Methodenbericht

Statistik der sozialmedizinischen Institutionen 1999–2004 und Krankenhausstatistik 1999–2002

Einsetzungen für fehlende Daten



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Neuchâtel, 2008

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**

Statistik der sozialmedizinischen Institutionen 1999–2004 und Krankenhausstatistik 1999–2002

Einsetzungen für fehlende Daten

Autor **Daniel Kilchmann**
Bundesamt für Statistik

Herausgeber **Bundesamt für Statistik**

Neuchâtel, 2008

Vorwort

Der Dienst Statistische Methoden des Bundesamts für Statistik wurde mit der Erarbeitung von Einsetzungsverfahren für die Statistik der sozialmedizinischen Institutionen und Krankenhausstatistik beauftragt. Die Arbeiten wurden zwischen 2004 und 2006 durch Daniel Kilchmann durchgeführt. Dabei wurde er von Tania Andreani und André Meister von der Sektion Gesundheit begleitet.

Die Angaben über Resultate der Erhebungen 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 Tania Andreani und André Meister für die sehr gute Zusammenarbeit, sowie Dr. Beat Hulliger, ehemals Dienst Statistische Methoden, nun von der Fachhochschule Nordwestschweiz, und Dr. Monique Graf vom Dienst Statistische Methoden für die fachliche Unterstützung und die aufmerksame Durchsicht des Berichts.

Zusammenfassung

In der Statistik der sozialmedizinischen Institutionen und der Krankenhausstatistik traten in den Jahren 1999 bis 2004, respektive 1998 bis 2002, wenige Antwortausfälle von Institutionen oder Krankenhäusern auf. Zusätzlich ergaben sich aber auch Merkmalsausfälle aufgrund von unvollständig gelieferten Daten. Beide Ausfallstypen wurden mittels einer Nächster-Nachbar-Einsetzungsmethode behandelt. Das Ziel, die multivariaten Zusammenhänge soweit möglich zu berücksichtigen und konsistente Entwicklungen abzubilden, gab den Ausschlag für die Anwendung der Nächster-Nachbar- Einsetzungsmethode. Diese Ziele konnten denn auch grösstenteils erreicht werden. Ausreisser wurden in der Krankenhausstatistik mit einer robustifizierten Version der Nächster-Nachbar-Einsetzungsmethode berücksichtigt. In beiden Datensätzen sind aber auch nach den Einsetzungen Ausreisser und Inkonsistenzen vorhanden. Grund dafür ist die grosse Anzahl sozialmedizinischer Institutionen, resp. Krankenhäuser, die über all die Jahre mindestens einmal unvollständige Daten geliefert hatten und somit nicht als Datenspender verwendet werden konnten. Deshalb wurden nicht alle Variablen eingesetzt, um die Anzahl Datenspender zu vergrössern.

Schlüsselwörter

Methodenbericht; Nächster Nachbar Einsetzung; Transformed Rank Correlations; fehlende Werte; Ausreisser; Statistik der sozialmedizinischen Institutionen; Krankenhausstatistik

Auskunft: Daniel Kilchmann, Tel. 032 713 60 84
Daniel.Kilchmann@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: 338-0050

Preis: gratis

Reihe: Statistik der Schweiz

Fachbereich: 0 Statistische Grundlagen und Übersichten

Originaltext: Deutsch

Graphik/Layout: BFS

Copyright: BFS, Neuchâtel 2008
Abdruck – ausser für kommerzielle Nutzung –
unter Angabe der Quelle gestattet.

ISBN: 978-3-303-00402-9

Inhaltsverzeichnis

1 Einführung **5**

2 Ausreisserlokalisierung **5**
2.1 *Huber-type skipped mean* . . . 6
2.2 *Hidiroglou&Berthelot* Ausreisser Lokalisierung . . . 7

3 Methoden für die Einsetzung **8**
3.1 Qualitäts-Kriterien der Einsetzungen . . . 8
3.2 Modell für die Antwortausfälle und Einsetzungsklassen . . . 8
3.3 Deduktive Einsetzungen . . . 9
3.4 Historisch iterative Einsetzung . . . 9
3.5 *Multiple imputation:* `proc mi` . . . 10
3.6 Nächster-Nachbar Einsetzung (NNE) . . . 10

4 Statistik der sozialmedizinischen Institutionen **14**
4.1 Daten, Variablen und Hilfsvariablen . . . 14
4.2 Antwortausfälle und Spenderpool . . . 15
4.2.1 Antwortausfälle von Einheiten und einzelnen Variablen . . . 16
4.2.2 Definition des Spenderpools . . . 20
4.3 Einsetzungen . . . 25
4.3.1 Resultate der Konservativen HIE und der deduktiven Einsetzung . . . 25
4.3.2 Datenaufbereitung für die NNE . . . 25
4.3.3 Verteilung der Totale in die Unterrubriken . . . 26
4.3.4 Vorgehen und Resultate NNE . . . 27
4.4 Schlussfolgerungen der Einsetzungen in der SM . . . 31

5 Krankenhausstatistik **32**
5.1 Daten, Variablen und Hilfsvariablen . . . 32
5.2 Antwortausfälle, Inkonsistenzen und Spenderpool . . . 33
5.3 Einsetzungen . . . 34
5.4 Schlussfolgerungen der Einsetzungen in der KS . . . 35

6 Schlussfolgerung **36**

Anhang **38**

A Statistik der sozialmedizinischen Institutionen **38**
A.1 Typologie und Variablen . . . 38
A.2 Kontrollregeln . . . 43
A.3 Einfluss der Einsetzungen . . . 46
A.3.1 Ausreisseranalysen . . . 46
A.3.2 Standardtabellen SM 2002 . . . 47
A.3.3 Entwicklung von Indikatoren . . . 50

B Krankenhausstatistik **56**
B.1 Typologie und Variablen . . . 56
B.2 Kontrollregeln . . . 57
B.3 Einsetzungsklassen . . . 58
B.4 Fehlende Werte und Spenderpool . . . 60

B.5 Einfluss der Einsetzungen 60
B.5.1 Entwicklung von Indikatoren 64

Literatur **71**

1 Einführung

Die Sektion 'Gesundheit' (GES) hat die Sektion 'Dienst statistische Methoden' (METH) beauftragt, die fehlenden Werte der 'Statistik der sozialmedizinischen Institutionen' (SM) und die fehlenden Werte der Krankenhausstatistik (KS) für die Jahre 1998 bis 2002 einzusetzen. Anhand der Einsetzungen sollen ausschliesslich Zeitreihen Analysen von Indikatoren ermöglicht werden. Nachdem diese Einsetzungen vorgenommen worden waren, wurde entschieden, die fehlenden Werte der SM für die Jahre 1999 bis 2004 mit der erarbeiteten Methode einzusetzen. Der vorliegende Methodenbericht behandelt vor allem die Erarbeitung der Methoden für die Einsetzungen in der SM und deren Anwendung in der SM und der KS.

In den ersten beiden Abschnitten werden die verwendeten Methoden für die Ausreisserlokalisierung und die Einsetzung vorgestellt. Für die Ausreisserlokalisierung wurden vor allem die *Huber-type skipped mean* Verwerfungsregel und die *Hidiroglou& Berthelot* Methode aufgezeigt. Eine so genannte 'deduktive' und eine so genannt 'Historisch iterative' Einsetzung werden definiert. Darauf folgen Überlegungen, ob die SAS Prozedur `proc mi` verwendet werden könnte. Die ersten beiden Methoden kamen schlussendlich teilweise zur Anwendung. Danach wird eine 'Nächster-Nachbar' Einsetzungsmethode beschrieben, die für die endgültigen Einsetzungen verwendet wurde.

Im Abschnitt 4 wird die Anwendung der eingeführten Methoden an die SM 1998-2002 diskutiert. Innerhalb dieses Abschnitts wird auf die grundlegende Definition der Einheiten, Variablen und des Spenderpools eingegangen, bevor die verschiedenen Einsetzungsverfahren angewandt werden.
Hierauf folgt eine kurze Diskussion der Anwendung der erarbeiteten Methoden auf die SM 1999-2004.

Im Abschnitt 5 wird die Anwendung der Nächster-Nachbar Einsetzungsmethode an die KS der Jahre 1998-2002 aufgezeigt.

Zur Kontrolle der Einsetzungen wurden die Kontrollen und die Methoden zur Ausreisserlokalisierung verwendet. GES wurde zusätzlich zu den vervollständigten Daten die KS-Standardtabellen der Jahre 1998 bis 2002, die SM-Standardtabellen 1999-2004 und die Entwicklung einiger Indices zur Kontrolle geliefert. Diese werden teilweise im Anhang aufgeführt.

Dem Anhang geht eine kurze allgemeine Schlussfolgerungen voraus.

2 Ausreisserlokalisierung

Ein wichtiger Schritt innerhalb der Datenkontrolle ist die Ausreisserlokalisierung. Anhand der Ausreisserlokalisierung soll verhindert werden, dass Ausreisser eingesetzt werden. Sie kann auch dazu verwendet werden, um Einsetzungen für offensichtliche Ausreisser vorzunehmen. Vier Typen von univariaten Ausreissern werden unterschieden

1. Univariate Ausreisser im Erhebungsjahr t
2. Ungewöhnliche Grössenverhältnisse zweier Variablen im Erhebungsjahr t
3. Variablen, die aussergewöhnliche Entwicklungen von einer Periode zur nächsten aufweisen
4. Ungewöhnliche Entwicklung der Grössenverhältnisse zweier Variablen von einer Periode zur nächsten.

Die Typen 1 und 3 wurden mit statistischen Methoden lokalisiert. Die Typen 2 und 4 wurden graphisch veranschaulicht. Der Typ 1 wurde anhand der Methode des Abschnitts 2.1, welche in (Ruckstuhl 2004) '*Huber-type skipped mean* Verwerfungsregel' genannt wird, eruiert. Der Typ 3. wurde mit der Hidiroglou&Berthelot Methode, die im Abschnitt 2.2 beschrieben ist, lokalisiert.
Die verwendeten Methoden sind so parametrisiert, dass anhand der Parameter die Anzahl Werte, die als Ausreisser bezeichnet werden, mit beeinflusst wird. Es gibt kein objektives Kriterium, um Ausreisser zu erkennen. Oft kann aber aufgrund einer Visualisierung der Ausreisser eine Übereinkunft mit den Fachexperten erzielt werden, um eine Grenze für die Entdeckung von Ausreissern festzulegen.
In den beiden folgenden Methoden werden die Variablenwerte so transformiert, dass univariate Abweichungen von den 'üblichen' Werten inklusive zulässiger Abweichung eruiert werden können. Der Entscheid, wann ein Wert einer ausserordentliche, sprich signifikante, Abweichung vom 'üblichen' Werte darstellt, ist dabei wie bereits erwähnt, nicht objektiv und hängt von wählbaren Parametern ab.
Im vorliegenden Bericht geht es weniger darum, Ausreisser zu entdecken, sondern darum, den Einfluss der Einsetzungen auf die Datenqualität abzuschätzen. Damit soll verhindert werden, dass sich die Datenqualität durch die Einsetzungen verschlechtert.

2.1 *Huber-type skipped mean*

Die *Huber-type skipped mean* Verwerfungsregel sollte nur an Variablen ohne strukturell fehlende Werte angewandt werden. Ansonsten sollten die Einheiten mit strukturell fehlenden Werten ausgeschlossen werden. Der Median und die MAD (mediane absolute Abweichung) [1] der Variablen x zum Zeitpunkt t sind innerhalb der Klasse C folgendermassen definiert:

$$\mathrm{med}_C(x) \doteq \underset{k\in C}{\mathrm{med}}(x) \doteq \begin{cases} x_{[\frac{n+1}{2}]} & \text{wenn die Klasse } C \text{ eine ungerade Anzahl Elemente enthält} \\ \frac{1}{2}\left(x_{[\frac{n}{2}]} + x_{[\frac{n+2}{2}]}\right) & \text{andernfalls.} \end{cases} \quad (1)$$

$$\mathrm{MAD}_C(x) \doteq \underset{k\in C}{\mathrm{MAD}}(x) \doteq \underset{k\in C}{\mathrm{med}}(|x_k - \mathrm{med}_C(x)|) \quad (2)$$

$$s_{MAD,C}(x) \doteq \frac{\mathrm{MAD}_C(x)}{0.6745} \quad (3)$$

wobei$x_{[j]} =$ der Variablenwert der Variablen x zum Zeitpunkt t der Einheit, die den j-grössten Wert hat.
Die Korrektur $\frac{1}{0.6745}$ in (3) wird benötigt, damit $s_{MAD,C}$ ein konsistenter Schätzer der Standardabweichung einer Normalverteilung ist.
Der robust zentriert und standardisierte Wert der Variablen x der Einheit $k \in C$ ist somit folgendermassen definiert:

$$\tilde{x}_k = \frac{x_k - \mathrm{med}_C(x)}{s_{MAD,C}(x)} \quad (4)$$

Wenn $|\tilde{x}_k| > \kappa(x)$, dann wird die Einheit k für die Variable x und die Periode t als Ausreisser gekennzeichnet. $\kappa(x)$ ist ein Parameter, der für jede Variable x gewählt werden muss. Als Robustheitsgewicht ausgedrückt, $u_k = \min(1, \kappa(x)/|\tilde{x}_k|)$, kann (4) als ein Ein-Schritt M-Schätzer betrachtet werden. Bei rechtsschiefen Verteilungen ist oft $\kappa(x) \geq 5$, cf. (Hulliger 1999). Normalerweise muss $\kappa(x)$ mit Fachspezialisten gewählt werden, der die Effizienz der Ausreisserentdeckung abschätzen kann.

[1] MAD='Median Absolute Deviation'

2.2 *Hidiroglou&Berthelot* Ausreisser Lokalisierung

Die folgende Methode für die Lokalisierung von Ausreissern der Entwicklung basiert auf der Methode, die von Hidiroglou&Berthelot (Hidiroglou and Berthelot 1986) entwickelt wurde. Eine leicht abgeänderte Form der H&B Methode wurde in einem SAS Macro implementiert.
Es sei $x_{k,t}$: der Variablenwert der Variablen x der Einheit k zum Zeitpunkt t. Wobei k der Klasse C angehört, welche so gebildet wurden, dass ihre Mitglieder sich gleichen, zum Beispiel anhand einer Klumpen-Methode. Die relative Veränderung der Variablen x von der Periode $t-1$ zu t ist für die Einheit k

$$r_{k,t} = \frac{x_{k,t}}{x_{k,t-1}}. \tag{5}$$

sofern $x_{k,t-1} \neq 0$. Um an beiden Enden der Verteilung der Variablen x Ausreisser zu erkennen, haben Hidiroglou und Berthelot $r_{k,t}$ transformiert:

$$s_C(r_{k,t}) = \begin{cases} 1 - \frac{m}{r_{k,t}}, & \text{wenn } 0 < r_{k,t} < m \\ \frac{r_{k,t}}{m} - 1, & \text{wenn } r_{k,t} \geq m \end{cases} \tag{6}$$

Wobei $m = \mathrm{med}_C(r_{k,t})$ ist.
Um die Grössenordnung der Variablen x mit einzubeziehen haben H&B folgende Transformation vorgenommen

$$E_{k,t}^{HB} = s_C(r_{k,t}) \left[\max(x_{k,t}; x_{k,t-1})\right]^u \tag{7}$$

wobei $0 \leq u \leq 1$.
u ist ein Parameter, der erlaubt, die Wichtigkeit der Grössenordnung zu kontrollieren. Anhand dieser Transformation kann mehr Gewicht auf eine kleine Änderung einer grossen Institution gelegt werden (u nahe 1) als auf eine grosse Veränderung einer kleinen Institution (u nahe 0). Eine leicht verschiedene Transformation wurde verwendet, damit $s_C(r_{k,t})$ und das Argument des Logarithmus in etwa die gleiche Grössenordnung aufweisen:

$$E_{k,t} = s_C(r_{k,t}) \log\left(\left[\max(x_{k,t}; x_{k,t-1})\right]^u\right) \tag{8}$$

Aufgrund des Medians, dem 1. Quartil ($\mathrm{Q1}_C()$) und dem 3. Quartil ($\mathrm{Q3}_C()$) werden zwei Abweichungen für die Variable x, der Klasse C, für die Periode t definiert

$$d_{C,1}(E_{k,t}) = \max\left(\mathrm{med}_C(E_{k,t}) - \mathrm{Q1}_C(E_{k,t}), |A\,\mathrm{med}_C(E_{k,t})|\right) \tag{9}$$
$$d_{C,3}(E_{k,t}) = \max\left(\mathrm{Q3}_C(E_{k,t}) - \mathrm{med}_C(E_{k,t}), |A\,\mathrm{med}_C(E_{k,t})|\right) \tag{10}$$

wobei A ein Parameter ist, der Probleme verhindern soll, wenn $\mathrm{med}_C(E_{k,t}) - \mathrm{Q1}_C(E_{k,t})$ oder $\mathrm{Q3}_C(E_{k,t}) - \mathrm{med}_C(E_{k,t})$ sehr klein ist. Wenn in diesem Falle die Werte von $E_{k,t}$ nahe um einen Wert versammelt sind und nur einzelne um Weniges abweichende Werte vorhanden sind, sollen diese nicht als Ausreisser bezeichnet werden. Als Ausreisser für die Variable x wird für die Periode t schlussendlich jede Einheit k gekennzeichnet, wenn

$$E_{k,t} \notin \left(\mathrm{med}_C(E_{k,t}) - Ld_{C,1}(E_i)\ ;\ \mathrm{med}_C(E_{k,t}) + Ld_{C,3}(E_i)\right)$$

wobei L ein veränderbarer Parameter ist, der die Grösse des Intervalls mitbestimmt.

3 Methoden für die Einsetzung

3.1 Qualitäts-Kriterien der Einsetzungen

In (EUREDIT Project 2004) sind objektive Evaluationskriterien von Einsetzungsmethoden definiert. Diese setzen voraus, dass für eine Stichprobe der Daten die 'wahren' Werte bekannt sind. Das heisst, dass diese Stichprobe so aufgearbeitet wurde, dass sie keine fehlenden Werte, Inkonsistenzen und Ausreisser mehr aufweist. Da keine solche Stichprobe zur Verfügung steht, wurden die Daten vor und nach der Einsetzung auf Konsistenz und Ausreisser untersucht. Somit sind keine objektiven Kriterien vorhanden, anhand derer bestimmt werden kann, ob die Einsetzungen von guter Qualität sind oder nicht.
Einsetzungen sollten folgende Eigenschaften haben

1. fehlende Werte vollständig einsetzen
2. realistische Werte im Vergleich zu den bestehenden Werten für jede Einheit einsetzen
3. bestehende Inkonsistenzen lösen
4. keine neuen Inkonsistenzen hinzufügen
5. die multivariate Verteilung der Einsetzungsvariablen nicht verletzen
6. keine Verzerrung kreieren
7. keine Ausreisser kreieren.

Ein grosser Anteil an inkonsistenter oder fehlender Daten kann dazu führen, dass die Kriterien 3., 4., 6. und 7. nicht erfüllt werden können, da die Information, welche für die Einsetzungen verwendet werden kann, zu beschränkt ist. Insbesondere wenn nicht alle Variablen eingesetzt werden, kann auch das Kriterium 4. nicht erfüllt werden.

3.2 Modell für die Antwortausfälle und Einsetzungsklassen

Aufgrund eines Modells für die Antwortausfälle soll die Verzerrung, die durch das Antwortverhalten hervorgerufen wird, minimiert werden. Dabei werden Daten, die das gleiche Antwortverhalten haben in **Einsetzungsklassen** gruppiert, innerhalb derer die Einsetzungen durchgeführt werden.
Bei diesem Vorgehen wird davon ausgegangen, dass das Antwortverhalten durch die verwendeten Hilfsvariablen, zu einem grossen Teil modelliert werden kann. In diesem Abschnitt werden die Notationen aus (Haziza 2002) und (Haziza 2003) verwendet.

Modell für die Antwortausfälle
Es sei die Indikator Variable $a_{k,t}$, die das Vorhandensein einer Antwort folgendermassen beschreibt

$$a_{k,t} = \begin{cases} 1 & \text{wenn die Einheit } k \text{ für die Periode } t \text{ eine Antwort gegeben hat} \\ 0 & \text{andernfalls.} \end{cases}$$

Die Antwortwahrscheinlichkeit, $p_{k,t} = P(a_{k,t} = 1)$, wird anhand eines logistischen Regressionsmodells modelliert

$$\log \frac{p_{k,t}}{1 - p_{k,t}} = \mathbf{z}_{k,t}^{'} \gamma \tag{11}$$

wobei $\mathbf{z}_{k,t}^{'}$ der Vektor der Hilfsvariablen ist, die für sämtliche Einheiten zum Zeitpunkt t vorhanden sind. γ ist der Vektor der Regressionsparameter, die geschätzt werden müssen.

Vorausgesetzt dass die Einheiten unabhängig voneinander antworten, so ist der Antwortmechanismus durch eine *Bernoulli*-Verteilung charakterisiert

$$a_{k,t} \sim B(1, p_{k,t}), \quad k = 1, \ldots, N_t$$

Einsetzungsklassen
Aufgrund des Modells (11) wird eine Vorhersage für die Antwortwahrscheinlichkeit $p_{k,t}$ gemacht, die mit $\hat{p}_{k,t}$ bezeichnet wird. Die Vorhersagen $\hat{p}_{k,t}$ der N_t Einheiten werden der Grösse nach geordnet, so dass $\hat{p}_{[1],t} \leq \hat{p}_{[2],t} \leq \ldots \leq \hat{p}_{[N_t],t}$. Danach werden Einsetzungsklassen C gebildet, die in etwa die gleiche Anzahl Einheiten enthalten. Diese Methode wird in (Haziza 2003) 'méthode des quantiles égaux' genannt.

3.3 Deduktive Einsetzungen

Eine **Deduktive Einsetzung** bezeichnet die Ableitung eines Variablenwerts von anderen Werten, welche ebenfalls einzelne Variablenwerte sein können oder auch Aggregate. Es gibt drei Voraussetzungen, die eine deduktive Einsetzung erlauben:

1. Die einzusetzende Variable wird mehrfach erhoben.
2. Die Aggregationsniveaus der mehrfach erhobenen Variablen müssen übereinstimmen.
3. Der vorhandene Wert muss als 'wahr' angesehen werden und darf somit mit keinen anderen Variablen im Widerspruch stehen.

Beispiel: In der Rubrik 2.b des SM Fragebogens wird der Personalbestand nach Tätigkeitsbereich pro Ausbildungsniveau gefragt. Um den Personalbestand nach Ausbildungsniveau zu erhalten, können die Totale pro Ausbildungsniveau über die Tätigkeitsbereiche summiert werden. Somit kann der Personalbestand (Vollzeitäquivalente) nach Ausbildungsniveau in der Rubrik 2.a von der Rubrik 2.b abgeleitet werden. Die umgekehrte Ableitung, von 2.a nach 2.b, ist nicht möglich, da das Aggregationsniveau der beiden Fragen nicht übereinstimmt.
In der Frage 3 kann die Anzahl Bewohner in den einzelnen Rubriken (3.1, 3.3, 3.4) nicht von den anderen Rubriken abgeleitet werden, da das Aggregationsniveau bei jeder Rubrik verschieden ist.

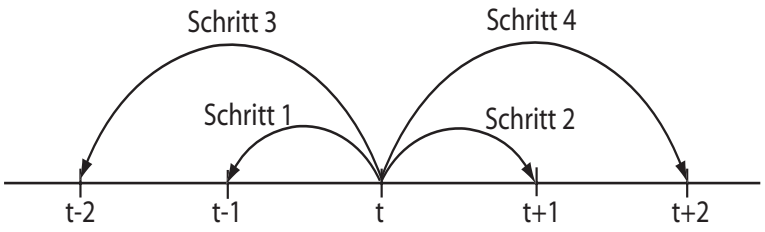
3.4 Historisch iterative Einsetzung

Ein iteratives Einsetzungsverfahren, welches der fehlerhaften Institution den zeitlich am nächsten liegenden Wert derselben Institution einsetzt, wird **historisch iterative Einsetzung** (HIE) genannt. Die Voraussetzung für eine HIE sind:

1. Eine eindeutige Identifikation aller Institutionen Einheiten über sämtliche Perioden hinweg, währen denen sie in der Erhebung sind.
2. Der eingesetzte Wert sollte keine Widersprüche hervorrufen.

Die zweite Voraussetzung kann meist nicht garantiert werden, da ein Wert aus einer anderen Erhebungsperiode eingesetzt wird, der mit der aktuellen nicht übereinstimmen muss. Widersprüche können oft nicht verhindert werden. Diese Methode eignet sich vor allem für Variablen, die über die Zeit relativ konstant sind, wie zum Beispiel die Typologie.
Die HIE kann in einer konservativen und einer progressiven Form angewandt werden. In der konservativen wird die Vergangenheit der Zukunft vorgezogen, wobei aber der zeitliche Abstand des Spenders immer minimiert wird (siehe Abbildung 1). In der progressiven Form wird bei minimalem zeitlichem Abstand die Zukunft der Vergangenheit vorgezogen.

Abbildung 1 Darstellung der konservative Form der historisch iterativen Einsetzung, wobei in der Periode t ein Variablenwert fehlt und nur in der Periode $t + 2$ ein Wert vorhanden ist.



3.5 *Multiple imputation:* `proc mi`

Die Mehrfacheinsetzung, die in SAS implementiert ist, verwendet standardmässig eine Markov-Ketten-Monte-Carlo (MCMC[2]) Methode. Für weitere Erläuterungen wird auf die Hilfedatei von SAS unter `proc mi` verwiesen und auf (Schafer 2000). Dabei wird von ignorierbaren fehlenden Werten ausgegangen, das heisst, dass die Verteilung der fehlenden Werte nur von den Variablen abhängt, die keine fehlenden Werte aufweisen (*missing at random: MAR*) und dass die Parameter des Modells für die fehlenden Werte verschieden von den Parametern der Datenmodelle sind (*distinctness*), die für die Schätzungen verwendet werden. Hinzu kommt, dass die MCMC Methode von einer multivariaten Normalverteilung der Variablen ausgeht. Durch diese Hypothesen werden diskrete Variablen bereits ausgeschlossen. Sämtliche Variablen, die eingesetzt werden sollen, sind nach unten mit der Grenze 0 begrenzt. Dies stellt ein Problem für die Hypothese der multivariaten Normalverteilung dar. Es gilt auch zu beachten, dass die Normalverteilung sehr anfällig auf Ausreisser ist, da sie durch zwei nicht robuste Parameter definiert ist. Wie in den Abschnitten 4.2 und 5.2 ausgeführt wird, weisen die Daten der SM und der KS einige Ausreisser auf.
Aufgrund dieser Überlegungen wurde `proc mi` nicht angewandt.

3.6 Nächster-Nachbar Einsetzung (NNE)

Eine Einheit wird **Spender** genannt, wenn alle seine Daten vollständig und konsistent sind. Diese Einheiten können ohne eine Veränderung zu erfahren, für Analysen und Datenauswertungen verwendet werden. Eine Einheit, die kein Spender ist, und die nur durch Veränderung seiner Variablenwerte in die Analysen und Datenauswertungen miteinbezogen werden kann, wird **Empfänger** genannt. Eine Einheit, die weder Spender noch Empfänger ist, wird **Aussenseiter** genannt. Hierbei kann es sich um Einheiten handeln, die Inkonsistenzen aufweisen, die nicht gelöst werden können. Oder es kann sich um Einheiten handeln, die Ausreisser oder ungewollte Variablenwerte aufweisen, welche vor den Analysen nicht verändert werden oder in einem speziellen Verfahren behandelt werden. Um zu verhindern, dass solche Variablenkombinationen eingesetzt werden, werden diese Einheiten nicht den Spendern zugeordnet. Die Menge aller Spender wird **Spenderpool** genannt.
Bei der Nächster-Nachbar Einsetzung wird anhand von Distanzen eine Nachbarschaft definiert, aus der ein Spender nach gewissen Kriterien ausgesucht wird. Wenn es sich dabei um den nächsten Nachbar handelt, so spricht man allgemein von einer **nächsten Nachbar Einsetzung**. Dieses Verfahren hat zum Vorteil, dass auf individueller Ebene immer die 'optimalste' Einsetzung vorgenommen wird. Je mehr derselbe Spender für Einsetzungen verwendet wird, desto eher führen die Einsetzungen zu einem Bias und je mehr wird die Varianz unterschätzt. Um diesen Nachteilen der Einsetzung des nächsten Nachbars entgegenzuwirken, kann eine

[2] Markov chain Monte Carlo.

