Methodenbericht

Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000

0

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

Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren

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

OFS BFS UST

Neuchâtel, 2003

Statistik der Schweiz

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

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

Statistique de la Suisse

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

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

Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000

Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren

Autor **Beat Hulliger**
Bundesamt für Statistik

Herausgeber **Bundesamt für Statistik**

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

OFS BFS UST

Neuchâtel, 2003

Vorwort

Die Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000 wurde im Jahr 2001 vom Bundesamt für Statistik in Zusammenarbeit mit dem Verband der Schweizer Unternehmen économiesuisse durchgeführt. Der Dienst Statistische Methoden des Bundesamts für Statistik wurde mit der Entwicklung der Schätzverfahren beauftragt. Die Arbeiten wurden im September und Oktober 2001 von Dr. Beat Hulliger durchgeführt. Dabei wurde er von Dr. Ruth Dumitrica, Pierre Zangger und Elisabeth Pastor von der Sektion Hochschule und Wissenschaft begleitet.

Die Angaben über Resultate der Erhebung in diesem Bericht dienen allein zur Illustration der Methoden. Es können sich Abweichungen gegenüber den offiziellen Ergebnissen ergeben, die auf der Wahl der Datenbasis und ihrem Stand, leichten Anpassungen der Methoden, und der Darstellung basieren. Nur die im offiziellen Bericht veröffentlichten Ergebnisse sind massgebend.

Der Autor dankt Dr. Ruth Dumitrica, Elisabeth Pastor und Pierre Zangger für die sehr gute Zusammenarbeit sowie Dr. Jean-Pierre Renfer vom Dienst Statistische Methoden für die guten Unterlagen zum Stichprobenplan und für die aufmerksame Durchsicht des Berichts.

Zusammenfassung

Die Erhebung über Forschung und Entwicklung in der Privatwirtschaft 2000 hat eine hohe Gesamt-Antwortquote erbracht. Der Anteil der Unternehmen, die Forschung und Entwicklung betreiben, ist klein. Umso wichtiger war es, bei den Unternehmen, die F+E betreiben, fehlende Merkmale mit Hilfe von einfachen Modellen zu ersetzen. Der Einfluss dieser Einsetzungen auf die Endresultate ist beträchtlich. Da einige grosse Unternehmensgruppen nur einen Fragebogen für die gesamte Gruppe eingereicht haben, mussten sie mit Hilfe von fiktiven Unternehmen auf die Wirtschaftsbranchen der Erhebung aufgeteilt werden. Die Schätzverfahren sind geschichtete Mittel mit einer Korrektur für die Antwortausfälle. Die Varianzen sind relativ gross. Zwei Grade der Detaillierung der Ergebnisse und die entsprechenden Varianzen werden diskutiert.

Schlüsselwörter

Methodenbericht; Fiktive Einheiten; Datenersatz; Einsetzungen; Merkmalsausfall; Schätzverfahren; Nachschichtung; Varianzschätzung; Forschung; Entwicklung.

Auskunft: Beat Hulliger, Tel. 032 713 60 15
Beat Hulliger@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: 338-0018

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Grafik/Layout: BFS

Copyright: BFS, Neuchâtel 2003
Abdruck – ausser für kommerzielle Nutzung–
unter Angabe der Quelle gestattet.

ISBN: 3-303-00264-9

Inhaltsverzeichnis

1 Einführung **5**

2 Unternehmensgruppen und Nettostichprobe **5**

2.1 Rücklauf, Daten und Ausgangsstichprobe 5

2.2 Unternehmensgruppen in der selben Branche 7

2.3 Unternehmensgruppen mit Zweigunternehmen in verschiedenen Branchen . . . 8

2.4 Nettostichprobe und Berichtigung der Populationstotale 9

3 Datenersatz **10**

3.1 Hauptvariablen 10

3.2 Vollzeitäquivalente 13

4 Schätzverfahren **14**

5 Punkt- und Varianzschätzungen **21**

5.1 Gesamttotale 21

5.2 Branchen-Totale 22

6 Anhang **25**

6.1 SAS-Programme 25

Bibliographie **30**

1 Einführung

Die Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft im Jahr 2000 (FE00) wurde im Jahr 2001 durch das Bundesamt für Statistik, Sektion Hochschule und Wissenschaft, in Zusammenarbeit mit dem Verband der Schweizer Unternehmen économiesuisse durchgeführt. Der Dienst Statistische Methoden entwickelte den Stichprobenplan, erstellte den Stichprobenrahmen aus dem Betriebs- und Unternehmensregister und zog die Stichprobe von 6299 Unternehmen (RENFER 2003). Die Resultate der Erhebung wurden im Dezember 2001 veröffentlicht (DUMITRICA et al. 2001).

In diesem Bericht werden die Probleme der vollständigen und teilweisen Antwortausfälle behandelt und die Erzeugung von fiktiven Einheiten aus den Antworten für Unternehmensgruppen. Die Schätzung der wichtigsten Totale sowie die entsprechenden Varianzschätzungen werden entwickelt. Die Qualität der Ergebnisse sowie der Einfluss der Qualität auf die Möglichkeiten zur Publikation von Ergebnisse für Untersuchungsbereiche werden diskutiert.

Die Arbeiten haben sich dabei auf die wichtigsten Variablen konzentriert (vgl. Kapitel 2.1).

2 Unternehmensgruppen und Nettostichprobe

2.1 Rücklauf, Daten und Ausgangsstichprobe

Die Fragebogen wurden am 28.2.2001 an die Bruttostichprobe der Grösse 6299 versandt. Abgabe-Termin war der 28.3.2001. Eine erste Mahnung wurde am 10.4.2001 versandt und eine zweite am 11.5.2001. Dann wurden vom 20.5.-30.6.2001 telefonische Mahnungen durchgeführt. Das Feld wurde am 12.7.2001 abgeschlossen. Die letzten in der Datenbank berücksichtigten Fragebogen trafen am 5.8.2001 ein.

Bis am 23.3.2001 waren 1915 Fragebogen zurück gekommen . Der weitere Rücklauf ist aus der Figur 1 ersichtlich. Der Effekt der Mahnungen ist sichtbar.

Die globale Antwortquote von 71.5% ist gut (cf. Tabelle 2). Der Rücklauf nach Schichten ist in den Tabellen 6 und 7 dokumentiert, wobei diese Tabellen bereits der umstrukturierten Stichprobe entsprechen. Eine leichte Tendenz zu steigender Antwortquote mit der Grössenklasse kann festgestellt werden: Die kleinsten Unternehmen haben 69.1%, die mittleren 68.4%,die grossen 78.2% und die ganz grossen Unternehmen 82.1% Antwortquote. Die kleinste Antwortquote in einer Schicht ist 49%. Es gibt also keine Schichten, bei denen eine extrem schlechte Antwortquote zu verzeichnen ist.

Die Daten der FE00 lagen als SAS-Datei vor (Meth_que200029sep01_versionb.sas7bdat, erstellt am 29. September 2001 um 22:21:32). Es handelt sich um die Bruttostichprobe mit einem Umfang von $n_b = 6299$ Beobachtungen.

Die methodischen Arbeiten konzentrierten sich in Absprache mit der Sektion HSW auf die Variablen in Tabelle 1.

Die Variable totfep ist die Summe der Fragebogenvariablen I0600F+I0600H+I0600FH (also Forschungs-Personal Frauen, Männer und Geschlecht unbekannt). Um jeweils eine Kontrolle

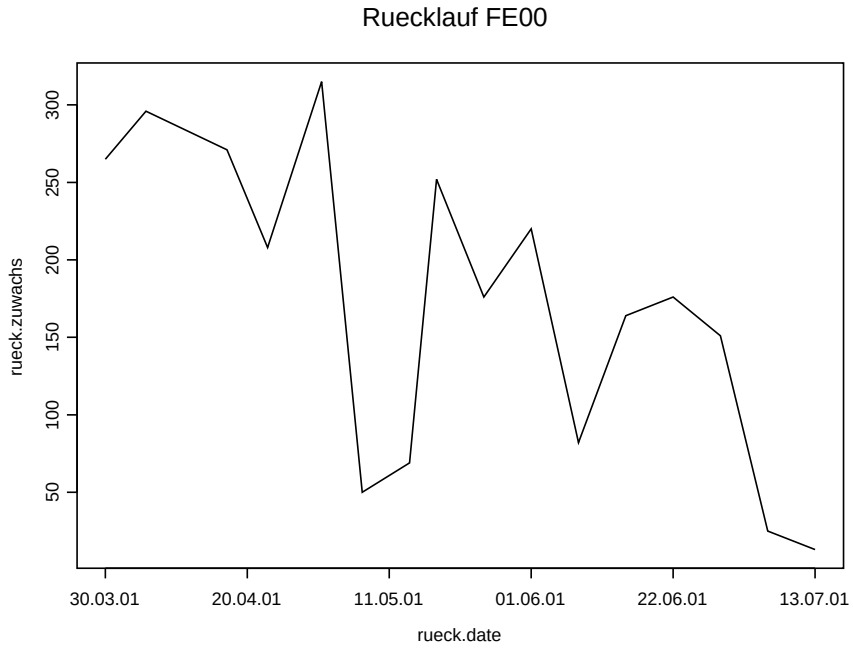


Abbildung 1: Rücklauf FE00 ab 23.3.01

Tabelle 1: Variablen für die Entwicklung der Methoden.

Name	Erläuterung
B0240T	Totaler Aufwand intramuros in der Schweiz
B0244T	Totale Investitionen intramuros in der Schweiz
F0291T	Totaler Aufwand extramuros in der Schweiz
F0276T	Totaler Aufwand extramuros im Ausland
K0205T	Totaler Aufwand “intramuros” im Ausland
totfep	Totale des F+E Personals

über die Grösse der Unternehmen und Branchen zu haben, für eventuelle Quotientenschätzer und als Hilfsgrösse bei der Erzeugung der fiktiven Unternehmen wurde zusätzlich zu den obigen Analysevariablen die Variable DUEMPLOI verwendet. Diese stammt aus dem BUR und gibt die Anzahl Beschäftigte (DUEMPLOI = Vollzeit + Teilzeit1 + Teilzeit 2) des Unternehmens an. Die Variable DUEMPLOI wurde bei der Erstellung des Stichprobenrahmens berechnet.

Die für die Auswertung relevante Branchennomenklatur ist ein Aggregierung von NOGA-Klassen, die durch die Sektion Hochschule und Wissenschaft (HSW) erstellt wurde. Die Stichprobe war nach diesen Branchen und Grössenklassen geschichtet worden (RENFER 2003).

Für die Definition der Nettostichprobe wurden die Variablen c_saisie_xxx benutzt. Der Erfassungsstatus (Variablen “c_saisie_xxx”) wurde für die Auswertung neu kodiert. Aus den 3 Variablen c_saisie_i_n für intramuros, c_saisie_e_n für extramuros und c_saisie_a_n für ausländische F+E Ausgaben wird ein globaler Erfassungsstatus c_saisie_n abgeleitet. Die Herleitung

Tabelle 2: Erfassungsvariable c_saisie_n und Nettostichprobe vor der Behandlung der Unternehmensgruppen.

Code	Erläuterung	Ausfall	Antwort	Total
1	fait RD / sans dép. RD (1)	0	8	8
2	fait RD / sans chiffres (2)	0	179	179
3	non RD (3)	0	3715	3715
4	quest. blanc (4)	110	0	110
5	n'existe plus (5)	59	0	59
6	n'existe plus/RD (6)	0	0	0
7	n'éxiste plus/pas de RD (7)	0	0	0
8	quest. pas retourné	1626	0	1626
9	supprimé (D)	0	0	0
10	dép. RD, avec pers. RD (O)	0	593	593
11	dép. RD, sans pers. RD (M)	0	9	9
12	Erreur d'échantillonage (8)	0	0	0
Total		1795	4504	6299

von c_saisie_n wurde durch die Sektion HSW durchgeführt. Die Codes von c_saisie_n und ihre Bedeutung ist in Tabelle 2 aufgeführt. Nur die Beobachtungen mit c_saisie_n $\in \{10, 11\}$ sollten bei mindestens einer der Hauptvariablen B0240T, B0244T, F0291T, F0276T, K0205T positive Angaben enthalten.

