Methodenbericht

Betriebszählung 2001

0

Statistische Grundlagen und Übersichten
Bases statistiques et produits généraux
Basi statistiche e presentazioni generali

Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse

OFS BFS UST

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica

Neuchâtel, 2002

Statistik der Schweiz

Die vom Bundesamt für Statistik (BFS) herausgegebene Reihe «Statistik der Schweiz» gliedert sich in folgende Fachbereiche:

0 Statistische Grundlagen und Übersichten
1 Bevölkerung
2 Raum und Umwelt
3 Arbeit und Erwerb
4 Volkswirtschaft
5 Preise
6 Industrie und Dienstleistungen
7 Land- und Forstwirtschaft
8 Energie
9 Bau- und Wohnungswesen
10 Tourismus
11 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, Medien, Zeitverwendung
17 Politik
18 Öffentliche Verwaltung und Finanzen
19 Rechtspflege
20 Einkommen und Lebensqualität der Bevölkerung
21 Nachhaltige Entwicklung und regionale Disparitäten

Statistique de la Suisse

La série «Statistique de la Suisse» publiée par l'Office fédéral de la statistique (OFS) couvre les domaines suivants:

0 Bases statistiques et produits généraux
1 Population
2 Espace et environnement
3 Vie active et rémunération du travail
4 Economie nationale
5 Prix
6 Industrie et services
7 Agriculture et sylviculture
8 Energie
9 Construction et logement
10 Tourisme
11 Transports et communications
12 Monnaie, banques, assurances
13 Protection sociale
14 Santé
15 Education et science
16 Culture, médias, emploi du temps
17 Politique
18 Administration et finances publiques
19 Droit et justice
20 Revenus et qualité de vie de la population
21 Développement durable et disparités régionales

Statistik der Schweiz | **Statistique de la Suisse**

Methodenbericht | **Rapport de méthodes**

Betriebszählung 2001

Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse

Autoren **Jann Potterat und Paul-André Salamin**
Bundesamt für Statistik

Herausgeber **Bundesamt für Statistik**

OFS BFS UST

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica

Neuchâtel, 2002

Vorwort

Die Betriebszählung des Bundesamtes für Statistik (BFS) ist eine gesamtschweizerische Befragung aller Betriebe und Unternehmen des Industrie- und Dienstleistungssektors. Im September 2001 hat die jüngste Erhebung stattgefunden. Das BFS hat sich zum Ziel gesetzt, nach erfolgter Erhebung möglichst rasch erste Ergebnisse der Betriebszählung der Öffentlichkeit zur Verfügung zu stellen. Für die Ermittlung von provisorischen Ergebnissen, wurde erstmals ein spezielles Hochrechnungsverfahren angewandt, welches vom Dienst Statistische Methoden des BFS entwickelt worden war.

Die Arbeiten wurden durch Jann Potterat und Paul-André Salamin vom Dienst Statistische Methoden ausgeführt und von Peter Buchs, Ernst Matti und Pascal Utz von der Sektion Unternehmen und Beschäftigung begleitet.

Zusammenfassung

In diesem Bericht werden die verwendeten Methoden zur Berechnung der provisorischen Ergebnisse der Betriebszählung 2001 beschrieben.

Unter Berücksichtigung von erhofften Genauigkeiten der Resultate und dem gewünschten Publikationstermin wurde aus den an der Betriebszählung teilnehmenden Arbeitsstätten eine geschichtete Bernoulli-Stichprobe gezogen. Diese Betriebe der Stichprobe wurden wenn möglich während des ganzen Prozesses der Betriebszählung, wie insbesondere bei der Datenbereinigung, gegenüber den anderen Betrieben priorisiert. Mitte April 2002 war der Rücklauf der Stichprobe ausreichend gut, so dass die provisorischen Resultate berechnet werden konnten.

Die Methode zur Berechnung der Schätzer basierte auf einer Kombination eines einfachen Schätzers und eines Quotientenschätzers, für welche zugleich auch passende Varianzschätzer berechnet wurden.

Mit den verwendeten Methoden war es möglich, die Anzahl der Beschäftigten wie auch die Anzahl Arbeitsstätten in der Schweiz und für gewisse Untersuchungsbereiche mit genügender Genauigkeit zu schätzen.

Schlüsselwörter

Methodenbericht; Betriebszählung; BZ01; provisorische Ergebnisse; Stichprobe; Schichtung; Bernoulli; Schätzung; Quotientenschätzer; Gewichtung.

Auskunft: Jann Potterat, Tel. 032 713 65 90
Jann.Potterat@bfs.admin.ch

Bearbeitung: Dienst Statistische Methoden, BFS

Vertrieb: Bundesamt für Statistik
CH-2010 Neuchâtel
Tel. 032 713 60 60 / Fax 032 713 60 61
Order@bfs.admin.ch

Internet: http://www.statistik.admin.ch

Bestellnummer: 338-0011

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Grafik/Layout: BFS

Copyright: BFS, Neuchâtel 2002
Abdruck – ausser für kommerzielle Nutzung –
unter Angabe der Quelle gestattet.

ISBN: 3-303-00248-7

Inhaltsverzeichnis

1 Einleitung **5**

2 Stichprobenrahmen **5**

3 Stichprobenplan **6**
3.1 Definition der Schichtung 6
3.2 Stichprobengrösse und Aufteilung der Stichprobe 8
3.3 Ziehung der Stichprobe 10
3.4 Ergänzungen zum Stichprobenplan 11

4 Schätzverfahren **13**
4.1 Datenbereinigungen 13
4.2 Notationen 16
4.3 Gewichtung 18
4.4 Schätzmethoden 18
4.4.1 Einfacher Schätzer 18
4.4.2 Quotientenschätzer (Separate Ratio) 21
4.4.3 Kombination der Methoden 25
4.4.4 Beispiel 26

5 Resultate **28**

Anhang **30**

A Stichprobenplan je Schicht **30**

B Verwendete Daten, Tabellen und Programme **35**

C SAS-Programm zur Hochrechnung **37**

Literatur **38**

1 Einleitung

Die Betriebszählung (BZ) des Bundesamtes für Statistik (BFS) ist eine gesamtschweizerische Befragung aller Betriebe und Unternehmen des Industrie- und Dienstleistungssektors (also ohne Landwirtschaft). Sie wird im Abstand von 3 bis 4 Jahren durchgeführt. Mit Stichtag 28. September 2001 hat die jüngste Erhebung, die BZ01, stattgefunden. Es wurden gegen eine halbe Million Unternehmen und Betriebe angeschrieben und über ihre wirtschaftliche Tätigkeit, die Anzahl der beschäftigten Personen und die internationale Verflechtung befragt.
Das BFS hat sich zum Ziel gesetzt, nach erfolgter Erhebung möglichst rasch erste Ergebnisse der Betriebszählung der Öffentlichkeit zur Verfügung zu stellen. Für die Ermittlung von provisorischen Ergebnissen, wurde erstmals ein spezielles Hochrechnungsverfahren angewandt, welches im Auftrag der Sektion Unternehmen und Beschäftigung (UNT) vom Dienst Statistische Methoden (METH) entwickelt worden war. In diesem Bericht werden die verwendeten Methoden zur Berechnung dieser provisorischen Ergebnisse beschrieben.
Im zweiten und im dritten Kapitel wird das Verfahren zur Bestimmung der Stichprobe beschrieben, im Kapitel vier werden die angewendeten Schätzverfahren erläutert. Die provisorischen Resultate der BZ erschienen schlussendlich in einer Pressemitteilung des BFS am 3. Mai 2002 (siehe [1]), diese sind auch am Ende dieses Berichts im fünfen Kapitel aufgeführt.

2 Stichprobenrahmen

Die Konstruktion des Stichprobenrahmens, woraus die Stichprobe gezogen wurde, erfolgte in zwei Schritten.



Aus dem Betriebs- und Unternehmensregister BUR wurden nach bestimmten Kriterien Unternehmen bzw. Arbeitsstätten (AST) ausgewählt und extrahiert, um die (Oracle-) Datenbank der Betriebszählung (die BZ-DB) zu kreieren. Diese Kriterien sind ausführlich in einem Dokument der Sektion UNT beschrieben (siehe [2]). Grundsätzlich wurden nur die *aktiven* Unternehmen mit ihren AST, die einen Fragebogen der Betriebzählung ausfüllen mussten, in die BZ-DB geladen. Zum Zeitpunkt der Extraktion im August 2001 waren 487'180 AST in der BZ-DB.
Als zweiter Schritt wurden anhand von vier Variablen der BZ-DB die AST bestimmt, welche in den Stichprobenrahmen aufgenommen wurden oder nicht (siehe Tabelle 1).

Tabelle 1: Kriterien zur Aufnahme einer AST in den Stichprobenrahmen

Variable BZ-DB	Bedeutung	Kriterium zur Aufnahme	Anzahl AST Total	Anzahl AST ausserhalb	Anzahl AST im Rahmen
workflow_status_cd	Workflow-Status	$\neq$ 'Z'	487'180	29'935	457'245
noga_cd	Wirtschaftsaktivität	$\in [10, 93]$	457'245	4'312	452'933
canton_cd	Kanton	$\neq$ 'LI'	452'933	1	452'932
domain_cd	Aufarbeitungssektor	$\neq$ 'Missing'	452'932	39	452'893

Die Variable 'Workflow_status_cd' oder kurz Workflowstatus gibt an, in welchem Stadium des Arbeitsablauf-Prozesses sich der Fragebogen zu einem bestimmten Zeitpunkt befindet. Zu Beginn der Erhebung waren nur zwei Ausprägungen möglich:

- 'A'= AST in BZ-DB geladen, bereit für Versand
- 'Z'= AST in BZ-DB geladen, nicht BZ relevant (gelöscht)

Weitere Ausprägungen dieser Variablen folgten später während der Phase der Datenbereinigung der BZ (siehe dazu Kapitel 4.1).
Die 29'935 AST mit Status 'Z' sollten also nicht die Chance haben in die Stichprobe gezogen zu werden. Ebenfalls ausgeschlossen wurden die 4'312 AST, die nicht im zweiten oder im dritten Wirtschaftssektor tätig sind, wie auch eine AST aus Liechtenstein, sowie die 39 AST, die keinem Aufarbeitungssektor (Einzelbetrieb (EB), Unternehmen (UNT) oder öffentlicher Betrieb (ÖFF)) zugewiesen wurden.
Der definitive Stichprobenrahmen bestand aus **452'893** AST.

3 Stichprobenplan

3.1 Definition der Schichtung

Bei einer Stichprobe ist es meistens von Vorteil eine Schichtung vorzunehmen. In Fall der BZ-Stichprobe wurden folgende vier Variablen zur Bestimmung der Schichtung verwendet:

1. NOGA2
Eine übliche Schichtungsvariable bei Unternehmenserhebungen ist die wirtschaftliche Aktivität eines Betriebes, die NOGA. Da die Resultate u.a. nach diesem Kriterium publiziert werden sollten, wurde die NOGA auf Niveau 2-steller (NOGA2) gemäss BUR als erste Schichtungsvariable gewählt.

2. Aufarbeitungssektor
Wie bereits erwähnt, gab es bei der BZ drei Aufarbeitungssektoren:

- Einzelbetriebe (EB)
- Unternehmen (UNT)
- Öffentliche Betriebe (ÖFF)

Die Fragebögen (FB) der AST in den drei Sektoren wurden während der BZ von verschiedenen Arbeitsgruppen bearbeitet. Da ein Unternehmen aus sehr vielen AST bestehen kann, und die FB der AST innerhalb eines Unternehmens nicht unabhängig voneinander bearbeitet werden konnten, bedeutete dies für die FB der Unternehmen einen mehrheitlich grösseren Arbeitsaufwand als für diejenigen der Einzelbetriebe. Ebenso kann die Zeit bis die FB retourniert werden von Sektor zu Sektor variieren. Von den Arbeitsgruppen der BZ wurde daher vorgegeben, wieviele FB je Sektor sie ungefähr bis zum Termin der Hochrechnung bearbeiten können. So wurde also der Aufarbeitungssektor als zweite Schichtungsvariable gewählt.

3. BetotBURFlag
Die dritte Schichtungsvariable (BetotBURFlag) ist eine Indikatorvariable, die für jede AST angibt, ob die Anzahl Beschäftigte (Betot) gemäss BUR positiv ist. Die Anzahl Beschäftigte wird berechnet als Summe der Vollzeit- (VZ) und der Teilzeitbeschäftigten (Teilzeit I (TZ1) und Teilzeit II (TZ2)), also BetotBUR=VZ+ TZ1+ TZ2. Der Indikator BetotBURFlag wird folgendermassen definiert:

$$\text{BetotBURFlag} = \begin{cases} 1, & \text{falls BetotBUR} > 0 \\ 0, & \text{falls BetotBUR} = 0 \text{ or } = \text{'Missing'} \end{cases}$$

Für die AST im Stichprobenrahmen ergaben sich 50'706 AST mit BetotBURFlag=0 und 402'187 AST mit BetotBURFlag=1. Aus Gründen der beabsichtigten Hochrechnungsmethode (siehe Kapitel 4.4.2), wurde diese Variable als dritte Schichtungsvariable berücksichtigt.

