Methodenbericht

Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004

0

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

Stichprobenplan, Gewichtung und Schätzverfahren

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

OFS BFS UST

Neuchâtel, 2006

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, Informationsgesellschaft, Sport
17 Politik
18 Öffentliche Verwaltung und Finanzen
19 Kriminalität und Strafrecht
20 Wirtschaftliche und soziale Situation der Bevölkerung
21 Nachhaltige Entwicklung und Disparitäten auf regionaler und internationaler Ebene

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, société de l'information, sport
17 Politique
18 Administration et finances publiques
19 Criminalité et droit pénal
20 Situation économique et sociale de la population
21 Développement durable et disparités régionales et internationales

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

Methodenbericht | **Rapport de méthodes**

Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004

Stichprobenplan, Gewichtung und Schätzverfahren

Autor **Jann Potterat**
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
Swiss Federal Statistical Office

Neuchâtel, 2006

Vorwort

Die Erhebung über die Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004 wird nach 2000 bereits zum zweiten Mal durchgeführt. Das Bundesamt für Berufsbildung und Technologie (BBT) beauftragte die Forschungsstelle für Bildungsökonomie an der Universität Bern, diese Erhebung durchzuführen.

Für die Planung und Ziehung der Stichprobe sowie für die Gewichtung und Schätzverfahren kontaktierte die Uni Bern wiederum den Dienst Statistische Methoden (METH) des Bundesamts für Statistik (BFS), wobei die Arbeiten von Jann Potterat von METH ausgeführt wurden.

Der Autor dankt den Herren Samuel Mühlemann und Prof. Dr. Stefan C. Wolter von der Uni Bern für die sehr gute Zusammenarbeit, Dr. Beat Hulliger von METH für die methodische Unterstützung und die aufmerksame Durchsicht des Berichts, sowie Frau Kathrin Holenstein von der Sektion SCHUL des BFS für die Gesamtkoordination der Arbeiten des BFS.

Zusammenfassung

In diesem Methodenbericht werden die statistischen Methoden (Stichprobenplan und Schätzverfahren) der Erhebung über die Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004 beschrieben.

Aus dem Betriebs- und Unternehmensregister des Bundesamts für Statistik wurde eine nach Wirtschaftsaktivität und Grössenklassen geschichtete Zufallsstichprobe von 40'000 Arbeitsstätten gezogen. Diese Betriebe wurden in zwei Phasen (einem Screening und einer Haupterhebung) befragt, ob sie Lehrlinge ausbilden und falls ja, welche Kosten diese verursachen bzw. Nutzen sie erbringen. Die nicht-ausbildenden Betriebe wurden in einem separaten Fragebogen über die Gründe der Nichtausbildung befragt.

Die Methoden für die Schätzverfahren wurden grösstenteils aus der letzten Erhebung im Jahr 2000 übernommen. Die ermittelte Gewichtung für die antwortenden Betriebe berücksichtigt neben der üblichen Korrektur für die unterschiedlichen Antwortsquoten unter anderem auch die bekannte Verteilung der Berufe gemäss der jährlichen BFS-Lehrlingsstatistik.

Ein Abschnitt widmet sich dem Vergleich der verwendeten Statistik-Softwares SAS und STATA zur Schätzung der provisorischen Resultaten. Am Ende des Berichts sind einige dieser Resultate aufgeführt, genau die Lehrlingslöhne nach Beruf, welche nur einen Teil aller Kosten der Lehrlinge ausmachen. Die offiziellen und definitiven Ergebnisse, welche die Schätzung aller Kosten und Nutzen enthalten werden, erscheinen im Laufe des Jahres 2006.

Schlüsselwörter

Methodenbericht; Kosten Nutzen Berufsbildung; Ausbildung; Lehrling; Stichprobenplan; Schichtung; Gewichtung; Kalibrierung; Calmar; Schätzverfahren; Beruf.

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: 338-0036

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Grafik/Layout: BFS

Copyright: BFS, Neuchâtel 2006
Abdruck – ausser für kommerzielle Nutzung–
unter Angabe der Quelle gestattet.

ISBN: 3-303-00331-9

Inhaltsverzeichnis

1 Einleitung **5**

2 Stichprobenrahmen **5**

3 Stichprobenplan und Stichprobenziehung **8**
3.1 Analyse von Administrativdaten und der letzten Erhebung 8
3.1.1 Lehrlingsstatistik des BFS 9
3.1.2 Verteilung der Berufe auf die NOGA 10
3.1.3 Analyse Rücklauf KNBB00 12
3.2 Definition der Schichtung 13
3.2.1 NOGA Unterabschnitte 13
3.2.2 Grössenklassen / Anzahl Lehrlinge BZ 2001 14
3.3 Stichprobengrösse und Allokation 15
3.4 Ziehung der Stichprobe 17
3.5 Auswahl des Screening-Berufes für die Hauptbefragung 17

4 Analyse Rücklauf KNBB04 und Anpassungen **19**
4.1 Analyse Rücklauf KNBB04 19
4.2 Zusammenlegen Schichten 20
4.3 Direktkontakte 20

5 Gewichtung **21**
5.1 Notationen 21
5.2 Initialgewichtung 21
5.3 Nachgewichtung 21
5.4 Analyse Verteilung der Gewichte 25

6 Schätzverfahren **25**
6.1 Allgemeines 25
6.2 Formeln für Punktschätzer 26
6.3 Vergleich SAS versus STATA 28

7 Provisorische Resultate **29**

Anhang **31**

A SAS-Programme und Daten **31**

B Tabellen **31**

Literatur **34**

1 Einleitung

Die Erhebung über die Kosten und Nutzen der Berufsbildung aus der Sicht der Betriebe im Jahr 2004 soll zeigen, ob sich der finanzielle Aspekt der Ausbildung von Lehrlingen in Betrieben gegenüber der letzten Erhebung vor vier Jahren verändert hat.
Das Bundesamt für Berufsbildung und Technologie (BBT) beauftragte die Forschungsstelle für Bildungsökonomie an der Universität Bern, diese Erhebung durchzuführen.
Für die Planung und Ziehung der Stichprobe sowie für die Gewichtung kontaktierte die Uni Bern wiederum den Dienst Statistische Methoden (METH) des Bundesamts für Statistik (BFS).
In diesem Bericht werden die methodischen Arbeiten beschrieben, welche im Herbst 2004 und 2005 von Jann Potterat/METH mit Unterstützung des Projektteams (S. Mühlemann und St. Wolter) von der Uni Bern ausgeführt wurden.

2 Stichprobenrahmen

Als Grundlage für die Erstellung des Stichprobenrahmens, welcher die Basis für die Planung und Ziehung der Stichprobe ist, diente der Auszug aus dem Betriebs- und Unternehmensregisters (BUR) vom 30. Oktober 2004.
Dieses umfasste zu diesem Zeitpunkt **538'522** aktive Arbeitsstätten (AST), wovon für den Stichprobenrahmen nur die in der Tabelle 1 enthaltenen Betriebstypen berücksichtigt wurden.

Tabelle 1: Betriebstypen Stichprobenrahmen

Betriebstyp	Beschreibung
L13	aktive örtliche Einheit des privaten Sektors (plus SBB)
L20	öffentlicher Sektor: Bund (ohne SBB, POST)
L21	öffentlicher Sektor: Kanton
L22	öffentlicher Sektor: Bezirk
L23	öffentlicher Sektor: Gemeinde
L24	andere öffentliche Körperschaften
L25	Staatlich anerkannte Kirche (exkl. Kt. GE)
L26	öffentlicher Sektor: örtliche Einheit - nur gültig für die Betriebzählung (BZ)
L27	öffentliche Betriebe (inkl. POST)

Diese Selektion entspricht der 'üblichen' Abgrenzung der Population bei einer Erhebung bei Betrieben. Ausgeschlossen wurden also damit die Arbeitsstätten mit den Betriebstypen 'L01' und 'L02' (86'053 Betriebe des Landwirtschaftsektors), sowie 'L14', 'L28' und 'L30' (52'782 Betriebe ohne Beschäftigte bzw. mit Arbeitszeit $<$ 20h bzw. ausländische Botschaften / Konsulate).
Ebenso nicht berücksichtigt wurden die Neuaufnahmen ('L18') bzw. die reaktivierten Betriebe ('L17'), da die Chance, dass diese Betrieben, welche maximal ca. 8-12 Monate alt sein können, Lehrlinge haben, als eher klein eingestuft wurde und wenn sie trotzdem welche hätten, wären diese wohl höchstens im ersten Lehrjahr.

Es blieben also **399'687** Arbeitsstätten übrig, welche als potentielle Population für den Stichprobenrahmen in Frage kamen. Zusätzlich zur Angabe des Betriebstyps der AST im BUR, wurden analog zur letzten Erhebung im Jahr 2000 (siehe dazu auch Renfer [1]) drei weitere Variablen betrachtet, um den definitiven Stichprobenrahmen zu erstellen. Diese Variablen sind:

- Die Wirtschaftsaktivität der AST (NOGA auf Niveau 2)
- Der Betriebstyp auf Niveau Unternehmen
- Die Anzahl Beschäftigte der AST

Wirtschaftsaktivität (NOGA 2):
Die NOGA der AST gibt die Wirtschaftsaktivität des Betriebes an. Von Interesse für diese Erhebung sind generell nur die Lehrberufe des zweiten und dritten Wirtsschaftssektors, obwohl man z.B. den auch zu betrachtenden Beruf des Gärtners mehrheitlich in Betrieben des ersten Sektors findet. Es wurde aber beschlossen, nur die AST mit $\mathbf{10} \leq \mathbf{NOGA\ 2} \leq \mathbf{93}$ zu betrachten.
Obwohl die Betriebstypen aus dem ersten Sektor 'L01' und 'L02' bereits ausgeschlossen wurden, blieben **1'063** zusätzliche AST übrig, welche sich ausserhalb der interessierenden NOGA-Kodierung befinden und noch ausgeschlossen werden mussten.

Betriebstyp auf Niveau Unternehmen:
Weiter wurde der Betriebstyp auf Niveau Unternehmen betrachtet. Dieser ist analog zum Betriebstyp auf Niveau AST in der Tabelle 1 definiert, nur dass bei der Kodierung statt 'Lxx', 'Exx' geschrieben wird. Bei Einzelbetrieben stimmen diese beide Typen überein, bei Mehrbetriebsunternehmen jedoch, wird der Betriebstyp auf Niveau Unternehmen durch denjenigen Betriebstypen mit den meisten Beschäftigten in den einzelnen Arbeitsstätten bestimmt.
Es gibt also z.B. durchaus Mehrbetriebsunternehmen mit z.B. Betyp='E01' (Agrar), welche einzelne AST haben, die im zweiten oder dritten Sektor tätig sind. Diese AST wurden im Stichprobenrahmen belassen, auch aus dem Grund, dass es darunter noch einige AST geben könnte, welche z.B. Gärtner ausbilden.
Es konnte festgestellt werden, dass in **8** Fällen der Unternehmens-Betyp='E91' (Löschungen) war, der AST-Betyp jedoch (noch) gleich 'L13' ist. Diese Fälle können als Fehler betrachtet werden und wurden daher aus der Population ausgeschlossen.

