

Public Use Samples (PUS): Public-Use-Stichproben der schweizerischen Volkszählungen 1970, 1980, 1990 und 2000

Autorinnen und Autoren:

Christine Breitenstein, SIDOS

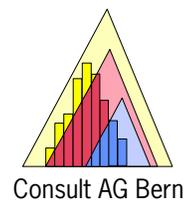
Dominique Joye, SIDOS

Monique Graf, BFS

Claude Joye, BFS

Rolf Kaufmann, Consult AG

S I D O S
S I D O S
S I D O S
S I D O S
S I D O S



VORWORT	4
1. EINFÜHRUNG.....	5
1.1 ÜBERBLICK.....	5
1.2 WOZU UND WEM DIENEN "PUBLIC USE SAMPLES" (PUS)?	6
1.3 WELCHE DATEN STEHEN ZUR VERFÜGUNG?	7
1.4 TEXTAUFBAU.....	7
2. VON DER VOLKSZÄHLUNG ZU DEN PUS.....	8
2.1 KURZABRISS DER GESCHICHTE DER SCHWEIZER VOLKSZÄHLUNG	8
2.1.1 <i>Inhalt</i>	8
2.1.2 <i>Grundlegende Definitionen</i>	8
2.1.3 <i>Methode</i>	11
2.1.4 <i>Akzeptanz</i>	12
2.1.5 <i>Auswertung der Fragebogen</i>	12
2.2 DIE VERWENDUNG DER VOLKSZÄHLUNGEN	13
2.2.1 <i>Einleitung</i>	13
2.2.2 <i>Publikationen des BFS</i>	14
2.3 METHODIK	15
2.3.1 <i>Einleitung: Vorteile und Nachteile von Stichproben wie PUS</i>	15
2.3.2 <i>Vier Datenebenen: Einzelpersonen, Haushalte, Wohnungen, Gebäude</i>	16
2.3.3 <i>Unmögliche und mögliche Analysen</i>	17
2.4 BEISPIELE AUS DEM AUSLAND.....	18
2.5 LITERATUR	19
3. STICHPROBEN UND DATENSCHUTZ.....	20
3.1. STICHPROBE.....	20
3.1.1 <i>Grundgesamtheit</i>	20
3.1.2 <i>Stichprobentechnik</i>	20
3.1.3 <i>Geographische Gebiete</i>	20
3.1.4 <i>Zielgrößen Haushalt, Gebäude und Wohnung – Grenzen der Auswertbarkeit</i>	21
3.2. METHODEN DES DATENSCHUTZES	21
3.2.1 <i>Angewandte Verfahren</i>	21
3.2.2 <i>Schutz der vorliegenden Stichproben</i>	22
3.3. STREUUNGEN	22
3.3.1 <i>Schätzen von Anteilen und Streuungen</i>	22
3.3.2 <i>Standardabweichung</i>	22
3.3.3 <i>Vertrauensintervalle</i>	23
3.3.4 <i>Anteile an der gesamten Population</i>	23
3.3.5 <i>Anteile in Subpopulation</i>	23
3.4. STREUUNGSKOMPONENTEN	23
3.5. LITERATUR	25
4. BESCHREIBUNG DER DATEN	26
4.1. VORWORT	26
4.2. DIE VARIABLENLISTEN	26
4.2.1 <i>Übersichtstabelle</i>	27
4.2.2 <i>Variablen nach Themen</i>	31
4.2.3 <i>Variablen in alphabetischer Reihenfolge</i>	37
4.3. DOKUMENTATIONSMODELL UND BEISPIEL.....	44
4.4. CODIERUNG DER FEHLENDEN DATEN	44
NÄHERE AUSKÜNFT.....	45

VORWORT

Die seit 1850 alle zehn Jahre durchgeführte Volkszählung ist ein Grundpfeiler der amtlichen Statistik der Schweiz. Sie liefert demographische Grunddaten sowie zeitlich und räumlich vergleichbare Schlüsselinformationen über die Gesamtheit der Einwohnerinnen und Einwohner, der Haushalte, der Arbeitsplätze und der Wohnungen.

Der Zugriff auf die Volkszählungsdaten ist jedoch nicht ohne Weiteres möglich. Einerseits ist eine Datei aus mehreren Millionen Einzeldatensätzen auch mit modernsten Informatikmitteln nur schwer zu verwalten, und andererseits muss die Vertraulichkeit der Daten unbedingt gewahrt werden.

Aus diesen Gründen stellt das Bundesamt für Statistik (BFS) in Zusammenarbeit mit dem wissenschaftspolitischen Rat für die Sozialwissenschaften (WRS) der Schweizerischen Akademie der Geistes- und Sozialwissenschaften der interessierten Öffentlichkeit Stichproben mit Einzeldatensätzen – so genannte “Public Use Samples” (PUS) – zur Verfügung. Diese PUS setzen sich aus Daten zusammen, die per Zufallsverfahren gezogen wurden und 5% der Bevölkerung abdecken. Sie können ohne Einschränkung verbreitet werden, da spezielle Massnahmen zur Gewährleistung der Anonymität getroffen wurden.

Die vorliegende CD-ROM enthält vier PUS. Die darauf gespeicherten Daten stammen aus den Volkszählungen 1970, 1980, 1990 und 2000.

Ziel der PUS ist es, allen interessierten Personen (Forschenden, Studierenden, wissenschaftlichen Mitarbeitenden usw.) die Analyse der Volkszählungsdaten im Rahmen eines “reduzierten Modells” zu ermöglichen, insbesondere für die Zwecke der Lehre und Forschung. Indem sie den vorhandenen Informationsreichtum sowie die Auswertungs- und Analysemöglichkeiten der Volkszählung aufzeigen, werden die PUS hoffentlich dazu beitragen, den Zugriff auf die Volkszählungsdaten zu vereinfachen.

Zahlreiche nationale statistische Ämter stellen bereits “Public Use Samples” für Forschende bereit. Zu erwähnen sind insbesondere Frankreich (mit der 5%-Stichprobe des statistischen Amtes INSEE), die USA, die Niederlande oder auch Kanada. Die Autorinnen und Autoren der vorliegenden Studie haben sich im Wesentlichen an das kanadische Beispiel gehalten. Bei den Anonymisierungsarbeiten zum Schutz der Mikrodaten dienten ihnen die Forschungsarbeiten der Expertengruppe für das europäische Projekt Esprit/SDC1 als Modell.

Mein Dank geht an die Autorenschaft, deren konstruktive Zusammenarbeit diese neue Art der Datendiffusion überhaupt erst ermöglicht hat.

Bundesamt für Statistik

Werner Haug

1. EINFÜHRUNG¹

1.1 ÜBERBLICK

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.² Dieses Verfahren wird in Kapitel 3 dieses Berichts detailliert beschrieben.

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.

Zwei Vorbemerkungen zur getroffenen Auswahl.

- Weshalb werden die Daten nahezu 5 Jahre nach der Volkszählung 2000 zur Verfügung gestellt? Die Aufbereitung von Volkszählungsdaten nimmt beträchtliche Zeit in Anspruch, da zahlreiche Kontrollen durchzuführen sind. Ausserdem lassen sich Daten, die vor einigen Jahren gesammelt wurden, einfacher anonymisieren als aktuelle Angaben, wie in Kapitel 3 noch zu sehen sein wird.
- 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

¹ Dieser Text stimmt weitgehend mit der Vorläuferversion «Public Use Samples (PUS), Public-Use-Stichproben der schweizerischen Volkszählungen» aus dem Jahr 2001 überein. Er wurde lediglich anhand der Informationen der Volkszählung 2000 aktualisiert.

² 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.

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.³ Die vorliegende Beschreibung soll einfach und verständlich sein und gleichzeitig Informationen aus zahlreichen Publikationen zusammenfassen.

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

Nicht alle Analysemöglichkeiten, die ein Datensatz bietet, sind im Voraus absehbar. Wir werden an einer späteren Stelle (Abschnitt 0) einige Beispiele aufführen und die stichprobenbedingten Grenzen erörtern. Ohne weitere Anwendungen auszuschliessen, seien hier einige Nutzungsmöglichkeiten erwähnt, an die wir bei der Erarbeitung dieser Dokumentation gedacht haben.

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.

³ Die nach jeder Volkszählung veröffentlichten Ergebnistabellen enthalten stets auch eine Beschreibung der verwendeten Indikatoren. Im vorliegenden Fall haben wir uns auf die Dokumentation zu den harmonisierten Datenblättern der Volkszählungen von 1970, 1980, 1990 und 2000 gestützt.

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.

1.3 WELCHE DATEN STEHEN ZUR VERFÜGUNG?

Die ausgewählten Variablen werden in Kapitel 4 näher beschrieben. Vorgängig möchten wir jedoch kurz die Grundidee hinter diesen Daten und ihrer Bereitstellung in Erinnerung rufen.

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. Im Abschnitt 3 werden die methodischen Aspekte dieser Auswahl besprochen. In der Fachsprache wird deshalb von einem "Public Use Sample" von Einzelpersonen und nicht von einem "Public Use Sample" von Haushalten gesprochen.

Die Stichprobe wird durch einfache Zufallsauswahl von Personen gebildet, die als repräsentative Stichprobe 5% der Bevölkerung abdeckt. Die Eigenschaften dieser Stichprobe werden in Kapitel 3 erläutert.

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.

1.4 TEXTAUFBAU

Die vorliegende Dokumentation behandelt alle einschlägigen Fragen im Zusammenhang mit der Verwendung dieser Daten. Zum Auftakt (Kapitel 2) werden die wesentlichen Merkmale einer Volkszählung rekapituliert, um mindestens ansatzweise die Besonderheiten, Stärken und Schwächen einer solchen Datensammlung aufzuzeigen und auf die Anpassungen im Laufe der Jahre hinzuweisen.

Als Zweites wird die Frage des Datenschutzes näher untersucht (Kapitel 3): Ein heikles und wichtiges Thema, nicht zuletzt darum, weil gewisse Entscheide in diesem Bereich direkte Auswirkungen auf die Analysemöglichkeiten haben. Der letzte Teil (Kapitel 4) enthält Variablentabellen, welche die verfügbaren Informationen aus unterschiedlicher Perspektive beleuchten.

Im Katalog der Variablen werden die einzelnen Merkmale vorgestellt. Im Vordergrund steht die Beschreibung ihrer Eigenschaften, daneben kommen aber auch die Harmonisierungsanstrengungen zur Sprache, die mit der gleichzeitigen Verwendung von Daten aus vier Volkszählungen verbunden sind. Schliesslich findet sich im Anhang eine Reihe von Dokumenten mit technischen Einzelheiten und Namen von Kontaktpersonen, die den Benutzerinnen und Benutzern bei einer geplanten Vertiefung der Datenanalysen beratend zur Seite stehen können.

2. VON DER VOLKSZÄHLUNG ZU DEN PUS

2.1 KURZABRISS DER GESCHICHTE DER SCHWEIZER VOLKSZÄHLUNG

Die Geschichte der Volkszählungen in der Schweiz ist bereits beschrieben worden (T. Busset, 1993) und soll hier nicht wiederholt werden. Wir möchten lediglich ein paar wichtige Punkte rekapitulieren.

Wie in den meisten europäischen Ländern wurden die ersten modernen Volkszählungen Mitte des 19. Jahrhunderts durchgeführt, zu einem Zeitpunkt, an dem der "Staat" im heutigen Sinne entstand.⁴ Dies zeigt deutlich, dass die Statistik eng mit dem Staat als solchem und mit der Beobachtung des sozialen Gefüges einer bestimmten Gesellschaft verknüpft ist.

2.1.1 Inhalt

Die Hauptthemen der Volkszählung sind im Laufe der Zeit mehr oder weniger gleich geblieben. Änderungen gab es vielmehr bei den Fragestellungen oder – noch ausgeprägter – bei der Art und Weise, wie die Daten erhoben und kodifiziert werden. Zur Erinnerung:

1. Bevölkerung: Hauptziel der Volkszählung ist die Ermittlung der Zahl der Wohnbevölkerung im Staatsgebiet. Verschiedene Faktoren können diesen Vorgang erschweren. Einerseits sind nicht alle Auskunftsspflichtigen geneigt zu antworten, und andererseits muss klar definiert sein, wer zur erhobenen Bevölkerung gehört.
2. Beschäftigung: Die Bevölkerung soll nicht nur gezählt, sondern es soll auch ermittelt werden, wie sie ihren Lebensunterhalt bestreitet und zur Wertschöpfung des Landes beiträgt. T. Busset (1993) unterstreicht die deutliche Entwicklung dieses Analysebereichs zu Beginn des 20. Jahrhunderts.⁵
3. Demografie: Dieses Thema erhielt um 1940 eine vordringliche Bedeutung, sahen sich die Statistiker (nach T. Busset, 1993) doch veranlasst, Alarm zu schlagen und die Politiker auf das zunehmende Ungleichgewicht zwischen Geburten und Todesfällen hinzuweisen.

Diese wenigen Zeilen machen bereits deutlich, dass die Volkszählung nicht nur als Zählinstrument dient, sondern auch eine Art Sozialberichterstattung impliziert. Neben dem bereits erwähnten Beispiel der Demographie sei in diesem Zusammenhang auch auf die Wohnungserhebung verwiesen, die um 1920 – als eine gravierende Wohnungsnot herrschte – in die Volkszählung aufgenommen wurde (T. Busset, 1993, S. 68). Als weitere Priorität kam ab den 1960er Jahren die Analyse des Pendlerverkehrs hinzu.⁶

2.1.2 Grundlegende Definitionen

Auf den ersten Blick mag eine Volkszählung als etwas Einfaches erscheinen. Schon beim Versuch, die zu erfassende Bevölkerung genau einzugrenzen, tauchen jedoch Probleme auf. Der Einsatz von Definitionen ist deshalb unerlässlich.⁷

⁴ Weitere Informationen über die Volkszählungen in Europa und den Zusammenhang mit dem Aufbau eines Verwaltungsapparates finden sich insbesondere in zwei vom INSEE herausgegebenen Bänden zu diesem Thema (1987a et 1987b).