4. VZAeqFlag
Die Anzahl Vollzeitäquivalente (VZAeq) resultieren aus einer Umrechnung von Teilzeit- auf Vollzeitstellen. Sie werden ermittelt, indem die nach drei Kategorien unterteilten Beschäftigten (VZ, TZ1 und TZ2) mit dem mittleren Beschäftigungsgrad der jeweiligen Kategorie multipliziert werden.

$$\text{VZAeq=VZ} + \alpha\times\text{TZ1} + \beta\times\text{TZ2}$$

Die Koeffizienten α und β werden je nach der wirtschaftlichen Aktivität jedes Jahr neu festgelegt und betragen $\alpha \approx 0.63$ bzw. $\beta \approx 0.30$.
Die vierte Schichtungsvariable (der VZAeqFlag) ist ein Indikator, der für jede AST angibt, ob die Anzahl Vollzeitäquivalente laut BUR 'gross' ist. Diese Variable wurde als vierte Schichtungsvariable berücksichtigt, da beabsichtigt wurde, alle grossen Betriebe in die Stichprobe zu ziehen.
Nun musste noch definiert werden, was genau unter 'gross' zu verstehen ist. Für jede Kreuzung Aufarbeitungssektor×NOGA2 wurde bei den AST mit BetotBURFlag=1 im Stichprobenrahmen ein bestimmtes p-Quantil $x_{[p]}$ für die Variable VZAeq wie folgt berechnet:

EB:		
NOGA2 mit	100 bis 10'000 AST	$\rightarrow p = 90\%$
NOGA2 mit	mehr als 10'000 AST	$\rightarrow p = 95\%$

UNT / ÖFF:		
NOGA2 mit	100 bis 1'000 AST	$\rightarrow p = 90\%$
NOGA2 mit	mehr als 1'000 AST	$\rightarrow p = 95\%$

Alle diese Quantile je Aufarbeitungssektor×NOGA2 wurden auf denjenigen der Werte 20, 50 oder 100 gerundet, der am nächsten vom entsprechenden Quantil liegt:

$$\tilde{q} = \underset{q \in \{20,50,100\}}{\arg\min} |x_{[p]} - q|$$

Der VZAeqFlag wurde nun für jede AST wie folgt definiert:

$$\text{VZAeqFlag} = \begin{cases} 1, & \text{falls VZAeq} > \tilde{q} \\ 0, & \text{sonst} \end{cases}$$

Bemerkungen:

1. Für die AST in den NOGA's mit insgesamt weniger als 100 AST (NOGA's 10, 11 und 23) wurde die obige Berechnung nicht durchgeführt, sondern es wurde VZAeqFlag=0 gesetzt.

2. Insgesamt resultierten 14'049 AST mit VZAeqFlag=1 und 438'844 AST mit VZAeqFlag=0

3.2 Stichprobengrösse und Aufteilung der Stichprobe

Die Berechnung der Stichprobengrösse und die anschliessende Aufteilung (Allokation) zu den Schichten war von zwei Faktoren abhängig:

- Bis Ende Januar 02, dies war ursprünglich der ungefähre Termin für die Hochrechnung, plante man folgende Anzahl AST bearbeiten zu können:
  Total ca. 100-110'000 AST, davon höchstens 30'000 aus dem Sektor UNT und 15'000 von ÖFF.

- Die Genauigkeit sollte genügend gut sein, d.h. der Variationskoeffizient für den Schätzer für das Total Anzahl Beschäftigte je NOGA2 dürfte höchstens 5% betragen.

Wie im Abschnitt 3.1 beschrieben, wurden die AST des Stichprobenrahmens in die Kreuzung folgender vier Schichten eingeteilt:

NOGA2×Aufarbeitungssektor×BetotBURFlag×VZAeqFlag

Diese Kreuzung ergab insgesamt 355 Schichten.
Es wurde zusätzlich berücksichtigt, ob eine Schicht h klein war oder nicht, wobei die Grenze bei N_h=100 AST festgelegt worden war (siehe Tabelle 2).

Tabelle 2: Typen von Schichten

Nummer	VZAeqFlag	BetotBURFlag	N_h	Anzahl AST	Anzahl Schichten
1	1	1	> 100	11'475	34
	1	1	≤ 100	2'574	60
	0	0	≤ 100	1'899	71
	0	1	≤ 100	1'274	49
2	0	1	> 100	386'864	93
3	0	0	> 100	48'807	48
Total:				452'893	355

Anhand dieser Vorgaben wurde nun die Berechnung der Stichprobengrösse vorgenommen.
Bei der anschliessenden Aufteilung auf die Schichten wurden je nach Nummer (siehe Tabelle 2) eine andere Methode angewandt:

Nummer=1: Take-all Schichten

Nummer=2: Schichten mit optimaler Aufteilung

Nummer=3: Optimale Aufteilung nach Imputation

1. Take-all Schichten
Bei den Schichten mit der Nummer=1, dies sind die grossen Arbeitsstätte und diejenigen mit einer kleiner Anzahl AST je NOGA2, wurde eine Take-all Allokation vorgenommen, d.h. *sämtliche* AST in diesen Schichten wurden erhoben (Vollerhebung). Insgesamt sind dies 214 Schichten mit 17'222 AST. Die grossen AST sollten möglichst alle in einer Stichprobe sein, denn so ist die Gefahr von grob verzerrten Resultaten kleiner. Die AST in den NOGA's mit weniger als 100 AST wurden ebenfalls voll erhoben, da es in zu kleinen Schichten nicht unbedingt sinnvoll ist, eine Stichprobe zu ziehen.

2. Schichten mit optimaler Aufteilung
Bei den Schichten mit Nummer=2, also den 93 Schichten, für die im BUR eine Hilfsinformation vorliegt, wurde eine optimale Allokation nach Neyman vorgenommen. Mit Hilfe eines S-Plus Programms wurde unter Vorgabe von Genauigkeiten zwischen 2-8%, die man für den Schätzer für die Anzahl Beschäftigte je NOGA2 erreichen wollte, die Stichprobengrössen in diesen Schichten berechnet.

3. Optimale Aufteilung nach Imputation
Ein erster Ansatz zur Aufteilung der Stichprobe bei den 48 Schichten mit Nummer=3 war die proportionale Aufteilung. Doch diese war nicht befriedigend (n zu gross). So wurde folgender Ansatz gewählt. Für die in diesen Schichten ja nicht existierende Hilfsvariable (Betot gemäss BUR) wurde ein Wert berechnet und eingesetzt. Bei der Imputation wurde die Annahme getroffen, dass die Varianzen und die Erwartungswerte dieser fiktiven Hilfsvariablen innerhalb einer NOGA×Aufarbeitungssektor gleich sind, wie bei der Schicht (Nummer=2), in der eine Hilfsvariable vorhanden war. Nun wurde die Stichprobengrösse so gewählt, dass die Genauigkeit halb so gross ist, wie bei der entsprechenden Schicht mit Nummer=2.

Die Kombination der verschiedenen Methoden ergab eine provisorische Stichprobengrösse $n_h^\star$. Zusätzlich zu diesen Methoden wurden folgende Regeln bei der Bestimmung der Stichprobengrössen je Schicht angewandt.

1. $n_h^\star \geq 100$ in den Schichten mit Nummer=2 und Nummer=3

   Grund: keine zu kleinen Stichproben je Schicht.

2. $f_h^{-1} := \frac{N_h}{n_h^\star} \leq 40 \; \forall h$

   Das Stichprobengewicht $N_h/n_h^\star$, welches definiert ist als das Inverse des Stichprobenanteils f_h, sollte nicht zu gross sein.

3. $f_h \leq 0.9$ in den Schichten mit Nummer=2 und Nummer=3

   Der Stichprobenanteil je Schicht soll bei den Schichten, die nicht voll erhoben werden, nicht in der Nähe von 1 sein. Falls dies trotzdem der Fall sein sollte, rundet man ihn lieber auf, d.h. nimmt alle AST in die Stichprobe.

4. $f_h \leq 0.8$ je NOGA2, falls nicht ganze NOGA2 ein take-all ist.

   Begründung analog zu 3., aber je NOGA.

Wurde mindestens eine dieser vier Regeln verletzt, musste das $n_h^\star$ um so viele AST erhöht werden, bis alle erfüllt waren.

3.3 Ziehung der Stichprobe

Für die Ziehung der Stichprobe wurde die Methode der Bernoulli-Ziehung verwendet, welche den folgenden Algorithmus verwendet:
Sei $n_h^\star$ die (gemäss Berechnung aus Abschnitt 3.2) *erwartete* Stichprobengrösse und N_h die Populationsgrösse je Schicht h. Jeder AST k in der Schicht h wurde eine Zufallszahl $u_k \in [0, 1]$ zugeordnet. Falls diese Zahl die Regel $u_k < n_h^\star/N_h$ erfüllte, war die AST in der Stichprobe, sonst nicht. Zu erwähnen ist, dass bei dieser Methode die Stichprobengrösse nicht fix ist, d.h. die *effektive* Stichprobengrösse, bezeichnet mit n_h, kann von $n_h^\star$ abweichen.
Auf Wunsch der Sektion UNT sollte die Stichprobe möglichst viele AST enthalten, die auch in der BESTA-Stichprobe sind. Da der Stichtag der BESTA des dritten Quartals 2001 identisch ist wie derjenige der BZ, könnte man so gewisse Vergleiche ziehen, wenn eine AST in beiden Stichproben enthalten wäre.
Da die BESTA auch eine Bernoulli-Stichprobe benutzt, konnte diesem Wunsch entsprochen werden, indem bei den AST, die sowohl im Stichprobenrahmen der BZ wie

auch in dem der BESTA sind, die Zufallszahl aus der BESTA übernommen wurde. Für die AST, die sich nur im BZ-Rahmen befanden, welches v.a. die neu gegründeten AST waren, wurde eine neue Zufallszahl generiert. Der Effekt der Übernahme der Zufallszahlen aus der BESTA war, dass diejenigen AST in der BESTA-Stichprobe, welche also allesamt tiefe Zufallszahlen besitzen, eine höhere Chance hatten auch in die BZ-Stichprobe gezogen zu werden, obwohl natürlich die Schichtung und die Stichprobenanteile der beiden Stichproben unterschiedlich sind.

3.4 Ergänzungen zum Stichprobenplan

Die erste Stichprobenziehung, welche mit den oben beschriebenen Methoden am 27. August 2001 erfolgte, umfasste insgesamt 99'193 AST. Diese Grösse hätte einer Genauigkeit von 2% je NOGA2 entsprochen. Anfangs Oktober, als der 'sample_flag', d.h. die Bezeichnung welche AST in die Stichprobe gezogen wurden, in der BZ-Datenbank implementiert war, musste festgestellt werden, dass die Stichprobe zu gross gewählt worden war. Insbesondere wurde der Vorgabe, dass höchstens 30'000 AST von Unternehmen und 15'000 AST von öffentlichen Betrieben dabei sein dürfen, zu wenig Rechnung getragen. Es wurde vernachlässigt, dass eine Unternehmung bzw. ein öffentlicher Betrieb als ganze Einheit verarbeitet werden musste, es also z.B. nicht möglich war, nur einen bestimmten Teil (einzelne AST) innerhalb der Unternehmung zu plausibilisieren. Für die geplanten 25'221 AST der Unternehmen (oder 12'790 AST der öff. Betriebe) in der Stichprobe, hätten also tatsächlich 54'638 (bzw. 33'011) AST verarbeitet werden müssen, was zusammen mit den 61'182 Einzelbetrieben zu viel gewesen wäre. Die Stichprobe musste also verkleinert werden.
Folgendes war der Ablauf bei der erneuten Ziehung der neuen Stichprobe am 15. Oktober 2001. Am Stichprobenrahmen bzw. an der Definition der 355 Schichten wurde nichts geändert. Jedoch wurde die Definition der Take-all-Schicht leicht abgeändert: Eine AST gehörte nun zur Take-all Schicht, falls $N_h \leq 80$ (neu, vorher 100) oder VZAeqFlag=1 ist. Dies hatte zur Folge, dass neu nur noch 16'743 AST oder 209 Schichten ein Take-all waren. Nun musste das Splus-Programm mit verschiedenen Vorgaben an die Genauigkeit (von 2-8%) an den Schätzer des Totals je NOGA neu laufen gelassen werden. Für den Sektor EB wurde die Variante 2.5% und für die Sektoren UNT und ÖFF die Variante 8% gewählt. Da bei der ersten Stichprobe für alle drei Sektoren die Variante 2% galt, hatte dies also zur Folge, dass die Stichprobe, insbesondere bei UNT und ÖFF massiv verkleinert wurde, wodurch natürlich gegenüber der ersten Stichprobe mit einem Genauigkeitsverlust zu rechnen war. Durch die Wiederholung der Bernoulliziehung wurde die definitive, neue Stichprobe kreiert, welches effektiv eine Teilstichprobe der ersten Stichprobe ist, da bei der Ziehung dieselben Zufallszahlen verwendet wurden. Die neue Stichprobengrösse betrug **61'716** AST. Die AST der Stichprobe wurden in der Datenbank markiert und während des ganzen Prozesses der BZ wenn möglich gegenüber den anderen AST priorisiert, d.h. bei eventuellen Mahnaktionen und insbesondere bei der Datenbereinigung wurden sie zuerst bearbeitet. Die Tabelle 3 zeigt den Stichprobenplan pro NOGA2. Der Stichprobenplan je Schicht ist im Anhang A zu finden.