Anzahl Beschäftigte der AST:
Wie bereits bei der letzten Erhebung im Jahr 2000, stellte man sich wieder die Frage, ob man die Betriebe mit nur wenig Beschäftigten (z.B. 1 oder 2) auch noch ausschliessen sollte, da diese ja auch eher wenig bis keine Lehrlinge haben. Diese Tatsache wird mit der folgenden Tabelle 2 sowie der Abbildung 1 aus der Betriebszählung 2001 unterlegt. Dabei wird darin die Verteilung der folgenden Variablen nach der Anzahl Beschäftigten (Vollzeit bzw. Betot=Vollzeit+Teilzeit1+Teilzeit2) dargestellt:

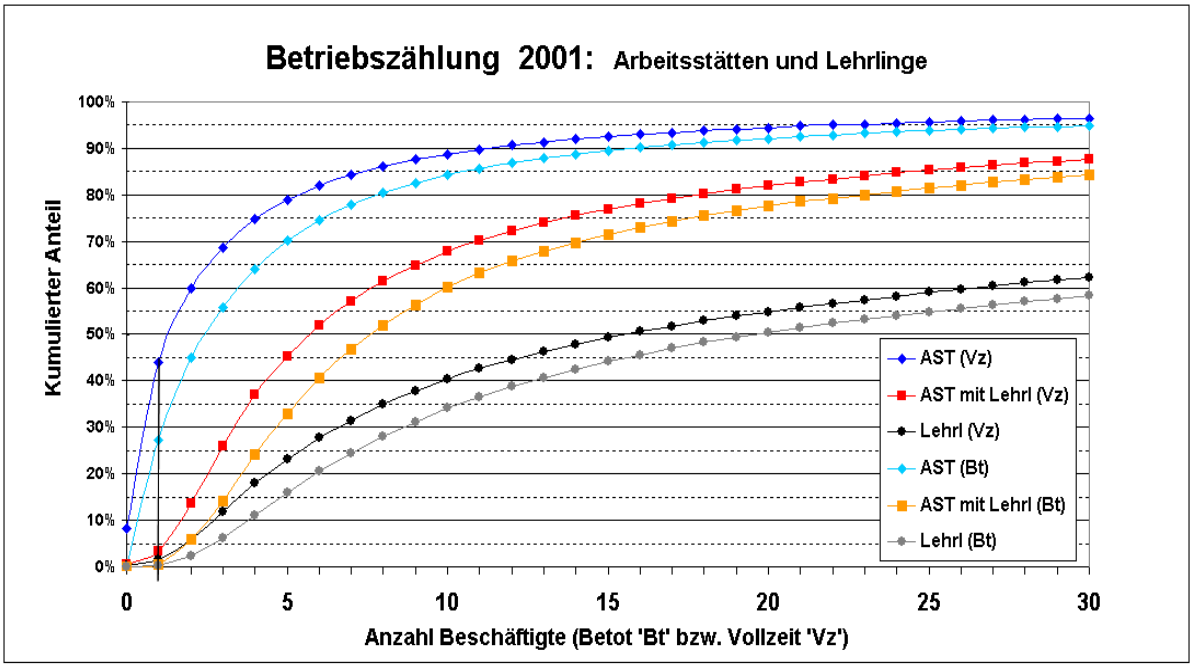
- Anzahl Arbeitsstätten (AST)
- Anzahl Arbeitsstätten mit Lehrlingen (ASTmL)
- Anzahl Lehrlinge (Lehrl)

Tabelle 2: BZ2001: Analyse der Verteilung der Lehrlinge auf die Arbeitsstätten 1

Betriebszählung 2001

Vollzeit		Betot				
		1	2	3	...	Total
0	AST	14'146	8'591	3'806		31'890
	ASTmL	9	60	52	...	271
	Lehrl	9	68	57		355
1	AST	90'689	26'167	9'706		136'655
	ASTmL	349	541	382	...	1'980
	Lehrl	349	562	416		2'187
2	AST		32'316	13'038		60'490
	ASTmL		2'960	1'641	...	6'763
	Lehrl		3'163	1'786		7'567
3	AST			15'449		34'180
	ASTmL			3'379	...	8'116
	Lehrl			4'157		10'214
:	:				:	:
Total	AST	104'835	67'074	41'999		382'979
	ASTmL	358	3'561	5'454	...	66'126
	Lehrl	358	3'793	6'416		170'832

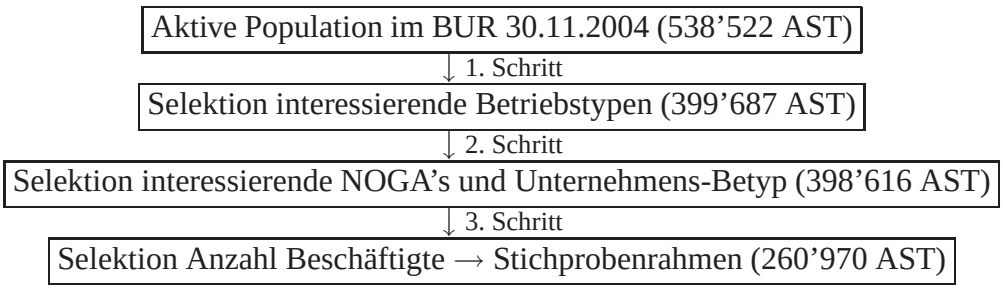
Abbildung 1: BZ2001: Analyse der Verteilung der Lehrlinge auf die Arbeitsstätten 2



Es wurde beschlossen, dass das Kriterium zum Ausschluss von Arbeitsstätten gegenüber der letzten Erhebung nicht verändert werden soll, d.h. alle AST mit **betot=1 oder vz=0** wurden nicht in den Stichprobenrahmen aufgenommen. In der BZ2001 würde dies den 122'579 Arbeitsstätten entsprechen, welche in der Tabelle 2 rot (dunkel) markiert sind. Diesen AST machen zwar 32.0% der Population aus, haben aber nur 704 Lehrlinge in 620 AST, welches wiederum nur 0.4% aller Lehrlinge ist.
Es ist zu bemerken, dass bei der BZ die Lehrlinge als Vollzeit Arbeitskräfte gezählt werden sollten. Dies wurde aber offensichtlich nicht immer so gehandhabt.

Wird das obige Kriterium auf die Arbeitsstätten der potentiellen Population vom 2004 angewendet, fallen dadurch weitere 137'646 AST weg (34.5%). Es bleiben also **260'970 AST** übrig, welche den Stichprobenrahmen bilden.

Zusammengefasst nochmals die Phasen zur Erstellung des Stichprobenrahmens:



3 Stichprobenplan und Stichprobenziehung

3.1 Analyse von Administrativdaten und der letzten Erhebung

Damit wiederum ein optimaler Stichprobenplan erstellt werden konnte, wurden im Voraus drei wichtige Punkte analysiert:

- Die am häufigsten ausgebildeten Berufe in der Schweiz im Jahr 2003 (Gemäss der Lehrlingsstatistik des Bundesamts für Statistik BFS, siehe [2]).
- Die Verteilung der Berufe auf die NOGA 2 gemäss des Screenings der letzten Erhebung im Jahr 2000.
- Die Analyse des Rücklaufs bei der letzten Erhebung im Jahr 2000.

3.1.1 Lehrlingsstatistik des BFS

Für die Erhebung über die Kosten / Nutzen der Berufsbildung sind nur die so genannten sekundären Lehrberufe von Bedeutung, also jene die direkt nach der obligatorischen Schulzeit gelernt werden können. Nicht darunter fallen also die Tertiärberufe, für die es z.B. einen Hochschulabschluss benötigt, aber auch Berufsmaturitäten oder die Anlehren werden nicht betrachtet.
Die Sektion Schul- und Berufsbildung des Bundesamts für Statistik publiziert jährlich eine Statistik über diese Lehrberufe in der Schweiz, wobei die Berufsbezeichnungen vom Bundesamt für Berufsbildung und Technologie (BBT) definiert und kodiert werden.
Die Tabelle 3 enthält die am häufigsten ausgebildeten Lehrberufe in der Schweiz im Jahr 2003, wobei nur die BBG-reglementierten Berufe sowie die SRK-Berufe mit mehr als 1000 Lehrlingen aufgeführt sind.

Tabelle 3: Häufigste Lehrberufe in der Schweiz im Jahr 2003

Rang	Berufname	Anzahl Lehrlinge	Rang	Berufname	Anzahl Lehrlinge
1	Kaufmännische/r Angestellte/r***	20'561	23	Elektroniker/in	2'268
2	Kaufmann/-frau**	11'791	24	Medizinische/r Praxisassistent/in	2'251
3	Elektromonteur/in	7'930	25	Dentalassistent/in	2'162
4	Verkäufer/in	7'419	26	Sanitärmonteur/in	2'139
5	Polymechaniker/in	7'187	27	Konstrukteur/in	1'823
6	Krankenschwester/-pfleger Niveau II*	6'654	28	Bäcker/in-Konditor/in	1'782
7	Informatiker/in	6'604	29	Metallbauer/in	1'764
8	Detailhandelsangestellte/r	6'374	30	Bauzeichner/in	1'575
9	Koch/Köchin	5'037	31	Pflegeassistent/in***	1'462
10	Schreiner/in	4'996	32	Büroangestellte/r***	1'395
11	Automechaniker/in	4'710	33	Landmaschinenmechaniker/in	1'303
12	Coiffeur/-euse	4'244	34	Drogist/in	1'291
13	Automonteur/in	2'951	35	Fachangestellte/r Gesundheit**	1'253
14	Krankenschwester/-pfleger Niveau I*	2'918	36	Heizungsmonteur/in	1'192
15	Gärtner/in	2'914	37	Servicefachangestellte/r	1'176
16	Hochbauzeichner/in	2'822	38	Polygraf/in	1'167
17	Maurer/in	2'620	39	Florist/in	1'123
18	Automatiker/in	2'590	40	Multimediaelektroniker/in	1'098
19	Pharma-Assistent/in	2'476	41	Autolackierer/in	1'076
20	Maler/in	2'433	42	Chemielaborant/in	1'026
21	Zimmermann/Zimmerin	2'421		übrige Lehrberufe (214 Stück)	34'171
22	Logistikassistent/in**	2'319		**Total**	184'468

* neu: tertiärer Lehrberuf, ** neue Berufsbezeichnungen,*** alte Berufsbezeichnungen

Bemerkungen:

- Das Ziel der Erhebung ist einerseits gute globale Informationen über die Kosten und Nutzen der Berufsbildung in der Schweiz zu erhalten, andererseits möchte man aber möglichst auch detaillierte Informationen über einzelne Berufe, darunter insbesondere diejenigen mit den meisten Lehrlingen. Die obige Tabelle 3 erlaubt es also zu bestimmen, welches die 'wichtigsten' Berufe sind.

- Die mit einem oder drei Sternen gekennzeichneten Berufen, hatte man bei der letzten Erhebung noch darin, wird man aber bei dieser oder einer späteren Erhebung nicht mehr analysieren wollen bzw. können, da sie zu tertiären Berufen geworden sind bzw. gar nicht mehr unter dieser Berufsbezeichnungen existieren. Die Kosten der gegenüber der letzten Erhebung neuen Berufstypen wie den Logistikassistenten (darunter fallen ua. die Postboten) oder den Fachangestellten Gesundheit (neuer

sekundärer Beruf im Gesundheitswesens) möchte man falls möglich ebenfalls analysieren können.

- Die Information über die Anzahl Lehrlinge aus der Lehrlingsstatistik wurde bereits bei der letzten Erhebung bei der Ermittlung der Resultate verwendet (siehe Potterat [3]). Auch bei dieser Erhebung können diese Werte bei einer Nachgewichtung wieder berücksichtigt werden.

3.1.2 Verteilung der Berufe auf die NOGA

Anhand der Angaben der antwortenden Betriebe beim Screening der letzten Erhebung 2000, wurde die Verteilung der Berufe auf die NOGA's analysiert. Diese Analyse sollte bei der Planung der Stichprobe verwendet werden, indem z.B. in einer NOGA, wo ein speziell interessierender Beruf vorkommt, überproportional mehr AST in die Stichprobe gezogen werden könnten.
In der Tabelle 4 sind die (ungewichtet) am häufigsten genannten Berufe je NOGA 2 aufgeführt, wobei aus Datenschutzgründen nur diejenigen NOGA 2 Klassen aufgelistet sind, von welchen mindestens 10 Antworten von ausbildenden AST eingetroffen sind. Die Variablen in dieser Tabelle sind wie folgt definiert:

Variable	Beschreibung
NOGA 2	Wirtschaftsaktivität der AST.
Anz. Ast NOGA	Anzahl antwortende ausbildende Arbeitsstätten in der NOGA 2.
Anz. Lehrl. NOGA	Anzahl Lehrlinge dieser ausbildenden AST in der NOGA 2.
Anz. Berufe NOGA	Anzahl unterschiedliche Berufe in der NOGA 2.
Häufigste Berufe Name	Die Berufsbezeichnung der häufigsten Berufe in der NOGA 2.
Häufigste Berufe Code	Der Berufscode (vom BBT) zu den entsprechenden Berufen.
Anz. Ast mit diesem Beruf	Anzahl AST, welche den bestimmten Beruf in der NOGA 2 ausbilden.
Anz. Lehrl. mit diesem Beruf	Anzahl Lehrlinge in diesen AST.
Rang Lehrl / Rang AST	Der Rang der Häufigkeit der Anzahl Lehrlinge bzw. der Anzahl AST mit dem entsprechenden Beruf.

