Bern, Februar 2000

Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung

Beat Hulliger und Philippe Eichenberger

Service de méthodes statistiques
Dienst Statistische Methoden
Servizio metodi statistici

Auskunft:

Beat Hulliger

Bundesamt für Statistik
Dienst Statistische Methoden

Espace de l'Europe 10, CH-2010 Neuchâtel
Tel.: (+41 31) 32 36015
Fax: (+41 31) 32 36375
Bestellnummer: 338-0000
© BFS/OFS

Am 1. Januar 1998 trat das neue Fernmeldegesetz in Kraft, welches die Telekommunikation liberalisierte. Die Telecom PTT wurde in ein privatwirtschaftlich organisiertes Unternehmen, die Swisscom, umgewandelt. Es musste eine neue Lieferungsvereinbarung für die Telefonnummern, welche die Grundlage für das Stichprobenregister für Haushalterhebungen (SRH) bilden, ausgehandelt werden. Der vorliegende Bericht beschreibt die methodischen, organisatorischen und informatischen Umstellungen des SRH. Dabei wurde aber die ursprüngliche Konzeption von Philippe Eichenberger und Michel Kolly übernommen. Der Bericht soll ausserdem eine Art Handbuch für die Erstellung des SRH und die Ziehung von Stichproben daraus sein. Die Arbeiten wurden von Michel Kolly (Sektion EWL) und Iba Ndiaye (Sektion PREIS) begleitet. Die Programmierarbeiten wurden von Angelo Cenci (BFI, Teil TERCO), Falk Möser (CSC Ploenzke, Teil MPAN) und Markus Eichenberger (BFI, SAS-Macros) durchgeführt.

INHALTSVERZEICHNIS

1 EINLEITUNG 1

1.1 AUSGANGSLAGE 1

2 ZWECK DES STICHPROBENREGISTERS FÜR HAUSHALTSERHEBUNGEN 2

2.1 ZUFALLSSTICHPROBEN UND STICHPROBENRAHMEN 2
2.2 ANFORDERUNGEN AN EINEN STICHPROBENRAHMEN FÜR HAUSHALTERHEBUNGEN 2

3 VORSCHLAG SWISSCOM 3

4 ANALYSE DER DATENLIEFERUNG SWISSCOM 6

5 STATISTISCHE PROBLEME 9

5.1 SCHICHTUNGEN NACH POSTLEITZAHLEN 9
5.2 AUSGESCHLOSSENE BEREICHE 11
5.3 DOPPELADRESSEN 11
5.4 EINHEITEN IM ETV UND IN DER ADRESSLIEFERUNG DER SWISSCOM 11

6 INFORMATIK-PROBLEME 14

6.1 UMSTELLUNG BISHERIGER INFORMATIK-APPLIKATIONEN 14
6.2 ETV-BATCHMAN 15
6.3 SAS-MACROS 15

7 ABLAUFSCHEMA 16

7.1 UEBERTRAGUNG AUF DEN HOST 16
7.2 VERARBEITUNG AUF DEM HOST 17

8 VORGEHEN BEIM ZIEHEN EINER STICHPROBE 18

9 OFFENE FRAGEN 19

9.1 BEI INTERVIEWS ZU STELLENDE FRAGEN 19
9.2 GEWICHTUNGSPROBLEME 19

10 ANHANG 20

10.1 FRÜHERE DATENLIEFERUNG 20
10.2 DATEI-ORGANISATION UND JOBS DER APPLIKATION TERCO 20
10.3 GNRMERGE 20
10.4 SAS-MACRO AUFBEREITUNG KORRESPONDENZTABELLE 21
10.5 SAS-MACRO UMWANDLUNG SCHICHTUNG NACH GNR IN SCHICHTUNG NACH PLZ6 23
10.6 PROGRAMME S-PLUS FÜR KORRESPONDENZ GNR-PLZ6 25
10.7 VERARBEITUNGSMELDUNG ZU ECPBTTEA (TEL991) BIS MPERRLD 25
10.8 READSRH FÜR DIE KONTROLLE DER SEQ. DATEI D663.A663SWC.ADR984 26
10.9 INDSRH ZUM LESEN UND AGGREGIEREN DES DATENBANK-INDEX 27
10.10 STAND DER DATENBANK (MIT DBDC-TOOLS IN COMPLETE TEST-UMGEBUNG) 27
10.11 UEBERSICHTSTABELLEN ÜBER DIE ZUGELASSENEN PROJEKTE 27
10.12 AUFBEREITUNG DER EINGANGSDATEIEN FÜR ECPBMEXT 28
10.13 OUTPUT VON JOB P663EXT IN TSO BETA 28
10.14 AUSGABE-FORMAT 29
10.15 UMWANDLUNG OUTPUT IN ASCII MIT PROGRAMM OUTREAD 29

1 Einleitung

1.1 Ausgangslage

Das Bundesamt für Statistik (BFS) benutzte seit 1990 eine elektronische Kopie des Telefonbuchs der Telecom PTT (zuerst das TERCO und später das elektronischen Telefon-Verzeichnisses ETV), um einen Stichprobenrahmen für Haushalterhebungen zu erstellen. Vermutlich sind über 95% der Schweizer Haushalte über einen Telefonanschluss erreichbar (Genauere Daten werden erst wieder dank der Volkszählung 2000 erhältlich sein). Der Eintrag der Telefonnummern im Telefonverzeichnis war lange obligatorisch (mit wenigen Ausnahmen). Der Stichprobenrahmen kann also beanspruchen, dass die Ueberdeckung der Bevölkerung gut ist, und dass er aktuell ist (4 mal jährlich aufdatiert).
Die Datenlieferung der Telecom PTT umfasste neben Fernkennzahl und Rufnummer u.a. auch Name, Vorname, Beruf, Strasse, Adresse, Werbesperrung, aber keine Unterteilung in private und geschäftliche Nummern. Das BFS bereinigte in einem relativ aufwendigen Prozess die Daten und erstellt vierteljährlich einen Stichprobenrahmen für Erhebungen bei Privat-Haushalten.
In den letzten Jahren haben folgende Erhebungen den Stichprobenrahmen für Haushalterhebungen (SRH) des BFS benützt:

- Schweizerische Arbeitskräfteerhebung (SAKE): jährlich ca. 24'000 mit CATI
- Mietpreiserhebung: vierteljährlich ca. 10'000 schriftlich
- Miestrukturerhebung: 1996 ca. 280'000 schriftlich
- Gesundheitsbefragung: 1993, 1998 ca. 18'000 mit CATI und schriftlich
- Verkehrsbefragung: 1994, 1999 ca. 15'000 mit CATI
- Familienbefragung: 1994, ca. 15'000 mit CATI und CAPI

Jährlicher Bedarf an Adressen

Jahr	Benützte Adressen (Schätzung)
1994	64'000
1995	34'000
1996	314'000
1997	52'000
1998	70'000

1.1.1 Das neue Fernmeldegesetz

Das Fernmeldegesetz vom 30. April 1997 wurde am 1. Januar 1998 in Kraft gesetzt. Es bricht mit dem Monopol der Telecom PTT für die Betreibung von Telefonnetzen. Im Januar 1999 waren ca. 170 Anbieter von Fernmeldedienstleistungen beim Bundesamt für Kommunikation (BAKOM) angemeldet. Neben der Telecom PTT, neu Swisscom genannt, haben die Firmen Diax und Orange Communication je Konzessionen für die Betreibung eines Netzes für Mobiltelefonie erhalten. Es werden vermutlich längerfristig neben der Swisscom zwei bis drei grössere Anbieter von Telefondiensten in der Schweiz entstehen. Das BAKOM ist zuständig für die Zuteilung von Telefonnummern an die Konzessionäre. Bis im Dezember 1999 hat kein Konzessionär Anschlüsse an ein festes Telefonnetz für Privat-Haushalte angeboten. Damit laufen alle Anschlüsse von Privat-Haushalten immer noch über das

Netz der Swisscom. Mindestens Diax und Sunrise wollen aber mittelfristig auch Festnetz-Anschlüsse an Private anbieten. Diese Anschlüsse könnten z.B. über das Stromnetz oder über Kabelfernseh-Anschlüsse erfolgen.
Der Eintrag von zugeteilten Telefonnummern in ein Telefonverzeichnis ist nicht mehr obligatorisch. Ausserdem sind nur die Konzessionäre für die Grundversorgung verpflichtet, Verzeichnisse zu veröffentlichen.
Es stellen sich mehrere Probleme:

1. In Zukunft wird es vermutlich kein zentrales Telefonverzeichnis mehr geben, sondern die verschiedenen Konzessionäre werden verschiedene Verzeichnisse herausgeben.
2. Das BFS kann sich längerfristig nicht mehr nur auf die veröffentlichten Telefonnummern abstützen, weil eine unbekannte Anzahl von Telefonnummern nicht mehr in den Verzeichnissen aufgeführt werden wird. Erfahrungen im Ausland zeigen, dass 10 bis 30% der Telefonnummern nicht veröffentlicht werden.
3. Die Anzahl Privat-Haushalte mit mehreren Festanschlüssen und mit zusätzlichen oder ausschliesslichen Mobiltelefonen nimmt zu.

2 Zweck des Stichprobenregisters für Haushaltserhebungen

2.1 Zufallsstichproben und Stichprobenrahmen

Eine methodisch saubere Stichprobe basiert auf einer Zufallsstichprobe aus der **Grundgesamtheit**. Der Zugang zu den Einheiten der Grundgesamtheit muss durch eine Liste, den **Stichprobenrahmen**, ermöglicht sein. Aus den Einträgen der Liste wird ein Teil, die **Stichprobe**, zufällig ausgewählt. Dabei muss jeder Eintrag eine positive Wahrscheinlichkeit haben, gewählt zu werden. Nur zu den Einheiten der Stichprobe wird der **Zugang** tatsächlich benötigt, weil nur über die Stichprobe Information gesammelt wird. Die Ergebnisse der Stichprobe können auf die Grundgesamtheit hochgerechnet werden. Voraussetzung für die Hochrechnung ist die Kenntnis des Auswahlprozesses in der Grundgesamtheit. Da sowohl Population als auch der Stichprobenrahmen sich ständig verändern, sprechen wir von einem Register.

2.2 Anforderungen an einen Stichprobenrahmen für Haushalterhebungen

Als Qualitätskriterien für einen Stichprobenrahmen können genannt werden:

1. Ueberdeckung der Grundgesamtheit
2. Eindeutigkeit (keine Mehrfacheinträge)
3. Aktualität
4. Richtigkeit
5. Strukturierbarkeit (insbes. regionale).

Ein Stichprobenrahmen mit möglichst guter Ueberdeckung der Grundgesamtheit ist eine wichtigte Voraussetzung, um eine Zufallsstichprobe zu ziehen und damit um eine wissenschaftlich saubere, sogenannt repräsentative Erhebungen durchzuführen.
Falls nicht alle Telefonnummern in den Stichprobenrahmen aufgenommen werden, dann werden gewisse Haushalte und Personen von der Erhebung nicht erfasst. Die Ueberdeckung der Population ist also nicht gewährleistet. Schätzer, die nur auf den im Stichprobenrahmen erfassten Haushalten und Personen basieren, können im Prinzip nichts über die unerfasste Population aussagen. Mit anderen Worten haben

die Schätzer systematische Fehler, falls sich die nicht erfassten Haushalte und Personen in ihren Charakteristiken vom Rest der Grundgesamtheit unterscheiden.