Die Ausgangs-Stichprobe wurde vorläufig definiert als die Menge der Beobachtungen in der Bruttostichprobe mit c_saisie_n $\in \{1, 2, 3, 10, 11\}$. Die Ausgangs-Stichprobe hat also einen Umfang von 4504 Unternehmen (vgl. Tabelle 2).

2.2 Unternehmensgruppen in der selben Branche

Während der Aufarbeitung war eine Liste von Unternehmensgruppen erstellt worden, bei denen statt der Antwort einzelner Unternehmen in der Stichprobe eine gesamthafte Antwort eingereicht wurde. Es stellte sich im Lauf der Arbeiten aber heraus, dass einzelne Unternehmen von Unternehmensgruppen trotzdem einen Fragebogen eingereicht hatten und dass andererseits die Antworten der Unternehmensgruppen u.U. nicht alle Unternehmen einschlossen.

Es können sich komplexe Situtation ergeben. Die generelle Politik der Sektion HSW war, die F+E Ausgaben von Zweigunternehmen unter dem jeweiligen Mutterunternehmen zu verbuchen und das Zweigunternehmen als "keine F+E" (c_saisie_n=3) zu klassieren. Dieses Vorgehen wurde anhand des aktuellen Stands des Betriebs- und Unternehmensregisters (insbesondere auch was die Zugehörigkeit zu Gruppen und Unternehmen sowie die Wirtschaftsklasse anbelangt) überprüft. In komplexen Fällen wurde die Strukutr der Unternehmungsgruppen direkt mit den Unternehmen zusammen geklärt.

Im Stichprobenrahmen wurden nur Einzelunternehmen (BEART=4) oder der Hauptsitz des Unternehmens (BEART=3) eingeschlossen (alte Typologie der Betriebsart).

Die Unternehmen wurden gebeten, in der Rubrik 0 alle ihre Arbeitsstätten und Zweigunterneh-

men namentlich aufzuführen. Ausserdem wurden sie gebeten, in der Rubrik C die Arbeitsstätten und Zweigunternehmen, die in der Schweiz F+E betreiben, aufzuführen.

Bei den Unternehmen, die auf ihrem Fragebogen im Teil C angegeben haben, dass sie die F+E-Angaben auch für ein anderes Unternehmen liefern, handelt es sich meistens um Mutterunternehmen, deren Zweigunternehmen in derselben Branche tätig sind wie das Mutterunternehmen. Falls ein solches Zweigunternehmen ebenfalls in der Bruttostichprobe war, wurden die Variablen des Zweigunternehmens auf fehlend gesetzt und die Variable c_saisie_n des Zweigunternehmens wurde auf "keine F+E"(Code 3) gesetzt. Dies hat zur Folge, dass allfällige Angaben über Forschungsausgaben von Zweig-Unternehmen ignoriert werden. Mit anderen Worten wird angenommen, dass das Mutter-Unternehmen die F+E-Ausgaben des Zweig-Unternehmens richtig angegeben hat. In allen Fällen ist das Zweig-Unternehmen in der selben Branche tätig wie das Mutterunternehmen. Das Zweig-Unternehmen bleibt gemäss der Definition der Variablen c_saisie_n immer noch in der Netto-Stichprobe.

Das beschriebene Vorgehen betrifft 17 Zweigunternehmen in 12 Mutterunternehmen mit insgesamt 3943 Beschäftigten. Für die Auswertung wurden nun diese Fälle so behandelt: Die Beschäftigten (DUEMPLOI) des Zweig-Unternehmens werden zum Mutterunternehmen addiert. Das Zweigunternehmen wird aus der Netto-Stichprobe ausgeschlossen. Der Stichprobenrahmen wird nicht geändert, hingegen wurde bei der Berechnung der Einschlusswahrscheinlichkeiten die Veränderung der Bruttostichprobengrösse aufgrund der Zweigunternehmen berücksichtigt (siehe Abschnitt 2.4). In zwei Fällen war das aufgeführte Zweigunternehmen nicht in unserem Stichprobenrahmen aufgeführt (Ausserhalb der Zielpopulation oder neue Unternehmung). Der Stichprobenrahmen wurde aber nicht entsprechend angepasst.

2.3 Unternehmensgruppen mit Zweigunternehmen in verschiedenen Branchen

Bei 3 Unternehmensgruppen mit 25 Zweigunternehmen verteilten sich die Zweigunternehmen auf verschiedene Branchen. Es handelt sich um grosse Unternehmen mit zum Teil grossem Forschungsaufwand. Damit die Resultate der Erhebung auf Branchenebene ausgewiesen werden können, müssen die Gruppentotale auf die Branchen der Gruppe verteilt werden. Das Vorgehen ist wie folgt:

Sei U_G die Menge der Unternehmen, die zu einer bestimmten Gruppe gehören. Seien U_{Gbt} mit $b = 1, \ldots, B$ und $t = 1, 2, 3$ die Unternehmen der Gruppe in den Schichten (Der Index b bezieht sich auf die Branche und der Index t auf die Grössenklasse). Wir betrachten eine Hilfsvariabel x_{bti}, die Anzahl Beschäftigte DUEMPLOI, und eine interessierende Variable y_{bti} pro Unternehmen. Für eine Gruppe is das Total einer Variable für die Branche b

$$x_{Gb+} = \sum_t \sum_{i \in U_{Gbt}} x_{bti}. \tag{1}$$

Das Gesamt-Total für die Gruppe über die Branchen ist dann $x_{G+} = \sum_b x_{Gb+}$ und analog für die Variable y.

Die Idee ist, für jede Gruppe eine fiktive Einheit pro Branche zu bilden. Dabei soll das Gesamt-Total der Variablen y proportional zu der Anzahl Beschäftigte einer Gruppe in einer Branche

verteilt werden. Wir müssen also die Branchen-Totale der Beschäftigten x_{Gb+} berechnen und die Gesamttotale der Beschäftigten und der interessierenden Variablen y.

Die Variable y der fiktiven Einheiten ist dann

$$\hat{y}_{Gb+} = y_{G+}(x_{Gb+}/x_{G+}). \tag{2}$$

Diese Schätzung ist korrekt, wenn die Forschungsausgaben pro Beschäftigter in den Branchen der Gruppe gleich ist. Diese Annahme wird natürlich höchstens annähernd zutreffen und entsprechend fehlerhaft sind die eingesetzten Werte bei den fiktiven Einheiten.

Falls die interessierende Variable eine Prozentzahl ist, kann ein analoges Einsetzungsverfahren verwendet werden. Dabei müssen aber die Anteile mit der Variable DUEMPLOI oder einer geeigneten Variable gewichtet werden. Sei p_{bti} eine Prozentzahl für die Unternehmung i in der Branche b und der Grössenklasse t, z.B. sei p_{bti} der Anteil der Intramuros-Ausgaben für Grundlagenforschung. Wir können nun den Anteil mit dem absoluten Wert des Ganzen multiplizieren (z.B. Anteil Intramuros-Ausgaben für Grundlagenforschung mal Total der Intramuros-Ausgaben) und erhalten damit eine neue Variable $y'_{bti} = p_{bti}y_{bti}$. Diese können wir dann gleich behandeln wie oben beschrieben.

Diese Einsetzung wurde für 3 Gruppen mit insgesamt 8 fiktiven Einheiten durchgeführt. Die fiktiven Einheiten erhielten die Branche b zugewiesen und wurden in die Grössenklasse 3 (Grosse Unternehmen) eingeteilt. Da in jeder der auftretenden Branchen mindestens 1 Unternehmen der Gruppe grösser als DUEMPLOI=100 war, wäre die Einschlusswahrscheinlichkeit der fiktiven Einheiten 1 gewesen, d.h. gleich wie die Einschlusswahrscheinlichkeiten der jeweils grössten betroffenen Zweigunternehmen.

Die Zweigunternehmen dieser Mehrbranchen-Gruppen wurden aus der Nettostichprobe entfernt.

2.4 Nettostichprobe und Berichtigung der Populationstotale

Die Nettostichprobe umfasst nach der Behandlung der Unternehmensgruppen

$$n = 4504 - 17 - 25 + 8 = 4470$$

Unternehmen. Die Veränderungen der Nettostichprobengrössen durch die Behandlung der Unternehmensgruppen wurde bei den Populationstotalen je Schicht nachvollzogen:

$$N_h^n = N_h^v - (n_h^v - n_h^n), \tag{3}$$

wobei der obenstehende Index ^{v} vor und ^{n} nach der Behandlung der Zweigunternehmen bedeutet. Entsprechend wurden die Populationstotale der Variable DUEMPLOI angepasst.

Die korrigierten Populationsgrössen sowie Netto- und Bruttostichprobengrössen sind in den Tabellen 6 und 7 aufgeführt.

3 Datenersatz

3.1 Hauptvariablen

Im Auswertungsdatensatz der Grösse $n = 4470$ hat es viele fehlende Werte und viele Nullwerte. Tabelle 3 zeigt z.B., dass nur 593 Unternehmen positive Ausgaben für intramuros Forschung angegeben haben.

Tabelle 3: Anzahl fehlende Werte, Nullwerte und Werte mit Angaben >0

Variable	Anz. fehlende	Anzahl 0	Anzahl >0
B0240T	186	3691	593
B0244T	207	3941	322
F0291T	29	4139	302
F0276T	29	4315	126
K0205T	18	4390	62
totfep	181	3708	581

Die Anzahl Werte grösser als Null variieren mit der Grössenklasse stark. Unter den fehlenden Werten sind vermutlich viele mit tatsächlich Forschungsausgaben. Dies wurde durch eine telefonische Nachbefragung bestätigt. Dabei hatten nämlich 193 Unternehmen angegeben, dass sie Forschung betreiben. Sie konnten aber die genauen Ausgaben nicht beziffern.

Tabelle 4: Anteil (in %) der Werte mit Angaben >0 nach Grössenklassen

	Grössenklasse			
Variable	1	2	3	Alle Klassen
B0240T	6.41	14.97	25.65	13.27
B0244T	2.46	8.11	15.92	7.20
F0291T	1.98	6.57	16.16	6.76
F0276T	0.77	2.24	7.14	2.82
K0205T	0.32	0.84	3.76	1.39
totfep	6.21	14.83	25.18	13.00

Die Kolonne “Alle Klassen” in Tabelle 4 zeigt zugleich die Forschungsintensität, d.h. den (ungewichteten) Anteil der Unternehmen in der Stichprobe, die Forschung betreiben an (bezüglich der betreffenden Variablen). Vor allem der Anteil Unternehmen, die Angaben für die Ausgaben von Tochterunternehmen im Ausland (Variable K205T) machen, ist sehr klein!

Die Verteilung der Variablen B0240T und DUEMPLOI ist in Abbildung 2 dargestellt. Typisches Bild ist, dass es sehr viele Fälle mit B0240T=0 hat: Es handelt sich dabei um die Punkte am unteren Bildrand. Ausserdem ist die Verteilung der positiven Werte sehr stark rechtsschief (siehe Normalplot in Abbildung 3). Wir müssen also mit Ausreissern rechnen, insbesondere, wenn aufgrund der vorhandenen Daten fehlende Werte ersetzt werden sollen.

Es wurden verschiedene Methoden für das Einsetzen von Werten bei Merkmals-Ausfällen untersucht.