Tabelle 4: Verteilung Berufe auf die NOGA, Erhebung 2000

NOGA 2	Anz. Ast NOGA	Anz. Lehrl. NOGA	Anz. Berufe NOGA	Häufigste Berufe Name	Code	Anz. Ast mit diesem Beruf	Anz. Lehrl. mit diesem Beruf	Rang Lehrl	Rang AST
15	216	897	35	KAUFM. ANGEST.	68001	117	355	1	1
				LEBENSMITTELTECHNOLOGE	21410	24	90	2	3
				KÄSER	21502	40	50	5	2
17	84	307	18	KAUFM. ANGEST.	68001	61	187	1	1
				TEXTILASSISTENT	25905	18	34	2	2
18	33	83	8	KAUFM. ANGEST.	68001	10	30	1	2
				DAMENSCHNEIDER	27110	15	20	3	1
19	12	33	6	ETUISMACHER	33302	2	15	1	3
				KAUFM. ANGEST.	68001	7	12	2	1
				SATTLER	36502	3	4	3	2
20	229	538	9	SCHREINER	30510	200	414	1	1
				ZIMMERMANN	30301	24	54	2	2
21	45	390	19	KAUFM. ANGEST.	68001	40	166	1	1
				PAPIERTECHNOLOGE	33002	8	45	2	4
				VERPACKUNGSTECHNOLOGE	33306	11	27	4	2
22	236	1268	25	POLYGRAF	34702	122	469	1	1
				KAUFM. ANGEST.	68001	92	299	2	3
				DRUCKER	34110	94	202	3	2
24	133	2086	25	ANLAGEN- UND APPARATEBAUER	44720	6	649	1	11
				CHEMIELABORANT	65307	47	445	2	2
				KAUFM. ANGEST.	68001	97	328	3	1
25	119	575	24	KAUFM. ANGEST.	68001	63	232	1	1
				KUNSTSTOFFTECHNOLOGE	38310	43	98	2	2
26	93	305	33	KAUFM. ANGEST.	68001	46	135	1	1
				POLYMECHANIKER	45702	9	31	2	3
				STEINBILDHAUER	39202	10	10	5	2
27	50	438	23	POLYMECHANIKER	45702	37	166	1	1
				KAUFM. ANGEST.	68001	37	138	2	2
28	306	1601	38	POLYMECHANIKER	45702	149	607	1	1
				KAUFM. ANGEST.	68001	115	312	2	2
29	383	5107	41	POLYMECHANIKER	45702	263	2074	1	1
				KONSTRUKTEUR	64205	187	935	2	3
				KAUFM. ANGEST.	68001	192	755	3	2
31	85	855	22	KAUFM. ANGEST.	68001	56	204	1	1
				POLYMECHANIKER	45702	28	191	2	3
				AUTOMATIKER	47408	41	174	3	2
32	43	400	12	ELEKTRONIKER	46502	22	139	1	2
				KAUFM. ANGEST.	68001	28	91	2	1
33	189	1681	34	POLYMECHANIKER	45702	83	552	1	2
				ELEKTRONIKER	46502	38	294	2	3
				KAUFM. ANGEST.	68001	91	270	3	1
34	48	260	17	FAHRZEUGSCHLOSSER	44902	29	96	1	1
				KAUFM. ANGEST.	68001	21	53	2	2
35	84	836	23	POLYMECHANIKER	45702	29	310	1	2
				ELEKTRONIKER	46502	6	140	2	7
				BOOTBAUER	30401	37	59	4	1
36	140	508	26	SCHREINER	30510	63	172	1	1
				KAUFM. ANGEST.	68001	48	139	2	2
40	159	1335	32	ELEKTROMONTEUR	47412	60	405	1	3
				KAUFM. ANGEST.	68001	87	274	2	1
				NETZELEKTRIKER	47406	85	224	3	2
41	17	41	10	KAUFM. ANGEST.	68001	9	22	1	1
				FORSTWART	19101	3	5	2	2
45	1146	5415	69	ELEKTROMONTEUR	47412	217	1753	1	3
				MAURER	51002	316	996	2	1
				KAUFM. ANGEST.	68001	221	354	3	2
50	444	1835	22	AUTOMECHANIKER	46304	225	798	1	1
				KAUFM. ANGEST.	68001	86	247	2	4
				AUTOMONTEUR	46306	166	239	3	2
51	458	2419	43	KAUFM. ANGEST.	68001	354	1316	1	1
				LAGERIST	95502	99	333	2	2
52	911	3571	52	VERKÄUFER	70400	365	1245	1	1
				DETAILHANDELSANGEST.	70200	224	781	2	2
55	384	2337	24	KOCH	79003	326	920	1	1
				SERVICEFACHANGEST.	78703	161	315	2	2
60	87	545	22	KAUFM. ANGEST.	68001	57	115	1	1
				POLYMECHANIKER	45702	11	100	2	4
				LASTWAGENFÜHRER	73302	24	94	3	2
63	138	1018	22	KAUFM. ANGEST.	68001	121	730	1	1
				LAGERIST	95502	22	72	2	3
				BÜROANGESTELLTER	68101	33	71	3	2
64	16	100	7	INFORMATIKER	47102	8	33	1	1
				KAUFM. ANGEST.	68001	7	22	3	2
65	143	2181	10	KAUFM. ANGEST.	68001	148	2073	1	1
				INFORMATIKER	47102	15	73	2	2

NOGA2	Anz. Ast NOGA	Anz. Lehrl NOGA	Anz. Berufe NOGA	Häufigste Berufe Name	Code	Anz. Ast mit diesem Beruf	Anz. Lehrl mit diesem Beruf	Rang Lehrl	Rang AST
66	113	1116	11	KAUFM. ANGEST.	68001	99	1041	1	1
				INFORMATIKER	47102	13	32	2	2
67	12	25	2	KAUFM. ANGEST.	68001	12	23	1	1
70	97	220	18	KAUFM. ANGEST.	68001	87	171	1	1
				HOCHBAUZEICHNER	64004	4	6	2	2
72	118	410	8	INFORMATIKER	47102	93	276	1	1
				KAUFM. ANGEST.	68001	41	99	2	2
73	42	329	24	CHEMIELABORANT	65307	13	59	1	2
				KAUFM. ANGEST.	68001	20	48	2	1
74	787	2094	59	KAUFM. ANGEST.	68001	328	797	1	1
				HOCHBAUZEICHNER	64004	179	255	2	2
75	542	2972	49	KAUFM. ANGEST.	68001	458	1872	1	1
				POLYMECHANIKER	45702	12	190	2	10
				BÜROANGESTELLTER	68101	64	131	4	2
80	243	8875	64	ELEKTROMONTEUR	47412	3	992	1	18
				KAUFM. ANGEST.	68001	160	737	2	1
				INFORMATIKER	47102	40	594	4	2
85	948	3340	70	KOCH	79003	476	887	1	1
				KRANKENPFLEGER NIVEAU I	86100	70	565	2	6
				KAUFM. ANGEST.	68001	252	538	3	2
90	16	33	11	KAUFM. ANGEST.	68001	6	12	1	1
				GÄRTNER	17007	3	7	2	2
91	53	131	10	KAUFM. ANGEST.	68001	45	106	1	1
				KOCH	79003	6	10	2	2
92	47	727	17	INFORMATIKER	47102	6	619	1	3
				INFORM. - + DOKUM. ASSISTISTENT	70506	10	36	2	2
				KAUFM. ANGEST.	68001	17	31	3	1
93	195	401	9	COIFFEUR	82010	170	338	1	1
				KOSMETIKER	82111	16	25	2	2

Bemerkungen:

- Wie man vermuten konnte, ist zu sehen, dass die kaufmännischen Angestellten Lehrlinge in jeder Branche vorkommen, wobei im dritten Wirtschaftssektor noch häufiger als im zweiten.
- Die Vielfalt der unterschiedlichen Berufe ist im Baugewerbe (NOGA2=45), sowie im Gesundheits- und Sozialwesen (NOGA2=85) am ausgeprägtesten.

3.1.3 Analyse Rücklauf KNBB00

Bei der Ermittlung der Schätzverfahren sowie der Beurteilung der Qualität einer Erhebung, ist es u.a. wichtig den Rücklauf zu analysieren.
Die Gründe für eine Nicht-Teilnahme können immer von unterschiedlicher Natur sein: Interesse am Thema der Erhebung, Inhalt / Länge des Fragebogens, Zeitpunkt der Erhebung etc. Bei der letzten Erhebung war es auch so, dass nur diejenigen Betriebe, welche sich beim Screening als Teilnahmebereit erklärt haben, schlussendlich auch einen Fragebogen der Hauptbefragung erhielten.
Der Rücklauf bei der Hauptbefragung im Jahr 2000 konnte global als gut bezeichnet werden (2'374 verwertbare Fragebogen von ausbildenden Betrieben), jedoch gab es einzelne Schichten, welche mangels genügenden Rücklaufs zusammengefasst werden mussten (siehe Bericht Potterat [3]).
Da sich das Design (Stichprobengrösse, Fragebogen für Screening und Hauptbefragung, Schichtung) bei der Erhebung im 2004 grundsätzlich nicht gross verändern sollte, konnte man also die Information des Rücklaufs aus der letzten Erhebung bei der Planung berücksichtigen.

3.2 Definition der Schichtung

Für die Bestimmung der Schichtung wurden drei Variablen verwendet. Darunter waren wie bei der letzten Erhebung eine Zusammenfassung von den NOGA 2 zu **Unterabschnitten**, sowie die Aufteilung der Betrieben in **Grössenklassen**. Als neue Schichtungsvariable kam die Information dazu, ob ein Betrieb gemäss der letzten Information (der BZ 2001) **viele Lehrlinge** ausbildet. Näheres wird nun in den folgenden Abschnitten beschrieben.

3.2.1 NOGA Unterabschnitte

Wie erwähnt sind im Stichprobenrahmen nur die (insgesamt 57) NOGA 2 - Klassen mit $\mathbf{10 \leq NOGA\ 2 \leq 93}$ von Interesse. Da die NOGA wie bekannt hierarchisch aufgebaut ist, lassen sich diese Klassen zusammenfassen zu Abschnitten und Unterabschnitten. Für die Schichtung wurde folgende in der Tabelle 5 aufgeführte Zusammenfassung verwendet, wobei diese erste Schichtungssvariable '`unterab`' benannt wurde.

Tabelle 5: Erste Schichtungsvariable: `unterab` (NOGA Unterabschnitte)

Nr.	`unterab`	Abschnitte/ Unterabschnitte	NOGA 2	Beschreibung	Anzahl AST
1	C-DC	CA, CB, DA, DB, DC	10-14, 15-19	Bergbau, Herstellung von Nahrungsmitteln, Getränken, Textilien und Lederwaren	3'859
2	DD	DD	20	Be- und Verarbeitung von Holz	4'410
3	DE	DE	21, 22	Papier-, Karton-, Verlags- und Druckgewerbe	3'389
4	DF-DI	DF, DG, DH, DI	23-26	Kokerei, chem. Industrie, Herstellung von Kunststoffwaren etc.	2'634
5	DJ	DJ	27, 28	Erzeugung und Bearbeitung von Metallerzeugnissen	6'273
6	DK	DK	29	Maschinenbau	2'835
7	DL	DL	30-33	Herstellung von elektr. und elektron. Geräten, Feinmechanik, Optik	4'056
8	DMDN	DM, DN	34-37	Fahrzeugbau, sonst. verarbeitendes Gewerbe	2'807
9	E	E	40, 41	Energie- und Wasserversorgung	787
10	F	F	45	Baugewerbe	25'990
11	G50	G-	50	Handel, Reparatur von Automobilien,Tankstellen	10'991
12	G51	G-	51	Handelsvermittlung und Grosshandel	14'564
13	G52	G-	52	Detailhandel, Reparatur von Gebrauchsgütern	35'620
14	H	H	55	Gastgewerbe	25'686
15	I	I	60-64	Verkehr und Nachrichtenübermittlung	12'695
16	J	J	65, 66, 67	Kredit und Versicherungsgewerbe	8'248
17	K7071	K-	70, 71	Immobilienwesen, Vermietung bewegliche Sachen ohne Bedienungspersonal	3'169
18	K72	K-	72	Informatikdienste	6'184
19	K7374	K-	73, 74	Forschung und Entwicklung, Erbringung Dienstl. für Unternehmen	34'263
20	L	L	75	Öffentl. Verwaltung, Landesverteidigung, Sozialversicherung	7'726
21	M	M	80	Unterrichtswesen	10'406
22	N	N	85	Gesundheits- und Sozialwesen	18'628
23	O	O	90-93	Erbringung von sonst. öffentl. und persönlichen Dienstleistungen	15'750
				Total	260'970

Verglichen mit der letzten Erhebung im 2000, wo 30 Unterabschnittsgruppen gebildet worden sind, gibt es also nun 7 Unterabschnittsklassen weniger. Diese 'Kürzung', welche nur einer gröberen Zusammenfassung der kleinen Schichten entspricht, wurde unter der

Annahme vorgenommen, dass der Rücklauf ähnlich zu erwarten ist wie bei der letzten Erhebung.

3.2.2 Grössenklassen / Anzahl Lehrlinge BZ 2001

Als zweite Variable der Schichtung wurde eine Kombination der Variablen Grössenklasse sowie der Information über die Anzahl Lehrlinge bei der BZ 2001 verwendet. Bei der BZ 2001 haben 66'126 Betriebe angegeben, dass sie Lehrlinge ausbilden (271'902 bilden nicht aus, 44'951 machten keine Angabe). Die Verteilung der Anzahl Lehrlinge je ausbildende AST bei der BZ2001 ist der Tabelle 6 zu entnehmen:

Tabelle 6: BZ 2001: Verteilung Anzahl Lehrlinge in den ausbildenden Arbeitsstätten

Percentil	Anz. Lehrlinge
0	1
25	1
50	2
75	3
90	4
95	6
97.5	10
99	17
100	725

Obwohl diese Information im 2004 nicht mehr aktuell ist und auch nicht mehr richtig sein muss, wurde bestimmt, dass alle AST aus dem Stichprobenrahmen, welche bei der BZ2001 mehr als 7 Lehrlinge hatten (Variable AnzLehrl01), in eine spezielle Schicht aufgenommen wurden, welche dann vollerhoben werden sollte. Die restlichen AST, also auch inkl. denjenigen, von welchen keine Angabe über die Lehrlinge im 2001 vorliegt, wurden wie bereits bei der letzten Erhebung anhand der Anzahl Beschäftigten (Betot) in vier Grössenklassen eingeteilt. Die Tabelle 7 gibt einen Überblick dieser zweiten Schichtungsvariable, welche '`grkllehr`' benannt wurde.