Beispiel: Aerzte beginnen systematisch ihre private Telefonnummer aus den Verzeichnissen zu streichen. Eine Haushalt-Erhebung auf einem Stichprobenrahmen, der nur auf den veröffentlichten Adressen basiert, kann also keine Haushalte von Aerzten erfassen. Es kann also keine Aussage über diese Haushalte gemacht werden. Falls sich Aerzte in ihrem Gesundheits-Verhalten vom Rest der Bevölkerung unterscheiden, dann sind auch Schätzer für die gesamte Bevölkerung verfälscht.

Die Einträge im Stichprobenrahmen sollten möglichst informativ sein. Das Minimum ist eine Telefonnummer ohne Zusätze. Besser ist eine Telefonnummer mit Namen und Vornamen. Um eine regionale Schichtung der Stichprobe vornehmen zu können sind Postleitzahlen und ev. weitere Adress-Elemente notwendig. Ein Code zur Unterscheidung von geschäftliche benützten Telefonnummern verhindert, dass Telefonnummern, die ausschliesslich geschäftlichen Zwecken dienen, für eine Erhebung über private Haushalte oder natürliche Personen benützt werden.

Am wichtigsten ist, dass eine geographische Zuordnung der Telefonnummer möglich ist. Nur damit kann eine Schichtung nach Regionen und eventuell ein Aufstockung für Kantone etc. durchgeführt werden. Die Telefonnummer sollen im Jahr 2001 auf 10 Ziffern erweitert werden, wobei die Fernkennzahl wegfallen würde. Die Telefonnummern sollen dann ohne Bindung an eine Zentrale verteilt werden. Auch können sich in Zukunft die Zuordnungen der Telefonnummern auf Zentralen verwischen, weil die Konzessionäre verschiedene Versorgungsgebiete, die nicht abgrenzbar sind, verwenden. Wir müssten also zu einer Telefonnummer z.B. noch eine Postleitzahl erhalten.

2.2.1 Aktualität

Das ETV ist einem ständigen Wandel unterzogen. Die nachfolgende Tabelle gibt einen Eindruck, wie hoch die Mutationsrate im vom ETV abhängenden SRH ist.

Tabelle 2: Aenderungen im SRH

	Einträge	
SRH vom 15.9.1998	3'640'294	
Verschwundene Nummern	155'850	4.3% von 3'640'294
Neue Nummern	132'210	3.7% von 3'616'654
SRH vom 12.12.1998	3'616'654	
Nummern mit neuer PLZ	8'254	0.2% von 3'616'654

Mit anderen Worten muss pro Quartal mit ca. 4% Erneuerung gerechnet. Eine vierteljährliche Erneuerung ist also angebracht.

3 Vorschlag Swisscom

Die Swisscom hat am 30. Oktober 1997 eine Preiserhöhung für die ETV-Adressen auf 30 Rappen angekündigt. Dies hätte zu unmöglichen Kosten für den Stichprobenrahmen geführt (ca. 5 Mio. CHF jährlich). In Verhandlungen zwischen

der Swisscom, den Marktforschungsinstituten und dem BFS wurde eine Lösung erarbeitet.

Ein erster Vorschlag des BFS war, neben einer niedrigen Pauschale nur die effektiv für die Stichproben gezogenen Adressen zu bezahlen. Dies hätte einen relativ grossen Aufwand bei der Verrechnung ausgelöst, hätte aber die Namen und Adressen gesichert. Die Swisscom machte den Vorschlag, zu einem Pauschalpreis alle Telefonnummern sowie eine Postleitzahl und klassifizierende Merkmale aber ohne identifizierende Merkmale wie Namen und Adressen zu liefern. Diese Lösung machte das eventuell heikle Problem hinfällig gemacht, dass eine Bezahlung pro Einheit vereinbart worden wäre.

Die Selektionsdaten für die Datenlieferung der Swisscom sind: 3. Montag im März, Juni und September, 2. Montag im Dezember. Das neue Stichprobenregister ist jeweils ca. 10 Tage nach der Selektion für die Stichprobenziehung verfügbar.

Pro Telefonnummer werden folgende Angaben gemacht: Fernkennzahl, Rufnummer, Werbesperrung (0,1), Postleitzahl (6-stellig), Dienst (TEL, NATEL), Sprache (D,F,I,R,E), Kundentyp (Privat=RES, Geschäft=BUS).

Pro Standort und Kunde wird nur eine Telefonnummer gemeldet. D.h. Geschäfte oder Private (z.B. ISDN), die am selben Standort mehrere Telefon auf dem fixen Netz betreiben, werden erkannt durch die Swisscom erkannt und bereinigt. Es werden aber alle NATEL-Nummern geliefert. Privatkunden, die am selben Standort eine oder mehrere Telefonnummern privat und geschäftlich benützen werden doppelt geliefert (einmal als privat, einmal als geschäftlich). Da bei älteren Telefonnummern nicht konsequent nach privat oder geschäftlich unterschieden worden ist und auch heute die Unterscheidung durch den Kunden auf dem Antragsformular erfolgt, muss hier mit einer gewissen Unschärfe gerechnet werden.

Die Daten stammen aus dem elektronischen Telefonverzeichnis der Swisscom und aus dem Verzeichnis der NATEL-Nummern. Die Eintragung ins NATEL-Verzeichnis wird aber oft nicht gemacht. Insbesondere werden NATEL-Easy Nummern (Mobiltelefonie mit Vorauszahlung) nicht eingetragen. Ein NATEL-Easy Anschluss wird in einem Detail-Handelsgeschäft bezogen und bezahlt und ist unmittelbar geschaltet. Der Kunde kann sich freiwillig bei der Swisscom registrieren lassen, wird aber nicht in ein Verzeichnis eingeschreiben.

Eine Arbeitsgruppe (Kolly, Bender, Ndiaye, Spöcker, Eichenberger, Hulliger) hat den Vorschlag der Swisscom beurteilt. Sie kommt zum Schluss, dass auch ohne Namen und Adressen ein Stichprobenrahmen hergestellt werden kann, der qualitativ mit der bisherigen Lösung vergleichbar ist. Dies gilt, solange das ETV der Swisscom praktisch die ganze Schweiz abdeckt. Da die Swisscom auch die Telefonnummern liefert, die nicht im ETV eingetragen sind, ergibt sich hier sogar eine Verbesserung gegenüber dem bisherigen Zustand. Die Arbeitsgruppe empfiehlt, den Vorschlag der Swisscom anzunehmen.

Die Umstellung auf die reduzierte Datenlieferung der Swisscom wurde am 30. März 1999 von der Geschäftsleitung des BFS bewilligt. Ebenso wurden die für die Umstellung benötigten Kredite gesprochen. Das Kostendach für die Umstellung und

wurde auf Fr. 78'000 festgelegt (Je Fr. 25'000 für die Umstellung der Applikationen TERCO und MPAN, Fr. 25'000 für eine eventuelle neue Applikation zum Herunterladen von Adressen, Fr. 3'000 für einen neuen Datensatz für eine Korrespondenztabelle zwischen Gemeindenummern und Postleitzahlen. Weitere Kosten für die Umstellung von einzelnen Erhebungen, z.B. bei der Applikation PRESTA, sind nicht in den obigen Zahlen enthalten). Die tatsächlichen Kosten sind wesentlich höher, da v.a. interne Arbeitsleistungen des BFS und des BFI anfielen. Der Dienst Statistische Methoden wurde mit der Federführung für die Umstellung betraut. Die Umstellung des TERCO wurde durch A. Cenci vom BFI durchgeführt. Der Kredit von Fr. 25'000 für die Umstellung von MPAN wurde ausgeschöpft (Falk Möser, CSC Ploenzke). Die Anschaffung von ETV-Batchman kostete Fr. 14'000. Die Installation der Klienten für ETV-Batchman funktioniert nicht.

Die Swisscom wollte gegenüber dem BFS keine Lieferverpflichtung einzugehen. Mit Brief vom 22. April 1998 wurde das BAKOM angefragt, ob es eine Möglichkeit gibt, die Swisscom und eventuelle andere Anbieter aufgrund des FMG zur Lieferung der Telefonnummern zu verpflichten. In zweiter Linie interessierte, ob man sogar auf einer kostenlosen Lieferung bestehen könnte. Laut Brief BAKOM vom 18. Mai 1998 und weiteren Abklärungen besteht im FMG weder für eine Lieferverpflichtung noch für die Gratislieferung eine Rechstgrundlage. Längerfristig sollte versucht werden, die Anbieter von Fernmeldediensten aufgrund des Bundesstatistikgesetzes zur Lieferung zu verpflichten. Dabei sollte auch die Form der Lieferung festgelegt werden.

Das BFS kann auch keine Stichproben für Dritte ziehen, nicht einmal, wenn Gebühren an die Swisscom abgeführt würden. Hingegen ist klargestellt, dass die gesamte Bundesstatistik das Stichprobenregister benützen kann.

Die Swisscom hat dem BFS noch eine Lieferung von vollständigen Adressen per 15. April 1998 zugestellt.

Am 17. Juli 1999 hat der Preisüberwacher die Erhöhung der Verkaufspreise für Adressdaten durch die Swisscom als überhöht und willkürlich abgelehnt. Da die Marktpreise für Adressdaten immer noch bei ca. 220'000 Franken pro Lieferung liegen, bleibt die Einschätzung der Situation für das BFS unverändert.

4 Analyse der Datenlieferung Swisscom

Lieferung vom 12. September 1998:
Verteilung der Vorwahl (FKZ)

FKZ	Absolut	%
1	577541	13.3
20	15	0
207	1	0
209	19	0
21	255043	5.9
22	254960	5.9
24	73000	1.7
26	117030	2.7
27	157542	3.6
31	223762	5.2
32	296872	6.8
33	118959	2.7
34	40616	0.9
40	0	0.0
41	294488	6.8
50	662	0
51	15	0
52	168442	3.9
53	2	0
55	97346	2.2
56	130543	3
59	1	0
61	283508	6.5
62	210304	4.9
71	276290	6.4
74	0	0.0
75	18329	0.4
77	124347	2.9
79	245627	5.7
81	167400	3.9
91	202364	4.7
92	1	0

Bei den Nummern mit Vorwahl 077 handelt es sich um alte NATEL-Nummern, die nicht mehr gültig sind.

Nummern mit Werbesperre

Werbesperre	Frequency	Percent
0	3301895	76.2
1	1033134	23.8

Die im ETV nicht eingetragenen Nummern erhalten ebenfalls eine Werbesperre.

Sprache

SPR	Frequency	Percent
D	3087861	71.2
F	1031052	23.8
I	215934	5.0
R	183	0.0

Später kommt dann Englisch als Sprache hinzu.