Zufallsauswahl in einer eingeschränkten Nachbarschaft verwendet werden. Da die 'Optimalität' der Einsetzung mit steigender Distanz abnimmt, wird die Zufallsauswahl normalerweise umgekehrt proportional zur Distanz durchgeführt.
Aus diesen Gründen wird eine Methode vorgezogen, bei der ein Parameter λ definiert ist, aufgrund dessen die maximale Grösse der Nachbarschaft festgelegt werden kann. Bei diesem Auswahlverfahren wird näheren Nachbarn somit die grössere Wahrscheinlichkeit zugeordnet, ausgewählt zu werden, als weniger nahen Nachbarn. Dieses Vorgehen wird im folgenden mit **NNE** bezeichnet.
Mit der verwendeten NNE Methode werden nur die Variablen eingesetzt, die fehlend sind. Dadurch wird riskiert, dass durch die Einsetzung neue Inkonsistenzen mit vorhandenen Werten auftreten. Würden hingegen sämtliche Daten des Spenders übernommen, so wäre die eingesetzte Einheit immer konsistent. Einen grosser Teil der vorhandenen Information dieser Einheit würde aber verloren gehen. Zudem würde die Varianz der Variablen künstlich verringert. Um die Konsistenz der Entwicklung soweit möglich beizubehalten, wird die Institutionen Einheit als Einheit festgelegt.

Distanz Berechnung und Spenderauswahl

$\mathbf{L_1}$-Distanz

Damit alle Variablen das gleiche Gewicht, respektive die gleiche Grössenordnung, bei der Distanz Berechnung haben, wurden sie zuerst zentriert und standardisiert. Die L_1 Distanz wird anhand der transformierten Variablen berechnet.

Der zentrierte und standardisierte Wert der Variablen x_i der Institutionen Einheit k der Periode t, die sich in der Einsetzungsklasse C befindet, ist

$$\tilde{x}_{i,k,t} = \begin{cases} . & \doteq \text{fehlend, wenn } x_{i,k,t} \text{ fehlend ist} \\ \min(\kappa, x_{i,k,t}) & \text{wenn } \mathrm{MAD}_{C,t}(x_i) = 0, 0 < \kappa \ll 1 \\ \frac{x_{i,k,t} - \mathrm{med}_{C,t}(x_i)}{\mathrm{MAD}_{C,t}(x_i)} & \text{sonst.} \end{cases} \tag{12}$$

Die folgende Distanzdefinition basiert auf der Jahres Einheit. Sie hängt davon ab, ob der Empfänger j in der Periode t nur fehlende Werte aufweist, oder ob er in der Periode t nur einzelne fehlende Werte besitzt. Wenn j nur einzelne fehlende Werte aufweist, ist die Distanz zwischen $x_{i,j,t}$ und $x_{i,k,t}$ wie folgt definiert, wobei j der Empfänger und k der Spender ist

$$d_{m,i,j,k,t} \doteq \begin{cases} 0 & \text{wenn } x_{i,j,t} = . \text{ und } x_{i,k,t} \neq . \text{ und } j,k \in C \\ \infty & \text{wenn } x_{i,j,t} = . = x_{i,k,t} \text{ oder } j \in C, k \notin C \\ |\tilde{x}_{i,j,t}| & \text{wenn } x_{i,j,t} \neq . \text{ und } x_{i,k,t} = . \text{ und } j,k \in C \\ |\tilde{x}_{i,j,t} - \tilde{x}_{i,k,t}| & \text{andernfalls.} \end{cases} \tag{13}$$

Mit der Distanz d_m in (13) wird verhindert, dass ein fehlender Wert durch einen fehlenden Wert ersetzt wird, indem die Distanz in diesem Fall auf ∞ gesetzt wird. Hingegen ist die Distanz 0, wenn ein fehlender Wert durch einen vorhandenen ersetzt werden kann. Weist der Spender einen fehlenden Wert auf, so wird dieser in der Distanzberechnung als 0 angesehen.
Ist der Empfänger j eine fehlende Einheit, so ist die Distanz zwischen $x_{i,j,t}$ und $x_{i,k,t}$ wie folgt definiert, wenn die Einheiten j und k der Einsetzungsklasse C angehören, und k im Spenderpool ist

$$d_{u,i,j,k,t} \doteq \begin{cases} 0 & \text{wenn } x_{i,j,t} = . \text{ und } j,k \in C \\ \infty & \text{wenn } j \in C, k \notin C \\ |\tilde{x}_{i,j,t}| & \text{wenn } x_{i,j,t} \neq . \text{ und } x_{i,k,t} = . \text{ und } j,k \in C \\ |\tilde{x}_{i,j,t} - \tilde{x}_{i,k,t}| & \text{andernfalls.} \end{cases} \tag{14}$$

Im Gegensatz zu d_m wird mit d_u erlaubt, dass fehlende Werte vom Spender übernommen werden. In den allermeisten Fällen ist $d_u = 0$, da nur wenige Empfänger Variablenwerte aufweisen. Die Anzahl Variablen, auf welcher die Berechnung der Distanz beruht, wird mit $n_{j,t}$ bezeichnet. Dies ist die Anzahl Variablen, welche beim Empfänger in der Periode t vorhanden sind, d.h. $n_{j,t} = \max\left(1, \sum_{i \in C} 1_{x_{i,j,t} \neq .}\right)$. Die durchschnittlichen Distanzen zwischen dem Empfänger j und dem Spender k für die Periode t ist somit folgendermassen definiert:

$$\bar{d}_{m,j,k,t} = \frac{1}{n_{j,t}} \sum_i d_{m,i,j,k,t} \tag{15}$$

$$\bar{d}_{u,j,k,t} = \frac{1}{n_{j,t}} \sum_i d_{u,i,j,k,t} \tag{16}$$

Es sei die Indikator Variable, die indiziert, ob dem Empfänger j während der Periode t nur einzelne Variablen fehlen oder sämtliche Variablen fehlen (*unit nonresponse*):

$$I_{j,t} = \begin{cases} 1 & \text{wenn } j \text{ während } t \text{ nur einzelne Variablen fehlen} \\ 0 & \text{andernfalls.} \end{cases}$$

Eine zweite Variable zeigt an, ob der Spender k während jeder Periode in der Erhebung war, während derer j präsent war. Anhand dieser Variablen soll sicher gestellt werden, dass sämtliche Perioden des Empfängers gleichzeitig eingesetzt werden. Zudem soll damit die Entwicklung miteinbezogen werden.

$$E_{j,k} = \begin{cases} 1 & \text{wenn die Einheit } k \text{ in allen Perioden vorhanden ist,} \\ & \text{in denen } j \text{ in der Erhebung ist.} \\ \infty & \text{sonst.} \end{cases}$$

Die Totale Distanz zwischen der Einheit j, dem Empfänger, und der Einheit k, dem Spender, ist dann

$$D_{j,k} = E_{j,k} \left[\sum_t I_{j,t} \bar{d}_{m,j,k,t} + \sum_t (1 - I_{j,t}) \bar{d}_{u,j,k,t} \right] \tag{17}$$

Die Einsetzung der Werte vom Spender k in den Empfänger j ist nur dann zulässig, wenn $D_{j,k} < \infty$.

Mahalanobis-Distanz

Im Folgenden werden zwei Einheiten j und k zur Veranschaulichung verwendet, die p Variablenwerte x und y besitzen, wobei bei der Einheit j nur q Variablenwerte vorhanden sind. Die Untermenge der vorhandenen Variablen wird mit O bezeichnet ($|O| = q$), x_o sind die Komponenten von x mit vorhandenen Werten, x_m jene mit fehlenden Werten. Die Distanz zwischen j und k wird mit $d_{j,k}$ bezeichnet.

Die **marginale Mahalanobis Distanz** ist $d_{j,k} = \frac{p}{q}(x_o - y_o)^T S_{oo}^{-1}(x_o - y_o)$. S_{oo} wird anhand der TRC[3]-Methode geschätzt, cf. (Béguin and Hulliger 2004). Ist S_{oo} singulär, so werden die Variablen, die für die Singularität verantwortlich sind, aus der Distanzberechnung ausgeschlossen, und die Mahalanobis Distanz wird auf den verbleibenden Variablen berechnet. Die Totale Distanz (17) ist dann

$$D_{j,k} = E_{j,k} d_{j,k} \tag{18}$$

[3] *Transformed Rank Correlations*

Spender Auswahl
Es sei m die Anzahl Spender, die gefunden wurden. Diese werden zuerst zufällig geordnet. Dies garantiert eine zufällige Einsetzung bei gleichen Distanzen. Danach werden die Spender nach aufsteigender Distanz $D_{j,k}$ geordnet, d.h. $D_{j,[1]} \leq D_{j,[2]} \leq \ldots < D_{j,[m]} < \infty$. Für die Einsetzung kommen λ Spender in Frage, wenn $\lambda \leq m$. Es sei $\xi = \min(\lambda, m)$. Die Spender sollen proportional zu ihrer Nähe zum Empfänger eingesetzt werden. Hierfür sei

$$T_j \doteq \sum_{k=1}^{\xi} \tilde{D}_{j,[k]} \doteq \sum_{k=1}^{\xi} \frac{1}{\tau + D_{j,[k]}}, \qquad 0 < \tau \text{ klein.}$$

Die Konstante τ im Nenner dient dazu, dass die Distanz immer definiert ist. Bei sehr kleinen $D_{j,[k]}$ wird das Auswahlverfahren durch die Konstante τ beeinflusst. Der Einfluss von τ steigt, je kleiner alle Distanzen $D_{j,[k]}$ sind. In diesem Fall ist dies aber nicht erheblich. Denn bei kleinen Distanzen aller Spender, sind alle Spender sehr nahe beim Empfänger und die Auswahl des Spenders wird keine grosse Auswirkungen haben.
Die Wahrscheinlichkeit, dass der Spender k, dem Empfänger j wirklich als Spender dient ist

$$p_{j,[k]} = \begin{cases} 0 & \text{wenn } k > \xi \\ \frac{\tilde{D}_{j,[k]}}{T_j} & \text{wenn } k \leq \xi \end{cases} \tag{19}$$

Der Spender wird schlussendlich durch eine Bernoulli Zufallsauswahl aufgrund von $p_{j,[k]}$ bestimmt. Durch die Zufallsauswahl soll zumindest teilweise verhindert werden, dass die Varianz aufgrund der Einsetzungen stark unterschätzt wird.

4 Statistik der sozialmedizinischen Institutionen

4.1 Daten, Variablen und Hilfsvariablen

Die Statistik der sozialmedizinischen Institutionen (SM) ist eine obligatorische Vollerhebung, die seit 1997 erhoben wird. Die Daten aller betroffenen Institutionen sind in einer ORACLE Datenbank gespeichert. Die Identifikatoren der Institutionen sind auch für Institutionen vorhanden, die nicht geantwortet haben. Für die Einsetzungen wurde METH ein Auszug in Form einer ACCESS Datenbank gegeben, der die Variablen enthält, die von GES vom thematischen Standpunkt aus als notwendig und hinreichend angesehen werden, um die Einsetzungen durchzuführen. Insbesondere enthält die Datenlieferung sämtliche Variablen, die einen direkten Bezug mit den Variablen haben, die eingesetzt werden sollen. Die erste Datenlieferung umfasst die Daten der Jahre 1998 bis 2002. Fehlende Werte werden in einem ersten Schritt ebenfalls für diese Zeitspanne eingesetzt. In einem zweiten Schritt werden die Methoden aus dem ersten Schritt auf die Jahre 1999-2004 angewandt, welche in einer zweiten Datenlieferung geliefert wurden.

Es wird zwischen zwei Definitionen der Einheit unterschieden, die im vorliegenden Methodenbericht verwendet werden:

1. Institution pro Jahr (Jahres Einheit)
2. Institutionen Einheit

Ist die Einheit **Jahres Einheit**, so zählt jede Institution für jedes Jahr einzeln. Wenn eine Institution 1998 bis 2002 in der SM ist, dann entspricht dies somit 5 Jahres Einheiten. Insgesamt umfasst der Datensatz von 1998 bis 2000 11'646 Jahres Einheiten. Ist die Einheit die **Institutionen Einheit**, so zählt jede Institution über die gesamte Erhebungsperiode nur einmal. Somit weist eine Institution, die 1998 bis 2002 in der SM war, 5-mal mehr Variablen auf, gegenüber einer Institution, die nur während eines Jahres in der SM war. Beide Institutionen Einheiten zählen aber als eine Einheit. Insgesamt befinden sich 2'565 Institutionen Einheiten in den Daten. Institutionen, die während eines Jahres eröffnet wurden, werden für das ganze Jahr als vorhanden angesehen. Diese Vereinfachung wurde gewählt, um die Komplexität der Einsetzungen in Grenzen zu halten. Diese Annahme führt aber sicherlich zu Verzerrungen.

Die Variablen, die eingesetzt werden sollen, sind in der Tabelle 19 im Anhang A.1 in Form der Fragebogenabschnitte zusammengefasst und in der Tabelle 20 wird die Notation gegeben, welche in den darauf folgenden Tabellen verwendet wird. Aus den Tabellen 21 bis 24 geht hervor, dass die die Konsistenz mehrerer Fragen untereinander über die Totale kontrolliert werden kann. Somit kann für diese Variablen auf aggregiertem Niveau festgestellt werden, ob mindestens eine Variable einer Unterkategorie fehlt oder nicht. Für die Einsetzungen müssen Totale einzelner Rubriken neu gebildet werden.

Variablen, die zusätzliche Information für die Einsetzungen liefern, aber nicht eingesetzt werden, werden **Hilfsvariablen** genannt. Zum Beispiel kann ein Mass für die Grösse einer SM-Institution als Hilfsvariable angesehen werden. Hilfsvariablen sind für die Einsetzung von grösster Wichtigkeit, da sie hilfreich sind um ein Modell für die Antwortausfälle zu schätzen und die Einsetzungsmodelle auf sie abgestützt werden können. Hilfsvariablen müssen für sämtliche Einheiten vorhanden sein, ob diese fehlend sind, fehlende Werte aufweisen oder vollständig vorhanden sind.

Da beinahe sämtliche Variablen des Fragebogens nur teilweise vorhanden sind, ist die Verwendung von Hilfsvariablen eingeschränkt. Die verwendeten Hilfsvariablen sind in Tabelle 1 aufgeführt. Anhand der Hilfsvariablen wurden Klassen gebildet, wie später im Abschnitt 3.2 aufgezeigt wird.
Zu den in der Tabelle 1 aufgeführten Hilfsvariablen, wurde zusätzlich eine Grössenklasse, 'Cluster' genannt, als Hilfsvariable verwendet. Diese Grössenklassen Variable wurde aufgrund der Anzahl Angestellter und der Anzahl Betten anhand der SAS Prozeduren `proc cluster` und `proc tree` gebildet. In 'Cluster'=1 sind die kleinsten Jahres Einheiten gruppiert und 'Cluster'=5 die grössten.
Fehlten die Hilfsvariablen in einem Jahr, so wurde die Klasse eines anliegenden Jahres mittels einer HIE (cf. 3.4) eingesetzt.

Tabelle 1 Hilfsvariablen, die für die Einsetzungen verwendet wurden.

Variable	vollständig vorhanden
Kanton (Verwaltungssitz[4])	ja
Jahr	ja
Anzahl Angestellte	nein
Anzahl Betten	nein
Typologie (von 1999) der SM-Institution, cf. Tabelle 18	nein
Korrespondenzsprache	nein
Rechtlich-wirtschaftlicher Status	nein

4.2 Antwortausfälle und Spenderpool

Seit 1999 können die sozialmedizinischen Institutionen Einheiten den Fragebogen mittels einer speziellen Software ausfüllen. Kontrollen, welche Totale vergleichen oder das vorhanden sein von notwendiger Information testen, sind in diesem Programm integriert. Der Benutzer ist jedoch nicht gezwungen die Tests durchzuführen, geschweige denn, die Inkonsistenzen zu beseitigen, bevor er die Daten dem BFS zukommen lässt. Werden die Tests durchgeführt, wird darauf hingewiesen, welche Antworten untereinander im Widerspruch stehen und welche fehlend sind.
Es bestehen 'fixe Kontrollregeln', die zur Überprüfung der Totale in verschiedenen Rubriken dienen, die gleich sein müssen. Weitere Regeln, die im elektronischen Fragebogen implementiert sind, weisen auf fehlende Variablenwerte bei sonst ausgefüllten Fragebogen hin. Eine dritte Art Regeln, dient zur Erkennung von ungültigen Werten. Im Weiteren verwendet GES bei der Datenaufbereitung 'variable Kontrollregeln', die sich auf die Veränderungen der Werte zum Vorjahr beziehen. Die Variabilität besteht darin, dass nicht genau definiert ist, welche Veränderungen 'inakzeptabel' sind. Die Institutionen, deren Werte bei diesen Tests als suspekt erachtet werden, werden rückgefragt oder dem Kanton gemeldet.
Für die Lokalisierung fehlender Werte und von inkonsistenten Daten wurden die Regeln, die in der Software der Erhebung `SM3.0` integriert sind, verwendet (siehe Tabellen 25 bis 28 im Anhang A.2).
Es werden zwei Arten von Antwortausfällen unterschieden, die Ausfälle von Jahres Einheiten (*unit nonresponse*) und die Ausfälle einzelner Variablen (*item nonresponse*). Vor der Formalisierung der Antwortausfälle, wird die Häufigkeit der Antwortausfälle in der SM 1998 - 2002 aufgezeigt, wobei die beiden Arten von Antwortausfällen unterscheiden werden.

[4] Eine Institution hat den Kanton geändert, da der Verwaltungssitz geändert hat.

4.2.1 Antwortausfälle von Einheiten und einzelnen Variablen

Antwortausfälle von Einheiten
Die Antwortausfälle von Jahres Einheiten ist GES bekannt, da aufgrund von Registerdaten bestimmt werden kann, welche Institutionen antworten müssen.
In den Daten der SM von 1998 bis 2002 sind 397 Antwortausfälle von Jahres Einheiten auf ein Total von 11'646 Jahres Einheiten. Diese wurden anhand der Indikatorvariable 'Qual' als fehlend erkannt. Hinzu kommt eine Institution, von der gewisse Angaben vorhanden sind, aber sämtliche Variablen, die eingesetzt werden sollen, nicht vorhanden sind.
Auffällig ist, dass im Jahr 1998 163 der 171 sozial-medizinischen Institutionen nicht geantwortet haben. In den folgenden Jahren verbesserte sich das Antwortverhalten merklich und erreicht im Jahr 2001 noch 2 sozialmedizinische Institutionen die nicht geantwortet haben. Auf Kantonsniveau weisen die Kantone Bern respektive Waadt für die Jahre 1998 und 1999 rund 30 ($\sim$8%) respektive 20 ($\sim$12%) totale Antwortausfälle auf. Für diese Kantone ist damit zu rechnen, dass die Einsetzungen das grösste Ausmass haben werden.

Antwortausfälle von einzelnen Variablen
Einige Variablen müssen bei allen Jahres Einheiten einen Wert haben, bei anderen Variablen wird eine Antwort nur verlangt, wenn in einer zweiten Variablen eine entsprechende Antwort gegeben wurde. So muss jede Jahres Einheit zwar Personal aufweisen, aber nicht unbedingt Personal mit einem Universitätsabschluss. Für die Variablen $\left\{u^t_{i,38}, \ldots, u^t_{i,43}\right\}$ aus der Tabelle 21 kann somit nur die Summe als fehlend erkannt werden, wenn in der Rubrik 2a Personal mit Universitätsabschluss angegeben wurde, und nicht die einzelnen Summanden. Es ist keine deterministische Regel vorhanden, um fehlende Angaben in 2a und 2b zu lokalisieren, wenn in beiden Fragen keine Antworten vorhanden sind. Diese Art Antwortausfälle wird '**bedingte Antwortausfälle**' einzelner Variablen genannt. Diese sind dadurch charakterisiert, dass sie durch eine Inkonsistenz zwischen mehreren Variablen, mindestens zwei, definiert werden können. Die '**zwingenden Antwortausfälle**' einzelner Variablen sind hingegen dadurch charakterisiert, dass sie unabhängig von den Werten anderer Variablen vorhanden sein müssen. So handelt es sich zum Beispiel zwingend um einen Antwortausfall, wenn keine Angabe für die Variable 'Personalaufwand' vorhanden ist. Für die Lokalisierung bedingter Antwortausfälle einzelner Variablen wird davon ausgegangen, dass sämtliche vorhandene Information beibehalten werden muss.
Die Anzahl Jahres Einheiten mit zwingenden Antwortausfällen ist für die Frage 5.1 mit rund 1.4% der Jahres Einheiten am höchsten (siehe Tabelle 2). Hingegen ist die Anzahl Jahres Einheiten mit bedingten Antwortausfällen in den Fragen 3.1.1.1 und 3.1.1.3 sowie in den Fragen 4.1 und 4.3 übermässig gross, wie aus den Tabellen 3 bis 5 ersichtlich ist. Es stellt sich die Frage, ob die Institutionen die Verbindung der beiden Fragen wirklich gemacht haben, oder ob sie diese Fragen nicht verstanden haben oder nicht beantworten konnten. Der Interpretationsspielraum dieser Fragen wurde in die Kontrollregeln miteinbezogen, indem Konstellationen zugelassen wurden, die in den ursprünglichen Kontrollregeln nicht erlaubt waren. Diese Konstellationen entsprachen den Möglichkeiten, wie die Fragen auch verstanden werden konnten. In den oben genannten Tabellen wurden diese Konstellationen nicht mitgezählt, d.h. sie wurden bereits als zulässig erachtet. Aus der Tabelle 6 ist ersichtlich, dass 9'041 der 11'646 Jahres Einheiten mindestens einen fehlenden Wert aufweisen. Wobei bei 4 Jahres Einheiten die Frage 2a durch Übernahme der Totale aus der Frage 2b vervollständigt werden konnte. Weitere 83 Jahres Einheiten wurden durch die grosszügigere Handhabung der Fragen 3.1, 4.1 und 4.3 als konsistent erachtet. Somit hatten immer noch rund 77% aller Jahres Einheiten von 1998 bis 2002 mindestens eine fehlende Angabe. Eine Beschränkung der Einsetzungen, wie im nächsten Abschnitt erläutert, drängte sich auf.

Tabelle 2 Total der 'zwingenden Antwortausfälle' einzelner Variablen von 1998 bis 2002. Dabei wurden die Antwortausfälle von Jahres Einheiten ausgenommen. '#' bedeutet Anzahl, der prozentuale Anteil der Ausfälle bezieht sich auf die 11'646 Jahres Einheiten, welche in der SM 1998 bis 2002 sind.

Frage	Variable	# Ausfälle	[%] Ausfälle
2.a	Gesamttotal Personen	26	0.22
2.a	Gesamttotal Stellen	48	0.41
5.1	Personalaufwand	10	0.09
5.1	Anderer Betriebsaufwand	10	0.09
5.1	Gesamte Rubrik	158	1.36

Tabelle 3 Total der 'bedingten Antwortausfälle' einzelner Variablen über die Periode 1998 bis 2002. Dabei wurden die Antwortausfälle von ganzen Jahres Einheiten (*unit nonresponse*) ausgenommen. Es wurden die weniger restriktiven Regeln verwendet. '#' bedeutet Anzahl. Der prozentuale Anteil der Ausfälle bezieht sich auf die 11'646 Jahres Einheiten, welche in der SM 1998 bis 2002 sind.

Frage	Variable	Fehlend aufgrund der Angaben in der Variablen	Anzahl Ausfälle	[%] Ausfälle
2.a	# Stellen i.e.B. mit universitärer Ausbildung	Personal mit universitärer Ausbildung 2.b	19	0.16
2.a	# Stellen i.e.B. mit höherer Ausbildung	Personal mit höherer Ausbildung 2.b	34	0.29
2.a	# Stellen i.e.B. mit Berufsausbildung	Personal mit Berufsausbildung 2.b	6	0.05
2.b	Gesamttotal mit universitärer Ausbildung	Stellen mit universitärer Ausbildung in 2.a	64	0.55
2.b	Gesamttotal mit höherer Ausbildung	Stellen mit höherer Ausbildung in 2.a	186	1.60
2.b	Gesamttotal mit Berufsausbildung	Stellen mit Berufsausbildung in 2.a	156	1.34
3.1	Gesamttotal Bewohner	Gesamttotal Bewohner 3.3 oder 3.4	17	0.15
3.3	Gesamttotal Bewohner	Gesamttotal Bewohner 3.1 oder 3.4	20	0.17
3.3	Gesamttotal Externe	Gesamttotal Externe 3.1 oder 3.4	11	0.09

Tabelle 4 Total der 'bedingten Antwortausfälle' einzelner Variablen über die Periode 1998 bis 2002. Dabei wurden die Antwortausfälle von ganzen Jahres Einheiten (*unit nonresponse*) ausgenommen. Es wurden die weniger restriktiven Regeln verwendet. '#' bedeutet Anzahl. Der prozentuale Anteil der Ausfälle bezieht sich auf die 11'646 Jahres Einheiten, welche in der SM 1998 bis 2002 sind.

Frage	Variable	Fehlend aufgrund der Angaben in der Variablen	Anzahl Ausfälle	[%] Ausfälle
3.1.1.1	# Betagte/Chronischkranke	Betagte/Chronischkranke in 4.1	306	2.63
3.1.1.1	# Betten für Betagte/Chronischkranke	Betagte/Chronischkranke in 4.1 oder 3.1.1.1 oder Aufenthaltstage in 3.1.1.1	333	2.86
3.1.1.1	# Aufenthaltstage für Betagte/Chronischkranke	Betagte/Chronischkranke in 4.1 oder 3.1.1.1 oder Betten in 3.1.1.1	312	2.68
3.1.1.2	# Betten für Psychiatriepatienten	Aufenthaltstage für Psychiatriepatienten in 3.1.1.2	7	0.06
3.1.1.2	# Aufenthaltstage für Psychiatriepatienten	Betten für Psychiatriepatienten in 3.1.1.2	1	0.01
3.1.1.3	# Behinderte	Behinderte in 4.3	3981	34.18
3.1.1.3	# Betten für Behinderte	Behinderte in 4.3 oder 3.1.1.3 oder Aufenthaltstage in 3.1.1.3	3986	34.23
3.1.1.3	# Aufenthaltstage für Behinderte	Behinderte in 4.3 oder 3.1.1.3 oder Betten in 3.1.1.3	3983	34.20
3.1.1.4	# Betten für Drogenabhängige	Aufenthaltstage für Drogenabhängige in 3.1.1.4	1	0.01
3.1.1.4	# Aufenthaltstage für Drogenabhängige	Betten für Drogenabhängige in 3.1.1.4	2	0.02
3.1.1.4	# Betten für Alkoholkranke	Aufenthaltstage für Alkoholkranke in 3.1.1.4	3	0.03
3.1.1.4	# Aufenthaltstage für Alkoholkranke	Betten für Alkoholkranke in 3.1.1.4	3	0.03
3.1.1.4	# Betten für psychosoz. Fälle	Aufenthaltstage für psychosoz. Fälle in 3.1.1.4	2	0.02
3.1.1.4	# Aufenthaltstage für psychosoz. Fälle	Betten für psychosoz. Fälle in 3.1.1.4	4	0.03
3.1.1.4	# Betten für Andere	Aufenthaltstage für Andere in 3.1.1.4	1	0.01
3.1.1.4	# Aufenthaltstage für Andere	Betten für Andere in 3.1.1.4	4	0.03

Tabelle 5 Total der 'bedingten Antwortausfälle' einzelner Variablen über die Periode 1998 bis 2002. Dabei wurden die Antwortausfälle von ganzen Jahres Einheiten (*unit nonresponse*) ausgenommen. Es wurden die weniger restriktiven Regeln verwendet. '#' bedeutet Anzahl. Der prozentuale Anteil der Ausfälle bezieht sich auf die 11'646 Jahres Einheiten, welche in der SM 1998 bis 2002 sind.

Frage	Variable	Fehlend aufgrund der Angaben in der Variablen	Ausfälle #	[%]
3.4	Gesamttotal männl. Bewohner	Gesamttotal Bewohner 3.1 oder 3.3 vorhanden	26	0.22
3.4	Gesamttotal Bewohnerinnen	Gesamttotal Bewohner 3.1 oder 3.3 vorhanden	26	0.22
4.1	Total Bewohner und Externe	Gesamttotal Bewohner 3.1 vorhanden oder 4.3 fehlend	1037 [5]	8.90
4.3	Total Bewohner und Externe	Gesamttotal Bewohner 3.3 vorhanden oder 4.1 fehlend	1037 [6]	8.67
5.1	Beiträge IV	SM Institution mit IV Beiträgen 0.1 vorhanden	102	0.88

Tabelle 6 Anzahl Jahres Einheiten der Total 11'646 Jahres Einheiten pro Anzahl fehlende Werte.