Tabelle 3: Stichprobenplan je NOGA2 BZ 2001

NOGA2	Pop.grösse N_h	Sp.grösse n_h	Sp.anteil f_h	NOGA2	Pop.grösse N_h	Sp.grösse n_h	Sp.anteil f_h
10	3	3	1.00	41	301	140	0.47
11	3	3	1.00	45	43'000	3'769	0.09
14	445	367	0.82	50	17'471	2'103	0.12
15	3'344	875	0.26	51	29'992	2'791	0.09
16	21	21	1.00	52	61'883	2'798	0.05
17	855	456	0.53	55	32'850	2'429	0.07
18	1'100	820	0.75	60	10'101	1'838	0.18
19	301	301	1.00	61	128	128	1.00
20	7'480	1'910	0.26	62	255	255	1.00
21	273	273	1.00	63	5'282	1'206	0.23
22	5'786	1'743	0.30	64	5'259	810	0.15
23	8	8	1.00	65	5'288	958	0.18
24	1'085	735	0.68	66	3'578	761	0.21
25	946	606	0.64	67	3'027	1'388	0.46
26	1'712	644	0.38	70	5'880	2'004	0.34
27	300	300	1.00	71	1'160	636	0.55
28	8'805	2'224	0.25	72	13'849	2'471	0.18
29	4'234	1'681	0.40	73	703	703	1.00
30	173	173	1.00	74	72'529	3'761	0.05
31	1'252	805	0.64	75	9'704	784	0.08
32	713	713	1.00	80	17'246	1'590	0.09
33	3'855	1'390	0.36	85	28'301	2'806	0.10
34	200	200	1.00	90	1'666	612	0.37
35	518	518	1.00	91	7'123	1'187	0.17
36	4'493	1'713	0.38	92	10'058	2'535	0.25
37	355	355	1.00	93	17'161	2'022	0.12
40	838	394	0.47				
				Total	452'893	61'716	

4 Schätzverfahren

4.1 Datenbereinigungen

In der Datenbank der BZ wurde wie erwähnt jede AST mit der Variablen Workflowstatus gekennzeichnet, die bestimmt, in welchem Stadium des Arbeitsablaufs sich der Fragebogen während der Aufarbeitung befindet. Nachdem es zu Beginn nur die Ausprägungen 'A' und 'Z' gab, kamen später die in der Tabelle 4 aufgelisteten Ausprägungen dazu.

Tabelle 4: Definition des Workflowstatus

Workflowstatus	Beschreibung
A	Erfasst BZ-DB
B	Versandt
C	Postretouren
D	Caisse Morte
E	FB erfasst
F	Potentieller L15
G	Spezialkodierung
H	Adressmutation
I	Fehler Stufe 1: NOGA-Plaus
J	Fehler Stufe 2: formelle Plaus
K	Fehler Stufe 3: logische Plaus
L	Fehler Stufe 4: BUR-Plaus
M	Fehler Stufe 5: UNT-Plaus
N	Löschkontrolle
O	OK für Plaus
Y	Plausibilisiert
Z	Gelöscht

Nach der Erfassung eines FB, der Kodierung der Wirtschaftsaktivität (NOGA) und einer eventuellen Adressmutation beginnt die Datenbereinigung, auch Plausibilisierung genannt. Die Fragebögen durchlaufen in einem System fünf Typen von Tests, welche aus mehreren Subtests bestehen. Einige dieser Subtests wurden von METH entwickelt (siehe Methodenbericht [3]). Die genaue Beschreibung dieser und der übrigen Tests befinden sich auch in einem Dokument bei der Sektion UNT (siehe [4]) und werden hier nicht mehr genauer erklärt.
Das Schema in der Abbildung 1 zeigt den Durchlauf der fünf Tests. Zur Erläuterung folgendes: Zu Beginn der Datenbereinigung erhält die AST den Status 'O-1'. Dies bedeutet: Workflowstatus 'O' mit der Angabe welcher Test folgt ('1' = der erste). Beim fehlerfreien Durchlauf erhält die AST nacheinander die Status O-2, O-3, O-4, O-5 und Y. Falls eine AST einen bestimmten Test nicht besteht, erhält er je nach Test den Status I-1, J-2, K-3, L-4 oder M-5, d.h. eine nochmalige Überprüfung mit einer eventuellen Anpassung der Daten ist notwendig, bevor der nächste Test durchlaufen werden kann.

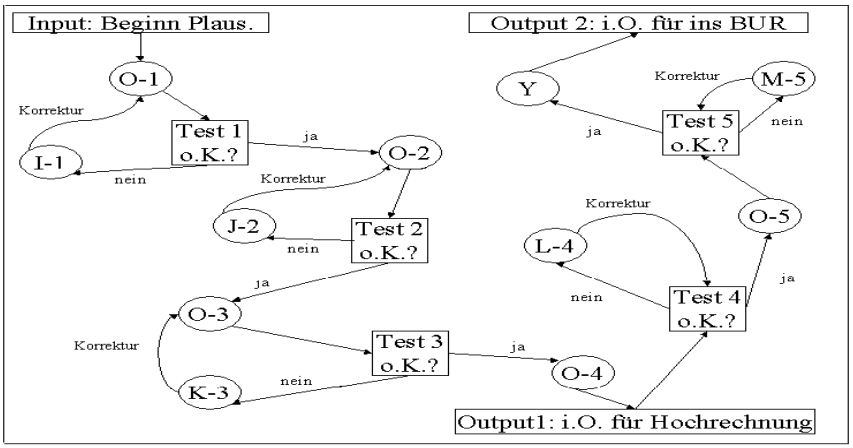


Abbildung 1: Workflow-System BZ 2001.

Am 15. April 2002, dies war der Zeitpunkt wo die Hochrechnung durchgeführt wurde, befanden sich nicht alle AST der Stichprobe in einem 'Endzustand', d.h. wo sie entweder fertig plausibilisiert waren oder gelöscht wurden, weil sie nicht mehr BZ relevant waren. Für die Hochrechnung wurden grundsätzlich alle AST berücksichtigt, die die ersten drei Plausibilisierungstests bestanden haben, also diejenigen mit Workflowstatus O-4 (siehe Abbildung 1). Da die Resultate ja auf Niveau AST publiziert werden sollten, war es unbedeutend, dass die AST die Tests vier und fünf noch nicht durchlaufen hatten, da diese nur das Unternehmen, nicht aber die Arbeitsstätte betrafen. Jede AST wurde einem von insgesamt sechs Status zugeordnet, deren Bedeutung in der Tabelle 5 beschrieben werden.

Tabelle 5: Beschreibung der verschiedenen Status

Status	Workflowstatus	Beschreibung
1	D, F, N, Z	nicht mehr aktiv, bzw. BZ relevant.
2	A, B, C	Fragebogen noch nicht eingetroffen (Missing)
3	E, G-K, O-1 O-2, O-3	FB eingetroffen, aber noch nicht fertig plausibilisiert
4	L, M, O-4, Y	FB hat Plausibilisierungstests bestanden
5	div. mit Status 3	FB, die ausserhalb des BZ-Systems plausibilisiert wurden, aber darin noch nicht als O-4 bezeichnet sind.
6	O-4	FB von AST, die nicht im Stichprobenrahmen sind, aber schon plausibilisiert. Dies sind AST, die zwischen der Ziehung der Stichprobe und dem Termin der Hochrechnung in die Datenbank aufgenommen wurden, also entweder neu gegründet wurden oder wegen einer Umstrukturierung eine neue Burnummer zugeteilt bekamen.

Die Tabelle 6 zeigt, wie viele AST sich bei der Hochrechnung in welchem Status befanden.

Tabelle 6: Anzahl AST in den einzelnen Status

Status	absolute Häufigkeit	relative Häufigkeit in % der Stichprobe
1	10'274	16.6
2	1'348	2.2
3	1'644	2.7
4	48'083	77.9
5	367	0.6
6	2'282	·

Die Summe der ersten fünf Häufigkeiten gibt 61'716 (=100%), welches die Stichprobengrösse ist.

Damit das Hochrechnungsprogramm, welches als Makro für die Statistiksoftware SAS geschrieben wurde und im Anhang aufgeführt wird, die Berechnungen der Resultate korrekt durchführen konnte, mussten die Daten vor der Hochrechnung aufbereitet werden. Die Aufbereitung wurde in Abhängigkeit des Status vorgenommen. So mussten die Daten der unplausibilisierten Fragebogen auf 'Missing' und diejenigen der nicht mehr aktiven AST auf '0' gesetzt werden. Obwohl die AST mit Status=2 theoretisch noch keine Einträge in der Datenbank haben sollten, gab es einige Fälle wo eine Variable oder ein Untersuchungsbereich (engl. Domain) wie z.B. die NOGA schon zum voraus kodiert wurden, d.h. bevor der Fragebogen zurück kam. Diese mussten für die Hochrechnung ebenfalls auf 'Missing' gesetzt werden. Die Tabelle 7 fasst diese Modifikationen zusammen.

Tabelle 7: Modifikation der BZ-DB in Abhängigkeit des Status

Status	Workflowstatus	Aktiv	(Datenbank)		(Hochrechnung)	
			Domain	Variable	Domain	Variable
1	D, F, N, Z	0	⋆	·	0	0
			·	·	0	0
			⋆	⋆	0	0
2	A, B, C	·	⋆	·	·	·
			·	·	·	·
3	E, G-K, O-1	1	⋆	⋆	·	·
	O-2, O-3		·	⋆	·	·
			⋆	·	·	·
4	L, M, O-4, Y	1	⋆	⋆	⋆	⋆
5	diverse	1	⋆	⋆	⋆	⋆
6	O-4	1	⋆	⋆	⋆	⋆
'⋆' = vorhandener, gültiger Wert, '·' ='Missing value'						

4.2 Notationen

Wir betrachten den Stichprobenrahmen F der Betriebszählung 2001 mit Grösse $|F| = N = 452'893$. Dieser Rahmen ist unterteilt in die 355 Schichten F_h. Es lässt sich also schreiben:

$$F = \bigcup_h F_h \quad \text{und} \quad |F| = N = \sum_h N_h$$

wobei $N_h = |F_h|$ die Anzahl Einheiten in der Schicht h sind.
Die aktiven AST der BZ bilden einen Untersuchungsbereich (engl. Domain) $U \subseteq F$ des Stichprobenrahmens. Ebenfalls von Interesse ist die Unterteilung in gewisse Sub-Domains $U = \bigcup_{d \in D} U_d$, wie zum Beispiel die Wirtschaftsaktivität (NOGA).
Folgende Notationen werden verwendet:

U	$M = \|U\|$
U_d	$M_d = \|U_d\|$
$U_h = F_h \cap U$	$M_h = \|U_h\|$
$U_{h,d} = F_h \cap U_d$	$M_{h,d} = \|U_{h,d}\|$

Es gelten die Beziehungen

$$M = \sum_h M_h \quad \text{und} \quad M_d = \sum_h M_{h,d}$$

In jeder Schicht wurde eine Bernoulli-Stichprobe $S_h \subseteq F_h$ gezogen. Es lässt sich schreiben

$$S = \bigcup_h S_h \quad \text{und} \quad |S| = n = \sum_h n_h$$

wobei n=61'716 die effektive Anzahl AST in der Stichprobe, und $|S_h| = n_h$ die Anzahl AST je Schicht ist. Die erwartete Anzahl AST in der Stichprobe wird mit $n^\star$, bzw. mit $n_h^\star$ für die Anzahl je Schicht, notiert.
Zusätzlich müssen die totalen Antwortausfälle (engl. unit nonresponse) berücksichtigt werden. Mit $R_h \subseteq S_h \subseteq F_h$ wird die Menge der AST je Schicht bezeichnet, die geantwortet haben. Es ist

$$R = \bigcup_h R_h \quad \text{und} \quad |R| = r = \sum_h r_h$$

Die Anzahl der antwortenden AST ist $r = 58'724$, was einer Antwortquote von 95.2% entspricht. Mit $|R_h| = r_h$ wird die entsprechende Anzahl je Schicht bezeichnet. Zu betrachten gilt es die Kreuzung der Stichprobe mit einem Untersuchungsbereich U.
Folgende Notationen werden verwendet:

$V = R \cap U$	$m = \|V\|$
$V_d = R \cap U_d$	$m_d = \|V_d\|$
$V_h = R \cap U_h$	$m_h = \|V_h\|$
$V_{h,d} = R \cap U_{h,d}$	$m_{h,d} = \|V_{h,d}\|$

Die Abbildung 2 gibt eine Übersicht der Schichten, Untersuchungsbereiche und Stichproben.

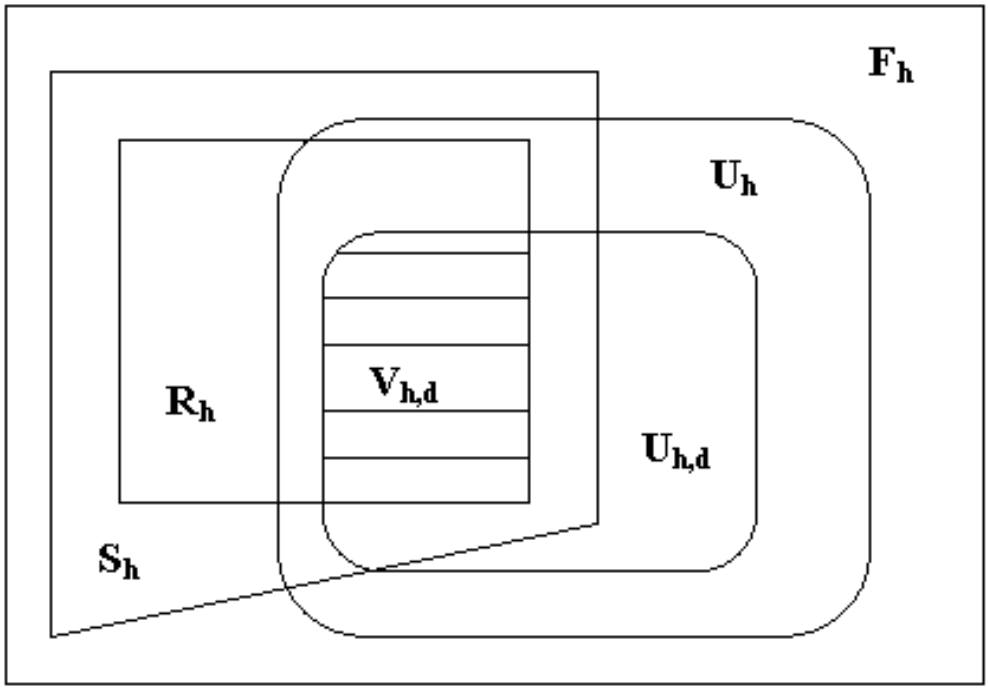


Abbildung 2: Übersicht der Schichten, Untersuchungsbereiche und Stichproben

Die Parameter von Interesse sind einerseits die Grössen der Untersuchungsbereiche $M = |U|$ und $M_d = |U_d|$, also die Anzahl aktiven AST, wie anderseits auch die Anzahl Beschäftigten der aktiven AST Y insgesamt, wie auch innerhalb von Untersuchungsbereichen Y_d. Diese werden wie folgt bezeichnet:

$Y = \sum_{k \in U} y_k$	$Y_d = \sum_{k \in U_d} y_k$
$Y_h = \sum_{k \in U_h} y_k$	$Y_{h,d} = \sum_{k \in U_{h,d}} y_k$

wobei die Beziehungen

$$Y = \sum_h Y_h \quad \text{und} \quad Y_d = \sum_h Y_{h,d}$$

gelten.

