

Autor: Dr. Egon Jung

Beschreibung des Verfahrens der Münchener Haushaltsstatistik

1. Zusammenfassung der Historie

Das alte Verfahren

Die Münchener Haushaltsstatistik kann auf eine nunmehr fast 30-jährige Tradition zurückblicken. Bereits für eine Verkehrs-Befragung wurde eine Vorstufe des Verfahrens entwickelt, wonach Familienverbände aus den damaligen Personenkenziffern des Einwohner-Melderegisters für den automatisierten Lohnsteuerkarten-Ausdruck konstruiert wurden. Hierbei waren, wie auch heute der Fall, jeder Person, falls vorhanden, die Kennziffern von Ehegatten und Kindern beigeordnet. Suchte man nun pro Gebäude alle anderen Personen ab, deren Kennziffer in diesen Anhängen vorkamen, so ließen sich Familien erstellen, und zwar auch über mehrere Generationen hinweg. Mit Hilfe einer speziellen Entscheidungstabelle wurde vermieden, dass eine Familie mehr als ein Ehepaar enthielt. In den Heften 2, Jahrgang 1980 und 5, Jahrgang 1983 der Münchener Statistik ist dieses Verfahren genau beschrieben.

Frühere Erweiterungen

Von der Einführung eines sogenannten gemeinsamen Ordnungsmerkmals je Person, das auf einen Elternteil verweist und dann die Familienzusammenführung durch einen einfachen Sortiervorgang ermöglicht, wurde bewusst abgesehen. Diese die Programmierung erheblich vereinfachende Maßnahme birgt nämlich einen Informationsverlust in sich, der die Bildung von Mehrgenerationen-Familien unmöglich macht.

Als Ergebnis kam eine Datei mit Datensätzen je Familie heraus, welche zusätzlich mit kleinräumigen Gliederungsmerkmalen wie Stadtbezirk, Teil, Viertel und Block versehen wurden. Dazu kam ursprünglich eine Gebäude-datei aus der Planung, später eine entsprechende, aber stets aktualisierte aus dem Vermessungswesen zur Verwendung.

Die Einführung verschärfter Datenschutzbestimmungen drohte das Verfahren erheblich zu verschlechtern. Es konnten nicht mehr beliebig lange die besagten Kennziffern-Anhänge gespeichert bleiben, sondern nur noch jene für Kinder bis zu 18 Jahren. Eine Simulation in 1985 ergab, dass sich dadurch die Größe der theoretischen Familien um ca. 10% verkleinern würde – eine statistisch unzumutbare Zahl. Es wurde daher überlegt, in wie weit ersatzweise Zuführungen von Einzelpersonen zu Familien diesen Informationsverlust ausgleichen könnten. Man kam überein, dass dies bei gleichem Herkunftsort und/oder Einzugsdatum sowie gemeinsamem Namen zur Wirkung kommen sollte und sprach seitdem von Haushalten.

Eine exakte Beschreibung dieser erweiterten Methode siehe Heft 6, Jahrgang 1991 der Münchener Statistik.

Haushaltsbefragungen allgemein

Die so erhaltene Haushaltsstatistik lief jedes Halbjahr ab. Darüber hinaus kam eine Methode zur Auffindung einer Zielperson je Haushalt für eine Vielzahl von Erhebungen zur Anwendung. Nach den Ziehungen geeigneter Stichproben konnten in einem weiteren Verfahren ergänzende Informationen aus dem Melderegister in datenbanktechnischer Manier gewonnen werden. Mit jeweiliger Kennziffer-Eingabe pro Person lautet der Inhalt des dazugehörigen Datensatzes:

Standarddatensatz für Adressenauswahl

Länge	Bedeutung
2	Stadtbezirk
5	Straßenschlüssel
4	Hausnummer
2	Buchstabe
5	Familienname gekürzt

Länge	Bedeutung	
26	Akad. Grad, Namensbestandteile	
26	Familiename	Anschrift, auf Wunsch infopostgerecht aufbereitet
26	Vorname	
26	evtl. Wohnungsgeber	
26	Straße, Hausnummer, Buchstabe	
26	evtl. Zusatz (z. B. Rückgebäude, etc.)	
26	PLZ, Ort	
8	Geburtsdatum (JJJJMMTT)	
1	Geschlecht (M/W)	
2	Familienstand	
1	D falls deutsch, sonst leer	
3	1. ausländische Staatsangehörigkeit	
3	2. " "	
3	3. " "	
1	Auskunftssperre 1, 3, 4, 6, 9	
1	" 5	
1	" 7	
1	" 8	
1	Wahlausschlussgrund	
1	Wohnungsart (Haupt-, Nebenwohnsitz)	
8	Einzug in die Wohnung (JJJJMMTT)	
4	Wahl/Stimmbezirk	
11	leer	