Anzahl fehlende Werte	Anzahl Jahres Einheiten
0	87[7]
1	113
2	873
3	5169
4	2012
5	88
6	579
7	53
8	21
9	23
10	4
11	3
12	4
13	7
14	1
15	1
16	1
17	1
19	1
Total	**9041**

[5] 944 mit fehlender Antwort in 4.3, 93 mit Antwort in 3.1.

[6] 944 mit fehlender Antwort in 4.1, 66 mit Antwort in 3.3.

[7] 4 dieser Institutionen Einheiten erlangten durch deterministische Einsetzungen in der Frage 2a eine vollständige Konsistenz. 83 Institutionen Einheiten werden aufgrund der weniger restriktiv gehandhabten Regeln bezüglich der Fragen 3.1, 4.1 und 4.3 als konsistent erachtet.

4.2.2 Definition des Spenderpools

Die potentiellen Spender eines Empfängers werden einerseits durch Kontrollregeln und die Ausreisserlokalisierung und andererseits durch das Modell für die Antwortausfälle bestimmt. Aufgrund der ersten Kriterien soll verhindert werden, dass Inkonsistenzen und extreme Werte eingesetzt werden. Das zweite Kriterium soll die Homogenität der Einsetzungen begünstigen.

Aus der Tabelle 7 geht hervor, dass bei der Anwendung der *Huber-type skipped mean* Verwerfungsregel eine realistische Grösse für κ bei $\kappa = 10$ liegt. Im Durchschnitt muss somit pro Jahr mit rund 25 sozialmedizinischen Einheiten gerechnet werden, die eingehender bearbeitet werden sollten.
Ausreisser in den Angaben des Personalbestandes rühren vor allem daher, dass die Angaben in Prozent gemacht wurden. Gleichermassen wurden in der Finanzbuchhaltung die Angaben oft mit einem Faktor 10 angegeben. In einem Fall wurden sämtliche sozialmedizinischen Institutionen einer grösseren Schweizer Stadt als eine einzige angegeben. Solcherlei Ausreisser sollten bei der Datenaufbereitung entdeckt und bearbeitet werden, da sie die Analysen massgeblich beeinflussen können.

Aus der Tabelle 8 geht hervor, dass einerseits vor allem beim Personalaufwand Ausreisser bestehen. Die Entwicklung vom Jahr 2000 zum Jahr 2001 weist eine zu den anderen Jahren vergleichsweise uniforme Verteilung der Anzahl Ausreisser bezüglich der analysierten Totale auf. Meist sind es mehr Ausreisser.
Ausreisser der Entwicklung in den Angaben des Personalbestandes rühren vor allem daher, dass zum Beispiel in einem Jahr die Mehrheit der Beschäftigten eine Berufsausbildung hatte, und im folgenden Jahr hatte eine ähnliche Anzahl Beschäftigter eine höhere Ausbildung.

Tabelle 7 Anzahl Institutionen Jahre mit mindestens einem Ausreisserwert über alle Variablen, die aufgrund einer *Huber-type skipped mean* Verwerfungsregel mit Parameter κ entdeckt wurden. '[%]' bezieht sich auf die Jahres Einheiten ohne deren Antwort Ausfälle.

κ	**Anzahl**	**[%]**
3	1848	16.40
4	1057	9.38
5	632	5.61
6	414	3.67
7	298	2.64
10	133	1.18
20	33	0.29
30	16	0.14
40	13	0.12
50	6	0.05

Die Zusammensetzung des Spenderpools wird durch die Definition der Einheit grundlegend beeinflusst. Ist die Einheit die Jahres Einheit, so gehören sämtliche Jahres Einheiten zum Spenderpool, welche keine Antwortausfälle aufweisen, keine Kontrollen verletzen und die keine Variable aufweisen, welche extreme Werte oder extreme Veränderungen aufweist. Ist die Einheit als Institution definiert, so müssen sämtliche Jahres Einheiten derselben Institution dem obigen Spenderpool angehören, damit sie als Spender Institution definiert wird.
Die Zusammensetzung des Spenderpools hängt stark davon ab, welche Kontrollen angewandt werden, respektive welche Variablen in die Kontrollen miteinbezogen werden, wie später aufgezeigt wird.
Als Hilfsvariablen wurden die Variablen, die im Abschnitt 4.1 aufgeführt sind, in das Modell (11)

Tabelle 8 Anzahl Institutionen Jahre mit mindestens einer 'ungewöhnlichen' Veränderung eines Totals. Parameter der *Hidiroglou-Berthelot* Ausreisser Erkennung: $u = 0.8$, $A = 0.05$, $L = 50$.

	Jahr			
Variable	1998-1999	1999-2000	2000-2001	2001-2002
Total Personen 2a	8	4	11	14
Vzae univ 2a	1	.	.	.
Vzae beruf 2a	7	8	14	7
Vzae hoehere 2a	5	8	11	7
Vzae univ 2b	1	.	.	.
Vzae hoehere 2b	6	8	11	6
Vzae beruf 2b	9	5	15	6
Besch. hoeh 2b	2	1	10	2
Besch. beruf 2b	5	3	10	4
Tot. Bew. (kt) 3.3	6	19	10	6
Tot. Ext. (kt) 3.3	1	5	2	2
Tot. Bet. (3.4)	2	3	2	1
Tot. Bew (3.4)	7	21	11	6
BETAFW 5.1	17	9	13	11
PERSONAL 5.1	23	14	14	9

aufgenommen. Da diese fehlende Werte aufweisen, wurde zuerst eine 'konservativ historisch iterative' Einsetzung (siehe Abschnitt 3.4) durchgeführt, um die Hilfsvariablen zu vervollständigen. Bei 73 Jahres Einheiten konnten die Hilfsvariablen mit dieser Methode nicht vervollständigt werden. Die entsprechenden Daten wurden von den folgenden Modellen ausgeschlossen. Durch den Einbezug der Variablen 'Cluster' in das Modell musste auch auf die Daten der Jahres Einheiten verzichtet werden, die überhaupt keine Daten aufweisen, da die Grössenklasse nicht gebildet werden konnte. Das Basismodell beschrieb das Ereignis 'Spender'='nein' mit den Variablen der Tabelle 1 als erklärende Variablen. Das heisst, 'Kanton', 'Jahr', 'TYPOL99', 'juroec', 'korrspra' und 'CLUSTER'.

Die Parameter des logistischen Regressions-Modells wurden anhand der SAS Prozedur `proc logistic` geschätzt und die signifikanten Variablen mit der Option `backward` eruiert, cf. Tabelle 9. Diese wurden danach verwendet um Einsetzungsklassen zu bilden, womit die potentiellen Spender für jeden Empfänger sich auf Spender innerhalb derselben Einsetzungsklasse beschränken.

Jahres Einheit

Werden die Fragen 0.5, 2.a, 2.b, 3.1, 3.3, 3.4, 4.1, 4.3 und 5.1 in die Kontrollen miteinbezogen, so befinden sich rund 20% der Jahres Einheiten im Spenderpool. Da ein grosser Anteil an Jahres Einheiten fehlende Werte und/oder Inkonsistenzen aufweisen, wurde in einem ersten Schritt das Jahr 1998 aus den Daten entfernt, um den Anteil der Spender zu erhöhen. Es verblieben somit 9'252 Jahres Einheiten im Datensatz. 1'962 (21%) davon gehören dem Spenderpool an, 7'290 (79%) nicht. Der Ausschluss der Daten von 1998 hatte nicht die gewünschte Wirkung auf den Anteil an Jahres Einheiten im Spenderpool. Aufgrund der logistischen Regression geht hervor, dass die Kantone 'BS', 'OW', 'SO', 'TI' und 'ZH' eine überdurchschnittlich hohe Quote an Institutionen aufweisen, welche unvollständige oder widersprüchliche Fragebogen eingereicht haben. Im Kanton 'OW' handelt es sich dabei ausschliesslich um Institutionen Einheiten, welche Probleme mit den Fragen 3.1, 4.1 und 4.3 haben. Sozial-medizinische Institutionen ohne IV Beiträge sind bei den Spendern untervertreten, ebenso wie kleine Institutionen

Tabelle 9 SAS-Output der Prozedur `proc logistic`, die angewendet wurde, um ein Modell für die Wahrscheinlichkeit der Zugehörigkeit zum Spenderpool zu schätzen. Als Einheit wurde die Jahres Einheit verwendet. Das Jahr 1998 wurde aus den Daten ausgeschlossen. Die Variable 'juroec' bezeichnet den wirtschaftlich-rechtlichen Status der Institution, 'Typol99' bezeichnet die Typologie von 1999 und 'korrspra' steht für die Korrespondenzsprache.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
KANTON	25	314.9676	<.0001
JAHR	3	72.1507	<.0001
juroec	1	5.4007	0.0201
TYPOL99	6	199.1225	<.0001
korrspra	2	13.4200	0.0012

und Institutionen deren Korrespondenzsprache italienisch ist. Letztere sind vielleicht auf das Fehlen eines italienischsprachigen Fragebogens zurückzuführen.

Da die Entfernung des Erhebungsjahrs 1998 nicht die gewünschten Auswirkungen hatte, wurden die Fragen 3.1, 4.1 und 4.3 aus den Kontrollen ausgeschlossen und das Jahr 1998 wieder aufgenommen. In der Frage 3.1.1.1 wurde dennoch aufgrund der Regel 36 (Tabelle 27) eingesetzt.
Sämtliche Jahres Einheiten der Kantone 'AI', 'BL', 'JU', 'NW', 'OW' und 'UR' hatten unter oben genannten Bedingungen keine fehlenden Daten. Diese wurden deshalb aus dem Modell ausgeschlossen, um die Schätzung des logistischen Regressionsmodells zu ermöglichen.
Es verblieben 11'067 Jahres Einheiten im Datensatz. Hiervon wiesen 3'067 Jahres Einheiten mindestens einen fehlenden Wert auf und 8'000 keine fehlenden Daten. Davon wurden 20 wegen Inkonsistenzen und 127 wegen Ausreissern aus dem Spenderpool ausgeschlossen, siehe Tabelle 10.
Wird das Jahr 1998 in die Daten miteinbezogen und die Fragen 3.1, 4.1 und 4.3 aus den Kontrollen ausgeschlossen, so hat die Variable 'Jahr' keinen signifikanten Einfluss auf die Zugehörigkeit zum Spenderpool. Diese hatte jedoch einen signifikanten Einfluss, wenn das Jahr 1998 ausgeschlossen wird und die oben genannten Fragen in die Kontrollen miteinbezogen werden. Die Variable 'CLUSTER' hat hingegen unter den erst genannten Voraussetzungen einen signifikanten Einfluss, cf. Tabelle 11. Vor allem kleinere Jahres Einheiten gehören weniger oft dem Spenderpool an.
Die signifikanten Variablen können für die Bildung von Einsetzungsklassen verwendet werden, um Verzerrungen aufgrund dieser Charakteristiken zu berücksichtigen.

Tabelle 10 Aufteilung der Daten in die Kantone, welche keine fehlenden Werte aufweisen ('ausgeschlossene Kt.); Einheiten, die nur Inkonsistenzen oder Ausreisser hatten ('nur Inkonsistenzen'); Einheiten, die nur Ausreisser hatten ('nur Ausreisser'); den Spenderpool; Einheiten, bei denen mindestens eine Jahres Einheit ganz fehlte ('*unit nonresponse*') und Einheiten, die in jedem Jahr, in welchem sie in der SM waren mindestens eine Angabe hatten, aber in mindestens einem Jahr eine fehlende Angabe hatten ('*item nonresponse*').

	Anzahl Einheiten: 1998-2002	
Datencharakteristik	**Jahres Einheit**	**Institutionen Einheit**
ausgeschlossene Kt.[8]	579	126
nur Inkonsistenzen	20	4
nur Ausreisser	127	29
Spenderpool	7'853	1'706
unit nonresponse	1'482	359
item nonresponse	1'585	341
Total	11'646	2'565

Tabelle 11 SAS-Output der Prozedur `proc logistic`, die angewendet wurde, um ein Modell für die Wahrscheinlichkeit der Zugehörigkeit zum Spenderpool zu schätzen. Als Einheit wurde Jahres Einheit verwendet. Die Fragen 3.1, 4.1 und 4.3 wurden nicht in die Kontrollen miteinbezogen. Die Variable 'Cluster' steht für die Grössenklasse, 'juroec' bezeichnet den wirtschaftlich-rechtlichen Status der Institution, 'korrspra' steht für die Korrespondenzsprache und 'Typol99' ist die Typologie von 1999.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
KANTON	19	1621.0812	<.0001
juroec	1	76.7283	<.0001
TYPOL99	6	77.4839	<.0001
CLUSTER	4	32.8892	<.0001

Einheit: Institution

Werden die Fragen 0.5, 2.a, 2.b, 3.1, 3.3, 3.4, 4.1, 4.3 und 5.1 mit einbezogen und das Erhebungsjahr 1998 weglassen wird, so befinden sich 8.3% der Institutionen Einheiten im Spenderpool. Die Mehrheit der fehlerhaften Institutionen Einheiten verletzt Regeln, in welchen die Totale der Fragen 3.1, 4.1 und 4.3 einander gegenübergestellt werden oder die Totale von 2.a und 2.b stimmen nicht überein. Von den 2'465 Institutionen Einheiten der Jahre 1999 bis 2002 weisen 874 für mindestens eine Erhebungsperiode Inkonsistenzen auf, aber keine fehlenden Werte.

Wie weiter oben erwähnt, wurde das Jahr 1998 wieder in die Einsetzungen miteinbezogen. Hingegen wurden die Fragen 3.1, 4.1 und 4.3 ausgeschlossen. Mit der Ausnahme, dass die Regel 36 der Tabelle 27 miteinbezogen wurde.

Durch die Reduktion der zu bereinigenden Variablen, wiesen sämtliche Institutionen Einheiten der Kantone 'AI', 'BL', 'JU', 'NW', 'OW' und 'UR' Daten ohne fehlende Werte für alle Erhebungsjahre auf. Diese Institutionen Einheiten wurden aus dem Spenderpool entfernt. Institutionen Einheiten ohne fehlende Werte aber mit Inkonsistenzen oder Ausreissern wurden aus dem Spenderpool ebenfalls ausgeschlossen (siehe Tabelle 10), damit wird so weit möglich verhindert, dass inkonsistente Daten eingesetzt werden. 1'706 Institutionen Einheiten (67%) bildeten den Spenderpool, bei 700 Institutionen Einheiten (27%) mussten Einsetzungen vorgenommen

[8] 'AI', 'BL', 'JU', 'NW', 'OW' und 'UR'.

werden. Die Hilfsvariablen der Tabelle 1 für die Institutionen Einheit wurden von der letzten Erhebungsperiode übernommen. Nach einer HIE (Abschnitt 3.4) der Hilfsvariablen wiesen noch 62 Institutionen Einheiten fehlende Werte in den Hilfsvariablen auf. Diese mussten aus der Modellierung ausgeschlossen werden. Die Parameter des logistischen Modells wurden anhand der SAS Prozedur `proc logistic` geschätzt, wobei die Option `backward` verwendet wurde, um die signifikanten Variablen im Modell zu eruieren. Die Variablen 'Kanton', 'Typol99' und wirtschaftlich-rechtlicher Status hatten einen signifikanten Einfluss auf die Zugehörigkeit einer Institution zum Spenderpool, c.f. Tabelle 12. Diese Variablen wurden anschliessend verwendet, um Einsetzungsklassen zu bilden (Abschnitt 4.3.2).

Tabelle 12 SAS-Output der Prozedur `proc logistic`, die angewendet wurde, um ein Modell für die Wahrscheinlichkeit der Zugehörigkeit zum Spenderpool zu schätzen. Als Einheit wurde die Institution verwendet. Die Fragen 3.1, 4.1 und 4.3 wurden nicht in die Kontrollen miteinbezogen. Die Hilfsvariablen wurden vom letzten Erhebungsjahr übernommen, in dem die Institution in der SM war. Die Variable 'juroec' bezeichnet den wirtschaftlich-rechtlichen Status der Institution, 'korrspra' steht für die Korrespondenzsprache, 'Cluster' steht für die Grössenklasse und 'Typol99' ist die Typologie von 1999.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
KANTON	19	493.0569	<.0001
juroec	1	20.7125	<.0001
TYPOL99	6	21.5756	0.0014
CLUSTER	4	13.8221	0.0079

4.3 Einsetzungen

4.3.1 Resultate der Konservativen HIE und der deduktiven Einsetzung

Unter Berücksichtigung aller Fragen und aller Jahre wurden zuerst die fehlenden Jahres Einheiten durch eine konservative HIE eingesetzt, bevor die Merkmalsausfälle durch deduktive Einsetzungen und eine konservative HIE eingesetzt wurden.
Mit der konservativen HIE konnte von den 397 Jahres Einheiten die vollständig fehlten 338 eingesetzt werden. Die 59 übrigen hatten zwischen 1998 und 2002 keinerlei Antworten.
Insgesamt konnte bei 69 Jahres Einheiten die Frage 2.a von der Frage 2.b anhand einer deduktiven Einsetzung übernommen werden, ohne einen Widerspruch hervorzurufen. In nur gerade 5 Fällen führten diese Ableitungen zu vollständigen Angaben in diesen Fragen. Die Anzahl Jahres Einheiten mit mindestens einer fehlenden Angaben konnte durch die deduktiven Einsetzungen und die konservative HIE um rund 1'700 auf 3'000 vermindert werden. Aus der Tabelle 13 ist ersichtlich, dass die Einsetzungen vor allem in den Fragen 3.1, 4.1 und 4.3 nicht die gewünschten Auswirkungen hatten. Hinzu kommt, dass vor den Einsetzungen rund 1'500 Jahres Einheiten mindestens zwei widersprüchliche Angaben aufwiesen. Nach den Einsetzungen stieg dieser Wert auf rund 2'500. Der Grund für diesen Anstieg liegt darin, dass einerseits bestehende Widersprüche selten gelöst werden konnten und andererseits wurden durch die Einsetzungen widersprüchliche Angaben übernommen und neu kreiert.
Dieses Vorgehen ist nicht befriedigend, da eine HIE zusätzlich zu den oben genannten Problemen, die Längsschnitt-Analyse beeinflusst, indem Veränderungen abgeschwächt werden.

Tabelle 13 Einfluss der konservativen HIE auf die Rubriken mit den meisten fehlenden Angaben. In den Rubriken 3.1.1.1 und 3.1.1.3 haben die Tage und Anzahl Betten ähnliche Werte, wie die Anzahl Bewohner, deshalb wurde nur letztere aufgeführt. 'Andere' bezeichnet die übrigen Rubriken.

Rubrik	Vor der Einsetzung fehlend	Nach der Einsetzung fehlend	Anzahl eingesetzte Werte	[%] eingesetzte an den Fehlenden vor der Eins.
3.1.1.1 Bewohner	681	405	276	40.5
3.1.1.3 Bewohner	2'870	1'973	897	31.3
4.1	2'067	1'003	1'064	51.5
4.3	740	330	410	55.4
Andere	<180	<40		

4.3.2 Datenaufbereitung für die NNE

Die Datenaufbereitung für die NNE besteht aus mehreren Schritten und dient der Definition des Spenderpools und der Empfänger. Zuerst werden fehlende Werte und Inkonsistenzen für jede Einheit lokalisiert. Daraus resultiert ein erster Spenderpool. Anhand der *Huber-type skipped mean* Verwerfungsregel werden die Einheiten bestimmt, welche aus dem ersten Spenderpool ausgeschlossen werden.

Für die Ausreisserlokalisierung wurden vor und nach der NNE für jedes Jahr Gruppen aufgrund der Typologie von 1999, des juristisch-wirtschaftlichen Statuses, der Anzahl Betten und der Grössenklassen der Anzahl Beschäftigte gebildet. Die Gruppen wurden wie im Abschnitt 3.2 beschrieben gebildet.
Innerhalb der Gruppen wurde ein Normalitätstest *Kolmogorov-Smirnov* für $\tilde{x}_{i,k,t}$ durchgeführt, anhand dessen die Nullhypothese, dass sie einer Normalverteilung folgen, bei einem Fehler

erster Art von 5%, nicht verworfen werden konnte. Aufgrund der zum Teil kleinen Gruppen muss die Aussagekraft des Normalitätstests in Frage gestellt werden. Aus Zeitgründen wurde dieser Frage nicht weiter nachgegangen, sondern die *Huber-type skipped means* Verwerfungsregel mit $\kappa = 10$ angewandt. Hiermit wurden 133 Jahres Einheiten als Einheiten mit mindestens einem Ausreisser markiert.
Vor der endgültigen Kreierung des Spenderpools, wurde eine konservative HIE für die Variablen 'juroec' und 'Recordart' durchgeführt, da diese Variablen als Hilfsvariablen verwendet wurden. Von 97 fehlenden Werten von Jahres Einheiten in der Variablen 'juroec' konnten 38 eingesetzt werden. In der Variablen 'Recordart' blieben von den 396 fehlenden Werten in Jahres Einheiten nach der konservativen HIE noch 70 fehlend. Bei den übrig gebliebenen Jahres Einheiten mit fehlenden Werten in diesen Variablen handelte es sich um Institutionen, die nur in einem Jahr in der SM waren und überhaupt keine Angaben aufwiesen. Für diese wurde eine Zufallseinsetzung aufgrund des Kantons und der Typologie 1999 gemacht, da die Einsetzungen innerhalb der Einsetzungsklassen vorgenommen wurden, welche die Variable 'juroec' enthalten und die 'Recordart=IV' einen direkten Einfluss auf die Einsetzungen hat.

Schichtung und Einsetzungsklassen
Es wurden Hilfsvariablen verwendet, die beim Spender und beim Empfänger übereinstimmen müssen. Durch Kreuzung dieser Hilfsvariablen wurden Schichten gebildet. Dabei wurden vor allem die Variablen 'Kanton' und 'Jahr' als Schichtvariablen verwendet, wobei die Variable 'Jahr' als Vektor betrachtet werden muss. Jedes Jahr, während dem die Institution in der SM war, bildet dabei eine Komponente, die durch das Jahr indexiert ist. Zusätzlich wurde die Einsetzungsklasse als Schichtvariable verwendet, die wie im Abschnitt 3.2 beschrieben gebildet wurden. Als Grundlage für die Bildung der Einsetzungsklassen diente das Modell der Tabelle 12. Die Einsetzungsklassen wurden zweimal bestimmt, da innerhalb der ersten Einsetzungsklassen nicht die nötigen Spender vorhanden waren, um bei sämtlichen Einheiten alle fehlenden Werte vollständig einzusetzen. Die erste Einsetzungsklassen C_1 wurden anhand der Hilfsvariablen 'Kanton', 'Juristisch-wirtschaftlicher Status', 'Cluster' und der 'Typologie 1999' gebildet. Aus diesem Modell wurden anhand der Vorhersagen für die Antwortwahrscheinlichkeit 10 Einsetzungsklassen gebildet. Für die Bildung der zweiten Einsetzungsklassen wurde die Variable 'Cluster' aus dem Modell von C_1 entfernt, da sie am wenigsten signifikant war. Zusätzlich wurden die Jahres Einheiten mit 'Typol99' $\in$ {S212, S222} aus der Bildung von C_2 ausgeschlossen, da diese keine fehlenden Daten mehr aufwiesen. Um die Anzahl Spender innerhalb der C_2 zu erhöhen, wurden noch drei Einsetzungsklassen gebildet.

4.3.3 Verteilung der Totale in die Unterrubriken

Nach den Einsetzungen von Totalen wurden zusätzlich Sub-totale und Variablen von Unterrubriken anhand des Spenders eingesetzt. Für die Fragen 2a und 2b wurden zudem Einsetzungen durch Ableitungen, wie in 3.3 beschrieben, durchgeführt. Fehlte die gesamte Frage 5.1, so wurden auch die Variablen übernommen, welche in einem ersten Schritt nicht eingesetzt wurden. Der Ertrag und der Aufwand wurden neu berechnet. Konkret wurde folgende sequentielle Prozedur durchgeführt:

1. Die Anzahl Stellen nicht Ausgebildeter in 2.a wird neu berechnet.

2. Die gesamte Anzahl Stellen in 2.a wird neu berechnet.

3. Wenn in 2.a die Stellen nach Ausbildung eingesetzt wurden, so werden die entsprechenden Unterrubriken in 2.b ebenfalls eingesetzt, sofern das entsprechende Total in 2.b nicht vorhanden war oder eingesetzt wurde.

4. Wenn ein Total in 2.b eingesetzt wurde, so werden die entsprechenden Unterrubriken ebenfalls vom Spender übernommen.

5. Wenn in 2.a die Anzahl Stellen nicht Ausgebildeter 0 ist und die gesamte Anzahl Stellen Ausgebildeter in 2.a mehr als 0.5 grösser ist als die gesamte Anzahl Stellen aus 2.b, dann wird die Differenz dieser Totale in die Anzahl Stellen nicht Ausgebildeter (2.a) eingesetzt.

6. Die Unterrubriken der Frage 3.3 (Verteilung der Patienten nach Wohnort) werden vom Spender übernommen, wenn das Total in 3.3 eingesetzt wurde.

7. Die Unterrubriken der Frage 3.4 (Verteilung der Patienten nach Alter und Geschlecht) werden vom Spender übernommen, wenn das Total in 3.4 eingesetzt wurde.

8. Wenn in der Frage 5.1 (Betriebsrechnung) der Personalaufwand und der 'andere Betriebsaufwand' eingesetzt wurde, so werden die übrigen Unterrubriken ebenfalls eingesetzt, sofern sie keinen gültigen Wert enthalten.

9. Wurde die Recordart 'IV' mit einer HIE eingesetzt, so wird der Beitrag seitens der IV auch eingesetzt, falls nicht schon vorhanden.

10. In der Frage 5.1 werden Betriebsgewinn, Betriebsverlust, Betriebsaufwand und Betriebsertrag neu berechnet.

4.3.4 Vorgehen und Resultate NNE

Vorgehen
Als Einheit wurde die Institutionen Einheit gewählt und in einem ersten Schritt wurden Totale mittels NNE eingesetzt. Erst in einem zweiten Schritt wurden die Totale auf die einzelnen Unterrubriken aufgeteilt (cf. 4.3.3). So wurde zum Beispiel das Total der Stellen mit universitärer Ausbildung der Frage 2b zuerst eingesetzt, bevor in einem zweiten Schritt das Total auf die 'Ausbildung' in der Frage 2b aufgeteilt wurde. Dieses Vorgehen drängte sich auf, da es nicht möglich war, fehlende Werte in den Unterrubriken zu lokalisieren, denn die Kontrollregeln vergleichen nur die Totale auf aggregiertem Niveau. Die Aufteilung auf die Unterrubriken wurde durch die Verteilung beim zuvor ausgewählten Spender definiert.
Diese Einsetzungen bergen jedoch das Risiko, dass Abhängigkeiten der Variablen in den Unterrubriken und den restlichen Variablen gebrochen werden, da diese nicht explizit in den Einsetzungsprozess eingebunden werden konnten.
Die NNE (3.6) wurde iterativ durchgeführt, wobei sich einzig die Einsetzungsklassen von Schritt zu Schritt änderten. Sofern vollständige Daten verlangt werden ist ein iteratives Vorgehen nötig, da einige Jahres Einheiten fehlende Werte in den Hilfsvariablen aufwiesen oder es in der Einsetzungsklasse keine Spender gab, aber so viel wie möglich Information in den Einsetzungen verwendet werden sollte.
Insgesamt waren 6 Einsetzungsschlaufen nötig, um bei sämtlichen Institutionen Einheiten alle Variablen einzusetzen, siehe Tabelle 14. Dabei wurde in Institutionen Einheiten ohne fehlende Jahres Einheit ('i_nr') und solche mit fehlenden Jahres Einheiten ('u_nr') unterschieden. Nach der vierten NNE wurde eine konservative HIE für die Frage 3.1.1.1 durchgeführt, da fehlenden Werte in dieser Frage Einsetzungen der restlichen fehlenden Variablen verhinderten (Regel 36 der Tabelle 27).

