Neuchâtel, Juni 2001

Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL

Jann Potterat und Beat Hulliger

Dienst Statistische Methoden
Service de méthodes statistiques
Servizio metodi statistici

Auskunft:

Jann Potterat

Bundesamt für Statistik
Dienst Statistische Methoden
Espace de l'Europe 10, CH-2010 Neuchâtel
Tel.: (+41 32) 71 36590
Fax: (+41 32) 71 36093
Bestellnummer: 338-0004
© BFS/OFS

Seit 1995 werden vom Bundesamt für Statistik im Rahmen der Erhebung PAUL auch Daten über die Einschnittmengen der Sägereien erhoben, aber nicht ausgewertet und veröffentlicht. Die Sektion Land- und Forstwirtschaft (AGR) hat den Dienst Statistische Methoden beauftragt, eine Schätzmethode zu entwickeln und die Genauigkeit der Schätzer abzuklären. Die Arbeiten wurden durch Jann Potterat und Beat Hulliger vom Dienst Statistische Methoden ausgeführt und von Arthur Zesiger von der Sektion AGR begleitet.

INHALTSVERZEICHNIS

1 EINFÜHRUNG 1

2 GRUNDGESAMTHEIT UND STICHPROBE 1

2.1 GRUNDGESAMTHEIT 1
2.2 STICHPROBENZIEHUNG 2
2.2.1 Allgemeines 2
2.2.2 Erneuerung der Stichprobe 2
2.2.3 Schichtung der Stichprobe 2
2.2.4 Struktur der Stichprobe der Sägerei-Erhebung 2
2.3 DATENLIEFERUNG UND –QUALITÄT 3
2.3.1 Datenlieferung 3
2.3.2 Datenqualität 3
2.4 ANALYSE DER MISSING VALUES 3
2.5 DATENBEREINIGUNG / IMPUTATION 4
2.5.1 Datenbereinigung 4
2.5.2 Imputation 4

3 METHODIK ZUR SCHÄTZUNG DES HOLZVOLUMENS 5

3.1 FORMELN 5
3.2 BEMERKUNGEN ZU DEN BERECHNUNGEN 6
3.3 VARIABLENLISTE 6
3.4 DARSTELLUNG DER DATEN 7

4 RESULTATE 8

4.1 ÜBERSICHT DER NETTOSTICHPROBENGRÖSSEN 8
4.2 HOCHRECHNUNGEN 9
4.2.1 Methode 1: 9
4.2.2 Methode 2: 10
4.2.3 Graphiken zu den Resultaten 11
4.3 DISKUSSION DER RESULTATE 12

5 ANHANG 13

5.1 TABELLEN 13
5.2 PROGRAMM FÜR SAS VERSION 8 14

1 Einführung

Die Sektion Produktion und Umsatz (PROD) des Bundesamts für Statistik (BFS) führt die Erhebung "Produktions-, Auftrags-, Umsatz- und Lagerindizes (PAUL) im sekundären Sektor" durch. Im Rahmen dieser quartalsweise stattfindenden Befragung, wird bei den Arbeitstätten (AST), die eine Wirtschaftsaktivität NOGA 20.1 (Säge-, Hobel- und Holzimprägnierwerke) besitzen, neben den erzielten Frankenbeträge in den obigen vier Kategorien zusätzlich noch die eingeschnittene Menge von Nadel- und Laub-Stammholz in m^3 nachgefragt.
Die Sektion Land- und Forstwirtschaft (AGR) beauftragte den Dienst Statistische Methoden (METH) anhand dieser seit 1995 vorliegenden Daten, eine Schätzung des jährlich in der Schweiz eingeschnittenen Holzvolumens zu liefern.
Die letzte statistische Erhebung ausschliesslich über das eingeschnittenen Holzvolumen in der Schweiz fand im Jahre 1996 statt. Es war eine Vollerhebung bei 731 Sägereien, welche insgesamt ein Volumen von 1'936'225 m^3 Rundholz eingeschnitten haben (davon Nadelholz 1'743'826 m^3, Laubholz 192'399 m^3). Diese und weitere Resultate sind in „Holzverarbeitung in der Schweiz 1996" [(1)] zu finden. Die Grundgesamtheit kann aber nicht direkt mit der Grundgesamtheit der NOGA 20.1 verglichen werden (siehe unten).

2 Grundgesamtheit und Stichprobe

2.1 Grundgesamtheit

Die eigentliche Zielpopulation wären alle AST, die Holz einsägen. Diese Population ist schwer zu fassen, da sie auch AST umfasst, die nur gelegentlich oder nur als Nebentätigkeit Holz sägen. Es wurde daher hier mit der Population der NOGA 20.10A als Zielpopulation gearbeitet.
Mit den Betriebszählungen 1995 und 1998 wurden die Anzahl der AST in diesen zwei Jahren in der NOGA 20.1 (Säge-, Hobel- und Holzimprägnierwerke) nach Grössenklassen eingeteilt. Das Kriterium für die Grössenklassen ist die Anzahl Mitarbeiter (Vollzeitäquivalente). Folgende Tabelle veranschaulicht diesen Sachverhalt:

Tabelle 1: Anzahl Arbeitstätten in den NOGA 20.1 gemäss BZ 95 und 98

Jahr	NOGA	Anzahl Arbeitsstätten nach Grössenklassen in der NOGA 20.1			
		0-19	20-49	50+	Summe
1995	20.10A	571	20	5	596
	20.10B	92	11	3	106
	20.1	**663**	**31**	**8**	**702**
1998	20.10A	556	19	4	579
	20.10B	80	8	1	89
	20.1	**636**	**27**	**5**	**668**

[(1)] Holzverarbeitung in der Schweiz 1996, A. Zesiger, BFS, A. Hurst, BUWAL, Neuchâtel, 1999

2.2 Stichprobenziehung

2.2.1 Allgemeines

Im Rahmen einer Revision, wurde die Stichprobe der Statistik PAUL 1995 neu gezogen. Als Grundlage für diese Stichprobenziehung diente das BUR, u.a. basierend auf der BZ91, mit der damals noch gültigen "Allgemeine Systematik der Wirtschaftszweige 1985 (ASWZ 85)" als Branchenkodierung (siehe Methodenbericht von R. Peters 1996: "Schéma de pondération des Indices PAUL").
Der Teil der Stichprobe, der für die Sägerei-Erhebung verwendet wird, wurde aus der ASWZ Klasse 2611 (Säge-, Hobel-, Furnier-, und Imprägnierwerke) gezogen. Mit der Einführung der NOGA 1995 musste also die Branchenzugehörigkeit jeder Firma neu definiert werden. Die ASWZ 2611 entspricht mit wenigen Ausnahmen der NOGA 20.1, so wurde bei den Berechnungen neben der NOGA 20.10A zusätzlich noch die NOGA 20.1 als Zielpopulation verwendet

2.2.2 Erneuerung der Stichprobe

Die Stichprobe wurde seit 1995 nicht mehr erneuert. Ausserdem werden den AST, die über längere Zeit nicht mehr antworten, keine Fragebögen mehr verschickt!