4.3 Gewichtung

Wenn $\pi_h = n_h^\star / N_h$ die Einschlusswahrscheinlichkeit der AST in der Schicht h ist, bezeichnet $d_h := \pi_h^{-1}$ das Anfangsgewicht für diese Schicht. Das Hochrechnungs-Gewicht w_h je Schicht h setzt sich aus drei Faktoren zusammen:

$$w_h = g_h^{(1)} g_h^{(2)} d_h = \frac{n_h}{r_h} \frac{n_h^\star}{n_h} \frac{N_h}{n_h^\star} = \frac{N_h}{r_h}$$

Der Faktor $g_h^{(2)} = n_h^\star / n_h$ bewirkt eine Kalibrierung des Anfangsgewichts d_h auf die effektive Grösse der Schicht und der Faktor $g_h^{(1)} = n_h / r_h$ eine Anpassung für die Nonresponse.

4.4 Schätzmethoden

Es wurde eine Kombination von zwei Schätzmethoden für die Berechnung der Resultate angewendet, wobei die Theorie aus Särndal et al. [5] entnommen wurde.

4.4.1 Einfacher Schätzer

Schätzer der Anzahl Beschäftigter:
Sei y_k der Wert einer Variable von Interesse für die AST k. Zu Schätzen ist das Total dieser Variablen aller *aktiven* AST

1. insgesamt (Y)
2. innerhalb eines Untersuchungsbereichs (Y_d)

Die Schätzer werden folgendermassen berechnet.

1. Total Beschäftigte aller aktiven AST:

$$Y = \sum_{k \in U} y_k = \sum_h \sum_{k \in U_h} y_k = \sum_h Y_h$$

Schätzer von Y

$$y[U]_k = \begin{cases} y_k & \text{falls } k \in U \\ 0 & \text{falls } k \notin U \Leftrightarrow k \in F \setminus U \end{cases}$$

$$\hat{Y}_h = \sum_{k \in R_h} w_k y[U]_k = \frac{N_h}{r_h} \sum_{k \in V_h} y_k$$

$$\Longrightarrow \hat{Y} = \sum_h \hat{Y}_h = \sum_h \frac{N_h}{r_h} \sum_{k \in V_h} y_k$$

Schätzung der Varianz von $\hat{Y}$

$$\widehat{\mathbf{Var}}\left(\hat{Y}_h\right) = N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} S^2_{y[U],R_h}$$

wobei

$$S^2_{y[U],R_h} = \frac{1}{r_h - 1} \sum_{k \in R_h} \left(y\,[U]_k - \overline{y\,[U]}_{R_h}\right)^2$$

mit

$$\overline{y\,[U]}_{R_h} = \frac{1}{r_h} \sum_{k \in R_h} y\,[U]_k$$

$$\Longrightarrow \widehat{\mathbf{Var}}\left(\hat{Y}\right) = \sum_h \widehat{\mathbf{Var}}\left(\hat{Y}_h\right)$$

2. Total Beschäftigte innerhalb eines Untersuchungsbereichs der aktiven AST:

$$Y_d = \sum_{k \in U_d} y_k = \sum_h \sum_{k \in U_{h,d}} y_k = \sum_h Y_{h,d}$$

Schätzer von Y_d

$$y[U_d]_k = \begin{cases} y_k & \text{falls } k \in U_d \\ 0 & \text{falls } k \notin U_d \Leftrightarrow k \in F \setminus U_d \end{cases}$$

$$\hat{Y}_{h,d} = \sum_{k \in R_h} w_k y[U_d]_k = \frac{N_h}{r_h} \sum_{k \in V_{h,d}} y_k$$

$$\Longrightarrow \hat{Y}_d = \sum_h \hat{Y}_{h,d} = \sum_h \frac{N_h}{r_h} \sum_{k \in V_{h,d}} y_k$$

Schätzung der Varianz von $\hat{Y}_d$

$$\widehat{\mathbf{Var}}\left(\hat{Y}_{h,d}\right) = N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} S^2_{y[U_d],R_h}$$

wobei

$$S^2_{y[U_d],R_h} = \frac{1}{r_h - 1} \sum_{k \in R_h} \left(y\,[U_d]_k - \overline{y\,[U_d]}_{R_h}\right)^2$$

mit

$$\overline{y\left[U_d\right]}_{R_h} = \frac{1}{r_h} \sum_{k \in R_h} y\left[U_d\right]_k$$

$$\Longrightarrow \widehat{\mathbf{Var}}\left(\hat{Y}_d\right) = \sum_h \widehat{\mathbf{Var}}\left(\hat{Y}_{h,d}\right)$$

Grösse der aktiven Population:
Ebenfalls von Interesse ist die Schätzung der Anzahl aktiven AST

1. insgesamt (M)
2. innerhalb eines Untersuchungsbereichs (M_d)

Hier die Berechnung der Schätzer

1. Anzahl aktive AST insgesamt:

$$M = \sum_h M_h$$

Schätzer von M

$$\chi\left[U\right]_k = \begin{cases} 1 & \text{falls } k \in U \\ 0 & \text{falls } k \notin U \Leftrightarrow k \in F \setminus U \end{cases}$$

$$\hat{M}_h = \sum_{k \in R_h} w_k \chi\left[U\right]_k = \frac{N_h}{r_h} m_h$$

$$\Longrightarrow \hat{M} = \sum_h \hat{M}_h = \sum_h \frac{N_h}{r_h} m_h$$

Schätzung der Varianz von $\hat{M}$

$$\widehat{\mathbf{Var}}\left(\hat{M}_h\right) = N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} \frac{m_h}{r_h - 1} \left(1 - \frac{m_h}{r_h}\right)$$

$$\Longrightarrow \widehat{\mathbf{Var}}\left(\hat{M}\right) = \sum_h \widehat{\mathbf{Var}}\left(\hat{M}_h\right)$$

2. Grösse eines Untersuchungsbereichs der aktiven Population:

$$M_d = \sum_h M_{h,d}$$

Schätzer von M_d

$$\chi\left[U_d\right]_k = \begin{cases} 1 & \text{falls } k \in U_d \\ 0 & \text{falls } k \notin U_d \Leftrightarrow k \in F \setminus U_d \end{cases}$$

$$\hat{M}_{h,d} = \sum_{k \in R_h} w_k \chi\left[U_d\right]_k = \frac{N_h}{r_h} m_{h,d}$$

$$\Longrightarrow \hat{M}_d = \sum_h \hat{M}_{h,d} = \sum_h \frac{N_h}{r_h} m_{h,d}$$

Schätzer der Varianz von $\hat{M}_d$

$$\widehat{\mathbf{Var}}\left(\hat{M}_{h,d}\right) = N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} \frac{m_{h,d}}{r_h - 1} \left(1 - \frac{m_{h,d}}{r_h}\right)$$

$$\Longrightarrow \widehat{\mathbf{Var}}\left(\hat{M}_d\right) = \sum_h \widehat{\mathbf{Var}}\left(\hat{M}_{h,d}\right)$$

4.4.2 Quotientenschätzer (Separate Ratio)

Für den Fall, dass aus einer externen Quelle (z.B. dem BUR) eine Hilfsgrösse bekannt ist, die mit der Variablen von Interesse positiv korreliert ist, kann die Methode des Quotientenschätzers verwendet werden.
Sei y_k der interessierende Wert einer Variable der AST k (z.B. die Anzahl Beschäftigte gemäss BZ). Mit x_k wird der entsprechende Wert der Hilfsgrösse (z.B. die Anzahl Beschäftigte gemäss BUR) bezeichnet. Je nachdem wie sich die Korrelation zwischen y_k und x_k in den einzelnen Schichten unterscheidet, kann die Methode des kombinierten (combined) - oder des getrennten (separate) - Quotientenschätzers gewählt werden (siehe Särndal et al. [5]). Es wurde die letztere gewählt. Für die Schätzung der Anzahl AST wurde auf den Quotientenschätzer verzichtet.
Die Quotientenschätzer für eine geschichtete Stichprobe, für

1. die Schätzung $\hat{Y}_{SR}$ des Totals der aktiven AST.
2. die Schätzung der Varianz von $\hat{Y}_{SR}$
3. die Schätzung $\hat{Y}_{SR,d}$ von Totalen innerhalb von Subdomains
4. die Schätzung der Varianz von $\hat{Y}_{SR,d}$

werden wie folgt berechnet.

1. Schätzer für das Total:

$$Y = \sum_{k \in U} y_k = \sum_h \sum_{k \in U_h} y_k = \sum_h Y_h$$

Quotientenschätzer $\hat{Y}_{SR}$

$$y\,[U]_k = \begin{cases} y_k & \text{falls } k \in U \\ 0 & \text{falls } k \notin U \Leftrightarrow k \in F \setminus U \end{cases}$$

$$x\,[U]_k = \begin{cases} x_k & \text{falls } k \in U \\ 0 & \text{falls } k \notin U \Leftrightarrow k \in F \setminus U \end{cases}$$

$$X_{F_h} = \sum_{k \in F_h} x_k, \qquad \hat{X}_{F_h} = \frac{N_h}{r_h} \sum_{k \in R_h} x_k$$

$$\hat{Y}_{SR,h} = X_{F_h} \frac{\sum_{k \in R_h} w_k y\,[U]_k}{\sum_{k \in R_h} w_k x_k} = X_{F_h} \frac{\sum_{k \in V_h} y_k}{\sum_{k \in R_h} x_k} = \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \hat{Y}_h$$

$$\Longrightarrow \hat{Y}_{SR} = \sum_h \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \hat{Y}_h = \sum_h \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \frac{N_h}{r_h} \sum_{k \in V_h} y_k$$

Der Quotientenschätzer verändert also die Gewichte w_h

$$w_{SR,h} = \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) w_h = \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \frac{N_h}{r_h}$$

Der Quotient $X_{F_h}/\hat{X}_{F_h}$ wird auch als g-factor bezeichnet.

2. Schätzung der Varianz von $\hat{Y}_{SR}$:

$$\widehat{\mathbf{Var}} \left(\hat{Y}_{SR,h} \right) = N_h^2 \left(1 - \frac{r_h}{N_h} \right) \frac{1}{r_h} S^2_{\hat{z}_h[U], R_h}$$

wobei

$$\hat{z}_h\,[U]_k = y[U]_k - \hat{\rho}_h x[U]_k, \quad k \in R_h$$

mit

$$\hat{\rho}_h = \frac{\sum_{k \in R_h} y_k}{\sum_{k \in R_h} x_k}$$

und

$$S^2_{\hat{z}_h[U],R_h} = \frac{1}{r_h - 1} \sum_{k \in R_h} \left(\hat{z}_h \left[U\right]_k - \overline{\hat{z}_h[U]}_{R_h} \right)^2$$

mit

$$\overline{\hat{z}_h \left[U\right]}_{R_h} = \frac{1}{r_h} \sum_{k \in R_h} \hat{z}_h \left[U\right]_k$$

$$\Longrightarrow \widehat{\mathbf{Var}} \left(\hat{Y}_{SR} \right) = \sum_h \widehat{\mathbf{Var}} \left(\hat{Y}_{SR,h} \right)$$

3. Schätzung des Totals eines Untersuchungsbereichs der aktiven AST:

$$Y_d = \sum_{k \in U_d} y_k = \sum_h \sum_{k \in U_{h,d}} y_k = \sum_h Y_{h,d}$$

Quotientenschätzer $\hat{Y}_{SR,d}$

$$y[U_d]_k = \begin{cases} y_k & \text{falls } k \in U_d \\ 0 & \text{falls } k \notin U_d \Leftrightarrow k \in F \setminus U_d \end{cases}$$

$$x[U_d]_k = \begin{cases} x_k & \text{falls } k \in U_d \\ 0 & \text{falls } k \notin U_d \Leftrightarrow k \in F \setminus U_d \end{cases}$$

$$X_{F_h} = \sum_{k \in F_h} x_k, \quad \hat{X}_{F_h} = \frac{N_h}{r_h} \sum_{k \in R_h} x_k$$

$$\begin{aligned} \hat{Y}_{SR,hd} &= X_{F_h} \frac{\sum_{k \in R_h} w_k y[U_d]_k}{\sum_{k \in R_h} w_k x_k} \\ &= X_{F_h} \frac{\sum_{k \in V_{h,d}} y_k}{\sum_{k \in R_h} x_k} = \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \hat{Y}_{h,d} \end{aligned}$$

$$\Longrightarrow \hat{Y}_{SR,d} = \sum_h \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \hat{Y}_{h,d} = \sum_h \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \frac{N_h}{r_h} \sum_{k \in V_{h,d}} y_k$$

Bemerkung: Der Quotientenschätzer eines Untersuchungsbereiches ist arithmetisch konsistent, im Sinne dass

$$\sum_d \hat{Y}_{SR,d} = \hat{Y}_{SR}$$

Beweis: Nach Definition von $\hat{Y}_{SR,d}$ haben wir

$$\sum_d \hat{Y}_{SR,d} = \sum_d \sum_h \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \hat{Y}_{h,d} = \sum_h \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \sum_d \hat{Y}_{h,d}$$

Da andererseits

$$\sum_d \hat{Y}_{h,d} = \sum_d \frac{N_h}{r_h} \sum_{k \in V_{h,d}} y_k = \frac{N_h}{r_h} \sum_d \sum_{k \in V_{h,d}} y_k = \frac{N_h}{r_h} \sum_{k \in V_h} y_k = \hat{Y}_h$$

folgt schliesslich

$$\sum_d \hat{Y}_{SR,d} = \sum_h \left(\frac{X_{F_h}}{\hat{X}_{F_h}} \right) \hat{Y}_h = \hat{Y}_{SR}$$

q.e.d.