⁵ Neben der Volkszählung erlaubt auch die Schweizerische Arbeitskräfteerhebung (SAKE) detaillierte Untersuchungen zu Entwicklung und Stand der Beschäftigung in der Schweiz. Für Einzelheiten zur Schweizerischen Arbeitskräfteerhebung siehe BFS, 2004 und R. Feusi Widmer, 2004.

⁶ Die Bedeutung dieses Aspektes zeigt sich schon an der Tatsache, dass im Jahr 2000 jede zweite erwerbstätige Person ausserhalb ihrer eigenen Wohngemeinde arbeitete.

⁷ Busset spricht im Zusammenhang mit der Bevölkerungsdefinition von einer «Kernfrage» und widmet sich diesem Thema ausführlich (T. Busset, 1993, S. 61ff.)

BEVÖLKERUNG

1970 wurde die Wohnbevölkerung in Artikel 5 der Verordnung über die eidgenössische Volkszählung wie folgt definiert⁸: «Zur Wohnbevölkerung einer Gemeinde gehört jede Person, die sich im Gemeindegebiet mit der Absicht dauernden Verbleibens aufhält oder aufzuhalten gedenkt. Über besondere Fälle entscheidet das Eidgenössische Statistische Amt».

«Im Allgemeinen fällt der Wohnsitz mit dem Ort zusammen, wo jemand die Schriften hinterlegt hat und die Steuern bezahlt. Von diesem Grundsatz wurde aber wie bisher für einzelne Personengruppen abgewichen. So wurden Familienmitglieder, die auswärts studierten oder arbeiteten (mit Ausnahme der verheirateten Berufstätigen) und nicht täglich nach Hause zurückkehrten, in jener Gemeinde als wohnhaft betrachtet, von der aus sie täglich zur Schule oder zur Arbeit gingen. Ebenso wurden die Insassen von Armen- und Waisenanstalten, Altersasylen, Bürgerheimen, Klöstern und Zwangserziehungsanstalten, ebenso wie die in Internaten von Erziehungs- und Unterrichtsinstituten lebenden Zöglinge der Bevölkerung jener Gemeinde zugeordnet, in der sich die betreffende Anstalt befand.

Personen, die weder berufstätig waren noch eine Schule besuchten, galten selbst dann als in der Zählgemeinde wohnhaft, wenn die Absicht dauernden Verbleibens nicht bestand, sobald sie sich seit mehr als 6 Monaten in dieser Gemeinde aufgehalten hatten. Diese Frist von 6 Monaten galt vor allem für Kurgäste, Ferienkinder sowie die Insassen folgender Anstalten: Spitäler, Heil- und Pflegeanstalten, Kinderheime, Gefängnisse und Strafanstalten».⁹

Nach den Richtlinien des Bundesamtes für Statistik werden Personen ohne ständigen Wohnsitz in der Gemeinde erfasst, in der sie zum Zeitpunkt der Erhebung niedergelassen sind.

«Geschäftsreisende, Monteure und übrige Personen, die einen Beruf ausübten, der einen häufigen Wechsel des Aufenthaltsortes bedingt, wurden, auch wenn sie nicht verheiratet waren, am Wohnsitz ihrer Angehörigen gezählt, wenn sie noch nicht sechs Monate am Aufenthaltsort wohnten.

Alle berufstätigen Ausländer, die am 1. Dezember in der Schweiz arbeiteten und eine Arbeitsbewilligung besaßen, also auch die Saisonarbeiter, wurden zur Wohnbevölkerung unseres Landes gezählt. [...] Zur Wohnbevölkerung einer Gemeinde rechnete man auch die an einer schweizerischen Schule eingeschriebenen Schüler und Studenten ausländischer Herkunft, die Flüchtlinge, ebenso alle nicht berufstätigen Personen aus dem Ausland, die vor dem 1. Juni in die Schweiz eingereist waren und sich seither ununterbrochen hier aufgehalten hatten. Umgekehrt wurden Personen, die zum Studium oder zur Ausübung ihrer Berufstätigkeit aus der Schweiz ausgereist waren, vom ersten Tag an nicht mehr zur Wohnbevölkerung der Schweiz gezählt, wenn es sich nicht um eine Geschäftsreise oder Montage für eine Firma mit Sitz in der Schweiz handelte. Für andere Personen galt auch hier die 6-Monate-Grenze».¹⁰

Eine Person kann dabei mehrere Wohnsitze haben. In der Volkszählung von 1990 wurde zum ersten Mal zwischen wirtschaftlichem und zivilrechtlichem Wohnsitz unterschieden. Diese Unterscheidung wurde in der Volkszählung 2000 beibehalten.

- «Als wirtschaftlicher Wohnsitz einer Person gilt die Gemeinde, in der sie während des grössten Teils der Woche wohnt, deren Infrastruktur sie

⁸ Verordnung über die eidgenössische Volkszählung 1970, zitiert in: «Eidgenössische Volkszählung 1970. Wohnbevölkerung der Gemeinden 1850-1970», 1971. Heft 467. S. 5ff. Bern: Bundesamt für Statistik.

⁹ Quelle: 1970: *Definition der Wohnbevölkerung* (Heft 467, S. 6-8).

¹⁰ Quelle: 1970: *Definition der Wohnbevölkerung* (Heft 467, S. 6-8).

hauptsächlich beansprucht und von der aus sie den Weg zur täglichen Arbeit oder zur Schule aufnimmt.

- Als zivilrechtlicher Wohnsitz einer Person gilt die Gemeinde, in der sie mit dem Heimatschein angemeldet ist, die Steuern bezahlt und die politischen Rechte ausübt bzw. für die ihr, wenn sie über eine ausländische Staatsbürgerschaft verfügt, die entsprechende Bewilligung ausgestellt worden ist».¹¹

Seit der Volkszählung 1990 ist deshalb auf Grund der Verordnung vom 26. Oktober 1988 über die eidgenössische Volkszählung 1990 der wirtschaftliche Wohnsitz für alle Belange im Zusammenhang mit der Erhebung ausschlaggebend.¹² Obwohl mit der Volkszählung von 1990 der doppelte Wohnsitz eingeführt wurde, ist der wirtschaftliche Wohnsitz de facto bereits seit 1941 entscheidend. Den hier berücksichtigten Erhebungen liegt also eine kohärente Definition des Wohnsitzes zu Grunde.¹³ In der Volkszählung 2000 war dieser Punkt praktisch gleich geregelt wie in der vorherigen Erhebung.¹⁴

Wie bereits erwähnt, wird gleichzeitig mit der Volkszählung auch die Gebäude- und Wohnungserhebung durchgeführt. Sie erfasst nicht alle Gebäude, sondern lediglich solche, die für Wohnzwecke vorgesehen sind oder zum Zeitpunkt der Erhebung so genutzt werden. Deshalb muss auch hier klar definiert werden, welche Gebäude und Wohnungen für die Volkszählung berücksichtigt werden.

GEBÄUDE

Nach der für die Volkszählung verabschiedeten Definition gilt als Gebäude «jedes freistehende oder durch Brandmauern von einem anderen getrennte Bauwerk, das am Stichtag bewohnt oder bewohnbar ist».¹⁵

«Bei Doppel-, Gruppen- und Reihenhäusern zählt jedes Einzelne von dem anderen durch eine vom Keller bis zum Dach reichende Trennmauer geschiedene Bauwerk als selbständiges Gebäude. Nach dieser Definition sind auch Gebäude zu erfassen, die keine Wohnungen im eigentlichen Sinne, sondern nur Räume zur Unterbringung und Verpflegung von Personen enthalten können, wie Alters- und Pflegeheime, Bürgerheime, Waisen- und Armenhäuser, Erziehungsinstitute und Unterrichtsanstalten mit Internaten, Klöster, Zwangserziehungsanstalten, Hotels und ihre Dependancen, Gasthöfe, Pensionen, Spitalbauten, Kliniken, Ferien und Erholungsheime, Kinderheime sowie Einrichtungen für den Straf- und Massnahmenvollzug.

Definitionsgemäss sind somit Kultusgebäude und Ökonomietrakte von der Zählung ausgeschlossen, ebenso Fabrik- oder Verwaltungsgebäude, Geschäftshäuser, Turn- oder Sporthallen, Hallenbäder und Schulhäuser, sofern sie nicht bewohnt sind oder mindestens eine leer stehende Wohnung enthalten. Weiter eingeschränkt wurde die Zahl der Gebäude dadurch, dass sonstige Unterkünfte wie Baracken, Maiensäss- und Alpgebäude, Gartenhäuschen, Abbruchobjekte oder wegen ihres baulichen Zustandes oder ihrer

¹¹ Quelle: 1990: *Definition der Wohnbevölkerung* (Heft 9012, S. X und Heft 9048, S. XIII).

¹² Obwohl in der Volkszählung 1990 der wirtschaftliche Wohnsitz ausschlaggebend war, wurden auch Daten zum zivilrechtlichen Wohnsitz erhoben. Deshalb kann das BFS Interessierten auch die Originaldaten der Volkszählung 1990 mit den Angaben zum zivilrechtlichen Wohnsitz als Referenzkriterium liefern.

¹³ Mit einer Ausnahme: 1970 und 1980 wurden Ehepaare zusammen erhoben, auch wenn ein Ehepartner den wirtschaftlichen Wohnsitz in einer anderen Gemeinde hatte. Diese Bestimmung wurde 1990 aufgehoben. Betroffen sind etwa 1.5% – sicherlich vorwiegend getrennt lebende – Ehepaare.

¹⁴ Selbstverständlich war auch die Volkszählung 2000 Gegenstand einer eigenen Verordnung (vom 13. Januar 1999).

¹⁵ Quelle: 1990: *Gebäude- und Wohnungserhebung: Definitionen* (Heft 9012, S. X-XI).

Abgeschiedenheit nicht mehr bewohnbare Gebäude nur unter die Erhebung fielen, wenn sie im Zeitpunkt der Zählung bewohnt waren».¹⁶

WOHNUNGEN

«Auch der Begriff Wohnung unterliegt im Rahmen dieser Statistik gewissen Einschränkungen. Als Wohnung gilt die Gesamtheit der Räume, die zur Unterbringung einer oder mehrerer Privathaushalte (deren Definition sich im Laufe der Zeit geändert hat) bestimmt sind, und die zum Zeitpunkt der Zählung nicht ausschliesslich anderen als Wohnzwecken dienen. Von der Zählung ausgeschlossen sind somit die zweckentfremdeten Wohnungen (Büros, Arzt-, Rechtsanwaltspraxen usw.). Erfasst wurden andererseits Gebäude, in denen Kollektivhaushaltungen untergebracht waren (oder Gebäude, die zur Unterbringung von solchen bestimmt sind); doch wurden deren Räumlichkeiten nicht als Wohnungen betrachtet. Ebenfalls nicht als Wohnungen im Sinne der Zählung gelten bewohnte Räumlichkeiten in sonstigen oder mobilen Unterkünften».¹⁷

HAUSHALTE

In den letzten vier Volkszählungen war die Beziehung zwischen Privathaushalten und Wohnungen nicht einheitlich geregelt. Wurde der Privathaushalt 1970 und 2000 als "Einheit aller Personen, die in der gleichen Wohnung leben" definiert, konnte gemäss der in den Erhebungen von 1980 und 1990 verwendeten Definition "eine Wohnung mehrere Haushalte umfassen". Obschon der Bestand der kleinen Wohnungen durch diese Korrektur der Definition geringe Veränderungen erfahren haben könnte, wird die Vergleichbarkeit der Daten aus den vier Erhebungen insgesamt nicht beeinträchtigt, denn fast alle Wohnungen (nahezu 99%) umfassen nur einen Haushalt".

2.1.3 Methode

Die für eine Volkszählung anzuwendende Methode ist auf den ersten Blick einfach: es werden alle Einwohnerinnen und Einwohner, Wohnungen oder anderen Einheiten eines Landes gezählt. De Facto war aber der Aufwand für die Durchführung einer zuverlässigen Erhebung zu einem bestimmten Zeitpunkt schon immer beträchtlich. Überdies haben das Streben nach einem optimalen Kosten-Nutzen-Verhältnis, bei dem insbesondere den immer umfassenderen administrativen Personenregistern Rechnung getragen wird, sowie der Wunsch gewisser Länder, über laufend aktualisierte Daten zu verfügen, in den meisten europäischen Ländern zu einer eigentlichen Revolution bei der Durchführung von Volkszählungen geführt.

In der Schweiz kamen bis und mit 1990 für die Volkszählung «Zählpersonen» zum Einsatz, welche die Fragebogen bei den Haushalten vorbeibrachten und sie nach einiger Zeit wieder abholten, wobei sie sich gleichzeitig von der Vollständigkeit der Daten überzeugten. Bei der Erfassung der Bevölkerung und des Gebäude- und Wohnungsbestandes kam es 2000 zu einer grundlegenden Neuerung: Die meisten Fragebogen wurden mit der Post versendet und kamen auf diesem Weg zurück. Das war für die Gemeinden zwar mit beträchtlichen Vorbereitungsarbeiten verbunden, mussten sie doch die für die Versendung der Fragebogen erforderlichen Adressen bereitstellen. Andererseits waren die Fragebogen aber bereits mit den wichtigsten Merkmalen versehen, die in den Einwohnerregistern der Gemeinden erfasst sind. Diese Neuerung markierte in der Schweiz den Beginn der systematischen Nutzung administrativer Personenregister für eine erleichterte Durchführung der Volkszählungen. Diese Erhebungsmethode soll im Rahmen der Volkszählung 2010 weiter ausgebaut werden.