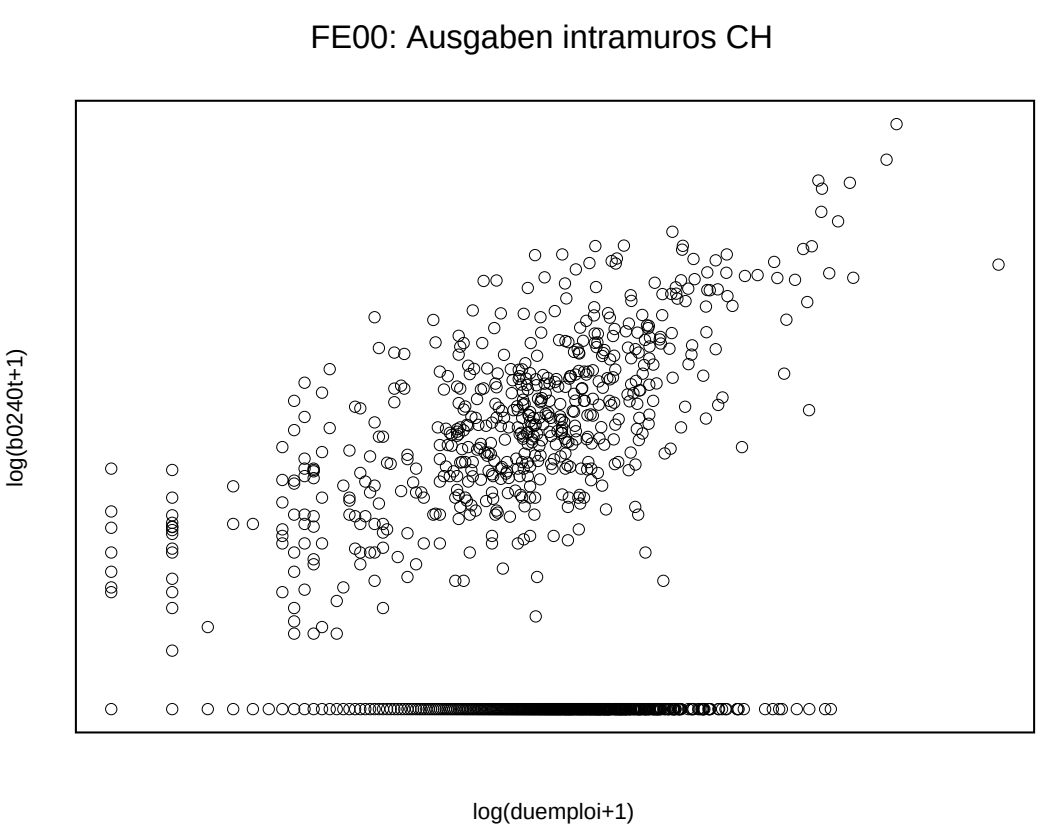


Abbildung 2: Diagramm B0240T gegen DUEMPLOI (je logarithmiert).

1. Einsetzung des Werts 0.
2. Einsetzung des ersten Quartils der Beobachtungen mit Werten > 0.
3. Einsetzung des Medians der Beobachtungen mit Werten > 0 (cf. Gleichung (15)).
4. Einsetzung des Mittelwerts der Beobachtungen mit Werten > 0.
5. Einsetzung des Mittelwerts über alle Beobachtungen.
6. Einsetzung des Mittelwerts der Beobachtungen mit Werten > 0 oder Einsetzung von 0 mit Wahrscheinlichkeit $1 - p_0$ and p_0, wobei

$$p_0 = \begin{cases} 1 & n_p = 0 \\ 0 & n_m \leq 0.5\, n_p, \\ (n_m - 0.5\, n_p)/n_m & \text{andernfalls.} \end{cases}$$

   Dabei bezeichnet n_m die Anzahl der fehlenden Werte und n_p die Anzahl Werte > 0.

Die letzte Version des Datenersatzes versucht, die Anzahl der neu erzeugten positiven Werte zu begrenzen, falls die Anzahl fehlende Werte gross ist verglichen mit der Anzahl positiver Werte. Alle Einsetzungen wurden für jede der Analyse-Gruppen (definiert durch die Variable gruppind,

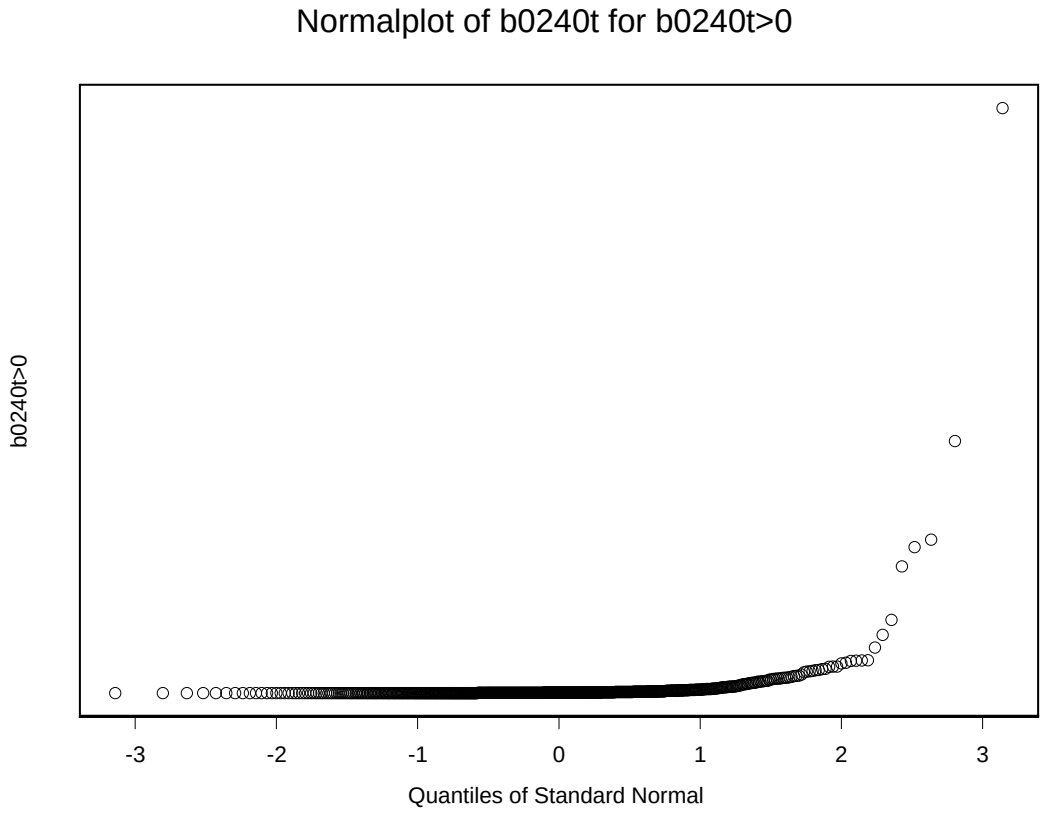


Abbildung 3: Normalplot der Variable B0240T für Beobachtungen mit B0240T>0.

siehe Kapitel 4) und für jede Variable separat durchgeführt. Mittelwerte und Mediane werden dabei pro Gruppe ungewichtet berechnet, weil die Einheiten der Gruppen gleiche Gewichte haben. Um die Einsetzungen zu überprüfen wird jeweils eine Schätzung des entsprechenden Totals berechnet. Die dabei verwendete Gewichtung wird im Abschnitt 4 erläutert.

Tabelle 5: Total der intramuros Ausgaben nach Einsetzungsmethode

Nr	Methode	Anzahl Werte	Total B0240T
1	Keine Einsetzung	4284	7’229’334
2	Einsetzen erstes Quartil der “>0"	4470	7’413’308
3	Einsetzen Median der “>0"	4470	7’710’178
4	Einsetzen Mittelwert der “>0"	4468	8’360’968
5	Einsetzen Mittelwert über alle	4470	7’709’314
6	Einsetzen von 0 oder Mittelwert mit Wk. p_0	4470	8’281’375

Da SAS bei der Aggregierung die fehlenden Werte ignoriert, entspricht die Methode “Keine Einsetzung” einer Einsetzung des Wertes 0.

Die Endresultate reagieren sehr empfindlich auf die Einsetzungsmethode. Z.B. ist die Schätzung für die gesamten Intramuros-Ausgaben (B0240T) je nach Einsetzungs-Methode zwischen

7.2 Mia und 8.4 Mia (vgl. Tabelle 5). Die Einsetzung des Mittelwerts über alle Beobachtungen einer Schicht (inklusive 0) ergab ein sehr ähnliches Resultat wie die Einsetzung des Medians der Beobachtungen mit Werten grösser als 0. Die Einsetzung des Medians ist dabei vorzuziehen, weil sie robustere Schätzungen liefert, also nicht durch einzelne Ausreisser in den Schichten stark beeinflusst wird. Die Einsetzung des Mittelwerts der Beobachtungen mit Werten grösser als 0 ist sicher keine robuste Methode und scheint auch im Vergleich mit externen Grössen zu hoch auszufallen. Die schlussendlich gewählte Methode ist, dass ein fehlender Wert durch den Median derselben Variable über die Beobachtungen der eigenen Gruppe, welche eine Angabe grösser als 0 gemacht haben, ersetzt wird. Falls keine Angaben grösser als 0 sind, wird eine 0 eingesetzt. Für jede eingesetzte Variable wurde eine entsprechende Indikator-Variable erzeugt (z.B. für B0240T wurde impb0240t erzeugt), welche angibt, ob der Wert eingesetzt ist (impb0240t=1) oder nicht (impb0240t=0).

Komplexere Methoden, die auf Modellen aufbauen, oder die multivariate Struktur der Daten berücksichtigen, konnten aus Zeitgründen nicht untersucht werden. Neben den Totalen, für die eine Einsetzung vorgenommen wurde, treten auch viele fehlende Werte in Subtotalen und bei Prozentwerten auf. Wenn für Subtotale das gleiche Vorgehen wie bei Totalen angewendet würde, entstehen natürlich Inkonsistenzen (die Subtotale addieren nicht mehr zum Total). In dieser Arbeit konnten aus Ressourcengründen nur Totale eingesetzt werden.

3.2 Vollzeitäquivalente

Ueber die ursprünglich vorgesehenen Kernvariablen hinaus musste für die Variable "Vollzeitäquivalente in F+E"(totfeept) ein spezieller Datenersatz durchgeführt werden, weil die Korrelation zwischen der Variable "Beschäftigte in F+E"(totfep) und der Variable totfeept sehr gross ist. Dabei dürfte die Korrelation je nach Grössenklasse des Unternehmens verschieden sein: Grosse Unternehmen können sich eher Vollzeit-Forscher leisten als kleine. Die Variable B0210T, “Aufwendungen für F+E Personal” hat eventuell darüber hinaus noch eine gewisse Vorhersagekraft. Die Anzahl Unternehmen, die von dieser zusätzlichen Einsetzung betroffen sind, beträgt nur 56 (D.h. es gibt 56 Unternehmen, bei denen totfep>0 ist, während totfeept fehlt). Trotzdem ist der Effekt natürlich in den Schätzungen für die Totale gut sichtbar.

Die Regressionsmodelle, die getestet wurden verwenden eine Gewichtung $w_{ri} = w_i/\sqrt{x_i}$, wobei x_i die Variable totfep, Anzahl Beschäftigte im F+E Bereich, ist. Damit soll gleichzeitig der Stichprobenplan und die Heteroskedastizität berücksichtigt werden. Die Daten, an die das Modell angepasst wurde, waren alle Daten mit totfep>0 und totfeept>0. Die Werte für totfeept der fiktiven Einheiten waren offensichtlich falsch (obwohl z.T. >0) und wurden allesamt eingesetzt.

Das vollständige Modell totfep=totfep*grkl+b0210t (ohne Achsenabschnitt, SAS-Notation ’*’ bedeutet vollständige Kreuzung) mit totfep, der Grössenklasse grkl und den Ausgaben für F+E-Personal b0210t ergab ein Bestimmtheitsmass von 0.980. Alle Schätzungen waren hochsignifikant. Es zeigte sich aber, dass vor allem die Gruppe der kleinsten Unternehmen eine unterschiedliche Steigung in totfep hatte, während die Unterschiede bei den anderen Grössenklassen klein waren. Ein Modell mit verschiedenen Steigungen für grkl=1 und grkl>1 aber ohne B0210T ergab ein Bestimmtheitsmass von 0.979, also praktisch gleich gut wie das grössere Modell. Das war darum das gewählte Einsetzungsmodell: Für Grössenklasse 1 wurde totfeept=0.6834*totfep gesetzt, während für die anderen Grössenklassen totfeept=0.92873*totfep gesetzt wurde.

Der Einfluss auf das Total von totfeept ist beachtlich: Dank der Imputation steigt das Total von 22'339 auf 35'445. Bei einer Gewichtung, die den fehlenden Werten in totfeept angepasst ist, wäre natürlich der erstere Wert auch viel höher ausgefallen. Ausserdem ist zu beachten, dass die Werte der fiktiven Einheiten offensichtlich falsch wahren. Der Wert 35'445 ist im übrigen plausibel, wenn man ihn mit den Werten für Vollzeitäquivalente aus anderen Erhebungen vergleicht.

