

Statistische Techniken und Datenverarbeitung bei Grosszählungen

Von A. Bugmann, Bern

In der Schweiz haben wir zwei Grosszählungen: die eidgenössischen Volkszählungen und die eidgenössischen Betriebszählungen. Durch Gesetz sind die Daten der letzteren auf die Fünfer- und der ersteren auf die mit 0 endenden Jahre festgelegt. In meinem Kurzreferat werde ich mich nur mit der Volkszählung befassen. Sinngemäss sind die hier gemachten Ausführungen auch für andere statistische Erhebungen anwendbar.

Aufgabe einer Volkszählung

In Artikel 5 seiner Verordnung über die Durchführung der eidgenössischen Volkszählung 1970 hält der Bundesrat fest, dass die Erhebung zur Aufgabe habe, «die Wohnbevölkerung der politischen Gemeinden und die Zahl der Wohnungen und Gebäude zu ermitteln und nach den erhobenen Merkmalen zu gliedern». Die Volkszählung soll also der öffentlichen Hand, der Wissenschaft und sonstigen Gremien zahlenmässige Unterlagen für die Lösung wirtschaftlicher, sozialpolitischer und kultureller Probleme auf Landes- und kommunaler Ebene zur Verfügung stellen. Um dieser Anforderung gerecht werden zu können, muss in einer modernen Volkszählung das Auswertungsprogramm gegenüber früher feiner gegliedert sein, d. h. es muss mehr Merkmalskombinationen enthalten, und dies nicht nur für das ganze Land, sondern – wenn immer möglich – bis in die unterste Erhebungseinheit – die politische Gemeinde – hinab. Dies wiederum bedingt eine eingehendere Überprüfung sämtlicher Merkmale eines jeden Formulars. Fehlkodifizierungen, die bei grösseren Gemeinden infolge des breiten Spielraumes unsichtbar bleiben, treten nämlich bei kleineren Gemeinden offen zutage. Im weiteren hat eine Volkszählung die Vergleichbarkeit mit früheren Zählungen zu wahren, denn nur so lässt sich die Entwicklung unseres Volkes in mannigfacher Hinsicht erforschen. Und last but not least sollen die Ergebnisse einer Zählung in möglichst kurzer Frist erhältlich sein.

Fehlerquellen einer Volkszählung

Die eigentliche Durchführung einer Volkszählung ist Sache der Gemeinden. Art. 11 und 14 der bundesrätlichen Verordnung bestimmen: «Die Gemeindebehörde ist verantwortlich für die richtige und vollständige Durchführung der

Zählung im Gemeindegebiet. Sie hat dafür zu sorgen, dass die Fragebogen vollständig und ordnungsgemäss ausgefüllt sind.» Schön und recht. Die Wirklichkeit sieht aber anders aus. Viele Gemeinden unterlassen eine Überprüfung auf Vollständigkeit des Materials, und noch mehr Gemeinden nehmen keine sachliche Überprüfung der Erhebungspapiere vor. Es ist dann Sache des Eidgenössischen Statistischen Amtes, dafür zu sorgen, dass fehlende Bogen noch nachträglich eingehen und fehlende Angaben ergänzt werden. Die eigentlichen Erhebungsformulare werden durch die Bevölkerung selbst ausgefüllt. Die Zähler sollen auf Wunsch wohl behilflich sein; sie sind aber keine Interviewers. Nun beachten die Leute aber oft nicht die Hinweise und Erläuterungen zu den einzelnen Fragen, die ihnen deren richtige Beantwortung erleichtern. Häufig erfüllen auch die Zähler ihre Pflicht nur nachlässig. So fehlt bei rund 15% der Zählformulare die Antwort auf eine Frage, oder sind Fragen falsch beantwortet. Aber auch während der Aufarbeitung werden Fehler entstehen. Hier ist vor allem an falsche Codifizierungen und an Übertragungsfehler beim Lochen zu denken.

Was kann bei einer Volkszählung vorgekehrt werden, dass die erwähnten Fehler und Mängel behoben und wirklichkeitsnahe Ergebnisse innert kürzester Frist erscheinen können?

Optisches Lesegerät

Die Übertragungsfehler fallen in der Volkszählung 1970 durch den Einsatz eines optischen Beleglesers IBM 1288 dahin. Da das Gerät die Daten direkt ab Erhebungsbogen auf Magnetband schreibt, wird das Übertragen der Daten der Fragebogen auf Lochkarten überflüssig. Das Lesegerät «liest» sowohl Strichmarkierungen wie auch Ziffern, und zwar mit einer Geschwindigkeit von rund 4000 Belegen der linken und 5000 der rechten Seite des individuellen Fragebogens pro Stunde.