¹⁶ Quelle: 1970: *Gebäude- und Wohnungserhebung: Definitionen* (Heft 537, S. 7-8).

¹⁷ Quelle: 1980: *Gebäude- und Wohnungserhebung: Definitionen* (Heft 705, S. 7-8, Heft 706, S. 7-8, Heft 707, S. 7-8 und Heft 713, S. 9-10).

Mit der Volkszählung 2000 wurde eine weitere «Revolution» eingeführt: Die Möglichkeit, die Fragebogen im Internet auszufüllen. Ohne dass diese neue Erhebungsmethode besonders gefördert worden wäre, machten bereits 4,2% der Bevölkerung von dieser Möglichkeit Gebrauch. Schliesslich wurde die Aufbereitung der Daten zentralisiert, womit Grössenvorteile (sog. Skaleneffekte) genutzt werden konnten.

2.1.4 Akzeptanz

Neben einer sorgfältigen Durchführung bildet auch die Akzeptanz der Volkszählung bei der Bevölkerung ein zentrales Qualitätskriterium. Dabei ist daran zu erinnern, dass es bei allen Volkszählungen gewisse Probleme hinsichtlich der Akzeptanz der Zählpersonen bei der Bevölkerung gab, da die Volkszählung teilweise als starke Einmischung des Staates in die Privatsphäre betrachtet wurde. Gewisse Personen befürchteten gar eine Verbindung zwischen Steuerbehörden und Volkszählung, zumindest wenn man einer Karikatur im Nebenspalter Glauben schenkt.¹⁸ Die meisten diesbezüglichen Probleme ergaben sich sicher 1990, im Anschluss an die «Fichenaffäre». Selbst damals war jedoch der Grad der Erfassung sehr hoch, nicht zuletzt dank eines beträchtlichen Aufwands der Gemeindebehörden zur Einbringung fehlender Daten. Die Volkszählung 2000 verlief dagegen ohne solche Zwischenfälle. Die Verteilung und Rücksendung der Fragebogen per Post wurde von der Bevölkerung sehr gut aufgenommen. Der Verzicht auf Zählpersonen erforderte allerdings spezifische Massnahmen, um die Vollständigkeit der Antworten sicherzustellen (telefonische Rückfragen).

2.1.5 Auswertung der Fragebogen

Die Daten der Volkszählung stellen insofern einen Sonderfall dar, als sie in mehreren Schritten aufbereitet werden müssen, um Richtigkeit und Vollständigkeit zu gewährleisten. Dieses Verfahren soll hier nicht detailliert beschrieben werden, wir möchten aber die wichtigsten der 2000 vorgenommenen Schritte kurz skizzieren.

Nach der Rücksendung der Fragebogen durch die Befragten wurde in den Verarbeitungszentren an den Formularen eine Reihe von Kontrollen durchgeführt, bevor sie für das Einscannen weitergeleitet wurden. Die elektronischen Bilder der Fragebogen wurden nun für die Datenerfassung verwendet (elektronische Texterkennung, manuelle Erfassung).¹⁹

Dann wurden verschiedene Plausibilitätstests durchgeführt, um die Daten hinsichtlich einer gewissen Logik zu prüfen. Vorrangiges Ziel der Kontrollen war es, die Vollständigkeit der Daten sicherzustellen. In der Volkszählung 2000 war die Angabe einer Reihe von Merkmalen obligatorisch. Wenn diese Angaben unvollständig waren, wurden die Befragten telefonisch kontaktiert, um die fehlenden Informationen einzuholen. Weitere Tests dienten dazu, unlogische oder widersprüchliche Daten aufzufindig zu machen. Diese Daten wurden darauf mittels automatischer Korrekturverfahren oder Einsetzungsmethoden (Ergänzung mit Hilfe ähnlicher Antworten nach der Methode des nächsten Nachbarn) ergänzt. Zum Beispiel wurde bei der automatischen Korrektur einer fehlenden Angabe zur Hauptsprache eines Kleinkindes die Hauptsprache der Mutter übernommen.

Die Arbeit umfasst jedoch noch weitere Schritte, da gewisse Variablen aus den bereinigten Informationen gebildet werden. Dies ist namentlich bei der Wahl der Referenzperson des Haushalts der Fall, die bei mehreren in Frage kommenden Personen in erster Linie auf Grund der beruflichen Tätigkeit und anhand des Beschäftigungsgrads bestimmt wird.

Damit die Daten verschiedener Volkszählungen verglichen werden können, mussten schliesslich umfangreiche Harmonisierungsarbeiten vorgenommen werden. In der Zeitspanne zwischen 1970 und 2000 haben nämlich zahlreiche Veränderungen stattgefunden. Zum Beispiel:

¹⁸ Diese Karikatur ist in der Publikation zur Geschichte der Volkszählungen abgebildet (T. Buset, 1993, S. 81).

¹⁹ Vor der Volkszählung 2000 waren diese Schritte von den Gemeinden vorgenommen worden.

1. Die Zahl der Gemeinden hat sich verändert: Es gab Zusammenschlüsse (in der Mehrzahl der Fälle) und Spaltungen; und es wurde sogar ein neuer Kanton gegründet – ein äusserst seltenes Ereignis in der Schweizer Geschichte.
2. Teilweise wurden die Fragen abgeändert, was im Übrigen auch für die Erfassung der Antworten gilt. So war zum Beispiel eine geringe Abänderung in der Frage zur Ausbildung bei der Analyse der Ergebnisse nicht ohne Auswirkungen (D. Joye und M. Schuler, 1995). Auch waren die bei den Pendlerbewegungen zur Auswahl stehenden Kategorien zur Einordnung der Transportmittel 2000 wesentlich zahlreicher als 1970.
3. Um der Entwicklung Rechnung zu tragen, wurden auch die Nomenklaturen geändert. So gab es im Informatikbereich 2000 andere Berufsbezeichnungen als 1970, und in gewissen Industriezweigen verschwanden bestimmte Berufe im Laufe der erfassten Zeitspanne beinahe.

Die hier vorgestellten Daten wurden diesen Bearbeitungsschritten unterzogen. Die Informationen im Kapitel 4 liefern dazu weitere Einzelheiten.

2.2 DIE VERWENDUNG DER VOLKSZÄHLUNGEN

2.2.1 Einleitung

Die Volkszählung weist neben ihrem Nutzen als Werkzeug für die Entscheidungsfindung²⁰ und für Prognosen sehr spezifische Eigenschaften auf:

Erstens liefert einzig die Volkszählung einen Datensatz, der vollständig nach Regionen aufgeschlüsselt werden kann. Im europäischen Vergleich weist die Schweiz ein sehr feines Netz von mitunter sehr kleinen Gemeinden auf. Die Volkszählung stellt diesbezügliche Informationen zur Verfügung und ist praktisch das einzige Instrument zur Messung soziodemografischer Komponenten in den verschiedenen Gemeinden. Zudem sind auf der Grundlage der Volkszählungsdaten weitere räumliche Analysen möglich, entweder durch eine Gruppierung von Gemeinden²¹ oder durch ein georeferenziertes System auf der Grundlage des Hektarrasters.²² Auf jeden Fall bietet die durch die Volkszählungsdaten mögliche Kombination von geografischen Angaben und sozialen Strukturen ein weites Feld von Analyseansätzen.

Dank den vollständigen Angaben zu den Regionen lassen sich auch Matrizen erstellen, welche verschiedene Bewegungen beschreiben: Migrationsströme zwischen Geburtsort oder Wohnort vor fünf Jahren und heutigem Wohnsitz oder auch Ströme zwischen Wohnort und Arbeitsort. Denkbar wären analog auch soziale Mobilitätstabellen, in denen die soziale Stellung der verschiedenen Haushaltsmitglieder hinsichtlich bestimmter Muster untersucht und vorhandene soziale Schranken erfasst würden. Die Rekonstruktion von Strömen im geografischen und sozialen Raum erfordert Datenmengen, die über das Volumen dessen hinausgehen, was von gängigen Umfragen bereitgestellt wird.

Die Volkszählung dient auch als Vergleichs- und Validierungsinstrument für Erhebungen. Angesichts der Vielzahl an Umfragen verschiedenster Art verweigern immer mehr Leute eine Auskunft. Da in der Volkszählung die gesamte Bevölkerung erfasst wird, dienen die damit erhobenen Daten als Validierungsgrundlage für Erhebungen oder als Gewichtungselement in Bezug auf die soziale Stellung.

²⁰ Es soll zum Beispiel daran erinnert werden, dass die Ergebnisse der Volkszählung dem Bundesrat vorgelegt werden und als Grundlage für die Berechnung der Anzahl Sitze der einzelnen Kantone im Nationalrat dienen.

²¹ Agglomerationen, MS-Regionen, Typologie der Gemeinden usw. (M. Schuler und C. Jemelin, 1996).

²² Aus Datenschutzgründen sind in diesem PUS nicht alle Angaben berücksichtigt. In den Originaldaten ab der Volkszählung 1990 sind diese Informationen aber enthalten.

Die Volkszählung liefert daneben auch Daten über Randgruppen. Wie könnte zum Beispiel eine Umfrage ein realistisches Bild der Haushalte zeichnen, in denen drei Generationen unter einem Dach wohnen? Wie soll eine geografische oder soziale Verteilung eines spezifischen Berufs in Erfahrung gebracht werden? Es gibt zahlreiche Beispiele, in denen die Volkszählung eine unersetzliche Grundlage darstellt. Ihre Ergebnisse können insbesondere den Publikationen des BFS zu diesem Thema entnommen werden.

2.2.2 Publikationen des BFS

Die Publikationen, die das BFS im Rahmen der Auswertung der Volkszählungsdaten 2000 veröffentlicht hat, lassen sich in drei Serien einteilen:

- Rote Serie: Sechs Publikationen stellen die Resultate in Form von Tabellen, Karten und Grafiken dar und enthalten als Beilage eine CD.
 - Bevölkerungsentwicklung der Gemeinden 1850-2000
 - Bevölkerungsstruktur, Hauptsprache und Religion
 - Gebäude, Wohnungen und Wohnverhältnisse
 - Haushalte und Familien
 - Raumgliederung in der Schweiz
 - Erwerbsleben, Pendlermobilität und Ausbildung
- Grüne Serie: Sie besteht aus 11 Publikationen, welche die grossen Themen der Volkszählung abdecken (diese Publikationen sind in deutscher und französischer Sprache erschienen).
 - Räumliche und strukturelle Bevölkerungsdynamik der Schweiz 1990-2000
 - Migration und Integration: Ausländerinnen und Ausländer in der Schweiz
 - Pendlermobilität in der Schweiz
 - Bildung und Arbeit: Entwicklung des Arbeitsmarktes in der Schweiz
 - Wohnversorgung und Wohnverhältnisse: Entwicklungen 1990-2000
 - Metropolisierung, Städte und Agglomerationen: Soziodemografische Struktur und Dynamik von urbanen Räumen
 - Religionslandschaft der Schweiz
 - Alter und Generationen: Das Leben in der Schweiz ab 50 Jahren
 - Fortschritte und Stagnation in der Gleichstellung der Geschlechter 1970-2000
 - Sprachenlandschaft in der Schweiz (auch in italienischer Sprache erschienen)
 - Familiäre Lebensformen im Wandel
- Gelbe Serie: Diese Publikationen ergänzen die Grüne Serie mit spezifischeren Themen auf der Basis von komplexen Analysen (diese Publikationen sind nur in der Originalsprache verfügbar).
 - Die Integration der ausländischen zweiten Generation und der Eingebürgerten in der Schweiz (zweisprachig Deutsch/Französisch)
 - Entwicklung der Sozialstruktur
 - Der Übergang von der Ausbildung in den Beruf
 - Familles et migration

- Interne Migration in der Schweiz
- Migrants et marché du travail
- Les générations face au marché du travail
- Arbeitsteilung in Paarhaushalten
- Der berufsstrukturelle Wandel der Beschäftigung in der Schweiz 1970-2000
- Facteurs influençant le comportement reproductif des Suissesses et des Suisses
- Les ségrégations sur le marché suisse du travail
- Die aktuelle Lage des Romanischen (auch in romanischer Sprache erhältlich)
- Die Bergregionen in der Schweiz
- Bevölkerungszusammensetzung, Integration und Ausgrenzung in den urbanen Zonen

2.3 METHODIK

2.3.1 Einleitung: Vorteile und Nachteile von Stichproben wie PUS

Die Daten einer Stichprobe bieten verschiedene Vorteile. Dazu gehört die elektronische Verarbeitung: Selbst ein leistungsfähiger PC benötigt zum Erstellen einer Kreuztabelle mit den Daten der sieben Millionen an der Volkszählung 2000 beteiligten Personen noch immer mehrere Minuten. Dieser Zeitaufwand für die Analyse mag gering erscheinen, vor allem wenn berücksichtigt wird, dass noch vor fünfzehn Jahren nur ein mehrere Millionen teures Rechenzentrum eine solche Aufgabe hätte bewältigen können. Trotzdem ist diese Verarbeitungszeit im Vergleich zu gängigen Applikationen lang.