Resultate NNE

Durch die NNE konnten sämtliche fehlenden Werte in den Fragen 0.5, 2a, 2b, 3.1.1.1, 3.3, 3.4 und 5.1 einsetzen werden, die als fehlend lokalisiert wurden. Aus der Tabelle 15 geht hervor, dass sich die Anzahl inkonsistente Daten durch die Einsetzung klar erhöht hat. Der Grund hierfür ist, dass nicht alle Fragen, die in die Tests miteinbezogen werden, verarbeitet wurden. Dies hatte zur Folge, dass Inkonsistenzen zu den nicht verarbeiteten Variablen eingesetzt wurden. Die Tests bezüglich der Variablen, die vollumfänglich behandelt werden konnten, werden nach der NNE nicht mehr verletzt.

Die Tests 8, 10 und 23 verzeichnen einen derart hohen Anstieg, da von der Frage 3 nur die Unterrubrik 3.1.1.1 eingesetzt wurden. Den Institutionen Einheiten, welche nie geantwortet haben, wurde hier notgedrungen eine Inkonsistenz eingesetzt.

Der Anstieg an Inkonsistenzen der Tests 1, 4, 5 und 6 ist damit zu erklären, dass in den betroffenen Variablen oft nicht alle eingesetzt werden mussten, was zu Inkonsistenzen zwischen den eingesetzten und den vorhandenen Variablen führte.

In der Tabelle 29 ist die Anzahl univariater Ausreisser aufgrund der *Huber-type skipped mean* Verwerfungsregel vor und nach der NNE aufgeführt. Zur Veranschaulichung wurde κ variiert, mit der Wahl $\kappa_{i,t} = \kappa, \forall i, t$. Bei $\kappa = 10$ muss damit gerechnet werden, dass jedes Jahr bei 20 bis 30 Institutionen aussergewöhnliche Antworten nachgefragt werden müssten. Der Anstieg von 8 Jahres Einheiten für $\kappa = 10$ bedeutet einen Anstieg von 6% gegenüber den 133 vor der NNE. Somit entspricht die Mehrheit der eingesetzten Variablenwerte dem Wertebereich der bereits vorhandenen.

Aus der Tabelle 30 geht hervor, dass die Anzahl Jahres Einheiten, die als Ausreisser der Entwicklung erkannt wurden (*Hidiroglou-Berthelot*), mit der NNE merklich gestiegen ist. Dies ist aufgrund der Jahres Einheiten zurückzuführen, die vor der NNE nicht in die Ausreisser Analyse miteinbezogen werden konnten, da sie nur fehlende Werte hatten. Somit ist die Konsistenz der Entwicklung der Daten nicht in allen Fällen gewährleistet. Vor allem die Veränderung von 1998 zu 1999 weist einen Anteil von mindestens 4.8%[9] von an Jahres Einheiten auf, die nach der NNE eine aussergewöhnliche Entwicklung im Personalaufwand aufweisen. Zeitreihen Analysen sollten deshalb erst ab 1999 durchgeführt werden. Auch die folgenden Jahre verzeichnen einen nicht vernachlässigbaren Anteil an Ausreissern auf.

[9] 111 Ausreisser von 2304 Jahres Einheiten, die 1999 in der SM waren.

Tabelle 14 Die 6 NNE-Schlaufen (Schlaufe=1,..,6) und die konservative HIE der Frage 3.1.1.1, die nach der 4. NNE-Schlaufe durchgeführt wurde. Der Datentyp bezeichnet *item nonresponse*, 'i_nr', und Institutionen Einheiten mit mindestens einem Jahr ohne Angaben *unit nonresponse*, 'u_nr'. Die Schichtung bezeichnet die Variablen, die bei Spender und Empfänger übereinstimmen müssen. 'C_1' und 'C_2' bezeichnen die Einsetzungsklassen. Unter 'Anzahl Jahres Einheiten' und 'Anzahl Institutionen Einheiten' sind das Total Empfänger der jeweiligen Einheit sowie die Anzahl Empfänger die eingesetzt wurden ('eingesetzt') und die Anzahl, die nicht eingesetzt werden konnte ('nicht eing.') aufgeführt.

Schlaufe	Daten-	Schichtung	Anzahl Jahres Einheiten			Anzahl Institutionen Einheiten		
	typ		Total	eingesetzt	nicht eing.	Total	eingesetzt	nicht eing.
1	i_nr	Kt, Jahr, C_1	1585	1493	92	341	320	21
2	i_nr	Kt, Jahr, C_2	92	43	49	21	10	11
3	i_nr	Kt, Jahr	49	24	25	11	6	5
4	u_nr	Kt, Jahr	1482	871	611	359	228	131
HIE 3.1.1.1	i_nr, u_nr		636	636	0	136	136	0
5	i_nr, u_nr	Kt, Jahr	636	617	19	136	126	10
6	i_nr, u_nr	Jahr	19	19	0	10	10	0

Tabelle 15 Anzahl Jahres Einheiten, die vor und nach der NN-Einsetzung den Test 'Test' nicht bestanden. 'Δ' bezeichnet die Veränderung.

Test	vorher	nachher	Δ	Δ [%]
1	12	49	37	308
2	201	190	-11	-5
3	205	51	-154	-75
4	517	547	30	6
5	466	499	33	7
6	25	30	5	20
8	81	495	414	511
10	88	508	420	477
12	469	469	0	0
14	23	23	0	0
15	524	524	0	0
16	1544	1544	0	0
17	1081	1081	0	0
18	1022	1022	0	0
19	410	410	0	0
21	39	39	0	0
22	180	180	0	0
23	46	435	389	846
25*	417	417	0	0
26*	3776	3776	0	0
30*	2552	2552	0	0
31	916	814	-102	-11
32	168	0	168	-100
33	168	0	168	-100
36	153	0	-153	-100

*: weniger restriktive Form als in SM3.0.

4.4 Schlussfolgerungen der Einsetzungen in der SM

Der Auftrag, fehlende Werte in der SM einzusetzen, konnte mit keiner der angewandten Methoden vollständig erfüllt werden. Der Hauptgrund hierfür war, dass die Fragen 3.1 (Beherbergung) und 4 (Personen nach Betreuungsstufen) von den sozialmedizinischen Institutionen offensichtlich nicht miteinander verbunden wurden. Hier besteht ein Handlungsbedarf in der Konzeption des Fragebogens. Die deduktive Einsetzung, sowie die 'historisch-iterative' Einsetzung konnten auch ohne die oben genannten Fragen mit einzubeziehen, nicht für alle fehlenden Werte einen Wert einsetzen. Die SAS-Prozedur `proc mi` wurde nicht angewandt, nachdem die Annahmen der in dieser Prozedur verwendeten Theorie untersucht wurden. Es stellte sich heraus, dass keine der Annahmen uneingeschränkt angenommen werden kann. Schlussendlich wurde eine Nächster-Nachbar Einsetzungsmethode angewandt, da es sich dabei um eine multivariate Einsetzungsmethode handelt und sie zum Ziel hat, die Abhängigkeiten der Variablen beizubehalten. Zudem garantiert uns diese Methode, dass, solange ein Spender vorhanden ist, alle fehlenden Werte eingesetzt werden. Vom Spender wurden jeweils nur die Daten übernommen, die beim Empfänger fehlten. Auf diese Weise wurden die bestehenden Angaben nie geändert.

Unabhängig von der Einsetzungsmethode war es nicht möglich, für alle Variablen, die eingesetzt werden sollten, eindeutig zu bestimmen, ob sie fehlend oder 0 sind. Fehlende Werte konnten nur auf aggregierten Niveaus, z.B. Total der Stellen, bestimmt werden. Deshalb sollten die Einsetzungen in den Unterkategorien nicht für Zeitreihen-Analysen verwendet werden, sondern nur die Totale, die in einem ersten Schritt eingesetzt wurden.
Aufgrund von univariaten Methoden zur Ausreisserlokalisierung (*Huber-type skipped mean* Verwerfungsregel, *Hidiroglou-Berthelot* Methode) kamen offensichtliche Ausreisser zu Tage, die vor den Einsetzungen bestanden haben. Mit dem Einbezug dieser oder ähnlicher Methoden in die Datenaufbereitung und mittels der gewählten Parameter, müsste jährlich mit ungefähr 30 sozialmedizinischen Institutionen gerechnet werden, für die Nachforschungen bezüglich der Ausreisser gemacht werden müssten. Die Einsetzungen haben die Problematik der Ausreisser noch verschärft. Vor allem die Anzahl der Ausreisser der Entwicklung stieg an. Dies bedeutet, dass bei Zeitreihen-Analysen, selbst auf dem Niveau der Totale, mit Verzerrungen gerechnet werden muss. Die Ausreisser wurden auf dem Niveau Schweiz gesucht. Durch eine Schichtung, welche mit der interessierenden Variable korreliert, steigt die Anzahl Ausreisser an. Denn für die Lokalisierung wurde keine weitere Hilfsvariable verwendet.
Da im Kanton SG im Jahr 1998 von den 171 sozialmedizinische Institutionen deren 164 mindestens einen fehlenden Wert aufwiesen, ist davon abzuraten, das Jahr 1998 in die Analysen mit einzubeziehen. Probleme dieses Jahres zeigten sich auch durch den Anstieg der Ausreisser der Entwicklung von 1998 zu 1999. In der Variablen 'Personalaufwand' wurden diese fast ausschliesslich durch die Einsetzungen im Kanton SG hervorgerufen.

Aufgrund dieser Schlussfolgerungen hat GES beantragt, die Einsetzungen mit derselben Methode für die gleichen Variablen und für die Jahre 1999-2004 zu wiederholen. Diese Arbeiten wurden von METH 2006 abgeschlossen. In den Abbildungen 2 bis 6 im Anhang A.3.2 ist exemplarisch der Einfluss auf die Standardtabellen von 2002 dieser Einsetzungen wiedergegeben. Die Abweichungen sind zum Teil nicht vernachlässigbar, vor allem im Kanton Bern, was vor allem auf Einheiten Ausfälle zurückzuführen ist. Da die Analyse der Entwicklung von Indikatoren das erklärte Ziel der Einsetzungen war, wurde die Aufmerksamkeit auf den Einfluss auf einige Indikatoren konzentriert, die von GES vorgeschlagen worden sind, cf. Abbildungen 7 bis 12. Der Einfluss der NNE ist in den meisten Fällen gering ausgefallen. Auf den Trend hatten sie ausser bei einem Indikator keinen Einfluss.

Nach Absprache mit GES wurden die Einsetzungen in dieser Form akzeptiert.

5 Krankenhausstatistik

Die Krankenhausstatistik (KS) ist eine obligatorische Vollerhebung, die seit 1997 erhoben wird. Die Daten aller betroffenen Krankenhäuser sind in einer ORACLE Datenbank gespeichert. Die Identifikatoren der Krankenhäuser sind auch für Krankenhäuser vorhanden, die nicht geantwortet haben. Für die Einsetzungen wurde METH ein Auszug in Form einer ACCESS Datenbank gegeben, der die Variablen enthält, die von GES vom thematischen Standpunkt aus als notwendig und hinreichend angesehen werden, um die Einsetzungen durchzuführen. Insbesondere enthält die Datenlieferung sämtliche Variablen, die einen direkten Bezug mit den Variablen haben, die eingesetzt werden sollen. Die Datenlieferung umfasst die Daten der Jahre 1998 bis 2002. Fehlende Werte werden ebenfalls für diese Zeitspanne eingesetzt.
Aufgrund der Erfahrungen, welche aus den Einsetzungen in der SM gewonnen werden konnten, wurde die Einheit von Anfang an als Krankenhaus Einheit, in der SM Institutionen Einheit genannt, definiert. Das heisst, dass jedes Krankenhaus über die Jahre 1998 bis 2002 als Einheit angesehen wird, unabhängig davon, in wievielen Jahren das Krankenhaus in der Grundgesamtheit der KS vorhanden war.
Zudem wurde versucht, die Problematik mit den Ausreissern verstärkt zu berücksichtigen. Deshalb wurde eine NNE mit einer robusten Distanzberechnung verwendet.

GES wurden die vollständig eingesetzten Daten geliefert. Sowie die Standardtabellen 1998 - 2002 mit Angaben zu den Werten vor der Einsetzung, nach der Einsetzung, dem Anteil des Zellwertes, welcher von Einheiten herrührt, die als Ausreisser markiert wurden und dem Anteil des Zellwertes, welcher eingesetzt wurde (cf. Auszug aus der Tabelle in der Abbildung 19). Hiermit wurde GES ein Anhaltspunkt gegeben, um Analysen zu relativieren, falls die Gefahr eines Bias besteht (grosse Anteile). Zusätzlich wurden GES Grafiken geliefert, welche den Einfluss der Einsetzungen für die Totale der Standardtabellen pro Kanton aufzeigen. Die Entwicklung einiger Indikatoren wurden GES ebenso geliefert, um die Einsetzungen beurteilen zu können und ebenso zur Unterstützung der Analysen, damit gewisse Interpretationen relativiert werden können (Abbildung 20 bis 37).
Aufgrund der Studie dieser Informationen wurden die Einsetzungen von GES gutgeheissen.

5.1 Daten, Variablen und Hilfsvariablen

In der Tabelle 32 im Anhang B.1 sind die Variablen aufgeführt, für welche GES prioritär vollständige Daten wünschte. Die Variablen bezüglich des Personals wurden für jeden Beschäftigten, der innerhalb des Erhebungsjahres im Krankenhaus gearbeitet hatte, geliefert. Diese Daten enthalten Merkmalsausfälle, die Vollständigkeit der Liste der Beschäftigten konnte jedoch nicht überprüft werden, da keinerlei Anhaltspunkte vorhanden waren. In einigen Fällen besteht der Verdacht, dass sie nur teilweise geliefert wurden, da in angrenzenden Jahren ausführlichere Listen geliefert wurden. Bei den Besoldungskonti konnten fehlende Werte aufgrund der Angaben im Finanzbuchhaltungsteil des Fragebogens eruiert werden. Fehlte der Finanzbuchhaltungsteil jedoch, war dies nicht möglich. Aus diesen Gründen wurde mit GES vereinbart, dass sich die Einsetzungen bezüglich des Personals auf die Totale basierend auf der Kreuzung der Personalvariablen Geschlecht und Beschäftigungsgrad mit den Variablen 'durchschnittlicher Beschäftigungsgrad', 'Besoldungskonto' und 'Anzahl Diensttage' beschränkt werden. Dies entsprach 27 Variablen: [keine Kreuzung, Geschlecht (Mann, Frau), Besolungskonti (Kto 30 - 35)] x [$\sum$ B.grad[10], $\sum$ Durchschnittlicher B.grad, $\sum$ Anstellungsdauer].
Für die Einsetzungen der Finanzbuchhaltung wurden zuerst nur die Totale mit einbezogen und die Subtotale in einem zweiten Schritt nachdem ein Spender gefunden worden war, vom selben

[10] B.grad steht für Beschäftigungsgrad

Spender übernommen.

Anhand der Variablen der Tabelle 16 wurden Einsetzungsklassen wie im Abschnitt 3.2 beschrieben gebildet. Dabei wurde jeweils der Wert des letzten Jahres übernommen, falls er vorhanden war. Ansonsten wurde eine konservative HIE angewandt (cf. 3.4). Es zeigte sich jedoch, dass die Anzahl Krankenhäuser mit fehlenden oder extremen Werten zu gross war, um anhand dieser Variablen Einsetzungsklassen mit genügend Spendern zu bilden. So war der Anteil Spender innerhalb der Krankenhaustypen immer kleiner als 50% und der Anteil Empfänger war bei 4 von 5 Typen grösser als der Anteil Spender.

Stattdessen wurde die robuste Mahalanobis-Distanz des *TRC*-Algorithmus verwendet, siehe 3.6, um Grössenklassen anhand eines multivariaten Masses zu definieren. Diese wurde mittels der durchschnittlichen robusten Mahalanobis-Distanz über die Jahre, in welchen das Krankenhaus in der KS war definiert. Die Grössenklassen wurden iterativ anhand der Grafiken der logarithmierten durchschnittlichen Mahalanobis-Distanz, dl_{maha}, eruiert (cf. Anhang B.3). Daraus entstanden 4 Grössenklassen mit den Grenzen u_{maha}:

$$u_{maha} = \begin{cases} 1 & \text{wenn } dl_{maha} < 2 \\ 2 & \text{wenn } 2 \leq dl_{maha} < 6.3 \\ 3 & \text{wenn } 6.3 \leq dl_{maha} < 8 \\ 4 & \text{andernfalls } (dl_{maha} \geq 8). \end{cases}$$

Tabelle 16 Hilfsvariablen, die für die Einsetzungen in Betracht gezogen wurden.

Variable	**vollständig vorhanden**
Kanton (Verwaltungssitz)	ja
Jahr	ja
Anzahl Angestellte	nein
Anzahl Betten	nein
Krankenhaus-Typologie (von 1999), siehe Tabelle 31	nein
Korrespondenzsprache	nein
Rechtlich-wirtschaftlicher Status	nein

5.2 Antwortausfälle, Inkonsistenzen und Spenderpool

Die Daten der KS 1998-2002 weisen insgesamt 439 Krankenhäuser auf, wovon 6 (1.4%) nie geantwortet haben. Weitere 33 Krankenhäuser (7.5%) haben mindestens während eines Jahres nicht geantwortet.
Vor allem die Variablen bezüglich des Bestandes und der Diensttage weisen mit rund 12% fehlender Werte in Jahres Einheiten die grössten Antwortausfallquoten auf, wie aus Tabelle 34 hervorgeht.

Vor allem zwei Tests, die Nummern 1343 und 1371 der Tabelle 33, brachten einige Inkonsistenzen zu Tage. Das Ausmass an Inkonsistenzen der SM wurde jedoch nicht erreicht.

Extreme Werte wurden anhand der Methoden lokalisiert, die im Abschnitt 2 besprochen wurden. Für die *Huber-type skipped mean* Verwerfungsregel wurde wie in der SM $\kappa = 10$ gewählt. Von den 439 Krankenhäusern der KS wiesen 120 (27.3%) mindestens einen extremen Wert in den Einsetzungsvariablen auf, siehe Tabelle 17. 76 von diesen Krankenhäusern wiesen ebenfalls mindestens einen fehlenden Wert auf. Wurden die Ausreisser aus dem Spenderpool

ausgeschlossen, so standen den 239 Krankenhäusern mit fehlenden Werten 156 Spender gegenüber, siehe Tabelle 35. Mit anderen Worten, auf jeden Spender entfielen 1.5 Empfänger. Aufgrund dieses 'Missverhältnisses' wurden die Einsetzungen in zwei Variablengruppen aufgeteilt. Die erste Variablengruppe bestand aus allen Variablen ausser den Variablen bezüglich des Personals. Die zweite Gruppe bestand ausschliesslich aus den Variablen bezüglich des Personals. Das Verhältnis Empfänger:Spender konnte damit in beiden Gruppen verbessert werden. In der ersten Gruppe auf 158:237=0.66, in der zweiten Gruppe auf 157:238=0.66. Mit anderen Worten mussten für beide Variablengruppen bei rund 40% der Krankenhäuser mindestens eine Einsetzung vorgenommen werden. Dies ist immer noch ein sehr hoher Wert. Die Wahrscheinlichkeit einer Verzerrung ist umso grösser, ebenso ist mit einer Unterschätzung der Varianz zu rechnen. Ausreisserwerte wurden auf fehlend gesetzt, damit sie in der Distanzberechnung kei-

Tabelle 17 Kreuztabelle der Indikatorvariablen 'Ausreisser' und 'Einsetzungscharakteristik'.

	Einsetzungscharakteristik							
	Empfänger		Aussenseiter		Spender		Total	
Ausreisser	**#**	**[%]**	**#**	**[%]**	**#**	**[%]**	**#**	**[%]**
Ja	76	17.31	44	10.02	0	0.00	120	27.33
Nein	163	37.13	0	0.00	156	35.54	319	72.67
Total	239	54.44	44	10.02	156	35.54	439	100

nen Einfluss hatten. Nach der Einsetzung wurden wieder die Ausreisserwerte eingesetzt. Der Grund hierfür war, dass die meisten Ausreisserwerte als richtig aber nicht-repräsentativ erachtet wurden. Das heisst, dass es sich um einzigartige Werte handelte, die nicht 'vervielfältigt' werden dürfen, cf. (Chambers 1986).

5.3 Einsetzungen

Aufgrund der Erfahrungen, welche aus den Einsetzungen in der SM gewonnen wurden, wurden keine Gegenüberstellungen verschiedener Methoden durchgeführt. Dennoch wurde die Methode gegenüber der SM angepasst und eine robuste Distanz verwendet, cf. 3.6. Der Grund hierfür war, dass einerseits in der KS mehr Ausreisser existierten und andererseits, dass mit der Distanzdefinition, die in der SM verwendet wurde, einzelne Indikatoren nach der Einsetzung unrealistische Schwankungen aufwiesen. Diese Schwankungen konnten anhand der Distanz aus dem *TRC*-Algorithmus teilweise gedämpft werden, wie weiter unten näher erklärt wird.

Für die erste Variablengruppe wurden die Einsetzungen in einem ersten Schritt innerhalb der Einsetzungsklassen, die im Abschnitt 5.1 definiert sind, mittels einer NNE mit robuster Mahalanobisdistanz durchgeführt. Siehe Abbildungen 13 bis 17 für die Verteilung der logarithmierten Mahalanobisdistanz in den vier Einsetzungsklassen. In einem zweiten Schritt wurde die NNE mit robuster Mahalanobisdistanz für die erste Variablengruppe der restlichen 3 Krankenhäuser ohne Grössenklassen durchgeführt. Die zweite Variablengruppe wurde in einem Schritt innerhalb der Grössenklassen verarbeitet. Durch die Aufteilung der Variablen wurde die Kovarianzmatrix in zwei Teile aufgeteilt. Bei 4 Krankenhäusern wurde bei der Einsetzung für die erste Variablengruppe derselbe Spender verwendet, wie bei der zweiten Variablengruppe. Der Zusammenhang dieser beiden Variablengruppen wurde somit stark beeinträchtigt, weshalb diese unabhängig voneinander analysiert werden sollten.
Durch die Einsetzungen wurde die Standardabweichung zwischen -10% und +5% beeinflusst, siehe Abbildung 18. Eingehendere Analysen ergaben, dass die Standardabweichung vor allem bei den Personalvariablen verringert wurde, da dort am meisten Einsetzungen vorgenommen wurden.

Die Einsetzungen hatten nicht zum Ziel, vorhandene Inkonsistenzen zu lösen. Es musste sogar ein Anstieg beobachtet werden, der, ausgenommen vom Test T.1511, darauf zurückzuführen ist, dass die Frage 3.4 nicht eingesetzt worden ist, siehe Tabelle 36. Der Anstieg beim Test T.1511 hält sich mit 5 Jahres-Einheiten in Grenzen, weshalb die Ursache des Anstiegs nicht weiter untersucht wurde.
Die Anzahl univariater Ausreisser, aufgrund der *Huber-type skipped mean* Verwerfungsregel mit $\kappa = 10$, wurde durch die Einsetzungen nicht beeinflusst. Hingegen stieg die Anzahl Ausreisser, die mittels der *Hidiroglou-Berthelot* Methode mit $u = 0.8, A = 0.05$ und $L = 50$ entdeckt wurde an (cf. Tabellen 37 und 38). Die Anzahl Krankenhäuser, welche mindestens eine Variable hat, die eine aussergewöhnliche Entwicklung nach *Hidiroglou-Berthelot* innerhalb der Grössenklassen aufweist, stieg durch die Einsetzungen von 167 auf 220. Somit konnte das Ziel, die Entwicklung nicht zu verzerren nur bedingt erreicht werden. Die grosse Anzahl Variablen im Gegensatz zur kleinen Anzahl Krankenhäuser und somit die beschränkte Grösse des Spenderpools hatte einen entscheidenden Einfluss auf diese Verzerrungen der Entwicklung.
Die Entwicklung einiger Indikatoren, die von GES vorgeschlagen wurden, zeigt auf, dass die Einsetzungen allgemein einen kleinen Einfluss hatten, siehe Abbildungen 20 bis 37. Die Indikatoren bezüglich des Krankenhaustyps 23, 'Andere Spezialkliniken', wurden jedoch durch die Einsetzungen zum Teil stark beeinflusst. Ebenso sind für die Jahre 1998 und 1999 starke Veränderungen beim Typ 22, 'Rehabilitationskliniken', bezüglich der Vollzeitäquivalenten auf 1'000 Pflegetage zu beobachten. Es ist nicht auszuschliessen, dass diese Veränderungen auf einen Einsetzungsbias hinweisen. Dieser kann jedoch nicht abgeschätzt werden. Eine vorsichtige Interpretation, respektive ein Verzicht auf eine Interpretation der Entwicklung dieser Indices scheint angebracht zu sein.

5.4 Schlussfolgerungen der Einsetzungen in der KS

Die fehlenden Werte in der KS wurden mittels einer Nächster-Nachbar Einsetzungsmethode mit robuster Distanzfunktion eingesetzt. Allgemein betrachtet, konnten die Einsetzungen die angestrebten multivariaten Abhängigkeiten beibehalten und das Ausmass an neu kreierten Inkonsistenzen konnte in Grenzen gehalten werden.
In Bezug auf die Ausreisser waren die Einsetzungen jedoch nur bedingt erfolgreich. Obwohl die Anzahl univariater Ausreisser unverändert blieb, konnten zusätzliche Ausreisser der Entwicklung auch durch die Verwendung einer robusten Distanzfunktion nicht verhindert werden.
Dadurch, dass die Frage 3.4 nicht eingesetzt wurde, entstanden durch die Einsetzungen zusätzliche Inkonsistenzen bezüglich dieser Frage. Deshalb sollte diese Frage von den Analysen ausgeschlossen werden.
Der Einfluss auf die untersuchten Indikatoren war gering, abgesehen vom Krankenhaustyp 23 (Andere Spezialkliniken), bei welchem zum Teil aussergewöhnliche Entwicklungen festzustellen sind. Somit wurden bezüglich dieser Indikatoren die multivariaten Abhängigkeiten grösstenteils beibehalten. Dennoch sollten allfällige Veränderungen nur mit der gebotenen Vorsicht interpretiert werden.

6 Schlussfolgerung

Im vorliegenden Methodenbericht wurden verschiedene Einsetzungsmethoden untersucht, um in relativ kleinen Datensätzen mit vielen Variablen über mehrere Jahre gleichzeitig die fehlenden Werte zu ersetzen. Die Qualität der Daten war insofern problematisch, dass einige Institutionen und Krankenhäuser ein ganzes Jahr oder oft auch einzelne Rubriken nicht beantwortet haben. Ausserdem enthielten die Daten eine grosse Anzahl Inkonsistenzen und Ausreisser, denen in Zukunft mehr Aufmerksamkeit geschenkt werden sollte. Die Einsetzungen hatten aber nicht zum Ziel diese Inkonsistenzen und Extremwerte zu bearbeiten.
Allgemein kann bei Daten mit fehlenden, fehlerhaften, inkonsistenten und extremen Werten die Datenqualität durch Einsetzungen nicht verbessert werden. Ist der Anteil dieser Daten gross, muss unabhängig von der Einsetzungsmethode mit Verzerrungen und einer Unterschätzung der Varianz gerechnet werden. Ausreissern kann im besten Fall mit robusten Einsetzungsverfahren Rechnung getragen werden. Hingegen sollten Inkonsistenzen vor der Einsetzung bearbeitet werden, oder deren Bearbeitung sollte in den Einsetzungsprozess mit einbezogen werden. Dies wirft jedoch das Problem der Lokalisierung der Werte, die abgeändert werden müssen, auf. Zudem muss damit gerechnet werden, dass Variablenwerte abgeändert werden, die von thematischen Spezialisten als realistisch angesehen werden.

Die Einsetzungen haben in den beiden Statistiken vor allem ein Verbesserungspotential bei den Erhebungen aufgezeigt und gewisse Analysen ermöglicht, die zuvor nicht möglich gewesen sind. Die Analysen werden durch die Einsetzungen aber nicht vereinfacht, da die Einsetzungsflaggen in alle Analysen miteinbezogen werden müssen, um verzerrte Interpretationen möglichst zu verhindern. Hierfür wurden einige Vorschläge gemacht.
Zeitreihen sollten im vorliegenden Fall nur mit der nötigen Vorsicht analysiert werden. Das heisst, dass bei der Entwicklung von Totalen überprüft werden sollte, welcher Anteil der Veränderung durch die Einsetzungen hervorgerufen wurde und welcher Anteil von Ausreissern herrührt. Werden diese Anteile in Bezug auf die Veränderung als beträchtlich gewertet, so kann dies ein Hinweis auf eine Verzerrung sein. Klare Regeln für eine solche Abschätzung existieren jedoch nicht.