2.2.3 Schichtung der Stichprobe

Nach dem Kriterium der Anzahl Beschäftigten (Vollzeitbeschäftigte) wurden drei Schichten (Nachschichtungen) definiert:

Tabelle 2: Schichtungen

Schicht	Anzahl Beschäftigte
0	50+
1	20-49
2	0-19

2.2.4 Struktur der Stichprobe der Sägerei-Erhebung

Folgende Tabelle zeigt die Struktur der (Brutto-) Stichprobe zum Zeitpunkt der Ziehung 1995.

Tabelle 3: Struktur der Stichprobe

NOGA	Grössenklasse			
	0-19	20-49	50+	Summe
20.10A	55	8	3	66
20.10B	4	3	1	8
20.1	59	11	4	74
K. Angabe [(2)]	45	5	1	51
Andere[(3)]	7	0	3	10
Total	111	16	8	135

Die Erklärungen zu [(2)] und zu [(3)] folgen im nächsten Abschnitt.

Bemerkung:
Aufgrund welcher Kriterien genau die Auswahl der Einheiten aus der Gesamterhebung PAUL erfolgte, konnte leider nicht definitiv geklärt werden. Vermutlich umfassen die gelieferten Daten alle Firmen, die irgendeinmal Holz eingesägt bzw. einen Fragebogen erhalten haben.

2.3 Datenlieferung und –qualität

2.3.1 Datenlieferung

Die von PROD gelieferten Daten beinhalten u.a. folgende Variablen zu den obigen 135 Arbeitsstätten:

- Firmennummer
- Jahr und Quartal
- Produkt 2 = 1. Analysevariable (Einschnitt von Nadel-Stammholz in m^3)
- Produkt 5 = 2. Analysevariable (Einschnitt von Laub-Stammholz in m^3)
- Anzahl Beschäftigte

2.3.2 Datenqualität

Zur Qualität, der von PROD gelieferten Daten, gibt es folgende Punkte zu bemerken:

- **Firmen- / BUR-Nummer** [(2)]

Jede Arbeitsstätte wird bei PAUL nur durch die jeweilige Firmennummer charakterisiert, die BUR-Nummer hingegen fehlt. Die Verbindung der AST zu ihrer BUR-Nummer und so zu ihrer NOGA – Klasse war nur wenn überhaupt mit grossem Aufwand möglich (Vergleich der Adressen in PAUL mit denjenigen in der Vollerhebung 1996). Von 51 AST konnte darum die BUR-Nummer nicht eruiert werden. Es ist aber davon auszugehen, dass diese nicht (mehr?) zur interessierenden Population gehören, da diese sonst bei der Vollerhebung dabei gewesen wären.

- **ASWZ / NOGA** [(3)]

Bei 10 AST wurde eine NOGA – Klasse eruiert, die nicht der 20.1 entspricht. Der Grund dafür könnte die Umschlüsselung von der ASWZ zur NOGA sein, bei der einige ehemals in die ASWZ – Klasse 2611 zugeteilte AST nicht in die NOGA 20.1 gerutscht sind (z.B. die Furnierwerke).

- **Missing values**

Ein weiteres Problem waren die zahlreichen "Missing values" (fehlenden Werte), aufgeteilt nach "Unit- und item nonresponse" (=Gesamtausfälle bzw. einzelne fehlende Werte), wobei bei PAUL die Gesamtausfälle nicht als solche deklariert werden!
Bei einer Firma fehlte leider auch die Angabe der Anzahl Beschäftigten. Sie wurde daher aus der Stichprobe vorgängig eliminiert. Die fehlenden Werte werden im folgenden Abschnitt analysiert:

2.4 Analyse der Missing Values

Folgende Tabellen sollen das Problem der fehlenden Werte verdeutlichen:

Tabelle 4: Gesamtausfälle von 1995-2000

	NOGA				
Schicht	**20.10A**	**20.10B**	**andere**	**k. Angabe**	**Total**
0	0	1	2	0	3
1	0	3	0	4	7
2	4	2	0	30	36
Total	4	6	2	34	**46**

Von den 135 AST, die ursprünglich einen Fragebogen erhalten haben, bleiben also nur 89 mit mindestens einer gültigen Angabe übrig. Die Gesamtausfälle betragen also

34.1%. Wir gehen im folgenden davon aus, dass diese Ausfälle ignoriert werden können. Ein Problem besteht aber dann, wenn ein hoher Anteil dieser Nicht-Antwortenden eine Produktion von "0" haben. Dies führt zu einer Überschätzung des Holzvolumen (Bias!).
Die gültigen Antworten verteilen sich wie folgt auf die Produkte 2 und 5:

Tabelle 5 : Firmen mit mindestens einer gültigen Antwort von 1995-2000 im Produkt 2

	NOGA				
Schicht	**20.10A**	**20.10B**	**andere**	**k. Angabe**	**Total**
0	3	0	1	1	5
1	8	0	0	1	9
2	49	2	7	15	73
Total	60	2	8	17	**87**

Tabelle 6: Firmen mit mindestens einer gültigen Antwort von 1995-2000 im Produkt 5

	NOGA				
Schicht	**20.10A**	**20.10B**	**andere**	**k. Angabe**	**Total**
0	1	0	0	1	2
1	2	0	0	0	2
2	22	0	1	7	30
Total	27	0	1	8	**36**

Die Analyse der fehlenden Einzelwerte in jedem Jahr ist im Anhang (siehe Tabelle 12 ff.) zu finden. Aus diesen Tabellen ist zu sehen, dass es viele fehlende Einzelwerte gibt. Wir versuchen mit zwei Methoden diese Lücken zu füllen.

2.5 Datenbereinigung / Imputation

Die schlechte Qualität der Daten hatte zur Folge, dass diese vor der Schätzung bereinigt werden mussten und wo möglich Werte eingesetzt wurden.

2.5.1 Datenbereinigung

Die Datenbereinigung umfasste das Löschen folgender Firmen aus der Stichprobe:

- Die Firma, über die keine Angabe zur Anzahl Beschäftigte vorlag und so nicht zu einer Schicht zugeteilt werden konnte
- Die 46 Firmen die in keinem Jahr eine Angabe über das eingeschnittene Holzvolumen gemacht haben
- Die Firmen, die nicht zur NOGA 20.10A oder zur 20.10B gehören, da diese nicht mit unserer Grundpopulation verglichen werden können.