4 Schätzverfahren

Die Grundgesamtheit wurde in 4 Grössenklassen aufgeteilt: Von der bisherigen dritten Grössenklasse der Unternehmen mit DUEMPLOI grösser als 100 wurde eine vierte Grössenklasse der Unternehmen mit mehr als 2000 Beschäftigten abgetrennt. Diese Klasse der grössten Unternehmen wurde als sogenannt autorepräsentativ deklariert und ihr Stichprobengewicht w_i wurde auf 1 gesetzt. In den anderen Grössenklassen wurden jeweils neben der Stichprobenrate auch die Ausfallrate mit Gewichtung ausgeglichen.

Genauer: Sei N_h die Populationsgrösse laut Stichprobenrahmen. Dabei bezeichne h die Schicht, die gebildet wird durch Kreuzung der Branche (Definition nach Sektion HSW, siehe (RENFER 2003)) mit den 4 Grössenklassen (DUEMPLOI= 0 bzw. 6 bis 49, 50 bis 99,100 bis 1999, 2000 und mehr). Um diese neuen Schichten von den Schichten des Stichprobenplans zu unterscheiden, bezeichnen wir sie mit Analysegruppen (Variable gruppind).

Da die Unternehmen mit DUEMPLOI>=100 vollständig in die Stichprobe aufgenommen wurden, können im Prinzip beliebige Nachschichten innerhalb dieser Klasse gebildet werden. Die Grenze 2000 wurde gewählt, weil ab einer gewissen Grösse die Unternehmen fast nicht mehr vergleichbar, und damit nicht ersetzbar sind. Es scheint auch nicht sinnvoll, ein ausgefallenes Unternehmen der Grösse DUEMPLOI=120 durch Höhergewichtung von Unternehmen u.a. auch mit DUEMPLOI über 2000 zu kompensieren. Die Grenze ist natürlich immer bis zu einem gewissen Grad willkürlich. Es fallen 19 Unternehmen in diese Grössenklasse.

Die Veränderungen in der Bruttostichprobe wegen den fiktiven Einheiten und den Zweig-Unternehmen wurden in der Grundgesamtheit nachvollzogen (cf. Gleichung (3)). Die Struktur der Grundgesamtheit und der Stichprobe ist in den Tabelle 6 und 7 aufgeführt.

Der Gruppen-Indikator h zeigt in der Einer-Stelle die Grössenklasse und in der Zehner-Stelle die Branche an (z.B. ist 162 die Grössenklasse 2 in der Branche 16). Die Grösse der Population U_h (korrigiert für fiktive Einheiten und Mutterunternehmen) ist N_h. Die Anzahl Beschäftigte (Vollzeit plus Teilzeit 1 plus Teilzeit 2) ist M_h (ebenfalls korrigiert). Die korrigierte Bruttostichprobengrösse ist n_{bh} und die Grösse der Netto-Stichprobe S_h ist n_h. Die Gewichte w_i wurden wie folgt berechnet: $w_i = N_{h(i)}/n_{h(i)}$. Dabei bezeichnet $h(i)$ die Gruppe des Elements i. Die Gewichtung wird im folgenden genauer hergeleitet.

Die Einschlusswahrscheinlichkeiten nach Stichprobenplan sind

$$\pi_i = n_{bh(i)}/N_{h(i)}. \tag{4}$$

Da alle fiktiven Einheiten mindestens ein Unternehmen mit Einschlusswahrscheinlichkeit 1 gemäss dem Stichprobenplan enthalten, kann die Einschlusswahrscheinlichkeit auch nach der Ver-

Tabelle 6: Populationsgrössen, Anzahl Beschäftigte, Brutto- und Netto-Stichprobengrössen, Antwortraten und Gewichtung nach Analysegruppen: Branchen 1 bis 10

Gruppe	N	M	n_b	n_n	r%	w
11	164	2673	164	110	67	1.4909
12	12	775	12	10	83	1.2000
13	5	908	5	5	100	1.0000
21	783	12183	150	98	65	7.9898
22	96	6674	96	69	72	1.3913
23	113	30793	113	100	88	1.1300
24	5	14683	5	5	100	1.0000
31	483	8068	272	168	62	2.8750
32	69	4954	69	64	44	1.5682
33	59	12300	59	52	88	1.1346
41	3266	43304	565	366	65	8.9235
42	159	10917	159	105	66	1.5143
43	128	32309	128	122	95	1.0492
44	1	2049	1	1	100	1.0000
51	246	4780	72	46	64	5.3478
52	43	3127	43	21	49	2.0476
53	59	17531	59	45	76	1.3111
54	3	9187	3	3	100	1.0000
61	71	1376	71	41	58	1.7317
62	27	1951	27	19	70	1.4211
63	28	8816	28	17	61	1.6471
64	2	14320	2	2	100	1.0000
71	373	6384	137	91	66	4.0989
72	56	3837	56	38	68	1.4737
73	61	14488	61	53	87	1.1509
81	458	7382	330	226	68	2.0265
82	39	2573	39	20	51	1.9500
83	37	7884	37	35	95	1.0571
91	2635	39140	203	133	66	19.8120
92	218	15115	55	32	58	6.8125
93	145	32093	145	120	83	1.2083
101	203	3532	39	22	56	9.2273
102	27	1947	27	18	67	1.5000
103	38	14681	38	28	74	1.3571

Tabelle 7: Populationsgrössen, Anzahl Beschä ftigte, Brutto- und Netto-Stichprobengrössen, Antwortraten und Gewichtung nach Analysegruppen: Branchen 11 bis 20 und Total

Gruppe	N	M	n_b	n_n	r%	w
111	873	13222	60	34	57	25.6765
112	127	8727	61	40	66	3.1750
113	129	36317	129	91	71	1.4176
114	2	8137	2	2	100	1.0000
121	1823	30410	109	67	61	27.2090
122	241	16771	50	32	64	7.5313
123	309	88999	309	204	66	1.5147
124	9	26719	9	5	56	1. 8000
131	792	12015	317	219	69	3.6164
132	56	3674	56	44	79	1.2727
133	42	777 1	42	40	95	1.0500
141	97	2127	97	91	94	1.0659
142	18	1229	18	17	94	1.0588
143	29	10014	29	27	93	1.0741
151	4676	70719	265	182	69	25.6923
152	347	23742	41	23	56	15.0870
153	182	36178	182	125	69	1.4560
154	1	3359	1	1	100	1.0000
161	1281	17795	35	35	100	36.6000
162	94	6424	35	32	91	2.9375
163	58	15544	58	55	95	1.0545
164	1	2802	1	1	100	1.0000
171	162	2744	162	148	91	1.0946
172	11	669	11	10	91	1.1000
173	21	7575	21	16	76	1.3125
174	2	26557	2	2	100	1.0000
181	1214	15772	118	87	74	13.9540
182	50	3421	50	33	66	1.5152
183	47	12760	47	36	77	1.3056
191	501	2384	134	93	69	5.3871
192	17	1187	17	14	82	1.2143
193	15	3905	15	11	73	1.3636
201	6311	71259	290	222	77	28.4279
202	120	8224	120	92	77	1.3043
203	100	26607	100	73	73	1.3699
204	2	4611	2	1	50	1.0000
Total	29872	1023104	6265	4470	71.3	

teilung auf die Branchen mit 1 angenommen werden (die fiktiven Einheiten gehören alle zu den grossen Unternehmen).

Wir setzen innerhalb der Gruppen das übliche Modell an, dass wir annehmen, die Gesamt-Antwortausfälle fehlen vollständig zufällig, d.h. wir können annehmen, die Teilstichprobe der Antwortenden sei eine einfache Zufallsstichprobe der Bruttostichprobe. Die entsprechende Antwort-Wahrscheinlichkeit ist

$$r_{h(i)} = n_{h(i)}/n_{bh(i)}. \tag{5}$$

Die kombinierte Einschluss- und Antwort-Wahrscheinlichkeit ist nun

$$\tilde{\pi}_i = \pi_i r_{h(i)} = \frac{n_{bh(i)}}{N_{h(i)}} \frac{n_{h(i)}}{n_{bh(i)}} = n_{h(i)}/N_{h(i)}. \tag{6}$$

Die Horvitz-Thompson Gewichtung ist nun das Inverse der Einschluss- und Antwortwahrscheinlichkeiten:

$$w_i = 1/\tilde{\pi}_i = \frac{N_{h(i)}}{n_{h(i)}}. \tag{7}$$

Der Horvitz-Thompson Schätzer für das Total einer Variable y ist nun

$$T_y = \sum_{i \in S} w_i y_i. \tag{8}$$

Der Schätzer für einen Mittelwert ist

$$\hat{\bar{y}} = \sum_{i \in S} w_i y_i / \sum_{i \in S} w_i. \tag{9}$$

Bei den autorepräsentativen Unternehmen kann man annehmen, dass es keine entsprechenden anderen Unternehmen in der Population hat. Dies ist sicher richtig für alle Branchen ausser Branche 12 und Branche 20. In der Branche 12 fehlen 4 sehr grosse Unternehmen von denen man annehmen muss, dass sie Forschung betreiben. Darum setzen wir in der Gruppe 124 das normale Gewichtungsmodell mit Korrektur der Ausfallrate $r_{124} = 5/9$ an.

In der Branche 20 fehlt ein sehr grosses Unternehmen für das wir keine Korrektur des Antwort-Ausfalls vornehmen wollen, weil das Unternehmen aufgelöst worden sei. Hier setzen wir also $r_{204} = 1$ an.

Die Gewichte w_i sind in den Tabellen 6 und 7 aufgeführt. Die grössten Gewichte entstehen in den Schichten mit den kleinen Unternehmen, die eine tiefe Stichprobenrate haben. Diagramme je der Untersuchungs-Variablen gegen die Gewichte zeigen, dass keine extrem grossen Werte mit einem grossen Gewicht verbunden sind.

Das Minimum der Gewichte w_i ist 1 während das Maximum 36.6 beträgt. Der Quotient

$$q = \max(w_i)/\min(w_i) = 36.6 \tag{10}$$

ist relativ klein (verglichen mit anderen Unternehmenserhebungen und mit komplexen Haushalterhebungen). Hingegen ist der Variationskoeffizient eher gross, nämlich 132.3. Das 95% Quantil der Gewichte ist 28.4. Der Anteil des Gewichts der Beobachtungen mit $w_i \geq 28.4$ am Gesamtgewicht, die "5%-Dominanz", ist 0.25. Die Median-Dominanz, d.h. der kleinste Anteil

an Beobachtungen, die mit ihrem Gewicht mehr als die Hälfte des Gesamtgewichts ausmachen, ist nur 0.12! Wir haben es also trotz der kleinen Spannweite mit sehr asymmetrischen Gewichten zu tun. Die extremen Gewichte können bereits in Kombination mit mässigen Werten eine gewichtete Summe dominieren. Für B0240T zeigt sich aber, dass die grossen Gewichte alle mit kleinen Werten kombiniert sind. Trotzdem tragen die drei grössten gewichteten Werte von B0240T 26% zum Gesamttotal bei!

Eine Gewichtung, die auf separaten Quotientenschätzern mit der Hilfsvariable DUEMPLOI beruht, wurde ebenfalls berechnet. Das Gewicht w_{mh} einer Unternehmung in Analysegruppe h ist definiert als

$$w_{mh(i)} = \frac{\sum_{j \in U_{h(i)}} M_j}{\sum_{j \in S_{h(i)}} M_j} = \frac{\sum_{j \in U_{h(i)}} M_j}{\sum_{j \in S_{h(i)}} w_j M_j} \quad w_i. \tag{11}$$

Dabei ist M_j die Grösse der Unternehmung j gemessen mit der Variable DUEMPLOI. Die rechte Seite der Formel zeigt, dass w_{mi} einem separaten Quotientenschätzer entspricht. Mit anderen Worten entsteht w_{mi}, indem man w_i auf die Hilfsvariable DUEMPLOI, d.h. auf M_i pro Gruppe kalibriert. Die Abbildung 4 zeigt aber, dass zwar eine positive Korrelation vorhanden ist, dass sie aber durch die Fälle mit 0 stark gedämpft wird. Die Abbildung zeigt auch, dass das Modell für die Einsetzungen die Abhängigkeit der intramuros Ausgaben von der Anzahl Beschäftigte zu wenig fein berücksichtigt (es werden ja nur 4 Klassen unterschieden). Unter anderem wird ein klarer Ausreisser erzeugt (Hebelpunkt!). Aber auch ohne eingesetzte Werte sind Probleme mit Ausreissern zu erwarten. Es müssten also robuste Schätzungen für den Quotienten verwendet werden (vgl. z.B. (HULLIGER und KASSAB 1998) oder (HULLIGER 1999)). Dies war unter dem gedrängten Zeitplan der Erhebung nicht möglich. Ein Vergleich der obigen Gewichte mit den reinen Expansionsgewichten w_i zeigt im übrigen, dass der Unterschied zwischen den beiden Gewichten klein ist (siehe auch Tabelle 8). Aus diesen Gründen wurde die Gewichtung w_{mi} fallen gelassen.