Datenschutz und Stichprobe

Um einer unbeabsichtigten Verletzung des Personen-Datenschutzes etwa durch Programmierfehler vorzubeugen, werden die gesperrten Fälle der ersten Art mittlerweile erst gar nicht weitergegeben. Zu einem Standardverfahren für Haushaltsbefragungen kam es jedoch nicht. Es wurden nur einzelne Bausteine erstellt, die für die diversen Erhebungen ausgewählt werden konnten. Die Anforderungsspektren der Befragungen waren keineswegs in Einklang zu bringen. Die verschiedensten Schichtungen der Stichproben, Vorweg-Auswahlen aus der Grundgesamtheit oder Datenausgaben, wie z. B. Briefkuvert- oder Adressaufkleber-Ausdruck kamen vor. Speziell für die Erhaltungssatzung wurde eine zusätzliche Methode zur Optimierung der Erheberwege sowie Bereitstellung von Ersatzanschriften möglichst nahegelegener und typgleicher Haushalte entwickelt. Folgende Auswahlverfahren aus den Haushalten wurden durchgeführt:

Zusammenstellung der Haushaltsbefragungen

Inhalt	Zeit
Sanierung Westend	1977
Sanierung Sebastiansplatz	1978
Sanierung "Icho-Block"	1978
Stadtteilsanierung Haidhausen	1979
Reaktionen der Verbraucher auf die neuen rohstoff- und energiekosteninduzierenden Preissteigerungen (Studienprojekt)	1980
Untersuchung der Wohnwünsche und des Wohnungssuchprozesses	1980
Katastrophenschutz im Bereich kerntechnischer Anlagen	1980
Betreuung ausländischer Haushalte mit Kindern im 20. Stadtbezirk	1981
Mietspiegelerhebung	1980-1995
Kindergartensituation in München	1982
Verkehrsberuhigung in Altschwabing	1984
Erhaltungssatzung	1984-1997
Eröffnung des Gasteig Kulturzentrums	1985
Modernisierungsstudie	1985
Verdrängungsuntersuchung	1985
Einzugsstruktur in Neubauten	1986
Parksituation in der Altstadt	1989

Inhalt	Zeit
Verkehrsverhalten der Münchener Bevölkerung (KONTIMUC)	1990
Biotonne	1990
Krippenbedarfarfsplan	1991
Soziale Entwicklung und Lebenssituation der Münchener Bürger	1991
Fortschreibung des Sportstättenentwicklungsplanes	1991
Längsschnittstudie zur Wirkung von Fluglärm auf Kinder und Erwachsene	1991
Papiertonne	1991
Konsequenzen der Bindungsabläufe von Sozialwohnungen	1992
Mikrozensus 1993	1993
AusländerInnen in München	1995
Kinderbefragung in Obergiesing	1995
Fall-Kontrollstudie EHEC	1996
Auskunft für den Deutschen FamilienVerband	1996
Ernährungsverhalten und Ernährungszustand von Kindern und Jugendlichen	1997
Osteoporose-Studie	1997
Gesundheitsbezogene Lebensstile von Vätern	1998
Lärminderungsplan	
Pilotprojekt Teiluntersuchungsgebiet Leuchtenbergring	1999
Nutzung von hauswirtschaftlichen Dienstleistungen	1999
Elternbefragung zum außerhäuslichen Betreuungsbedarf für unter Dreijährige	1999, 2005
Bürgerbefragung 2000	2000
Studie im Rahmen des "Amtsarztlehrgangs"	2000
Forschungsprojekt MOBINET	2000
Wanderungsmotive für den Wegzug aus München in das Umland	2001
Wohnen in der Innenstadt als Lebensstil	2003

Änderungen im Melderegister

2. Verfahrenserweiterungen

Die ursprünglich vierstellig binäre PersonenKennziffer wurde durch ein zwölfstellig numerisches Personen-Ordnungsmerkmal abgelöst, das in der jetzigen Form keine Informationen über Geburtsdatum und Geschlecht mehr enthält. Auf die Familienzusammenführung hatte dies keinen Einfluss. Um aber diese Angaben nicht zu verlieren, haben die Eingabe-Datensätze nun folgenden Inhalt:

Ausgangsdatensatz

Länge	Bedeutung	
5	Straßenschlüssel	
4	Hausnummer	
2	Buchstabe	Wohnungsdaten
8	Einzugsdatum in die Wohnung (JJJJMMTT)	
8	Zuzugsdatum in die Gemeinde (JJJJMMTT)	
12	Gemeindeschlüssel, falls Inland, sonst Nationalitätskennzeichen	
5	Straßenschlüssel, falls in München	Herkunfts-Adresse
4	Hausnummer, falls Inland	
2	Buchstabe, falls Inland	
45	Familienname	
45	Geburtsname	
60	Vornamen	
3	Staatsangehörigkeit	
2	Familienstand	Daten zur Person
2	Religion	
1	Erwerbstätigkeit	
8	Geburtsdatum (JJJJMMTT)	
1	Geschlecht	