2.5.2 Imputation

Dem Problem der fehlenden Einzelwerte wurde mit einer Einsetzung (einer Imputation) Rechnung getragen. Bei der Imputation wurden zwei verschiedene Ansätze gewählt:

Methode 1:

Bei den 64 Firmen, die nach der Datenbereinigung übrig blieben, wurde versucht, die fehlenden Werte mit folgender Methode einzusetzen:
Fehlt für eine Firma i in einem bestimmten Quartal j im Jahr k beim Produkt 2 oder 5 die Volumenangabe y_{ij}, so wird, falls nicht gerade alle 4 Quartale im Jahr fehlen, das

Holzvolumen für diese Firma in diesem Jahr $\acute{y}_{ik}$ und für dieses Produkt durch folgende Formel bestimmt:

$$\acute{y}_{ik} = 4 * \frac{\sum_{j=1}^{4} y_{ij} * I(y_{ij} \text{ nicht fehlend})}{\sum_{j=1}^{4} I(y_{ij} \text{ nicht fehlend})}$$

Der Quotient "Summe der vorhandenen Werte" durch "Anzahl vorhandene Werte" wird also mit 4 multipliziert, was bedeutet, dass die fehlenden durch den Mittelwert der vorhandenen Werte ersetzt werden. Fehlen die Angaben aller vier Quartale in einem Jahr für ein Produkt einer bestimmten Firma, so konnte leider nichts imputiert werden, d.h. y_{ik} ist dann fehlend. Im Jahr 1995 wurde ebenfalls nichts imputiert, da die vorliegenden Daten keine Unterteilung in Quartale besitzen, sondern bereits die Jahrestotale bilden.
Zu Bemerken ist, dass diese einfache Imputationsmethode natürlich saisonale Effekte nicht berücksichtigen kann. Die Analyse der vorhandenen Daten zeigte allerdings, dass solche Effekte auch gar nicht zu vermuten sind.

Methode 2:

Bei der zweiten Methode wurden alle fehlenden Werte bei den 64 Firmen gleich „0" gesetzt.
Dieser Ansatz wurde gewählt, da nie ganz klar war, wie bei PAUL die fehlenden Werte kodiert werden.

3 Methodik zur Schätzung des Holzvolumens

3.1 Formeln

Bei der Hochrechnung wurden folgende bekannte Schätzer für einfach geschichtete Stichproben verwendet (siehe Cochran p.89 ff.[4]):

Der Schätzer für das Holzvolumentotal ist:

$$\hat{Y}_{tot} = \sum_{h=0}^{2} N_h \bar{y}_h = \sum_{h=0}^{2} \sum_{i \in S_h} \frac{N_h}{n_h} y_{ik} = \sum_{i \in S_h} w_i y_{ik}$$

wobei

$h = 0, 1, 2$ Schichtvariable
N_h = Populationsgrösse der Schicht h
n_h = Stichprobengrösse der Schicht h
w_i = Stichprobengewicht der Schicht h
$\bar{y}_h$ = Mittelwert in der Schicht h $\quad (= n_h^{-1} \sum_{i \in S_h} y_i)$

[4] Sampling Techniques, W. G. Cochran, Wiley, New York, 1977

Die geschätzte Varianz des obigen Schätzers ist:

$$v(\hat{Y}_{tot}) = \sum_{h=0}^{2} N_h^2 \frac{1-f_h}{n_h} S_{y_h}^2$$

wobei

f_h = Stichprobenrate in der Schicht h $(= n_h / N_h)$

$S_{y_h}^2$ = Varianz der Variablen y in der Schicht h $(= (n_h - 1)^{-1} \sum_{i \in S_h} (y_i - \bar{y}_h)^2)$

3.2 Bemerkungen zu den Berechnungen

- Da die Daten des 4 Quartals des Jahres 2000 im Zeitpunkt der Verfassung des Berichtes noch nicht vorlagen, wurde nur mit den Jahren 95-99 gerechnet.
- Da die Angaben über die Populationsgrösse N_h gemäss Betriebszählung nur für die Jahre 1995 und 1998 vorhanden sind, wurde für die anderen Jahre die Annahme getroffen, dass sich die Population zwischen den BZ nicht verändert. So wurden also für die Jahre 95, 96 und 97 die Angaben der BZ 1995 und für die Jahre 98 und 99 diejenigen der BZ 1998 verwendet.
- Für n_h wurde die bereinigte (Netto)-Stichprobengrösse in der Schicht h eingesetzt, d.h. bei der Methode 1 mussten von der effektiven Stichprobengrösse die Missing values abgezogen werden.
- Die obigen Formeln basieren auf dem Horvitz – Thompson Schätzer im Falle einer geschichteten Stichprobe.
- Bei der Schicht der grössten Firmen (h=0), ging man davon aus, dass alle Populationseinheiten befragt wurden, d.h. $N_0 = n_0$, was gleichbedeutend ist mit w_0 =1. Somit wurde versucht, die Auswirkung der zahlreichen Missing values (zu grosse Gewichte für die grössten Firmen bringt eine Überschätzung) in dieser Schicht zu dämpfen.
- Die Schätzungen wurden für die Produkte 2 und 5 je Methode getrennt berechnet.
- Ebenfalls je Methode wurde die Analyse zweimal durchgeführt: 1.) mit NOGA 20.1, 2.) mit NOGA 20.10A als Grundpopulation (→ N_h ändert, siehe Tabelle 1).

3.3 Variablenliste

Für jede Firma wurden die interessierenden Variablen in SAS v.8 wie folgt codiert:

1.) Firmen Variablen:

AnzBe	Anzahl Beschäftigte
BurNr	BUR - Nummer
Firma	Firmen - Nummer
NOGA	NOGA 5-steller
NOGA4	NOGA 4-steller
Schicht2	Schichtvariable h=0,1,2

2.) Variablen zu Produkt 2:

_219950			Volumen Jahrestotal 1995
_219961	bis	**_219964**	Volumen Quartal 1-4, Jahr 1996
_219971	bis	**_219974**	Volumen Quartal 1-4, Jahr 1997
_219981	bis	**_219984**	Volumen Quartal 1-4, Jahr 1998
_219991	bis	**_219994**	Volumen Quartal 1-4, Jahr 1999
_21995v			Volumen Jahrestotal 1995 (0=Missing)
_21996A	bis	**_22000A**	Anzahl Werte im Jahr 96 - 00
_21996imp	bis	**_21999imp**	Jahrestotal 96 - 99 imputiert (Methode1)
_21996imp2	bis	**_21999imp2**	Jahrestotal 96 - 99 imputiert (Methode2)

3.) Variablen zu Produkt 5:

_519950			
_519961	bis	**_519964**	analog zu Produkt 2
_519971	bis	**_519974**	
_519981	bis	**_519984**	
_519991	bis	**_519994**	
_51995v			
_51996A	bis	**_52000A**	
_51996imp	bis	**_51999imp**	
_51996imp2	bis	**_51999imp2**	

3.4 Darstellung der Daten

Die folgenden Scatterplots sollen grob einen Überblick geben über die Struktur der Streuung des Holzvolumens. Aus Gründen des Datenschutzes wurde auf die Achsenbeschriftung verzichtet.
Da die Abbildungen in allen Jahren pro Produkt mehr oder weniger ähnlich aussehen, wurde auf die Darstellung der Graphiken von 1997-1999 verzichtet.