Verteilung des Kundentyps und des Dienstes

	Dienst			
Kundentyp	NATEL	TEL	Total	In Prozent
BUS (geschäftlich)	70'204	537'642	607'846	14.02
RES (privat)	300'467	3'426'717	3'727'184	85.98
CH	370'671	3'964'359	4'335'030	
In Prozent	8.55	91.45		100.00

Die Lieferung vom 15. Dezember 1998: Dienst und Kundentyp

	NATEL	TEL	Total	In Prozent
BUS	71'553	537'219	608'772	14.24
RES	258'855	3'406'302	3'665'157	85.76
CH	330'408	3'943'521	4'273'929	
In Prozent	7.73	92.27		100.00

Möglicherweise hat die Swisscom bereits einige Kunden verloren, vor allem bei den NATEL oder aber die Swisscom hat begonnen das ETV zu bereinigen (ungültige NATEL wurden gelöscht). Eine weitere mögliche Erklärung ist, dass die Bereinigung der Adressen vor der Lieferung an das BFS besser geworden ist.

Die Daten der Swisscom werden für das SRH aufbereitet, d.h. ungültige, doppelte und rein geschäftliche Telefonnummern werden eliminiert. Das SRH, das auf der Lieferung nach neuem Modus vom Dezember 1998 beruht, wurde mit dem letzten SRH, welches auf einer Lieferung nach altem Modus beruht, verglichen.

Vergleich SRH vom 12.12.1998 und vom 15.4.1998:

Vorwahl	Absolut	%	Absolut	%
1	497755	13.8	466620	14.6
20	13	0	0	0
21	222284	6.1	205476	6.4
22	221766	6.1	201359	6.3
24	62674	1.7	56823	1.8
26	100453	2.8	91968	2.9
27	129078	3.6	113303	3.5
31	195873	5.4	180173	5.6
32	255345	7.1	238578	7.5
33	102013	2.8	90348	2.8
34	35470	1	33073	1.0
40	0	0	1	0.0
41	251919	7	236325	7.4
51	4	0	4	0.0
52	146880	4.1	138227	4.3
55	83424	2.3	78194	2.4
56	112788	3.1	107076	3.4
61	244600	6.8	230647	7.2
62	181273	5	170523	5.3
71	237505	6.6	221309	6.9
74	0	0	377	0.0
75	14855	0.4	12828	0.4
77	15	0	0	0.0
79	211471	5.8	40437	1.3
81	139608	3.9	126295	4.0
91	169588	4.7	152266	4.8
CH	3616654		3192230	

Die Zahl der Einträge hat um 13.3% zugenommen. Dabei betrug die Zunahme ohne die FKZ 079 (NATEL) 8.0%, während die NATEL-Nummern um mehr als das 5-fache zugenommen haben. Im SRH erhalten rein private Nummern den Telefon-Code 1, während gemischte Nummern den Telefon-Code 2 erhalten. Im neuen SRH können weniger Telefone eindeutig als privat klassiert werden.

Aufteilung nach Dienst und Telefon-Code vom

Datum	Dienst	Tel.-Code 1	Tel.-Code 2	Total	%
12.12.98					
	NATEL	211499	0	211499	5.85
	TEL	3026354	378801	3405155	94.15
	Total	3237853	378801	3616654	
	%	89.53	10.47		100
15.4.98					
		3109597	82633	3192230	
	%	97.4	2.6		

Die Zunahme von Tel-Code 2 ist also vermutlich hauptsächlich auf das Fest-Netz zurückzuführen. Mit anderen Worten konnten viel weniger Nummern klar als

Geschäftsnummern erkannt werden als im alten SRH! Vermutlich werden also relativ viele nicht eingetragene Nummern sowohl als BUS als auch als RES geführt.

Verteilung der Sprachen im SRH

Sprache	Absolut 15.12.98	%	Absolut 15.4.98	%
D	2570058	71.1	2288139	71.7
E	40	0		
F	871189	24.1	746726	23.4
I	175206	4.8	157365	4.9
R	160	0		
X	1	0		

Die Verschiebungen in der Verteilung der Sprachen sind also minim.

5 Statistische Probleme

5.1 Schichtungen nach Postleitzahlen

Da keine Adressen mehr vorhanden sind, kann der Stichprobenrahmen nicht mehr direkt nach politischen Gemeinden geschichtet werden. Dies ist aber z.B. notwendig, um für Kantone eine Aufstockung der Stichprobe zu ermöglichen. Statt dessen müssen Schichtungen nach politischen Gemeinden mit Hilfe einer gewichteten Korrespondenztabelle in Schichtungen nach Postleitzahlen übertragen werden (zu Korrespondenztabellen vgl. Hulliger, B. (1998) Linking of Classifications by Linear Mappings, Journal of Official Statistics Vol. 14 No. 3, Statistics Sweden). Beim Uebergang zu Postleitzahlen ergibt sich eine gewisse Unschärfe in der Schichtung. Dies hat auf die Validität der Resultate keinen Einfluss. Hingegen steigt die Varianz u.U. leicht an. Die Korrespondenztabelle sollte idealerweise 4 Mal pro Jahr mit neuen Postleitzahlen- und Gemeinde-Daten aktualisiert werden. Im Moment steht aber höchstens einmal pro Jahr eine aktualisierte Postleitzahlen-Datei zur Verfügung. Der Dienst METH hat auf der Basis von Hektar-Daten aus dem Jahr 1997, welche die Sektion RAUM bereit gestellt hat, eine provisorische Korrespondenztabelle erstellt. Dabei wurden von RAUM pro Hektar eine Zugehörigkeit zu einer politischen Gemeinde und zu einer Postleitzahl (6-stellig) sowie verschiedene Grössen aus der Volkszählung 1990, insbesondere die Population, geliefert.

5.1.1 Resultate Auswertungen

Um die Unschärfe bei der Verwendung der Korrespondenztabelle quantitativ abzuschätzen, wurde die Schichtung der SAKE nach Gemeindenummern und nach aus Gemeindenummern abgeleiteten Postleitzahlen untersucht. Die untenstehende Tabelle geht davon aus, dass in Kolonne deck.strat.gnr die korrekte Grösse der Population dargestellt ist. Die Kolonne deck.strat.ok gibt die Grösse der Population an, die auch mit der Schichtung nach PLZ korrekt klassiert ist. Der prozentuale Verlust ist in Kolonne deck.loss. Die Schicht erhält dafür einen Teil der Population, der nach der Gemeinde-Schichtung eigentlich nicht dazu gehört. Kolonne deck.strat.plz6 gibt die nach PLZ in die Schicht klassierte Population an und deck.gain zeigt den prozentualen Ueberhang der "falsch" klassierten Population verglichen mit der korrekt klassierten Population an. Die Schicht 0 umfasst

Gemeinden, die keiner Schicht zugeordnet werden konnten. Die Gewinne und Verluste sind fast immer kleiner als 5%. Der grösste Verlust (8.3%) und Gewinn (10.5%) findet sich in Schicht 1901, die sehr klein ist.

Vergleich Ueberdeckung der Schichtung nach GNR und PLZ6.

schicht	deck.strat.gnr	deck.strat.ok	deck.strat.plz6	deck.loss	deck.gain
0	42009	39043	39315	-7.1	0.6
101	448100	446255	448206	-0.4	0.4
102	617135	613210	616907	-0.6	0.6
103	48481	48262	49631	-0.5	2.8
104	49675	48033	48481	-3.3	0.9
201	223601	220272	224931	-1.5	2.1
202	313294	305783	312537	-2.4	2.2
203	171672	169253	173965	-1.4	2.7
204	223501	217029	219547	-2.9	1.1
300	321836	321151	321508	-0.2	0.1
400	33932	33932	34030	0	0.3
500	107242	106730	107121	-0.5	0.4
600	59202	59199	59280	0	0.1
800	38368	38286	38325	-0.2	0.1
900	83178	83108	83559	-0.1	0.5
1000	209764	209463	211188	-0.1	0.8
1100	202300	201115	203372	-0.6	1.1
1400	71433	71028	72193	-0.6	1.6
1500	65316	64070	64230	-1.9	0.2
1701	74045	73302	73452	-1	0.2
1702	109602	103645	107722	-5.4	3.7
1703	175686	172493	179989	-1.8	4.3
1704	63090	61295	62905	-2.8	2.6
1800	171322	171260	171265	0	0
1901	32197	29528	32899	-8.3	10.5
1902	255468	245099	252031	-4.1	2.7
1903	71950	68725	74458	-4.5	8
1904	87720	85297	87294	-2.8	2.3
2000	192782	191231	192702	-0.8	0.8
2100	139000	139000	139117	0	0.1
2201	177185	174630	187104	-1.4	7
2202	255120	241945	245457	-5.2	1.4
2203	68644	68014	69220	-0.9	1.8
2204	80226	78725	79440	-1.9	0.9
2300	239154	239064	239097	0	0
2400	161394	161216	161517	-0.1	0.2
2501	168826	167821	174287	-0.6	3.8
2502	203330	196860	198164	-3.2	0.6
2600	65455	65376	65458	-0.1	0.1
2701	175925	173145	176289	-1.6	1.8
2702	260573	256782	260075	-1.5	1.3
2703	36351	35677	36821	-1.9	3.1
2704	55194	54390	55189	-1.5	1.4
CH	6650278	6549742	6650278	-1.5	1.5

5.2 Ausgeschlossene Bereiche

Die Lieferung umfasst das gesamte Gebiet der Schweiz plus Campione (I) und Büsingen (D) und Liechtenstein. Campione und Büsingen werden aus dem SRH eliminiert. Hingegen wird das Fürstentum Liechtenstein im SRH belassen. Es kann natürlich bei der Ziehung ausgeschlossen werden.

5.3 Doppeladressen

In der alten Applikation TERCO konnten eine Telefon-Nummer, die an verschiedenen Orten eingetragen war, mittels eines Namen - Adress - Abgleichs eruiert werden und alle ausser dem ersten Eintrag wurden gelöscht.

5.4 Einheiten im ETV und in der Adresslieferung der Swisscom

Pro Standort liefert die Swisscom eine Nummer des festen Netzes. Wenn also an einem Standort mehrere Nummern eines festen Netzes eingetragen sind, wählt die Swisscom eine davon aus. Dies gilt insbesondere für ISDN-Anschlüsse mit verschiedenen Nummern. Falls unklar ist, ob eine Nummer privat oder geschäftlich genutzt wird, oder wenn tatsächlich eine Nummer sowohl als privat wie geschäftlich benutzt angegeben worden ist, dann wird die Nummer zweimal geliefert, einmal mit dem Kundentyp RES und einmal mit dem Kundentyp BUS. Hingegen kann der Abgleich mit den NATEL-Nummern nicht durchgeführt werden. Bei NATEL-Nummern gilt die Adresse für die Rechnungsstellung als Adresse. Es gilt immer noch, dass für Privat-Haushalte, bei denen meherer unabhängige Telefon-Anschlüsse existieren (Konkubinat, Wohngemeinschaft), alle Telefonnummern im SRH aufgeführt sind. Damit wird also immer noch die Einschlusswahrscheinlichkeit eines Haushalts ungefähr proportional zur Anzahl Anschlüsse wachsen.

In einem ersten Verarbeitungsschritt werden die Einträge nach Telefonnummern sortiert. Telefonnummern, die nur einmal im Register auftreten und nur geschäftliche genutzt werden (BUS), werden eliminiert. Telefonnummern, die einmal als privat (RES) und einmal als geschäftlich (BUS) eingetragen sind, erhalten den Code 2 (gemischte Telefonnumer) und werden nur noch einmal im SRH geführt. Rein private Telefonnummern werden mit Code 1 ins SRH aufgenommen.