Der zweite Vorteil liegt in der sofortigen Verfügbarkeit: Während für die vollständige Datensammlung ein Vertrag unterschrieben werden muss, ist die hier vorgestellte Stichprobe frei zugänglich, da die Anonymität sichergestellt ist.

Selbstverständlich stehen dem Vorteil eines freien Zugangs auch verschiedene Nachteile gegenüber:

1. Die Genauigkeit ist geringer – allerdings nur beschränkt: Verschiedenste wissenschaftliche Studien basieren auf einer Stichprobe von lediglich tausend Personen und ermöglichen trotzdem aussagekräftige Prognosen. Die Stichprobe zu den Volkszählungen ist 350 Mal grösser und gewährleistet somit auch eine höhere Genauigkeit.²³ Zudem besteht praktisch keine Stichprobenverzerrung, da bei Volkszählungen wesentlich höhere Antwortquoten als bei anderen Umfragen erreicht werden.
2. Wesentlich einschneidender ist die Elimination gewisser Detaildaten der Variablen. Eine der herausragenden Qualitäten der Volkszählung besteht darin, dass detaillierte Analysen möglich sind, zum Beispiel nach Berufen oder Branchen. Diese Informationen mussten aus Datenschutzgründen zum Teil entfernt werden. Auch geografische Detailinformationen gehören zu den besonderen Qualitäten der Volkszählungsdaten. Die Sicherstellung der Anonymität setzt dabei jedoch Grenzen, weshalb sich mögliche Analysen auf Kantone und grosse räumliche Typologien beschränken.

²³ Zur Erinnerung: Die Genauigkeit ist abhängig von der Quadratwurzel des Stichprobenumfangs und steht bei gleicher Varianz nicht in Zusammenhang mit der Grösse der beschriebenen Population.

3. Wenn vollständige Daten vorhanden sind, können gleichzeitig Analysen auf verschiedenen Ebenen durchgeführt werden, zum Beispiel für Einzelpersonen und Haushalte, was bei einer Stichprobe nicht unbedingt möglich ist.

2.3.2 Vier Datenebenen: Einzelpersonen, Haushalte, Wohnungen, Gebäude

Wie bereits erwähnt, beschränkt sich die Volkszählung nicht auf das Zählen von Einzelpersonen, sondern umfasst zusätzlich Informationen über Haushalte, Wohnungen und allenfalls Gebäude, wenn sich in diesen mindestens eine Wohnung befindet. Ein Kapitel befasst sich mit der detaillierten Beschreibung der im PUS verfügbaren Variablen. Wichtig sind diese Daten über Haushalte und Wohnungen, weil sich damit eine Einzelperson einem sozialen Umfeld zuordnen lässt und Angaben zu den Lebensbedingungen erschlossen werden.

Deshalb ist die Versuchung gross, die PUS als Stichproben von Haushalten oder Wohnungen zu verwenden. Aus statistischer Sicht ist jedoch vor einer solchen Verwendung zu warnen:

1. Einerseits besteht ein Gewichtungproblem: Ein von mehreren Personen bewohntes Gebäude ist mit einer grösseren Wahrscheinlichkeit im PUS berücksichtigt als ein Gebäude, in dem nur eine Person lebt. Dasselbe gilt für die Haushalte. Zudem kann hier kein Gewichtungskoeffizient angegeben werden, da sich daraus die Zahl der im Haushalt oder in der Wohnung lebenden Personen ableiten liesse. Genau diese Information wurde aus Datenschutzgründen eliminiert.
2. Im Zusammenhang mit der Anonymisierung ergibt sich noch ein weiteres Problem: Die Anonymisierung wurde im Hinblick auf eine Stichprobe mit Daten zu Einzelpersonen vorgenommen; die Auswirkungen auf die Struktur der Haushalte oder der Wohnungen wurden nicht berücksichtigt.
3. Schliesslich basiert die Stichprobe immer auf den Daten zu den Einzelpersonen. Dies bedeutet, dass sich keine Daten zum Haushalt rekonstruieren lassen, mit Ausnahme der Variablen zur Struktur.

Da die der Stichprobe zu Grunde liegenden Daten ausschliesslich auf den Angaben zu den Einzelpersonen und nicht zu den Haushalten basieren, ergeben sich gewisse Folgen für die Auslegung. Ein fiktives Beispiel: Eine Gesellschaft sei ausschliesslich aus Familien mit zwei Elternteilen und Kindern gebildet. Bei der einen Hälfte der Haushalte sei der Mann Referenzperson, bei der andern die Frau. Aus verschiedenen Gründen könnten Haushalte mit einem Mann als Referenzperson drei Kinder und Haushalte mit einer Frau als Referenzperson nur ein einziges Kind haben. Daraus ergäbe sich, dass 62,5% der Personen in Privathaushalten einen Mann und lediglich 37,5% eine Frau als Referenzperson haben. Diese Art der Stichprobenentnahme ist deshalb für Angaben in Bezug auf Einzelpersonen geeignet, nicht aber für Informationen zu den Haushalten.

Die Daten zu den Haushalten und Wohnungen liefern demnach ausschliesslich Rahmeninformationen zu den Lebensbedingungen und dürfen nicht isoliert betrachtet werden. Dies ändert jedoch nichts daran, dass die Daten einen Informationswert in Bezug auf Lebensweise und Umfeld haben: Die Kombination von individuellen und kontextuellen Informationen in einem Datensatz hat immer eine grosse Bereicherung dargestellt, die übrigens in der Vergangenheit vielleicht nicht immer voll ausgeschöpft wurde.²⁴

Der Einbezug von Informationen aus anderen Bereichen erfordert zusätzliche Massnahmen im Bereich der Antwortausfälle. So haben zum Beispiel Informationen über Wohnungen nicht dieselbe Bedeutung für Privathaushalte wie für Kollektivhaushalte: Um hervorzuheben, dass ein Vergleich

²⁴ Die Berücksichtigung von regionalen Kontextdaten ist in der Soziologie, wo soziale und nicht geografische Dimensionen überwiegen, ungewohnt. Die Volkszählung erlaubt jedoch grundsätzlich die Einbindung aller Klassifikationen – seit einiger Zeit auch geo-referenzierter Indikatoren – und ist hier besonders nützlich. Durch die Anonymisierung kann dieses Ziel nur teilweise erreicht werden, während die gesamte Datensammlung der Volkszählung auch eine solche Anwendung ermöglicht.

zwischen den Daten unmöglich ist, sind die Informationen zu den Kollektivhaushalten durch einen besonderen Code verschlüsselt. Analog sind nur Informationen über den Partner der Referenzperson zu finden, wenn eine solche Person wirklich vorhanden ist. Auch dieser Vorbehalt wird durch den Code signalisiert. Allgemein bedeutet dies, dass bei einer Analyse den Voraussetzungen, unter denen die Angaben gemacht wurden, sowie fehlenden Daten unbedingt Rechnung zu tragen ist.

2.3.3 Unmögliche und mögliche Analysen

Wie soeben gezeigt, sind gewisse Feinanalysen auf der Grundlage von Stichproben nicht möglich. Dies gilt zum Beispiel für die Haushaltsstruktur, da nicht mehr die gesamten Informationen über alle im Haushalt lebenden Personen enthalten sind. Für solche Analysen muss der gesamte Datensatz der Volkszählung verwendet werden.

Dies gilt auch für verschiedene weitere Analysen: So wurde zum Beispiel der Zusammenhang zwischen der Verstärkung und der Entwicklung der Pendlerbewegungen in Agglomerationen für höhere Kaderkräfte zwischen 1970 und 1990 untersucht (P.- A. Rumley et al., 2000). Diese Analyse bedingte nicht nur detaillierte Daten zum Beruf, sondern auch zum Arbeits- und zum Wohnort. Zur Sicherstellung der Anonymität mussten bei der Genauigkeit der Kategorien und bei den geografischen Angaben Abstriche gemacht und Daten entfernt werden.²⁵ Die beibehaltenen Daten schliessen jedoch eine Analyse über die Entwicklung des Pendlerverkehrs – wenn auch auf einer weniger detaillierten Ebene – nicht aus.

Es gibt noch andere Beispiele für Forschungsfragen, die sich mit den Daten der PUS beantworten lassen. Die Publikationsreihe des BFS vermittelt ein Bild davon. Ohne Anspruch auf Vollständigkeit seien hier einige Beispiele aufgezählt, über die die wissenschaftliche und politische Diskussion bei weitem noch nicht abgeschlossen ist:

1. Kantonsspezifische Analyse der Ausbildungen – insbesondere an Universitäten – und der entsprechenden Entwicklung. Es ist bekannt, dass in der Schweiz im internationalen Vergleich nur ein sehr geringer Prozentsatz der Bevölkerung über einen Mittelschulabschluss verfügt, zudem besteht dabei ein starkes Ost-West-Gefälle. Diese Unterschiede spielen in einer mobilen Welt sicher eine grosse Rolle.
2. Analyse der Migrationsbewegungen und der sozialen Stellung der Migranten. Die Migration und die damit verbundenen Folgen gehören seit Jahren zu den wichtigsten politischen Themen. Deshalb ist es sehr wichtig, diesbezügliche Daten in Erfahrung zu bringen. In diese Richtung gingen vier Erhebungen, in denen Daten zu Herkunft, Wohnort vor 5 Jahren, Ausbildung und aktueller beruflicher Stellung erhoben wurden. Damit lassen sich zumindest teilweise Geschichte und Auswirkungen dieser Wanderungen in Abhängigkeit des unterschiedlichen sozialen Umfelds nachvollziehen.
3. Analyse der sprachlichen Verhältnisse in der Schweiz. Ausschlaggebend für die sprachliche Verteilung sind nicht nur die Wanderungen und die breite Palette verwendeter Sprachen, sondern auch die Arbeitsgewohnheiten, im Zuge derer Englisch immer mehr als Alternative auftaucht, möglicherweise generationenspezifisch unterschiedlich ausgeprägt. Auch hier ist es wichtig, die Entwicklung in Bezug auf das Umfeld zu analysieren. Deshalb wurde diesem Punkt auch in der Broschüre zur Volkszählung 2000 grosse Aufmerksamkeit geschenkt.
4. Analyse der Erwerbstätigkeit von Frauen mit Kindern. Es ist bekannt, dass sich nicht nur die Familienstrukturen wandeln, sondern dass immer mehr Frauen mit Kindern berufstätig sind, was unter anderem die Angaben über alternative Berufsmodelle mit kürzeren Arbeitswegen, Teilzeitpensen usw. zeigen.

²⁵ Eine Analyse auf der Grundlage der Gemeinden ist allerdings angesichts der äusserst geringen Grösse gewisser Gemeinden mit zum Teil nur 20 Einwohnern ohnehin schwierig.

5. Analyse der Entwicklung des Pendlerverkehrs in Bezug auf die Verstädterung. Zahlreiche Fachpersonen sprechen davon, dass die Globalisierung die Entstehung von Ballungsgebieten fördert. Diese Tendenzen dürften das Verhältnis zwischen Zentren und Peripherien verändern. Mit Daten zur Entwicklung der regionalen Beschäftigungssituation und zu Veränderungen im Pendlerverkehr lassen sich in diesem Zusammenhang verschiedene Hypothesen überprüfen.

Dies sind nur einige Beispiele aus so unterschiedlichen Bereichen wie Soziologie, Bildung, Familienstrukturen, Lebensgewohnheiten oder Arbeitsmarkt. Anhand der Daten aus vier aufeinander folgenden Volkszählungen lässt sich die Entwicklung eines beobachteten Phänomens über einen längeren Zeitraum verfolgen. Zudem spricht nichts dagegen, die Informationen aus den Volkszählungen mit anderen Datensammlungen zu verknüpfen, zum Beispiel mit Abstimmungsergebnissen. Die Daten sind zudem detailliert genug, dass damit Informationen aus anderen Umfragen validiert und Stichproben überprüft werden können, was ebenfalls für eine Freigabe des PUS spricht.

Schliesslich eignen sich die Grundvariablen auch für internationale Vergleiche, die in der Wissenschaft immer wichtiger werden. Aus diesem Grund wurde zur Einteilung der Berufe das ISCO-System²⁶ herangezogen.

Selbstverständlich stellen die genannten Beispiele nur eine Auswahl möglicher Analysen dar. Wir vertrauen darauf, dass die Benutzer das Potenzial dieser Daten bestens ausschöpfen werden.

2.4 BEISPIELE AUS DEM AUSLAND

In zahlreichen Ländern stellen die nationalen statistischen Ämter Stichproben ihrer Volkszählungen zu Forschungszwecken zur Verfügung. Der Zugang zu den Daten ist je nach Land, Stichprobengrösse und Detailinformationen unterschiedlich geregelt. Zu erwähnen sind die «5%-Stichprobe» des statistischen Amtes Frankreichs INSEE oder auch die Stichproben des U.S. Census Bureau²⁷ und der statistischen Ämter der Niederlande oder Kanadas. In den Niederlanden wurden Fragen zu den PUS und zum Datenschutz in jüngerer Zeit in Publikationen des nationalen statistischen Amtes diskutiert.²⁸

Wir haben uns im vorliegenden Fall wesentlich von der Regelung in Kanada inspirieren lassen. Dort werden regelmässig Stichproben der Volkszählungsdaten veröffentlicht, und zwar sowohl über Einzelpersonen als auch über Haushalte. Die Datensammlung zur kanadischen Volkszählung und die zur weiten Verbreitung bestimmten Datensätze mit Mikrodaten (Statistique Canada, 1998) umfassen sogar das Einkommen (diese Angabe wird in der Schweizer Volkszählung nicht erfasst).