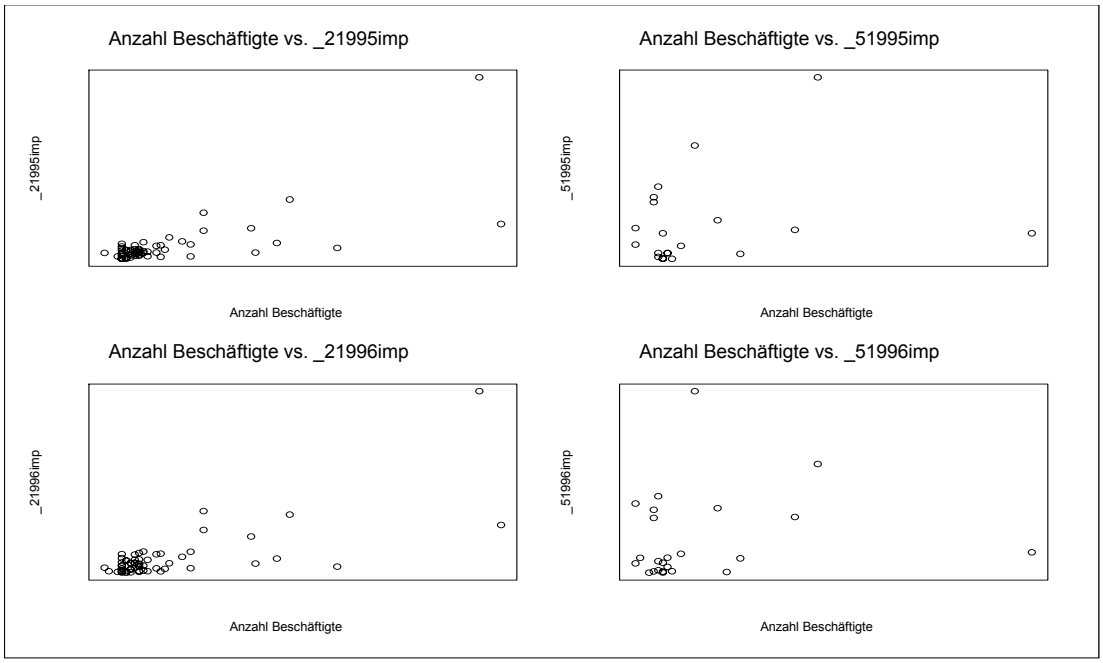

Abbildung 1: Scatterplots

Kommentar:
Die Abbildungen für das Produkt 2 zeigen, dass es eventuell möglich wäre, die Methode des Quotientenschätzers bei der Schätzung des Holzvolumens anzuwenden, da es eine mehr oder weniger lineare Beziehung zwischen der Anzahl Beschäftigten und des eingeschnittenen Holzvolumens gibt. Bei Produkt 5 ist dieser Zusammenhang offensichtlich zu schwach. Da für beide Produkte ein einheitliches Schätzverfahren angewendet werden sollte, wurde auf diese Möglichkeit verzichtet.

4 Resultate

4.1 Übersicht der Nettostichprobengrössen

Folgende Tabelle vergleicht die Nettostichproben- (nach Methode 1) mit den Populationsgrössen:

Tabelle 7: Vergleich der Nettostichproben- mit den Populationsgrössen

NOGA 20.1	**Schicht**	**1995**	**1996**	**1997**	**1998**	**1999**
Populations-	0	8	8	8	5	5
grösse	1	31	31	31	27	27
	2	663	663	663	636	636
Total		**702**	**702**	**702**	**668**	**668**
Nettostich-	0	3	3	1	2	2
probengrösse	1	8	8	8	8	7
Produkt 2	2	48	48	38	34	29
Total		**59**	**59**	**47**	**44**	**38**
Nettostich-	0	1	1	1	1	1
probengrösse	1	3	4	4	4	3
Produkt 5	2	17	19	16	11	10
Total		**21**	**24**	**21**	**16**	**14**
NOGA 20.10A						
Populations-	0	5	5	5	4	4
grösse	1	20	20	20	19	19
	2	571	571	571	556	556
Total		**596**	**596**	**596**	**579**	**579**
Nettostich-	0	3	3	1	2	2
probengrösse	1	8	8	8	8	7
Produkt 2	2	47	46	36	33	28
total		**58**	**57**	**45**	**43**	**37**
Nettostich-	0	1	1	1	1	1
probengrösse	1	3	4	4	4	3
Produkt 5	2	17	19	16	11	10
Total		**21**	**24**	**21**	**16**	**14**

4.2 Hochrechnungen

In den folgenden Tabellen 8 bis 11 sowie den Abbildungen 1 und 2 sind die Resultate der Hochrechnungen nach beiden Methoden, aufgeteilt in die zwei NOGA – Klassen 20.1 und 20.10A für die Jahre 1995 bis 1999 für die Produkte 2 und 5, zu finden.

4.2.1 Methode 1:

Tabelle 8: Resultate der Hochrechnung für die NOGA 20.1

Variable	Variations-koeffizient	Summe in m^3	Standard-abweichung	Untere 95% Vertrauensgrenze für die Summe	Obere 95% Vertrauensgrenze für die Summe
_21995v	0.08	3'800'071	324'217	3'150'586	4'449'555
_21996imp	0.08	3'581'628	295'740	2'989'191	4'174'066
_21997imp	0.11	3'386'320	356'529	2'667'782	4'104'858
_21998imp	0.10	3'641'737	364'037	2'906'549	4'376'926
_21999imp	0.13	3'614'358	470'569	2'659'052	4'569'663
_51995v	0.27	1'916'133	524'342	814'532	3'017'735
_51996imp	0.28	1'713'128	478'940	717'118	2'709'137
_51997imp	0.29	1'729'752	509'950	658'387	2'801'117
_51998imp	0.29	1'245'926	364'413	458'661	2'033'192
_51999imp	0.28	916'076	259'022	345'973	1'486'180