Die Lieferung der Swisscom enthält u.a. auch sehr viele ungültige Rufnummern (z.B. die Notrufnummern 118 je FKZ etc.) aber auch viele ungültige NATEL-Nummern (Lieferung vom 15. Dez. 1998 enthielt 85'494 ungültige Telefonnummer, der Grossteil davon alte NATEL-Nummern mit Vorwahl 077). Diese Nummern werden nicht ins SRH geladen.

5.4.1 NATEL

Die NATEL-Nummern des ETV stammen aus einem eigenen Register. Es scheint aber nicht so zuverlässig zu sein wie für die festen Anschlüsse. Mindestens eine eingetragene NATEL-Easy Nummer wurde im SRH nicht gefunden. Die NATEL-Nummern, die ein Abonnement bei der Swisscom haben, sind im SRH enthalten, es besteht aber ein gewisse zeitliche Verzögerung.

5.4.2 Nicht eingetragene Nummern, Unterscheidung von privaten und geschäftlichen Nummern

Die Mietpreiserhebung arbeitet mit einer Bernoulli-Stichprobe mit gleicher Auswahlwahrscheinlichkeit für alle Telefonnummern des SRH. Eine Auswertung der Stichprobe für die Mietpreiserhebung aus dem SRH vom 12. Dezember 1998 ergabe folgende Resultate:

Teile der Stichprobe	Absolut	in Prozent
Anzahl gezogene Telefonnummern	2030	
Anzahl entsprechende Einträge im ETV	2806	
Anzahl Telefonnummern ohne PLZ	127	6.3% von 2030 (SD 0.5%)
Davon mit Vorwahl 079 (NATEL)	37	29% von 127
Anzahl Telefonnummern ohne Vorname (mit PLZ)	100	4.9% von 2030 (SD 0.5%)

Die Anzahl gefundene Einträge pro Telefonnummer verteilt sich wie folgt:

Anz. Einträge	0	1	2	3	4	5	6	7	8	9	10
Anz. TN	128	1481	250	93	24	30	10	7	3	1	3
In Prozent	6.3	73.0	12.3	4.6	1.2	1.5	0.5	0.3	0.1	0.0	0.1

Bei 1609 Einträgen, bzw. 79.3% der gezogenen Telefonnummern stellt sich kein Problem mit mehreren Einträgen. Es hat aber bis zu 10 Einträge pro Telefonnummer.

Eine Auswertung durch die Sektion Preis (Bearbeitung jeder Telefonnummer einzeln) ergibt sehr ähnlich Resultate. Dabei wurde jeweils nur die erste gefundene Telefonnummer berücksichtigt. Es sind aber wesentlich mehr Geschäftsnummern identifiziert worden als durch das einfache Kriterium "kein Vorname".

Teile der Stichprobe	Absolut	in Prozent
Gezogene Telefonnummern	2030	100
NATEL	117	5.8
Liechtensteiner PLZ	8	0.4
Geschäftsnummern	135	6.7
Ohne Adresse ab ETV-Batchman	128	6.3
davon NATEL ohne Adressen	37	28.9

Es muss also im Moment mit ca. 6.3% nicht eingetragenen Nummern gerechnet werden. Ein grosser Teil davon (29%) sind NATEL-Nummern. Die Ausschöpfung dieser nicht eingetragenen Nummern ist gering: Von den 128 Nummern konnten nur 12 ausgewertet werden.

Bei der SAKE wurden für den Basis-Teil 10187 Telefonnummern gezogen. 368 davon haben keine Adresse mit ETV-Batchman bekommen, wovon 298 in der Erhebung verwendet wurden (70 waren Reserve). Es resultierten 60 realisiert Interview, also 20% Ausschöpfung.

5.4.3 Mehrfacheinträge

Ein fiktives Beispiel soll die Problematik erklären:

Person	Haushalt	Geschäft	Festnetz privat	Mobilnetz	Geschäfts-TN (Festnetz) im SRH mit Code 2 (RES/BUS)
Hans Meier	M	G	xxxx1	xxxx3 (NATEL swiss)	xxxx5
Rita Müller	M		xxxx1		
Peter Schmid	S	G	xxxx2	xxxx4 (NATEL-easy)	xxxx5

Sucht man mit den Nummern der linken Spalte im ETV, dann findet man die Einträge der Nummern in den entsprechenden Zeilen

Nummer ETV: xxxx+	1	2	3	4	5
xxxx1	x				
xxxx2		x[a]			
xxxx3	x		x		
xxxx4					
xxxx5	x[b]				x[b]

a: Die Nummer könnte z.B. auch nicht eingetragen sein.
b: Die Geschäftsnummer kann u.U. eliminiert werden.

Die Abfrage ans ETV mit der Geschäfts-Nummer liefert u.U. viel mehr Einträge. Im obigen Beispiel würden typischerweise je Teilhaber ein Eintrag unter ihrem Namen und unter einer Rubrik gefunden. Ausserdem wird unter Umständen noch ein Eintrag aufgeführt, der leer ist (z.B. ein Postfach).
Die NATEL-easy Nummer ist nicht im SRH enthalten. Alle anderen Nummern sind aufgeführt. Die Geschäfts-Nummer ist sowohl privat als auch geschäftlich aufgeführt, erhält also im SRH den Code 2 (gemischte Adressen).
Wenn man mit der Geschäftsnummer ins ETV geht, findet man auch den Haushalt M, da dieser neben seiner privaten Telefonnummer auch die Geschäftstelefonnummer eintragen liess (oder unter der Geschäftsnummer auch seine private Telefonnummer). Es ist aber vermutlich besser, Geschäftsnummern zu eliminieren, auch wenn man durch sie einen privaten Haushalt erreichen könnte. Damit wird vermieden, dass z.B. Rechtsanwälte oder Immobilienhändler mit sehr vielen Büros zu oft gezogen werden. Allerdings dürfen echte gemischte Telefonnummern nicht gelöscht werden.
Haushalt M kann also via 3 Telefonnummern erreicht werden, bzw. via 2 wenn die Geschäftsnummer eliminiert wird, hat also eine Wahrscheinlichkeit, die ca. 3-fach bzw. 2-fach der normalen ist. Haushalt B kann nur via 1 Nummer erreicht werden. Die Komplexität der Situation ist den Betroffenen im Normalfall nicht klar, bzw. sie erinnern sich nicht an ihre Einträge. Wie in der Befragung Klarheit über die Einschlusswahrscheinlichkeit gewonnen werden soll, muss noch weiter untersucht werden.

Ariane Bender (Sektion EWL) hat die Situation bei Mehrfacheinträgen für die SAKE näher untersucht. Ein grosser Teil dieser Fälle kann relativ einfach bereinigt werden.

6 Informatik-Probleme

6.1 Umstellung bisheriger Informatik-Applikationen

6.1.1 TERCO-BFS

Bis 2. Quartal 1998 wurden in der Applikation TERCO folgende Aufgaben durchgeführt:

- Einlesen der ETV-Adressen aus ASCII-Format inklusive Zeichensatz-Aenderung
- Elimination von doppelten Einträgen
- Elimination von Geschäftseinträgen und Kennzeichnung von gemischten Einträgen
- Codierung der amtlichen Gemeindenummer aufgrund der Adresse mit Hilfe der Adressdatenbank ADVER
- Uebernahme in Datenbank
- Ziehung von geschichteten und geklumpten Stichproben und Reservestichproben mit spezifizierter Grösse
- Eintrag der gezogenen Telefonnummern in ein Register, damit Mehrfachbefragungen vermieden werden.

6.1.1.1 Umstellung

Das Programm, um die Daten einzulesen wurde auf die neuen Datenformate umgestellt. An diversen Stellen mussten Datenformate geändert werden.
Die Codierung der Gemeindenummer ist nicht mehr möglich. Der entsprechende Teil der Applikation wurde desaktiviert.
Ueberall wurde die Rolle der Gemeindenummer durch die 6-stellige Postleitzahl übernommen.
Bei der Zuordnung der Postleitzahlen zu Schichten, wurde gleichzeitig eine Zuordnung von Gemeindenummern vorgesehen. Die Gemeindenummer muss mit der Datei, welche die Schichtung definiert, mitgeliefert werden.
Neu wird eine sequentielle Kopie der Datenbank DB-IND mit den Indices der Postleitzahlen erstellt. Damit kann der aktuelle Stand der Postleitzahlen überprüft werden und gleichzeitig können die Populationsgrössen der Schichten berechnet werden.

6.1.2 MPAN

Die Applikation MPAN für die Mietpreiserhebung greift in die Applikation TERCO hinein. Die Mietpreiserhebung ist ein rotierendes Panel, das auf einer Bernoulli-Stichprobe basiert. Die Bernoulli-Indikatoren werden im SRH gespeichert. Damit das Panel geführt werden kann, müssen die Bernoulli-Indikatoren jeweils von der alten auf die neue Ausgabe des SRH übertragen werden. Dies geschieht im Programm MPPDIFCR, welches organisatorisch der Applikation TERCO zugeordnet ist. Ausserdem wird das SRH nach der Mietpreiserhebung jeweils aufdatiert mit Aenderungen, die während der Erhebung festgestellt wurden (Mieterwechsel, Verweigerer).

6.1.2.1 Umstellung

Während im alten Programm MPPDIFCR die Einträge unter einer Telefonnummer auf Name, Adresse und Postleitzahl überprüft wurden (mit entsprechenden

Fehlermeldungen), wird im neuen Programm lediglich die Postleitzahl überprüft. Damit können wenigstens Mieterwechsel mit einem Umzug in einen anderen Zustellbezirk entdeckt werden.
Es wurde darauf verzichtet, weitergehende Informationen aus der Mietpreiserhebung beim Abgleich der SRH-Versionen zu verwenden.

6.2 ETV-Batchman

Die Applikation ETV-Batchman der Firma Nova Esprit dient zum Herunterladen von Namen und Adressen zu einer Telefonnummer. Der Vorteil von ETV-Batchman ist, dass eine Datei als Input und eine Datei als Output angegeben werden kann, während der Abgleich dann selbständig geschieht.

Die Applikation läuft unter Windows-NT und ist in einer Client-Server Architektur implementiert. Der ETV-Batchman Server greift via das X.28 Netzwerk auf den ETV-Server zu. Es hätten 5 Clients im BFS installiert werden solen. Nur ein Client ist effektiv operationell. Intern hält ETV-Batchman die Daten auf einer Microsoft Access Datenbank.

Das Input-Format für ETV-Batchman kann frei gewählt werden. Zwei Standards sollen unterhalten werden. Der eine für Stichprobenziehungen für Haushalt-Erhebungen, der andere für die Mietpreiserhebung. Der Input muss vom Host auf einen PC mit einem ETV-Klienten transferiert werden (z.B. mit WS-ftp).

Der Output kann relativ frei gestaltet werden. Input-Variablen können an den Output weitergegeben werden, ohne dass sie in die Suche auf dem ETV involviert wurden. Dies wird insbesondere für die eindeutige Identifikation der Einträge benützt. Es können ja zu einer Input-Telefonnummer auch andere Telefonnummern gefunden werden. Beim Output kann auch definiert werden, wieviele Einträge pro Input-Telefonnummer ausgegeben werden.