Tabelle 7: Zweite Schichtungsvariable: `grkllehr` (Grössenklassen / Anzahl Lehrlinge)

`grkllehr`	Bedingungen			Anzahl AST
	Betot		AnzLehrl01	
1	1-9	und	≤ 7	196'074
2	10-49	und	≤ 7	53'479
3	50-99	und	≤ 7	5'645
4	>100	und	≤ 7	3'337
5			>7	2'435
			Total	260'970

Die Idee der Bildung dieser Schichtungsvariablen sollte sein, dass neben den 2'435 AST mit `grkllehr`=5 auch die grossen AST mit vielen Beschäftigten (`grkllehr`= 3 oder 4) vollerhoben werden sollten.

Durch Kombination (Kreuzung) der beiden Schichtungsvariablen `unterab`$\times$`grkllehr` ergab sich die definitive Schichtung mit 23$\times$5=115 Schichten.

3.3 Stichprobengrösse und Allokation

Als maximale (Brutto-)Stichprobengrösse für das Screening wurde vom Projektteam n = 40'000 vorgegeben. Diese Grösse entspricht derselben Vorgabe wie bei der ersten Erhebung. Die effektive Stichprobengrösse betrug damals aber schlussendlich 39'800.
METH hat dem Projektteam einige Vorschläge einer Allokation, d.h. der Aufteilung der 40'000 AST auf die einzelnen (115) Schichten, präsentiert. Dabei wurden grundsätzlich folgende Annahmen gemacht:

- Vollerhebung der Schichten mit `grkllehr`=3,4,5
- Aufteilung proportional zu den Anzahl AST im Stichprobenrahmen
- Aufstockung der Stichprobe in den Schichten, wo die Antwortquote das letzte Mal mager war, mit dem Ziel je Schicht mindestens 10-15 Antworten von ausbildenden AST zu erhalten.

In einer ersten Phase wurde auch eine optimale Aufteilung (im Sinn von Neyman, siehe in Cochran [4], oder Särndal [5]) getestet. Dies würde bedeuten, dass man in den Schichten, wo es eine grössere Variabilität der Kosten/Nutzen gibt, der Stichprobenanteil eher grösser wäre als in den Schichten mit einer kleineren Variabilität.
Von dieser Idee hat man schlussendlich abgesehen, da die Information über die Variabilität je Schicht anhand der letzten Erhebung hätte berechnet werden müssen. Diese Berechnung wäre aber einerseits relativ aufwendig gewesen, da die alten und neuen Schichten nicht genau dieselben sind und einige Angaben nicht auf die Schichten hätten zugeteilt werden können (Direktkontakt Problemfälle), andererseits ist diese Erhebung auch schon vier Jahre zurück, was bedeutet, dass eine optimale Allokation mit vierjährigen Daten heute nicht mehr optimal zu sein braucht.

Aufgrund der Allokations-Vorschläge von METH, wurden vom Projektteam einige Änderungswünsche angebracht, welche Schichten zusätzlich noch aufgestockt (oder auch vermindert) werden sollten, da sich darin Berufe von speziellem Interesse befinden, die man mindestens so detailliert analysieren möchte wie beim letzten Mal wie z.B. die Berufe des Coiffeurs, des Hochbauzeichners und die Berufe des Baugewerbes.
Die Tabelle 8 enthält den definitiven Stichprobenplan der Erhebung der Kosten / Nutzen der Berufsbildung im Jahr 2004, wobei folgende Notationen verwendet werden:

N_h	Populationsgrösse je Schicht
n_h	Stichprobengrösse je Schicht
f_h	Stichprobenanteil je Schicht in Prozent (=$100 \times n_h/N_h$)

Tabelle 8: Stichprobenplan der Erhebung KNBB04

unterab	Grössenklasse (grkllehr)															Total		
	1			2			3			4			5					
	N_h	n_h	f_h	N_h	n_h	f_h	N_h	n_h	f_h	N_h	n_h	f_h	N_h	n_h	f_h	N_h	n_h	f_h
C-DC	2'581	442	17.1	919	340	37.0	169	169	100	141	141	100	49	49	100	3'859	1'141	29.6
DD	3'503	417	11.9	841	100	11.9	40	40	100	5	5	100	21	21	100	4'410	583	13.2
DE	2'361	281	11.9	785	95	12.1	97	97	100	68	68	100	78	78	100	3'389	619	18.3
DF-DI	1'374	625	45.5	837	238	28.4	200	200	100	151	151	100	72	72	100	2'634	1'286	48.8
DJ	4'219	503	11.9	1'668	199	11.9	184	184	100	99	99	100	103	103	100	6'273	1'088	17.3
DK	1'553	357	23.0	836	110	13.2	169	169	100	99	99	100	178	178	100	2'835	913	32.2
DL	2'480	309	12.5	1'008	119	11.8	247	247	100	197	197	100	124	124	100	4'056	996	24.6
DMDN	2'039	238	11.7	602	151	25.1	74	74	100	53	53	100	39	39	100	2'807	555	19.8
E	356	310	87.1	323	68	21.1	48	48	100	27	27	100	33	33	100	787	486	61.8
F	18'904	2'380	12.6	6'127	772	12.6	418	418	100	159	159	100	382	382	100	25'990	4'111	15.8
G50	9'210	860	9.3	1'566	186	11.9	36	36	100	11	11	100	168	168	100	10'991	1'261	11.5
G51	10'794	1'072	9.9	3'146	181	5.8	377	377	100	171	171	100	76	76	100	14'564	1'877	12.9
G52	29'470	3'711	12.6	5'407	681	12.6	381	381	100	110	110	100	252	252	100	35'620	5'135	14.4
H	19'505	2'233	11.4	5'687	651	11.4	280	280	100	65	65	100	149	149	100	25'686	3'378	13.2
I	8'517	670	7.9	3'353	264	7.9	405	405	100	324	324	100	96	96	100	12'695	1'759	13.9
J	5'322	464	8.7	2'211	187	8.5	289	289	100	275	275	100	151	151	100	8'248	1'366	16.6
K7071	2'625	301	11.5	479	55	11.5	38	38	100	20	20	100	7	7	100	3'169	421	13.3
K72	5'035	601	11.9	985	143	14.5	96	96	100	52	52	100	16	16	100	6'184	908	14.7
K7374	29'305	2'306	7.9	4'204	373	8.9	354	354	100	318	318	100	82	82	100	34'263	3'433	10.0
L	4'509	378	8.4	2'670	237	8.9	293	293	100	171	171	100	83	83	100	7'726	1'162	15.0
M	5'505	630	11.4	4'100	506	12.3	476	476	100	252	252	100	73	73	100	10'406	1'937	18.6
N	13'811	1'581	11.4	3'354	384	11.4	819	819	100	463	463	100	181	181	100	18'628	3'428	18.4
O	13'096	1'499	11.4	2'371	375	15.8	155	155	100	106	106	100	22	22	100	15'750	2'157	13.7
Total	196'074	22'168	11.3	53'479	6'415	12.0	5'645	5'645	100	3'337	3'337	100	2'435	2'435	100	260'970	40'000	15.3

3.4 Ziehung der Stichprobe

Die Ziehung der Stichprobe wurde mit der Prozedur `proc surveyselect` der Statistiksoftware SAS v8.2 vollzogen, indem gemäss des Plans der Tabelle 8 innerhalb jeder Schichten eine einfache Zufallsstichprobe gezogen wurde.

Die ASCII-Datei mit der definitiven Stichprobe `sampleKnbb2004.txt` (aufgeteilt in zwei Zip-Dateien) wurde dem Projektteam am Freitag, 19. November 2004 geliefert.
12 korrigierte Adressen - wegen fehlerhaften Postfachbezeichnung im BUR - mussten am 24.11.04 nachgeliefert werden.

3.5 Auswahl des Screening-Berufes für die Hauptbefragung

Ein wichtiger methodischer Aspekt, dem bei dieser Erhebung grosse Beachtung geschenkt wird, ist folgender:
Beim Screening wird jede der 40'000 AST gefragt, welche(n) Beruf(e) sie ausbilden und wieviele Lehrlinge je Beruf sie haben, wobei bei der letzten Erhebung maximal sieben (neu sechs) Berufe angegeben werden konnten.
Bei der Hauptbefragung wird von den Arbeitsstätten aus Gründen des relativ langen Fragebogens aber nur die Angabe der Kosten / Nutzen **eines einzigen** Berufes verlangt. Die Kosten für die anderen Berufe werden also ignoriert.
Die Tabelle 9 enthält die (ungewichtete) Verteilung der Anzahl Berufe je Arbeitsstätte gemäss des Screenings der letzten Erhebung.

Tabelle 9: Anzahl unterschiedliche Berufe je AST

Anzahl Berufe	Anzahl AST	
	absolut	in %
1	6'447	66.25
2	1'748	17.96
3	733	7.53
4	366	3.76
5	192	1.97
6	120	1.23
7	125	1.28
Total	9'731	100.00

Es ist zu sehen, dass 66% der ausbildenden Arbeitsstätten nur einen einzigen Beruf ausbilden. Bei 34% der AST musste jedoch der Beruf für die Hauptbefragung ausgewählt werden.
Es sind mehrere Möglichkeiten einer Auswahl des Berufes denkbar. Aus statistischer Sicht ist es aber wichtig, dass diese Auswahl möglichst **zufällig** geschieht.
Nicht ideal wäre es also z.B. einfach den ersten Beruf der Liste auf dem Fragebogen oder den Beruf mit den meisten Lehrlingen zu nehmen.

Aus methodischer Sicht wird für die Erhebung 2004 folgendes Vorgehen empfohlen:

Schritt 1:
Es wird für jede AST mit mehreren Berufen die Anzahl Lehrlinge je Beruf analysiert.

Schritt 2:
Falls eine AST nur einen Beruf ausbildet, wird dieser natürlich auch bei der Hauptbefragung gewählt.
Falls zwei oder mehr Berufe gemäss dem Screening ausgebildet werden, sollte die Auswahl nach folgendem Kriterium erfolgen:

- Der Beruf wird proportional zur Anzahl Lehrlinge innerhalb der AST zufällig ausgewählt.

Ein Beispiel zu dieser zufälligen proportionalen Bestimmung:

Ein Betrieb X hat insgesamt 10 Lehrlinge in 3 Berufen: 5 Zimmerleute, 3 Maurer und 2 Elektromonteure. Für diesen Betrieb wird mit der Funktion `ranuni()` im SAS - oder `=ZUFALLSZAHL()` im Excel - eine uniformverteilte Zufallszahl zwischen 0 und 1 generiert. In Abhängigkeit des Wertes der Zufallszahl wird der entsprechende Beruf gewählt:

Wert im Intervall		gewählter Beruf
$[0/10; 5/10)$	$= [0.0; 0.5)$	Zimmermann
$[5/10; 8/10)$	$= [0.5; 0.8)$	Maurer
$[8/10; 10/10]$	$= [0.8; 1.0]$	Elektromonteur

Dies entspricht also einer zufälligen Auswahl proportional zur Anzahl Lehrlinge.

Eine mögliche Alternative einer Auswahl wäre folgende gewesen:
Die Unternehmen wählen aus allen Lehrlingen zufällig einen aus, wobei alle Lehrlinge dieselbe Ziehungswahrscheinlichkeit haben, und wählen den Beruf des gezogenen Lehrlings aus.
Diese Alternative wurde aber nicht gewählt, da man die oben beschriebene Prozedur als die bessere befand und andererseits wäre es eventuell nicht für alle Unternehmen möglich gewesen, eine saubere zufällige Auswahl zu treffen.