²⁶ «International Standard Classification of Occupations» ausgearbeitet durch das Internationale Arbeitsamt (Internationale Arbeitsorganisation: <http://www.ilo.org>). Für eine detaillierte Beschreibung siehe Internationales Arbeitsamt, 1991.

²⁷ Diese Stichproben werden durch eine Universität zur Verfügung gestellt (siehe: <http://www.ipums.org>).

²⁸ Siehe Netherlands Official Statistics, 1999, Band 14, zum Thema «Statistical Disclosure Control».

2.5 LITERATUR

Ambühl M. (Consult AG), Baarli B. (BFS), Breitenstein Ch. (SIDOS), Eichhorn W. (BFS), Graf M. (BFS), Joye D. (SIDOS), Kaufmann R. (Consult AG), Spahn D. (BFS) (2001) *Public Use Samples (PUS), Public-Use-Stichproben der schweizerischen Volkszählungen 1970, 1980 und 1990*. Neuchâtel: Bundesamt für Statistik.

Internationales Arbeitsamt (1991) *Classification internationale type des professions, CITP-88*. Genf: Internationales Arbeitsamt.

Busset T. (1993) *Eidgenössische Volkszählung 1990. Zur Geschichte der eidgenössischen Volkszählung*. Bern: Bundesamt für Statistik.

Feusi Widmer R. 2004 *Die Schweizerische Arbeitskräfteerhebung (SAKE). Konzepte, methodische Grundlagen, praktische Ausführung*. Neuchâtel: Bundesamt für Statistik.

INSEE (1987a) *Pour une histoire de la statistique, Tome 1: Contributions*. Paris: Economica.

INSEE (1987b) *Pour une histoire de la statistique, Tome 2: Matériaux*. Paris: Economica.

Jouve B. und Lefèvre C. (Hg.) (1998) *Villes, métropoles. Les nouveaux territoires du politique*. Paris: Economica.

Joye D. und Schuler M. (1995) *Sozialstruktur der Schweiz. Sozio-professionelle Kategorien*. Bern: Bundesamt für Statistik.

Bundesamt für Statistik (2004) *Schweizerische Arbeitskräfteerhebung (SAKE) 2004*. Neuchâtel: Bundesamt für Statistik.

Racine J.-B. und Joye D. (1988) *Le système urbain helvétique à l'aune européenne*. S. 259-281 in *Les enjeux de l'urbanisation. Agglomerationsprobleme in der Schweiz*. Hg. M. Bassand, D. Joye und M. Schuler. Bern: P. Lang.

Rumley P.-A., Burkhalter G., Jémelin C., Joye D. und Schuler M. (2000) *Réseau des villes suisses*. Bern: Bundesamt für Raumentwicklung.

Schuler M. und Jemelin C. (1996) *Régions urbaines et agglomérations: multiplicité et diversité des définitions statistiques en Europe*. In: *Raisons et déraisons de la ville*, Hg. C. Jaccoud, M. Schuler und M. Bassand. Lausanne: Presses polytechniques et universitaires romandes.

Schuler M., Dessemontet P. und Joye D. (2005) *Raumgliederungen der Schweiz*. Neuchâtel: Bundesamt für Statistik.

Statistique Canada (1998) *FMGD de 1996 sur les particuliers, documentation de l'utilisateur*. Ottawa: Statistique Canada.

Siehe Netherlands Official Statistics, 1999, Band 14, zum Thema « Statistical Disclosure Control ».

Hefte zu den Volkszählungen 1970, 1980 und 1990 (Definitionen usw.).

3. STICHPROBEN UND DATENSCHUTZ

3.1. STICHPROBE

3.1.1. Grundgesamtheit

Die Stichproben wurden unabhängig voneinander aus den Volkszählungen der Jahre 70, 80, 90 und 2000 gezogen. Es handelt sich dabei je um die harmonisierten Personen- und Haushaltsdateien (inklusive Gebäude- und Wohnungsdaten). Technisch gesehen kamen die Grundgesamtheiten dadurch zustande, dass an die Personenrecords die entsprechenden Merkmale aus den Haushaltsdateien angehängt wurden (über die Identifizierungen Zählgemeinde, Zählkreis, Haushaltsnummer). Es ergeben sich folgende Umfänge der Grundgesamtheiten:

Jahr	Anzahl Personen
1970	6'269'783
1980	6'365'960
1990	6'873'687
2000	7'288'010

3.1.2. Stichprobentechnik

Aus den oben beschriebenen Grundgesamtheiten wurde unabhängig voneinander je eine Stichprobe gezogen. Die Stichprobeneinheit bildete dabei die Personen-Identifizierung, jede Person besass die gleiche Ziehungswahrscheinlichkeit und wurde nur einmal in die Stichprobe aufgenommen. Der Umfang der Stichprobe betrug je 5% der Grundgesamtheit. Im statistischen Jargon spricht man dabei von einer einfachen Zufallsstichprobe ohne Zurücklegen. Aus Gründen des Datenschutzes wurden in einem ersten Schritt gezogene Personen aus den beiden Gemeinden Ederswiler und Bosco/Gurin aus der Stichprobe gelöscht (Identifikationsrisiko aufgrund des Merkmals "Sprache"). Es ergaben sich so folgende Stichprobenumfänge:

Jahr	Stichprobenumfang (5%)	Anzahl gelöscht	Resultierender Stichprobenumfang	Resultierender Stichprobenanteil
1970	313'489	16	313'473	4.9997%
1980	318'298	8	318'290	4.9999%
1990	343'684	12	343'672	4.9998%
2000	364'401	0	364'401	5.0000%

3.1.3. Geographische Gebiete

Die Wahl der kleinsten geographischen Gebiete, welche man in einer öffentlichen Stichprobe ausweisen soll, ist aus Sicht des Datenschutzes eines der schwierigsten Probleme. Es ist natürlich klar, dass der Anwender der Stichproben möglichst viel der geographischen Information haben möchte, auf der anderen Seite erleichtert deren Vorliegen – zum Beispiel die Kenntnis der Wohngemeinde – das Identifizieren einer Person enorm. Im Bereich der "Disclosure Risk Analysis" (Lehre des Schutzes gegen Re-Identifikation von öffentlich zugänglich gemachten Stichproben aus Volkszählungen) gibt es dazu einige allgemein anerkannte und praktisch überall angewandte

Faustregeln für die Bildung dieser kleinsten auszuweisenden geographischen Einheiten. Sie sollen so gewählt sein, dass

a) sie mindestens 100'000 Einwohner enthalten

und

b) mindestens zwei Städte (wenn solche vorhanden) nicht mehr als 50% der Einwohnereiner ausgewiesenen Region enthalten.

Im Fall der Schweiz ist natürlich die erste Regel besonders problematisch. Wenn sie eingehalten werden soll, fällt nicht nur das Ausweisen von Gemeinde und Bezirk weg, auch das Merkmal Kanton und sogar die Grossregion erreichen univariat die gewünschte Häufigkeit von 100'000 nicht überall. Im Sinne eines Kompromisses – und weil das Alter der veröffentlichten Daten ja beträchtlich ist – wurde der Kanton noch in der Stichprobe behalten (mit der Einschränkung, dass die beiden Appenzell zusammengelegt werden).

3.1.4. Zielgrössen Haushalt, Gebäude und Wohnung – Grenzen der Auswertbarkeit

Wie oben ausgeführt, bildet die Person die Stichprobeneinheit der vorliegenden Stichproben, es ist unter den Merkmalen aber auch eine Haushaltsidentifikation vorhanden. Zu Auswertungen, welche die Zielgrösse Haushalt besitzen, ist folgendes anzumerken:

- Die Personen wurden *zufällig* aus der Grundgesamtheit gezogen, für die Haushalte ist dies *nicht* der Fall.
- Die Ziehungswahrscheinlichkeiten der Haushalte sind also verschieden.
- Es ist nicht trivial, die resultierende Ziehungswahrscheinlichkeit für einen Haushalt anzugeben, es ist lediglich klar, dass grosse Haushalte gegenüber kleinen in der Stichprobe übervertreten sind. Die exakte Wahrscheinlichkeit hängt von der Chronologie der Ziehung ab und müsste nach jedem Schritt neu berechnet werden. Es wurde deshalb auf eine Angabe solcher Gewichte verzichtet.

Als Fazit dürfte die Zusammensetzung der Haushalte, wie sie in der Stichprobe zu finden ist, das Bild aus der Grundgesamtheit *nicht korrekt* wiedergeben. Aus diesem Grunde sind Auswertungen, welche die Zielgrösse Haushalt besitzen (gilt auch für die Zielgrössen Wohnung und Gebäude), mit allergrösster Vorsicht zu interpretieren.

3.2. METHODEN DES DATENSCHUTZES

3.2.1. Angewandte Verfahren

Bei einer Veröffentlichung einer Stichprobe aus einer Volkszählung müssen die gesetzlichen Vorschriften bezüglich des Datenschutzes beachtet werden. Es ist insbesondere möglichst zu vermeiden, dass eine in der Stichprobe enthaltene Person aufgrund einer Teilmenge von Merkmalen (zum Beispiel Alter, Beruf, Wohnort) identifiziert werden kann und ihr so die Werte der restlichen Merkmale zugeordnet werden können. Auch wenn dieses Risiko nie ganz ausgeschlossen werden kann, so ist es doch das Ziel, es möglichst zu minimieren. Dazu sind in der Lehre und Praxis verschiedene Verfahren bekannt, deren wichtigste wir kurz ansprechen wollen:

- *“global recoding”* Verschiedene Klassen eines Merkmals, welche in der Regel kritisch tiefe Häufigkeiten aufweisen, werden zusammengelegt. *“Global”* wird

- “local suppression” das Verfahren genannt, weil es auf alle Individuen in der Stichprobe angewandt wird. Die Ausprägung einer Person (deshalb “local”) in einem Merkmal (zum Beispiel ein seltener Beruf) wird unterdrückt, das heisst, auf “nicht bekannt” (missing) gesetzt.
- “swapping” Die Werte zweier Personen werden für ein bestimmtes Merkmal vertauscht. Dieses Verfahren wird auf “möglichst ähnliche” Personen angewandt, die sich aber in einem Merkmal (zum Beispiel einem geographischen) unterscheiden.
- “adding noise” Einem stetigen Merkmal wird ein zufälliger Fehler beigefügt.

Diese Verfahren weisen bezüglich Verfälschung und Informationsverlust in der Stichprobe unterschiedliche Charakteristiken auf, welche aber hier nicht weiter besprochen werden (siehe dazu das Literaturverzeichnis am Ende dieses Kapitels, *Willenborg, De Waal, 1996*).

3.2.2. Schutz der vorliegenden Stichproben

Das Wesen der Stichprobe selbst, bietet den grundlegenden Schutz der veröffentlichten Daten: Bei einer behaupteten Re-Identifikation müsste natürlich nachgewiesen werden, dass in der Grundgesamtheit aller Personen (die nicht öffentlich zugänglich ist), keine Person existiert, welche bezüglich der Werte der betrachteten Merkmale mit der vermeintlich Identifizierten übereinstimmt. Nichtsdestotrotz haben wir aber die vier Stichproben mittels oben beschriebener Verfahren zusätzlich gegen Re-Identifikation geschützt. Nebst der augenfälligen globalen Recodierung (offensichtlich im Vergleich mit vom Bundesamt veröffentlichten Auswertungen), kamen dabei weitere der Verfahren zur Anwendung. Welche davon in welchem Ausmass, stellen wir hier – wieder aus Datenschutzgründen – natürlich nicht dar, ihr Einsatz ist aber aus leichten Abweichungen, welche sich wieder zu Auswertungen des Bundesamtes ergeben können, ersichtlich. Das Ausmass dieser (leichten) Stichprobenverfälschung wird unten im Abschnitt “Streuungen” beschrieben.

3.3. STREUUNGEN

3.3.1. Schätzen von Anteilen und Streuungen

Da es sich bei sämtlichen Variablen in den Stichproben um kategorielle Grössen handelt, beschränkt sich die folgende Diskussion auf die Präzision von Anteilen.

3.3.2. Standardabweichung

Betrachten wir eine Population von N Individuen, aus der eine Zufallsstichprobe vom Umfang $n < N$ vorliegt. Der Anteil der Individuen in der Population, die ein gewisses Merkmal aufweisen, laute π ($0 < \pi < 1$). Verfügt man nur über eine Stichprobe aus dieser Population, so ist es naheliegend, den unbekanntem Populationsanteil π mit dem entsprechenden Stichprobenanteil p zu schätzen. Der Erwartungswert von p ist dann π (d.h. p schätzt π *erwartungstreu*), und die Standardabweichung von p ist gegeben durch

$$S.A.(p) = \sqrt{\frac{N-n}{N} \cdot \frac{p(1-p)}{n}}$$

Ist n/N , der Anteil der Stichprobe an der Population, klein, so liegt der Faktor $\sqrt{(N-n)/N}$ (die sogenannte *Endlichkeitskorrektur*) nahe bei 1 und wird damit vernachlässigbar.

In diesem Fall gilt also:

$$S.A.(p) = \sqrt{\frac{p(1-p)}{n}}$$

Im vorliegenden Fall einer 5%-Stichprobe beträgt der Korrekturfaktor $\sqrt{(N-n)/N} = \sqrt{0.95} \approx 0.975$, und kann weggelassen werden.