Tabelle 9: Resultate der Hochrechnung für die NOGA 20.10A

Variable	Variations-koeffizient	Summe in m^3	Standard-abweichung	Untere 95% Vertrauensgrenze für die Summe	Obere 95% Vertrauensgrenze für die Summe
_21995v	0.08	3'131'050	266'045	2'597'885	3'664'216
_21996imp	0.08	2'918'454	242'695	2'431'879	3'405'028
_21997imp	0.10	2'775'819	293'520	2'183'472	3'368'165
_21998imp	0.10	3'096'639	310'326	2'469'446	3'723'831
_21999imp	0.13	3'083'126	400'179	2'269'864	3'896'388
_51995v	0.27	1'603'902	440'932	677'539	2'530'265
_51996imp	0.28	1'456'237	410'673	602'196	2'310'279
_51997imp	0.29	1'480'527	438'138	560'034	2'401'020
_51998imp	0.29	1'081'243	318'068	394'100	1'768'387
_51999imp	0.29	788'364	225'854	291'264	1'285'465

4.2.2 Methode 2:

Tabelle 10: Resultate der Hochrechnung für die NOGA 20.1

Variable	Variations-koeffizient	Summe in m^3	Standard-abweichung	Untere 95% Vertrauensgrenze für die Summe	Obere 95% Vertrauensgrenze für die Summe
_219950	0.09	3'504'148	322'783	2'858'704	4'149'592
_21996imp2	0.10	2'891'470	288'908	2'313'763	3'469'178
_21997imp2	0.14	2'359'395	331'047	1'697'426	3'021'364
_21998imp2	0.13	2'324'783	313'577	1'697'748	2'951'818
_21999imp2	0.16	1'991'930	309'252	1'373'543	2'610'317
_519950	0.31	628'028	194'333	239'435	1'016'622
_51996imp2	0.33	525'097	171'771	181'619	868'575
_51997imp2	0.32	352'220	112'355	127'553	576'888
_51998imp2	0.35	198'440	69'836	58'793	338'086
_51999imp2	0.37	154'209	57'671	38'890	269'528

Tabelle 11: Resultate der Hochrechnung für die NOGA 20.10A

Variable	Variations-koeffizient	Summe in m^3	Standard-abweichung	Untere 95% Vertrauensgrenze für die Summe	Obere 95% Vertrauensgrenze für die Summe
_219950	0.09	2'880'977	261'071	2'358'575	3'403'380
_21996imp2	0.10	2'334'217	238'203	1'857'574	2'810'860
_21997imp2	0.14	1'937'136	270'729	1'395'409	2'478'863
_21998imp2	0.13	1'965'800	263'198	1'439'142	2'492'457
_21999imp2	0.15	1'687'325	257'227	1'172'615	2'202'034
_519950	0.31	523'719	162'052	199'453	847'985
_51996imp2	0.33	442'859	146'064	150'585	735'133
_51997imp2	0.32	299'383	95'748	107'792	490'974
_51998imp2	0.36	169'958	60'387	49'124	290'792
_51999imp2	0.38	130'916	49'680	31'506	230'327

4.2.3 Graphiken zu den Resultaten

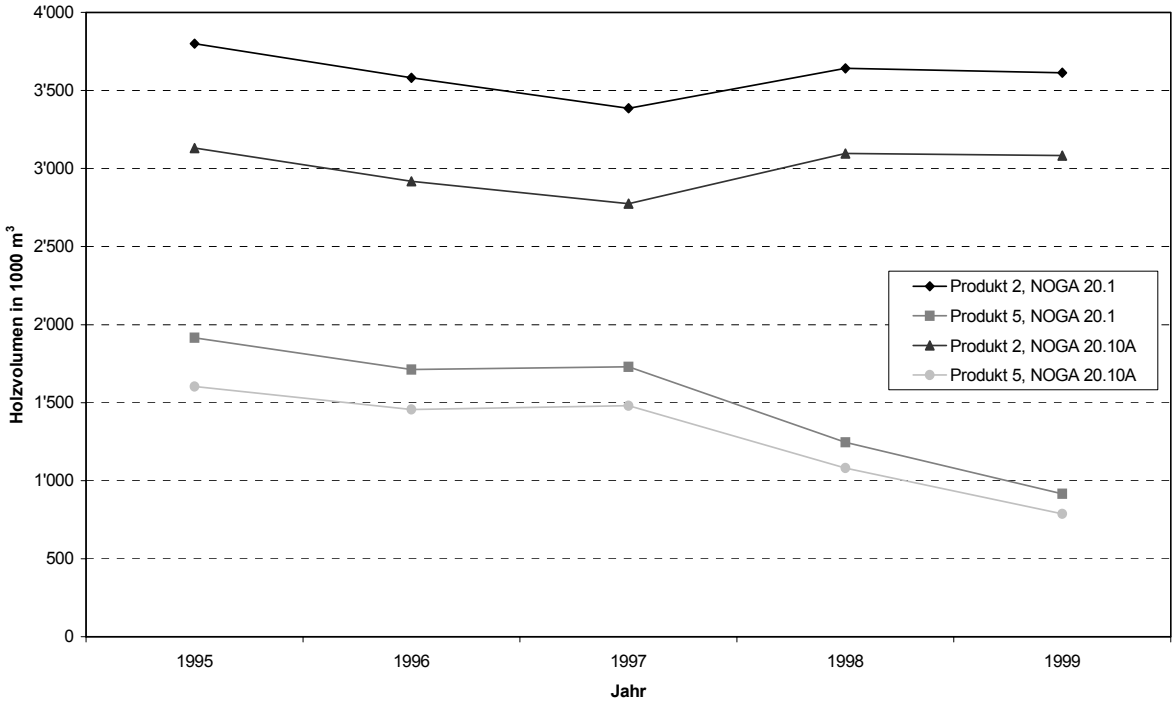

Abbildung 2: geschätztes Holzvolumen (Methode 1)