Länge	Bedeutung	
2	Anzahl der Ordnungsmerkmale (max. 22)	
12	eigenes Ordnungsmerkmal (OM)	Ordnungsmerkmale
12	evtl. OM des Ehegatten bzw des Kindes	
12	evtl. OM des Kindes	

Neuere Modifikationen

Eine mehrfache Ausgabe von Datensätzen für eine Person ist nicht mehr möglich, was bei mehreren Wohnsitzen in München bisher der Fall war. Die ausgegebenen Datensätze der Haushaltsdatei wurden inzwischen mit Ausnahme der Namen auf sämtliche Informationen der Einwohnersätze der Eingabe erweitert. An Stelle von ursprünglich beabsichtigten Koordinaten sind Grundschulsprenge- und Wahlbezirksnummern aus der Gebäudedatei getreten. Außerdem erhielten die Mitglieder der sogenannten Kernhaushalte, das sind jene, die sich bereits während der Ordnungsmerkmals-Verkettung ergeben, Kennungen, aus denen jeweils der Stand in der Familie hervorgeht. Die Haushalts-Datensätze haben nun folgenden Inhalt:

Ergebnisdatensatz

Länge	Bedeutung
2	Stadtbezirk
1	Stadtbezirksteil
1	Viertel
2	Block
5	Straßenschlüssel
3	Hausnummer
1	Buchstabe
1	leer
5	Grundschulsprenge
1	leer
5	Wahl-/Stimmbezirksnummer
2	leer
1	Kennung für nicht eindeutige Zuordnung
2	Anzahl der Haushaltsmitglieder

zu jedem Haushaltsmitglied oder einmalig für die Zielperson:

- 1 Geschlecht
- 8 Geburtsdatum (JJJJMMTT)
- 3 Staatsangehörigkeit
- 2 Familienstand
- 1 Erwerbstätigkeit
- 8 Einzugsdatum in die Wohnung (JJJJMMTT)
- 8 Zuzugsdatum in die Gemeinde (JJJJMMTT)
- 12 Gemeindegeschlüssel des Herkunftsortes, falls Inland, sonst Nationalitätskennzeichen
- 5 Straßenschlüssel der Herkunftsadresse, falls in München
- 4 Hausnummer der Herkunftsadresse, falls Inland
- 2 Buchstabe der Herkunftsadresse, falls Inland
- 2 Religion
- 12 Ordnungsmerkmal
 - 1 Kennung: E für Ehegatte/in
 - Kennung: K für Kind
 - Kennung: A für Alleinerzieher/in
 - Kennung: L für Alleinerziehendes Kind
 - Kennung: G für Großelternteil im Haushalt mit Ehepaar
 - Kennung: H für Großelternteil im Haushalt ohne Ehepaar
 - Einzelperson

Zusatzinformationen

Zusätzlich werden bei jedem Lauf weitere Dateien ausgegeben:
a) Ein allgemeines Fehlerprotokoll, worin die wenigen Ungereimtheiten der Einwohner-Eingabe aufgelistet sind und das zur Pflege des Melderegisters herangezogen werden kann.

- b) Eine Liste der ebenfalls wenigen Fälle, die aus dem Vergleich von Einwohner- und Gebäudedatei nicht eindeutig zugeordnet werden können. Damit lässt sich letztere Datei aktualisieren.
- c) Eine Datei kompliziert erscheinender Haushalte, die aus dem gewohnten Denkraum fallen, wie z. B. solche mit mehreren Alleinerziehern oder Ehepaaren und Alleinerziehern etc. in Verbindung mit mehr als einer Generation. Diese Fälle sind sehr selten, in München weniger als 100. Statistisch sind sie daher ohne Belang, könnten aber Hinweise auf versteckte Unstimmigkeiten im Verfahrensablauf geben.

Kurzauswertung

Die zusammenfassende, meist einseitige Kurzauswertung führt die auftretenden Haushalte mit mehr als 20 Personen auf. Durch die Zuführung von Einzelpersonen nach den Namens- und Zuzugskriterien können solche Fälle besonders bei Ausländern durchaus auftreten, fallen aber statistisch nicht ins Gewicht. Danach werden die Anzahlen von Einwohnern und Haushalten sowie deren auch geschlechtsweise getrennten Zielpersonen und weiter die der Zuordnungen von Einzelpersonen stufenweise getrennt angegeben. Außerdem erhält man die Gesamtzahlen der Alleinerziehungs-Haushalte, die der Haushalte, bei denen zwar Vater und Mutter im gleichen Gebäude wohnen, aber nicht untereinander, sondern nur über die Ordnungsmerkmale der Kinder verbunden sind, sowie jene von irregulären Fällen. Schließlich erfolgt eine Zahlenangabe und Häufigkeitsanalyse sogenannter Kinderhaushalte, die zwar unrealistisch sind, deren Zustandekommen aber vom Verfahren leider nicht vermeidbar ist, wenn z. B. Kinder bei Großeltern ohne die Eltern wohnen.