Die Schätzer wurden anschliessend mit der Prozedur SURVEYMEANS von SAS berechnet, um eine Varianzschätzung gleich mitgeliefert zu bekommen. Dabei konnte aber der Einfluss der Einsetzungen auf die Varianzen nicht berücksichtigt werden. Die angegebenen Variationskoeffizienten sind also sicher zu klein.

Die SAS Prozedur SURVEYMEANS verwendet eine Taylor-Linearisierung für die Berechnung der Varianz (SAS INSTITUTE 1999, S. 3198 ff.). Die Dokumentation über die Prozedur geht von einer zweistufigen Stichprobe mit Klumpung aus. Wir verwenden hier die Notation, dass der Index i in jeder Schicht h separat verläuft (Der Index j der SAS-Dokumenatation kann hier weggelassen werden). Für unseren Fall mit einer einfachen geschichteten Zufallsstichprobe lautet die Formel für den Schätzer der Varianz des Totals $T_y = \sum_h \sum_i w_{hi} y_{hi}$ einer Variable y

$$\hat{V}(T_y) = \sum_{h=1}^{H} \frac{n_h(1 - n_h/N_h)}{n_h - 1} \sum_{i=1}^{n_h} (e_{hi} - \bar{e}_h)^2, \tag{12}$$

wobei in unserem Fall $e_{hi} = w_{hi}(y_{hi} - \hat{\bar{y}})$ mit dem gewichteten Mittel $\hat{\bar{y}}$ und $\bar{e}_h = \sum_{i=1}^{n_h} e_{hi}/n_h$ ist.

Der Summand $e_{hi} - \bar{e}_h$ kann in unserem Fall einfacher berechnet werden.

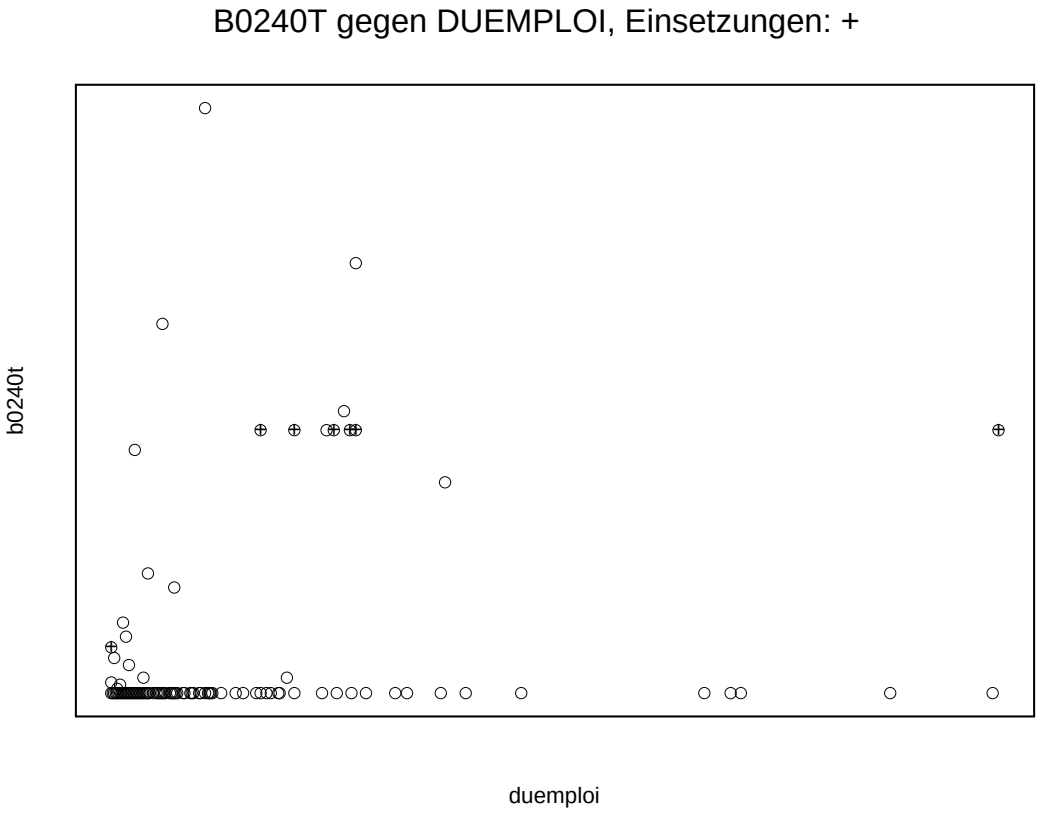


Abbildung 4: Intramuros Ausgaben (B0240T) gegen Anzahl Beschäftigte (DUEMPLOI) einer Branche. Die eingesetzten Werte sind mit + bezeichnet.

$$\begin{aligned} e_{hi} - \bar{e}_h &= w_{hi} - w_{hi}\hat{\bar{y}} - \frac{1}{n_h}\sum_{i=1}^{n_h} w_{hi}y_{hi} + \frac{1}{n_h}\sum_{i=1}^{n_h} w_{hi}\hat{\bar{y}} \\ &= \left(w_{hi}y_{hi} - \frac{1}{n_h}\sum_{i=1}^{n_h} w_{hi}y_{hi}\right) - \hat{\bar{y}}\left(w_{hi} - \frac{1}{n_h}\sum_{i=1}^{n_h} w_{hi}\right) \\ &= w_h(y_{hi} - \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}), \end{aligned} \tag{13}$$

weil hier $w_{hi} = w_h = N_h/n_h$ ist.

Damit wird die Varianzschätzung (12) zu

$$\hat{V}(T_y) = \sum_{h=1}^{H} \frac{n_h(1 - n_h/N_h)}{n_h - 1}\left(\frac{N_h}{n_h}\right)^2 \sum_{i=1}^{n_h}(y_{hi} - \frac{1}{n_h}\sum_{i=1}^{n_h} w_{hi}y_{hi})^2. \tag{14}$$

Dies ist die gewohnte Varianz des geschichteten Totals.

Um den Einfluss der Einsetzungen auf die Varianzschätzungen etwas näher zu untersuchen, wurden die Varianzen der Totale auch so berechnet, dass die fehlenden Werte einfach wegge-

Tabelle 8: Schätzung der totalen intramuros Ausgaben mit Expansions- und Quotienten-Gewichten

	Totale intramuros Ausgaben CH	
Branche	Mit w_i	Mit w_{mi}
1	723	724
2	392267	390550
3	10880	10706
4	31555	31890
5	594778	600892
6	1834027	1862948
7	46184	49902
8	8986	8934
9	171072	175760
10	354471	339789
11	644726	660950
12	2099530	2071283
13	22015	22023
14	355	381
15	4085	4190
16	93651	93001
17	54169	53750
18	262605	257687
19	876460	924850
20	207640	213119
All	7'710'178	7'773'330

lassen wurden. Um den Einfluss des dadurch entstehenden kleineren Totals auf den Variationskoeffizienten zu vermeiden, wurden die Variationskoeffizienten von gewichteten Mitteln betrachtet. Die zweite Kolonne in Tabelle 9 enthält den Variationskoeffizienten für die Schätzung des Mittelwertes, wie er ohne Modifikation von der SAS Prozedur SURVEYMEANS berechnet wird. Die dritte Kolonne zeigt den Variationskoeffizienten für die Schätzung des Mittelwerts, wenn nur die nicht imputierten Daten bei der Berechnung zugelassen werden. Diese letzteren Variationskoeffizienten sind durchwegs leicht grösser als die Variationskoeffizienten mit eingesetzten Werten. Die Unterschätzung ist zwar nicht dramatisch erreicht aber bei B0244T 26%. Der Grund für die Unterschätzung ist einerseits die Stichprobengrösse, die tatsächlich ohne die eingesetzten Werte berechnet werden müsste. Andrerseits ist auch die Variabilität der eingesetzten Werte gering, da überall der Median eingesetzt wurde.

Die tatsächliche Varianz ist schwieriger zu berechnen. Um den Einfluss der Einsetzungen zu klären, kann der Schätzer des Totals in einer Analysegruppe so geschrieben werden:

$$T = \frac{1}{n_{h(i)}} \left(\sum_{1}^{n_{rh(i)}} w_{h(i)} y_i + (n_{h(i)} - n_{rh(i)}) w_{h(i)} \operatorname{med}_{y_i>0} y_i \right) \tag{15}$$

Dabei ist $n_{h(i)}$ die Grösse der Netto-Stichprobe in der betreffenden Analysegruppe und $n_{rh(i)}$

Tabelle 9: Vergleich Variationskoeffizient mit und ohne Einsetzungen

Variable	Mit Einsetzungen	Ohne Einsetzungen
DUEMPLOI	0.8	-
B0240T	4.3	5.0
B0244T	7.2	9.1
F0291T	11.1	11.4
F0276T	4.0	4.1
totfep	3.9	4.7
K0205T	5.1	5.2

die Anzahl Beobachtungen, die für die interessierende Variable Werte geliefert haben. Für eine korrekte Schätzung der Varianz dieses Schätzers müsste auch die Korrelation zwischen den beiden Summanden berücksichtigt werden. Dies konnte aber nicht realisiert werden. Die grobe Abschätzung in Tabelle 9 zeigt aber, dass bei gebührender Vorsicht die Gefahr von falschen Schlüssen wegen den unkorrigierten Varianzen klein ist. Im folgenden werden also nur noch die unkorrigierten Varianzen wie sie von der SAS-Prozedur SURVEYMEANS berechnet werden verwendet.

5 Punkt- und Varianzschätzungen

5.1 Gesamttotale

Die Ergebnisse für die Gesamttotale basieren auf 70 Analysegruppen mit 4470 Beobachtungen (Tabelle 10). Das Total der Gewichte ist $T_w = \sum_i w_i = N = 29'871$, also die Populationsgrösse. Bei Abweichungen gegenüber den Ergebnissen in der Publikation für die Erhebungen ((DUMITRICA et al. 2001)) sind ausschliesslich die letzteren Werte massgebend.

Wir schlagen für die Publikation folgende Politik vor: Eine Schätzung muss einen Variationskoeffizienten kleiner als 5% haben, um kommentarlos veröffentlicht zu werden. Für Variationskoeffizienten zwischen 5% und 15% werden die entsprechenden Schätzungen in Klammern angegeben, um dem Leser anzuzeigen, dass eine gewisse Vorsicht geboten ist. Bei Schätzungen mit einem Variationskoeffizienten über 15% werden nur noch leere Klammern () veröffentlicht, um dem Leser anzuzeigen, dass auf Anfrage eine Schätzung erhältlich ist, dass sie aber sehr ungenau ist. Dieses Vorgehen wurde bereits für das statistische Jahrbuch verwendet.

Während die intramuros Ausgaben in der Schweiz (B0240T) und die Anzahl Forschungspersonal (totfep) verlässliche Schätzungen darstellen, können die Investitionen (B0244T), der Aufwand extramuros in der Schweiz (F0291T) und der Aufwand extramuros im Ausland (K205T) zwar angegeben werden, sind aber ungenauer. Mit anderen Worten: Wendet man die obige Politik der Veröffentlichung an, dann können die Gesamttotale B0240T, F0276T und totfep ohne Klammern veröffentlicht werden, während die Totale für B0244T, F0291T und K0205T in Klammern gesetzt werden müssen.