Die Erhebungspapiere der Volkszählung 1970 wurden dem Belegleser angepasst mit möglichst vielen vorgegebenen Antworten und der Möglichkeit des «Anstreichens». Dadurch fällt für diese Fragen die Arbeit des Codierens weg. Dies ist wichtig, da der Optical Reader stilisierte Ziffern verlangt. Schon die Probezählungen haben gezeigt, dass ein Teil der Signierkräfte, der die gebotene Ausführung der Ziffern nicht sicher genug zustande bringen konnte, ausgeschieden werden musste.

Um der Geheimhaltungspflicht zu genügen, werden die Daten von Grosszählungen anonym gespeichert. Es werden weder Namen noch Adressen von Personen, Firmen usw. auf Band übertragen. An deren Stelle tritt der sogenannte Kontrollbegriff, bestehend aus einer Nummer für die Zählgemeinde und den Zählkreis sowie einer Haushaltsungs-Nummer innerhalb des Zählkreises und einer Fragebogen-Nummer innerhalb der Haushaltung. Aneinandergefügt ergibt das

eine Ziffernreihe, die jeweils nur einmal vorkommt. Damit ist jeder Fragebogen identifiziert. Das Eintragen des Kontrollbegriffes – in stilgerechten Ziffern für den Belegleser – ist ein riesiger manueller Arbeitsaufwand und zudem sehr fehleranfällig. Wir erreichen für die Zählung 1970 eine grosse Erleichterung, indem wir vor jedem Zählkreis ein sogenanntes Meisterblatt legen, das Gemeinde-Nummer und Zählkreis-Nummer sowie die Anzahl der zum Zählkreis gehörenden Belege enthält. Der optische Leser überprüft nun nicht nur die Anzahl der eingelesenen Formulare, sondern übernimmt für alle dem Meisterblatt folgenden Einheiten die Gemeinde- und die Zählkreis-Nummer. Für die Individualkarten sieht die Ersparnis folgendermassen aus: Vorhandene Fragebogen 6,25 Millionen, vorhandene Zählkreise 35 000; Gemeinde- und Zählkreis-Nummer müssen rund 6,2 Millionen mal weniger signiert werden.

Plausibilität

Schwieriger gestaltet sich das Beheben jener Unstimmigkeiten, die sowohl von der Ausfüller- wie von der Aufarbeitungsseite aus vorhanden sind. Vor der Einführung der elektronischen Datenverarbeitung bei Grosszählungen – also vor 1960 und selbst in jener Zählung – wurde nach Abschluss sämtlicher Signierarbeiten eine manuelle Kontrolle und Korrektur der Daten eines jeden Volkszählungsbogens vorgenommen. Dies führte nun vielfach nicht nur zu einer Verbesserung einzelner Daten, sondern oft auch zu einer Verschlimmberung. Jeder Kontrolleur war nämlich bemüht, möglichst viel zu korrigieren. Fehler, die nur aus dem Zusammenhang heraus ersichtlich waren, entgingen dabei oft einer Korrektur. Sie machten sich erst in jenem Zeitpunkt bemerkbar, in welchem die Daten in Tabellenform vorlagen. Als Beispiel möchte ich hier erwähnen, dass in einer Tabelle 1960 plötzlich zwei männliche Hebammen auftauchten. Solche Fälle führten zu zeitraubenden Tabellenkorrekturen. Da der Computer druckfertige Manuskripttabellen liefert, die mittels Offsetverfahren in Buchform gebracht werden, ist heute eine Tabellenkorrektur verunmöglicht oder mit aufwendigen Schreibebeiten verbunden.

Durch den Einsatz elektronischer Datenverarbeitungsanlagen ergeben sich nun für die Kontrolle und Korrektur der Volkszählungsdaten neue Notwendigkeiten und Möglichkeiten; die manuellen Verfahren, die bisher angewendet wurden und die, wie wir gesehen haben, das Ausgangsmaterial unter Umständen nicht verbessern, sondern verschlechtern, werden in ihrer Bedeutung zurückgedrängt. Da manuelle Kontrollen zudem unrationell sind, sollte es oberstes Ziel sein, auf sie gänzlich zu verzichten.