3.3.3. Vertrauensintervalle

Der Stichprobenanteil p allein enthält keine Angaben über die Präzision dieser Schätzung, d.h. er sagt nichts darüber aus, in welcher Grössenordnung die Abweichung zwischen p und π liegen, mit denen gerechnet werden muss. Aus diesem Grund wird die Punktschätzung häufig ergänzt durch eine sogenannte *Intervallschätzung (Vertrauensintervall)*, d.h. es wird ein Intervall angegeben, welches den Populationsanteil mit grosser Wahrscheinlichkeit (z.B. 95%) enthält. Diese Wahrscheinlichkeit nennt man auch *Vertrauenswahrscheinlichkeit*.

Ist die Voraussetzung

$$np(1-p) > 10$$

erfüllt, so kann ein Vertrauensintervall mit Vertrauenswahrscheinlichkeit $1-\alpha$ für den Populationsanteil π bestimmt werden mit der Formel

$$p \pm z_{1-\alpha/2} \cdot S.A.(p)$$

wo $z_{1-\alpha/2}$ das $(1-\alpha/2)$ -Quantil der Standardnormalverteilung bezeichnet. Im Fall $\alpha=5\%$, der am häufigsten verwendet wird, lautet dieser Wert $z_{0.975} = 1.96$, und das entsprechende Vertrauensintervall berechnet man als $p \pm 1.96 \cdot S.A.(p)$.

3.3.4. Anteile an der gesamten Population

Ist man an einem Anteil an der gesamten Population interessiert, so ist $n = 343'000$ und $N = 20 \times n = 6'860'000$. In diesem Fall ist die Bedingung $np(1-p) > 10$ ausser bei Stichprobenanteilen sehr nahe bei 0 oder 1 (etwa p unter 0.01% und über 99.99%) erfüllt. Punktschätzungen für Häufigkeiten in der Population erhält man aus den entsprechenden Häufigkeiten in der Stichprobe durch Multiplikation mit 20, die zugehörigen Vertrauensintervalle aus denjenigen für den entsprechenden Anteil durch Multiplikation der Intervallsgrenzen mit $N = 20 \times n = 6'860'000$.

3.3.5. Anteile in Subpopulation

Interessiert man sich für einen Anteil in einer Subpopulation (z.B. Anteil der berufstätigen Frauen oder Anteil der unter 20-jährigen in einem bestimmten Kanton), so erfolgen Punkt- und Intervallschätzung völlig analog zum Vorgehen in der gesamten Population. Am Faktor der Endlichkeitskorrektur ändert sich nichts, da die Stichprobe durch einfache Ziehung ermittelt wurde. Diese kann also auch hier vernachlässigt werden. Punktschätzungen für absolute Populations-Häufigkeiten findet man weiterhin aus den entsprechenden Häufigkeiten in der Stichprobe durch Multiplikation mit 20. Vertrauensintervalle für Populations-Häufigkeiten erhält man aus den Intervallsgrenzen für den entsprechenden Anteil durch Multiplikation mit $N = 20 \times n$.

3.4. STREUUNGSKOMPONENTEN

Die Streuung in den Stichproben hat zwei Komponenten: Die erste besteht aus der vorne beschriebenen Stichprobenvariabilität. Sie ist für Anteile absolut in der Mitte (bei 50%) am grössten, an den Rändern (bei beispielsweise 1% oder 99%) am kleinsten. *Relativ* gesehen allerdings ergeben sich hohe Streuungen für kleine Anteile, dies stellt auch in den vorliegenden Stichproben die bedeutende Problematik dar. Über alle Merkmale gesehen ergeben sich zwischen den Grundgesamtheiten und den entsprechenden Stichproben absolute Abweichungen in der Grössenordnung von +/- 1 Promille (0.1%). Bei sehr kleinen Anteilen (unter einem Prozent

– solche werden aber in den Stichproben fast nicht erreicht), kann aber eine kleine absolute Abweichung relativ betrachtet ziemlich viel ausmachen und auch schon mal gegen die 10% erreichen. Die zweite Komponente ist diejenige, welche aus der Stichprobenverfälschung durch das Anwenden von Datenschutz-Verfahren (siehe Abschnitt 3.2.) resultiert. Über alle Merkmale gesehen liegen auch hier die Abweichungen (zusätzlich) bei plus minus einem Promille oder darunter, sind also absolut gesehen sehr klein, univariat vernachlässigbar. Aber auch hier können – vor allem wieder bei sehr kleinen Anteilen – *relative* Abweichungen zwischen Stichprobe und Grundgesamtheit von 10% oder darüber erreicht werden.

Zusammenfassend kann gesagt werden, dass die Streuungskomponenten univariat gesehen sehr klein sind, sich insgesamt im 2 - Promille-Bereich bewegen (jetzt mal als absolute Differenzen und nicht als Standardabweichung angesehen) und praktisch vernachlässigbar sind. Punktuell kann es aber grössere relative Abweichungen bei kleinen Anteilen geben und hier ist vor allem dann Vorsicht geboten, wenn solche Variablen Teil einer Mehrfeldertafel sind. Hier sollten die Resultate nur sehr vorsichtig interpretiert werden.

3.5. LITERATUR

- Bethlehem J. A., Keller W. J. und Pannekoek J. (1990), *Disclosure Control of Microdata*. JASA 85, 38-45.
- De Waal A. G. und Willenborg L. C. R. J. (1999). *Information loss through global recoding and local suppression*. Netherlands Official Statistics 14, 17-20.
- Elliot M. J., Dale A. und Skinner C. J. (1998). *Special uniques, random uniques and sticky populations: some counterintuitive effects of geographical detail on disclosure risk*. Statistical data protection: Proceedings of the conference of Lisbon, March 1998.
- Kim J. J. und Winkler E. (1995). *Masking Microdata files*. American Statistical Association, Proceedings of the Section on Survey Research Methods, 114-119.
- Moore R. A. Jr. (1996). *Preliminary Recommendations for Disclosure Limitation for the 2000 Census; Improving the 1990 Confidentiality Edit Procedure*. US Bureau of the Census, Statistical Research Division.
- Takemura A. (1998). *Some superpopulation models for estimating the number of population uniques*. Statistical data protection: Proceedings of the conference of Lisbon, March 1998.
- Willenborg L. C. R. J. und De Waal A. G. (1996). *Statistical Disclosure control in Practice*. New York: Springer.

4. BESCHREIBUNG DER DATEN

4.1. VORWORT

Dieses Kapitel beschreibt die verfügbaren Variablen und Dokumentationen sowie die Art, wie diese Informationen innerhalb der Datensammlung organisiert sind. Solche Informationen sind für die korrekte Benützung der Daten unerlässlich; dies gilt insbesondere für so umfangreiche Datensätze wie jene im Zusammenhang mit der Personen-, Haushalts- sowie Gebäude- und Wohnungserhebung im Rahmen der Volkszählung, für die komplexe Indikatoren gebildet wurden. Hinzu kommt, dass zwischen 1970 und 1990 einige Fragen abgeändert wurden, nicht zuletzt um dem sozialen Wandel Rechnung zu tragen (Arbeit und insbesondere Definition des Haushaltsvorstands). Vor diesem Hintergrund spielt die Dokumentation eine zentrale Rolle für die notwendige Anpassung der Messinstrumente.

Beginnen wir jedoch damit, wie die Variablen aufzufinden sind.

4.2. DIE VARIABLENLISTEN

Wie bereits erwähnt, haben wir ein Maximum an Variablen der Personen-, Haushalts-, Wohnungs- und Gebäudeerhebung beibehalten und können so rund hundert Variablen zur Verfügung stellen. Aus Übersichtsründen haben wir drei Listen erstellt, die verschiedene Herangehensweisen an die Daten erlauben:

- eine Übersichtstabelle, die der Logik der Volkszählungsfragebogen und auch der Reihenfolge der Variablen in der Datensammlung folgt;
- eine thematisch organisierte Variablenliste; dies ist wahrscheinlich die einfachste Zugangsmöglichkeit bei einem ersten Kontakt mit den Daten;
- eine alphabetisch sortierte Liste für die Benutzer, die sich bereits mit den Variablenbezeichnungen des Amtes auskennen.

4.2.1. Übersichtstabelle

Nr. der Variable	Variable	Bezeichnung	Anzahl Klassen[1]	ASCII Position
1	ZJHR	Zähljahr	4	1-5
Einzelperson				
Demographie / Religion				
2	ALTJ	Alter - Kategorien	18	6-10
3	GESL	Geschlecht	3	11-15
4	NATI	Nationalität	4	16-20
5	ZIVL	Zivilstand	5	21-25
6	KONF	Konfession	5	26-30
7	SPRA1	Hauptsprache	5	31-35
8	GORTK	Geburtsort - Kategorien	5	36-40
9	WO5K	Wohnort vor 5 Jahren	7	41-45
Ausbildung				
10	HABGH	Höchste abgeschlossene Ausbildung	9	46-50
11	ERLB1	Erlerner Beruf	6	51-55
Beruf / Soziale Stellung				
12	EZUX	Erwerbsschlüssel	9	56-60
13	PENS	Arbeitspensum	5	61-65
14	PBER1	Berufliche Tätigkeit	6	66-70
15	ISCO1	Berufliche Tätigkeit nach ISCO	11	71-75
16	STHB	Stellung im Beruf	7	76-80
17	WART1	Wirtschaftsart	5	81-85
18	REFO	Rechtsform der Unternehmens	6	86-90
19	AORT	Arbeits-, Schulort - Kategorien	5	91-95
20	WEGZ	Zeitbedarf für den Arbeits-/Schulweg	7	96-100
21	WEGH	Häufigkeit Arbeits-/Schulweg	4	101-105
22	VEMI	Hauptverkehrsmittel	8	106-110
23	SOPK	Sozio-professionelle Kategorie	11	111-115
24	TREI2	Prestige nach der Skala von Treiman	8	116-120
Referenzperson im Privathaushalt (PHH)				
Demografie / Religion				
25	RALTJ	Bevölkerung nach dem Alter der Referenzperson im PHH	5	121-125
26	RGESL	Bevölkerung nach dem Geschlecht der Referenzperson im PHH	3	126-130
27	RNATI	Bevölkerung nach der Nationalität der Referenzperson im PHH	3	131-135
28	RZIVL	Bevölkerung nach dem Zivilstand der Referenzperson im PHH	4	136-140
29	RKONF	Bevölkerung nach der Religion der Referenzperson im PHH	5	141-145
30	RSPRA1	Bevölkerung nach der Hauptsprache der Referenzperson im PHH	5	146-150
Ausbildung				
31	RHABGH	Bevölkerung nach der höchste abgeschlossene Ausbildung der Referenzperson im PHH	9	151-155

Referenzperson im Privathaushalt (PHH)				
Beruf / Soziale Stellung				
32	RESCH	Bevölkerung nach dem Erwerbsschlüssel der Referenzperson im PHH	8	156-160
33	RPENS	Bevölkerung nach dem Arbeitspensum der Referenzperson im PHH	4	161-165
34	RPBER1	Bevölkerung nach der berufliche Tätigkeit der Referenzperson	6	166-170
35	RSTHB	Bevölkerung nach der Stellung im Beruf der Referenzperson	4	171-175
36	RWART1	Bevölkerung nach der Wirtschaftsart der Referenzperson	5	176-180
37	RREFO	Bevölkerung nach der Rechtsform des Unternehmens in dem die Referenzperson des Privathaushaltes arbeitet	6	181-185
38	RAGDE	Bevölkerung nach dem Arbeits-, Schulort der Referenzperson im PHH	5	186-190
39	RWEGZ	Bevölkerung nach dem Zeitbedarf für den Arbeits-/Schulweg der Referenzperson im PHH	7	191-195
40	RWEGH	Bevölkerung nach der Häufigkeit Arbeits-, Schulweg der Referenzperson im PHH	4	196-200
41	RVEMI	Bevölkerung nach dem Hauptverkehrsmittel der Referenzperson im PHH	8	201-205
42	RSOPK	Bevölkerung nach der sozio-professionelle Kategorie der Referenzperson im PHH	11	206-210
(Ehe-)Partner/in der Referenzperson im Privathaushalt (PHH)				
Demografie / Religion				
43	ZALTJ	Bevölkerung nach dem Alter der Partnerin/des Partners der Referenzperson im PHH – Kategorie	5	211-215
44	ZGESL	Bevölkerung nach dem Geschlecht der Partnerin/des Partners der Referenzperson im PHH	3	216-220
45	ZNATI	Bevölkerung nach der Nationalität der Partnerin/des Partners der Referenzperson im PHH	3	221-225
46	ZZIVL	Bevölkerung nach dem Zivilstand der Partnerin/des Partners der Referenzperson im PHH	4	226-230
47	ZKONF	Bevölkerung nach der Religion der Partnerin/des Partners der Referenzperson im PHH	5	231-235
48	ZSPRA1	Bevölkerung nach der Hauptsprache der Partnerin/des Partners der Referenzperson im PHH	5	236-240
Ausbildung				
49	ZHABGH	Bevölkerung nach der höchste abgeschlossene Ausbildung der Partnerin/des Partners der Referenzperson im PHH	9	241-245