In den Einsetzungen der SM und der KS hat sich auch gezeigt, dass ein Bedarf an konzeptuellen Richtlinien für die Datenaufbereitung von derart vielen Variablen nötig ist, damit die Produktion der Statistik effizienter gesteuert und kontinuierlich verbessert werden kann.

Bezüglich der analysierten Methoden ergab die Übernahme von Daten aus angrenzenden Erhebungen (Historisch iterative Einsetzungen, HIE) für Variablen mit einer kleinen Variabilität über die Zeit akzeptable Resultate. Für Variablen mit einer grösseren Variabilität muss bei Erhebungen wie der SM mit einer beträchtlichen Anzahl zusätzlicher Inkonsistenzen gerechnet werden. Zudem kann mittels der Übernahme von Daten oft nicht für alle fehlenden Werte ein Ersatz gefunden werden. Diese Methode kam schlussendlich vor allem für die Einsetzung von Hilfsvariablen zur Anwendung.

Bei deduktiven Einsetzungen in der SM traten generell dieselben Probleme wie bei der HIE auf, wobei die Anwendung eingeschränkt war. Deshalb wurde diese Methode nicht weiter angewandt.

Die Nächster-Nachbar Einsetzungsmethode (NNE) ist allgemein geeignet, um multivariate Abhängigkeiten mit einzubeziehen, wobei nicht garantiert werden kann, dass immer ein Spender gefunden werden kann. So mussten z.B. in der SM 6 Iterationen durchgeführt werden, wobei die Bedingungen bezüglich der Hilfsvariablen jeweils gelockert wurden.
Es muss damit gerechnet werden, dass durch die NNE zusätzliche Ausreisser der Veränderung

kreiert werden auch wenn alle Jahre gleichzeitig eingesetzt werden. Denn es kann durchaus vorkommen, dass in jedem Jahr unterschiedliche Variablen eingesetzt werden müssen, was zu ungewöhnlichen Veränderungen führen kann.
Die NNE erlaubt keine Schätzungen auf Einheitenebene. Das heisst, die eingesetzten Werte dürfen nicht auf den Empfänger bezogen werden. Denn bei dieser Einsetzungsmethode geht es unter anderem darum, die multivariaten Charakteristiken zu berücksichtigen, nicht aber um die 'wahren' Werte herzustellen. Deshalb ist bei der Analyse von kleinen Bereichen, z.B. Institutionen des Typs S22 (Andere) oder Krankenhäuser des Typs K23 (Andere), Vorsicht geboten.
Durch die Verwendung der NNE mit zufälliger Auswahl aus der Nachbarschaft wurde die Unterschätzung der Variabilität reduziert. Bei sämtlichen Analysen sollte dennoch berücksichtigt werden, dass die Varianz für die meisten Variablen unterschätzt wird. Das Ausmass der Unterschätzung kann aber nicht beziffert werden.

Anhang

A Statistik der sozialmedizinischen Institutionen

A.1 Typologie und Variablen

Tabelle 18 Typologie der sozialmedizinischen Institutionen ('Typ99'). 'Typ' alleine steht für die Nummerierung, die im Bericht teilweise verwendet wurde.

Typ99	Typ	Bezeichnung
S 11	1	Heime für pflegebedürftige Betagte und/oder chronischkranke
S 12	2	Altersheime
S 21	3	Institutionen für Behinderte und Suchtkranke
S 22	4	Andere sozialmedizinische Insitutionen

Tabelle 19 Einsetzungsvariablen der SM. Der Zeitpunkt oder die Periode, auf welche die Variable sich bezieht, wird mit Stichtag bezeichnet.

Fragebogenabschnitt	Variable	Stichtag
0. Allgemeine Angaben	Juristisch-wirtschaftlicher Status	Jahr
2.a Beschäftigung	Total engerer Betrieb	31.12.
2.a Beschäftigung	Gesamttotal Betrieb	31.12.
2.b	Personalbestand nach Tätigkeitsbereich pro Ausbildungsniveau	31.12.
3.1 Beherbergung	Betten/Bewohner/Tage nach Unterbringungsart	31.12.
3.3 Beherbergung/Betreuung	Kanton des Wohnortes der Bewohner/Betreuten	31.12.
3.4 Beherbergung/Betreuung	Verteilung der Bewohner nach Geschlecht und Altersklasse	31.12.
4.1 Betreuungsstufe	Bewohner/Betreute nach Dauer der benötigten Pflege	31.12.
4.3 Betreuungsstufe	Bewohner/Betreute nach Art der Behinderung	31.12.
5.1 Betriebsaufwand	Personalkosten	Jahr
5.1 Betriebsaufwand	Anderer Betriebsaufwand	Jahr
5.1 Betriebsaufwand	Betriebsgewinn	Jahr
5.1 Betriebsaufwand	Total Aufwand	Jahr
5.1 Betriebsertrag	Pensionstaxen	Jahr
5.1 Betriebsertrag	Beiträge AHV	Jahr
5.1 Betriebsertrag	Beiträge IV	Jahr
5.1 Betriebsertrag	Beiträge Kanton	Jahr
5.1 Betriebsertrag	Beiträge Gemeinde	Jahr
5.1 Betriebsertrag	Betriebsverlust	Jahr
5.1 Betriebsertrag	Total Ertrag	Jahr

Tabelle 20 Einsetzungsvariablen. Ist eine Variable in der Datenbank vorhanden, so ist 'In der DB'='ja', ansonsten 'nein'. Durch 'Kreuzungsvariablen' wird das Niveau der Einsetzung definiert wird. '# Var.' steht für die Anzahl Variablen, die nach der Kreuzung (effektiv) eingesetzt werden müssen. Die Vollzeitäquivalenten sind mit 'VZÄ' abgekürzt. Die Abkürzungen der Variablennamen ist unter 'Abk.' aufgeführt, wobei gleichzeitig Indices für die Institution und das Jahr in die Notation aufgenommen wurden.

Abschnitt des Fragebogens	Variablen Beschreibung	In der DB	Kreuzungs-variablen	# Var.	Abk.
0. Allg. Angaben	Rechtlich-wirtschaftlicher Statuts	ja		1	$u^t_{i,0}$
2. Beschäftigung	Personen engerer Betrieb ohne Ausbildung	nein[11]		1	$u^t_{i,1}$
	VZÄ engerer Betrieb ohne Ausbildung	nein[11]		1	$u^t_{i,2}$
	Personal in Nebenbebetrieb od. unbezahlt	ja		1	$u^t_{i,3}$
	VZÄ in Nebenbebetrieb od. unbezahlt	ja	Ausbildungsniveau	4	$u^t_{i,4} - u^t_{i,7}$
	Personal mit Berufsausbildung	nein	Ausbildungsbereich	6	$u^t_{i,8} - u^t_{i,13}$
	Personal mit höherer Berufsausbildung	nein	Ausbildungsbereich	6	$u^t_{i,14} - u^t_{i,19}$
	Personal mit universitärer Ausbildung[12]	nein	Ausbildungsbereich	6	$u^t_{i,20} - u^t_{i,25}$
	VZÄ mit Berufsausbildung	nein	Ausbildungsbereich	6	$u^t_{i,26} - u^t_{i,31}$
	VZÄ mit höherer Berufsausbildung	nein	Ausbildungsbereich	6	$u^t_{i,32} - u^t_{i,37}$
	VZÄ mit universitärer Ausbildung[b]	nein	Ausbildungsbereich	6	$u^t_{i,38} - u^t_{i,43}$
3. Beherbergung	Anzahl Betten	nein	Unterbringungsart	7	$v^t_{i,1} - v^t_{i,7}$
	Anzahl Bewohner	nein	Unterbringungsart	6	$v^t_{i,8} - v^t_{i,13}$
	Anzahl Aufenthaltstage	nein	Unterbringungsart	7	$v^t_{i,14} - v^t_{i,20}$
Betreuung	Anzahl Bewohner	ja	Kanton Bewohner	27	$v^t_{i,21} - v^t_{i,47}$
	Anzahl Externe	ja	Kanton Externe	28	$v^t_{i,48} - v^t_{i,75}$
	Anzahl männliche Bewohner	ja	5-Jahres Alterskl.	21	$v^t_{i,76} - v^t_{i,96}$
	Anzahl weibliche Bewohner	ja	5-Jahres Alterskl.	21	$v^t_{i,97} - v^t_{i,117}$
4. Betreuungsstufe	Anzahl Personen	ja	Pflegedauer	4	$v^t_{i,118} - v^t_{i,121}$
	Anzahl Personen	ja	Art d. Behinderung	7	$v^t_{i,122} - v^t_{i,128}$
	Personalaufwand	ja		1	$w^t_{i,1}$
	Anderer Betriebsaufwand	ja		1	$w^t_{i,2}$
	Weiterer Betriebsaufwand	ja		1	$w^t_{i,3}$
	Pensionstaxen	ja		1	$w^t_{i,4}$
5. Buchhaltung	Beiträge AHV	ja		1	$w^t_{i,5}$
	Beiträge IV	ja		1	$w^t_{i,6}$
	Beiträge Kanton	ja		1	$w^t_{i,7}$
	Beiträge Gemeinde	ja		1	$w^t_{i,8}$
	Weitere Erträge	nein		1	$w^t_{i,9}$
Total Anzahl Einsetzungsvariablen				**181**	

[11] Kann bei vorhandenen Daten durch Deduktion berechnet werden.

[12] Die universitäre Ausbildung ist in den medizinischen und Pflegeausbildungen (II) nicht möglich: $u^t_{i,21} = u^t_{i,39} = 0 \quad \forall i, t$.

Tabelle 21 Ausprägungen und Berechnungen der Variablen, welche eingesetzt werden sollen. Die Variablen, die nicht eingesetzt werden müssen, aus technischen Gründen aber dennoch eingesetzt werden, sind in der Spalte 'Tabelle des Fragebogens' kursiv gedruckt.

Tabelle des Fragebogens	**Ausprägungen und Berechnungen**					
2.a **Personalbestand**	Gesamtbestand		Vollzeitäquivalente			
	Personen	VZÄ	o. Ausbild.	Berufsausb.	Höhere B.ausb.	Univ.
Total engerer Betrieb	$u^t_{i,1} + \sum_{j=8}^{25} u^t_{i,j}$	$u^t_{i,2} + \sum_{j=26}^{43} u^t_{i,j}$	$u^t_{i,2}$	$\sum_{j=26}^{31} u^t_{i,j}$	$\sum_{j=32}^{37} u^t_{i,j}$	$\sum_{j=38}^{43} u^t_{i,j}$
Total Nebenbetriebe	$u^t_{i,4}$	$\sum_{j=4}^{7} u^t_{i,j}$	$u^t_{i,4}$	$u^t_{i,5}$	$u^t_{i,6}$	$u^t_{i,7}$
Gesamttotal Betrieb	$u^t_{i,1} + u^t_{i,3} + \sum_{j=8}^{25} u^t_{i,j}$	$u^t_{i,2} + \sum_{j=4}^{7} u^t_{i,j} + \sum_{j=26}^{43} u^t_{i,j}$	$u^t_{i,2} + u^t_{i,4}$	$\sum_{j=26}^{31} u^t_{i,j} + u^t_{i,5}$	$\sum_{j=32}^{37} u^t_{i,j} + u^t_{i,6}$	$\sum_{j=38}^{43} u^t_{i,j} + u^t_{i,7}$
2.b **Personalbestand**	Berufsausbildung		Höhere Berufsausbildung		Universität[a]	
	Personen	VZÄ	Personen	VZÄ	Personen	VZÄ
Med.- und Pflegeausb. (I)	$u^t_{i,8}$	$u^t_{i,26}$	$u^t_{i,14}$	$u^t_{i,32}$	$u^t_{i,20}$	$u^t_{i,38}$
Med.- und Pflegeausb. (II)	$u^t_{i,9}$	$u^t_{i,27}$	$u^t_{i,15}$	$u^t_{i,33}$	$u^t_{i,21}$	$u^t_{i,39}$
Soz. u. and. Ausb.	$u^t_{i,10}$	$u^t_{i,28}$	$u^t_{i,16}$	$u^t_{i,34}$	$u^t_{i,22}$	$u^t_{i,40}$
Therapeutische Ausb.	$u^t_{i,11}$	$u^t_{i,29}$	$u^t_{i,17}$	$u^t_{i,35}$	$u^t_{i,23}$	$u^t_{i,41}$
Admin. u. Hotel Ausb.	$u^t_{i,12}$	$u^t_{i,30}$	$u^t_{i,18}$	$u^t_{i,36}$	$u^t_{i,24}$	$u^t_{i,42}$
Restauration u. and. Ausb.	$u^t_{i,13}$	$u^t_{i,31}$	$u^t_{i,19}$	$u^t_{i,37}$	$u^t_{i,25}$	$u^t_{i,43}$
3.1 **Beherbergung**	Betten	Bewohner	Aufenth. Tage			
Betagte, chron. Kranke	$v^t_{i,1}$	$v^t_{i,8}$	$v^t_{i,14}$			
Psychiatriepatienten	$v^t_{i,2}$	$v^t_{i,9}$	$v^t_{i,15}$			
Behinderte	$v^t_{i,3}$	$v^t_{i,10}$	$v^t_{i,16}$			
Drogenabhängige	$v^t_{i,4}$	$v^t_{i,11}$	$v^t_{i,17}$			
Alkoholkranke	$v^t_{i,5}$	$v^t_{i,12}$	$v^t_{i,18}$			
Psychosoziale Fälle	$v^t_{i,6}$	$v^t_{i,13}$	$v^t_{i,19}$			
andere Beherbergung	$v^t_{i,7}$	$\sum_{j=76}^{117} v^t_{i,j} - \sum_{j=8}^{13} v^t_{i,j}$	$v^t_{i,20}$			

[a] Die universitäre Ausbildung ist in den medizinischen und Pflegeausbildungen (II) nicht möglich: $u^t_{i,21} = u^t_{i,39} = 0 \quad \forall i, t$.

Tabelle 22 Ausprägungen und Berechnungen der Variablen, welche eingesetzt werden sollen.

Tabelle des Fragebogens	**Ausprägungen und Berechnungen**	
3.3 **Herkunft der Patienten**	Bewohner	Externe
Kanton AG - ZH (alphabetisch geordnet)	$\{v^t_{i,21}, \ldots, v^t_{i,46}\}$	$\{v^t_{i,48}, \ldots, v^t_{i,73}\}$
Unbekannter Kanton	$v^t_{i,47}$	$v^t_{i,74}$
Wohnort Ausland	$\sum_{j=76}^{117} v^t_{i,j} - \sum_{j=21}^{47} v^t_{i,j}$	$v^t_{i,75}$
Total Kantone	$\sum_{j=21}^{46} v^t_{i,j}$	$\sum_{j=48}^{73} v^t_{i,j}$
Total unbekannter Kanton u. Ausland	$\sum_{j=76}^{117} v^t_{i,j} - \sum_{j=21}^{46} v^t_{i,j}$	$v^t_{i,74} + v^t_{i,75}$
Gesamttotal	$\sum_{j=76}^{117} v^t_{i,j}$	$\sum_{j=48}^{75} v^t_{i,j}$
3.4 **Alter der Bewohner**	männl. Bewohner	weibl. Bewohner
20 5-Jahres Klassen (0-4,...,95+)	$\{v^t_{i,76}, \ldots, v^t_{i,95}\}$	$\{v^t_{i,97}, \ldots, v^t_{i,116}\}$
Alter unbekannt	$v^t_{i,96}$	$v^t_{i,117}$
Total mit Altersangabe	$\sum_{j=76}^{95} v^t_{i,j}$	$\sum_{j=97}^{116} v^t_{i,j}$
Gesamttotal	$\sum_{j=76}^{96} v^t_{i,j}$	$\sum_{j=97}^{117} v^t_{i,j}$

Tabelle 23 Ausprägungen und Berechnungen der Variablen, welche eingesetzt werden sollen.

Tabelle des Fragebogens	**Ausprägungen und Berechnungen**
4.1 **Pflegedauer**	Bewohner und Externe
weniger als 1 h	$v^t_{i,118}$
1 h bis 3 h	$v^t_{i,119}$
3h bis 5 h	$v^t_{i,120}$
mehr als 5 h	$v^t_{i,121}$
Total	$\sum_{j=118}^{121} v^t_{i,j}$
4.3 **Art der Behinderung**	Bewohner und Externe
ohne besondere Beh.	$v^t_{i,122}$
physisch	$v^t_{i,123}$
psychisch	$v^t_{i,124}$
geistig	$v^t_{i,125}$
sinnesbeh.	$v^t_{i,126}$
suchtbehindert	$v^t_{i,127}$
integrationsgeschädigte	$v^t_{i,128}$
andere	$\sum_{j=118}^{121} v^t_{i,j} - \sum_{j=122}^{128} v^t_{i,j}$
Total	$\sum_{j=118}^{121} v^t_{i,j}$

Tabelle 24 Ausprägungen und Berechnungen der Variablen, welche eingesetzt werden sollen. Die Variablennamen der Variablen, die nicht eingesetzt werden müssen, aus technischen Gründen aber dennoch eingesetzt werden, sind kursiv gedruckt.

5.1 **Betriebsrechnung**			
Betriebsaufwand		Betriebsertrag	
Personal	$w_{i,1}^t$	Pensionstaxen	$w_{i,4}^t$
And. Betriebsaufwand	$w_{i,2}^t$	Beiträge AHV	$w_{i,5}^t$
Weiterer Betriebsaufw.	$w_{i,3}^t$	Beiträge IV	$w_{i,6}^t$
		Beitr. Kt	$w_{i,7}^t$
		Beitr. Gemeinde	$w_{i,8}^t$
		Weit. Erträge	$w_{i,9}^t$
Betriebsgewinn	$\max\left(0, \sum_{j=4}^{9} w_{i,j}^t - \sum_{j=1}^{3} w_{i,j}^t\right)$	Betriebsverlust	$\max\left(0, \sum_{j=1}^{3} w_{i,j}^t - \sum_{j=4}^{9} w_{i,j}^t\right)$
Aufwandtotal	$\sum_{j=1}^{3} w_{i,j}^t + \max\left(0, \sum_{j=4}^{9} w_{i,j}^t - \sum_{j=1}^{3} w_{i,j}^t\right)$	Ertragstotal	$\sum_{j=4}^{9} w_{i,j}^t + \max\left(0, \sum_{j=1}^{3} w_{i,j}^t - \sum_{j=4}^{9} w_{i,j}^t\right)$

A.2 Kontrollregeln

Tabelle 25 Mehrere Totale, die eingesetzt werden sollen, müssen gleich sein. Die Spalte 'Inkons.' beinhaltet die Anzahl Institutionen Einheiten, welche die Regel vor den Einsetzungen verletzen. Die Prozentangaben beziehen sich auf das Total der 11646 Jahres Einheiten, der Jahre 1998 bis 2002. Wenn die Regel nicht programmiert wurde, steht in den Kolonnen 'Inkons. vor Imp.' und '[%]' ein 'x'. Fehlende Werte in 'Nr. in SM3.0' bedeuten, dass der Test nicht in die Software ist.

Test Nr.	Nr. in SM	Beschreibung	Inkons.	[%]
1	2a 25	Das Total der Anzahl Personen im engeren Betrieb muss mindestens so gross sein, wie die Anzahl Stellen im engeren Betrieb.	12	0.10
2		Das Total der Anzahl Stellen im engeren Betrieb muss gleich der Summe der Anzahl Stellen nach Ausbildungsniveau im engeren Betrieb sein.	201	1.73
3		Das Gesamttotal der Anzahl Personen, muss mindestens so gross sein, wie das Gesamttotal der Anzahl Stellen.	205	1.76
4	2b6 32	Das Total engerer Betrieb der Anzahl Stellen nach Ausbildung (2.a) muss gleich dem Total der Stellen nach Ausbildung in 2.b sein.	517	4.44
5		Das Total der Anzahl Personen im engeren Betrieb muss mindestens so gross sein, wie die Summe der Anzahl Personen in 2.b.	466	4.00
6	2.b.* 29	Sämtliche Totale der Anzahl Personen in 2.b müssen mindestens so gross sein, wie die entsprechenden Totale der Anzahl Stellen.	25	0.21
7	3.2.1.4 50	Das Gesamttotal der Anzahl Bewohner (3.1) muss gleich sein, wie das Gesamttotal der Anzahl betreuter Bewohner in 3.2	x	x
8	3.3 51	Das Gesamttotal der Anzahl Bewohner (3.1) muss gleich dem Gesamttotal der Anzahl Bewohner in 3.3 sein.	81	0.70
9	3.3 52	Das Gesamttotal der Anzahl betreuter Externer (3.2) muss gleich dem Gesamttotal der Anzahl Externen in 3.3 sein.	x	x
10	3.4 53	Das Gesamttotal der Anzahl Bewohner (3.1) muss gleich der Summe des Gesamttotals der männlichen Bewohner und des Gesamttotals der weiblichen Bewohner in 3.4. sein.	88	0.76
11	3.4 54	Das Gesamttotal der Anzahl betreuter Externer (3.2) muss gleich der Summe des Gesamttotals der männlichen Externen und der weiblichen Externen in 3.4. sein.	x	x
12		In der Fragebogen-Maske: Das Total in 4.1 muss gleich dem Total in 4.3 sein.	469	4.03
13		Die Summe der Anzahl betreuter Bewohner und der Anzahl betreuter Externer unter 3.2 muss gleich der Anzahl Betreuter unter 4.1 sein.	x	x
14	5.4 65	Das Total des Betriebsaufwands muss gleich dem Total des Betriebsertrags sein (ausgeglichene Bilanzsumme).	23	0.20

Tabelle 26 Konsistenzkontrollen unter den Variablen. Die Spalte 'Flag' beinhaltet die Anzahl Institutionen Einheiten, welche die Regel vor den Einsetzungen verletzen. Die Prozentangaben beziehen sich auf das Total der 11646 Jahres Einheiten, welche über den Zeitrahmen von 1998 bis 2002 vorhanden sein müssen. Wenn die Regel nicht programmiert wurde, steht in den Kolonnen 'Flag' und '[%]' ein 'x'. 'U.bringungsart' steht für 'Unterbringungsart'.

Test Nr.	Nr. in SM	Beschreibung	Flag	[%]	Bemerkung
15	2.a 28	Eine Angabe im med. pfleg. Dienst unter 2.a verlangt ein Angebot solcher Leistungen für Bewohner oder/und Externe in 1.1.1, und umgekehrt, wenn keine Angabe vorhanden ist, dass diese Dienste von Externen erbracht werden.	524	4.50	354 mit Antwort in 1.1.1
16	2.a 28	Eine Angabe in Beratung und soz. Dienste unter 2.a verlangt ein Angebot solcher Leistungen für Bewohner oder/und Externe in 1.1.2, und umgekehrt, wenn keine Angabe vorhanden ist, dass diese Dienste von Externen erbracht werden.	1544	13.26	1498 mit Antwort in 1.1.2
17	2.a 28	Eine Angabe in Körperpflege unter 2.a verlangt ein Angebot solcher Leistungen für Bewohner oder/und Externe in 1.1.3, und umgekehrt, wenn keine Angabe vorhanden ist, dass diese Dienste von Externen erbracht werden.	1081	9.28	1026 mit Antwort in 1.1.3
18	2.a 28	Eine Angabe therapeutischer Dienste unter 2.a verlangt ein Angebot solcher Leistungen für Bewohner oder/und Externe in 1.1.4, und umgekehrt, wenn keine Angabe vorhanden ist, dass diese Dienste von Externen erbracht werden.	1022	8.78	922 mit Antwort in 1.1.4
19	2.a 28	Eine Angabe in Infrastruktur unter 2.a verlangt ein Angebot solcher Leistungen für Bewohner oder/und Externe in 1.1.5, und umgekehrt, wenn keine Angabe vorhanden ist, dass diese Dienste von Externen erbracht werden.	410	3.52	353 mit Antwort in 1.1.5
20	2.b.1 31	Eine Angabe im med. Dienst unter 1.1.1 verlangt eine Angabe zu den ausgebildeten Ärzten in 2.b.1.	0	0.00	
21	2.b.1 30	Eine Angabe zur Anzahl Stellen in 2.b verlangt eine Angabe zur Anzahl Personen und umgekehrt.	39	0.33	63 Berufs-Codes
22	3.1.1.1 35 + 3.1.1.4 41	Anzahl Betten und Aufenthaltstage innerhalb der Fragebogenabschnitte in 3.1 müssen beide ausgefüllt oder beide gleichzeitig leer sein.	180	1.55	239 Codes der U.bringungsart
23	3.1.1.4 41	Mindestens ein Fragebogenabschnitt unter 3.1 muss ausgefüllt sein.	46	0.40	
25		Eine Angabe in 3.1.1.1 benötigt eine Antwort in 4.1 und umgekehrt.	4420	37.95	2353 Antwort in 4.1, 2067 in 3.1.1.1.
26		Eine Angabe in 3.1.1.3 benötigt eine Antwort in 4.3 und umgekehrt.	6805	58.43	6065 Antwort in 4.3, 740 in 3.1.1.3.
27	3.2.1.* 48	Eine Angabe von Betreuungstagen für Bewohner unter 3.2 zieht eine Angabe zu der Anzahl Plätze unter der gleichen Rubrik von 3.2 nach sich.	x	x	
28	3.2.1.* 49	Die Anzahl betreuter Bewohner unter 3.2 ist höchstens so gross wie die Anzahl Plätze der gleichen Rubrik von 3.2.	x	x	
29	3.2.1.* 44	Eine Angabe zur Anzahl betreuter Externen unter 3.2 zieht eine Angabe zur Anzahl Betreuungstage für Externe unter der gleichen Rubrik nach sich.	x	x	

Tabelle 27 Konsistenztests unter den Variablen und Konsistenztests der Veränderung bezüglich dem vorhergehenden Jahr. Die Spalte 'Inkons. vor Imp.' beinhaltet die Anzahl Institutionen Einheiten, welche die Regel vor den Einsetzungen verletzen. Die Prozentangaben beziehen sich auf das Total der 11646 Jahres Einheiten, welche über den Zeitrahmen von 1998 bis 2002 vorhanden sein müssen. Wenn die Regel nicht programmiert wurde, steht in den Kolonnen 'Inkons. vor Imp.' und '[%]' ein 'x'. Steht unter 'Test Nr. in SM3.0' nichts, so ist der Test nicht in die Software integriert.

Test Nr.	Test Nr. in SM	Beschreibung	Inkons. vor Imp.	[%]	Bemerkung
30	4.1 68	Eine Angabe unter 4.1 zieht eine Angabe unter 1.2 nach sich und umgekehrt.	4017	34.49	
31	5.1 56	'Ist der Betrieb in 0.1. als SM-Institution mit IV-Beiträgen deklariert, so zieht dies eine Antwort in 5.1 unter 'Beiträge seitens der IV' nach sich.'	916	7.87	102 davon IV
32	5.1 57	Ein Personalaufwand unter 5.1 muss vorhanden sein.	168	1.44	
33	5.1 58	Anderer Betriebsaufwand unter 5.1 muss vorhanden sein.	168	1.44	
34	5.1 60	Der Betriebsertrag muss zwischen 0 und 99999 sein. (In Tausend Franken.)	0	0.00	
35	5.1 61	Die Angabe Betriebsbeitragsgarantie=Ja (in 0.5) zieht eine Angabe unter Beiträge seitens der AHV, der IV, der Kantone oder der Gemeinden nach sich, und umgekehrt.	x	x	
36		Typlo99 $\in$ {S111, S112, S122} verlangt eine Antwort in 3.1.1.1	153	1.31	

Tabelle 28 Konsistenzregeln bezüglich dem Gesamttotal Bewohner. 'Inkons.' steht für 'Inkonsistenzen'.