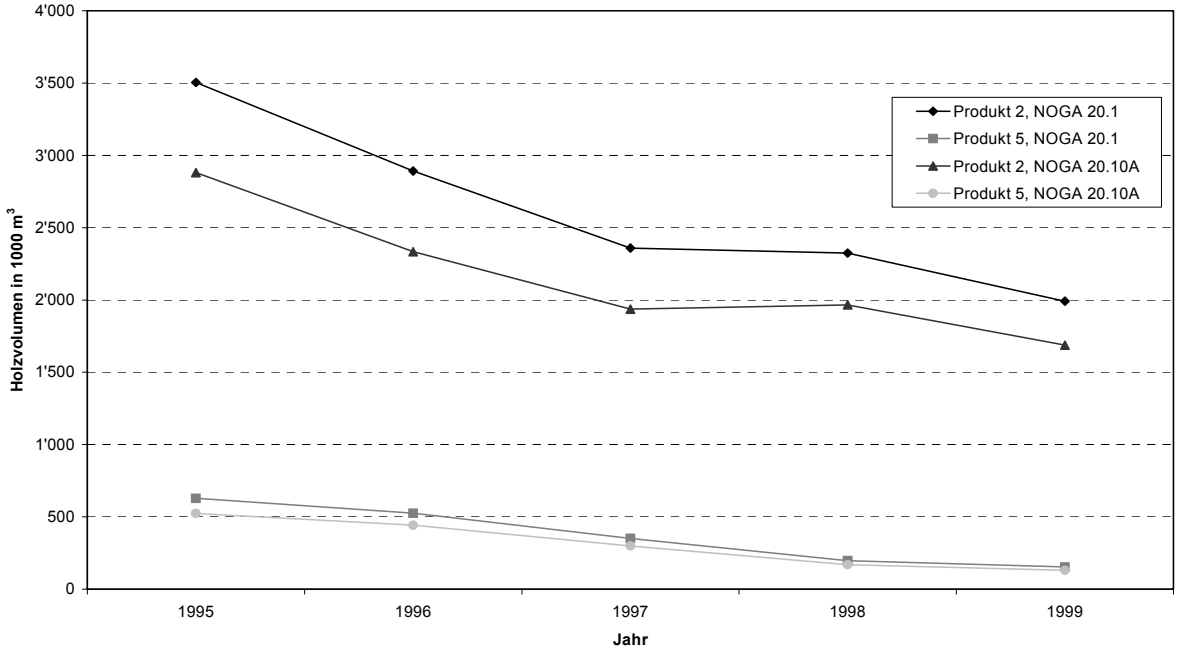

Abbildung 3: geschätztes Holzvolumen (Methode 2)

4.3 Diskussion der Resultate

Bei der NOGA 20.10A sind die Schätzungen des Holzvolumens bei beiden Methoden weniger hoch ausgefallen als bei der NOGA 20.1, da wir dort auf eine kleinere Grundgesamtheit hochgerechnet haben. Je nach NOGA – Stufe variiert die Schätzung z.B. beim Nadelholz im Jahr 1999 (bei Methode1) von 3'083'126 m^3 bis 3'614'358 m^3, was einer Abweichung von 17,2% entspricht!
Innerhalb einer NOGA liefert die Methode 2 tiefere Resultate als bei Methode 1. Dies lässt sich zu Beispiel beim Vergleich des Volumens für NOGA 20.10A, beim Nadelholz, im Jahr 1999 sehen: Methode 2 =1'687'325 m^3, Methode 1 =3'083'126 m^3. →Abweichung 82,7%! Dieser enorme Unterschied, der allein auf die Methode der Behandlung der fehlenden Werte zurückzuführen ist, zeigt, dass aufgrund dieser fehlenden Werte der Bias sehr gross sein kann.
Anhand des Variationskoeffizienten bzw. dem 95% Vertrauensintervall ist zu sehen, dass bei einer Differenzierung der Resultate nach Produkt die Variabilität beim Laubholz bei beiden Methoden und beiden NOGA's zu gross ist, um eine verlässliche Schätzung zu erhalten (mittlerer Variationskoeffizient >29%!, beim Nadelholz beträgt dieser ungefähr 10%). Es ist aber festzuhalten, dass eigentlich Schätzungen mit Variationskoeffizienten von über 0.05 (=5%) nur bedingt brauchbar sind!
Die Schätzung des Holzvolumens bei Methode 2 nimmt von 1995 bis 1999 ab. Dies ist aber nicht nur mit einem Rückgang der Produktion zu erklären, sondern auch mit der Zunahme der fehlenden Werte. Die Schätzungen sind also zu tief. Methode 1 scheint aber andererseits laut Einschätzungen von Experten zu hohe Schätzungen zu geben.
Welche der beiden vorgeschlagenen Methoden nun die besseren Schätzwerte liefert, kann nicht definitiv gesagt werden. Ein Vergleich unserer Resultate für das Jahr 1996 mit den Zahlen aus der Holzverarbeitungserhebung 1996 (HV 1996) zeigt, dass die Werte aus dieser Vollerhebung (Nadelholz: 1.744 Mio. m^3, Laubholz: 0.192 Mio. m^3) für das Laubholz bei Methode 2 (nicht jedoch bei Methode 1) im berechneten 95% Vertrauensintervall liegen. Die Schätzungen des eingeschnittenen Nadelholzes sind bei beiden Methoden höher als bei der HV 1996.
Wenn man in Betracht zieht, dass die Grundgesamtheit der HV 1996 vermutlich grösser ist, als diejenige der NOGA 20.1, dann scheint Methode 2 besser zu sein.
Zusammenfassend lässt sich also folgendes festhalten:

- Der grösste Einfluss auf die Schätzung hat die Behandlung der fehlenden Werte (v.a. weil sehr viele und immer mehr Werte fehlen).
- Der zweitgrösste Einfluss hat die Wahl der Population, auf die hochgerechnet wird.
- Der kleinste Einfluss hat die Varianz. Diese ist aber trotzdem nach unseren Einschätzungen noch zu gross.

5 Anhang

5.1 Tabellen

Folgende zwei Tabellen enthalten die Anzahl AST, die pro Jahr geantwortet haben, aufgeteilt nach Anzahl Antworten (für die Jahre 1995 bis 1999, die Produkte 2 und 5 (P2 bzw. P5), nach Elimination der Gesamtausfälle und aufgeteilt nach NOGA – Klasse). Zu beachten ist, dass im Jahr 1995 nur die Möglichkeiten "Antwort vorhanden" (Anz. Antworten = 4) oder "nicht vorhanden" (Anz. Antworten = 0) möglich sind.