6.3 SAS-Macros

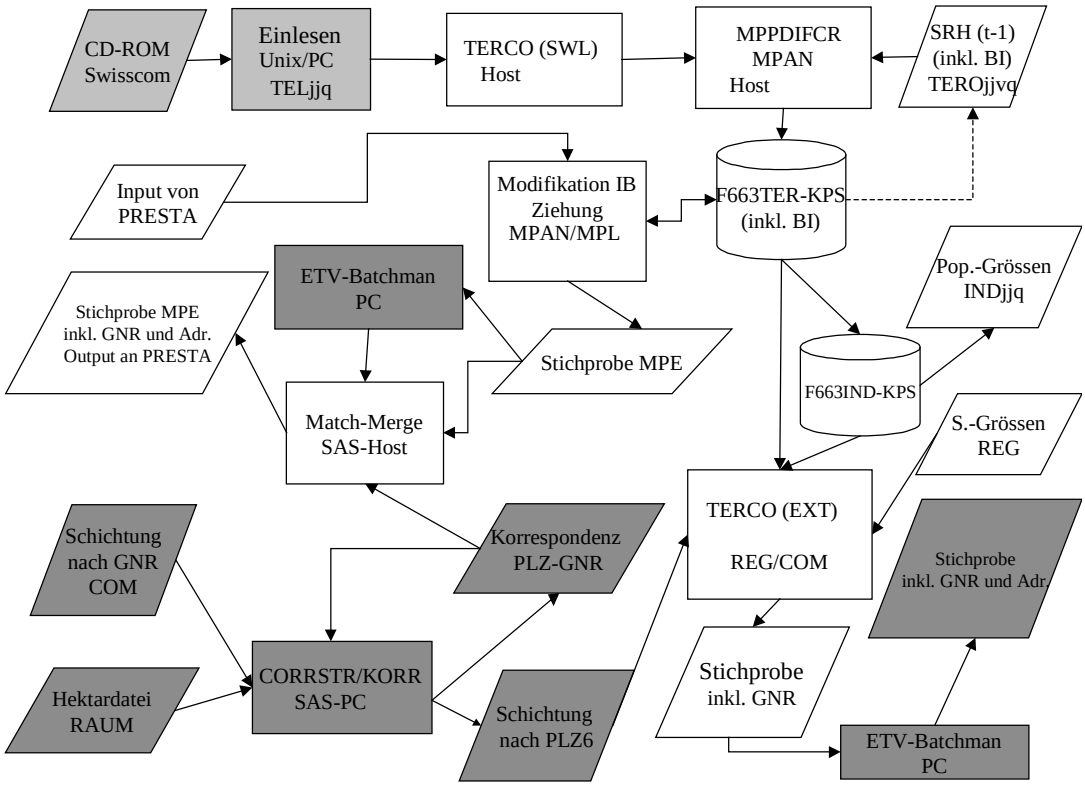
Um die Korrespondenztabelle zwischen Gemeindenummern und Postleitzahlen zu erstellen wurde durch M. Eichenberger (BFI) aufgrund der Spezifikationen von METH ein SAS-Macros KORR erstellt (siehe Anhang). Das SAS-Macro hilft gleichzeitig einen Ueberblick über die Korrespondenzen, insbesondere über Postleitzahlen mit tiefen Uebereinstimmungen zu bekommen. Die Korrespondenztabelle wird auf einer Diskette als SAS-Dataset abgespeichert. Ein weiteres SAS-Macro CORRSTR hilft, eine Schichtung nach Gemeindenummern in eine Schichtung nach Postleitzahlen zu verwandeln und gleichzeitig den Postleitzahlen eine Gemeindenummer zuzuordnen.

Die entstehende Datei mit der Schichtung nach Postleitzahlen muss mit dem neuesten SRH abgeglichen werden. Dazu werden die Postleitzahlen aus der sequentiellen Kopie von DB-IND mit der Schichtungs-Datei verglichen (siehe Programm GNRMERGE im Anhang). Eventuelle neue PLZ müssen in der Schichtungsdatei ergänzt werden. Veränderungen in den Postleitzahlen nach dem Stichdatum für die Korrespondenztabelle sind natürlich nicht in der Korrespondenztabelle enthalten. Unter Umständen entsprich natürlich auch die Schichtung nach Gemeinden nicht dem

Stand der Korrespondenztabelle. Für die entsprechenden Umstellungen bestehen im Moment im BFS keine Korrespondenztabellen.

7 Ablaufschema

Im nachfolgenden Schema wird ein grober Ablauf für eine Aufbereitung des SRH bis zu einer Stichprobe für die Mietpreiserhebung bzw. eine Standard-Haushalterhebung. Es kommen 3 Plattformen zum Zug: Unix oder PC zum Einlesen der CD-ROM von Swisscom und zur Uebertragung der Datei auf den Host. Der Hauptteil der Applikation läuft auf dem Host. PC unter Windows NT für die Applikation ETV-Batchman und für die SAS-Macros zur Aufbereitung und Verwendung der Korrespondenztabelle zwischen PLZ und Gemeindenummern.



7.1 Uebertragung auf den Host

Auf dem Host muss eine sequentielle Datei D663.A663SWC.TEL&JJQ alloziert werden (Management-Class MC10J, Space-Units BLOCK, Primary Quantity 7990, Secondary Quantity 382, Directory Blocks 0, Record Format FB, Record length 50, Block size 27950). Der Parameter JJQ steht für Jahr (zweistellig) und Quartal. Eine Kopie der CD-ROM wird mit ftp auf den Host übertragen (Kontroll-Taste, um Namen zu ändern, Uebertragungsformat ASCII). Grösse ist ca. 180MB. Bei der Lieferung vom 15. März 1999 fehlten die Carriage Return/Linefeed Zeichen am Ende der Records. Die Uebertragung musste mit einem speziellen SAS-Programm gemacht werden, welches in verdankenswerter Weise von M. Eichenberger bereitgestellt worden ist.

7.2 Verarbeitung auf dem Host

Ein Beispiel eines Verarbeitungs-Protokolls befindet sich im Anhang.

7.2.1 Eingangskontrolle

Mit Programm SASREAD wird die Anzahl Records und Grundstrukturen der Eingangsdatei überprüft. Vergleich mit den Resultaten des Vorquartals. Anzahl Records sollte mit den Angaben auf dem Lieferschein und mit den Angaben im Footer-Record der Datei überinstimmen.

7.2.2 Arbeitspläne

Plan-Name	Beschreibung
P663DWL	TERCO-DB alter Stand auf seq. Datei laden und sortieren (für Applikation MPAN)
P663SWC	Aufbereitung Swisscom-Datei: Eliminierung doppelte TN und nicht erwünschte Gebiete, Zuweisung Telefoncode 1 (privat) oder 2 (gemischt privat/geschäftlich), Eliminierung rein geschäftliche TN.
P663DIF	Vergleich neuer Stand mit altem Stand (für MPAN), Fehlerdateien, Laden der Datenbank des SRH (F663TER-KPS). Erstellen Index und sequentielle Kopie davon.
P663MPL	Laden der Fehlerdatei für MPAN
P663EXT	Ziehungsprogramm: Aufbereitung der Spezifikation Stichprobenplan (Schichtung, Stichprobengrössen) und Ziehung mit Historisierung.
P663LOE	Löschen temporärer Dateien

7.2.3 Kontrollen

Die Datei D663.A663SWC.ADR&JJQ sollte auf ihre Struktur untersucht werden (Z.B. mit Programm READSRH, siehe Anhang).

Die Index-Datei D663.A663SWC.IND&JJQ enthält pro 6-stellige Postleitzahl und Telefoncode die Anzahl Einträge im SRH. Die Populationsgrössen in den Postleitzahlen können mit dem Programm INDSUM aus der Index-Datei berechnet werden. Um die Populationsgrössen in den Schichten zu berechnen muss die Schichtungsdatei D663.A663SWC.COM&SECT (Input für P663EXT) hinzugemergt werden. Zusammen mit der Datei der Stichprobengrössen D663.A663SWC.COM&TEST lassen sich dann die Einschlusswahrscheinlichkeiten berechnen.

7.2.4 Sperrung von gezogenen Telefonnummern (Historisierung)

Das Ziehungssystem speichert die Telefonnummern der gezogenen Stichproben und sperrt sie für nächste Ziehungen. Damit wird eine negative Koordination der Stichproben erreicht. Mit anderen Worten wird die Befragungslast gleichmässiger auf die Bevölkerung verteilt. Damit nicht zuviele Telefonnummern gesperrt werden, sollten in der betreffenden Datenbank TER-HIS die Stichproben, die älter als zwei Jahre sind, gelöscht werden. Dies geschieht in der Online-Applikation U663ONL in der Umgebung COMPLETE-PRODUKTION. Eine noch zu erstellende Applikation soll die Ueberwachung und Aufdatierung von TER-HIS vereinfachen.

8 Vorgehen beim Ziehen einer Stichprobe

Nr.	Schritt	Ergebnis
1	Stichprobenplan definieren mit Schichten nach Gemeindenummer. Die Nomenklatur sollte wenn möglich dem Stand der in Schritt 2 benützten Korrespondenztabelle entsprechen. Falls Liechtenstein (oder ein anderer Bereich) ausgenommen werden soll, dürfen die entsprechenden PLZ6S bzw. GNR nicht auftreten.	1. Datei COM mit Variablen Gemeindenummer und Schichtnummer. 2. Datei REG mit Variablen Schichtnummer, Telefoncode von, Telefoncode bis, Stichprobe Nr., Stichprobengrösse (bis zu 5 Stichproben).
2	Mit Hilfe der Korrespondenztabelle und dem Macro corrstr eine Schichtung nach PLZ erzeugen. Formate siehe Programm GNRMERGE.	Datei COM mit Variablen Postleitzahl (6-stellig), Schichtnummer, Gemeindenummer (fakultativ).
3	Merge der Datei COM mit der Datei DBIND (Index des SRH). Kontrolle, welche PLZ6 nicht mit der Datenbank übereinstimmen, eventuell manuelle Ergänzung der Datei COM.	Datei COM, deren PLZ6 mit den PLZ6 aus DBIND übereinstimmen.
4	Kopie der Datei DBIND archivieren, um später mit Hilfe von COM die Populationsgrössen und mit Hilfe von REG die Einschlusswahrscheinlichkeiten bestimmen zu können.	Kopie DBIND.
5	Bereitstellung der Dateien COM und REG auf dem Host. Der letzte Teil der Datei-Namen (***APPL***) kann individuell gewählt werden (4 Stellen alphanumerisch)	D663.A663SWC.REG***APPL*** D663.A663SWC.COM***APPL***
6	Die Applikation mit Namen ***APPL*** muss in der Tabelle der Applikationen angemeldet sein. Mit U663ONL, Gestion des Tables, Applications. Als Kommentar sollte in der Tabelle das Ziehungsdatum angegeben werden. Damit wird die spätere Löschung vereinfacht.	Bestätigung in der Tabelle
7	Der Produktionsauftrag für Plan P663EXT an AVOR2/BFI faxen (Autorisiert: P. Eichenberger, B. Hulliger und M. Kolly für SAKE). Parameter ist der Applikationsname.	Beispiel des Outputs in TSO Beta im Anhang.
8	Kontrolle des Ergebnisses und Uebertragung in ASCII-Format mit Programm OUTREAD.	ASCII-Datei und Tabellen. Die Werbesperre sollte immer mitgliefert werden.
9	Uebertragen der Datei auf PC mit ftp.	Datei der Stichprobe (ohne Namen und Adressen) auf PC.
10	Einlesen der Datei in ETV-Batchman, Herunterladen Namen und Adressen, Ausgabe auf Datei.	Stichprobe mit Namen und Adressen (ev. mehrere).