4. Schätzung der Varianz von $\hat{Y}_{SR,d}$:

$$\widehat{\mathbf{Var}} \left(\hat{Y}_{SR,hd} \right) = N_h^2 \left(1 - \frac{r_h}{N_h} \right) \frac{1}{r_h} S^2_{\hat{z}_h[U_d],R_h}$$

wobei

$$\hat{z}_h[U_d]_k = y[U_d]_k - \hat{\rho}_h x[U_d]_k, \quad k \in R_h$$

mit

$$\hat{\rho}_h = \frac{\sum_{k \in R_h} y_k}{\sum_{k \in R_h} x_k}$$

und

$$S^2_{\hat{z}_h[U_d],R_h} = \frac{1}{r_h - 1} \sum_{k \in R_h} \left(\hat{z}_h[U_d]_k - \overline{\hat{z}_h[U_d]}_{R_h} \right)^2$$

mit

$$\overline{\hat{z}_h[U_d]}_{R_h} = \frac{1}{r_h} \sum_{k \in R_h} \hat{z}_h[U_d]_k$$

$$\Longrightarrow \widehat{\mathbf{Var}} \left(\hat{Y}_{SR,d} \right) = \sum_h \widehat{\mathbf{Var}} \left(\hat{Y}_{SR,hd} \right)$$

4.4.3 Kombination der Methoden

Für die Berechnung der provisorischen Resultate der BZ 2001 wurde aus einer Kombination der in den Abschnitten 4.4.1 und 4.4.2 beschriebenen Methoden der definitive Schätzer bestimmt. Für gewisse Schichten liegen tatsächlich geeignete Hilfsinformationen aus dem BUR vor. Dies ist der Fall in den Schichten mit den AST, die einen BetotBurFlag=1 besitzen (siehe Abschnitt 3.1).
Da es viele kleine take-all Schichten gab, wurden diese Schichten innerhalb von NOGA2×BetotBURFlag×VZAeqFlag über die Aufarbeitungssektoren zusammengelegt, so dass es neu 264 Schichten gab.
Zur Notation wird mit $\mathcal{H}$ die Menge aller Schichten bezeichnet, wobei $|\mathcal{H}| = H$. Diese wird in zwei Teilmengen $\mathcal{H}_\alpha$ und $\mathcal{H}_\beta$ unterteilt, wobei $\mathcal{H}_\alpha$ die Menge aller Schichten mit und $\mathcal{H}_\beta$ die Menge aller Schichten ohne Hilfsgrösse identifiziert. Es gilt $\mathcal{H}_\alpha \cap \mathcal{H}_\beta = \varnothing$ und $\mathcal{H}_\alpha \cup \mathcal{H}_\beta = \mathcal{H}$.
Für die Schätzung der interessierenden Variablen wurde nun in den Schichten mit Hilfsinformation $\mathcal{H}_\alpha$ der Quotientenschätzer, und in den Schichten ohne Hilfsinformation $\mathcal{H}_\beta$ diejenige des einfachen Schätzers angewendet.

1.) Schätzer der Totale

$$\hat{Y}_{KM} = \sum_{h \in \mathcal{H}_\alpha} \frac{X_{F_h}}{\hat{X}_{F_h}} \frac{N_h}{r_h} \sum_{k \in V_h} y_k \quad + \quad \sum_{h \in \mathcal{H}_\beta} \frac{N_h}{r_h} \sum_{k \in V_h} y_k$$

Der zweite Terme, also derjenige für $h \in \mathcal{H}_\beta$, kann so interpretiert werden, dass der Quotient $X_{F_h}/\hat{X}_{F_h} = 1$ gesetzt wurde.

2.) Schätzer der Varianz der Totale

$$\widehat{\mathbf{Var}}(\hat{Y}_{KM}) = \sum_{h \in \mathcal{H}_\alpha} N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} S^2_{\hat{z}_h[U],R_h} \quad + \quad \sum_{h \in \mathcal{H}_\beta} N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} S^2_{y[U],R_h}$$

3.) Schätzer für die Totale innerhalb von Subdomains

$$\hat{Y}_{KM,d} = \sum_{h \in \mathcal{H}_\alpha} \frac{X_{F_h}}{\hat{X}_{F_h}} \frac{N_h}{r_h} \sum_{k \in V_{h,d}} y_k \quad + \quad \sum_{h \in \mathcal{H}_\beta} \frac{N_h}{r_h} \sum_{k \in V_{h,d}} y_k$$

4.) Varianzschätzung innerhalb von Subdomains

$$\widehat{\mathbf{Var}}(\hat{Y}_{KM,d}) = \sum_{h \in \mathcal{H}_\alpha} N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} S^2_{\hat{z}_h[U_d],R_h} \quad + \quad \sum_{h \in \mathcal{H}_\beta} N_h^2 \left(1 - \frac{r_h}{N_h}\right) \frac{1}{r_h} S^2_{y[U_d],R_h}$$

4.4.4 Beispiel

Anhand von fiktiven Beispiels-Daten in der Tabelle 8 sollen die Berechnungen erläutert werden. Die Tabelle enthält Informationen von nach Status agregierten AST-Daten in vier Schichten der Stichprobe. Diese Schichten enthielten bei einem provisorischen Auszug der BZ-Datenbank mindestens 1 AST mit einem NOGA2-Code 100. Folgendes sind die Bedeutungen der Spalten in der Tabelle:

h	Schicht Nummer
schicht_neu	Nummerierung im Hochrechnungs File
status	Workflowstatus gemäss Tabelle 5
aktiv	Indikator, ob AST aktiv ist, $k \in U$
noga	fiktive NOGA2 nach BZ01
subdom	Indikator, ob AST im Untersuchungsbereich ist, $k \in U_d \subset U$
n_h	Anzahl AST in Stichprobe
Betot_DB	Beschäftigtentotal in BZ-DB
Betot_Hochr	für Hochrechnung modifiziertes Beschäftigtentotal y_k
Betot_BUR	Beschäftigtentotal im BUR
w_h	Hochrechnungsgewicht

In der unteren Hälfte der Tabelle 8 befinden sich ein paar Variablen, die zur Berechnung hilfreich sind. Die Notation dieser Variablen wurde bereits in den Kapiteln 4.4.1 bis 4.4.3 eingeführt.
Zur Schätzung von Betot im Untersuchungsbereich aktive AST in NOGA 100, müssen wie erwähnt bei den AST, welche nicht mehr aktiv sind (status=1, bzw. aktiv=0) oder nicht zum Untersuchungsbereich NOGA100 gehören, die Variable Betot_DB modifiziert werden, indem die Werte der Variablen auf '0' gesetzt wurden (Betot_Hochr). Bei den erhaltenen Fragebogen, die noch nicht vollständig plausibilisiert sind (status=3) oder bei den fehlenden FB (status=2), wurde Betot_Hochr='.' und noga= '.' gesetzt. Im Beispiel dieser NOGA gab es keine AST mit Status 5 oder 6. Für die Hochrechnung werden also die AST mit Status=1 oder 4 verwendet.
Die Berechnung kann mit der SAS-Prozedur `proc surveymeans` durchgeführt werden. Dazu sind einerseits im File `Daten` die Variablen `schicht_neu, wh, Betot_DB` und `noga` und andererseits im File `Stplan` die Variablen `_total_` und `schicht_neu` gegeben, wobei `_total_` die Populationsgrösse je Schicht N_h bezeichnet.
Der entsprechende Code lautet:

```
proc surveymeans data=Daten total=Stplan missing sum cv std CLSUM;
domain      noga;               /*Untersuchungsbereich*/
strata      schicht_neu;        /*Schichtungsvariable*/
var         Betot_DB;           /*Hochrechnungsvariable*/
weight      wh;                 /*Gewichte*/
ods output  domain=dataoutput1; /*Exportieren der Resultate*/
run;
```

Das vollständige SAS-Makro ist im Anhang C aufgelistet. Zur Kontrolle und zum Verstehen der SAS-Resultate kann folgende Berechnung hilfreich sein.

Tabelle 8: Beispiel zur Berechnung der Schätzer

h	schicht_neu	status	aktiv	noga	subdom	n_h	Betot_DB	Betot_Hochr	Betot_Bur	w_h
1	1019	4	1	100	1	1	1	1	0	1.0625
		4	1	$\neq 100$	0	7	248	0	0	1.0625
		1	0	0	0	8	0	0	0	1.0625
		2,3	·	·	·	1	·	·	0	·
					total	17	249	1	0	
2	1124	4	1	100	1	196	8'507	8'507	8'090	1.0658
		4	1	$\neq 100$	0	34	553	0	464	1.0658
		1	0	0	0	13	0	0	488	
		2,3	·	·	·	16	·	·	783	·
					total	259	9'060	8'507	9'825	1.0658
3	1126	4	1	100	1	7	66	66	56	7.1160
		4	1	$\neq 100$	0	892	4'295	0	3'960	7.1160
		1	0	0	0	84	0	0	201	7.1160
		2,3	·	·	·	37	·	·	125	·
					total	1'020	4'361	66	4'342	
4	1227	4	1	100	1	23	5'585	5'585	5'390	1
		4	1	$\neq 100$	0	1	264	0	229	1
		1	0	0	0	0	0	0	0	0
		2,3	·	·	·	0	·	·	0	·
					total	24	5'849	5'585	5'619	

h	N_h	r_h	$m_{h,d}$	w_h	$\sum_{k\in V_{h,d}} y_k$	X_{F_h}	X_{R_h}	$X_{F_h}/\hat{X}_{F_h}$	$S^2_{y[U_d],R_h}$	$S^2_{\hat{z}_h[U_d],R_h}$
1	17	16	1	1.0625	1	0	0	·	0.06	·
2	259	243	196	1.0658	8'507	9'825	9'042	1.01947	8'019.32	115.45
3	6'995	983	7	7.1160	66	30'870	4'217	1.02872	1.47	0.17
4	24	24	23	1.0000	5'585	5'619	5'619	1.00000	21'720.65	21'720.65

Berechnung der interessierenden Schätzer im Untersuchungsbereich NOGA 100.

$$\hat{Y}_d = \sum_h \hat{Y}_{h,d} = \sum_h \frac{N_h}{r_h} \sum_{k\in V_{h,d}} y_k$$

$$= \tfrac{17}{16}\cdot 1 + \tfrac{259}{243}\cdot 8'507 + \tfrac{6'995}{983}\cdot 66 + \tfrac{24}{24}\cdot 5'585 = \underline{15'123}$$

$$\widehat{\mathbf{Var}}(\hat{Y}_d) = \sum_h \widehat{\mathbf{Var}}(\hat{Y}_{h,d}) = \sum_h N_h^2 \left(1-\frac{r_h}{N_h}\right)\frac{1}{r_h} S^2_{y[U_d],R_h}$$

$$= 17^2\cdot(1-\tfrac{16}{17})\cdot\tfrac{1}{16}\cdot 0.06 + \ldots + 24^2\cdot(1-\tfrac{24}{24})\cdot\tfrac{1}{24}\cdot 21'720.65$$

$$= 0.07 + 136'757 + 63'063.91 + 0 = 199'817.9 = \underline{(447.01)^2}$$

$$\hat{Y}_{KM,d} = \sum_{h\in\{1\}} \frac{N_h}{r_h} \sum_{k\in V_{h,d}} y_k \quad + \quad \sum_{h\in\{2,3,4\}} \frac{X_{F_h}}{\hat{X}_{F_h}} \frac{N_h}{r_h} \sum_{k\in V_{h,d}} y_k$$

$$= \tfrac{17}{16}\cdot 1 + 1.01947\cdot 8'507 + 1.02872\cdot 66 + 1\cdot 5'585 = \underline{15'313}$$

$$\widehat{\mathbf{Var}}(\hat{Y}_{KM,d}) = \sum_{h\in\{1\}} N_h^2 \left(1-\frac{r_h}{N_h}\right)\frac{1}{r_h} S^2_{y[U_d],R_h} \quad + \sum_{h\in\{2,3,4\}} N_h^2 \left(1-\frac{r_h}{N_h}\right)\frac{1}{r_h} S^2_{\hat{z}_h[U_d],R_h}$$

$$= 0.07 + 1'968.83 + 7'312.72 + 0 = 9'281.4 = \underline{(96.23)^2}$$

$$\hat{M}_d = \sum_h \hat{M}_{h,d} = \sum_h \frac{N_h}{r_h} m_{h,d}$$

$$= \tfrac{17}{16}\cdot 1 + \tfrac{259}{243}\cdot 196 + \tfrac{6'995}{983}\cdot 7 + \tfrac{24}{24}\cdot 23 = \underline{283}$$

$$\widehat{\mathbf{Var}}(\hat{M}_d) = \sum_h \widehat{\mathbf{Var}}(\hat{M}_{h,d}) = \sum_h N_h^2 \left(1-\frac{r_h}{N_h}\right)\frac{1}{r_h}\frac{m_{h,d}}{r_h-1}\left(1-\frac{m_{h,d}}{r_h}\right)$$

$$= 0.066 + 2.671 + 302.786 + 0 = 305.52 = \underline{(17.479)^2}$$

5 Resultate

Anhand der beschriebenen Methoden wurden die Resultate der provisorischen Ergebnisse mit den Daten der BZ-Datenbank am 15. April berechnet. Die Resultate wurden am 3. Mai 2002 veröffentlicht (siehe [1]).
Es wurden diverse Tabellen und Grafiken publiziert. In den Tabellen 9 bis 12 sind ein paar der wichtigsten Resultate aufgelistet.