4 Analyse Rücklauf KNBB04 und Anpassungen

4.1 Analyse Rücklauf KNBB04

Ab Mitte Oktober 2005 wurde dem BFS die Daten des Screenings sowie provisorische Daten der Haupterhebung zur Verfügung gestellt. Zur Auswertung wurde schlussendlich der plausibilisierte Datensatz vom 5. Januar 2006 (`datensatz050106.sas7bdat`) verwendet. Die Tabelle 10 zeigt den Rücklauf:

Tabelle 10: Rücklauf KNBB04

		Anzahl AST		Anzahl Lehrlinge	
		Total	in %	Total	in Direktkontakten$^{(b)}$
Stichprobenrahmen		260'970		175'019$^{(a)}$	
Stichprobe Brutto		40'000			
Antworten Screening	ausbildend	10'931	38.19	67'772	13'896
	nicht ausbildend	17'689	61.81		
	Total	28'620			
Versandte Fragebogen	ausbildend	4'755	61.88		
Haupterhebung	nicht ausbildend	2'929	38.12		
(Teilnahmebereite AST)	Total	7'684			
Antworten Haupterhebung	ausbildend	2'413	56.43	13'454	5'236
	nicht ausbildend	1'863	43.57		
	Total	4'276			

$^{(a)}$ Wert aus der Lehrlingsstatistik 2004 ([6])
$^{(b)}$ Direktkontakte sind i.A. grosse Unternehmen, welche für mehrere AST geantwortet haben. (siehe Abschnitt 4.3)

Der Anteil retournierter Fragebogen beim Screening beträgt 71.6% und ist daher tiefer als der Wert aus der letzten Erhebung (80%). Der Fragebogen der Haupterhebung wurde wiederum nur an diejenigen AST geschickt, welche beim Screening einer Teilnahme zugestimmt haben. Dies waren 4'755 / 10'931 =43.5% der ausbildenden sowie 16.6% der nicht ausbildenden Betriebe.
Bei der Haupterhebung antworteten die nicht ausbildenden Betriebe mit 63.6% (=1'863 / 2'929) leicht besser als die ausbildenden Betriebe (50.7%). Global und im Vergleich zur letzen Haupterhebung gesehen, wo die Anzahl der ausgefüllten Fragebogen für die nicht ausbildenden sowie ausbildenden Betriebe je rund 2'300 betrug, ist nun die Anzahl der retournierten Fragebogen für die ausbildenden Betriebe höher.
Die globale Antwortquote dieser Erhebung beträgt gemäss obiger Tabelle also 4'276 / 40'000 = 10.7%.
Die Analyse des Rücklaufs für das Screening und die Haupterhebung je Schicht befindet sich im Anhang in der Tabelle 22.

4.2 Zusammenlegen Schichten

Die Analyse des Rücklaufs je Schicht in der Haupterhebung zeigte, dass es in einigen wenigen Schichten keine oder zu wenig antwortende ausbildende AST gab. Daher wurden diese Schichten für die Hochrechnung mit anderen zusammengelegt. Folgende in der Tabelle 11 aufgeführte Schichten sind von dieser Zusammenlegung betroffen:

Tabelle 11: Zusammenlegen von Schichten

Alte Schicht Nr. (schichtnr)	Nogaklasse (unterab)	Grössenklasse (grkllehr)	Neue Schichtnr. (schi_neu)
53	G50	3	53055
54	G50	4	53055
55	G50	5	53055
82	K7071	2	82085
83	K7071	3	82085
84	K7071	4	82085
85	K7071	5	82085
89	K72	4	89090
90	K72	5	89090

Durch diese Zusammenlegung wird gewährleistet, dass es je Schicht mindestens zwei ausbildende AST gibt, welche geantwortet haben. Dadurch kann also eine Varianz je Schicht geschätzt werden.

4.3 Direktkontakte

Die Direktkontakte sollten gleich behandelt werden wie bei der letzten Erhebung (siehe Potterat [3]). Zur Erinnerung: die Direktkontakte sind grosse Unternehmen, für welche es einfacher war, die Fragebogen für das ganze oder einen Teil des Unternehmens auszufüllen, als nur für die Arbeitsstätten aus der Stichprobe. Im Datensatz können die Antworten von Direktkontakten anhand einer fiktiven Burnummer identifiziert werden. Zusätzlich gibt eine Referenzvariable an, für welche Einheiten die Antwort gültig ist.
So wurde für diese Unternehmen wiederum zwei neue Schichten gebildet: Die Schicht 'XX' enthält die Antworten der Direktkontakte (genau die fiktiven Burnummern der blauen und gelben Fragebogen), welche auch ein positives Gewicht erhalten werden. Die Schicht 'ZZ' enthält die Informationen / AST, welche für die Hochrechnung nicht relevant ist und daher ein Gewicht von 0 erhalten wird. Es sind dies die Fragebogen, welche nur für das Screening gültig sind (mit fiktiver Burnummern $< 1'000'000$), die Antworten der weissen Fragebogen (da diese Informationen bereits auf die blauen Fragebogen übertragen wurden) sowie alle in der Schicht 'XX' referenzierten AST.
Die Information aus 'ZZ' werden verwendet, um die Populationstotale der bestehenden 'normalen' Schichten neu zu berechnen (Variable `AnzAst_neu`).

Nach dem Zusammenlegen der Schichten sowie der Behandlung der Direktkontakte gab es neu (115 - 6 + 2 =) 111 Schichten.

5 Gewichtung

5.1 Notationen

In der Tabelle 12 sind die wichtigsten Notationen aufgeführt, welche hilfreich sind, um die Theorie der Gewichtung beschreiben zu können.

Tabelle 12: Notationen: Anzahl Arbeitsstätten

	Total =	ausbildende AST (a)	+	AST ohne Lehrlinge (k)
Population (je Schicht)	$N_h =$	$\hat{N}_h^a$	+	$\hat{N}_h^k$
Stichprobe Brutto (je Schicht)	$n_h =$	$\hat{n}_h^a$	+	$\hat{n}_h^k$
Antworten Screening (je Schicht)	$r_{h,Sc} =$	$r_{h,Sc}^a$	+	$r_{h,Sc}^k$
Antworten Haupterhebung (je Schicht)	$r_h =$	r_h^a	+	r_h^k
In der Population und der Brutto Stichprobe ist die Aufteilung der AST in ausbildend / nicht ausbildend unbekannt! Das Symbol ˆ über den Anzahlen kennzeichnet die geschätzten Werte.				

5.2 Initialgewichtung

Das Initialgewicht $w_{i,h}^{(1)}$ einer AST i in der Schicht h berücksichtigt die Einschlusswahrscheinlichkeit ($d_{i,h}$) der AST sowie die globale Antwortquote (q_h) in der Haupterhebung und wird berechnet mit:

$$w_{i,h}^{(1)} = d_{i,h}^{-1} \times q_h^{-1} = \frac{N_h}{n_h} \times \frac{n_h}{r_h} = \frac{N_h}{r_h} \equiv w_h^{(1)} \tag{1}$$

Dieses Gewicht ist innerhalb einer Schicht also identisch für alle AST - egal ob ausbildend oder nicht.

5.3 Nachgewichtung

Eine Nachgewichtung der Initialgewichte, d.h. eine Anpassung / Korrektur derselben, kann von Vorteil sein, wenn die Antwortquoten in gewissen Untergruppen unterschiedlich sind und eine Korrelation zwischen diesen Untergruppen und den interessierenden Zielgrössen (Kosten / Nutzen) vermutet wird.
Eine erste Nachgewichtung ist für die Unterscheidung in ausbildenden bzw. nicht ausbildenden AST je Schicht angebracht. Es werden zwei unterschiedliche Gewichte definiert:

$$w_{i,h}^{(2),a} = w_h^{(2),a} = \frac{\hat{N}_h^a}{r_h^a}, \qquad w_{i,h}^{(2),k} = w_h^{(2),k} = \frac{\hat{N}_h^k}{r_h^k}$$

Eine einfache und gute Methode die benötigten Populationswerte zu schätzen, ist die Information aus dem Screening auszunutzen, da dort die Stichprobe grösser und die Antwortrate besser ist als in der Haupterhebung. Mit $\hat{N}_h^a = \lfloor \frac{r_{h,Sc}^a}{r_{h,Sc}} N_h \rfloor$ und $\hat{N}_h^k = \lfloor \frac{r_{h,Sc}^k}{r_{h,Sc}} N_h \rfloor$, wobei mit $\lfloor \ldots \rfloor$ die gerundeten Werte gemeint sind, werden die Populationsgrössen geschätzt. Daraus lassen sich die ersten Nachgewichte berechnen mit

$$w_{i,h}^{(2),a} = \frac{\hat{N}_h^a}{r_h^a} = \frac{\lfloor \frac{r_{h,Sc}^a}{r_{h,Sc}} N_h \rfloor}{r_h^a} \tag{2}$$

$$w_{i,h}^{(2),k} = \begin{cases} 0, & \text{falls } r_h^k = 0 \\ \frac{\hat{N}_h^k}{r_h^k} = \frac{\lfloor \frac{r_{h,Sc}^k}{r_{h,Sc}} N_h \rfloor}{r_h^k}, & \text{falls } r_h^k > 0 \end{cases} \tag{3}$$

Durch diese erste Nachgewichtung wird also berücksichtigt, dass die nicht ausbildenden AST schlechter geantwortet haben als die ausbildenden.
Aus der jährlichen Statistik (siehe BFS [6]) ist bekannt, wie viele Lehrlinge je Beruf in der ganzen Schweiz ausgebildet werden. Mittlerweile lagen also schon die Zahlen vom Ausbildungsjahr 2004 vor. Diese Angaben sollten für eine zweite Nachgewichtung verwendet werden. Konkret wurden mit einer Ausnahme die am häufigsten ausgebildeten Berufe mit mehr als 2000 Lehrlingen verwendet (cf. Tabelle 13). Die Ausnahme bildete der mit 3096 Lehrlingen eigentlich auf Rang 11 'klassierte' Gärtner, da die Gartenbaubetriebe des ersten Sektors nicht im Stichprobenrahmen enthalten waren. So wurde der Gärtner und die restlichen Berufen, welche mit wenigen Ausnahmen alles BBG-reglementierte Berufe sind, in der Kategorie 24 'übrige Berufe' zusammengefasst.

Tabelle 13: Berufe zur Nachgewichtung

Nr	Berufscode	Beruf	Total Lehrlinge 2004
1	68200	Kaufmann (B+ E + kfm. Ang.)	31'456
2	47412	Elektromonteur/in	7'841
3	70400	Verkäufer/in	7'352
4	45702	Polymechaniker/in	7'092
5	70200	Detailhandelsangestellte/r	6'751
6	47110	Informatiker/in	6'262
7	79003	Koch/Köchin	5'096
8	30503	Schreiner/in	5'041
9	46304	Automechaniker/in	4'811
10	82010	Coiffeur/-euse	4'217
11	46306	Automonteur/in	2'970
12	64004	Hochbauzeichner/in	2'881
13	51002	Maurer/in	2'862
14	47408	Automatiker/in	2'655
15	53001	Maler/in	2'556
16	30302	Zimmermann/Zimmerin	2'510
17	86906	Fachangestellte/r Gesundheit	2'505
18	70402	Pharma-Assistent/in	2'500
19	95503	Logistikassistent/in	2'408
20	47702	Sanitärmonteur/in	2'336
21	86903	Medizinische/r Praxisassistent/in	2'301
22	86905	Dentalassistent/in	2'237
23	46502	Elektroniker/in	2'223
24		übrige Berufe	56'156
		Total Population Lehrlinge	175'019

Mit dem SAS Macro 'Calmar2' können mit der Methode 'iterative proportional fitting' für die **ausbildenden AST** kalibrierte Gewichte $w_{i,h}^{cal}$ berechnet werden. Bei dieser Methode wird die Pseudo-Distanzfunktion $G(x) = x\,log(x) - x + 1$ gewählt, um die Distanz zwischen den Initial und den kalibrierten Gewichten zu bestimmen, d.h. $x = w_{i,h}^{cal}/w_{i,h}^{(2),a}$. Die kalibrierten Gewichte minimieren die Summe der gewichteten Distanzen:

$$\min_{w_{i,h}^{cal}} \sum_{i \in S} w_{i,h}^{(2),a} G(w_{i,h}^{cal}/w_{i,h}^{(2),a})$$

unter der Nebenbedingung $\sum_{i \in S} w_{i,h}^{cal} y_i = X$
Diese kalibrierten Gewichte entstehen im Prinzip durch eine Multiplikation der Initialgewichte $w_{i,h}^{(2),a}$ mit einem Faktor $c(g,h)$:

$$w_{i,h}^{cal} = w_{i,h}^{(2),a} \times c(g,h),$$

wobei der Faktor $c(g,h)$ abhängig von der Nachgewichtungsgruppe g (hier g = Beruf) sowie der Schicht h ist und macht, dass folgende Bedingungen erfüllt werden:

$$\sum_{i \in g} w_{i,h}^{cal}\ y_i = \sum_{i \in U_g} x_i\ \forall g. \quad \text{Sowie}$$
$$\sum_{g} \sum_{i \in g} w_{i,h}^{cal}\ y_i = \sum_{i \in U} x_i \equiv 175'019.$$

Die Variable y_i ist die Anzahl Lehrlinge im ausgewählten Beruf einer AST i gemäss Haupterhebung (Variable `anz001pt` im Datensatz) und $\sum x_i = X$ die Anzahl Lehrlinge aus der Lehrlingsstatistik.

Neben dem Beruf wurden noch andere Nachgewichtungsvariablen verwendet. Die Tabelle 14 zeigt eine Übersicht aller Variablen.

Tabelle 14: Alle Variablen für die Nachgewichtung

Nr	Variable	Variablentyp für Calmar2
1	Beruf	24 Variablen numerisch
2	Schichtgrössen: Anzahl AST in der Population je (zusammengefasste) Schicht	Charakter mit 69 Levels
3a	geschätzte Anzahl ausbildende resp. nicht ausbildende AST Total	Charakter mit 2 Levels
3b	geschätzte Anzahl ausbildende AST je (zusammengefasste) Schicht	69 Variablen numerisch

Bemerkungen:

- Eine Nachgewichtungsvariable vom Typ 'charakter' darf im Makro Calmar2 höchsten 99 Ausprägungen haben. Daher wurden für die Kalibrierung die voll erhobenen Schichten je Nogaklasse (`unterab`) zusammengefasst (Reduktion von 111 auf 69 Schichten).