(Ehe-)Partner/in der Referenzperson im Privathaushalt (PHH) (Fortsetzung)				
Beruf / Soziale Stellung				
50	ZESCH	Bevölkerung nach dem Erwerbsschlüssel der Partnerin/des Partners der Referenzperson im PHH	8	246-250
51	ZPENS	Bevölkerung nach dem Arbeitspensum der Partnerin/des Partners der Referenzperson im PHH	4	251-255
52	ZPBER1	Bevölkerung nach der berufliche Tätigkeit der Partnerin/des Partners der Referenzperson im PHH	6	256-260
53	ZSTHB	Bevölkerung nach der Stellung im Beruf der Partnerin/des Partners der Referenzperson im PHH	4	261-265
54	ZWART1	Bevölkerung nach der Wirtschaftsart der Partnerin/des Partners der Referenzperson im PHH	5	266-270
55	ZREFO	Bevölkerung nach der Rechtsform des Unternehmens in dem der/die Partner/in der Referenzperson des Privathaushaltes arbeitet	6	271-275
56	ZAGDE	Bevölkerung nach dem Arbeits-, Schulort der Partnerin/des Partners der Referenzperson im PHH	5	276-280
57	ZWEGZ	Bevölkerung nach dem Zeitbedarf für den Arbeits- oder Schulweg der Partnerin/des Partners der Referenzperson im PHH	7	281-285
58	ZWEGH	Bevölkerung nach der Häufigkeit Arbeits-/Schulweg der Partnerin/des Partners der Referenzperson im PHH, Anzahl Male	4	286-290
59	ZVEMI	Bevölkerung nach dem Hauptverkehrsmittel der Partnerin/des Partners der Referenzperson im PHH	8	291-295
60	ZSOPK	Bevölkerung nach der sozio-professionelle Kategorie der Partnerin/des Partners der Referenzperson im PHH	11	296-300
Haushalt				
61	HHTP	Haushaltstyp	4	301-305
62	RPHH	Referenzperson im Haushalt	5	306-310
63	STHH	Stellung im Haushalt	6	311-315
64	APER	Anzahl Personen im Haushalt	9	316-320
65	STH2	Anzahl (Schwieger-)Kinder im HH	6	321-325
66	STH3_4	Anwesenheit von (Schwieger-)Eltern und anderen Verwandten im PHH	3	326-330
67	STH5	Anwesenheit von Nichtverwandten im PHH	3	331-335
68	AL00_04	Anzahl 0-4-jähriger Kinder im PHH	4	336-340
69	AL05_09	Anzahl 5-9-jähriger Kinder im PHH	4	341-345
70	AL10_14	Anzahl 10-14-jähriger Kinder im PHH	4	346-350
71	AL15_19	Anzahl 15-19-jähriger Kinder im PHH	4	351-355
72	AL20	Anzahl lediger 20-24-jähriger Kinder im PHH	4	356-360
73	AL25	Anwesenheit von ledigen 25-29-jährigen Kindern im PHH	3	361-365

Haushalt (Fortsetzung)				
74	AL30	Anwesenheit von ledigen Kindern im Alter von 30 und mehr Jahren im PHH	3	366-370
75	AP65	Anzahl Personen im Haushalt im Alter von 65 und mehr Jahren	5	371-375
76	AP80	Anzahl Personen im Haushalt im Alter von 80 und mehr Jahren	4	376-380
77	IA14	Anzahl Schulpflichtiger im PHH	5	381-385
78	LEHR	Anzahl Lehrlinge im PHH	4	386-390
79	IAUS	Anzahl Personen in Ausbildung (auch wenn erwerbstätig) im PHH	6	391-395
80	ETTO	Anzahl Erwerbstätige im PHH	6	396-400
81	ELOS	Anwesenheit von Erwerbslosen im PHH	3	401-405
82	RENT	Anzahl (nicht erwerbstätige) Rentner im PHH	4	406-410
83	AUSL	Anzahl Ausländer im PHH	7	411-415
Wohnung				
84	WBTyp	Bewohnertypus	8	416-420
85	WMiET	Mietpreis pro Monat	28	421-425
86	WAPTo	Anzahl Personen im Wohnungsrecord	9	426-430
87	WAPRA	Anzahl Wohnräume pro Person im Wohnungsrecord	6	431-435
88	WAPFL	Fläche der Wohnung pro Person im Wohnungsrecord	7	436-440
Gebäude				
89	GGART	Gebäudeart	5	441-445
90	GBAUP	Bauperiode	8	446-450
91	GAZWT	Zahl der Wohneinheiten im Gebäude	10	451-455
92	GHEIZ	Heizungsart	6	456-460
Geografie				
93	KANT	Kanton	26	461-465
94	GROSSREG	Grossregion der Schweiz	7	466-470
95	SPR93	Sprachgebiet	3	471-475
96	AGGLO	Agglomeration nach Definition 2000	4	476-480
[1] Die Anzahl Klassen umfasst auch die Kategorie INAP				

4.2.2 Variablen nach Themen²⁹

Einzelperson			Referenzperson im Privathaushalt (PHH)			(Ehe-)Partner/in der Referenzperson im Privathaushalt (PHH)		
Demografie / Religion								
2	ALTJ	Alter - Kategorien	25	RALTJ	Bevölkerung nach dem Alter der Referenzperson im PHH	43	ZALTJ	Bevölkerung nach dem Alter der Partnerin/des Partners der Referenzperson im PHH – Kategorie
3	GESL	Geschlecht	26	RGESL	Bevölkerung nach dem Geschlecht der Referenzperson im PHH	44	ZGESL	Bevölkerung nach dem Geschlecht der Partnerin/des Partners der Referenzperson im PHH
4	NATI	Nationalität	27	RNATI	Bevölkerung nach der Nationalität der Referenzperson im PHH	45	ZNATI	Bevölkerung nach der Nationalität der Partnerin/des Partners der Referenzperson im PHH
5	ZIVL	Zivilstand	28	RZIVL	Bevölkerung nach dem Zivilstand der Referenzperson im PHH	46	ZZIVL	Bevölkerung nach dem Zivilstand der Partnerin/des Partners der Referenzperson im PHH
6	KONF	Konfession	29	RKONF	Bevölkerung nach der Religion der Referenzperson im PHH	47	ZKONF	Bevölkerung nach der Religion der Partnerin/des Partners der Referenzperson im PHH

²⁹ Die Zahl links des Variablennamens bezeichnet die Position der Variable in der Datendatei.

Einzelperson			Referenzperson im Privathaushalt (PHH)			(Ehe-)Partner/in der Referenzperson im Privathaushalt (PHH)		
Demografie / Religion (Fortsetzung)								
7	SPRA1	Hauptsprache	30	RSPRA1	Bevölkerung nach der Hauptsprache der Referenzperson im PHH	48	ZSPRA1	Bevölkerung nach der Hauptsprache der Partnerin/des Partners der Referenzperson im PHH
8	GORTK	Geburtsort - Kategorien						
9	WO5K	Wohnort vor 5 Jahren						
Ausbildung								
10	HABGH	Höchste abgeschlossene Ausbildung	31	RHABGH	Bevölkerung nach der höchste abgeschlossene Ausbildung der Referenzperson im PHH	49	ZHABGH	Bevölkerung nach der höchste abgeschlossene Ausbildung der Partnerin/des Partners der Referenzperson im PHH
11	ERLB1	Erlerner Beruf						
Beruf / Soziale Stellung								
12	EZUX	Erwerbsschlüssel	32	RESCH	Bevölkerung nach dem Erwerbsschlüssel der Referenzperson im PHH	50	ZESCH	Bevölkerung nach dem Erwerbsschlüssel der Partnerin/des Partners der Referenzperson im PHH
13	PENS	Arbeitspensum	33	RPENS	Bevölkerung nach dem Arbeitspensum der Referenzperson im PHH	51	ZPENS	Bevölkerung nach dem Arbeitspensum der Partnerin/des Partners der Referenzperson im PHH

Einzelperson			Referenzperson im Privathaushalt (PHH)			(Ehe-)Partner/in der Referenzperson im Privathaushalt (PHH)		
Beruf / Soziale Stellung (Fortsetzung)								
14	PBER1	Berufliche Tätigkeit	34	RPBER1	Bevölkerung nach der berufliche Tätigkeit der Referenzperson	52	ZPBER1	Bevölkerung nach der berufliche Tätigkeit der Partnerin/des Partners der Referenzperson im PHH
15	ISCO1	Berufliche Tätigkeit nach ISCO						
16	STHB	Stellung im Beruf	35	RSTHB	Bevölkerung nach der Stellung im Beruf der Referenzperson	53	ZSTHB	Bevölkerung nach der Stellung im Beruf der Partnerin/des Partners der Referenzperson im PHH
17	WART1	Wirtschaftsart	36	RWART1	Bevölkerung nach der Wirtschaftsart der Referenzperson	54	ZWART1	Bevölkerung nach der Wirtschaftsart der Partnerin/des Partners der Referenzperson im PHH
18	REFO	Rechtsform der Unternehmens	37	RREFO	Bevölkerung nach der Rechtsform des Unternehmens in dem die Referenzperson des Privathaushaltes arbeitet	55	ZREFO	Bevölkerung nach der Rechtsform des Unternehmens in dem der/die Partner/in der Referenzperson des Privathaushaltes arbeitet
19	AORT	Arbeits-, Schulort - Kategorien	38	RAGDE	Bevölkerung nach dem Arbeits-, Schulort der Referenzperson im PHH	56	ZAGDE	Bevölkerung nach dem Arbeits-, Schulort der Partnerin/des Partners der Referenzperson im PHH

Einzelperson			Referenzperson im Privathaushalt (PHH)			(Ehe-)Partner/in der Referenzperson im Privathaushalt (PHH)		
Beruf / Soziale Stellung (Fortsetzung)								
20	WEGZ	Zeitbedarf für den Arbeits-/Schulweg	39	RWEGZ	Bevölkerung nach dem Zeitbedarf für den Arbeits-/Schulweg der Referenzperson im PHH	57	ZWEGZ	Bevölkerung nach dem Zeitbedarf für den Arbeits- oder Schulweg der Partnerin/des Partners der Referenzperson im PHH
21	WEGH	Häufigkeit Arbeits-/Schulweg	40	RWEGH	Bevölkerung nach der Häufigkeit Arbeits-, Schulweg der Referenzperson im PHH	58	ZWEGH	Bevölkerung nach der Häufigkeit Arbeits-/Schulweg der Partnerin/des Partners der Referenzperson im PHH, Anzahl Male
22	VEMI	Hauptverkehrsmittel	41	RVEMI	Bevölkerung nach dem Hauptverkehrsmittel der Referenzperson im PHH	59	ZVEMI	Bevölkerung nach dem Hauptverkehrsmittel der Partnerin/des Partners der Referenzperson im PHH
23	SOPK	Sozio-professionelle Kategorie	42	RSOPK	Bevölkerung nach der Sozio-professionelle Kategorie der Referenzperson im PHH	60	ZSOPK	Bevölkerung nach der Sozio-professionelle Kategorie der Partnerin/des Partners der Referenzperson im PHH
24	TREI2	Prestige nach der Skala von Treiman						

Haushalt		
61	HHTP	Haushaltstyp
62	RPHH	Referenzperson im Haushalt
63	STHH	Stellung im Haushalt
64	APER	Anzahl Personen im Haushalt
65	STH2	Anzahl (Schwieger-)Kinder im HH
66	STH3_4	Anwesenheit von (Schwieger-)Eltern und anderen Verwandten im PHH
67	STH5	Anwesenheit von Nichtverwandten im PHH
68	AL00_04	Anzahl 0-4-jähriger Kinder im PHH
69	AL05_09	Anzahl 5-9-jähriger Kinder im PHH
70	AL10_14	Anzahl 10-14-jähriger Kinder im PHH
71	AL15_19	Anzahl 15-19-jähriger Kinder im PHH
72	AL20	Anzahl lediger 20-24-jähriger Kinder im PHH
73	AL25	Anwesenheit von ledigen 25-29-jährigen Kindern im PHH
74	AL30	Anwesenheit von ledigen Kindern im Alter von 30 und mehr Jahren im PHH
75	AP65	Anzahl Personen im Haushalt im Alter von 65 und mehr Jahren
76	AP80	Anzahl Personen im Haushalt im Alter von 80 und mehr Jahren
77	IA14	Anzahl Schulpflichtiger im PHH
78	LEHR	Anzahl Lehrlinge im PHH
79	IAUS	Anzahl Personen in Ausbildung (auch wenn erwerbstätig) im PHH
80	ETTO	Anzahl Erwerbstätige im PHH
81	ELOS	Anwesenheit von Erwerbslosen im PHH
82	RENT	Anzahl (nicht erwerbstätige) Rentner im PHH
83	AUSL	Anzahl Ausländer im PHH

Wohnung		
84	WBTyp	Bewohnertypus
85	WMIET	Mietpreis pro Monat
86	WAPTO	Anzahl Personen im Wohnungsrecord
87	WAPRA	Anzahl Wohnräume pro Person im Wohnungsrecord
88	WAPFL	Fläche der Wohnung pro Person im Wohnungsrecord

Gebäude		
89	GGART	Gebäudeart
90	GBAUP	Bauperiode
91	GAZWT	Zahl der Wohneinheiten im Gebäude
92	GHEIZ	Heizungsart

Geografie		
93	KANT	Kanton
94	GROSSREG	Grossregion der Schweiz
95	SPR93	Sprachgebiet
96	AGGLO	Agglomeration nach Definition 2000