Tabelle 9: Arbeitsstätten und Beschäftigte, Total und nach Wirtschaftssektor 2001

Sektor	Arbeitsstätten	CV	Total Beschäftigte	CV	Vollzeit-beschäftigte	CV	Teilzeit-beschäftigte	CV	Vollzeit-äquivalente	CV
Total	**389'600**	**0.3**	**3'641'800**	**0.5**	**2'645'700**	**0.5**	**996'100**	**1.1**	**3'117'000**	**0.5**
2. Sektor	83'700	0.7	1'043'600	0.4	925'500	0.4	118'100	1.1	983'700	0.4
3. Sektor	305'900	0.3	2'598'200	0.7	1'720'200	0.7	878'000	1.1	2'133'400	0.7

Betriebszählung 2001, provisorische Ergebnisse
CV = Variationskoeffizient in %

Tabelle 10: Arbeitsstätten und Beschäftigte nach Grossregion 2001

Grossregion	Arbeitsstätten	CV	Total Beschäftigte	CV	Vollzeit-beschäftigte	CV	Teilzeit-beschäftigte	CV	Vollzeit-äquivalente	CV
Total	**389'600**		**3'641'800**		**2'645'700**		**996'100**		**3'117'100**	
Genferseeregion	70'900	1.7	642'100	1.0	464'100	0.9	178'000	2.6	551'700	0.7
Espace Mittelland	81'800	1.5	778'900	1.7	553'700	1.6	225'200	2.7	660'600	1.6
Nordwestschweiz	50'700	2.1	504'400	0.9	364'700	0.9	139'700	2.1	431'000	0.7
Zürich	69'200	1.7	723'100	0.8	518'800	0.8	204'300	2.1	614'500	0.7
Ostschweiz	60'600	1.9	503'500	0.9	371'500	0.9	132'000	2.3	432'400	0.7
Zentralschweiz	38'200	2.4	338'300	2.2	249'800	2.2	88'500	2.8	290'200	2.2
Tessin	18'200	3.6	151'500	1.6	123'100	1.7	(28'400)	4.2	136'700	1.4

Betriebszählung 2001, provisorische Ergebnisse
CV = Variationskoeffizient in %
Angaben in Klammern: Statistisch nur bedingt zuverlässig (CV > 4%)

Tabelle 11: Arbeitsstätten und Beschäftigte nach Grössenklasse 2001

Grössenklasse	Arbeitsstätten	CV	Total Beschäftigte	CV	Vollzeit-beschäftigte	CV	Teilzeit-beschäftigte	CV	Vollzeit-äquivalente	CV
Total	**389'600**		**3'641'800**		**2'645'700**		**996'100**		**3'117'100**	
0 - 9	332'800	0.4	1'102'400	0.6	737'500	0.6	364'900	1.3	902'200	0.5
10 - 49	47'600	2.0	1'048'500	0.8	769'100	0.8	279'400	2.0	901'300	0.6
50 - 249	8'300	3.1	920'800	1.7	684'700	1.5	236'100	2.8	799'500	1.6
250 und mehr	900	0.9	570'100	0.4	454'400	0.3	115'700	1.1	514'100	0.4

Betriebszählung 2001, provisorische Ergebnisse
CV = Variationskoeffizient in %

Tabelle 12: Arbeitsstätten und Beschäftigte nach NOGA-Unterabschnitt 2001

Wirtschaftszweig Unterabs.	NOGA2	Arbeitsstätten	CV	Total Beschäftigte	CV	Vollzeit-beschäftigte	CV	Teilzeit-beschäftigte	CV	Vollzeit-äquivalente	CV
Total		**389'600**		**3'641'800**		**2'645'700**		**996'100**		**3'117'100**	
2. Sektor		**83'700**		**1'043'600**		**925'500**		**118'100**		**983'700**	
C	10,11,12,13,14	400	3.3	5'100	1.6	4'700	1.7	400	3.9	4'900	1.7
DA	15,16	(3'600)	6.0	70'500	0.9	57'400	1.1	(13'100)	4.2	63'900	0.9
DB	17,18	(1'800)	6.1	21'600	0.9	17'400	0.9	4'200	3.4	19'500	0.8
DC	19	(300)	8.5	2'500	0.5	2'000	0.5	500	1.7	2'300	0.5
DD	20	7'000	2.4	42'200	1.7	38'300	1.8	(3'900)	4.2	40'100	1.7
DE	21,22	5'000	2.0	67'900	0.6	54'500	0.7	13'400	1.6	61'200	0.5
DF,DG	23,24	(1'300)	7.9	66'600	0.6	59'200	0.6	7'400	1.7	63'400	0.6
DH	25	(1'000)	7.6	26'500	0.8	23'700	0.8	2'800	3.2	25'200	0.8
DI	26	(1'700)	5.8	19'400	1.6	17'500	1.7	(1'900)	7.0	18'400	1.5
DJ	27,28	8'700	1.8	105'900	1.0	94'300	1.1	11'600	2.4	100'100	1.1
DK	29	3'900	3.9	106'400	0.8	97'600	0.8	8'800	2.1	102'200	0.6
DL	30,31,32,33	5'400	2.6	136'300	0.4	120'500	0.4	15'800	1.4	129'000	0.4
DM	34,35	(700)	10.2	22'200	0.9	20'800	0.9	1'400	3.8	21'500	0.9
DN	36,37	4'700	3.6	32'600	1.5	27'700	1.1	(4'900)	8.2	30'200	1.2
E	40,41	1'000	2.6	23'100	1.9	21'100	2.0	(2'000)	5.2	22'200	2.0
F	45	37'200	1.2	294'800	1.3	268'800	1.4	26'000	4.0	279'600	1.1
3. Sektor		**305'900**		**2'598'200**		**1'720'200**		**878'000**		**2'133'400**	
G	50,51,52	89'600	0.8	589'600	0.9	420'900	1.0	168'700	2.2	503'000	0.5
H	55	28'600	1.2	224'700	1.9	159'300	1.8	(65'400)	4.9	185'500	1.6
I	60,61,62,63,64	18'500	0.9	237'200	1.3	179'400	1.3	57'800	3.0	205'100	1.2
J	65,66,67	11'200	1.8	201'200	0.9	165'200	0.9	36'000	2.2	183'600	0.8
K	70,71,72,73,74	77'900	0.8	425'600	1.3	292'100	1.1	133'500	2.8	346'800	1.0
L	75	9'300	1.7	(142'200)	4.8	(109'300)	5.0	(32'900)	6.7	(126'600)	4.8
M	80	15'700	1.3	(229'100)	4.9	(104'200)	5.9	(124'900)	4.6	(164'000)	5.3
N	85	25'800	1.0	402'100	2.1	199'600	3.1	202'500	2.0	303'000	2.4
O	90,91,92,93	29'300	1.0	146'500	1.4	90'200	1.3	56'300	2.8	115'800	1.1

Betriebszählung 2001, provisorische Ergebnisse
CV = Variationskoeffizient in %
Angaben in Klammern: Statistisch nur bedingt zuverlässig (CV > 4%)

Anhang

A Stichprobenplan je Schicht

Tabelle 13: Stichprobenplan (1) BZ 2001

SCHICHT	NOGA2	domain sek	Betot Flag	VZAeq Flag	N_h	n_h	f_h
1	10	1	1	0	2	2	1.00
2	10	3	1	0	1	1	1.00
3	11	1	0	0	2	2	1.00
4	11	2	1	0	1	1	1.00
5	14	1	0	0	38	38	1.00
6	14	1	1	0	192	192	1.00
7	14	1	1	1	26	26	1.00
8	14	2	0	0	22	22	1.00
9	14	2	1	0	143	65	0.45
10	14	2	1	1	10	10	1.00
11	14	3	1	0	14	14	1.00
12	15	1	0	0	278	119	0.43
13	15	1	1	0	2'187	331	0.15
14	15	1	1	1	222	222	1.00
15	15	2	0	0	24	24	1.00
16	15	2	1	0	547	93	0.17
17	15	2	1	1	69	69	1.00
18	15	3	1	0	17	17	1.00
19	16	1	0	0	2	2	1.00
20	16	1	1	0	7	7	1.00
21	16	2	1	0	5	5	1.00
22	16	2	1	1	7	7	1.00
23	17	1	0	0	64	64	1.00
24	17	1	1	0	617	237	0.38
25	17	1	1	1	59	59	1.00
26	17	2	0	0	9	9	1.00
27	17	2	1	0	105	86	0.82
28	17	3	1	0	1	1	1.00
29	18	1	0	0	98	98	1.00
30	18	1	1	0	892	612	0.69
31	18	1	1	1	40	40	1.00
32	18	2	0	0	10	10	1.00
33	18	2	1	0	57	57	1.00
34	18	3	1	0	3	3	1.00
35	19	1	0	0	23	23	1.00
36	19	1	1	0	246	246	1.00
37	19	1	1	1	22	22	1.00
38	19	2	1	0	10	10	1.00
39	20	1	0	0	471	315	0.67
40	20	1	1	0	6'403	1167	0.18
41	20	1	1	1	220	220	1.00
42	20	2	0	0	72	72	1.00
43	20	2	1	0	251	73	0.29
44	20	2	1	1	49	49	1.00
45	20	3	0	0	2	2	1.00
46	20	3	1	0	12	12	1.00
47	21	1	0	0	19	19	1.00
48	21	1	1	0	171	171	1.00
49	21	1	1	1	32	32	1.00
50	21	2	0	0	4	4	1.00
51	21	2	1	0	47	47	1.00
52	22	1	0	0	508	313	0.62
53	22	1	1	0	4'373	930	0.21
54	22	1	1	1	297	297	1.00
55	22	2	0	0	46	46	1.00
56	22	2	1	0	486	81	0.17
57	22	2	1	1	69	69	1.00
58	22	3	1	0	7	7	1.00
59	23	1	1	0	7	7	1.00
60	23	2	1	0	1	1	1.00

Tabelle 14: Stichprobenplan (2) BZ 2001

SCHICHT	NOGA2	domain sek	Betot Flag	VZAeq Flag	N_h	n_h	f_h
61	24	1	0	0	138	117	0.85
62	24	1	1	0	640	364	0.57
63	24	1	1	1	101	101	1.00
64	24	2	0	0	20	20	1.00
65	24	2	1	0	127	74	0.58
66	24	2	1	1	57	57	1.00
67	24	3	1	0	2	2	1.00
68	25	1	0	0	45	45	1.00
69	25	1	1	0	668	374	0.56
70	25	1	1	1	81	81	1.00
71	25	2	0	0	17	17	1.00
72	25	2	1	0	121	75	0.62
73	25	2	1	1	14	14	1.00
74	26	1	0	0	106	80	0.75
75	26	1	1	0	1'106	301	0.27
76	26	1	1	1	108	108	1.00
77	26	2	0	0	47	47	1.00
78	26	2	1	0	325	88	0.27
79	26	2	1	1	19	19	1.00
80	26	3	1	0	1	1	1.00
81	27	1	0	0	13	13	1.00
82	27	1	1	0	213	213	1.00
83	27	1	1	1	24	24	1.00
84	27	2	0	0	4	4	1.00
85	27	2	1	0	46	46	1.00
86	28	1	0	0	577	330	0.57
87	28	1	1	0	6'995	1020	0.15
88	28	1	1	1	701	701	1.00
89	28	2	0	0	61	61	1.00
90	28	2	1	0	431	72	0.17
91	28	2	1	1	35	35	1.00
92	28	3	1	0	5	5	1.00
93	29	1	0	0	305	226	0.74
94	29	1	1	0	3'089	957	0.31
95	29	1	1	1	252	252	1.00
96	29	2	0	0	57	57	1.00
97	29	2	1	0	438	96	0.22
98	29	2	1	1	93	93	1.00
99	30	1	0	0	27	27	1.00
100	30	1	1	0	102	102	1.00
101	30	1	1	1	20	20	1.00
102	30	2	0	0	3	3	1.00
103	30	2	1	0	21	21	1.00
104	31	1	0	0	103	94	0.91
105	31	1	1	0	880	492	0.56
106	31	1	1	1	71	71	1.00
107	31	2	0	0	23	23	1.00
108	31	2	1	0	132	82	0.62
109	31	2	1	1	43	43	1.00
110	32	1	0	0	89	89	1.00
111	32	1	1	0	504	504	1.00
112	32	1	1	1	44	44	1.00
113	32	2	0	0	17	17	1.00
114	32	2	1	0	59	59	1.00
115	33	1	0	0	302	210	0.70
116	33	1	1	0	2'858	665	0.23
117	33	1	1	1	362	362	1.00
118	33	2	0	0	27	27	1.00
119	33	2	1	0	246	66	0.27
120	33	2	1	1	60	60	1.00
121	34	1	0	0	14	14	1.00
122	34	1	1	0	151	151	1.00
123	34	1	1	1	13	13	1.00
124	34	2	0	0	4	4	1.00
125	34	2	1	0	18	18	1.00
126	35	1	0	0	52	52	1.00
127	35	1	1	0	358	358	1.00
128	35	1	1	1	31	31	1.00
129	35	2	0	0	7	7	1.00
130	35	2	1	0	69	69	1.00
131	35	3	1	0	1	1	1.00
132	36	1	0	0	337	261	0.77
133	36	1	1	0	3'667	1127	0.31
134	36	1	1	1	177	177	1.00
135	36	2	0	0	42	42	1.00
136	36	2	1	0	242	78	0.32
137	36	2	1	1	27	27	1.00
138	36	3	1	0	1	1	1.00