- Es wurde beschlossen, dass die AST der Direktkontakten ein Hochrechnungsgewicht von 1 haben sollen. Daher wurden diese aus der Nachgewichtung ausgeschlossen. Das bedingte, dass die Variable 'Beruf' für die Nachgewichtung der restlichen Schichten modifiziert wurde (Subtrahieren der Lehrlinge aus den Direktkontakten).

- Da es mit der Kalibrierung mit Calmar2 auf die vier Variablen möglich ist, dass es Gewichte einiges kleiner als 1 geben kann (z.B. 0.003), wurde im Calmar die Option 'PONDQK=q_i^{-1}' gesetzt mit $q_i^{-1} = w_{i,h}^{(2),a}/[1-(1/w_{i,h}^{(2),a})]$. Dies entspricht der Brewer-Methode und gewährleistet, dass die ganz kleinen und die ganz grossen Gewichte durch die Nachgewichtung nicht stark verändert werden.

Das Gewicht für die **nicht ausbildenden** AST in der Stichprobe wurde bei den Nachgewichtungen mit Calmar2 nicht mehr verändert, da die Information aus der Population - nämlich die geschätzte Anzahl nicht ausbildende AST - bereits bei der ersten Nachgewichtung korrekt berücksichtigt wurde.

Zusammenfassend lässt sich also das definitive Gewicht für die AST i in der Schicht h - notiert mit $w_{i,h}$ - wie folgt schreiben:

$$w_{i,h} = \begin{cases} w_{i,h}^{cal}, & \text{falls AST ausbildend und Schicht} \neq \text{(XX, ZZ)} \\ w_{i,h}^{(2),k}, & \text{falls AST nicht ausbildend und Schicht} \neq \text{(XX, ZZ)} \\ 1, & \text{falls Schicht = XX} \\ 0, & \text{falls Schicht = ZZ} \end{cases} \tag{4}$$

Durch die Nachgewichtung ist es im Gegensatz zur Initialgewichtung der Fall, dass die Gewichte der AST innerhalb einer Schicht unterschiedlich sind.

5.4 Analyse Verteilung der Gewichte

Um den Einfluss der Nachgewichtung auf die Gewichte beurteilen zu können, sollte die Verteilung vor und nach der Kalibrierung betrachtet werden. Die Tabelle 15 zeigt die Verteilung der Initialgewichte, der kalibrierten Gewichte, sowie des Quotienten der beiden.

Tabelle 15: Verteilung der Gewichte

	Initialgewichte `gewichtGlobal`	Kalibrierte Gewichte `Gewicht`	Quotient: `Gewicht/ gewichtGlobal`
100% Max	299.000	702.364	9.320
99%	214.625	214.625	2.267
95%	159.857	163.870	1.153
90%	137.000	155.794	1.044
75% Q3	106.476	106.969	1.000
50% Median	49.768	46.618	1.000
25% Q1	6.500	6.310	0.951
10%	4.771	4.478	0.847
5%	3.950	3.501	0.757
1%	1.000	1.000	0.502
0% Min	1.000	1.000	0.157
Mean	59.733	59.733	1.001
Std Deviation	56.367	58.044	0.348
CV	0.944	0.972	0.348
Max/Min	299.000	702.364	59.307
Sum Weights	255’416	255’416	4’281
Sum Observations	4’276	4’276	4’276

Die Nachgewichtung hat also die Eigenschaft, dass es die Initialgewichte verändert. Aus der Tabelle 15 ist aber zu sehen, dass es nur kleine Veränderungen sind, welche nicht stören. Im Gegenteil gibt es ja wie erwähnt den grossen Vorteil, dass mit diesen definitiven Gewichten gewährleistet wird, dass gewisse Anzahlen (Lehrlinge je Berufe, Anzahl ausbildende AST) unverfälscht geschätzt werden.

6 Schätzverfahren

6.1 Allgemeines

Die Schätzung der Resultate für die Kosten und Nutzen der Berufsbildung erfolgt im Wesentlichen analog zur letzten Erhebung im Jahr 2000 (siehe Kapitel 2 des Methodenberichts von Potterat [3]). Die Formeln für die neuen Berechnungen weichen im Vergleich dazu lediglich in einem Punkt dazu ab, nämlich dass es kein identisches Gewicht innerhalb einer Schicht mehr gibt. Daher werden im Abschnitt 6.2 die wichtigsten Formeln nochmals aufgeführt, allerdings nur für die Punktschätzer.
Nach erfolgten Imputationen oder Nachgewichtungen müssten zur Schätzung der Varianzen streng gesehen spezielle Methoden wie Jackknife oder Bootstrap angewandt werden. In unserem Fall werden aber mangels der zur Verfügung stehenden Prozeduren die

Standard-Formeln verwendet, wobei die entsprechenden Formeln im zitierten Bericht oder der Sampling-Literatur (cf. Cochran [4] oder Särndal [5]) konsultiert werden können.

6.2 Formeln für Punktschätzer

Für die Berechnung der Punktschätzer sind folgende in der Tabelle 16 aufgelistete Notationen von Nutzen.

Tabelle 16: Notationen

j, $j = 1, \ldots, p$	Index fürs j-te Lehrjahr, $p = \{2, 3 \text{ oder } 4\}$
i	Index für die AST.
d	Index für den Beruf.
$\bar{y}_{ij,d}$	Die Durchschnittskosten pro Lehrling im j-ten Lehrjahr mit Beruf d der i-ten AST.
$x_{ij,d}$	Die Anzahl Lehrlinge im j-ten Lehrjahr für die i-te AST im Beruf d.
$y_{ij,d} = \bar{y}_{ij,d}\, x_{ij,d}$	Totale Kosten der Lehrlinge im j-ten Lehrjahr der i-ten AST (Beruf d).
$y_{i,d} = \sum_j y_{ij,d}$	Totale Kosten der i-ten AST (Beruf d).
$x_{i,d} = \sum_j x_{ij,d}$	Anzahl Lehrlinge der i-ten AST (Beruf d).
$w_{i,h}$	Kalibriertes Hochrechnungsgewicht der i-ten AST in der Schicht h.

In der Tabelle 17 sind die Variablen aufgelistet, welche von Interesse sind und geschätzt werden sollen. Die Variable 'Kosten' steht dabei immer stellvertretend für eine der interessierenden Zielgrössen (Bruttokosten, Nutzen, Lehrlingslöhne etc.).

Tabelle 17: Variablen von Interesse

Lehrjahr (j)	Schätzgrösse	Beruf (d) 1	2	3	...	24	Total
1	Anzahl Lehrlinge	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	...	$X_{1,24}$	$X_{1,\cdot}$
	Total Kosten	$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	...	$Y_{1,24}$	$Y_{1,\cdot}$
	ø-Kosten pro Lehrling	$Q_{1,1}$	$Q_{1,2}$	$Q_{1,3}$	...	$Q_{1,24}$	$Q_{1,\cdot}$
2	Anzahl Lehrlinge	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	...	$X_{2,24}$	$X_{2,\cdot}$
	Total Kosten	$Y_{2,1}$	$Y_{2,2}$	$Y_{2,3}$	...	$Y_{2,24}$	$Y_{2,\cdot}$
	ø-Kosten pro Lehrling	$Q_{2,1}$	$Q_{2,2}$	$Q_{2,3}$	...	$Q_{2,24}$	$Q_{2,\cdot}$
3	Anzahl Lehrlinge	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	...	$X_{3,24}$	$X_{3,\cdot}$
	Total Kosten	$Y_{3,2}$	$Y_{3,2}$	$Y_{3,3}$	...	$Y_{3,24}$	$Y_{3,\cdot}$
	ø-Kosten pro Lehrling	$Q_{3,1}$	$Q_{3,2}$	$Q_{3,3}$	...	$Q_{3,24}$	$Q_{3,\cdot}$
4	Anzahl Lehrlinge			$X_{4,3}$	...	$X_{4,24}$	$X_{4,\cdot}$
	Total Kosten			$Y_{4,3}$	...	$Y_{4,24}$	$Y_{4,\cdot}$
	ø-Kosten pro Lehrling			$Q_{4,3}$	...	$Q_{4,24}$	$Q_{4,\cdot}$
Total	ø-Kosten pro Lehrling für die ganze Lehre	$Q_{\cdot,1}$	$Q_{\cdot,2}$	$Q_{\cdot,3}$	...	$Q_{\cdot,24}$	-
	Total Kosten pro Lehrjahr pro Beruf	$Y_{\bar{\cdot},1}$	$Y_{\bar{\cdot},2}$	$Y_{\bar{\cdot},3}$	...	$Y_{\bar{\cdot},24}$	$Y_{\bar{\cdot},\cdot}$
	Total Kosten über alle Lehrjahre pro Beruf	$Y_{\cdot,1}$	$Y_{\cdot,2}$	$Y_{\cdot,3}$	...	$Y_{\cdot,24}$	$Y_{\cdot,\cdot}$

Die monatlichen Durchschnittskosten pro Lehrjahr und Lehrling für einen Beruf $Q_{j,d}$ werden geschätzt mit

$$\hat{Q}_{j,d} = \frac{\hat{Y}_{j,d}}{\hat{X}_{j,d}} = \frac{\sum_h \sum_i w_{i,h}\, \bar{y}_{ij,d}\, x_{ij,d}}{\sum_h \sum_i w_{i,h}\, x_{ij,d}} \;, \quad \forall j \in \{1,\ldots,p\} \quad \text{und} \quad d \in \{1,\ldots,24\} \tag{5}$$

Das Produkt $\bar{y}_{ij,d}x_{ij,d}$ im Zähler wird für eine AST auf Null gesetzt, die keine Angaben über den interessierenden Beruf d hat oder gar keine Lehrlinge ausbildet. (Im SAS kann dazu eine zusätzliche Indikatorvariable als Domainvariable angegeben werden).

Die Durchschnittskosten pro Lehrling für die ganze Lehre $Q_{\cdot,d}$ wird geschätzt als Summe der Quotienten der einzelnen Lehrjahre, d.h.

$$\hat{Q}_{\cdot,d} = \sum_{j=1}^{p} \hat{Q}_{j,d} = \frac{\hat{Y}_{1,d}}{\hat{X}_{1,d}} + \ldots + \frac{\hat{Y}_{p,d}}{\hat{X}_{p,d}} \;, \quad \forall d \in \{1,\ldots,24\} \tag{6}$$

Die Anzahl der Summanden ist also abhängig von der Lehrdauer des entsprechenden Berufes.
Analog könnten die Durchschnittskosten je Lehrjahr über alle Berufe berechnet werden mit:

$$\hat{Q}_{j,\cdot} = \sum_{d} \hat{Q}_{j,d} = \frac{\hat{Y}_{j,1}}{\hat{X}_{j,1}} + \ldots + \frac{\hat{Y}_{j,d}}{\hat{X}_{j,d}} \;, \quad \forall j \in \{1,\ldots,4\} \tag{7}$$

Diese Grösse ist aber eher weniger von Interesse. Mehr interessieren schon die Totalkosten über die ganze Schweiz nach Beruf oder insgesamt. Diese werden wie folgt geschätzt:

$$\hat{Y}_{\cdot,d} = \sum_h \sum_i w_{i,h} \big(\sum_j \bar{y}_{ij,d}\, x_{ij,d} \big) = \sum_h \sum_i w_{i,h} y_{i,d} \;, \quad \forall d \in \{1,\ldots,24\} \tag{8}$$

sowie

$$\hat{Y}_{\cdot,\cdot} = \sum_d \hat{Y}_{\cdot,d} \tag{9}$$

Daraus lassen sich weitere Grössen schätzen:

$$\hat{\bar{Y}}_{\cdot,d} = \hat{Y}_{\cdot,d}/p \tag{10}$$

$$\hat{\bar{Y}}_{\cdot,\cdot} = \sum_d \hat{\bar{Y}}_{\cdot,d} \tag{11}$$

$$\hat{Q}_{\cdot,\cdot} = \hat{Y}_{\cdot,\cdot}/\hat{X}_{\cdot,\cdot} = \hat{Y}_{\cdot,\cdot}/175'019 \tag{12}$$

6.3 Vergleich SAS versus STATA

Im BFS wurden als Test einige Berechnungen mit der Statistiksoftware SAS (Version 9.1.3 auf Unix/GAUSS) durchgeführt. Dazu wurde das Makro smsub.sas (Quelle cf. Tabelle 20) verwendet, welches als Ergänzung von `proc surveymeans` dient. Die offiziellen Resultate wurden an der Universität Bern mit der Software STATA Version 7 berechnet. In beiden Softwares gibt es vordefinierte Prozeduren, womit Resultate aus Stichproben inklusive Varianzen geschätzt werden können. Die Darstellung in Tabelle 18 zeigt, welche Optionen in beiden Softwares gesetzt werden müssen, um dieselben, richtigen Resultate zu erhalten.