Frage	Verletzte Regel	Anzahl Inkons.	[%] Inkons.
3.1	Gesamttotal Bewohner 3.1 < Gesamttotal Bewohner 3.3 und 3.4	4	0.18
3.3	Gesamttotal Bewohner 3.3 < Gesamttotal Bewohner 3.1	35	0.30
3.4	Gesamttotal Bewohner 3.4 < Gesamttotal Bewohner 3.1	35	0.30

A.3 Einfluss der Einsetzungen

A.3.1 Ausreisseranalysen

Huber-type skipped mean

Tabelle 29 Anzahl Jahres Einheiten (Institutionen) mit mindestens einem Ausreisserwert vor der NNE ('n_outl_org') und nach der NNE ('n_outl_nne'), die aufgrund einer *Huber-type skipped mean* Verwerfungsregel mit Parameter κ entdeckt wurden. '[%]' bezieht sich auf die Jahres Einheiten ohne deren Antwort Ausfälle.

	n_outl_org		**n_outl_nne**		**Anstieg**	
κ	**Anzahl**	**[%]**	**Anzahl**	**[%]**	**Anzahl**	**[%]**
3	1848	16.40	1979	16.99	131	0.59
4	1057	9.38	1148	9.86	91	0.48
5	632	5.61	734	6.30	102	0.69
6	414	3.67	465	3.99	51	0.32
7	298	2.64	320	2.75	22	0.10
10	133	1.18	141	1.21	8	0.03
20	33	0.29	37	0.32	4	0.02
30	16	0.14	18	0.15	2	0.01
40	13	0.12	10	0.09	-3	-0.03
50	6	0.05	6	0.05	0	0.00

Hidiroglou&Berthelot

Tabelle 30 Anzahl Jahres Einheiten vor und nach der Einsetzung (NNE) mit mindestens einer Variablenentwicklung, die als Ausreisser entdeckt wurde. Es wurden nur die Variablen des ersten Einsetzungsschritts aufgeführt. Parameter der *Hidiroglou&Berthelot* Ausreisser Erkennung: $u = 0.8$, $A = 0.05$, $L = 50$.

	Jahr							
	1998-1999		1999-2000		2000-2001		2001-2002	
Variable	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher
Total Personen 2a	8	74	4	18	11	18	14	20
Vzae univ 2a	1	3	.	.	.	.	.	.
Vzae beruf 2a	7	27	8	13	14	17	7	9
Vzae hoehere 2a	5	23	8	12	11	13	7	9
Vzae univ 2b	1	1	.	.	.	.	.	.
Vzae hoehere 2b	6	23	8	14	11	14	6	7
Vzae beruf 2b	9	29	5	13	15	22	6	9
Besch. hoeh 2b	2	14	1	5	10	10	2	2
Besch. beruf 2b	5	20	3	11	10	15	4	5
Tot. Bew. (kt) 3.3	6	49	19	6	10	5	6	2
Tot. Ext. (kt) 3.3	1	5	5	3	2	2	2	.
Tot. Bet. (3.4)	2	4	3	2	2	2	1	.
Tot. Bew (3.4)	7	49	21	6	11	5	6	2
BETAFW 5.1	17	76	9	40	13	22	11	17
PERSONAL 5.1	23	111	14	51	14	24	9	15

A.3.2 Standardtabellen SM 2002

Abbildung 2 Anstieg und relativer Anstieg der Vollzeitäquivalenten nach Kanton, der durch die Einsetzungen hervorgerufen wurde. Typ1: H. für pflegebedürftige Betagte, Typ2: Alters- und Pflegeh., Typ3: Inst. für Behinderte und Suchtkranke, Typ4: Andere.

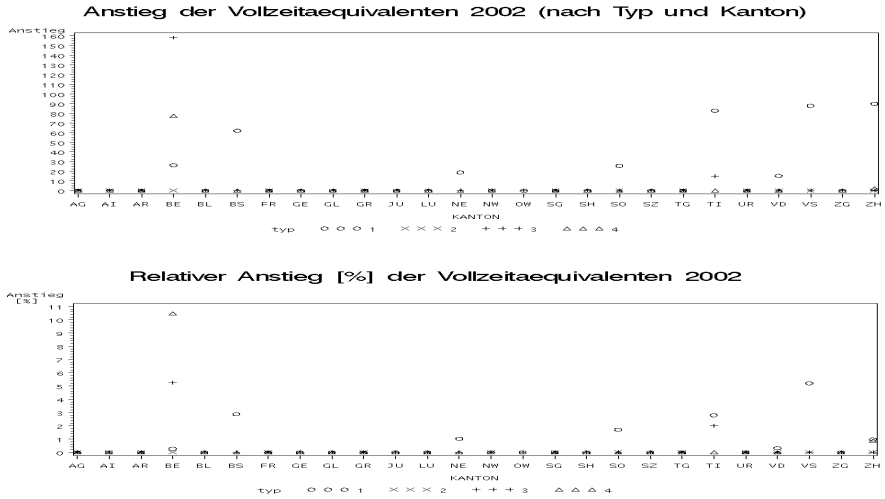


Abbildung 3 Anstieg und relativer Anstieg der Betreuten nach Kanton, der durch die Einsetzungen hervorgerufen wurde. Typ1: H. für pflegebedürftige Betagte, Typ2: Alters- und Pflegeh., Typ3: Inst. für Behinderte und Suchtkranke, Typ4: Andere.

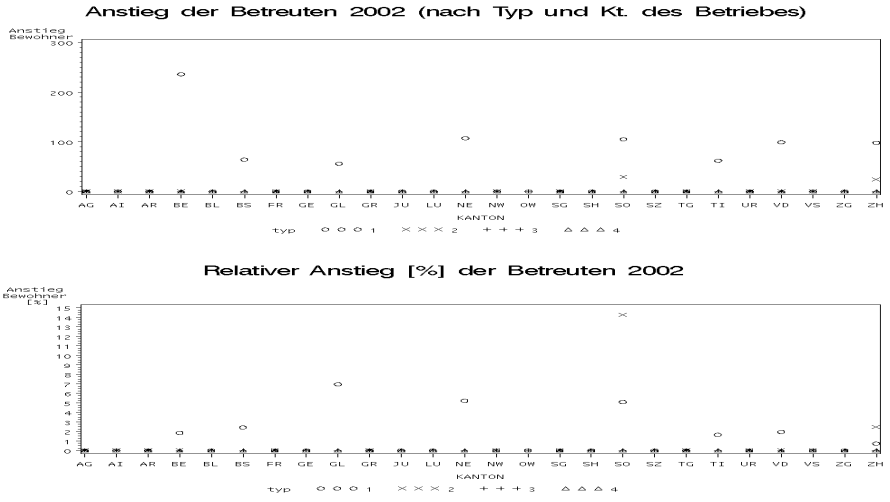


Abbildung 4 Anstieg und relativer Anstieg der Betreuten nach Kanton, der durch die Einsetzungen hervorgerufen wurde. Typ1: H. für pflegebedürftige Betagte, Typ2: Alters- und Pflegeh., Typ3: Inst. für Behinderte und Suchtkranke, Typ4: Andere.

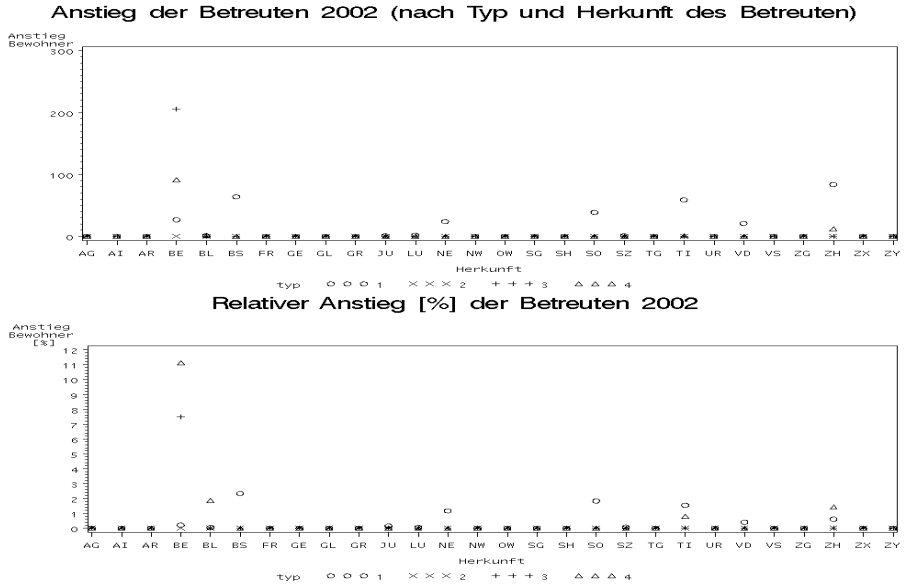


Abbildung 5 Anstieg und relativer Anstieg der Unterbringungen nach Kanton, der durch die Einsetzungen hervorgerufen wurde. Typ1: H. für pflegebedürftige Betagte, Typ2: Alters- und Pflegeh., Typ3: Inst. für Behinderte und Suchtkranke, Typ4: Andere.

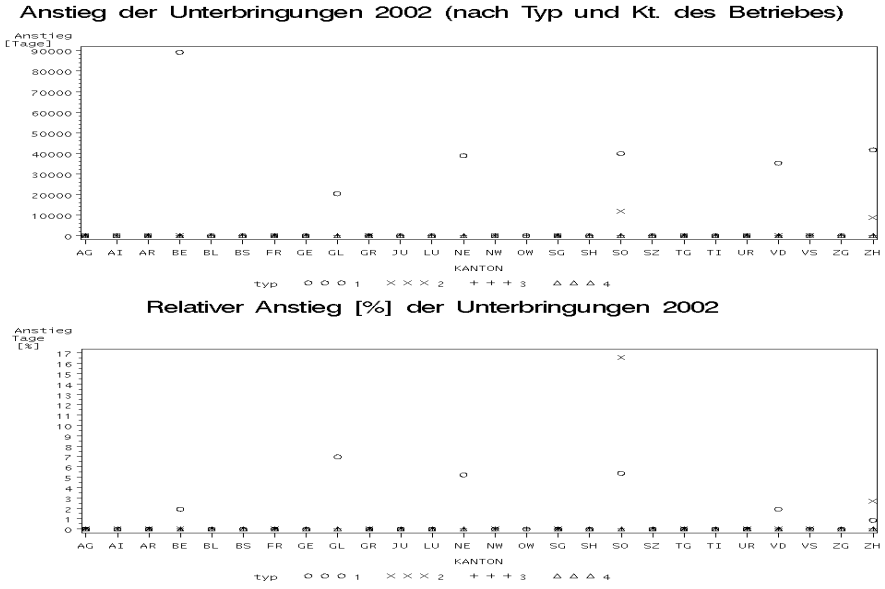
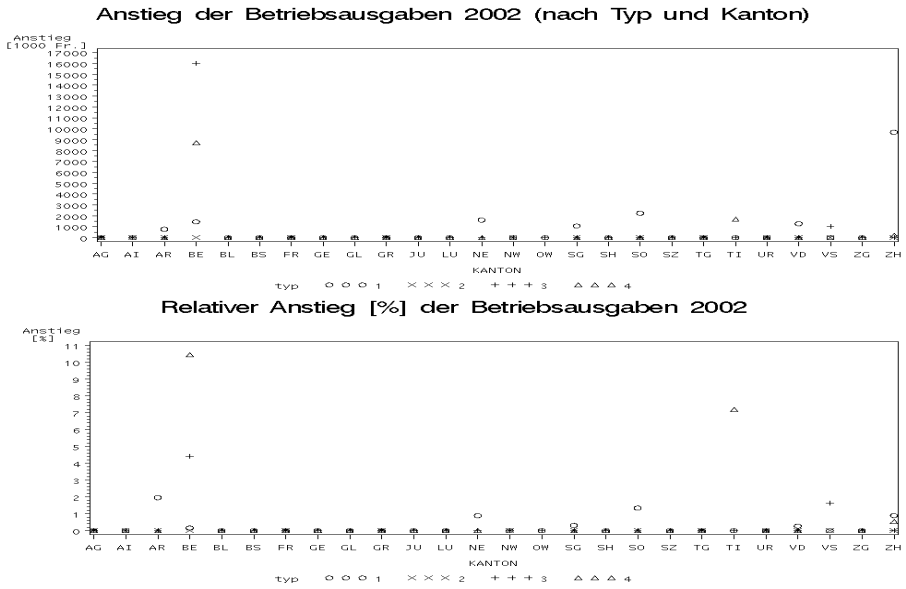
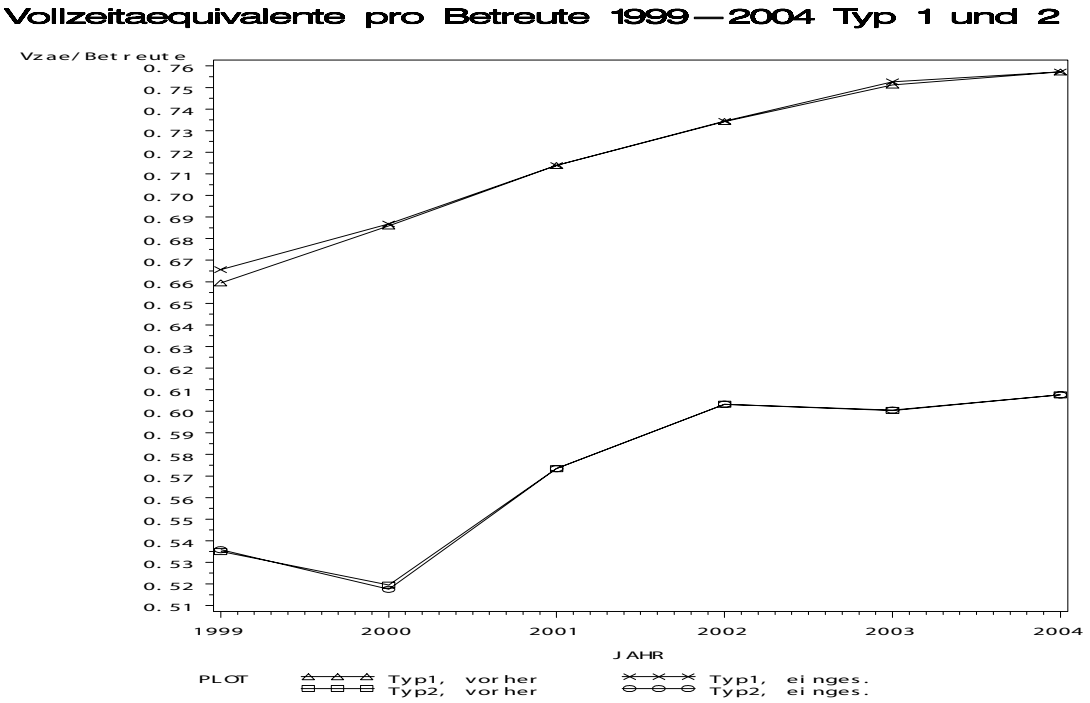


Abbildung 6 Anstieg und relativer Anstieg des Betriebsaufwands nach Kanton, der durch die Einsetzungen hervorgerufen wurde. Typ1: H. für pflegebedürftige Betagte, Typ2: Alters- und Pflegeh., Typ3: Inst. für Behinderte und Suchtkranke, Typ4: Andere.



A.3.3 Entwicklung von Indikatoren

Abbildung 7 Entwicklung der Anzahl Vollzeitäquivalenter pro Betreute nach Typ 1999-2004.



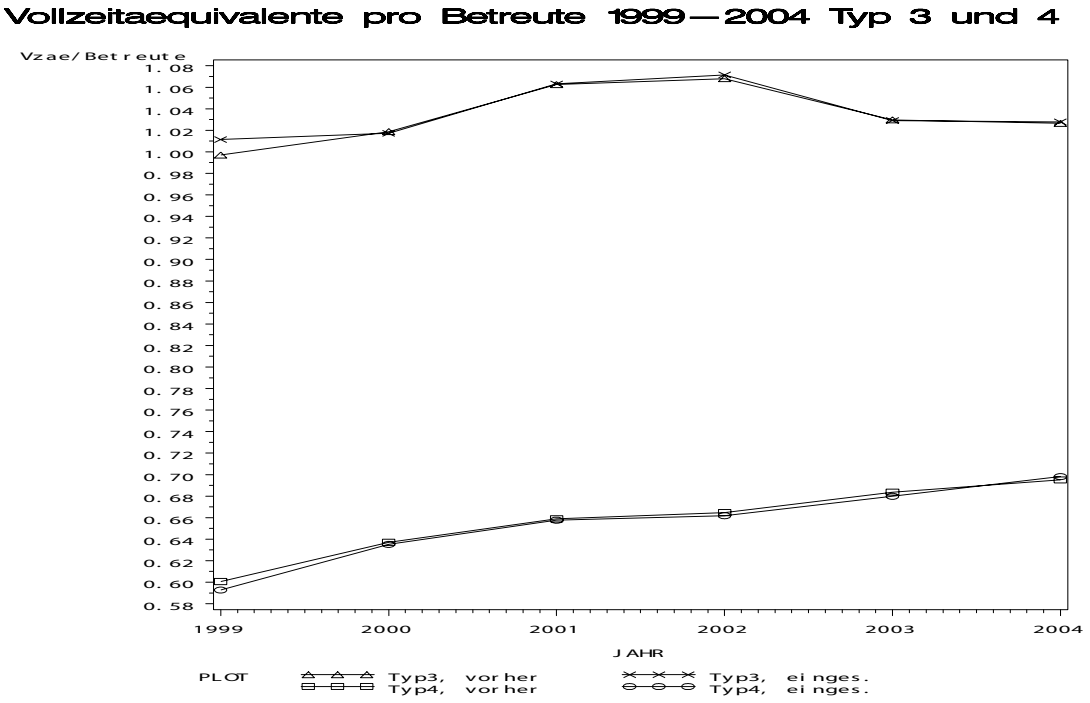
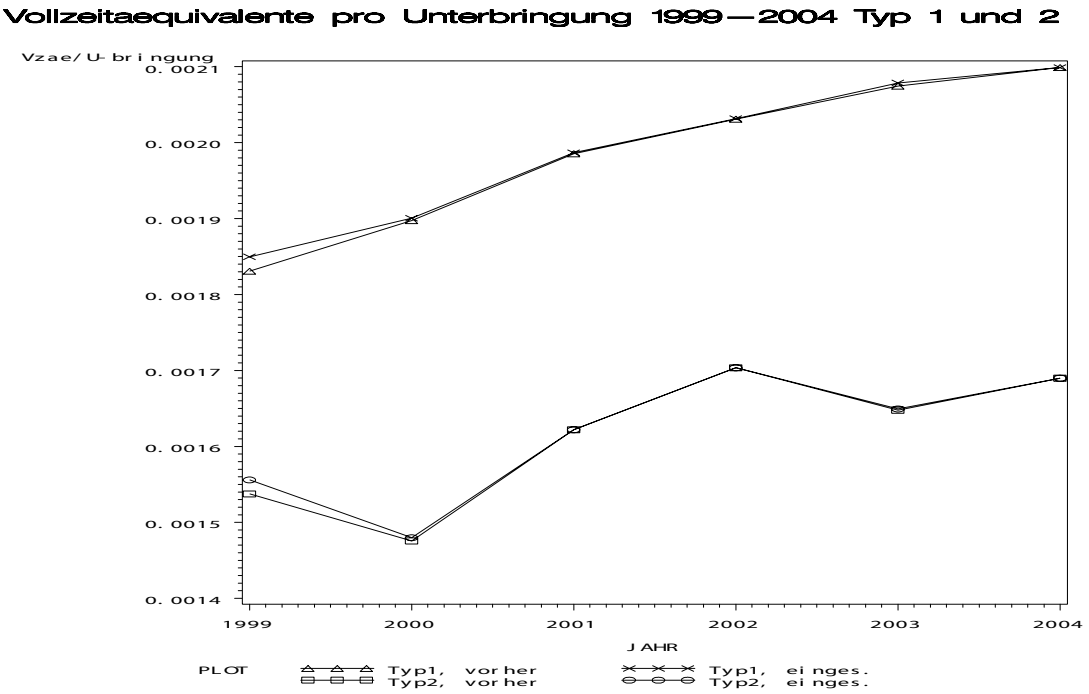


Abbildung 8 Entwicklung der Anzahl Vollzeitäquivalenter pro Unterbringung nach Typ 1999-2004.



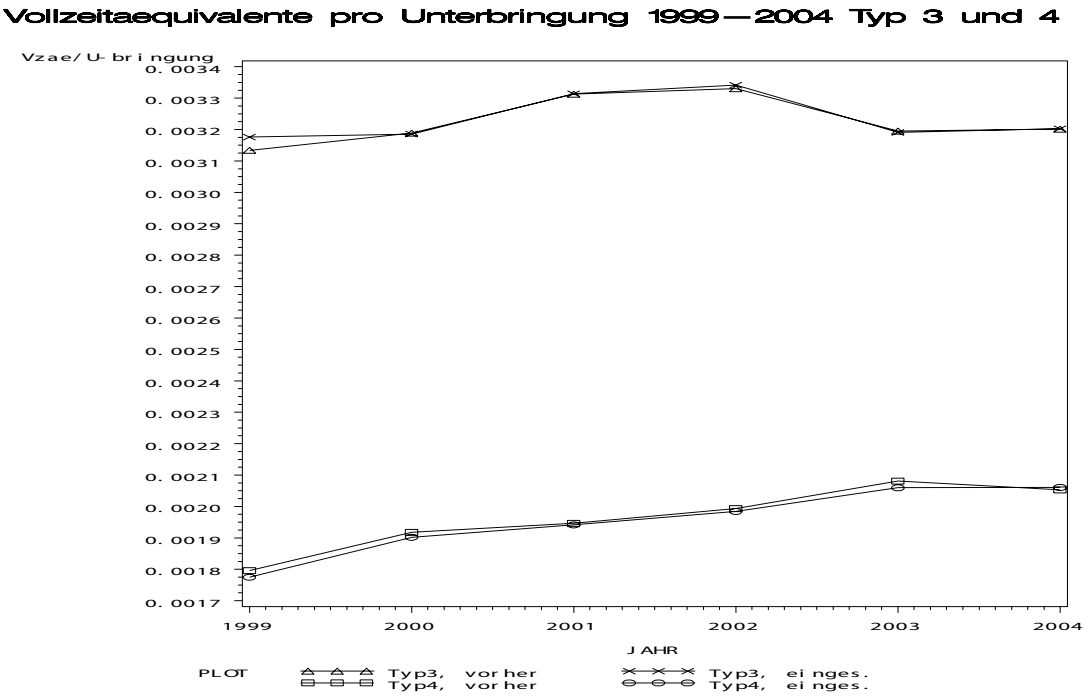
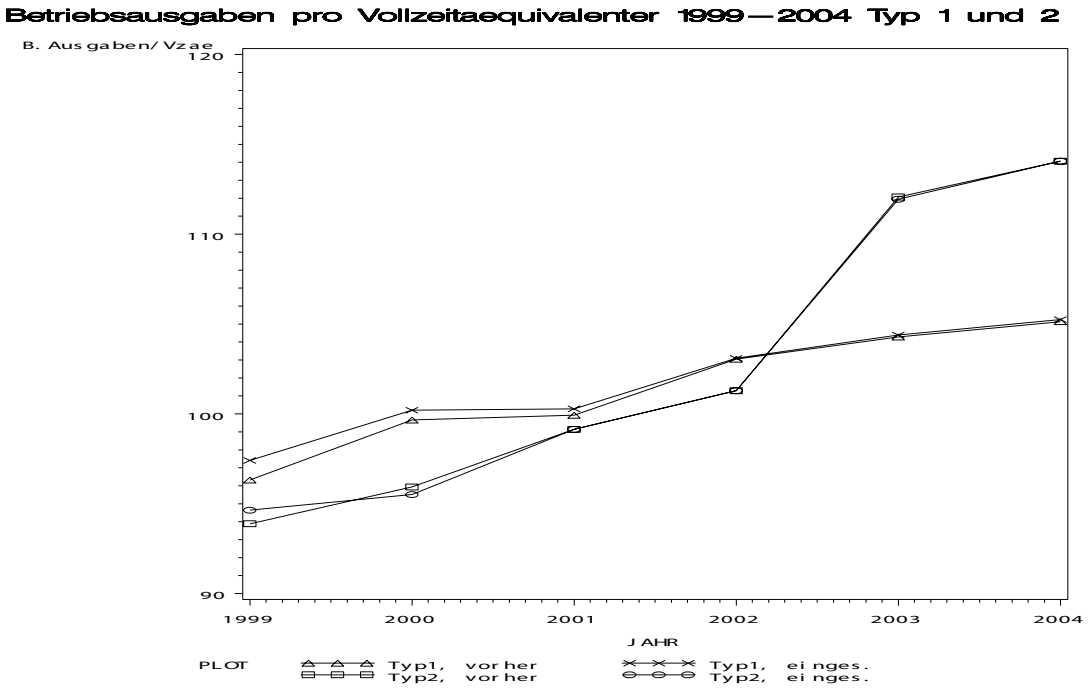


Abbildung 9 Entwicklung der Betriebsausgaben pro Vollzeitäquivalenter nach Typ 1999-2004.



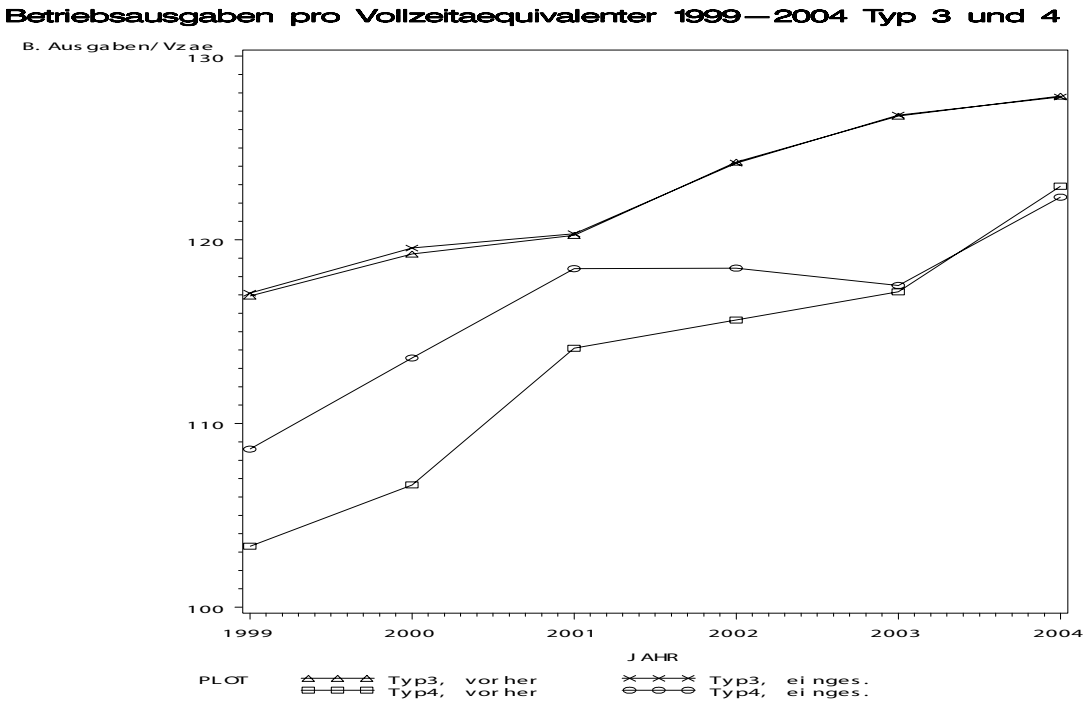
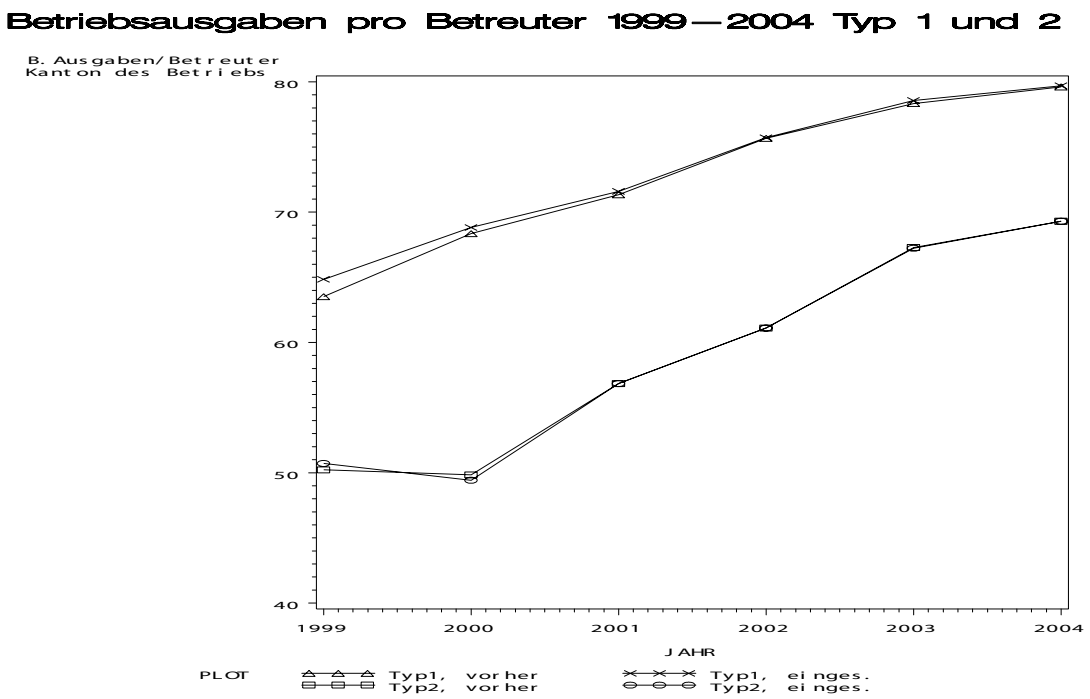


Abbildung 10 Entwicklung der Betriebsausgaben pro Betreuter nach Typ 1999-2004.