In der Volkszählung 1970 unterlassen wir die manuelle Schlusskontrolle. Auch sonst sollen dem Signierer eigentliche Kontrollaufgaben nicht zugemutet

werden. Es wird möglichst viel der maschinellen Kontrolle und Korrektur überlassen. Rückfragen werden weitgehend vermieden. Durch die maschinelle Plausibilitätskontrolle werden die fehlerhaften Angaben soweit wie möglich aufgedeckt. Ihre Bereinigung erfolgt teils maschinell, teils manuell.

Für die Datenverarbeitung werden die hierzu benötigten Unterlagen von der Fachabteilung in einer sogenannten « Basismappe » zusammengefasst, nämlich einerseits die Problemstellung für die Plausibilität und anderseits das Auszählprogramm auf Spacing-charts.

In der Problemstellung muss genau umschrieben sein, welche Codes bei jedem Merkmal vorkommen dürfen, welche Merkmalskombinationen fragwürdig oder unerlaubt sind und was bei unvollständigem bzw. widersprüchlichem Datensatz einer Person zu geschehen hat.

Grundzüge des Verfahrens

Anhand der auf Grund der Problemstellung durch die Fachleute der Datenverarbeitung erstellten Programme nimmt der Computer die vorgeschriebenen Überprüfungen vor. Zur Erledigung der fehlenden Merkmale bzw. der widersprüchlichen Merkmalskombinationen werden ihm zwei Möglichkeiten offengelassen:

1. die automatische Korrektur,
2. die Meldung von Fehlern oder Unwahrscheinlichkeiten mittels einer Liste.

Bei der automatischen Korrektur wird eine fehlende Merkmalsangabe mit Hilfe von Einsetzungsverfahren bereinigt. Hierbei werden zur Korrektur sonstige vorhandene Merkmalsangaben der Person selbst benutzt oder Informationen von anderen Personen des Haushalts mit herangezogen, d. h. die unvollständige Angabe bei der jeweiligen Person wird aus dem Zusammenhang des Haushalts bereinigt. Als Beispiel für den ersteren Fall möchte ich das Einsetzen fehlender Zivilstandsangaben auf Grund der Geburtsjahrangabe erwähnen. Beispiele für den zweiten Fall sind das Einsetzen fehlender Angaben in der Muttersprache, in der Konfession und in der Heimat bei Kindern auf Grund vorhandener Angaben dieser Merkmale bei den Eltern.

Das erste Beispiel zeigt aber auch die Schwierigkeiten auf, die bei der automatischen Korrektur mittels des Einsetzungsverfahrens auftreten können. Setzen wir nämlich für Personen mit fehlendem Zivilstand für die Jahrgänge 1910 bis 1970 ledig, für die vor 1910 Geborenen den Zivilstand verwitwet ein, dürften einige Korrekturen falsch sein. Das ist weiter nicht schlimm, ist doch dem Computer einprogrammiert, dass er bei der Überprüfung von Merkmalswidersprüchen diese Ledigen und Verwitweten in Verheiratete abändern soll, sofern es

sich um Eheleute handelt, was er anhand der Hinweisziffer Ehepartner ohne weiteres erkennen kann.

Da der Computer ein «Entweder-Oder» nicht kennt und automatische Korrekturen nur nach einer Seite vornehmen kann, müssen bei vielfachen Widersprüchen Prioritäten eingebaut sein. Das beste Beispiel hiezu ist die Berufsplausibilität. Zu jedem Beruf wird dem Computer ein «Ausweich-Code berufliche Stellung», eine «Ersatz-Berufs-Nummer» sowie die für den Beruf wahrscheinlichste Erwerbsart mitgegeben. Treten nun Widersprüche auf, muss zuerst die berufliche Stellung durch den Ausweichcode ersetzt werden. Ist immer noch eine Unstimmigkeit vorhanden, muss der persönliche Beruf in den Ersatzberuf abgeändert werden. Tritt nach dieser Korrektur immer noch ein Widerspruch auf, ist eine vertretbare automatische Korrektur nicht mehr möglich; es muss eine Fehlermeldung erfolgen.