Die gesamten Intramuros-Ausgaben betragen also nach dieser Schätzung 7.7 Mia mit einem

Tabelle 10: Gesamttotale, Standardabweichungen, Variationskoeffizienten

Variable	Beschreibung	Total	SA	VK
DUEMPLOI	Vollzeitäquivalente	1006940	8482	0.8
B0240T	Aufwand intramuros CH	7710178	330802	4.3
B0244T	Investitionen intramuros CH	(905026)	64812	7.2
F0291T	Aufwand extramuros CH	(572188)	63274	11.1
F0276T	Aufwand extramuros Ausl.	1137903	46039	4.0
totfep	Personal F+E	41250	1626	3.9
K0205T	Aufwand intramuros Ausl.	(6489703)	331860	5.1

95%-Vertrauensintervall von 7.1 Mia bis 8.4 Mia. Nehmen wir denselben Variationskoeffizienten (4.3%) für die Schätzung im Jahr 1996 an, dann ergibt sich für 1996 ein Vertrauensintervall von 6.8 Mia +/- 0.6 Mia, also 6.2 Mia bis 7.4 Mia. Die beiden Vertrauensintervalle überlappen sich. Es kann also nicht von einer signifikanten Veränderung gesprochen werden. Nehmen wir weiter an, dass die Stichproben und Antwortausfälle unabhängig waren für die beiden Erhebungen, dann ergibt sich für die Differenz von 7.7-6.8=0.9 Mia ein Vertrauensintervall von 0.9 ± 0.9 Mia also von 0 Mia bis 1.8 Mia. (Verfeinerungen müssten einerseits die Ueberlappung der Stichproben, anderseits aber die Teuerung und die Unterschätzung der Varianz berücksichtigen). Falls wir also von einer Zunahme der intramuros Ausgaben sprechen wollen, dann mit grosser Vorsicht und immer unter dem Hinweis, dass der Unterschied gerade nicht mehr signifikant ist. Aehnliche Untersuchungen können natürlich mit den anderen Totalen angestellt werden.

5.2 Branchen-Totale

Bei einer konsequenten Anwendung der Veröffentlichungspolitik mit Genauigkeits-Hinweisen (Klammerung) können für einige Branchen überhaupt keine Resultate veröffentlicht werden, während für andere Branchen sogar einige Schätzungen ohne Klammern gegeben werden können. Als Alternative könnten für die Branchen-Resultate die Grenzen für die Klammerung leicht gelockert werden, z.B. auf 10% und 25%. Diese Grenzen müssen selbsverständlich kommuniziert werden, um den Benutzer in die Lage zu setzen, die Qualität der Schätzungen für seinen Gebrauch zu beurteilen. Ein Variationskoeffizient von 25% bedeutet, dass das approximative Vertrauensintervall von der Hälfte der Schätzung bis zum 1.5 fachen reicht! Die Resultate für die Variablen B0240T und totfep sind aus der Tabelle 11 ersichtlich. Die letzte Kolonne in Tabelle 11 zeigt den Anteil von Forschungspersonal in den Branchen und illustriert die grosse Heterogenität der Forschungsintensität.

Die Sektion HSW hat sich entschlossen, die ursprünglich 20 Branchen auf 8 zu reduzieren und damit die Genauigkeit zu erhöhen. Der Schlüssel für die Regruppierung ist in Tabelle 12 zu sehen.

Die entsprechenden Resultate sind in Tabelle 13. Es zeigt sich, dass zwar die Variationskoeffizienten kleiner sind als bei 20 Branchen, dass aber immer noch 3 der 8 Branchen totale intramuros Ausgaben CH (B0240T) mit einem Variationskoeffizienten von über 10% haben, wobei keine

Tabelle 11: Branchentotale für B0240T und totfep. Die letzte Kolonne ist totfep/DUEMPLOI in Prozenten. Klammergrenzen sind 10% (einfache Klammern) und 25% (doppelte Klammern).

Branche	n	$\sum w_i$	DUEMPLOI	B0240T	VK	totfep	VK	fep
1	125	181	4334	(723)	13.3	(16)	18.0	0
2	272	997	64637	392267	4.8	(22120)	17.4	34
3	264	611	25737	(10880)	15.7	((203))	30.1	1
4	594	3554	86942	31555	8.8	(325)	21.2	0
5	115	351	33928	594778	4.1	3099	4.7	9
6	79	128	25509	1834027	3.0	5386	3.5	21
7	182	490	23124	(46184)	11.8	355	8.0	2
8	281	534	17895	(8986)	16.9	(161)	18.1	1
9	285	2998	84198	(171072)	18.0	(1217)	17.4	1
10	68	268	20880	(354471)	15.9	(2013)	12.6	10
11	167	1131	64143	(644726)	11.3	(4085)	10.3	6
12	308	2382	163858	(2099530)	13.4	11088	9.4	7
13	303	890	23411	22015	8.2	234	7.9	1
14	135	144	12086	(355)	13.3	(10)	12.6	0
15	331	5206	131149	((4085))	36.8	((192))	40.5	0
16	123	1434	42803	(93651)	13.9	(543)	16.3	1
17	176	196	38194	54169	6.6	196	5.0	1
18	156	1311	32807	((262605))	30.0	((2243))	26.9	7
19	118	533	7170	(876460)	10.5	(5662)	13.5	79
20	388	6532	104133	(207640)	23.1	(2130)	22.4	2

Branche einen VK über 25% hat. Die Anzahl Beschäftigte in F+E zeigt auch noch ziemliche gute Resultate (4 Branchen mit $10\% \leq VK \leq 25\%$). Die anderen Variablen haben immer noch zu grosse Variationskoeffizienten. Würde man der Politik folgen, Schätzungen mit Variationskoeffizient unter 10% zu veröffentlichen, bei Variationskoeffizient zwischen 10% und 25% in Klammern zu veröffentlichen und bei Variationskoeffizient über 25% nur noch leere Klammern zu veröffentlichen, dann hätten B0244T, F0276T und K0205T drei leere Klammern und F0291T hätte vier leere Klammern.

Die Schätzungen der FE00 sind natürlich trotz der z.T. grossen Unsicherheit wichtige Informationen für die Forschungs- und Innovationspolitik der Schweiz. Um die Benutzer (Politiker, Medien, Unternehmen, Vereinigungen) in die Lage zu setzen, die Unsicherheit der Schätzungen angemessen in ihren Entscheidungen berücksichtigen zu können, ist dringend empfohlen, Angaben über die Qualität der Informationen, insbesondere Varianzschätzungen, zusammen mit den Ergebnissen zu veröffentlichen.

Tabelle 12: Regruppierung der Branchen

Neuer Name	8 Branchen	20 Branchen
Maschinen, Metall	1	9, 11, 12
Elektrotechnik	2	10,14
Pharma, Chemie	3	5, 6, 7
Nahrungsmittel	4	2
ICT	5	17, 18
Forschungslabors	6	19, 20
Steine, Erde, Bau	7	1, 8, 15
Andere	8	3, 4, 13, 16, 21

Tabelle 13: Einteilung in 8 Branchen, Totale und Variationskoeffizienten (in Prozent)

Branche8		B0240T	B0244T	F0291T	F0276T	K0205T	totfep
1	Total	2’915’328	279’503	285’948	89’349	339’119	16’389
	VK	10.0	17.6	15.5	22.5	18.1	7.0
2	Total	354’826	53’859	13’771	15’125	28’443	2’024
	VK	15.9	31.1	16.9	45.1	37.4	12.6
3	Total	2’474’988	306’705	54’643	842’743	4’726’743	8’840
	VK	2.4	10.4	7.4	0.4	0.9	2.7
4	Total	392’267	75’171	59’500	22’120	292’000	2’092
	VK	4.8	1.8	29.1	17.4	0.0	5.2
5	Total	316’774	28’944	5’639	9’884	4’276	2’438
	VK	24.9	43.6	34.6	31.8	78.7	24.8
6	Total	1’084’101	97’635	136’615	155’735	1’053’432	7’792
	VK	9.6	16.0	29.9	26.0	30.7	11.5
7	Total	13’794	3’841	10’667	315	0	368
	VK	15.5	40.4	66.1	39.5	.	22.6
8	Total	158’102	59’367	5’404	2’631	45’689	1’305
	VK	8.6	14.8	14.6	13.4	15.0	9.9

6 Anhang

6.1 SAS-Programme

6.1.1 Bibliotheken und Optionen

Dieses Programm ist eine Aenderung des Programms runauto von P. Zangger. Es wird insbesondere für die korrekte Definition der Formate benötigt.

```
/* runauto_Hb.sas: Bibliotheken und Optionen */

libname labels "h:\daten\dok-01\fe00\livraison_RD\rddata" ;
options fmtsearch=(labels.formats); libname Chckdata
"h:\daten\dok-01\fe00\livraison_RD\Checkfile" ; libname Rddata
"h:\daten\dok-01\fe00\livraison_RD\rddata" ; libname rahmen
"h:\daten\dok-01\fe00\livraison_RD"; options linesize=72; options
pagesize=55; run;
```

6.1.2 Recodierungen und Abgleich mit fiktiven Einheiten

```
/* mergfikt.sas: Recodierungen und Abgleich
                 mit fiktiven Einheiten*/
/* Input: Meth_que200029sep01_versionb */
/* Output: rddata.spfe */
title1 'F+E 2000';
title2 'Abgleich fiktive Einheiten';
data d0;
  set rddata.Meth_que200029sep01_versionb;
  totbes=sum(A0109H , A0109F , A0109HF);
        /*sum ignoriert fehlende Werte*/
  totfep=sum(i0600h,i0600f,i0600hf);
  /* Taille ist alphanumerisch formatiert */
  if taille=5 then grkl=2;
  if taille=12 then grkl=1;
  if taille=14 then grkl=3;
/* Die Daten stammen aus der Datei
   liste-für-meth.xls vom 24.9.01 */
data gruppen;
  input bur feinh antw brafil;
  /* Keine Veroeffentlichung der betroffenen BUR-Nummern */
  cards;
...
...
...
;
proc sort data=gruppen;
  by bur;
proc sort data=d0;
  by bur;
/* Hier werden die fikt. Einh. dazugemischt */
data rddata.spfe;
  merge d0 gruppen;
  by bur;
/* Es folgen diverse Auswertungen zum Testen */
proc freq data=rddata.spfe;
  table brafil*branchea/nopercent nocol norow;
run;
proc freq data=rddata.spfe;
  table taille;
  table grkl;
  table branchea;
  table branchea*grkl;
  table strate;
run;
proc tabulate data=rddata.spfe;
  class feinh;
  var totbes duemploi b0240t;
  table feinh,sum*(totbes duemploi b0210t);
run;
proc tabulate data=rddata.spfe;
  class feinh;
  var totbes duemploi b9999t;
  table feinh,sum*(totbes duemploi b9999t);
run;
proc tabulate data=rddata.spfe;
  class feinh;
  var totbes duemploi b0240t;
  table feinh,sum*(totbes duemploi b0240t);
run;
proc tabulate data=rddata.spfe;
  class feinh;
  var totbes duemploi b0244t;
  table feinh,sum*(totbes duemploi b0244t);
run;
proc tabulate data=rddata.spfe;
  class feinh;
  var totbes duemploi f0291t;
  table feinh,sum*(totbes duemploi f0291t);
run;
proc tabulate data=rddata.spfe;
  class feinh;
  var totbes duemploi totfep;
  table feinh,sum*(totbes duemploi totfep);
run;
proc tabulate data=rddata.spfe;
  class feinh;
  var totbes duemploi f0276t;
  table feinh,sum*(totbes duemploi f0276t);
run;
```