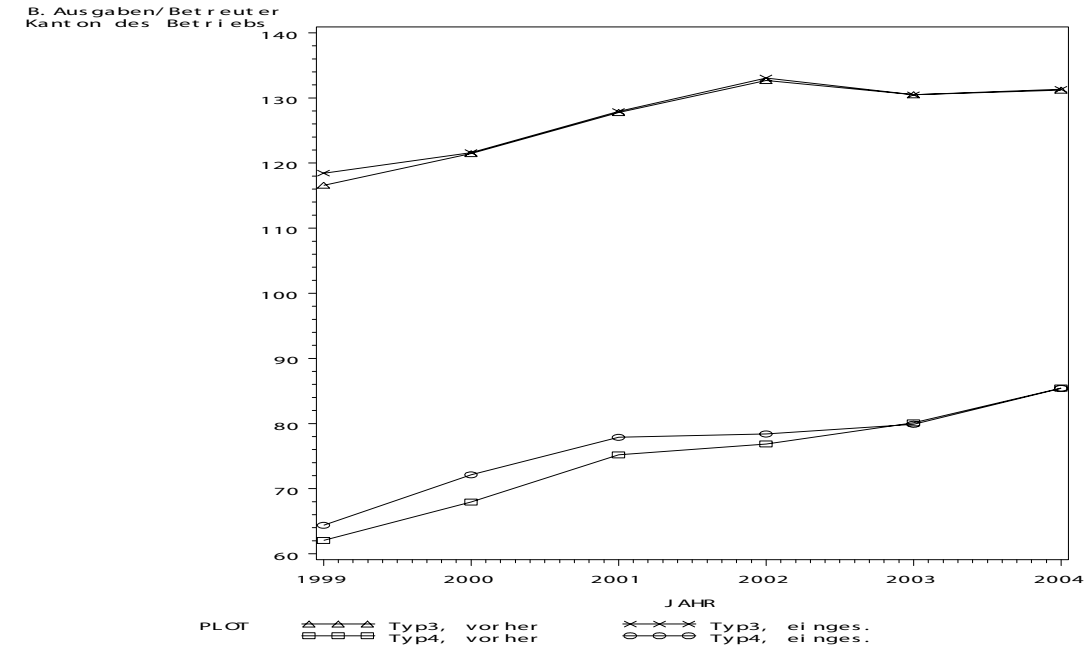
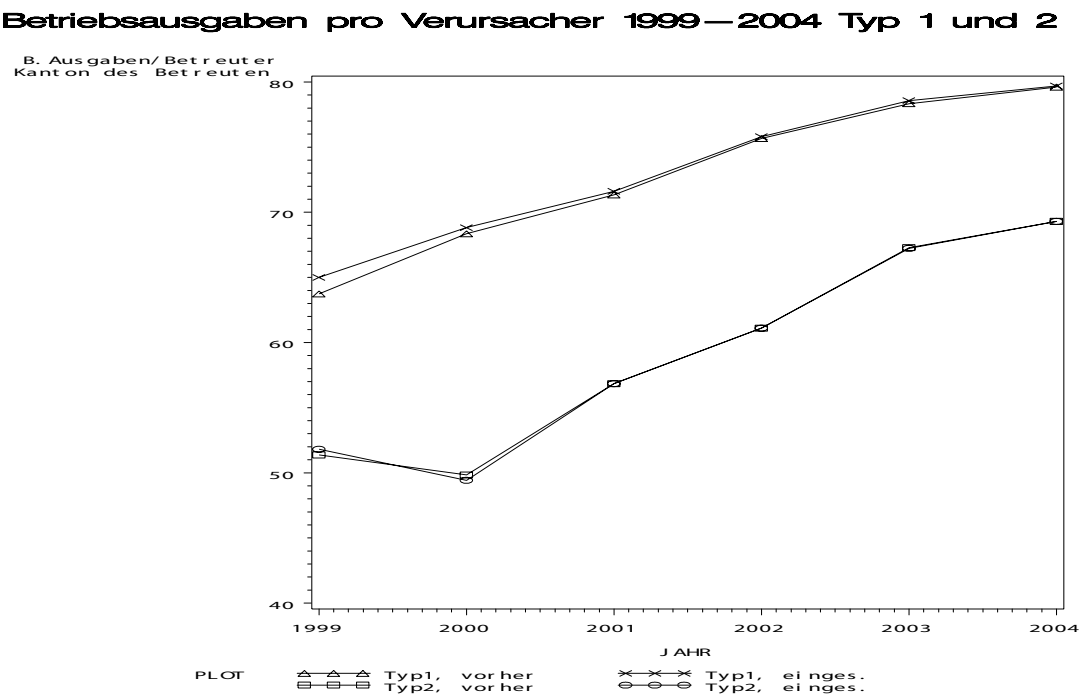


Abbildung 11 Entwicklung der Betriebsausgaben pro Verursacher nach Typ 1999-2004.



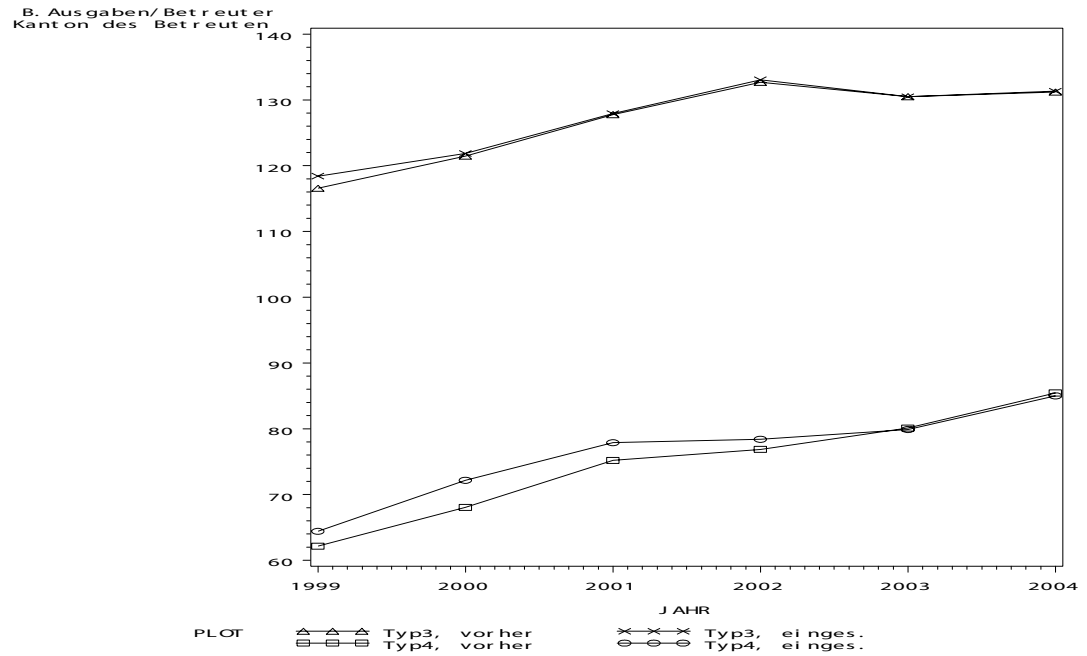
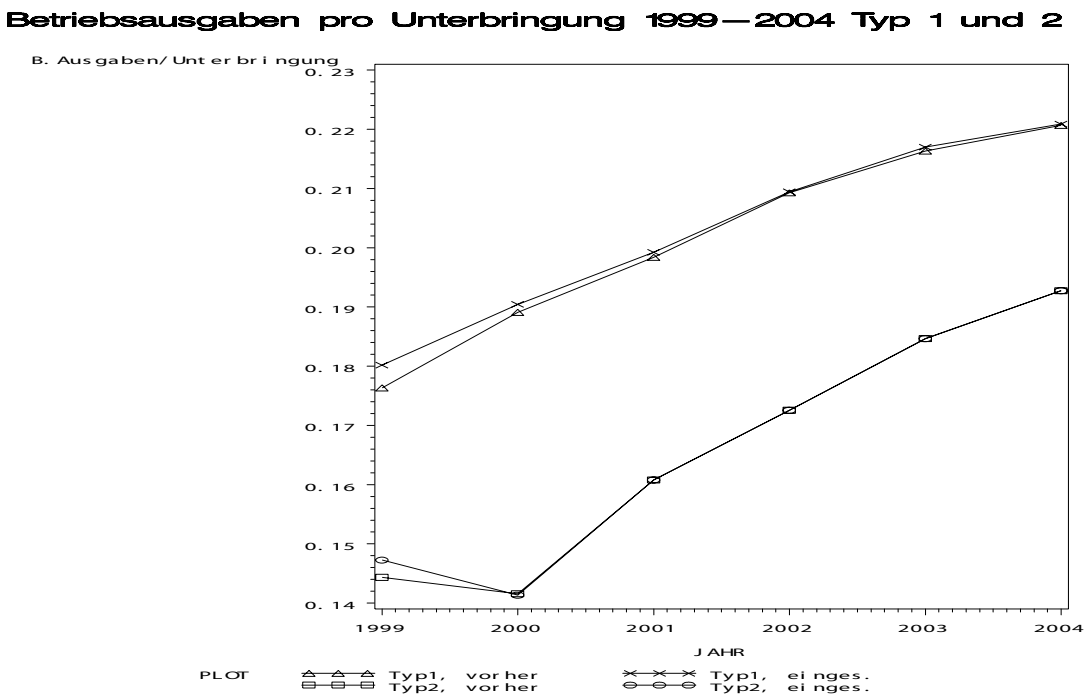
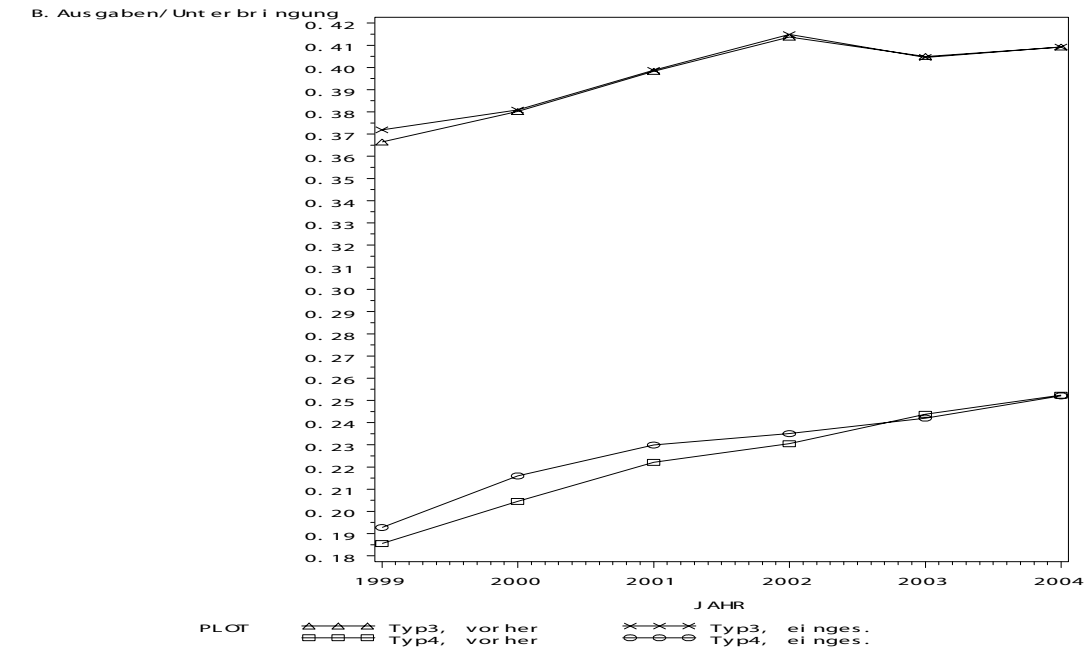


Abbildung 12 Entwicklung der Betriebsausgaben pro Unterbringung nach Typ 1999-2004.





B Krankenhausstatistik

B.1 Typologie und Variablen

Tabelle 31 Typologie der Krankenhäuser

Typ	Bezeichnung
K 11	Zentrumsversorgung
K 12	Grundversorgung
K 21	Psychiatrische Kliniken
K 22	Rehabilitationskliniken
K 23	Andere Spezialkliniken

Tabelle 32 Einsetzungsvariablen der KS.

Fragebogenabschnitt	Variable	Nr.
0.4 Allgemeine Angaben	Rechtsform	0.4.V01
	Wirtschaftsstatus	0.5
2.1 Beschäftigung	Beschäftigungsgrad	2.1.V04
	durchschn. Beschäftigungsgrad	2.1.V05
	Bezahlte Diensttage	2.1.V06
	Konto Besoldungen	2.1.V12
	Geschlecht	2.1.V14
3.1 Betten	Betriebstage von Betten für Hospitalisierte	3.1.1.V01
	Betriebstage von Betten für gesunde Säuglinge	3.1.1.V02
	Betriebstage von Betten für Gesunde	3.1.1.V03
	Betriebstage von Betten für teilstationäre Hosp.	3.1.1.V04
3.2 Hospitalisierte	Pflegetage insgesamt	3.2.1.1.V17
	Bestand der Hospitalisierten am 31.12.	3.2.2.1.V17
	Bestand der privat und halbprivat Hosp. am 31.12.	3.2.2.2.V17
	Aufnahmen während des Jahres	3.2.3.1.V17
	Austritte während des Jahres	3.2.5.1.V17
3.3 Teilstationäre Patienten	Teilstat. Behandlungsfälle	3.3.1.1.V15
5.1 Betriebsrechnung	Besoldung Ärzte und Akademiker	5.1.1.1.V01
	Besoldung Pflegepersonal im Pflegebereich	5.1.1.1.V02
	Besoldung Personal anderer med. Fachbereich	5.1.1.1.V03
	Besoldung Verwaltungspersonal	5.1.1.1.V04
	Besoldung Ökonomie-, Transport-, Hausdienstpers.	5.1.1.1.V05
	Besoldung Personal tech. Betriebe	5.1.1.1.V06
	Sozialleistungen	5.1.1.1.V07
	Arzthonorare	5.1.1.1.V08
	Personalnebenkosten	5.1.1.1.V09
	Total Besoldungen und Sozialleistungen	5.1.1.1.V10
	Medizinischer Bedarf	5.1.1.2.V01
	Aufwand für Anlagenutzung	5.1.1.2.V05
	Zinsaufwand und Umlaufvermögen	5.1.1.2.V07
	Total Übriger Betriebsaufwand	5.1.1.2.V10
	Übriger Betriebsaufwand	5.1.1.2.V10 - 5.1.1.2.V01 - 5.1.1.2.V05 - 5.1.1.2.V07
	Betriebsertrag Poli-, Tages- und Nachtkliniken	5.1.2.1.V05
	Total Betriebsertrag aus Hosp. und Pflege	5.1.2.1.V06
	Übriger Betriebsertrag aus Hosp. und Pflege	5.1.2.1.V06 - 5.1.2.1.V05
	Total Betriebsertrag aus ambulanter Pflege	5.1.3.V05
	Total Betriebsertrag	5.1.2.V01
5.2 Deckung des Betriebsverlustes	Ungedeckter Verlust	5.2.V06
	Total Deckung des Betriebsverlustes	5.2.V06

B.2 Kontrollregeln

Tabelle 33 Konsistenztests unter den Variablen Die Spalte 'Inkons. vor Imp.' beinhaltet die Anzahl Institutionen Einheiten, welche die Regel vor den Einsetzungen verletzen. Die Prozentangaben beziehen sich auf das Total der 1898 Jahres Einheiten, welche über den Zeitrahmen von 1998 bis 2002 vorhanden sein müssen.

Nr. in SES	Beschreibung	Inkons.	[%]
1343	Es gibt mehr Pflegetage als Betriebstage der Betten	52	2.74
1344	Es gibt mehr Pflegetage als Betriebstage der Betten für Neugeborene	7	0.37
1345	Es gibt mehr Aufenthaltstage für Gesunde als Betriebstage der Betten für Gesunde	11	0.58
1346	Die Gesamtzahl der Personen nach Wohnsitz unter 'privat und halbprivat' entspricht nicht der Anzahl der hospitalisierten Personen	31	1.63
1347	Die Gesamtzahl der Entlassungen nach Wohnsitz unter 'privat und halbprivat' entspricht nicht der Anzahl Entlassungen für das Jahr	24	1.26
1348	Die Anzahl der teilstationären Fälle entspricht nicht der Gesamtzahl teilstationären Fälle nach Wohnsitz	10	0.53
1349	Die Anzahl der teilstationären Fälle unter 'privat und halbprivat' entspricht nicht der Gesamtzahl teilstationären Fälle nach Wohnsitz	5	0.26
1358	Das Total der Hospitalisierungen nach Wohnsitz muss dem Total der Hospitalisierungstage während des Jahres ohne Neugeborene und Nichtkranke entsprechen	24	1.26
1359	Das Total der teilstationären Behandlungsfälle nach Wohnsitz muss dem Total der teilstationären Behandlungsfälle während des Jahres entsprechen	10	0.53
1360	Das Total der Austritte nach Wohnsitz muss dem Total der Austritte während des Jahres entsprechen	26	1.37
1368	Es kann nicht mehr Personen am 31.12 geben als Pflegetage während des Jahres.	5	0.26
1369	Es kann nicht mehr Eintritte als Pflegetage während eines Jahres geben.	6	0.32
1370	Es kann nicht mehr Austritte als Pflegetage während eines Jahres geben.	6	0.32
1371	Das Total der Eintritte durch interne Überweisungen muss gleich dem Total der Austritte durch interne Überweisungen sein.	114	6.01
1401-1442	Anzahl Pflegetage < Anzahl Personen Personen am 31.12./Eintritte/Austritte	31	1.63
1504	Personalaufwand $\neq$ Summe der Aufwände der einzelene Personalkonti	8	0.42
1511	Ertrag aus Pflege > Ertragssumme	7	0.37

B.3 Einsetzungsklassen

Abbildung 13 Verteilung der logarithmierten Mahalanobis Distanz `dl_maha`.

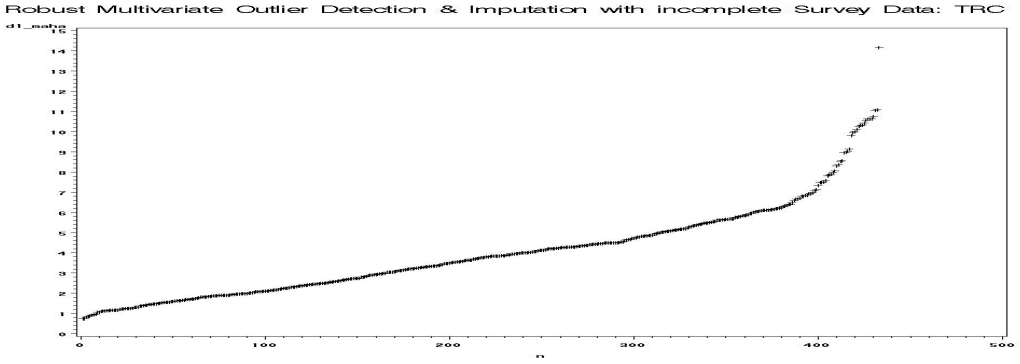


Abbildung 14 Verteilung der logarithmierten Mahalanobis Distanz `dl_maha` der ersten Einsetzungsklasse.

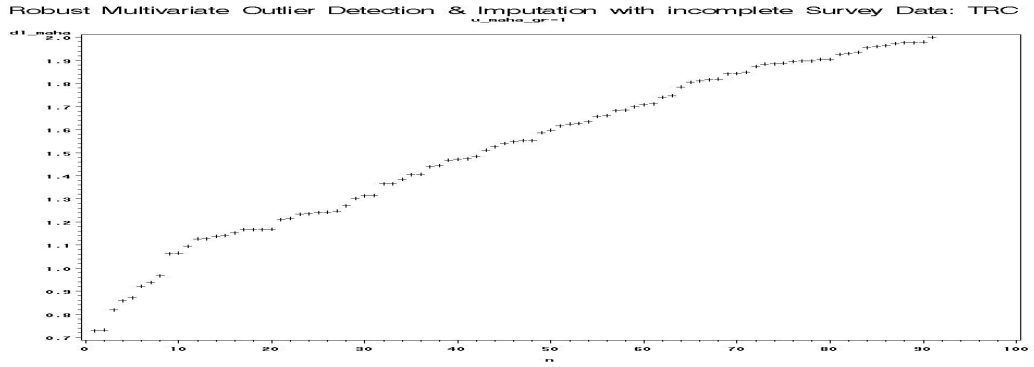


Abbildung 15 Verteilung der logarithmierten Mahalanobis Distanz `dl_maha` der zweiten Einsetzungsklasse.

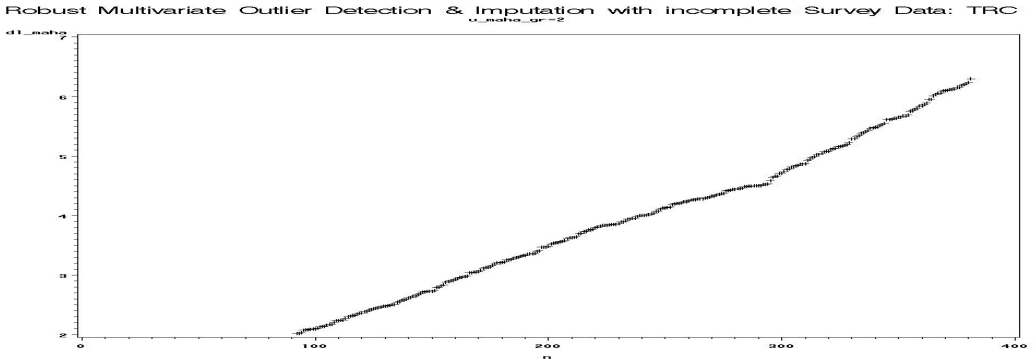


Abbildung 16 Verteilung der logarithmierten Mahalanobis Distanz `dl_maha` der dritten Einsetzungsklasse.

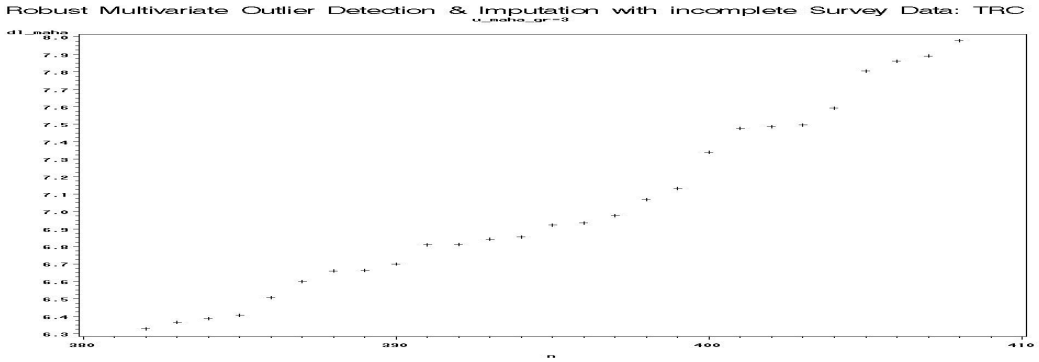
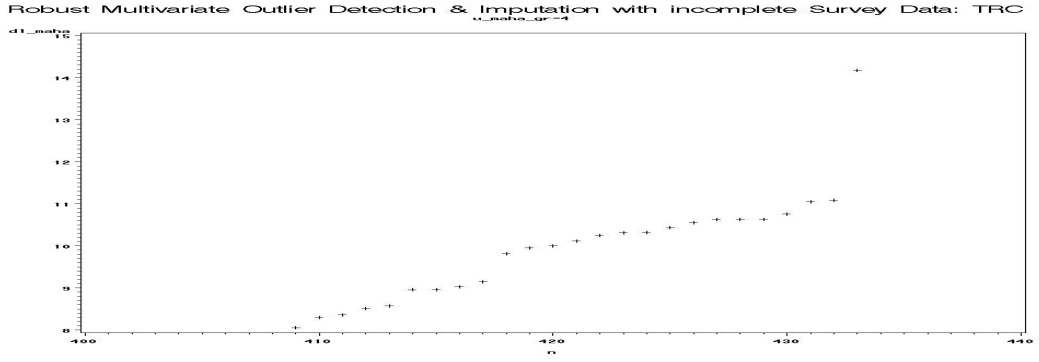


Abbildung 17 Verteilung der logarithmierten Mahalanobis Distanz `dl_maha` der vierten Einsetzungsklasse.



B.4 Fehlende Werte und Spenderpool

Tabelle 34 Anzahl fehlende Werte bei den 1898 Jahres Einheiten der KS. Aufwand wurde mit 'Aufw.' abgekürzt, Hospitalisiert mit 'Hosp.', Betriebsertrag mit 'BE', Vollzeitäquivalente mit 'VZ', Besoldungskonto mit 'Kto', Männer mit 'M', Frauen mit 'F', Durchschnittliche VZ mit $\overline{VZ}$ und Diensttage mit 'Tg'. Deckung2 steht für das Total der Deckung des Betriebsverlustes ohne den ungedeckten Verlust.

Variable	fehlend
Saeuglinge 3.1.1	48
Gesunde 3.1.1	52
Hospitalisierte 3.1.1	51
Teilstätionäre Hospitalisierung 3.1.1	95
Pflegetage 3.2.1	51
Bestand 3.2.2	240
Bestand PHP 3.2.2	240
Aufnahmen 3.2.3	71
Austritte 3.2.5	54
Teilstationäre 3.3	63
Personalaufw. 5.1.1	77
Übriger Betriebsaufw. 5.1.1	76
BE aus Hosp. & Pflege 5.1.2	113
BE 5.1.2	104

Variable	fehlend
Deckung2	67
VZ	137
VZ Kto30	153
VZ Kto31	153
VZ Kto32	153
VZ Kto33	153
VZ Kto34	153
VZ Kto35	153
VZ M	144
VZ F	144
$\overline{VZ}$	164
$\overline{VZ}$ Kto30	178
$\overline{VZ}$ Kto31	178
$\overline{VZ}$ Kto32	178

Variable	fehlend
$\overline{VZ}$ Kto33	178
$\overline{VZ}$ Kto34	178
$\overline{VZ}$ Kto35	178
$\overline{VZ}$ M	169
$\overline{VZ}$ F	169
Tg	215
Tg Kto30	227
Tg Kto31	227
Tg Kto32	227
Tg Kto33	227
Tg Kto34	227
Tg Kto35	227
Tg M	220
Tg F	220

Tabelle 35 Spendercharakteristik nach Krankenhaustyp ohne Variablensplit. 'Ausreisser' sind vollständige Einheiten, 'Empfänger' können hingegen auch extreme Werte aufweisen.

	Krankenhaustypologie											
Charak-teristik	**Zentrums-vers.**		**Grund-vers.**		**Psych. Kliniken**		**Rehab.-kliniken**		**Andere Spez.kliniken**		**Total**	
Empfänger	10	37.04	92	48.42	39	55.71	20	44.44	78	72.9	239	54.44
Ausreisser	5	18.52	14	7.37	9	12.86	6	13.33	10	9.35	44	10.02
Spender	12	44.44	84	44.21	22	31.43	19	42.22	19	17.76	156	35.54
Total	27	6.15	190	43.28	70	15.95	45	10.25	107	24.37	439	100

B.5 Einfluss der Einsetzungen

Tabelle 36 Anzahl Jahres Einheiten der 1898 Jahres Einheiten der KS, die vor und nachder NNE den Test 'Test' nicht bestanden. 'Δ' bezeichnet die Veränderung.