Tabelle 15: Stichprobenplan (3) BZ 2001

SCHICHT	NOGA2	domain sek	Betot Flag	VZAeq Flag	N_h	n_h	f_h
139	37	1	0	0	27	27	1.00
140	37	1	1	0	249	249	1.00
141	37	1	1	1	35	35	1.00
142	37	2	0	0	3	3	1.00
143	37	2	1	0	39	39	1.00
144	37	3	1	0	2	2	1.00
145	40	1	0	0	46	46	1.00
146	40	1	1	0	153	72	0.47
147	40	1	1	1	13	13	1.00
148	40	2	0	0	32	32	1.00
149	40	2	1	0	244	86	0.35
150	40	2	1	1	20	20	1.00
151	40	3	0	0	14	14	1.00
152	40	3	1	0	281	76	0.27
153	40	3	1	1	35	35	1.00
154	41	1	0	0	9	9	1.00
155	41	1	1	0	36	36	1.00
156	41	2	1	0	3	3	1.00
157	41	3	0	0	6	6	1.00
158	41	3	1	0	237	76	0.32
159	41	3	1	1	10	10	1.00
160	45	1	0	0	3'544	626	0.18
161	45	1	1	0	32'093	1035	0.03
162	45	1	1	1	1'683	1683	1.00
163	45	2	0	0	772	80	0.10
164	45	2	1	0	3'863	104	0.03
165	45	2	1	1	129	129	1.00
166	45	3	0	0	5	5	1.00
167	45	3	1	0	887	83	0.09
168	45	3	1	1	24	24	1.00
169	50	1	0	0	1'237	466	0.38
170	50	1	1	0	14'525	1088	0.07
171	50	1	1	1	269	269	1.00
172	50	2	0	0	147	82	0.56
173	50	2	1	0	1'188	93	0.08
174	50	2	1	1	76	76	1.00
175	50	3	0	0	2	2	1.00
176	50	3	1	0	27	27	1.00
177	51	1	0	0	4'311	601	0.14
178	51	1	1	0	20'517	1004	0.05
179	51	1	1	1	861	861	1.00
180	51	2	0	0	465	75	0.16
181	51	2	1	0	3'691	103	0.03
182	51	2	1	1	145	145	1.00
183	51	3	1	0	2	2	1.00
184	52	1	0	0	5'175	633	0.12
185	52	1	1	0	39'339	859	0.02
186	52	1	1	1	254	254	1.00
187	52	2	0	0	1'185	214	0.18
188	52	2	1	0	15'465	373	0.02
189	52	2	1	1	452	452	1.00
190	52	3	1	0	13	13	1.00
191	55	1	0	0	3'081	455	0.15
192	55	1	1	0	24'215	746	0.03
193	55	1	1	1	847	847	1.00
194	55	2	0	0	603	83	0.14
195	55	2	1	0	3'862	87	0.02
196	55	2	1	1	112	112	1.00
197	55	3	0	0	17	17	1.00
198	55	3	1	0	113	82	0.73
199	60	1	0	0	1'099	383	0.35
200	60	1	1	0	7'108	824	0.12
201	60	1	1	1	282	282	1.00
202	60	2	0	0	170	90	0.53
203	60	2	1	0	1'289	106	0.08
204	60	2	1	1	87	87	1.00
205	60	3	1	0	66	66	1.00
206	61	1	0	0	22	22	1.00
207	61	1	1	0	78	78	1.00
208	61	2	0	0	6	6	1.00
209	61	2	1	0	20	20	1.00
210	61	3	1	0	2	2	1.00
211	62	1	0	0	39	39	1.00
212	62	1	1	0	98	98	1.00
213	62	2	0	0	4	4	1.00
214	62	2	1	0	106	106	1.00
215	62	2	1	1	8	8	1.00

Tabelle 16: Stichprobenplan (4) BZ 2001

SCHICHT	NOGA2	domain sek	Betot Flag	VZAeq Flag	N_h	n_h	f_h
216	63	1	0	0	580	209	0.36
217	63	1	1	0	2'642	521	0.20
218	63	1	1	1	94	94	1.00
219	63	2	0	0	187	88	0.47
220	63	2	1	0	1'598	113	0.07
221	63	2	1	1	101	101	1.00
222	63	3	0	0	3	3	1.00
223	63	3	1	0	77	77	1.00
224	64	1	0	0	248	74	0.30
225	64	1	1	0	646	92	0.14
226	64	1	1	1	46	46	1.00
227	64	2	0	0	60	60	1.00
228	64	2	1	0	482	101	0.21
229	64	2	1	1	71	71	1.00
230	64	3	0	0	27	27	1.00
231	64	3	1	0	3'548	208	0.06
232	64	3	1	1	131	131	1.00
233	65	1	0	0	297	79	0.27
234	65	1	1	0	1'267	122	0.10
235	65	1	1	1	109	109	1.00
236	65	2	0	0	147	77	0.52
237	65	2	1	0	2'762	239	0.09
238	65	2	1	1	207	207	1.00
239	65	3	0	0	9	9	1.00
240	65	3	1	0	451	77	0.17
241	65	3	1	1	39	39	1.00
242	66	1	0	0	96	80	0.83
243	66	1	1	0	660	173	0.26
244	66	1	1	1	103	103	1.00
245	66	2	0	0	145	84	0.58
246	66	2	1	0	2'382	129	0.05
247	66	2	1	1	111	111	1.00
248	66	3	0	0	1	1	1.00
249	66	3	1	0	74	74	1.00
250	66	3	1	1	6	6	1.00
251	67	1	0	0	610	377	0.62
252	67	1	1	0	2'097	798	0.38
253	67	1	1	1	45	45	1.00
254	67	2	0	0	52	52	1.00
255	67	2	1	0	185	78	0.42
256	67	2	1	1	32	32	1.00
257	67	3	0	0	1	1	1.00
258	67	3	1	0	5	5	1.00
259	70	1	0	0	1'257	584	0.46
260	70	1	1	0	4'095	1137	0.28
261	70	1	1	1	66	66	1.00
262	70	2	0	0	60	60	1.00
263	70	2	1	0	318	73	0.23
264	70	2	1	1	37	37	1.00
265	70	3	0	0	13	13	1.00
266	70	3	1	0	34	34	1.00
267	71	1	0	0	153	127	0.83
268	71	1	1	0	779	387	0.50
269	71	1	1	1	2	2	1.00
270	71	2	0	0	30	30	1.00
271	71	2	1	0	183	77	0.42
272	71	2	1	1	11	11	1.00
273	71	3	0	0	1	1	1.00
274	71	3	1	0	1	1	1.00
275	72	1	0	0	3'099	797	0.26
276	72	1	1	0	9'727	1272	0.13
277	72	1	1	1	153	153	1.00
278	72	2	0	0	152	72	0.47
279	72	2	1	0	626	85	0.14
280	72	2	1	1	81	81	1.00
281	72	3	1	0	11	11	1.00

Tabelle 17: Stichprobenplan (5) BZ 2001

SCHICHT	NOGA2	domain sek	Betot Flag	VZAeq Flag	N_h	n_h	f_h
282	73	1	0	0	139	139	1.00
283	73	1	1	0	401	401	1.00
284	73	1	1	1	40	40	1.00
285	73	2	0	0	9	9	1.00
286	73	2	1	0	61	61	1.00
287	73	3	0	0	3	3	1.00
288	73	3	1	0	50	50	1.00
289	74	1	0	0	9'179	1014	0.11
290	74	1	1	0	56'433	1580	0.03
291	74	1	1	1	643	643	1.00
292	74	2	0	0	849	73	0.09
293	74	2	1	0	4'959	130	0.03
294	74	2	1	1	209	209	1.00
295	74	3	0	0	14	14	1.00
296	74	3	1	0	228	83	0.36
297	74	3	1	1	15	15	1.00
298	75	1	0	0	11	11	1.00
299	75	1	1	0	3	3	1.00
300	75	3	0	0	540	117	0.22
301	75	3	1	0	8'722	225	0.03
302	75	3	1	1	428	428	1.00
303	80	1	0	0	631	166	0.26
304	80	1	1	0	4'442	309	0.07
305	80	1	1	1	252	252	1.00
306	80	2	0	0	120	86	0.72
307	80	2	1	0	702	78	0.11
308	80	2	1	1	59	59	1.00
309	80	3	0	0	474	101	0.21
310	80	3	1	0	10'259	232	0.02
311	80	3	1	1	307	307	1.00
312	85	1	0	0	1'191	398	0.33
313	85	1	1	0	20'542	878	0.04
314	85	1	1	1	953	953	1.00
315	85	2	0	0	294	81	0.28
316	85	2	1	0	2'560	83	0.03
317	85	2	1	1	100	100	1.00
318	85	3	0	0	187	88	0.47
319	85	3	1	0	2'321	72	0.03
320	85	3	1	1	153	153	1.00
321	90	1	0	0	40	40	1.00
322	90	1	1	0	404	218	0.54
323	90	1	1	1	44	44	1.00
324	90	2	0	0	16	16	1.00
325	90	2	1	0	59	59	1.00
326	90	3	0	0	47	47	1.00
327	90	3	1	0	958	90	0.09
328	90	3	1	1	98	98	1.00
329	91	1	0	0	544	242	0.44
330	91	1	1	0	2'408	507	0.21
331	91	1	1	1	72	72	1.00
332	91	2	0	0	75	75	1.00
333	91	2	1	0	1'330	71	0.05
334	91	2	1	1	64	64	1.00
335	91	3	0	0	34	34	1.00
336	91	3	1	0	2'563	89	0.03
337	91	3	1	1	33	33	1.00
338	92	1	0	0	1'443	658	0.46
339	92	1	1	0	6'646	1317	0.20
340	92	1	1	1	141	141	1.00
341	92	2	0	0	108	78	0.72
342	92	2	1	0	752	93	0.12
343	92	2	1	1	70	70	1.00
344	92	3	0	0	52	52	1.00
345	92	3	1	0	809	89	0.11
346	92	3	1	1	37	37	1.00
347	93	1	0	0	1'208	488	0.40
348	93	1	1	0	14'511	1202	0.08
349	93	1	1	1	39	39	1.00
350	93	2	0	0	111	76	0.68
351	93	2	1	0	1'135	83	0.07
352	93	2	1	1	31	31	1.00
353	93	3	0	0	9	9	1.00
354	93	3	1	0	98	75	0.77
355	93	3	1	1	19	19	1.00

B Verwendete Daten, Tabellen und Programme

Die Tabellen 18 und 19 beschreiben alle Files (Daten, Tabellen, Programme), die für das Projekt BZ01 verwendet wurden.

Tabelle 18: Übersicht der verwendeten Files (1)

Filename	Beschreibung
BeschreibungFilesBz 2001.doc	Beschreibung aller Files.
Stichprobenrahmen.sas, StichprobenrahmenUnix.sas	Kreiert aus den (ursprünglichen) Oracle-Tabellen ein Sasfile (Bzflatfile.sas7bdat) mit den wichtigsten Variablen (Stichprobenrahmen Stufe 1)
Bzflatfile.sas7bdat	Basisfile für alle Berechnungen
Bestaneu.sas7bdat	Enthält alle Bernoullizahlen der Besta in einem File. Dies ist ein merge der Tabellen Besta.sas7bdat, vaud.sas7bdat, stgall.sas7bdat, neuch.sas7bdat, genf.sas7bdat.
Base2000Besta.sas	Erstellt das obige File Bestaneu.sas7bdat
Berechnungen.sas (1.Teil)	Verarbeitet das File Bzflatfile.sas7bdat und erstellt (den definitiven Stichprobenrahmen) das Export File alle.csv, das ins Splus gelesen wird.
alle.csv /Alle.xls	Ausgangsfile im Splus, für die Berechnung des Stichprobenplans
BerechnungNHs.ssc/ BerechnungNHsSplus.txt	Berechnet mit dem File alle.xls den (provisorischen, d.h. ohne Bernoullizahlen.) Stichprobenplan. Resultat: 4 Exportfiles: Stplan.sd2, StplanAlle2%.xls, StplanproNoga2%.xls, StplanAllelang2%.xls
Berechnungen.sas (2.Teil)	Mergt die Files Stplan.sas7bdat, Bestaneu.sas7bdat mit dem Bzflatfile. Berechnung des ursprünglich definitiven Stichprobenplans (mit Bernoulliziehung). (Bemerkung: zu ursprünglich. siehe unten BerechnungenNEU.sas), Export: folgende 6 Files werden kreiert:
BzflatfileEnd.sas7bdat	Neues Flatfile (mit allen Variablen, inkl. Sampleflg)
StplanproSchicht.sas7bdat StplanproSchicht.xls	Stichprobenplan pro Schicht (analog dem File StplanAllelang2%.xls, aber mit neuen nh, alt!!)
sampleBurnr2.sas7bdat sampleBurnr2.txt	Alle Burnummern mit der Angabe ob die AST in der Stichprobe ist oder nicht. (Alte Stichprobe!)
sampleBurnr.sas7bdat sampleBurnr.txt	Alle Burnr. der Stichprobe (Alte Stichprobe!)
Sampleburworkfl.sas7bdat	Enthält den Sampleflg und den (ursprünglichen) Workflow_status_code (Alte Stichprobe!)
Flagrahmen.sas7bdat	Enthält den Flag ob AST ursprünglich im Stichprobenrahmen ist.
BerechnungenNEU.sas (1.Teil)	Da die (ursprünglich) gewählte Stichprobe (99'193) zu gross ist, erstellt dieses File die neue Stichprobe
BerechnungNHsNEU2.ssc/	Dieses (im Vergleich zum BerechnungNHs.ssc leicht abgeänderte) Splus- Programm berechnet nochmals die Aufteilung der Schichten. Resultat: SAS-Exportfiles: stplan8sd2 stplan25.sd2
BerechnungenNEU.sas (2.Teil)	Fortsetzung vom 1 Teil. Zusätzlich werden noch die Genauigkeiten neu berechnet. Die 6 Files werden neu erstellt! (BzflatfileEnd, StplanproSchicht, sampleBurnr2, sampleBurnr, Sampleburworkfl, Flagrahmen).