Das Vorgehen im Schritt 10, insbesondere die genauen Input und Output-Spezifikationen für ETV-Batchman, wird in einem späteren Bericht beschrieben. Die automatische Bereinigung der Mehrfachadressen mit Hilfe eines Programms muss noch weiter untersucht werden.

9 Offene Fragen

9.1 Bei Interviews zu stellende Fragen

In Anbetracht der Komplexität der Situation ist vermutlich die beste Möglichkeit, die Namen der Mitglieder des Haushaltes zu erfragen. Damit könnten im ETV die entsprechenden Nummern erfragt werden und anschliessend kontrolliert werden, welche dieser Nummern tatsächlich im SRH enthalten sind. Die SAKE hat eine Frage nach der Anzahl fester Anschlüsse und der Anzahl Mobiltelefone (ohne NATEL-easy) gestellt. Es müssten aber eigentlich nicht nur die Namen, sondern auch die Telefonnummern der Haushaltsmitglieder erfragt werden, um auch nicht eingetragene Nummern im SRH wieder finden zu können.

Da Mobiltelefone eher als persönliche Telefone betrachtet werden und nicht als Telefone für den ganzen Haushalt sollte das Verhältnis von Stichprobeneinheit (Telefonnummern) und Untersuchungseinheiten (Haushalt, Personen) näher untersucht werden.

9.2 Gewichtungsprobleme

Mit der Zunahme von Mobiltelefonen und später eventuell von Festnetz-Anschlüssen muss dem Problem der Einschlusswahrscheinlichkeiten mehr Aufmerksamkeit gewidmet werden. Theoretisch ist die Einschlusswahrscheinlichkeit ungefähr proportional zur Anzahl Anschlüsse pro Haushalt. Die Kontaktierbarkeit nimmt aber nur unterproportional zu. Eine Untersuchung über die Erreichbarkeit von Haushalten in Abhängigkeit der Anzahl Anschlüsse wäre notwendig.

10 Anhang

10.1 Frühere Datenlieferung

Die Datenlieferung der Telecom PTT umfasste am 13.6.1997 folgende Variablen: Recordtyp (Grund- oder Folgeeintrag), Rubriktitel und/oder Organisationsstrukturtitel, Rubriktyp, Text nach Name, Männernahme, Vorname, Text nach Vorname, Frauenname, Text nach Frauenname, Beruf, Text nach Beruf, Hausbezeichnung, Strasse, Hausnummer, Stockwerk, Wohnung, Text nach Strasse, Zusatzort, Text nach Zusatzort, PLZ 4-stellig, PLZ 6-stellig, Ort, Postfachadresse, Gültig ab, Gültig bis, Text vor Rufnummer, Merkmal (Sperrmerkmal: *=wünscht keine Werbung), Fernkennzahl, Rufnummer, Umnummerierungs-FKZ, Umnummerierungs-Rufnummer, Umnummerierungs-Text, Text nach Rufnummer.

10.2 Datei-Organisation und Jobs der Applikation TERCO

Die verschiednen Programme und Beispiele für die JCL befinden sich in t7bcj.tera.cntl. Der Ablauf für die Aufbereitung der Datenlieferung vom 12. Dez. 1998, d.h. 4. Quartal 1998 wird hier beschrieben. Die Arbeitspläne wurden bereits weiter oben erklärt.

Plan	Step	JCL: J663xxx	Natural-Programm	Input	Output
DWL	1	UNL	ECPBTUNL	F663TER-KPS	D196.A196MPAN.TERO983
DWL	2	TR3		D196.A196MPAN.TERO983	D663.A663WC.TEMP01
SWC	3	TEA2	ECPBTTEA	D663.A663TER.STEUER.DATAF D663.A663SWC.TEL984	D663.A663WC.TEMP02
SWC	4	TR1		D663.A663WC.TEMP02	D663.A663WC.TEMP03
SWC	5	TEL	ECPBTTEL	D663.A663WC.TEMP03	D663.A663WC.TEMP04
DIF	6	DIF	MPPDIFCR	D663.A663WC.TEMP04 (neu) D663.A663WC.TEMP01 (alt) F196MPAN	D663.A663WC.TEMP05 D196.A196MPAN.ERR
DIF	7	TR42		D663.A663WC.TEMP05	D663.A663SWC.ADR984
DIF	8	LOD		D663.A663SWC.ADR984	DB-TEL (=F663TER-KPS)
DIF	9	IDX	ECPBTIDX	F663TER-KPS	D663.A663SWC.IND984 (=F663IND-KPS)
EXT	10	REG	ECPBSMEXT	D663.A663SWC.REGTEST D663.A663SWC.COMTEST	D663.A663WC.TEMP06
EXT	11			D663.A663WC.TEMP06	D663.A663WC.SELTEST
EXT	12	EXT	ECPBTEXT	D663.A663WC.SELTEST F663TER-KPS F663IND-KPS F663HIST-KPS F663ERR-KPS	D663.A663SWC.ECHTEST D663.A663SWC.ECHCOM

Eine vollständige Beschreibung der Abläufe und Pläne inklusive Diagramme befindet sich bei METH.

10.3 GNRMERGE

```
//T7BCJS JOB (317,04,D397000),'BCJ-HULL',
//         MSGLEVEL=(1,1),CLASS=J,
//         NOTIFY=T7BCJ,MSGCLASS=T
//SAS    EXEC SAS
  /* GEMEINDENUMMERN ZU PLZ6 */;
FILENAME OUT  'T7BCJ.TERCO.REG.DATAF' DISP=(NEW,CATLG);
```

```
FILENAME INP 'T7BCJ.TERCO.DBIND' DISP=OLD;
FILENAME GNR 'T7BCJ.TERCO.PLZ6GNR.DATAF' DISP=OLD;
OPTIONS LINESIZE=80;
DATA P;
 INFILE INP;
 INPUT ID 8. PLZ6S 8. NTEL 10.;
PROC SORT DATA=P;
 BY PLZ6S;
DATA G;
 INFILE GNR;
 INPUT PLZ6S 8. GNR 8.;
PROC SORT DATA=G;
 BY PLZ6S;
DATA REG;
 MERGE P (DROP=NTEL) G;
 BY PLZ6S;
DATA REG0;
 SET REG;
 FILE OUT;
 IF GNR=. THEN GNR=0;
 PUT _N_ 8. PLZ6S 8. GNR 8.;
PROC FREQ;
 TABLE GNR;
RUN;
```

10.4 SAS-Macro Aufbereitung Korrespondenztabelle

```
%macro korr(hekdata=,c1var=,c2var=,sumvar=,out1=,out2=,pk=);
   /*****************************************************************************
      hekdata=  Hektardatei (Datei mit Einzelbeobachtungen bzw. Hektardaten)required
      c1var=    1. Klassierungsvariabe (z.Bsp plz,Gem_nr)                   required
      c2var=    2. Klassierungsvariabe (z.Bsp Gem_nr,plz)                   required

      sumvar=   Summierungsvariable                                         required

      out1=     Output Dataset auf Grund des MAX-Kriteriums                 required
      out2=     Output Dataset mit Verteilung der Summierungsvariablen
                ohne MAX-Kriterium   (Korrespondenztabelle)                 optional
      pk=       Listen der Fälle deren Anteil < pk, 0 < pk < 1, Default:0.9 optional
   ******************************************************************************/
   /* Markus Eichenberger, Juli 1998
      Beat Hulliger, März 1999 Reduktion auf Korrtab und Max, Kommentare  */
options pageno=1  linesize=76 pagesize=54 nodate;
%if &pk = %then %let pk=0.9;
%let cl1=class &c1var;
%let unique1=;
%if &sumvar ne %then %do;
 %let unique1=if first.&sumvar;
 %let ifdel=if &sumvar = 0 then delete;
%end;
%let bye=by &c1var &c2var;
%let cle=class &c1var &c2var;
%let ide=&c2var;
%let vare=var &sumvar;
%let sume=sum=&sumvar;
%let maxe=max=&sumvar;
%if &c1var= and &c2var=  %then  %do;
    %let cle=;
    %let bye=;
%end;
%if &sumvar=             %then  %do;
    %let sume=;
    %let maxe=;
    %let vare=;
%end;
data _hekt_;
  set &hekdata;
  if &c1var>0 and &c2var>0;
  *Faelle mit negativen plz oder gnr ausscheiden;
run;
*Kontingenztafel c1*c2;
proc sort data=_hekt_   out=_t0_;
     &bye ;
run;
proc summary nway data=_t0_  sum;
```

```
  &cle;
  &vare;
  output out=_t1_   &sume;
run;
%if &sumvar ne %then %do;
data _t1_;
 set _t1_;
 &ifdel;
run;

* Summierung fuer c1;
proc summary nway data=_t1_  sum;
  &cl1;
  &vare;
  output out=_t1t_   &sume;
run;

* Total fuer c1;
proc summary nway data=_t1_  sum;
  &vare;
  output out=_t1to_   &sume;
run;
title "Gesamtsumme der Variablen &sumvar" ;
proc print data=_t1to_  noobs uniform;
     var &sumvar;
run;

* Summen fuer c1 wieder auf Kontingenztafel aufblasen;
*data _t1t_;
*drop _type_  i;
*rename  &sumvar=_tot_;
* set _t1t_;
*  do i=1 to _freq_;
*    output;
*  end;
*run;
data _t2t_;
drop _freq_ _type_ ;
merge _t1t_(in=in1 drop=_freq_ rename=(&sumvar=_tot_))  _t1_(in=in2);
by &c1var;
if in1 and in2;
if _tot_ > 0  then  p=&sumvar/_tot_;
label p="Proportion";
run;
proc sort data=_t2t_  out=_t2t_;
     by &c1var &sumvar;
run;
proc means data=_t2t_  nway  noprint;
     var p;
     class &c1var;
     output out=_tmp_  max=p;
run;
data _tmp_;
  set _tmp_;
 drop _type_  _freq_;
  if p < &pk  then output;
run;
data _tmp_;
 merge _tmp_(in=in1)  _t2t_(in=in2);
 by &c1var;
 if in1;
run;
title "Liste der &c1var deren maximaler Anteil < &pk ist";
proc print data=_tmp_  noobs label uniform;
     var  &ide  &sumvar p;
     sum p;
     by &c1var;
     sumby &c1var;
run;
 %if &out2 ne %then %do;
  data &out2;
   drop _tot_;
   set _t2t_;
  run;
 %end;
%end;
```

```
proc sort data=_t1_  out=_t1_;
     by &c1var &sumvar;
run;
%if &vare ne  %then %do;
proc means data=_t1_  nway  noprint;
     &vare;
     by  &c1var;
     output out=_t2_  &maxe;
run;
data  _t2_;
drop _type_ _freq_;
set _t2_;
run;
data &out1;
  drop _tot_;
  merge _t2t_(in=in1)  _t2_(in=in2);
  by &c1var &sumvar;
  if in2;
  &unique1;
run;
%end;
%else %do;
data &out1;
 set _t1_;
run;
%end;
title;
%mend;


libname  srh  '/home/bhu/srh';
/***********************************************/
data hektar;
 set srh.hektar;
if kt=11;
run;


%korr(hekdata=hektar,c1var=plz6,c2var=gnr,
         sumvar=p9btot,out1=co1,out2=co2,pk=0.95);
title "Kanton Solothurn mit MAX-Kriterium";
proc print data=co1  noobs;   run;
title "Kanton Solothurn ohne MAX-Kriterium";
proc print data=co2  noobs;   run;

/*%korr(hekdata=hektar,c1var=plz6,c2var=gnr,
         sumvar=p9btot,out1=co3,out2=co4,pk=0.95); */
```