Tabelle 18: Vergleich SAS mit STATA

Funktion	**SAS**	**STATA**
a) Quotient	%smsub(data=knbb04, statistics=nobs ratio seratio /noprint, popsize=AnzAst_neu, weight=Gewicht, strata=schi_neu, subpop= flag_1=1, ratio=Lohn_tot1/anz001p1, tables=Beruf, output= tables= lohn_1);	svyset strata schi_neu svyset pweight Gewicht svyset fpc AnzAst_neu svyratio Lohn_tot1/anz001p1, subpop(flag_1) by(Beruf) ci obs size
b) Mittelwert	%smsub(data=knbb04, statistics=nobs mean semean /noprint, popsize=AnzAst_neu, weight=wh_1, strata=schi_neu, subpop= flag_1=1, var=Lohn_lehr_1, tables=Beruf, output= tables=lohn_1b);	svyset strata schi_neu svyset pweight wh_1 svyset fpc AnzAst_neu svymean Lohn_lehr_1, subpop(flag_1) by(Beruf) ci obs size
c) Total	%smsub(data=knbb04, statistics=nobs total setotal /noprint, popsize=AnzAst_neu, weight=Gewicht, strata=schi_neu, subpop= ausb=1, var=Lohn_tot_t, tables=Beruf, output= tables=lohn_tot);	svyset strata schi_neu svyset pweight Gewicht svyset fpc AnzAst_neu svytotal Lohn_tot_t, subpop(ausb) by(Beruf) ci obs size
d) Linearkombinationen von Schätzern	nicht möglich	svylc [1]1 + [2]1 + [3]1 (für Kaufmann: Beruf=1, Lehrdauer=3)

Bemerkungen / Erläuterungen:

- In den Beispielen a) und b) werden auf zwei unterschiedliche Arten die Durchschnittslöhne im ersten Lehrjahr pro Beruf ($\hat{Q}_{1,d}$) geschätzt. Bei a) werden für jede AST die Lohnsumme für alle Lehrlinge (Lohn_tot1) und die Anzahl Lehrlinge (anz001p1) als 'ratio-Variablen' eingegeben. Bei b) wird hingegen der Durchschnittslohn pro Lehrling (Lohn_lehr_1) eingesetzt, dafür wird eine modifizierte

Gewichtungsvariable verwendet: wh_1= Gewicht $\times$ anz001p1. Mit beiden Methoden werden aber identische Resultate geschätzt.

- Die Variable schi_neu kennzeichnet die Schichtzugehörigkeit für jede AST.
- Die Variable AnzAst_neu gibt pro AST die Schichtgrösse der Population an, welches für die Endlichkeitskorrektur in der Varianzschätzung benötigt wird.
- Mit der 'noprint'- Option kann der Output unterdrückt werden. Als Ersatz werden die Resultate je Beruf mit der 'output= tables='- Option in eine SAS-Datei geschrieben.
- Mit der 'subpop='- Option wird die Auswertung nur für eine Subpopulation berechnet, nämlich die AST mit mindestens einem Lehrling im ersten Lehrjahr (flag_1=1) bzw. für alle ausbildenden AST (ausb=1).
- Bei c) werden alle in der Schweiz ausbezahlen Lehrlingslöhne über alle Lehrjahre pro Beruf geschätzt ($\hat{Y}_{\cdot,d}$).
- Im STATA werden zuerst mit der 'svyset='-Funktion die Schichten, die Gewichte und die Endlichkeitskorrektur definiert, danach können mit einer der Funktionen svyratio, svymean oder svytotal die Resultate geschätzt werden.
- Um die Varianzen der Grössen $\hat{Q}_{\cdot,d}$ als Summe von Schätzern über die einzelnen Lehrjahre korrekt zu schätzen, gibt es im STATA die Funktion svylc. Eine solche Funktion gibt es im SAS leider (noch) nicht.

7 Provisorische Resultate

Die Tabelle 19 zeigt einige provisorische Resultate der Variablen Lehrlingslohn unterteilt nach den ausgewählten Berufen und den Lehrjahren. Zu jedem geschätzten Durchschnittslohn (Variable: ø-Lohn) ist die Genauigkeit mit der halben Länge eines 95% Vertrauensintervalls (Variable: halb-VI) angegeben. Diese Länge entspricht der doppelten (genau der 1.96fachen) Standardabweichung der Schätzung.

Die definitiven und offiziellen Resultaten, welche von den hier präsentierten provisorischen Resultaten abweichen können und dann nicht nur die Löhne, sondern alle Kosten aber auch die Nutzen der Lehrlinge für die Unternehmen zeigen werden, erscheinen im Verlaufe des Jahres 2006.

Tabelle 19: Provisorische Resultate: Lehrlingslöhne[1]

BERUF	Berufname	Stichprobe		Population	Lehrlinge 1. Lehrjahr			Lehrlinge 2. Lehrjahr			Lehrlinge 3. Lehrjahr			Lehrlinge 4. Lehrjahr			Total alle Lehrjahre	
		♯ AST	♯ Lehrlinge	♯ Lehrlinge	♯ AST	ø-Lohn	halb-VI	♯ AST	ø-Lohn	halb-VI	♯ AST	ø-Lohn	halb-VI	♯ AST	ø-Lohn	halb-VI	ø-Lohn	halb-VI
0	Total	2'413	13'454	175'019	1'702	651.38	32.97	1'479	843.06	39.04	1'316	1'107.63	47.76	460	1'172.68	22.34	-	-
1	Kaufmann[2]	600	4'855	31'456	443	668.05	9.73	402	859.18	11.06	347	1'184.33	16.00				2'711.5	26.98
2	Elektromonteur	85	848	7'841	67	482.53	24.99	65	649.24	26.69	64	857.58	24.89	63	1'173.09	44.71	3'162.4	82.59
3	Verkäufer	45	114	7'352	35	654.49	49.32	28	835.32	55.92							1'489.8	88.28
4	Polymechaniker	107	992	7'092	86	551.78	18.27	78	737.29	26.97	80	977.64	39.70	79	1'230.05	48.18	3'496.8	112.07
5	Detailhandelsangest.	65	558	6'751	47	623.06	44.15	46	825.98	41.11	45	1'079.19	54.30				2'528.2	108.95
6	Informatiker	78	945	6'262	39	639.88	28.39	52	832.36	45.26	46	1'090.67	49.01	57	1'340.91	60.42	3'903.8	144.24
7	Koch	137	309	5'096	102	992.83	27.36	82	1'179.36	27.76	64	1'436.08	38.98				3'608.3	75.48
8	Schreiner	51	131	5'041	30	537.59	32.23	29	691.58	60.93	23	977.33	103.52	20	1'183.85	60.93	3'390.3	154.25
9	Automechaniker	34	312	4'811	27	542.76	37.81	24	683.38	34.88	22	907.59	62.11	25	1'182.43	88.38	3'316.2	167.40
10	Coiffeur	26	68	4'217	22	390.90	14.29	17	570.89	144.45	19	721.06	191.19				1'682.8	333.64
11	Automonteur	26	43	2'970	11	539.04	55.11	13	652.12	52.20	15	897.10	54.71				2'088.3	115.51
12	Hochbauzeichner	47	85	2'881	19	436.53	80.97	24	565.24	59.10	15	711.69	79.52	20	976.24	29.84	2'689.7	135.98
13	Maurer	66	257	2'862	58	997.91	62.97	44	1'332.37	79.05	43	1'976.29	154.46				4'306.6	216.72
14	Automatiker	23	165	2'655	18	507.65	80.65	19	712.89	95.34	17	901.96	88.71	16	1'174.54	163.14	3'297.1	411.42
15	Maler	34	86	2'556	21	574.57	39.47	21	789.85	53.27	15	1'317.75	110.69				2'682.2	167.40
16	Zimmermann	29	95	2'510	19	699.92	74.68	23	1'011.84	102.80	19	1'329.80	83.03				3'041.6	197.05
17	Fachangest. Gesundheit	55	222	2'505	53	673.45	118.24	30	908.82	25.32	6	1'298.87	66.05				2'881.1	141.94
18	Pharma-Assistent	27	62	2'500	20	554.49	27.37	15	776.05	19.68	19	1'026.13	33.96				2'356.7	63.16
19	Logistikassistent	28	656	2'408	19	636.45	57.61	14	839.22	115.88	18	1'107.53	106.95				2'583.2	262.47
20	Sanitärmonteur	19	46	2'336	14	594.59	17.20	12	785.90	21.02	10	1'030.25	48.65				2'410.7	56.63
21	Mediz. Praxisassistent	42	78	2'301	21	385.56	81.54	16	819.93	54.09	25	1'227.19	55.25				2'432.7	143.92
22	Dentalassistent	37	51	2'237	22	511.24	13.65	13	858.99	22.41	13	1'199.88	59.56				2'570.1	70.98
23	Elektroniker	23	163	2'223	15	551.35	14.45	17	701.31	44.93	19	905.84	50.79	17	1'182.80	91.63	3'341.3	178.11
242	übrige mit 2 Lehrjahren	33	64	1'385	26	1'031.09	27.49	27	1'251.09	26.15							2'282.2	49.13
243	übrige mit 3 Lehrjahren	400	1'211	32'579	283	770.37	113.12	218	969.61	130.56	208	1'223.74	178.04				2'963.7	395.03
244	übrige mit 4 Lehrjahren	296	1'038	22'193	185	521.70	25.80	150	700.05	46.71	164	902.33	45.11	163	1'125.09	38.62	3'249.2	106.49

[1] Monatslöhne Brutto inkl. Sozialversicherungsbeiträge Arbeitnehmer und monatlich ausbezahlte Zulagen, exkl. unregelmässig ausbezahlte Lohnbestandteile (13./14. Monatslohn, Gratifikationen etc.)

[2] Beim Kaufmann wurde die Basisausbildung B, die erweiterte Grundausbildung E sowie die 'alte' Berufsbezeichnung 'kaufm. Angestellter' KV zusammengefasst.

Anhang

A SAS-Programme und Daten

Folgende in der Tabelle 20 aufgeführten SAS-Programme, alle gespeichert auf PC im Verzeichnis `'...\Projekte\Beruf04\SAS-Programme'` sowie die SAS-Daten gespeichert auf dem Unix Server GAUSS unter `'.../Projekte/Beruf04'` wurden zur Planung und Ziehung der Stichprobe sowie zur Gewichtung und Schätzung der Resultate verwendet:

Tabelle 20: Verwendete Daten und Programme

Programme / Daten	**Beschreibung**
AnalyseKNBB00.sas	Analyse der letzten Erhebung
BearbeitungBUR.sas	Bearbeitung BUR-Auszug, Lehrlinge der BZ01, Erstellung des Stichprobenrahmens
StplanBeruf04.sas	Erstellung des Stichprobenplans und Ziehung der Stichprobe
week_loc041030.sas7bdat	BUR-Auszug AST-File vom 30.10.2004
week_ent041030.sas7bdat	BUR-Auszug Unt-File vom 30.10.2004
stprahmen.sas7bdat	Stichprobenrahmen
stplan.sas7bdat	Stichprobenplan
sampleKnbb04.sas7bdat / *.txt	Stichprobe
sampKnbb04_1.txt / *_2.txt	Zweiteilung der Stichprobe zwecks Versand
datensatz050106.sas7bdat	Definitiver Datensatz Screening und Haupterhebung erhalten am 5.1.2006
Gewichtung.sas	Programm zu Erstellung der Gewichtung
Gewicht_AlternX.sas	Tests mit alternativen Gewichtungsmöglichkeiten
Direktkontakte.sas	Behandlung der Direktkontakte, wird von Gewichtung.sas aufgerufen
Hochrechnung.sas	Programm zu Schätzung der Resultate
smsub.sas	SAS-Makro (v.3.0) von sas.com als Ergänzung von `proc surveymeans`, u.a. zur Schätzung von Quotienten innerhalb von Domains.
Gewi_KNBB04.sas7bdat	Definitive Gewichte geliefert an S.Mühlemann am 12.1.2006

B Tabellen

Die Tabelle 22 zeigt eine Übersicht einiger Informationen innerhalb der zusammengefassten Schichten. Die Bedeutung der Spalten ist in der Tabelle 21 beschreiben.
Die Schicht 'XX' enthält wie erwähnt die Direktkontake. Die Schicht 'ZZ', worin die AST sind, dessen Antworten in der Schicht XX enthalten sind, ist hier aufgeteilt in zwei Schichten 'ZZ1': die AST der Stichprobe und 'ZZ2': die AST ausserhalb der Stichprobe, aber im Rahmen. Die Summe der Variablen 'AnzAst_alt' bzw. 'AnzAst_neu' über alle Schichten ist daher gleich gross.