3. Programmiertechnik und Laufzeitverhalten

Rechenzeitverhalten

Es mag antiquiert erscheinen, aber die seit jeher ausgeübte Methode, nämlich die ausschließliche Verwendung sequenzieller Dateien und die des Hauptspeichers, blieben erhalten. Die Möglichkeiten einer Datenbank mit einer darauf orientierten Sprache wie SQL sind zwar verlockend und würden die Programmierung erheblich vereinfachen. Man muss aber stets bedenken, dass dann z. B. eine einzige Suchanweisung nicht nur den um Größenordnungen langsameren Zugriff auf den Hintergrundspeicher, sondern je nach Geschick der Datenbank-Einrichtung auch Rechenzeiten von u. U. polynomialer Laufzeitcharakteristik auslösen kann. Die eigentlich für die Einzelsatzverarbeitung geschaffene Datenbanktechnik stellt sich somit für Massendaten-Auswertungen aus dem Melderegister als ungeeignet in Hinblick auf brauchbare Rechenzeiten heraus.

Der hier eingeschlagene Weg führt, ausgehend von einem sequentiellen Auszug aus dem Melderegister in ca. 150-160 CPU-Sekunden auf einer Siemens S140 Anlage, Ausbaustufe 10C, zu einer wiederum sequentiellen Haushaltsstatistik. Allerdings ist bei einer Einwohnerzahl von 1,4 Mill. wie in München ein verfügbarer Hauptspeicher von 175 Megabytes nötig. Da Speicher wie Laufzeit sich weitgehend proportional zur Anzahl der verarbeiteten Personen verhalten, lässt sich der Bedarf je Gemeinde abschätzen. Bei entsprechendem Rechner-Hintergrund ließe sich die Methode auch bundesweit anwenden, eine einheitliche Dateneingabe vorausgesetzt. Für den Fall, dass kein Adressraum in o. g. Größe vorhanden ist, kann auf Banddateien ausgewichen werden. Dies allerdings auf Kosten weiterer CPU-Zeit, im Falle Münchens wären dann etwa 200 CPU-Sekunden mehr zu veranschlagen.

Auch mit wachsender Einwohnerzahl je Gebäude, in München bis zu etwas über 1 000, verhält sich das Laufzeitverhalten quasi linear ($n \cdot \log(n)$). Grund dafür ist, dass sämtliche Suchverfahren nach der Bisektionsmethode ablaufen. Die hierzu benötigten Adressfelder zur Beschreibung der aufsteigenden Ordnungen werden in Sortierungen im Hauptspeicher in einfacher Weise durch wiederholtes Trennen und Mischen geordneter Sequenzen gewonnen.

Als Programmiersprache kommt Fortran-90 zur Anwendung. Dies hat vor allem historische Gründe, da schon 1979 die Münchener Haushaltsstatistik lief, als diese Sprache noch weitere Verbreitung hatte. Eine Umsetzung des etwa 1 000 Zeilen langen Programmcodes in eine andere Sprache wäre problemlos. Es sollte dabei jedoch bedacht werden, dass im Sinne kurzer Laufzeiten das Programm als ganzes in die Maschinensprache übersetzt und nicht befehlsweise interpretiert wird. Außerdem lässt sich Fortran in der Regel vom Compiler optimieren, was erfahrungsgemäß Rechenzeitvorteile von über 50% ergibt.

Künftige Aspekte

4. Ausblick

Wie es derzeit den Anschein hat, kann das Haushaltsgenerierungsverfahren als gesättigt angesehen werden. Die bereits jahrzehnte lange Anwendungsreihe gibt der Erforschung demografischer Trends weitere Möglichkeiten. Einer größeren Verbreitung als bis jetzt steht nur wenig im Wege, da sowohl für die Datenbestände einfachste Normen bestehen, als auch die Anforderungen an die Rechner-Umgebung äußerst gering sind. Weder die Menge der Daten noch die entstehenden Rechenzeiten lassen hier eine Barriere erkennen. Sollten sich in der Zukunft haushaltsrelevante Faktoren zeigen, die heute noch nicht bekannt sind, so lässt sich das kurz und überschaubar gehaltene Programm in einfacher Weise darauf anpassen.

Anmerkung der Redaktion:

Der Autor, Dr. Egon Jung, ist als Dipl.-Mathematiker im Direktorium der Stadt München, Haptabteilung III – ITD – 2.02, im Sachgebiet Technisch-Wissenschaftliche Aufgaben tätig.