Tabelle 19: Übersicht der verwendeten Files (2)

Filename	Beschreibung
SampleBurWorkfl.sas7bdat	Dieses File enthält neu beide Sampleflags! (auch noch den der alten Stichprobe.
StplanproSchicht.sas7bdat StplanproSchicht.xls	Enthält den definitiven Stichprobenplan pro Schicht mit den zu erwartenden Genauigkeiten.
PlausOK.txt	Enthält AST, die eigentlich i.O. sind für die Hochrechnung, aber diese sind in DB noch nicht so gekenntzeichnet sind. Diese werden mit dem zusPlausFlg=1 gekenntzeichnet
keyneu.sas7bdat	enthält neue Schicht-Variablen (schicht_neu), und der neuen Populations+Stichprobengrössen je Schicht:Idee vergrössern der einzelnen Schichten
GenauigkeitproNoga.xls	Enthält den definitiven Stichprobenplan pro NOGA2
GenauigkeitproNogakurz.xls	mit den zu erwartenden Genauigkeiten.
Bindungen.xls NeueStichprobeBind.xls	Enthält Erklärung wieso alte Stichprobe im UNT/Oeff- Bereich zu gross war.
Präsentation.doc	Entalten die Folien mit dem Ablauf der Konstruktion des Stp.rahmens und -plans, die wir UNT präsentiert haben (aber neuere Version)
Ueberblick.doc/rtf	Vergleich der Stichprobengrössen/Anz.AST/fh pro Grossregion*Noga2 (erzeugt in berechnungen.sas)
Rücklauf.sas	Wird verwendet um einen BZ-DB Auszug mit dem aktuellsten Stand (während der BZ-Erhebung) zu kreieren. Dazu werden zusätzlich die 5 Files: plausok, keyneu, Sampleburworkfl, Flagrahmen und StplanproSchicht verwendet.
Checks.sas	Programm welches Tabellen erzeugt für den Überblick über den Stand des Rücklaufs
Hochrechung.sas	1.Programm für die Hochrechnung (für einfachen Schätzer, 1. Domainvar)
HochrDomain3-4.sas	2.Programm für die Hochr (für einfachen Schätzer, keine od. 2-3 Domainvar).
HochrechnungQuot.sas	3. Programm für die Hochr (einfacher+ Quotientenschätzer, 1. Domainvar).
HochrDomain3-4Quot.sas	4.Programm für die Hochr (einfacher+ Quotientenschätzer, keine od. 2-3 Domainvar).
Rücklauf.doc	Outputs von checks.sas zu gewissen Zeitpunkten
VerschResultate2002-04-15.xls	Prov. Hochrechnungsresultate
Vergleich***.xls	Vergleich der prov. Hochrechnungsresultate mit Besta
Bericht.tex	Bericht Plausiblisierungen
Hochrechung.tex	Bericht Stplan/Hochrechnung

C SAS-Programm zur Hochrechnung

```
/* Hochrechnungprogramme mit Verwendung der SAS-prozedur
"Proz surveymeans", möglich ab SAS-Version 8.2*/

/*es gibt vier Hochrechnungs Makros:
1. Hochrechnung.sas
(für einfacher Schätzer, 1. Domainvar)
2. HochrDomain3-4.sas
(für einfacher Schätzer, keine od. 2-3 Domainvar)
3. HochrechnungQuot.sas
(einfacher+ Quotientenschätzer, 1. Domainvar)
4. HochrDomain3-4Quot.sas
(einfacher+ Quotientenschätzer, keine od. 2-3 Domainvar)
*/
folgend der Code von 3.HochrechnungQuot.sas:

libname   bz01  '/home/jpt/BZ01/BZDB/AuszugApr1502';
options linesize=97  pagesize=65  number pageno=1;

%macro Hochrech1a(dataimput,Pfad,dataoutput,Variable,Domain);

data outp; set &dataimput (rename=(nogaimp=noga));
if ((noga2_bz in (1,2,5,95) and stat in (1,4,5,6))
or Gregion=. or stat=1)  then do;
&Variable=0;
&Domain=.;
end;
else if stat in (2,3) then do;
&Variable=.;
&Domain=.;
end;
else if (stat in (4,5,6) and (&Domain=. or &Variable=.))
then do;
&Variable=.;
&Domain=.;
end;

keep    &Variable
        Xr&Variable
        Xu&Variable
        Xrh&Variable
        yr&Variable
        mr&variable
        schicht_neu
        schicht_begin
        noga
        kanton
        empl_total_bur
        zussamp
        Gregion
        Wsektor_neu
        Unterab
        grklasse
        AnzAST
        stat
        BetotBurflg
        noga2
        _total_
        vzaequivburflg
        aktiv;
label noga='NOGA2_BZ';
run;

/*zählen der mh's (Antworten) pro Schicht
(Anzahl Ast, die geantwortet haben und NOGA vorhanden ist)*/

proc means data=outp n noprint;
class schicht_neu;
var &Variable;
output out= outp2   n(&Variable)=mh&Variable;
run;

data outp3; set outp2;
if _type_=1;
drop _type_ _freq_;
run;

/*Anmergen der Summen (Antworten) ans Ausgangsfile*/
data outp4;
merge outp outp3;
by schicht_neu;
if (&Variable eq .)  then wh=.;
else if stat=6 then wh=1;
else  wh=_total_ / mh&Variable;

/*Definition Quotientenschätzer*/
if (zussamp=1 or BetotBurflg=0 or xrh&variable=0) then g_factor=1;
  else g_factor =Xu&Variable /((_total_/mh&Variable)*xrh&Variable);
  if (zussamp=1 or BetotBurflg=0 or xrh&variable=0) then ratio=0;
  else ratio = yr&Variable / xrh&Variable;
  &Variable._sr  = &Variable * g_factor;
  &Variable._res = &Variable - ratio*Xr&Variable;
run;

options linesize=95  pagesize=120  pageno=1;
title "Hochrechnung der Variablen &Variable je Domäne &Domain";

/*Durchführung der Hochrechnung*/
/*Das File "bz01.totale" enthält den
Stichprobenplan mit den Variablen
1)_total_, die Populationsgrösse je Scicht (Nh)
2)Schicht_neu, die 264 zusammengelegten Schichtungsvariablen
*/

proc surveymeans data=outp4 total=bz01.totale  missing sum nobs std;
domain &domain;
strata schicht_neu;
var &Variable &Variable._sr &Variable._res;
weight wh;
ods output  domain=dataoutput1;
run;

/*Berechnung des 95% Vertrauensintervall
Verschönerung des zu exportierenden Outputs*/
data dataoutput2; set dataoutput1;
if &domain ne .;
&Variable._HT_Sum=      round(&Variable._Sum);
&Variable._HT_CV =      round(100*(&Variable._StdDev /
                              &Variable._Sum),0.1);
&Variable._HT_StdDev=   round(&Variable._StdDev,0.01);
&Variable._HT_LCl=      floor(max(0,&Variable._Sum -1.96*
                              &Variable._StdDev));
&Variable._HT_UCl=      ceil(&Variable._Sum +1.96*
                               &Variable._StdDev);
&Variable._SR_Sum=      round(&Variable._sr_Sum);
&Variable._SR_CV =      round(100*(&Variable._res_StdDev /
                              &Variable._sr_Sum),0.1);
&Variable._SR_StdDev=   round(&Variable._res_StdDev,0.01);
&Variable._SR_LCl=      floor(max(0,&Variable._sr_Sum -1.96*
                              &Variable._res_StdDev));
&Variable._SR_UCl=      ceil(&Variable._sr_Sum +1.96*
                              &Variable._res_StdDev);
format  &Variable._HT_CV &Variable._SR_CV 4.1;
run;
data dataoutput3; set dataoutput2;
keep
&domain
&Variable._HT_Sum
&Variable._HT_CV
&Variable._HT_StdDev
&Variable._HT_LCl
&Variable._HT_UCl
;
run;
data dataoutput4; set dataoutput2;
if 0<=&Variable._SR_CV<=1 then V='A';
if 1<&Variable._SR_CV<=2 then V='B';
if 2<&Variable._SR_CV<=4 then V='C';
if &Variable._SR_CV>4 then V='D';
keep
&domain
&Variable._SR_Sum
&Variable._SR_CV
&Variable._SR_StdDev
&Variable._SR_LCl
&Variable._SR_UCl
V
;
run;
data &dataoutput;
merge dataoutput3 dataoutput4;
by &domain;
run;
/*Zur Formatierung der Variablen Unterabschnitt*/
%include '/home/jpt/BZ01/BZDB/Mergen/Formate.txt';run;
data &dataoutput; set &dataoutput;
format unterab unterab. kanton kanton.;
run;
/*Export der Resultate*/
PROC EXPORT DATA=&dataoutput
          OUTFILE="&Pfad./&dataoutput..csv"
          DBMS=CSV Replace;
run;
%mend Hochrech1a;
/*Aufruf des Makros*/

%Hochrech1a(bz01.sample,/home/jpt/BZ01/BZDB/AuszugApr1502,Bt_Unterab,betot,noga);
```

Literatur

[1] *Betriebszählung 2001: Provisorische Ergebnisse*. Pressemitteilung Nr. 45, BFS, Neuchâtel, 2002.

[2] *Betriebszählung 2001: Detailkonzept, Séléction pour la BZ 2001*. BFS, Neuchâtel, 2001.

[3] Potterat, J., Salamin, P.-A.: *Betriebszählung 2001: Methoden für die Datenbereinigung*. BFS, Neuchâtel, 2002.

[4] *Betriebszählung 2001 Detailkonzept, Use-Case 23*. BFS, Neuchâtel, 2001.

[5] Särndal, C.E., Swensson, B. Wretman, J.: *Model Assisted Survey Sampling*. Springer, New York, 1992.

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodological reports of the Statistical Methods Unit of SFSO

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. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statistische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln:

Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

Diffusionsmittel	*Kontakt* *N° à composer*	*Moyen de diffusion*
Individuelle Auskünfte	032 713 60 11	Service de renseignements individuels
Das BFS im Internet	http://www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	032 713 60 11	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette)	032 713 60 60	Publications: information approfondie (certaines sont disponibles sur disquette)
Online-Datenbank	032 713 60 86	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das alle 2 Jahre nachgeführte **Publikationsverzeichnis.** Es ist auf dem Internet abzurufen oder kann gratis über Tel. 032 713 60 60 bezogen werden.

La **Liste des publications**, qui est mise à jour tous les deux ans, donne davantage de détails sur les divers moyens de diffusion. Elle est proposée sur Internet et peut être obtenue gratuitement en téléphonant au n° 032 713 60 60.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodological reports of the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicher zu stellen, und sie sollen andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

In diesem Bericht werden die verwendeten Methoden zur Berechnung der provisorischen Ergebnisse der Betriebszählung 2001 beschrieben.

Unter Berücksichtigung von erhofften Genauigkeiten der Resultate und dem gewünschten Publikationstermin wurde aus den an der Betriebszählung teilnehmenden Arbeitsstätten eine geschichtete Bernoulli-Stichprobe gezogen. Diese Betriebe der Stichprobe wurden wenn möglich während des ganzen Prozesses der Betriebszählung, wie insbesondere bei der Datenbereinigung, gegenüber den anderen Betrieben priorisiert. Mitte April 2002 war der Rücklauf der Stichprobe ausreichend gut, so dass die provisorischen Resultate berechnet werden konnten.

Die Methode zur Berechnung der Schätzer basierte auf einer Kombination eines einfachen Schätzers und eines Quotientenschätzers, für welche zugleich auch passende Varianzschätzer berechnet wurden.

Mit den verwendeten Methoden war es möglich, die Anzahl der Beschäftigten wie auch die Anzahl Arbeitsstätten in der Schweiz und für gewisse Untersuchungsbereiche mit genügender Genauigkeit zu schätzen.

Bestellnummer:
338-0011

Bestellungen:
032 713 60 60 Fax: 032 713 60 61

Preis:
gratis

ISBN 3-303-00248-7