Tabelle 12 : Häufigkeiten nach NOGA 20.10A

Anzahl	1995		1996		1997		1998		1999	
Antworten	P2	P5	P2	P5	P2	P5	P2	P5	P2	P5
0	4	41	5	38	17	41	19	46	25	48
1	-	-	8	3	4	4	2	3	2	2
2	-	-	3	5	3	3	4	3	3	2
3	-	-	8	6	5	3	6	2	5	3
4	58	21	38	10	33	11	31	8	27	7
Total	62	62	62	62	62	62	62	62	62	62
davon mind. 1 Antwort	58	21	57	24	45	21	43	16	37	14

Tabelle 13: Häufigkeiten nach NOGA 20.1

Anzahl	1995		1996		1997		1998		1999	
Antworten	P2	P5	P2	P5	P2	P5	P2	P5	P2	P5
0	5	43	5	40	17	43	20	48	26	50
1	-	-	8	3	5	4	2	3	2	2
2	-	-	3	5	3	3	4	3	3	2
3	-	-	9	6	5	3	6	2	5	3
4	59	21	39	10	34	11	32	8	28	7
Total	64	64	64	64	64	64	64	64	64	64
davon mind. 1 Antwort	59	21	59	24	47	21	44	16	38	14

5.2 Programm für SAS Version 8

Untenstehend ist ein Auszug des Programmiercodes von SAS v.8 zu finden.

```
libname holz  'h:\projekte\hv01';     * Setzen libname;
options linesize=95 pagesize=55 nodate;
run;
title 'Holzproduktion HV01';

 data holz.variante2   ;              *Import der Daten;
    infile 'H:\Projekte\HV01\Variante2.csv' delimiter = ';.' MISSOVER DSD  firstobs=2 ;
       format VAR1 best12. ;
       format VAR2 best12. ;
       format VAR3 $2. ;
       format Produkt best12. ;
       format Quart best12. ;
       format Firma best12. ;
       format Menge best12. ;
       informat VAR1 best32. ;
       informat VAR2 best32. ;
       informat VAR3 $2. ;
       informat Produkt best32. ;
       informat Quart best32. ;
       informat Firma best32. ;
       informat Menge best32. ;
    input VAR1 VAR2 VAR3$ Produkt Quart Firma Menge;
    drop VAR1 VAR2 VAR3;
run;

data var2;                             /*Aufteilung der Variablen Quart in Jahr und
                                      Quartal*/
set holz.variante2;
if Quart > 20000 then b=20000;
else b=19900;
c=Quart-b;
d=floor(c/10);
Jahr=b/10+d;
Quartal=c-10*floor(c/10);
drop  b c d;
run;


proc means data=var2 sum MAXDEC=0;       *Berechnung der Summen je Kategorie;
var Menge;
class Firma Jahr Quartal Produkt;
output out= results sum=Summe;
run;

PROC EXPORT DATA=RESULTS
            OUTFILE= "H:\Projekte\HV01\Output\OutVar2.xls"
            DBMS=EXCEL5 Replace;     *Export der Resultate (Summen je Kategorie);
RUN;

data var2;                            *Zusammenfügen der Variablen Produkt, Jahr und
                                      Quartal;
set var2;                              *in eine neue, zwecks Transposition;
ProdJahrQuart= Produkt*100000 + Quart;
run;
```

```
proc transpose data=var2 label=none  let
              out=var2_transp;          *Transposition der Daten Zwecks Erkennung ;
        by firma;                      *der Missing values;
        id ProdJahrQuart;
       var Menge;
run;

data holz.schicht;                     *Import der Schichtvariblen (2Klassen);
    infile 'H:\Projekte\HV01\Schichten.csv' delimiter = ';' MISSOVER DSD  firstobs=2 ;
       format Firma best12. ;
       format AnzBe best12. ;
       format Schicht best12. ;
       format BurNr best12. ;
       format NOGA $5. ;
       informat Firma best32. ;
       informat AnzBe best32. ;
       informat Schicht best12. ;
       informat BurNr best32. ;
       informat NOGA $5. ;
    input Firma  AnzBe Schicht BurNr NOGA$;
    run;
data schicht;                              *Einführung einer zusaetzlichen Schichtung;
    set holz.schicht;                      *(mit 3 Klassen);
       if AnzBe <20 then Schicht2=2;
       else if AnzBe >=20 and AnzBe <50 then Schicht2=1;
       else if AnzBe >=50 then Schicht2=0;
       else Schicht2=.;
run;

proc sort data=schicht
       out=schicht;              *Sortieren nach Firma, damit gemergt werden kann;
       by firma;
run;

data manip;                            *Mergen der Datensaetzte;
 merge  var2_transp schicht; by firma;
 run;

data manip;    *Löschen der Variablen, die nicht interessieren;
set manip;
drop
_119950 _119961 _119962 _119963 _119964 _119971 _119972 _119973 _119974
_119981 _119982 _119983 _119984 _119991 _119992 _119993 _119994 _120001 _120002 _120003
_319950 _319961 _319962 _319963 _319964 _319971 _319972 _319973 _319974
_319981 _319982 _319983 _319984 _319991 _319992 _319993 _319994 _320001 _320002 _320003
_419950 _419961 _419962 _419963 _419964 _419971 _419972 _419973 _419974
_419981 _419982 _419983 _419984 _419991 _419992 _419993 _419994 _420001 _420002 _420003
_619950 _619961 _619962 _619963 _619964 _619971 _619972 _619973 _619974
_619981 _619982 _619983 _619984 _619991 _619992 _619993 _619994 _620001 _620002 _620003
;
if schicht2=. then delete;       *Records, die keine Schicht haben, werden gelöscht;
run;

PROC EXPORT DATA=manip
            OUTFILE= "H:\Projekte\HV01\Output\grunddaten"
            DBMS=EXCEL5 Replace;     *Export der Grunddaten;
RUN;




data manip;
```

```
 set manip;
 NOGA4= substr(NOGA,1,4);
 NOGA5= substr(NOGA,5,1);
 if NOGA4 ne 2010 then delete;           *Records, die die falsche NOGA (nicht 20.1);
 if NOGA5 = "A" then relev=1;            * haben, werden gelöscht;
 if NOGA5 = "B" then relev=2;
run;

* Berechnung der Antworthäufigkeiten bei Variablen CH2  ;

data manip;
set manip;
_21995A=0;
if (_219950 ne .)
    then _21995A=1;
    if (_219950 = .)
    then _219950 = 0;
    else _219950 =_219950;
run;

%macro cool1b(n);
data manip;
set manip;
_21996A=0;
%do i=1 %to &n;
    if (_21996&i ne .)
    then _21996A + 1;
    if (_21996&i = .)
    then _21996&i = 0;
    else _21996&i =_21996&i;
%end;
run;
%mend;
%cool1b(4);
run;

etc..

* Berechnung der Antworthäufigkeiten bei Variablen CH5  ;

data manip;
set manip;
_51995A=0;
    if (_519950 ne .)
    then _51995A=1;
    if (_519950 = .)
    then _519950 = 0;
    else _519950 =_519950;
run;

%macro cool1e(n);
data manip;
set manip;
_51996A=0;
%do i=1 %to &n;
    if (_51996&i ne .)
    then _51996A + 1;
