
EF103	Haushaltsgröße: Bevölkerung in Privathaushalt + Anstalten (W7)	
EF104	Anzahl Einkommensbezieher im HH	
EF105	Anzahl Einkommensbezieher im Haushalt mit nur Deutschen	
EF106	Anzahl Einkommensbezieher im Haushalt mit Ausländern	

EF107	Maximales Einkommen im Haushalt mit nur Deutschen	
EF108	Überwiegender Lebensunterhalt der Haushaltsmitglieder	
EF109	Anzahl Erwerbspersonen im Haushalt	
EF110	Anzahl mithelfender Familienangehöriger/Haushalt	
EF111	Anzahl lediger Kinder im HH insgesamt	
EF112	Anzahl lediger Kinder 18 Jahre und älter	
EF113	Anzahl lediger Kinder unter 18 Jahre	
EF114	Anzahl lediger Kinder unter 15 Jahre	
EF115	Anzahl lediger Kinder unter 10 Jahre	
EF116	Anzahl lediger Kinder unter 6 Jahre	
EF117	Anzahl lediger Kinder unter 3 Jahre	
EF118	Familientyp (W7) 1 Byte	
EF119	Familientyp (W7) 2 Byte	
EF120	Familiengröße (W7)	
EF121	Anzahl Einkommensbezieher in Familie nur Deutsche (W8)	
EF122	Anzahl Einkommensbezieher in Familie mit Ausländern	
EF123	Maximales Einkommen in Familie mit nur Deutschen	
EF124	Anzahl ernährter Kinder/Familie	
EF125	Anzahl ernährter Kinder in Ausbildung/Familie	
EF126	Religionszugehörigkeit der Kinder	
EF127	Anzahl lediger Kinder in Familie	
EF128	Anzahl lediger Kinder in Familie 21 Jahre und älter	
EF129	Anzahl lediger Kinder in Familie 18-21 Jahre	
EF130	Anzahl lediger Kinder in Familie 15-18 Jahre	
EF131	Anzahl lediger Kinder in Familie 14-15 Jahre	
EF132	Anzahl lediger Kinder in Familie 10-14 Jahre	
EF133	Anzahl lediger Kinder in Familie 6-10 Jahre	
EF134	Anzahl lediger Kinder in Familie 4-6 Jahre	
EF135	Anzahl lediger Kinder in Familie 3-4 Jahre	
EF136	Anzahl lediger Kinder in Familie 2-3 Jahre	
EF137	Anzahl lediger Kinder in Familie 1-2 Jahre	
EF138	Anzahl lediger Kinder in Familie unter 1 Jahr	
EF139	Alter jüngstes Kind in Familie	
EF140	Geschl. des Ernährers (W1)	
EF141	Alter des Ernährers (W1)	
EF142	Beteiligung am Erwerbsleben des Ernährers	
EF143	Überwiegender Lebensunterhalt Ernährer	
EF144	Stellung im Beruf d. Ernährers (W1)	
EF145	Wirtschaftszweig d. Ernährers (W1)	
EF146	Einkommen des Ernährers	
EF147	Berufsgruppe des Ernährers	
EF148	Anzahl Ernährte männlich (W1)	
EF149	Anzahl Ernährte weiblich (W1)	
EF150	Anzahl Ernährte männlich (W2)	
EF151	Anzahl Ernährte weiblich (W2)	
EF152	Geburtsjahr des Ehepartners	
EF153	Alter des Ehepartners	
EF154	Ehepartner vor/nach Stichtag geboren	
EF155	Religion des Ehepartners	
EF156	Ehepartner schon vorher verheiratet?	
EF157	Beteiligung am Erwerbsleben des Ehepartners	
EF158	Einkommen des Ehepartners	
EF159	Überwiegender Lebensunterh. des Ehepartners	
EF160	Stellung im Beruf des Ehepartners	
EF161	Geschäftszweig des Ehepartners	

EF162	Schulabschluss des Ehepartners	
EF163	Schulbesuch des Ehepartners	
EF164	Vertriebenengruppe des Ehepartners	
EF165	Geburtsjahr des Familienvorstands (FV)	
EF166	Alter des Familienvorstands	
EF167	Geschlecht des Familienvorstands	
EF168	Familienstand des Familienvorstands	
EF169	Familienvorstand schon vorher verheiratet?	
EF170	Beteiligung am Erwerbsleben des Familienvorstands	
EF171	Überwiegender Lebensunterhalt des Familienvorstands	
EF172	Einkommen des Familienvorstands	
EF173	Stellung im Beruf des Familienvorstands	
EF174	Geschäftszweig des Familienvorstands	
EF175	Sozio-ökonomische Gliederung des Familienvorstands	
EF176	Schulbesuch des Familienvorstands	
EF177	Schulabschluss des Familienvorstands	
EF178	Vertriebenengruppe des Familienvorstands	
EF179	Familienstand der Mutter	
EF180	Beteiligung am Erwerbsleben der Mutter	
EF181	Stellung im Beruf der Mutter	
EF182	Geschäftszweig der Mutter	
EF183	Überwiegender Lebensunterhalt der Mutter	
EF184	Schulbesuch der Mutter	
EF185	Schulabschluss der Mutter	
EF186	Hausnummer-Zusatz	gelöscht
EF187	Prioritätsmerkmal	gelöscht
EF188	Belegart	gelöscht
EF189	Hausnummer-Zusatz	gelöscht
EF190	Prioritätsmerkmal	gelöscht
EF191	Belegart	gelöscht
EF192	Korrekturnummer	gelöscht
EF193	Bündelnummer	gelöscht
EF194	Zählerlistennummer	gelöscht
EF195	Anstaltslisten-Nr 1-Byte	gelöscht
EF196	Anstaltslisten-Nr 2-Byte	gelöscht
EF197	Zählbezirksnummer	gelöscht
EF198	Haushaltsnummer	gelöscht
EF199	Paginiernummer	gelöscht
EF200	VZ-Kenn-Nummer	gelöscht
EF201	Wohnsitzgemeinde/Kreis	gelöscht
EF202	Wohngemeinde/Gemeinde	gelöscht
EF203	Zielgemeinde/Land	gelöscht
EF204	Zielgemeinde/Gemeinde	gelöscht
EF205	Auspendler über Landesgrenzen	gelöscht