10.5 SAS-Macro Umwandlung Schichtung nach GNR in Schichtung nach PLZ6

Das Macro corrstr bildetete die Ausgangslage für das Macro korr und ist abgesehen von den Kommentaren und einer Vereinfachung bei korr identisch.

```
%macro corrstr(corrdat=,c1var=,c2var=,strdata=,s1var=,strvar=,out=,pk=);
   /*************************************************************************
      corrdat=  Korrespondenzfile PLZ - Gemeinde_NR                    required
      c1var=    plz6 (Postleitzahl)                                    required
      c2var=    gnr  (Gemeindenummer)                                  required

      strdata=  Schichtdatei (z.Bsp gem_nr <--> Schicht-Nr)            optional
      s1var=    Gemeindenummer in Schichtfile                          optional
      strvar=   Schichtvariable                                        optional

      out=      Output Dataset auf Grund des MAX-Kriteriums            required
      pk=       Listen der Fälle deren Anteil < pk, 0 < pk < 1, Default:0.9 optional
   *************************************************************************/
/* Markus Eichenberger, Juli 1998 */
options pageno=1  linesize=76 pagesize=54 nodate;
%if &pk = %then %let pk=0.9;
%if &strdata ne %then %do;
%let ide=;
%let bye=by &c1var &strvar;
```

```
%let cle=class &c1var &strvar;
proc sort  data=&strdata  out=&strdata;
     by &s1var;
run;
data &strdata;
  set &strdata;
  rename &s1var=&c2var;
run;
%end;
%else %do;
%let bye=by &c1var &c2var;
%let cle=class &c1var &c2var;
%let ide=&c2var;
%end;

proc sort  data=&corrdat  out=&corrdat;
     by &c2var;
run;
%if &strdata ne %then %do;
data  _t_;
merge &corrdat(in=in1)   &strdata(in=in2);
if in1;
by &c2var;
run;
proc sort data=_t_   out=_t0_;
     &bye ;
run;
proc summary data=_t0_ nway sum;
     var p;
     &cle;
     output out=_t1_ sum=p;
run;
%end;

%else %do;
proc sort data=&corrdat  out=_t1_;
     &bye ;
run;
%end;

proc means data=_t1_  nway  noprint;
     var p;
     class &c1var;
     output out=_tmp1_  max=p;
run;
data _tmp_;
  set _tmp1_;
  if p < &pk  then output;
run;
data _tmp_;
 merge _tmp_(in=in1)  _t1_(in=in2);
 by &c1var;
 if in1;
run;
title "Liste der &c1var deren maximaler Anteil < &pk ist";
proc print data=_tmp_  noobs label uniform;
     var  &ide  &strvar  p;
     sum p;
     by &c1var;
     sumby &c1var;
run;
proc sort data=_tmp1_;
     by &c1var p;
run;
proc sort data=_t1_;
     by &c1var p;
run;
data &out;
  drop _freq_  _type_;
  merge _tmp1_(in=in1)  _t1_(in=in2);
   by &c1var p;
   if in1;
run;
title;
%mend;
```

```
libname  indat  'h:\daten\mysas\korrtab';
filename out 'h:\daten\mysas\korrtab\plz6s2gnr.dat';
data schicht;
infile 'h:\daten\mysas\korrtab\gschicht.dat';
input  st  gnr;
run;

%corrstr(corrdat=indat.corrdat,c1var=plz6,c2var=gnr,
         strdata=schicht,s1var=gnr,strvar=st,out=co1,pk=0.80);

%corrstr(corrdat=indat.corrdat,c1var=plz6,c2var=gnr,
         strdata=,s1var=,strvar=,out=co2,pk=0.80);


data;
  set co2;
  file out;
  put plz6 8. gnr 8.;
run;
```

10.6 Programme S-Plus für Korrespondenz GNR-PLZ6

Vier Programme wurden als Prototypen für die SAS-Macros entwickelt:

f.corrplz	Zeile der Korrespondenztabelle
f.corrtab	Korrespondenzmatrix
f.gnr2plz	Umwandlung der Schichtung nach gnr in Schichtung nach plz6
f.deck	Berechnung der Ueberdeckung der Schichten nach f.grn2plz

Die Programme sind bei B. Hulliger erhältlich.

10.7 Verarbeitungsmeldung zu ECPBTTEA (TEL991) bis MPERRLD

```
...
Record mit ungültiger Rufnummer, Rec-Nr.:    3599183
Record mit ungültiger Rufnummer, Rec-Nr.:    3614971
Record mit ungültiger Rufnummer, Rec-Nr.:    3635537
Record mit ungültiger Rufnummer, Rec-Nr.:    3639377
Record mit ungültiger Rufnummer, Rec-Nr.:    3750174
Record mit ungültiger Rufnummer, Rec-Nr.:    3769597
Record mit ungültiger Rufnummer, Rec-Nr.:    3899708
Record mit ungültiger Rufnummer, Rec-Nr.:    3973247
Record mit ungültiger Rufnummer, Rec-Nr.:    3973248
Record mit ungültiger Rufnummer, Rec-Nr.:    3973250
Record mit ungültiger Rufnummer, Rec-Nr.:    3973254
Record mit ungültiger Rufnummer, Rec-Nr.:    3973255
Record mit ungültiger Rufnummer, Rec-Nr.:    3973256
Record mit ungültiger Rufnummer, Rec-Nr.:    4136796
Record mit ungültiger Rufnummer, Rec-Nr.:    4136797



Verarbeitungsmeldungen:
----------------------
Anzahl eingelesene Records.........................:    4332141
Anzahl akzeptierte Records.........................:    4332137
Anzahl Records mit ungültigen Rufnummern...........:      80370
Anzahl abgelehnte Records wegen Netz...............:          1
Anzahl abgelehnte Records wegen PLZ................:       1356
Anzahl auf seq. File geschriebene Records..........:    4250410


ENDE PGM ECPBTTEA: 04.05.99 UM: 12:28:46.9

Verarbeitungsmeldung zu ECPBTTEL (TEL991)

 START PGM ECPBTTEL: 04.05.99 UM: 12:46:55.3
```

```
 Verarbeitungsmeldungen:
 ----------------------
 Anzahl INPUT-Records :    4250410
 Anzahl OUTPUT-Records:    3673949


 ENDE PGM ECPBTTEL: 04.05.99 UM: 12:59:38.1


START OF PROGRAM MPPDIFCR
NEW BERNOULLI DE    1506000
NEW BERNOULLI A     1999000
RD-START    753674023
OLD BERNOULLI DE    1506000
OLD BERNOULLI A     1999000
FIN:NO ALEATOIRE   1388467788

Page      2

 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
              PROGRAMME : MPPDIFCR
              --------------------
   NOMBRE D'ENREGISTREMENTS DU TERCO PRECEDENT :   3616654

   NOMBRE DE NUMEROS TROUVES ANCIEN-NOUVEAU    :   3504743
   NOMBRE DE NUMEROS DISPARUS DANS LE NOUVEAU  :    111911
   NOMBRE D'ENREGISTREMENTS DIFFERENTS         :      8086
   NOMBRE D'ENREGISTREMENTS RECONDUITS         :   3496657
   NOMBRE D'ENREGISTREMENTS NOUVEAUX           :    169204
   NOMBRE D'ENREGISTREMENTS ERREURS            :       478

 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
   ECHANTILLON E2 A SUPPRIMER                  :         0

   ECHANTILLON E3 RECONDUITS                   :     12315
                  DISPARUS                     :       434
                  DIFFERENTS                   :        44
                  NOUVEAUX                     :       840

   ECHANTILLON E4 A AJOUTER                    :         0
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 END OF PROGRAM MPPDIFCR
FIN NORMALE DU PROGRAMME ECPBTIDX
    8377 ENREGISTREMENTS CHARGES DANS LE FICHIER
    8377 Anzahl auf seq. File geschriebene Records von IND-DB
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
              PROGRAMME : MPPERRLD
              --------------------

  NOMBRE D'ENREGISTREMENTS SUPPRIMES T - 1    :       931
  NOMBRE D'ENREGISTREMENTS CRREES    T        :       478
```

10.8 READSRH für die Kontrolle der seq. Datei D663.A663SWC.ADR984

```
//T7BCJS JOB (317,04,D397000),'BCJ-HULL',
//           MSGLEVEL=(1,1),CLASS=N,
//           NOTIFY=T7BCJ,MSGCLASS=T
//SAS    EXEC SAS,WORK='50000,10000'
  /* LIEST AUS DEM SEQUENTIELLEN, BEREINIGTEN TERCO
     WELCHES IN DIE ADABAS DB GESPEICHER WIRD */;
  /* KLEINE AUSWERTUNG VON HAEUFIGKEITEN  */;
FILENAME INP  'D663.A663SWC.ADR991' DISP=OLD;
OPTIONS LINESIZE=80;
DATA D1;
  INFILE INP;
  INPUT  BERN PD5. GNR PD5. KT PD2. TELCODE 1. SPR $1.
        @177 PLZ4 4.
        @206 IND1 3. TEL1 PD4.
        @220 DIENST $ 8. @283 PLZ6 6. WSP 8.;
PROC FREQ;
  TABLE KT;
  TABLE TELCODE;
  TABLE SPR;
  TABLE IND1;
```

```
  TABLE DIENST*TELCODE;
  TABLE WSP;
RUN;
```

10.9 INDSRH zum Lesen und Aggregieren des Datenbank-Index

```
//T7BCJS JOB (317,04,D397000),'BCJ-HULL',
//           MSGLEVEL=(1,1),CLASS=L,
//           NOTIFY=T7BCJ,MSGCLASS=T
//SAS    EXEC SAS
  /* POPULATIONSGROESSEN AUS INDEX-DATEI */;
FILENAME INP  'D663.A663IND.IND984' DISP=OLD;
OPTIONS LINESIZE=80;
DATA PLZ6;
  INFILE INP;
  INPUT PLZ6S 6. GENRTEL 1. ISN 10. NTEL 10.;
PROC MEANS SUM;
  VAR NTEL;
RUN;
```