Tabelle 21: Beschrieb von Tabelle 22

Variable	Bedeutung
schi_neu	Schichtnummer
unterab	Nogaklasse
grkla3_ast	zusammengefasste Grössenklassen
AnzAst_alt	Anzahl AST gemäss Plan (zusgef. Schichten)
nsize	Stichprobengrösse
AnzAst_neu	Anzahl Ast nach Elimination der Direktkontakte.
RespScree	Anzahl Antworten Screening
ausbScree	Anzahl Ausbildene AST Screening
anzlscree	Anzahl Lehrlinge Screeningantworten
Response	Anzahl Antworten Haupterhebung
ausbild	davon ausbildende in der Haupterhebung
Anzl	Anzahl Lehrlinge Haupterhebung
respquote	Antwortquote Haupterhebung in Prozent

Tabelle 22: Übersicht Schichtinfos

schi_neu	unterab	grkla3_ast	AnzAst_alt	_nsize_	AnzAst_neu	RespScree	ausbScree	anzlscree	Response	ausbild	Anzl	respquote
0	XX	0			67	85	85	13’896	67	67	5’236	100.00
1	ZZ1	0			1’002							-
2	ZZ2	0			4’619							-
1000	C-DC	1	2’581	442	2’581	329	44	57	30	9	12	6.79
2000	C-DC	2	919	340	919	270	80	163	37	16	31	10.88
3000	C-DC	3	169	169	169	138	77	216	20	14	35	11.83
4000	C-DC	4	141	141	141	122	91	392	24	17	56	17.02
5000	C-DC	5	49	49	49	45	43	543	15	15	73	30.61
6000	DD	1	3’503	417	3’503	326	128	193	35	18	31	8.39
7000	DD	2	841	100	841	75	58	156	11	11	25	11.00
8000	DD	3	40	40	40	34	23	71	6	5	12	15.00
9000	DD	4	5	5	5	4	4	11	2	2	2	40.00
10000	DD	5	21	21	21	20	20	123	5	5	32	23.81
11000	DE	1	2’361	281	2’355	192	38	57	21	9	12	7.47
12000	DE	2	785	95	783	79	47	112	15	9	21	15.79
13000	DE	3	97	97	96	77	52	187	4	3	12	4.12
14000	DE	4	68	68	67	60	44	259	11	11	44	16.18
15000	DE	5	78	78	78	69	67	978	16	15	100	20.51
16000	DF-DI	1	1’374	625	1’374	491	52	55	36	5	5	5.76
17000	DF-DI	2	837	238	837	185	57	167	26	12	24	10.92
18000	DF-DI	3	200	200	200	172	94	284	32	20	41	16.00
19000	DF-DI	4	151	151	147	116	93	508	22	18	62	14.57
20000	DF-DI	5	72	72	69	56	55	813	14	14	121	19.44
21000	DJ	1	4’219	503	4’219	355	67	107	45	13	25	8.95
22000	DJ	2	1’668	199	1’668	153	75	194	25	15	45	12.56
23000	DJ	3	184	184	184	159	99	385	23	16	63	12.50
24000	DJ	4	99	99	98	82	64	380	18	17	46	18.18
25000	DJ	5	103	103	101	88	87	1’108	19	19	116	18.45
26000	DK	1	1’553	357	1’552	271	70	111	33	12	22	9.24
27000	DK	2	836	110	830	92	37	91	16	9	21	14.55
28000	DK	3	169	169	164	137	91	314	24	19	40	14.20
29000	DK	4	99	99	93	81	60	431	18	15	78	18.18
30000	DK	5	178	178	174	152	150	3’758	34	33	361	19.10
31000	DL	1	2’480	309	2’479	240	51	71	36	11	17	11.65
32000	DL	2	1’008	119	1’001	95	26	60	14	2	4	11.76
33000	DL	3	247	247	241	202	100	323	41	25	61	16.60
34000	DL	4	197	197	187	152	105	655	33	27	97	16.75
35000	DL	5	124	124	119	99	96	1’803	35	35	313	28.23
36000	DMDN	1	2’039	238	2’039	186	39	53	24	9	16	10.08
37000	DMDN	2	602	151	602	116	52	137	17	9	22	11.26
38000	DMDN	3	74	74	74	64	34	124	11	6	13	14.86
39000	DMDN	4	53	53	53	45	32	151	4	2	7	7.55
40000	DMDN	5	39	39	39	27	27	665	9	9	151	23.08
41000	E	1	356	310	353	256	53	130	41	18	47	13.23
42000	E	2	323	68	321	49	27	153	10	7	21	14.71
43000	E	3	48	48	40	25	21	149	9	9	38	18.75
44000	E	4	27	27	26	20	18	148	2	2	11	7.41
45000	E	5	33	33	31	26	26	526	9	9	66	27.27
46000	F	1	18’904	2’380	18’888	1’649	477	852	188	88	133	7.90
47000	F	2	6’127	772	6’107	606	405	1’182	107	82	248	13.86

. . . Fortsetzung der Tabelle . . .

schi_neu	unterab	grkla3_ast	AnzAst_Alt	_nsize_	AnzAst_neu	RespScree	ausbScree	anzlscree	Response	ausbild	Anzl	respquote
48000	F	3	418	418	418	349	294	1’154	65	52	146	15.55
49000	F	4	159	159	159	134	121	700	29	29	150	18.24
50000	F	5	382	382	382	321	315	4’381	72	72	639	18.85
51000	G50	1	9’210	860	9’208	638	271	442	82	48	76	9.53
52000	G50	2	1’566	186	1’566	152	123	398	14	13	35	7.53
53055	G50	345	215	215	209	166	148	1’878	24	24	177	11.16
56000	G51	1	10’794	1’072	10’790	794	61	95	88	11	16	8.21
57000	G51	2	3’146	181	3’141	153	44	109	19	5	11	10.50
58000	G51	3	377	377	376	296	179	620	56	38	101	14.85
59000	G51	4	171	171	165	135	90	390	20	15	49	11.70
60000	G51	5	76	76	76	64	61	796	19	19	126	25.00
61000	G52	1	29’470	3’711	29’464	2’395	696	1’353	251	119	207	6.76
62000	G52	2	5’407	681	5’390	422	284	836	72	62	144	10.57
63000	G52	3	381	381	374	155	123	530	19	16	44	4.99
64000	G52	4	110	110	107	56	47	333	7	6	22	6.36
65000	G52	5	252	252	248	125	118	1’250	20	20	152	7.94
66000	H	1	19’505	2’233	19’504	1’320	76	134	129	19	30	5.78
67000	H	2	5’687	651	5’684	447	132	351	65	28	50	9.98
68000	H	3	280	280	276	225	130	498	28	20	37	10.00
69000	H	4	65	65	65	55	37	161	10	7	28	15.38
70000	H	5	149	149	149	128	125	1’084	31	31	113	20.81
71000	I	1	8’517	670	6’166	325	24	38	24	2	3	3.58
72000	I	2	3’353	264	2’177	133	41	79	18	9	17	6.82
73000	I	3	405	405	298	193	92	391	20	10	26	4.94
74000	I	4	324	324	243	103	60	342	12	9	31	3.70
75000	I	5	96	96	71	55	52	882	9	9	74	9.38
76000	J	1	5’322	464	4’573	290	51	101	30	7	17	6.47
77000	J	2	2’211	187	1’783	109	50	157	22	15	51	11.76
78000	J	3	289	289	204	141	52	293	21	9	28	7.27
79000	J	4	275	275	146	96	47	383	13	8	57	4.73
80000	J	5	151	151	112	66	64	1’225	12	12	196	7.95
81000	K7071	1	2’625	301	2’595	250	27	36	20	3	7	6.64
82085	K7071	2345	544	120	531	90	45	203	8	6	14	6.67
86000	K72	1	5’035	601	5’020	471	42	55	62	9	13	10.32
87000	K72	2	985	143	977	118	33	86	31	11	28	21.68
88000	K72	3	96	96	93	75	37	140	21	12	65	21.88
89090	K72	45	68	68	63	51	35	251	9	8	70	13.24
91000	K7374	1	29’305	2’306	29’232	1’788	313	480	287	86	138	12.45
92000	K7374	2	4’204	373	4’170	308	128	343	72	50	111	19.30
93000	K7374	3	354	354	342	270	113	568	46	30	99	12.99
94000	K7374	4	318	318	313	233	106	553	37	21	84	11.64
95000	K7374	5	82	82	75	54	51	752	13	13	67	15.85
96000	L	1	4’509	378	4’499	279	112	355	48	31	64	12.70
97000	L	2	2’670	237	2’665	157	92	316	25	20	48	10.55
98000	L	3	293	293	293	183	104	476	48	35	105	16.38
99000	L	4	171	171	171	105	72	441	24	21	115	14.04
100000	L	5	83	83	83	66	58	677	33	31	361	39.76
101000	M	1	5’505	630	5’495	383	24	33	33	9	16	5.24
102000	M	2	4’100	506	4’091	358	54	112	45	17	34	8.89
103000	M	3	476	476	474	347	92	287	52	27	44	10.92
104000	M	4	252	252	251	168	76	229	32	18	49	12.70
105000	M	5	73	73	73	54	24	478	8	4	21	10.96
106000	N	1	13’811	1’581	13’807	1’237	304	407	193	77	110	12.21
107000	N	2	3’354	384	3’349	302	122	316	70	32	116	18.23
108000	N	3	819	819	817	696	408	1’252	134	93	162	16.36
109000	N	4	463	463	461	397	303	1’504	75	67	165	16.20
110000	N	5	181	181	181	152	133	1’746	42	40	237	23.20
111000	O	1	13’096	1’499	13’094	1’104	197	388	118	38	75	7.87
112000	O	2	2’371	375	2’370	296	64	165	44	16	34	11.73
113000	O	3	155	155	154	115	49	133	25	14	30	16.13
114000	O	4	106	106	106	80	47	158	22	15	38	20.75
115000	O	5	22	22	22	20	17	146	4	2	12	18.18
Total			260’970	40’000	260’970	28’620	10’931	67’772	4’276	2’413	13’454	10.69

Literatur

[1] Renfer, J.-P.: *Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage*. BFS, Neuchâtel, 2002. Bestell-Nr. 338-0012.

[2] *BFS aktuell. Statistik der Lehrverträge, Lehrabschlussprüfungen und Berufsmaturitäten 2003*. BFS, Neuchâtel 2004. Bestell-Nr.: 365-0003.

[3] Potterat J.: *Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren*. BFS, Neuchâtel, 2003. Bestell-Nr.: 338-0016.

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

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

[6] *BFS aktuell, Statistik der Lehrverträge und Lehrabschlussprüfungen 2004*. Neuchâtel, 2005. Bestell-Nr.: 365-0400.

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

Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0036

Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. Bestellnummer: 338-0035

Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0034

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. Bestellnummer: 338-0033

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. Bestellnummer: 338-0032

Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031

Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030

Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029

Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028

Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro de commande: 338-0012

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 info@bfs.admin.ch	Service de renseignements individuels
Das BFS im Internet	www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette/CD-Rom)	032 713 60 60 order@bfs.admin.ch	Publications: information approfondie (certaines sont disponibles sur disquette/CD-Rom)
Online-Datenbank	032 713 60 86 www.statweb.admin.ch	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das laufend nachgeführte Publikationsverzeichnis im Internet unter der Adresse www.statistik.admin.ch→Aktuell→Publikationen.

La liste des publications, mise à jour régulièrement, donne davantage de détails sur les divers moyens de diffusion. Elle se trouve sur Internet à l'adresse www.statistique.admin.ch →Actualités→Publications.

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.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar.

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.

Les rapports de méthodes sont disponibles sous forme électronique sur le site internet de l'OFS.

In diesem Methodenbericht werden die statistischen Methoden (Stichprobenplan und Schätzverfahren) der Erhebung über die Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004 beschrieben.

Aus dem Betriebs- und Unternehmensregister des Bundesamts für Statistik wurde eine nach Wirtschaftsaktivität und Grössenklassen geschichtete Zufallsstichprobe von 40'000 Arbeitsstätten gezogen. Diese Betriebe wurden in zwei Phasen (einem Screening und einer Haupterhebung) befragt, ob sie Lehrlinge ausbilden und falls ja, welche Kosten diese verursachen bzw. Nutzen sie erbringen. Die nicht-ausbildenden Betriebe wurden in einem separaten Fragebogen über die Gründe der Nichtausbildung befragt.

Die Methoden für die Schätzverfahren wurden grösstenteils aus der letzten Erhebung im Jahr 2000 übernommen. Die ermittelte Gewichtung für die antwortenden Betriebe berücksichtigt neben der üblichen Korrektur für die unterschiedlichen Antwortsquoten unter anderem auch die bekannte Verteilung der Berufe gemäss der jährlichen BFS-Lehrlingsstatistik.

Ein Abschnitt widmet sich dem Vergleich der verwendeten Statistik-Softwares SAS und STATA zur Schätzung der provisorischen Resultaten. Am Ende des Berichts sind einige dieser Resultate aufgeführt, genau die Lehrlingslöhne nach Beruf, welche nur einen Teil aller Kosten der Lehrlinge ausmachen. Die offiziellen und definitiven Ergebnisse, welche die Schätzung aller Kosten und Nutzen enthalten werden, erscheinen im Laufe des Jahres 2006.

Bestellnummer: 338-0036
Bestellungen: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Preis: gratis
ISBN 3-303-00331-9