6.1.3 Definition Netto-Stichprobe

```
/* missings3.sas: Analyse fehlende Werte und Rekodierung,
   insbesondere Definition der Nett-Stichprobe */
/* Input: rddata.spfe */
/* Output: rddata.spfe */
/* proc univariate data=rddata.spfe;
  var b0210t b0220t b0230t b0240t b0244t b9999t; */
data rddata.spfe;
  set rddata.spfe;
  if (1<=c_saisie_n<=3 or 10<=c_saisie_n<=11) then netto=1;
    else netto=0;
  if b0240t=. then mb0240t=1;
  if b0240t=0 then mb0240t=2;
  if b0240t>0 then mb0240t=3;
  if totfep=. then mtotfep=1;
  if totfep=0 then mtotfep=2;
  if totfep>0 then mtotfep=3;
  if f0291t=. then mf0291t=1;
  if f0291t=0 then mf0291t=2;
  if f0291t>0 then mf0291t=3;
  if f0276t=. then mf0276t=1;
  if f0276t=0 then mf0276t=2;
  if f0276t>0 then mf0276t=3;
  if f0280t=. then mf0280t=1;
  if f0280t=0 then mf0280t=2;
  if f0280t>0 then mf0280t=3;
  if k0205t=. then mk0205t=1;
  if k0205t=0 then mk0205t=2;
  if k0205t>0 then mk0205t=3;
run;
proc freq;
  table c_saisie_i_n*netto;
  table c_saisie_e_n*netto;
  table c_saisie_a_n*netto;
  table c_saisie_n*netto;
```

```
proc freq;
  table mb0240t*netto;
  table mtotfep*netto;
proc freq data=rddata.spfe(where=(netto=1));
  table mb0240t*mtotfep;
  table mb0240t*c_saisie_n;
  table mb0240t*c_saisie_i_n;
  table mf0291t*c_saisie_e_n;
  table mk0205t*c_saisie_a_n;
run;
proc print data=rddata.spfe(where=(mmbur>0));
  var bur mmbur feinh totbes duemploi b0240t totfep;
run;
```

6.1.4 Einsetzungen für fiktive Einheiten

```
/* fillfikt.sas: Einsetzen von Werten
                 für die fiktiven Einheiten */
/* Input: rddata.spfe */
/* Output: rddata.spfe2, rddata.spfe3b, rddata.spfe3n */
/* Zuerst die Aggregate der Zweig-Unternehmen
   der fiktiven Einheiten
   auf Ebene der Branchen, damit die Verteilung
   nachher mit DUEMPLOI gemacht werden kann */
title1 'F+E 2000';
title2 'Einsetzen von Werten für fiktive Einheiten ';
data d0;
  set rddata.spfe(where=(feinh^=.));
  gruppe=int(feinh/100);
proc sort data=d0;
  by feinh; /* Sortiert auch nach Gruppe */
proc means data=d0 sum;
  by feinh; /* Entspricht den Branchen-Totalen der Gruppen*/
  var totbes duemploi totfep b0240t b0244t f0291t
      f0276t f0280t k0205t;
  output out=d1
    sum(totbes duemploi totfep b0240t b0244t f0291t f0276t
        f0280t k0205t)=
    sftotbes sfduemploi sftotfep sfb0240t sfb0244t
    sff0291t sff0276t
    sff0280t sfk0205t;
data d2; /* Branchentotale zu den fiktiven Einheiten mischen */
  merge d0 d1;
  by feinh;
/* Jetzt die Gruppen-Totale berechnen und hinzumischen */
proc sort data=d2(where=(bur<1000));
/* Nur die fiktiven Einheiten */
  by gruppe;
proc means data=d2 mean sum;
  by gruppe;
  var sftotbes sftotfep sfduemploi sftotfep sfb0240t sfb0244t
      sff0291t sff0276t sff0280t sfk0205t;
  output out=d3 sum=/autoname;
data d4; /* Gruppentotale zu fiktiven Einheiten mischen */
  merge d2 d3;
  by gruppe;
proc print data=d4; /* Test-Ausdruck für b0240t */
  var gruppe
      bur
      feinh
      duemploi
      sfduemploi
      sfduemploi_sum
      b0240t
      sfb0240t
      sfb0240t_sum;
/* Einsetzen der Werte bei den fiktiven Einheiten
   prop. zu DUEMPLOI */
data d5;
  set d4;
  duemploi=sfduemploi;
  totbes=sfduemploi*sftotbes_sum/sfduemploi_sum;
  totfep=sfduemploi*sftotfep_sum/sfduemploi_sum;
  b0240t=sfduemploi*sfb0240t_sum/sfduemploi_sum;
  b0244t=sfduemploi*sfb0244t_sum/sfduemploi_sum;
  totfep=sfduemploi*sftotfep_sum/sfduemploi_sum;
  f0291t=sfduemploi*sff0291t_sum/sfduemploi_sum;
  f0276t=sfduemploi*sff0276t_sum/sfduemploi_sum;
  f0280t=sfduemploi*sff0280t_sum/sfduemploi_sum;
  k0205t=sfduemploi*sfk0205t_sum/sfduemploi_sum;
proc print data=d5; /* Test-Ausdruck für alle Var. */
  var gruppe bur feinh
      totbes sftotbes sftotbes_sum
      duemploi sfduemploi sfduemploi_sum
      b0240t sfb0240t sfb0240t_sum
      b0244t sfb0244t sfb0244t_sum
      totfep sftotfep sftotfep_sum
      f0291t sff0291t sff0291t_sum
      f0276t sff0276t sff0276t_sum
      f0280t sff0280t sff0280t_sum
      k0205t sfk0205t sfk0205t_sum;
/* Aufdatieren der Ausgangsdatei */
proc sort data=rddata.spfe;
  by bur;
data rddata.spfe2;
  update rddata.spfe d5;
  by bur;
  aktiv=1;  /* Alle werden aktiv gesetzt */
  /* Zweig-Unt. der fiktiven Einheiten inaktiv setzen*/
  if (feinh^=. and bur>1000) then aktiv=0;
  /* Fiktive Einheiten zu Netto-Stichprobe */
  if (feinh^=. and bur<1000) then netto=1;
  if grkl=. then grkl=3;
  if branchea=. then branchea=mod(feinh,100);
  /* Zweig-Unt. mit nur einer Mutter inaktiv */
  if (mmbur>0 and feinh=.) then aktiv=0;
proc freq data=rddata.spfe2;
  table aktiv*netto;
run;
proc freq data=rddata.spfe2;
  table mb0240t*c_saisie_i_n
   mf0291t*c_saisie_e_n
   mk0205t*c_saisie_a_n
   mb0240t*netto
   mf0291t*netto
   mk0205t*netto
   mtotfep*netto /nopercent nocol norow;
/* Ueberprüfung, ob richtige Unternehmen
   nicht in Netto-Stichprobe */
proc print data=rddata.spfe2 (where=(netto=0 and mb0240t^=1));
  var bur feinh totbes duemploi totfep
      b0240t b0244t f0291t f0276t k0205t;
run;
proc print data=rddata.spfe2 (where=(netto=0 and mk0205t^=1));
  var bur feinh totbes duemploi totfep
      b0240t b0244t f0291t f0276t k0205t;
run;
proc print data=rddata.spfe2 (where=(netto=0 and mf0291t^=1));
  var bur feinh totbes duemploi totfep
      b0240t b0244t f0291t f0276t k0205t;
run;
/* Uebertragen von duemploi der Zweig-Unt. auf Mütter */
data d10; /* Zweig-Unternehmen und Mütter herausziehen */
  set rddata.spfe2;
  /* Aendern Mutter xxx in Mutter yyy mit Zweig zzz */
  if bur=... then filiale='M';
  if bur=... then filiale='S';
  if bur=... then mmbur=...;
  if filiale='M' or mmbur>0;
proc print data=d10;
  var bur feinh totbes duemploi b0240t filiale mmbur;
run;
data d11;
  set d10;
  if feinh=.;
  if mmbur=0 then gbur=bur; else gbur=mmbur;
proc sort data=d11;
  by gbur filiale;
proc means data=d11 sum;
  by gbur;
  var totbes totfep duemploi totfep
      b0240t b0244t f0291t f0276t f0280t k0205t;
  output out=d12 sum=/autoname;
data d13;
  merge d11 d12;
  by gbur;
proc print data=d13;
  var netto bur gbur filiale mmbur totbes duemploi totfep
      b0240t b0244t f0291t f0276t k0205t
      totbes_sum duemploi_sum totfep_sum
      b0240t_sum b0244t_sum f0291t_sum f0276t_sum
      f0280t_sum k0205t_sum;
run;
/* Aufdatierung der Besch. der Mütter */
data d14;
  set d13;
  if filiale='M';
  duemploi=duemploi_sum;
  keep bur duemploi;
proc sort data=d14;
  by bur;
/* Die Mutterunternehmen aufdatieren,
   bereinigte Brutto-Stichprobe */
data rddata.spfe3b;
  update rddata.spfe2 d14;
  by bur;
```

```
  branche=branchea;
  stratind=branche*10+grkl;
  /* Quest. blanc Rubrik k implizert k0205t=0 */
  if c_saisie_a_n=4 then k0205t=0;
  if aktiv=1; /* Zweig-Unt. raus */
/* Netto-Stichprobe */
data rddata.spfe3n;
  set rddata.spfe3b;
  if netto=1;
  keep bur
    branche
    grkl
    stratind
    totbes
 totfep
 duemploi
 b0240t
 b0244t
 f0291t
 f0276t
 f0280t
 k0205t;
run;
```

6.1.5 Korrektur Populationsgrössen

```
/* mergepop.sas: Abgleich mit den Populationsgrössen N und M */
/* Input: rahmen.csrd2000, rddata.Meth_que200029sep01_versionb,
          rddata.spfe3b, rddata.spfe3n */
/* Output: rddata.spfe4, rddata.poptotals */
title1 'F+E 2000';
title2 'Populations-Totale, Korrekturen, Gewichtung' ;
/* N und M im Stichprobenrahmen */
data d0;
  set rahmen.csrd2000;
  if duemploi>=2000 then taille=4; /* autorepräsentativ */
proc freq data=d0;
  table branchea*taille /nopercent nocol norow;
run;
proc means nway sum data=d0;
  class branchea taille ;
  var duemploi;
  output out=d1 sum=;
run;
proc print data=d1;
run;
data rahmen.totals;
  set d1;
  gruppind=branchea*10+taille;
  npop=_FREQ_;
  mpop=duemploi;
  drop _TYPE_ _FREQ_ duemploi;
proc sort data=rahmen.totals;
  by gruppind;
run;
/* n und m in der Brutto-Stichprobe */
data d10;
  set rddata.Meth_que200029sep01_versionb;
  groe=1;
  if duemploi>=2000 then groe=4; /* autorepräsentativ */
  if 2000>duemploi>=100 then groe=3;
  if 100>duemploi>=50 then groe=2;
proc freq data=d10;
  table branchea*groe /nopercent nocol norow;
run;
proc means nway sum data=d10;
  class branchea groe ;
  var duemploi;
  output out=d11 sum=;
run;
proc print data=d11;
run;
data rddata.bruttot;
  set d11;
  gruppind=branchea*10+groe;
  nbrutto=_FREQ_;
  mbrutto=duemploi;
  drop _TYPE_ _FREQ_ duemploi;
proc sort data=rddata.bruttot;
  by gruppind;
run;
/* n und m nach Behandlung der fiktiven und Zweige */
title2 'Brutto-Totale nach Behandlung fiktive E. und Zweig-U.';
data d20;
  set rddata.spfe3b;
  groe=1;
  if duemploi>=2000 then groe=4; /* autorepräsentativ */
  if 2000>duemploi>=100 then groe=3;
  if 100>duemploi>=50 then groe=2;
proc freq data=d20;
  table branche*groe grkl*groe/nopercent nocol norow;
run;
proc means nway sum data=d20;
  class branche groe ;
  var duemploi;
  output out=d21 sum=;
run;
proc print data=d21;
run;
data rddata.bruttotb;
  set d21;
  gruppind=branche*10+groe;
  nbruttob=_FREQ_;
  mbruttob=duemploi;
  drop _TYPE_ _FREQ_ duemploi;
proc sort data=rddata.bruttotb;
  by gruppind;
run;
data d30;
  merge rahmen.totals rddata.bruttot rddata.bruttotb;
  by gruppind;
proc print data=d30;
run;
data d31;
  set d30;
  diffn=nbrutto-nbruttob;
  diffm=mbrutto-mbruttob;
  npopb=npop-diffn;
  mpopb=mpop-diffm;
proc print data=d31;
  var gruppind npop diffn npopb mpop diffm mpopb nbruttob;
run;
proc means sum data=d31;
  var npop diffn npopb mpop diffm mpopb;
run;
data rddata.poptotals;
  set d31;
  _total_=npopb;
  keep gruppind _total_ mpopb nbruttob;
run;
data d100;
  set rddata.spfe3n;
  groe=1;
  if duemploi>=2000 then groe=4; /* autorepräsentativ */
  if 2000>duemploi>=100 then groe=3;
  if 100>duemploi>=50 then groe=2;
  gruppind=branche*10+groe;
proc freq data=d100;
  table branche*groe grkl*groe/nopercent nocol norow;
run;
proc means nway sum data=d100;
  class branche groe ;
  id gruppind;
  var duemploi;
  output out=d101 sum=;
run;
data rddata.poptotals;
  merge rddata.poptotals d101;
  by gruppind;
  ns=_FREQ_;
  ms=duemploi;
  drop _TYPE_ _FREQ_ duemploi;
  nbweight=_total_/nbruttob;
  nweight=_total_/ns;
  mweight=mpopb/ms;
  if groe=4 and branche=20 then do;
    nweight=1;
 mweight=1;
 nbweight=1;
  end;
proc print data=rddata.poptotals;
run;
data rddata.spfe3n;
  set rddata.spfe3n;
  groe=1;
  if duemploi>=2000 then groe=4; /* autorepräsentativ */
  if 2000>duemploi>=100 then groe=3;
  if 100>duemploi>=50 then groe=2;
  gruppind=branche*10+groe;
proc sort data=rddata.spfe3n;
  by gruppind;
data rddata.spfe4;
  merge rddata.spfe3n rddata.poptotals;
  by gruppind;
proc tabulate data=rddata.spfe4;
  class gruppind;
```

```
  var duemploi mpopb;
  weight nbweight;
  table (gruppind all), (n sum*duemploi mean*mpopb);
proc tabulate data=rddata.spfe4;
  class gruppind;
  var duemploi mpopb;
  weight nweight;
  table (gruppind all), (n sum*duemploi mean*mpopb);
proc plot data=rddata.poptotals;
  plot nbweight*nweight;
run;
```