10.10 Stand der Datenbank (Mit DBDC-Tools in Complete Test-Umgebung)

```
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
10:37:47        ***  A D A B A S  -  S u b s y s t e m  ***          11.05.1999
T7BCJ                       -  File-information  -                    ADMINFO1

Database-ID : 250    corresponding to : 2                   User   : T7317
File-no.    : 14                      14                   Date   : 02.02.99
File-name   : F663TERC-KPS        FDT : F663TER   in  WANLIB5
Status      : 3   file used normal
Project     : 663

Reservation :                  loaded : 06.05.99
Release     :                 deleted :
                         last update :
Top-ISN     :   3673947
Rec. loaded :   3673947       Sortseq :
Max. ISN    :   3876487

Allocation           AC        UI          NI          DS        DSST
- extents   :         1         1           1           1           1
- original  :                  70 B     20700 B     14000 B
- now alloc.:                  70 B     20700 B     14000 B
- now unused:                   5 B      1297 B      7284 B
```

10.11 Uebersichtstabellen über die zugelassenen Projekte

```
CVPAP1MF              Gestion des tables               mercredi     09/06/99
T7BCJ          Mise à jour de la table des applications             15:14:56.4
--------------------------------------------------------------------------------
* --> M=Modifier, D=Supprimer                  /  Pour  Créer, presser  <PF4>
--------------------------------------------------------------------------------
*    Code         Nom de l'enquête        Groupe  Du F Tir. Ménage  No.Aléat.
- -------- ---------------------------- -------- -- - ---- ------- ----------
. BAG      BAG - Tabakbefragung          D663     10 T    1    5500
. BUWA     Enquête sur le sort des batte D663     10 A    1    2000
. EVE1     EVE98                         D663     10 T    5   81048
. KPS      Supervision                   G397     10 T    1     100  797009023
. KPS1     Tirage normal                 G397     10 T   28  122375
. MZFF     MZ Familie - Frauen           G397     10 A    1   17091
. MZFM     MZ Familie - Hommes           G397     10 A    2   14636
. SAKE     Enquête suisse s. pop. active G624     10 T   32  119650
. SGB2     SGB 97 - 2. Welle             D663      1 T    1    8078
. SGB3     SGB 97 - 3. Welle             D663      1 T    1    8407
. SGB4     SGB 97 - 4. Welle             D663      1 T    1    9810
. TEST     test de la base de sondage    G397     10 T    4   12706
. WSL1     wsl 19.5.1999                 D663      1 T    1    5000
--------------------------------------------------------------------------------
```

10.12 Aufbereitung der Eingangsdateien für ECPBMEXT

```
//T7BCJS JOB (317,04,D397000),'BCJ-HULL',
//           MSGLEVEL=(1,1),CLASS=J,
//           NOTIFY=T7BCJ,MSGCLASS=T
//LOESCH  EXEC ERZTSM,CL='*'
//SYSIN    DD   *
  DELETE D663.A663SWC.REGWSL1
  DELETE D663.A663SWC.COMWSL1
//SAS    EXEC SAS
  /* ERZEUGT DIE EINGANGS-DATEIEN FUER DIE ZIEHUNG */;
  /* SCHICHTUNG WSL */;
FILENAME INP 'T7BCJ.TERCO.WSLREG2.DATAF' DISP=OLD;
FILENAME SIZES 'D663.A663SWC.REGWSL1' DISP=(NEW,CATLG)
   RECFM=FB LRECL=50;
FILENAME STRATA 'D663.A663SWC.COMWSL1' DISP=(NEW,CATLG)
   RECFM=FB LRECL=24;
OPTIONS LINESIZE=80;
DATA D1;
  INFILE INP;
  INPUT ID 8. PLZ6S 8. NTEL 10.;
PROC SORT DATA=D1; BY ID PLZ6S;
DATA;
  SET D1;
  GNR=9999;
  FILE STRATA;
  PUT ID 8. PLZ6S 8. GNR 8.;
DATA;
  INPUT SIND NB;
  FILE SIZES;
  GT1=1; GT2=2; S1=1; NB1=1; S2=2; S3=3; S4=4; S5=5; NB0=0;
  PUT SIND 8. GT1 1. GT2 1. S1 1. NB 7.
                            S2 1. NB0 7.
                            S3 1. NB0 7.
                            S4 1. NB0 7.
                            S5 1. NB0 7.
                            ;
  CARDS;
  1 1984
  2 300
  3 300
  4 300
  5 300
  6 1816
  ;
RUN;
```

10.13 Output von Job P663EXT in TSO Beta

```
NEXT LOGON U663BAT
1Logon accepted to library U663BAT.
1NEXT EX ECPBTEXT
1#CAPPLI
 DATA WSL1
1Page      1
-REGION :         1
    NBRE DE RECORDS DEMANDES SEL 0:     1984
    ** CODE POSTALE / GENRE D'EXTRACTION INVALIDE :    332100 / 1
     1984 RECORDS SELECTIONNES POUR LA REGION         1
0REGION :         2
    NBRE DE RECORDS DEMANDES SEL 0:      300
      300 RECORDS SELECTIONNES POUR LA REGION         2
0REGION :         3
    NBRE DE RECORDS DEMANDES SEL 0:      300
      300 RECORDS SELECTIONNES POUR LA REGION         3
0REGION :         4
    NBRE DE RECORDS DEMANDES SEL 0:      300
      300 RECORDS SELECTIONNES POUR LA REGION         4
0REGION :         5
    NBRE DE RECORDS DEMANDES SEL 0:      300
      300 RECORDS SELECTIONNES POUR LA REGION         5
0REGION :         6
    NBRE DE RECORDS DEMANDES SEL 0:     1816
    ** CODE POSTALE / GENRE D'EXTRACTION INVALIDE :    703101 / 1
    ** CODE POSTALE / GENRE D'EXTRACTION INVALIDE :    742802 / 1
```

```
    ** CODE POSTALE / GENRE D'EXTRACTION INVALIDE :    771002 / 1
     1816 RECORDS SELECTIONNES POUR LA REGION         6
0FIN DE TRAITEMENT : ECPBTEXT
 NOMBRE DE RECORDS DE L'ECHANTILLON ACTUEL     :    99655
 NOMBRE DE RECORDS TOTAL LUS (TYPE 0 ET 1)     :     4340
 NOMBRE DE NOUVEAUX RECORDS SELECTIONNES       :     5000
 NOMBRE DE REGIONS SELECTIONNEES               :        6
 NOMBRE DE RECORDS ECRITS APPLICATION STANDARD :     5000
 NOMBRE DE RECORDS ECRITS APPLICATION SAKE     :        0
1NAT9995 NATURAL SESSION TERMINATED NORMALLY
```

10.14 Ausgabe-Format

```
006300 DEFINE SUBROUTINE WRITE-STD     /*ARS1
006310 * -------------------------
006320 WRITE WORK FILE 3
006330   SEQ3.NTIRAGE                 /*        P  4.0
006340   SEQ3.NOMENAGE                /*        P  7.0
006350   SEQ3.DEXTRACT                /*        D
006360   SEQ3.CEXTRACT                /*        A    1
006370   SEQ3.NOCOMMUNE               /*        P  8.0
006380   SEQ3.NOREGION                /*        P  8.0  N
006390   SEQ3.NOCANTON                /*        P  3.0
006400   SEQ3.GENRTEL                 /*        A    1
006410   SEQ3.CLANGUE                 /*        A    1
006420   SEQ3.NOM                     /*        A   30  N
006430   SEQ3.PRENOM                  /*        A   25  N
006440   SEQ3.ALLIANCE                /*        A   25  N
006450   SEQ3.PROF                    /*        A   25  N
006460   SEQ3.RUE                     /*        A   25  N
006470   SEQ3.NMAISON                 /*        A    7  N
006480   SEQ3.COMPL                   /*        A   25  N
006490   SEQ3.CPOSTAL                 /*        A    4
006500   SEQ3.LOCALITE                /*        A   25  N
006510   SEQ3.IND1                    /*        A    3  N
006520   SEQ3.NTEL1                   /*        P  7.0  N
006530   SEQ3.IND2                    /*        A    3  N
006540   SEQ3.NTEL2                   /*        P  7.0  N
006550   SEQ3.DMUTATION               /*        D
006560   SEQ3.CSINTERVIEW             /*        A  2.0
006570   SEQ3.CSADRESSE               /*        A  8.0
006580   SEQ3.GADAT                   /*        A    8  CE2
006590   SEQ3.MNAME                   /*        A   30  CE2
006600   SEQ3.HAUSB                   /*        A   25  CE2
006610   SEQ3.PLZ6S                   /*        A    6  CE2
006620   SEQ3.GBDAT                   /*        A    8  CE2
006630   SEQ3.VZORT                   /*        A   39  CE2
006640   SEQ3.UNTXT                   /*        A   10  CE2
006650 * SEQ3.RES1                    /*        A   75  CE2  CE5
006660   SEQ3.RES1                    /*        A   67  CE2  CE5
006670   SEQ3.RESULT-E663RA           /*        N   1.7      CE5
006680 *
```

Achtung: Die Variable DIENST ist in GADAT mit Format A5 und die Variable WERBESP in GBDAT mit Format A1 abgespeichert.

10.15 Umwandlung output in ASCII mit Programm OUTREAD

```
//T7BCJS JOB (317,04,D397000),'BCJ-HULL',
//           MSGLEVEL=(1,1),CLASS=N,
//           NOTIFY=T7BCJ,MSGCLASS=T
//LOESCH  EXEC ERZTSM,CL='*'
//SYSIN    DD   * *
  DELETE D663.A663SWC.ECHWSL1.TEXT
//SAS    EXEC SAS
  /* LIEST DIE AUSGABE VON ECPBTEXT */;
FILENAME INP  'D663.A663SWC.ECHWSL1' DISP=OLD;
FILENAME OUT 'D663.A663SWC.ECHWSL1.TEXT'
            DISP=(NEW,CATLG) RECFM=FB LRECL=80;
OPTIONS LINESIZE=80;
DATA DATA1;
  INFILE INP;
  INPUT  NTIR PD3. NMEN PD4. @12 CEXT $1.
  NCOM PD5. NREG PD5.
```

```
  NCANT PD2. GENRTEL $1. CLANGUE $1.
  @218 IND1 $3. NTEL1 PD4.
  @246 DIENST $5.
  @309 PLZ6S 6. WERBESP 1.;
PROC FREQ;
  TABLES NREG;
  TABLES IND1;
  TABLES DIENST;
  TABLES GENRTEL;
  TABLES CLANGUE;
DATA;
  SET DATA1;
  FILE OUT;
  PUT NMEN 7. NREG 8. GENRTEL 1. CLANGUE $1.
      IND1 3. NTEL1 7. DIENST $5. PLZ6S 6.;
RUN;
```

Methodenberichte des Dienstes Statistische Methoden

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.

Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.

Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Eichenberger, P. (1999). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

Rapports de méthodes du Service des méthodes statistiques

Une liste complète des titres de cette série de publications figure à la fin de ce rapport. Les rapports de méthodes peuvent être commandés auprès de:

Office fédéral de la statistique
Service des publications
Espace de l'Europe 10
CH-2010 Neuchâtel
Tél.: (+41 32) 71 36060
Fax: (+41 32) 71 36061
Internet: Order@bfs.admin.ch

Methodenberichte des Dienstes Statistische Methoden

Ein vollständige Liste der Titel dieser Serie befindet sich am Ende dieses Berichts. Die Methodenberichte können bestellt werden bei:

Bundesamt für Statistik
Publikationsdienst
Espace de l'Europe 10
CH-2010 Neuchâtel
Tel.: (+41 32) 71 36060
Fax: (+41 32) 71 36061
Internet: Order@bfs.admin.ch