Bei einem Merkmalswiderspruch müssen bei automatischer Korrektur zur endgültigen Abklärung noch weitere Personenmerkmale hinzugezogen werden, während bei einer Erledigung durch Fehlermeldung das Betrachten der beiden widersprüchlichen Merkmale auf dem Beleg in der Regel eine Bereinigung gestattet. Nehmen wir als Beispiel eine Person mit Geburtsjahr 1955 und Zivilstand verheiratet. Durch Konsultation der beiden Merkmale auf dem Beleg kann bei Fehlermeldung die Richtigstellung ohne grossen Aufwand vorgenommen werden. Soll die Korrektur automatisch erfolgen, müssen noch die Merkmale «Stellung in der Haushaltung» und «Erwerbstätigkeit» mitberücksichtigt werden. Ist die Stellung in der Haushaltung Kind und dieses nicht erwerbstätig, kann verheiratet automatisch auf ledig geändert werden. Ist die Stellung aber Haushaltungsvorstand oder Zimmermieter und zudem eine Erwerbstätigkeit vorhanden, ist eine automatische Korrektur unmöglich, da die Angabe im Geburtsjahr offensichtlich falsch ist. Es muss also trotz Beizuges weiterer Merkmale eine Fehlermeldung und damit eine manuelle Korrektur erfolgen.

Um vor unangenehmen Überraschungen gefeit zu sein, soll die Fachabteilung grundsätzlich die Möglichkeit erhalten, maschinell bereits korrigierte Fälle zu überprüfen und gegebenenfalls die Korrekturen zu ändern. Jede Art von maschineller Änderung wird zudem anzahlmässig aufgelistet, so dass bei einem Verdacht auf Unzulässigkeiten jederzeit eingegriffen werden kann.

Fehlermeldungen, d. h. manuelles Korrigieren von Unstimmigkeiten, kommen in der Volkszählung 1970 zur Hauptsache nur noch in der Vorplausibilität vor. In dieser prüft der dem Optical Reader angeschlossene Kleincomputer die eingelesebenen Belege auf Vollständigkeit der Angaben, Fehlermeldungen erscheinen allerdings bei Fehlen von Angaben nur insofern, als für diese keine automatischen Korrekturen vorgesehen sind. Im weiteren wird noch eine Meldung erstattet bei Rückweisungen, d. h. für den Leser nicht lesbare Zeichen und Ziffern, sowie bei unerlaubten Schlüsselnummern. Zu jedem auf der Fehlerliste

aufgeführten Fehler wird eine Verbundkarte geliefert, die neben dem eingestanzten Kontrollbegriff die Art des Fehlers aufzeigt. In jede Verbundkarte wird nun zuerst die korrigierte Angabe im Feld des entsprechenden Merkmals handschriftlich eingetragen und anschliessend eingelocht. Durch Eingabe der so behandelten Verbundkarte wird die Korrektur an die zutreffende Stelle des Magnetbandes übertragen.

Probespiel

Zur Kontrolle der vom Rechnungszentrum erstellten Programme wird ein sogenanntes Probespiel erstellt, und zwar getrennt für Plausibilitätsprogramm und Tabellenprogramm. Die Probespiele bestehen aus fiktiven Erhebungsformularen. Im Probespiel für das Tabellenprogramm müssen weitgehend alle in den Auswertungen auftauchenden Elemente, d. h. Merkmalskombinationen, enthalten sein. Anhand der fiktiven Bogen wird einmal von Hand, das andere Mal maschinell tabelliert und die beiden Ergebnisse miteinander verglichen. Bei Unstimmigkeiten muss sowohl auf der einen wie auf der andern Seite nach den Gründen gesucht werden, bis eine hundertprozentige Übereinstimmung zwischen Hand- und maschineller Auszählung herrscht.

Im Probespiel für die Plausibilität müssen möglichst alle in der Problemstellung aufgeführten Fakten eingebaut sein. Systematisch werden Formulare fehlerhaft ausgefüllt; in einer Buchhaltung wird vermerkt, welche maschinellen Änderungen, welche Fehler- oder Unwahrscheinlichkeitsmeldungen entsprechend der Problemstellung auftauchen müssen. Dann wird wie vorher so lange verglichen und verbessert, bis eine vollständige Angleichung eingetreten ist.

Weil in einem Volkszählungs-Probespiel niemals sämtliche in Wirklichkeit vorkommenden Kombinationen von Fehlern eingebaut werden können, sind wir trotz sorgfältiger Vorbereitung der Zählung nie sicher vor Überraschungen. Hoffen wir für die Volkszählung 1970 das Beste und für ihre Nachfolgerin eine noch eingehendere Vorbereitung.