6.1.6 Analyse der Gewichte

```
/* gewan.sas: Analyse der Gewichte */
/* Input: rddata.spfe4 rddata.spfe5 */
/* Ouput: - */
proc univariate data=rddata.spfe4 plots robustscale;
  var nweight nbweight;
*title1 'B0240T vs DUEMPLOI';
*title2 ;
proc gplot data=rddata.spfe5(where=(branche=8));
  *by branche;
  plot b0240t*duemploi/noaxes;
run;
proc gplot data=rddata.spfe4;
  plot mweight*nweight;
run;
proc tabulate data=rddata.spfe4;
  class gruppind;
  var nweight nbweight mweight;
  table (gruppind),(mean*(nbweight nweight  mweight));
run;
/* Dominanzen */
data d0;
  set rddata.spfe4;
  one=1;
proc sort data=d0;
  by nweight;
data d1;
  set d0;
  cumsum+nweight;
proc means data=d1 p95 nway;
  class one;
  var nweight;
  output out=d2 p95=nwp95 sum=nwsum;
data d3;
  merge d1 d2;
  by one;
  if nweight>=nwp95 then large=nweight; else large=0;
  if cumsum>nwsum*0.5 then medom=1; else medom=0;
proc summary data=d3 nway;
 class one;
 var nweight large medom;
 output out=d7 sum= mean= /autoname;
data d8;
  set d7;
  dominance=large_sum/nweight_sum;
proc print data=d8;
run;
proc tabulate data=rddata.spfe4;
  class gruppind;
  var duemploi mpopb;
  weight nbweight;
  table (gruppind all), (n sum*duemploi mean*mpopb);
proc tabulate data=rddata.spfe4;
  class gruppind;
  var duemploi mpopb;
  weight nweight;
  table (gruppind all), (n sum*duemploi mean*mpopb);
proc tabulate data=rddata.spfe4;
  class gruppind;
  var duemploi mpopb;
  weight mweight;
  table (gruppind all), (n sum*duemploi mean*mpopb);
run;
proc tabulate data=rddata.spfe5;
  class branche;
  var b0240t;
  weight nbweight;
  table (branche all), (sum*b0240t);
proc tabulate data=rddata.spfe5 format=12.0;
  class branche;
  var b0240t;
  weight nweight;
  table (branche all), (sum*b0240t);
proc tabulate data=rddata.spfe5 format=12.0;
  class branche;
  var b0240t;
  weight mweight;
  table (branche all), (sum*b0240t);
run;
```

6.1.7 Einsetzung Median

```
/* einsetz_median.sas: Einsetzung des Medians */
/* Input: rddata.spfe4 */
/* Output: rddata.spfe5 */
title1 'F+E 2000';
title2 'Einsetzungen von Medianen pro Analyse-Gruppe';
/* Mediane berechnen pro Analyse-Gruppe für positive Variable */
proc summary data=rddata.spfe4(where=(b0240t>0)) nway;
  class gruppind;
  var b0240t ;
  output out=d1 median=/autoname;
proc summary data=rddata.spfe4(where=(b0244t>0)) nway;
  class gruppind;
  var b0244t;
  output out=d2 median=/autoname;
run;
proc summary data=rddata.spfe4(where=(f0291t>0)) nway;
  class gruppind;
  var f0291t;
  output out=d3 median=/autoname;
proc summary data=rddata.spfe4(where=(f0276t>0)) nway;
  class gruppind;
  var f0276t;
  output out=d4 median=/autoname;
proc summary data=rddata.spfe4(where=(f0280t>0)) nway;
  class gruppind;
  var f0280t;
  output out=d4a median=/autoname;
proc summary data=rddata.spfe4(where=(totfep>0)) nway;
  class gruppind;
  var totfep;
  output out=d5 median=/autoname;
proc summary data=rddata.spfe4(where=(k0205t>0)) nway;
  class gruppind;
  var k0205t;
  output out=d6 median=/autoname;
/* Median einsetzen. Falls er fehlt 0 einsetzen. Flag setzen */
data rddata.spfe5;
  merge rddata.spfe4 d1 d2 d3 d4 d4a d5 d6;
  by gruppind;
  if b0240t=. then do;
     b0240t=b0240t_median;
  if b0240t_median=. then b0240t=0;
     impb0240t=1;
  end;
  else impb0240t=0;
  if b0244t=. then do;
     b0244t=b0244t_median;
  if b0244t_median=. then b0244t=0;
     impb0244t=1;
  end;
  else impb0244t=0;
  if f0291t=. then do;
     f0291t=f0291t_median;
  if f0291t_median=. then f0291t=0;
     impf0291t=1;
  end;
  else impf0291t=0;
  if f0276t=. then do;
     f0276t=f0276t_median;
  if f0276t_median=. then f0276t=0;
     impf0276t=1;
  end;
  else impf0276t=0;
  if f0280t=. then do;
     f0280t=f0280t_median;
  if f0280t_median=. then f0280t=0;
     impf0280t=1;
  end;
  else impf0280t=0;
  if totfep=. then do;
```

```
    totfep=totfep_median;
  if totfep_median=. then totfep=0;
    imptotfep=1;
  end;
  else imptotfep=0;
  if k0205t=. then do;
    k0205t=k0205t_median;
  if k0205t_median=. then k0205t=0;
    impk0205t=1;
  end;
  else impk0205t=0;
/* Test-Berechnung Gruppen-Totale */
proc tabulate data=rddata.spfe5;
  class gruppind;
  var duemploi b0240t b0244t f0291t f0276t f0280t totfep k0205t;
  weight nweight;
  table (gruppind all), ((n sum)*b0240t);
  table (gruppind all), ((n sum)*b0244t);
  table (gruppind all), ((n sum)*f0291t);
  table (gruppind all), ((n sum)*f0276t);
  table (gruppind all), ((n sum)*f0280t);
  table (gruppind all), ((n sum)*totfep);
  table (gruppind all), ((n sum)*k0205t);
run;
```

6.1.8 Schätzer mit PROC SURVEYMEANS

```
/* totals.sas: Schätzt die Branchen-Totale und Varianzen */
/* Input: rddata.spfe5, rddata.poptotals */
/* Output: ---- */
title1 'F+E 2000';
title2 'Branchentotale und Varianzen';
proc surveymeans data=rddata.spfe5 total=rddata.poptotals
                 sum cv clsum min max alpha=0.1;
  strata gruppind/list;
  weight nweight;
  var duemploi b0240t b0244t f0291t f0276t
      f0280t totfep k0205t;
run;
/* Vergleich mit einfacher gewichteter Varianz */
proc surveymeans data=rddata.spfe5 total=29871
    sum cv clsum min max alpha=0.1;
  weight nweight;
  var duemploi b0240t b0244t f0291t f0276t
      f0280t totfep k0205t;
run;
/* Test der Varianzen: fehlende bei Mittelwerten weglassen */
data d1;
  set rddata.spfe5;
  if impb0244t=0;
proc surveymeans data=d1 total=rddata.poptotals
                 mean sum cv ;
  strata gruppind/list;
  weight nweight;
  var b0244t;
run;
/* Zum Vergleich die Statistiken mit
   dem imputierten Datensatz */
proc surveymeans data=rddata.spfe5 total=rddata.poptotals
                 mean cv ;
  strata gruppind/list;
  weight nweight;
  var f0276t;
run;
proc surveymeans data=rddata.spfe5 total=rddata.poptotals
                 sum cv clsum min max;
  by branche;
  strata gruppind;
  weight nweight;
  var duemploi b0240t b0244t f0291t f0276t f0280t totfep k0205t;
run;
data d2;
  set rddata.spfe5;
  if k0205t=0 and impk0205t=1 then k0205t=250;
proc surveymeans data=d2 total=rddata.poptotals
                 sum cv clsum min max alpha=0.1;
  strata gruppind/list;
  weight nweight;
  var k0205t;
run;
data d3;
  set rddata.spfe5;
  aggbra=branche;
  if aggbra=2 then aggbra=1;
data d4;
  set rddata.poptotals;
  aggbra=branche;
  if aggbra=2 then aggbra=1;
proc surveymeans data=d3 total=d4
                 sum cv clsum min max alpha=0.1;
  strata gruppind/list;
  by aggbra;
  weight nweight;
  var b0240t;
run;
/* Hier eine Summe über einen Untersuchungsbereich */
data d5;
  set rddata.spfe5;
  domain=0;
  if branche=19 then domain=1;
  b0240td=b0240t*domain;
proc surveymeans data=d5 total=rddata.poptotals
                 sum cv clsum min max alpha=0.1;
  strata gruppind/list;
  weight nweight;
  var b0240td;
run;
```

Bibliographie

DUMITRICA, R., M. LÉVY und E. PASTOR (2001). *Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000*. Technischer Bericht, Bundesamt für Statistik und économiesuisse.

HULLIGER, B. und M. KASSAB (1998). *Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes*. Methodenbericht, Bundesamt für Statistik.

HULLIGER, B. (1999). *Simple and Robust Estimators for Sampling*. In: *Proceedings of the Section on Survey Research Methods*, S. 54–63. American Statistical Association.

SAS INSTITUTE (1999). *SAS/STAT User's Guide, Version 8*. SAS Institute Inc., Cary NC (USA).

RENFER, J.-P. (2003). *Plan d'échantillonnage pour l'enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse*. Methodenbericht Bestellnummer: 338-0017, Bundesamt für Statistik.

Methodenberichte des Dienstes Statistische Methoden des BFS

Rapports de méthodes du Service de méthodes statistiques de l'OFS

Methodological reports of the Statistical Methods Unit of SFSO

Hulliger, B. (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 é lèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Num éro 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. Num éro 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>>News>>Neuerscheinungen.

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>>Nouvelles 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.

Die Erhebung über Forschung und Entwicklung in der Privatwirtschaft 2000 hat eine hohe Gesamt-Antwortquote erbracht. Der Anteil der Unternehmen, die Forschung und Entwicklung betreiben, ist klein. Umso wichtiger war es, bei den Unternehmen, die F+E betreiben, fehlende Merkmale mit Hilfe von einfachen Modellen zu ersetzen. Der Einfluss dieser Einsetzungen auf die Endresultate ist beträchtlich. Da einige grosse Unternehmensgruppen nur einen Fragebogen für die gesamte Gruppe eingereicht haben, mussten sie mit Hilfe von fiktiven Unternehmen auf die Wirtschaftsbranchen der Erhebung aufgeteilt werden. Die Schätzverfahren sind geschichtete Mittel mit einer Korrektur für die Antwortausfälle. Die Varianzen sind relativ gross. Zwei Grade der Detaillierung der Ergebnisse und die entsprechenden Varianzen werden diskutiert.

Bestellnummer: 338-0018

Bestellungen: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch

Preis: gratis

ISBN 3-303-00264-9