    if (_51996&i = .)
    then _51996&i = 0;
    else _51996&i =_51996&i;
%end;
run;
```

```
%mend;
%cool1e(4);
run;

etc...


data manip2;
 set manip;
test=_219950+ _219961+ _219962+ _219963+ _219964+ _219971+ _219972+ _219973+ _219974+
_219981+ _219982+ _219983+ _219984+ _219991+ _219992+ _219993+ _219994+
_220001+ _220002+ _220003+
_519950+ _519961+ _519962+ _519963+ _519964+ _519971+ _519972+ _519973+ _519974+
_519981+ _519982+ _519983+ _519984+ _519991+ _519992+ _519993+ _519994+
_520001+ _520002+ _520003
;
if test =0 then delete;                  *Records, die gar nie geantwortet haben werden
                                         gelöscht;
run;

proc freq data=manip2 formchar(1,2,7)='|-+';  *Antworthäufigkeiten  bei Prod 2 und 5;
tables _21995A*NOGA /nocol norow nopercent;
tables _21996A*NOGA /nocol norow nopercent;
tables _21997A*NOGA /nocol norow nopercent;
tables _21998A*NOGA /nocol norow nopercent;
tables _21999A*NOGA /nocol norow nopercent;
tables _22000A*NOGA /nocol norow nopercent;

tables _51995A*NOGA /nocol norow nopercent;
tables _51996A*NOGA /nocol norow nopercent;
tables _51997A*NOGA /nocol norow nopercent;
tables _51998A*NOGA /nocol norow nopercent;
tables _51999A*NOGA /nocol norow nopercent;
tables _52000A*NOGA /nocol norow nopercent;
run;


/*Imputation, falls nicht alle 4 Einträge vorhanden sind.
Methode1: für fehlende Werte innerhalb einer Firma und eines Produkts,
wird arithm. Mittel der vorhandenen Werte (dieser Firma
und dieses Produkts) eingesetzt.
Falls alle 4 Umsaetze fehlen wird nichts imputiert*/

data manip2;
set manip2;
    if _21996A ne 0  then
    _21996imp=4*(_219961+_219962+_219963+_219964)/_21996A;
    else _21996imp=.;

    if _51996A ne 0  then
    _51996imp=4*(_519961+_519962+_519963+_519964)/_51996A;
    else _51996imp=.;
run;

data manip2;
set manip2;
    if _21997A ne 0  then
    _21997imp=4*(_219971+_219972+_219973+_219974)/_21997A;
    else _21997imp=.;

    if _51997A ne 0  then
    _51997imp=4*(_519971+_519972+_519973+_519974)/_51997A;
```

```
    else _51997imp=.;
run;

data manip2;
set manip2;
    if _21998A ne 0  then
    _21998imp=4*(_219981+_219982+_219983+_219984)/_21998A;
    else _21998imp=.;

    if _51998A ne 0  then
    _51998imp=4*(_519981+_519982+_519983+_519984)/_51998A;
    else _51998imp=.;
run;

data manip2;
set manip2;
    if _21999A ne 0  then
    _21999imp=4*(_219991+_219992+_219993+_219994)/_21999A;
    else _21999imp=.;

    if _51999A ne 0  then
    _51999imp=4*(_519991+_519992+_519993+_519994)/_51999A;
    else _51999imp=.;
run;


/*Methode 2: ohne Imputation, Missing=0
a)für Firmen mit mind. 1 Eintrag (64Firmen)(set manip2)
b)für alle (74Firmen) (set manip)
*/

data manip3;
set manip2;
    _21995imp2=_219950;
    _51995imp2=_519950;
    _21996imp2=_219961+_219962+_219963+_219964;
    _51996imp2=_519961+_519962+_519963+_519964;
    _21997imp2=_219971+_219972+_219973+_219974;
    _51997imp2=_519971+_519972+_519973+_519974;
    _21998imp2=_219981+_219982+_219983+_219984;
    _51998imp2=_519981+_519982+_519983+_519984;
    _21999imp2=_219991+_219992+_219993+_219994;
    _51999imp2=_519991+_519992+_519993+_519994;
run;

/* Hochrechnung*/

%macro
Hochrech(dataimput,dataoutput,Variable,Schicht,Poptotal0,Poptotal1,Poptotal2,NOGA);
title "Hochrechnung der Variablen &Variable mit NOGA&NOGA";

proc means data=&dataimput n noprint;    *Berechnung der Schichttotale;
var  &Variable;
class &Schicht;
output out=ttt N=samptot;
run;



data strattot;
    set ttt;
    if (&Schicht = 1)
```

```
       then do;
        sampliweigth=&Poptotal1/samptot;  *Eingabe Sampling Gewicht + Poptotal Schicht 1;
        _total_=&Poptotal1;
       end;
    else if (&Schicht = 2)
       then do;
       sampliweigth= &Poptotal2/samptot; * Berechnung Sampling Gewicht+ Eingabe Poptotal
Schicht 2;
        _total_=&Poptotal2;
       end;
    else if (&Schicht = 0)
       then do;
       sampliweigth= 1;              * Sampling Gewicht =1 und Eingabe Poptotal für
                                    (grösste)Schicht 0;
        _total_=&Poptotal0;
       end;
    if _N_=1 then delete;             *loeschen der ueberfl. spalten/Zeilen;
    drop _type_ _freq_ ;
run;

proc sort data=&dataimput
        out=&dataimput;
    by &schicht;
run;

data &dataoutput;                          *Mergen der Datensaetzte;
 merge  &dataimput strattot; by &schicht;
 if &NOGA=5 and relev=2 then delete;     *Löschen der NOGA 20.10B, falls nicht
                                          erwünscht;
run;


data &dataoutput;
 set &dataoutput;
if &Variable=. then delete ;              *Löschen der Analysevar mit missing values,;
run;                                    *damit Summe der Sampling Gewichte stimmt;

proc surveymeans data=&dataoutput total=strattot missing sum cv std CLSUM ;
var &Variable;                          *Summe aller Holzmengen hochgerechnet;
weight sampliweigth;
strata &Schicht/list;                   *Populgrösse nach BZ98;
run;
%mend;

/* Für NOGA 20.1 -> NOGA =3, für NOGA 20.10A -> NOGA=5 */


/* 1.)für die Daten von 1995
    a)Grundpopulation 20.1 aus BZ95)*/
%Hochrech(manip2,manip295,_21995v,schicht2,8,31,663,3);
%Hochrech(manip2,manip595,_51995v,schicht2,8,31,663,3);
/*  b)Grundpopulation 20.10A aus BZ95)*/
%Hochrech(manip2,manip295,_21995v,schicht2,5,20,571,5);
%Hochrech(manip2,manip595,_51995v,schicht2,5,20,571,5);

/* analog für die Daten von 1996-1999 */

/*für Methode 2 ist manip2 durch manip3 zu ersetzen, und die Variablen anzupassen*/
```

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

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

Rapports de méthodes du Service de 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

Eine 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