Jahr		
1	ZJHR	Zähljahr

4.2.3. Variablen in alphabetischer Reihenfolge

Nr. der Variable	Variable	Bezeichnung	Variablentypen									
			Einzelperson			Haushalt			Wohnung	Gebäude	Geografie	
			Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung				
96	AGGLO	Agglomeration nach Definition 2000										X
68	AL00_04	Anzahl 0-4-jähriger Kinder im PHH				X						
69	AL05_09	Anzahl 5-9-jähriger Kinder im PHH				X						
70	AL10_14	Anzahl 10-14-jähriger Kinder im PHH				X						
71	AL15_19	Anzahl 15-19-jähriger Kinder im PHH				X						
72	AL20	Anzahl lediger 20-24-jähriger Kinder im PHH				X						
73	AL25	Anwesenheit von ledigen 25-29-jährigen Kindern im PHH				X						
74	AL30	Anwesenheit von ledigen Kindern im Alter von 30 und mehr Jahren im PHH				X						
2	ALTJ	Alter – Kategorien	X									
19	AORT	Arbeits-, Schulort - Kategorien			X							
75	AP65	Anzahl Personen im Haushalt im Alter von 65 und mehr Jahren				X						
76	AP80	Anzahl Personen im Haushalt im Alter von 80 und mehr Jahren				X						
64	APER	Anzahl Personen im Haushalt				X						
83	AUSL	Anzahl Ausländer im PHH				X						
81	ELOS	Anwesenheit von Erwerbslosen im PHH				X						
11	ERLB1	Erlerner Beruf		X								

Nr. der Variable	Variable	Bezeichnung	Variablentypen									
			Einzelperson			Haushalt						
			Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Wohnung	Gebäude	Geografie	
80	ETTO	Anzahl Erwerbstätige im PHH				X						
12	EZUX	Erwerbsschlüssel			X							
91	GAZWT	Zahl der Wohneinheiten im Gebäude								X		
90	GBAUP	Bauperiode								X		
3	GESL	Geschlecht	X									
89	GGART	Gebäudeart								X		
92	GHEIZ	Heizungsart								X		
8	GORTK	Geburtsort - Kategorien	X									
94	GROSSREG	Grossregion der Schweiz										X
10	HABGH	Höchste abgeschlossene Ausbildung		X								
61	HHTP	Haushaltstyp				X						
77	IA14	Anzahl Schulpflichtiger im PHH				X						
79	IAUS	Anzahl Personen in Ausbildung (auch wenn erwerbstätig) im PHH					X					
15	ISCO1	Berufliche Tätigkeit nach ISCO			X							
93	KANT	Kanton										X
6	KONF	Konfession	X									
78	LEHR	Anzahl Lehrlinge im PHH					X					
4	NATI	Nationalität	X									
14	PBER1	Berufliche Tätigkeit			X							
13	PENS	Arbeitspensum			X							
38	RAGDE	Arbeits-, Schulort der Referenzperson im PHH - Kategorien						X				

Nr. der Variable	Variable	Bezeichnung	Variablentypen										
			Einzelperson			Haushalt							
			Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Wohnung	Gebäude	Geografie		
25	RALTJ	Bevölkerung nach dem Alter der Referenzperson im PHH - Kategorien				X							
18	REFO	Rechtsform des Unternehmens			X								
82	RENT	Anzahl (nicht erwerbstätige) Rentner im PHH				X							
32	RESCH	Bevölkerung nach dem Erwerbsschlüssel der Referenzperson im PHH						X					
26	RGESL	Bevölkerung nach dem Geschlecht der Referenzperson im PHH				X							
31	RHABGH	Bevölkerung nach der höchste abgeschlossene Ausbildung der Referenzperson im PHH					X						
29	RKONF	Bevölkerung nach der Religion der Referenzperson im PHH				X							
27	RNATI	Bevölkerung nach der Nationalität der Referenzperson im PHH				X							
34	RPBER1	Bevölkerung nach der berufliche Tätigkeit der Referenzperson						X					
33	RPENS	Bevölkerung nach dem Arbeitspensum der Referenzperson im PHH						X					
62	RPHH	Referenzperson im Haushalt				X							
37	RREFO	Bevölkerung nach der Rechtsform des Unternehmens in dem die Referenzperson des Privathaushaltes arbeitet						X					

Nr. der Variable	Variable	Bezeichnung	Variablentypen								
			Einzelperson			Haushalt					
			Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Wohnung	Gebäude	Geografie
42	RSOPK	Bevölkerung nach der sozio-professionelle Kategorie der Referenzperson im PHH						X			
30	RSPRA1	Bevölkerung nach der Hauptsprache der Referenzperson im PHH				X					
35	RSTHB	Bevölkerung nach der Stellung im Beruf der Referenzperson						X			
41	RVEMI	Bevölkerung nach dem Hauptverkehrsmittel der Referenzperson im PHH						X			
36	RWART1	Bevölkerung nach der Wirtschaftsart der Referenzperson im PHH						X			
40	RWEGH	Bevölkerung nach der Häufigkeit Arbeits-, Schulweg der Referenzperson im PHH						X			
39	RWEGZ	Bevölkerung nach dem Zeitbedarf für den Arbeits-/Schulweg der Referenzperson im PHH						X			
28	RZIVL	Bevölkerung nach dem Zivilstand der Referenzperson im PHH				X					
23	SOPK	Sozio-professionelle Kategorie			X						
95	SPR93	Sprachgebiet									X
7	SPRA1	Hauptsprache	X								
65	STH2	Anzahl (Schwieger-) Kinder im PHH				X					
66	STH3_4	Anwesenheit von (Schwieger-)Eltern und anderen Verwandten im PHH				X					

			Variablentypen								
			Einzelperson			Haushalt					
Nr. der Variable	Variable	Bezeichnung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Wohnung	Gebäude	Geografie
67	STH5	Anwesenheit von Nichtverwandten im PHH				X					
16	STHB	Stellung im Beruf			X						
63	STHH	Stellung im Haushalt				X					
24	TREI2	Prestige nach der Skala von Treiman			X						
22	VEMI	Hauptverkehrsmittel			X						
88	WAPFL	Fläche der Wohnung pro Person im Wohnungsrecord							X		
87	WAPRA	Anzahl Wohnräume pro Person im Wohnungsrecord							X		
86	WAPTO	Anzahl Personen im Wohnungsrecord							X		
17	WART1	Wirtschaftsart			X						
84	WBTyp	Bewohnertypus							X		
21	WEGH	Häufigkeit Arbeits-/Schulweg			X						
20	WEGZ	Zeitbedarf für den Arbeits-/Schulweg			X						
85	WMIET	Mietpreis pro Monat							X		
9	WO5K	Wohnort vor 5 Jahren - Kategorien	X								
56	ZAGDE	Bevölkerung nach dem Arbeits-, Schulort der Partnerin/des Partners der Referenzperson im PHH						X			
43	ZALTJ	Bevölkerung nach dem Alter der Partnerin/des Partners der Referenzperson im PHH - Kategorie				X					
50	ZESCH	Bevölkerung nach dem Erwerbsschlüssel der Partnerin/des Partners der Referenzperson im PHH						X			

Nr. der Variable	Variable	Bezeichnung	Variablentypen									
			Einzelperson			Haushalt						
			Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Wohnung	Gebäude	Geografie	
44	ZGESL	Bevölkerung nach dem Geschlecht der Partnerin/des Partners der Referenzperson im PHH				X						
49	ZHABGH	Bevölkerung nach der höchste abgeschlossene Ausbildung der Partnerin/des Partners der Referenzperson im PHH					X					
5	ZIVL	Zivilstand	X									
1	ZJHR	Zähljahr										
47	ZKONF	Bevölkerung nach der Religion der Partnerin/des Partners der Referenzperson im PHH				X						
45	ZNATI	Bevölkerung nach der Nationalität der Partnerin/des Partners der Referenzperson im PHH				X						
52	ZPBER1	Bevölkerung nach der berufliche Tätigkeit der Partnerin/des Partners der Referenzperson im PHH						X				
51	ZPENS	Bevölkerung nach dem Arbeitspensum der Partnerin/des Partners der Referenzperson im PHH						X				
55	ZREFO	Bevölkerung nach der Rechtsform des Unternehmens in dem der/die Partner/in der Referenzperson des Privathaushaltes arbeitet						X				
60	ZSOPK	Bevölkerung nach der sozio-professionelle Kategorie der Partnerin/des Partners der Referenzperson im PHH						X				

Nr. der Variable	Variable	Bezeichnung	Variablentypen								
			Einzelperson			Haushalt					
			Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Demografie / Religion	Ausbildung	Beruf / Soziale Stellung	Wohnung	Gebäude	Geografie
48	ZSPRA1	Bevölkerung nach der Hauptsprache der Partnerin/des Partners der Referenzperson im PHH				X					
53	ZSTHB	Bevölkerung nach der Stellung im Beruf der Partnerin/des Partners der Referenzperson im PHH						X			
59	ZVEMI	Bevölkerung nach dem Hauptverkehrsmittel der Partnerin/des Partners der Referenzperson im PHH						X			
54	ZWART1	Bevölkerung nach der Wirtschaftsart der Partnerin/des Partners der Referenzperson im PHH						X			
58	ZWEGH	Bevölkerung nach der Häufigkeit Arbeits-/ Schulweg der Partnerin/des Partners der Referenzperson im PHH, Anzahl Male						X			
57	ZWEGZ	Bevölkerung nach dem Zeitbedarf für den Arbeits- oder Schulweg der Partnerin/des Partners der Referenzperson im PHH						X			
46	ZZIVL	Bevölkerung nach dem Zivilstand der Partnerin/des Partners der Referenzperson im PHH				X					

4.3. DOKUMENTATIONSMODELL UND BEISPIEL

Die Dokumentation jeder Variable basiert auf folgendem Standardmodell:

- Kürzel der Variable (fett)
- Bezeichnung der Variable (fett)
- kurze Beschreibung des Inhalts der Variable
- Hinweis "Referenzpopulation", der angibt, auf welche Grundgesamtheit sich die Variable bezieht
- Variablenresultate anlässlich der Volkszählungen von 1970, 1980 und 1990 (Anzahl Fälle und Prozentwerte)
- Abschnitt "Besonderes" mit Details über die Unterschiede oder Besonderheiten der einzelnen Volkszählungen
- Abschnitt "Hinweis" mit allgemeineren Informationen
- einfache grafische Darstellung für einen Vergleich zwischen den Jahren 1970, 1980, 1990 und 2000.

(siehe unter Dokumentation \ Beschreibung der Variablen)

4.4. CODIERUNG DER FEHLENDEN DATEN

Wie wir bereits gesehen haben, fehlen in der Datensammlung der Volkszählung wegen des mehrstufigen Plausibilitätskontrollverfahrens grundsätzlich nur sehr wenige Informationen. Trotzdem können aus folgenden Gründen Datenlücken entstehen:

1. Gewisse Daten – z.B. Angaben zum Beruf – sind manchmal zu ungenau, um exakt codiert zu werden;
2. Bestimmte Informationen werden manchmal zur Wahrung der Anonymität einer speziellen Kategorie zugeordnet; dies ist jedoch selten geschehen (weniger als 1% der Fälle für die sensibelste Variable) und hat deshalb die Resultate nicht massgeblich verändert;
3. Gewisse Informationen sind alleine wegen der Struktur der Datensammlung unmöglich zu ermitteln; so lassen sich z.B. die Angaben über die Referenzperson des Haushalts, deren Partner bzw. Partnerin oder die Wohnung nur für Privathaushalte, nicht aber für Kollektivhaushalte erheben; weiter machen Daten über die Pendlerbewegungen nur Sinn, falls eine berufliche Tätigkeit oder Ausbildung vorliegt.

Für all diese Fälle haben wir in den Tabellen auf der CD-ROM die Kategorie "INAP" mit dem Wert -1 gebildet³⁰. Die INAP-Fälle wurden grundsätzlich von den Analysen ausgeklammert; so z.B. haben wir sie jeweils nicht in die Berechnungen der Prozentwerte einbezogen und empfehlen auch allen Benutzern, sie systematisch auszuschliessen.

Angesichts unseres Vorgehens bei der Bezeichnung der INAP-Daten sollten sich die Benutzerinnen und Benutzer fragen, ob die von uns gefällten Entscheide für ihr Analysevorhaben sinnvoll sind. Gerade bei Untersuchungen mit Variablen rund um den Haushalt oder die Wohnung ist die Wahl der Referenzpopulation komplex.

³⁰ Die Dokumentationsstandards empfehlen eigentlich die Bildung einer separaten INAP-Kategorie für jeden Grund von nicht zu erhebenden Daten. Aus Datenschutzgründen haben wir jedoch auf diese stärkere Unterteilung verzichtet.

NÄHERE AUSKÜNFTE

über die “Public Use Samples” (PUS) und die Volkszählungen:

Auskunftszentrale für die Volkszählung
Bundesamt für Statistik
Espace de l'Europe 10
2010 Neuchâtel
Tel: 032/713 61 11
Fax: 032/713 67 52
[mailto : info.census@bfs.admin.ch](mailto:info.census@bfs.admin.ch)
internet <http://www.statistique.admin.ch>

über das Bundesamt für Statistik (BFS):

BFS
Espace de l'Europe 10
CH-2010 Neuchâtel
Zentraler Informationsdienst
Tel. +41 032 713 60 11
Fax +41 032 713 60 12
mailto: info@bfs.admin.ch
internet: <http://www.statistik.admin.ch>

über den schweizerischen Informations- und Datenarchivdienst für die Sozialwissenschaften (SIDOS):

SIDOS
Ruelle Vaucher 13
CH-2000 Neuchâtel
Tel. +41 032 721 18 21
Fax +41 032 721 20 74
E-Mail: sidos.service@unine.ch
<http://www.sidos.ch>

für Publikationen des BFS:

Publikationsdienst
Tel. +41 032 713 60 60
Fax +41 032 713 60 61

für Ausleihen bei der Bibliothek:

BFS - Espace public
Espace de l'Europe 10
CH-2010 Neuchâtel
Tel. +41 032 713 60 54
Fax +41 032 713 69 03