Test	vorher	nachher	Δ	Δ [%]
T. 1343	52	52	0	0
T. 1344	7	7	0	0
T. 1345	11	9	-2	-18
T. 1346	31	31	0	0
T. 1347	24	24	0	0
T. 1348	10	71	61	610
T. 1349	5	5	0	0
T. 1358	24	71	47	196
T. 1359	10	71	61	610

Test	vorher	nachher	Δ	Δ [%]
T. 1360	26	75	49	188
T. 1368	5	52	47	940
T. 1369	6	54	48	800
T. 1370	6	52	46	767
T. 1371	114	114	0	0
T. 1401 - 1442	31	31	0	0
T. 1504	8	8	0	0
T. 1511	7	12	5	71

Abbildung 18 Prozentuale Abweichungen der Standardabweichung nach der Einsetzung gegenüber der Standardabweichung vor der Einsetzung, 'orig. Standardabweichung'.

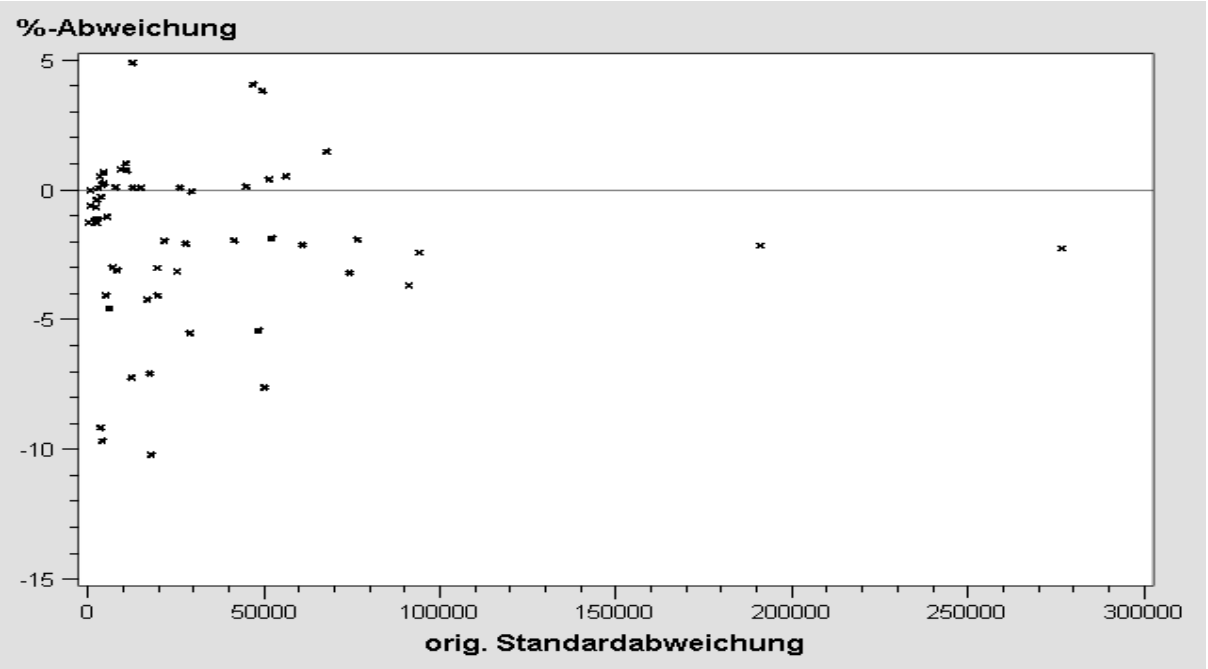


Tabelle 37 Anzahl Jahres Einheiten der KS vor und nach der NNE mit mindestens einer Variablenentwicklung, die mittels *Hidiroglou-Bertholot* Methode als Ausreisser entdeckt wurde. Parameter der Ausreisser Erkennung: $u = 0.8$, $A = 0.05$, $L = 50$.

	Jahr									
	1998-1999		1999-2000		2000-2001		2001-2002		Total	
Variable	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher
Saeuglinge 3.1.1	3	3	16	18	10	11	0	0	29	32
Gesunde 3.1.1	1	1	0	0	0	0	0	1	1	2
Hospitalisierte 3.1.1	2	6	4	9	30	44	1	2	37	61
Teilstätionäre Hospitalisierung 3.1.1	3	4	11	13	3	4	1	1	18	22
Pflegetage 3.2.1	0	7	0	5	0	3	0	0	0	15
Bestand 3.2.2	1	7	3	13	8	10	7	12	19	42
Privat und halbprivat Bestand 3.2.2	1	1	1	1	3	3	4	4	9	9
Aufnahmen 3.2.3	0	10	0	7	1	4	1	2	2	23
Austritte 3.2.5	0	9	0	6	1	3	1	1	2	19
Teilstationäre 3.3	1	3	1	4	3	3	3	3	8	13
Personalaufwand 5.1.1	0	8	0	9	0	4	2	2	2	23
Übriger Betriebsaufwand 5.1.1	1	7	0	6	0	2	1	1	2	16
Ertrag aus Hospitalisierung und Pflege 5.1.2	0	5	0	8	2	9	0	3	2	25
Übriger Betriebsertrag 5.1.2	3	3	5	5	4	4	4	4	16	16
Beiträge und Subventionen 5.1.2	2	2	5	5	6	6	8	8	21	21
Betriebsertrag 5.1.2	0	4	0	6	2	8	1	4	3	22
Total Deckung des Betriebsverlusts - Ungedeckter Verlust 5.2	6	15	6	11	7	7	12	13	31	46

Tabelle 38 Anzahl Jahres Einheiten der KS vor und nach der NNE mit mindestens einer Variablenentwicklung, die mittels *Hidiroglou-Bertholot* Methode als Ausreisser entdeckt wurde. Parameter der Ausreisser Erkennung: $u = 0.8$, $A = 0.05$, $L = 50$.

	Jahr									
	1998-1999		1999-2000		2000-2001		2001-2002		Total	
Variable	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher	vorher	nachher
Vollzeitäquivalente	6	15	4	9	0	2	1	4	11	30
Vollzeitäquivalente Besoldung 30	5	10	5	14	1	3	2	5	13	32
Vollzeitäquivalente Besoldung 31	9	12	2	8	0	4	1	3	12	27
Vollzeitäquivalente Besoldung 32	3	7	2	10	2	3	2	5	9	25
Vollzeitäquivalente Besoldung 33	4	6	3	8	0	0	1	1	8	15
Vollzeitäquivalente Besoldung 34	6	13	5	6	2	5	1	1	14	25
Vollzeitäquivalente Besoldung 35	3	2	2	2	2	3	2	2	9	9
Vollzeitäquivalente Männer	8	14	5	10	0	2	0	3	13	29
Vollzeitäquivalente Frauen	6	13	6	11	2	4	1	4	15	32
Durchschnittliche Vollzeitäquivalente	6	14	3	11	0	1	0	2	9	28
Durchschnittliche Vollzeitäquivalente Besoldung 30	4	10	5	13	2	3	2	4	13	30
Durchschnittliche Vollzeitäquivalente Besoldung 31	6	13	2	5	0	3	0	0	8	21
Durchschnittliche Vollzeitäquivalente Besoldung 32	3	7	1	8	1	2	2	2	7	19
Durchschnittliche Vollzeitäquivalente Besoldung 33	3	5	2	6	0	0	0	0	5	11
Durchschnittliche Vollzeitäquivalente Besoldung 34	4	6	1	3	1	3	0	1	6	13
Durchschnittliche Vollzeitäquivalente Besoldung 35	3	3	2	2	2	3	1	1	8	9
Durchschnittliche Vollzeitäquivalente Männer	7	12	4	8	0	1	0	0	11	21
Durchschnittliche Vollzeitäquivalente Frauen	5	10	5	11	1	3	1	1	12	25
Diensttage	6	11	4	11	4	4	1	3	15	29
Diensttage Besoldung 30	2	10	4	12	3	3	3	6	12	31
Diensttage Besoldung 31	6	8	3	7	1	4	0	1	10	20
Diensttage Besoldung 32	2	3	1	12	5	5	2	3	10	23
Diensttage Besoldung 33	3	3	2	4	2	4	0	0	7	11
Diensttage Besoldung 34	4	7	3	2	1	3	1	5	9	17
Diensttage Besoldung 35	7	4	6	9	3	4	2	3	18	20
Diensttage Männer	4	13	3	7	1	3	0	2	8	25
Diensttage Frauen	4	10	3	11	3	4	1	3	11	28

Abbildung 19 Einfluss der Einsetzung und der Ausreisser auf die Standardtabellen am Beispiel der Tabelle K.3 der KS 1998. 'i_' bezeichnet die eingesetzten Werte. 'i_Anteil_' bezeichnet den %-Anteil an Krankenhäusern, die eingesetzt wurden. 'Ausr_Anteil' bezeichnet den Anteil des Zellenwertes, der als HB-Ausreisser lokalisiert wurde. In der untersten Zeile haben diese beiden Spalten keine Aussagekraft. Gelb: 5%-10%, Rot: >10%.

K.3.	KTO_30	i_KTO_30	i_Anteil_kto_30	Anteil_Ausr_30	KTO_31	i_KTO_31	i_Anteil_kto_31	Anteil_Ausr_31
AG	55221.00	60475.00	0.38	0.06	103812.00	115718.00	0.38	0.00
AI	471.00	471.00	0.00	0.00	4093.00	4093.00	0.00	0.00
AR	5508.00	5508.00	0.00	0.00	9187.00	9187.00	0.00	0.00
BE	162613.00	164406.00	0.09	0.10	291703.00	296161.00	0.09	0.00
BL	30170.00	30170.00	0.00	0.00	62060.00	62060.00	0.00	0.00
BS	91681.00	91681.00	0.00	0.00	117988.00	117988.00	0.00	0.00
FR	23299.00	23299.00	0.00	0.00	54592.00	54592.00	0.00	0.00
GE	99021.00	175707.00	0.29	0.00	158716.00	248368.00	0.29	0.00
GL	3880.00	3880.00	0.00	0.00	7905.00	7905.00	0.00	0.00
GR	27213.00	27213.00	0.00	0.00	37645.00	37645.00	0.00	0.00
JU	10643.00	10643.00	0.00	0.00	19273.00	19273.00	0.00	0.00
LU	49874.00	49874.00	0.00	0.00	80366.00	80366.00	0.00	0.00
NE	23142.00	23142.00	0.00	0.00	48750.00	48750.00	0.00	0.00
NW	2519.00	2519.00	0.00	0.00	6502.00	6502.00	0.00	0.00
OW	2403.00	2403.00	0.00	0.00	5681.00	5681.00	0.00	0.00
SG	59664.00	59664.00	0.00	0.17	97396.00	97396.00	0.00	0.00
SH	9201.00	9201.00	0.00	0.00	17332.00	17332.00	0.00	0.00
SO	24377.00	24377.00	0.00	0.00	56292.00	56292.00	0.00	0.00
SZ	8977.00	8977.00	0.00	0.00	20133.00	20133.00	0.00	0.00
TG	20553.00	22956.00	0.50	0.00	29328.00	35009.00	0.50	0.00
TI	41039.00	41039.00	0.00	0.00	100420.00	100420.00	0.00	0.00
UR	2773.00	2773.00	0.00	0.00	8383.00	8383.00	0.00	0.00
VD	85774.00	85774.00	0.00	0.00	176304.00	176304.00	0.00	0.01
VS	23550.00	23550.00	0.00	0.00	66698.00	66698.00	0.00	0.00
ZG	5298.00	5298.00	0.00	0.00	16433.00	16433.00	0.00	0.00
ZH	166435.00	180972.00	0.30	0.00	292968.00	307420.00	0.30	0.00
All	1035299.00	1135972.00	1.55	0.33	1889960.00	2016109.00	1.55	0.01

B.5.1 Entwicklung von Indikatoren

Abbildung 20 Durchschnittliche Aufenthaltsdauer 1998-2002 Typ 11 und 12.

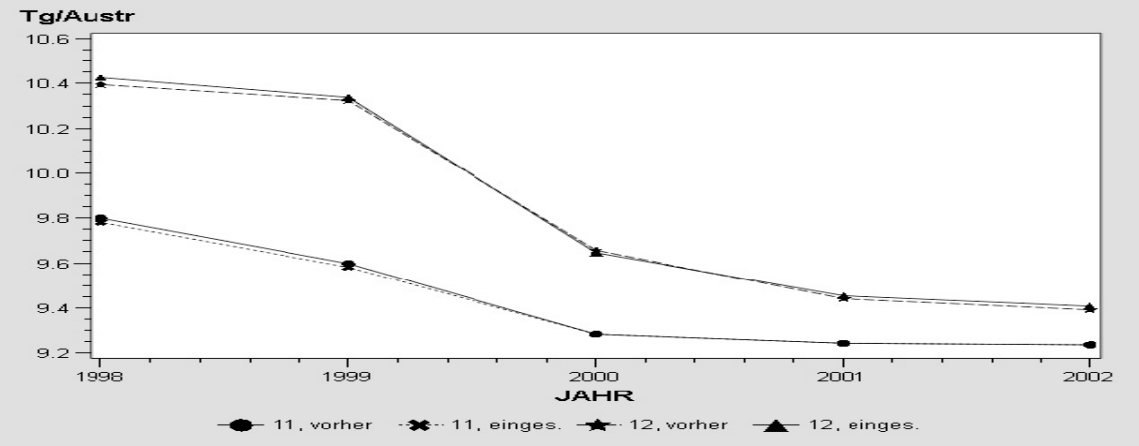


Abbildung 21 Durchschnittliche Aufenthaltsdauer 1998-2002 Typ 21 und 22.

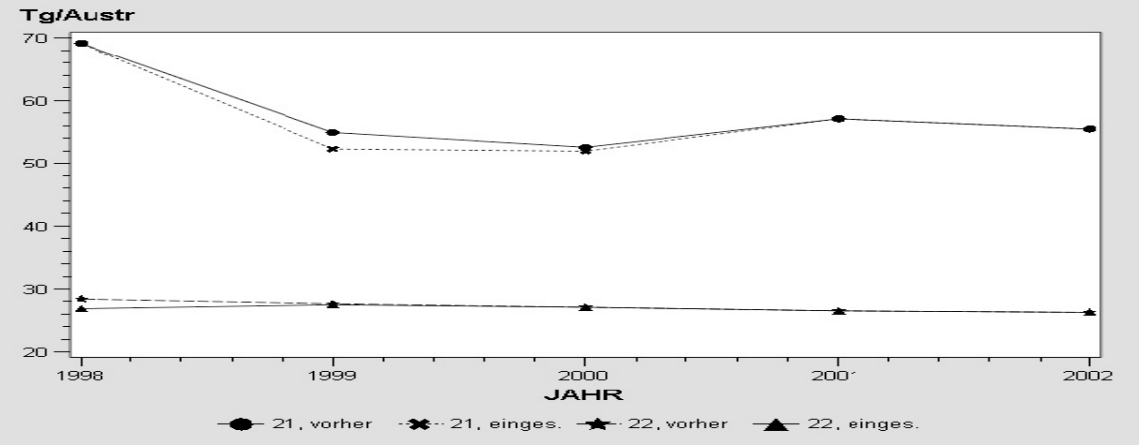


Abbildung 22 Durchschnittliche Aufenthaltsdauer 1998-2002 Typ 23.

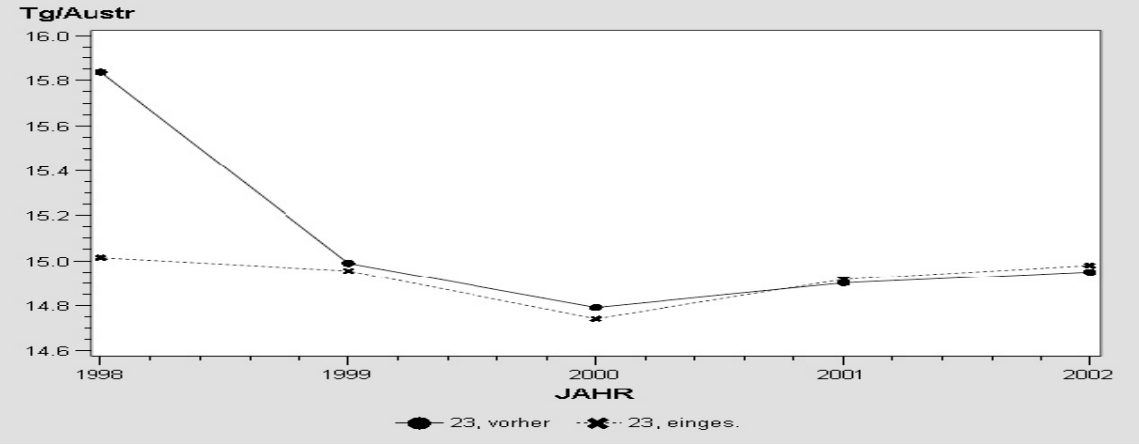


Abbildung 23 Durchschnittliche Kosten pro Tag 1998-2002 Typ 11 und 12.

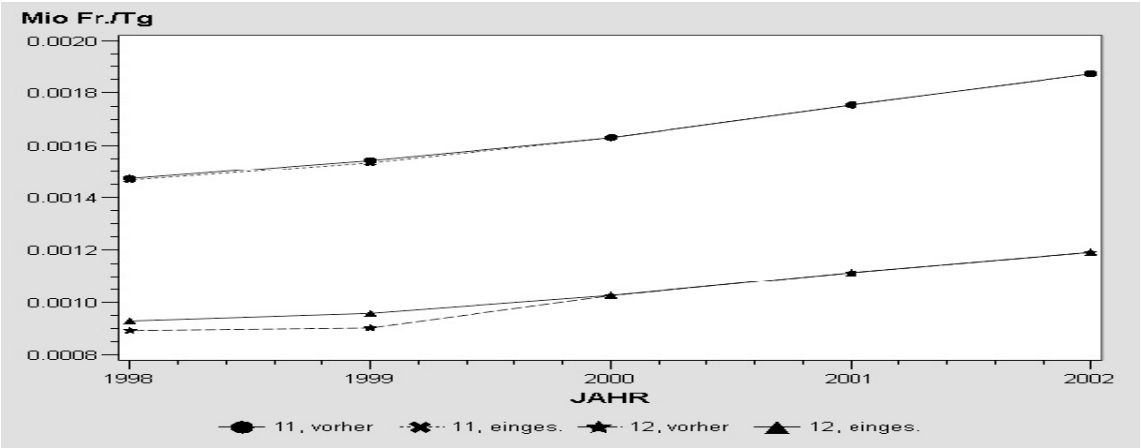


Abbildung 24 Durchschnittliche Kosten pro Tag 1998-2002 Typ 21 und 22.

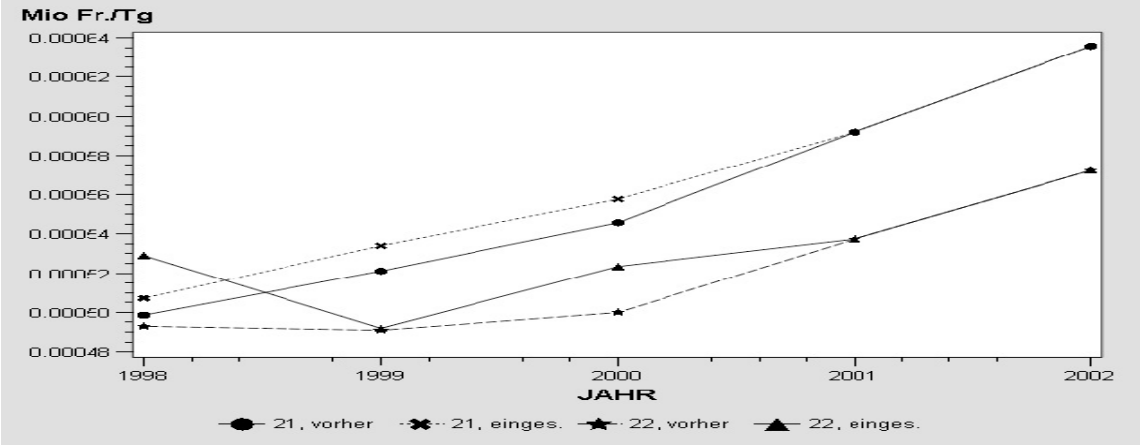


Abbildung 25 Durchschnittliche Kosten pro Tag 1998-2002 Typ 23.

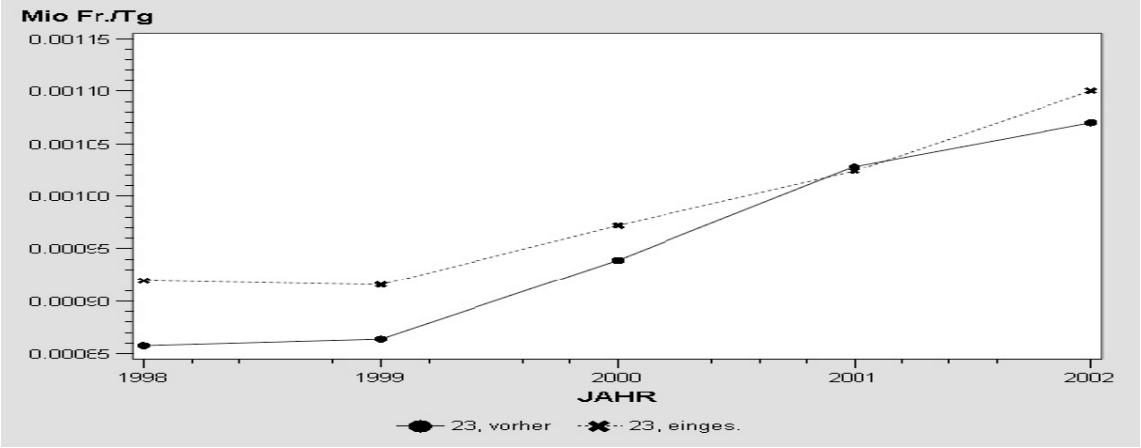


Abbildung 26 Durchschnittliche Kosten (o. Invest.) pro Tag 1998-2002 Typ 11 und 12.

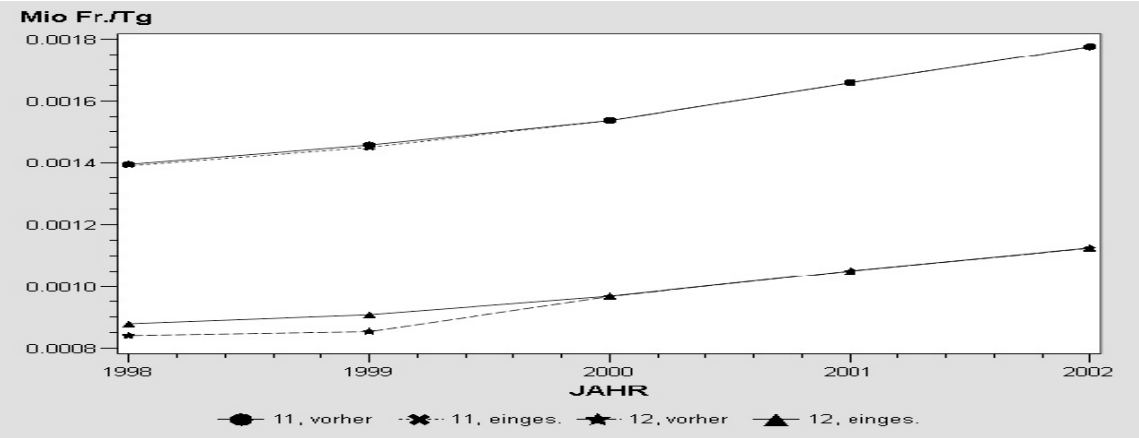


Abbildung 27 Durchschnittliche Kosten (o. Invest.) pro Tag 1998-2002 Typ 21 und 22.

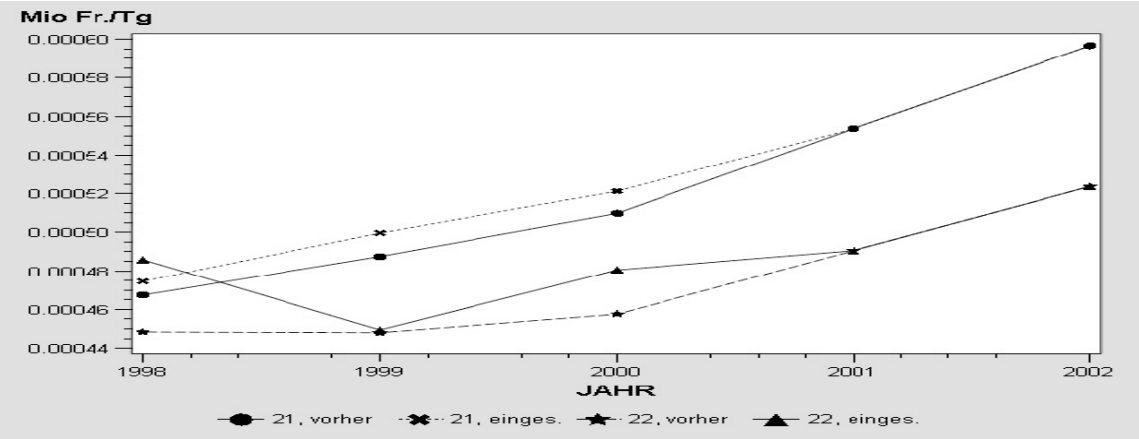


Abbildung 28 Durchschnittliche Kosten (o. Invest.) pro Tag 1998-2002 Typ 23.

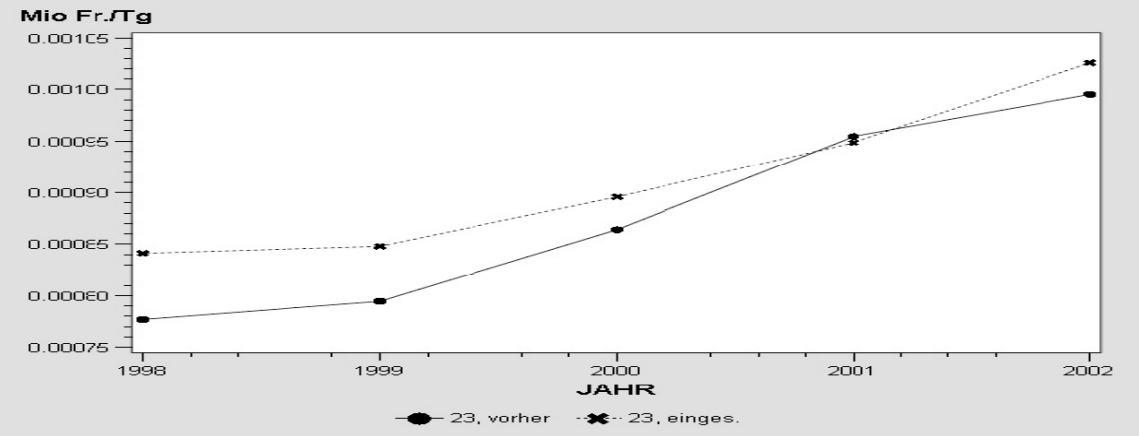


Abbildung 29 Durchschnittliche Kosten (o. Invest. & Amb.) pro Tag 1998-2002 Typ 11 und 12.

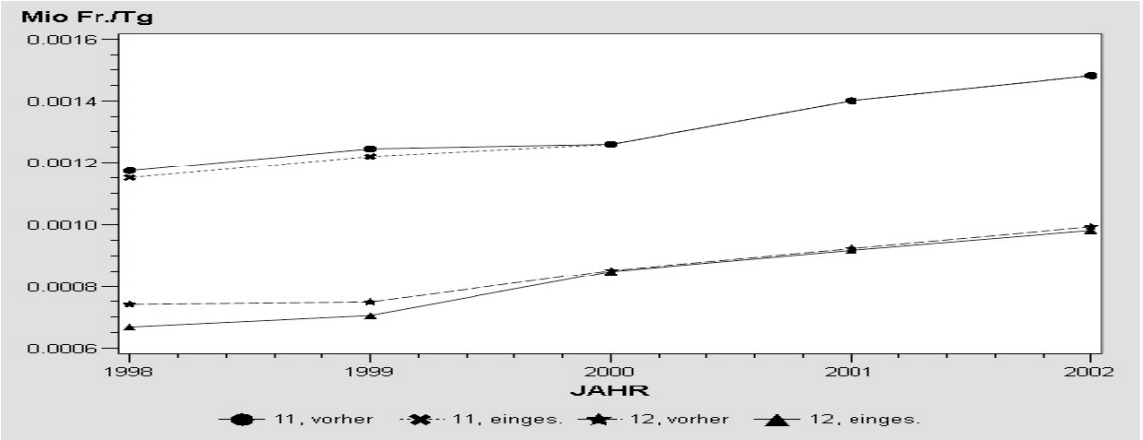


Abbildung 30 Durchschnittliche Kosten (o. Invest. & Amb.) pro Tag 1998-2002 Typ 21 und 22.

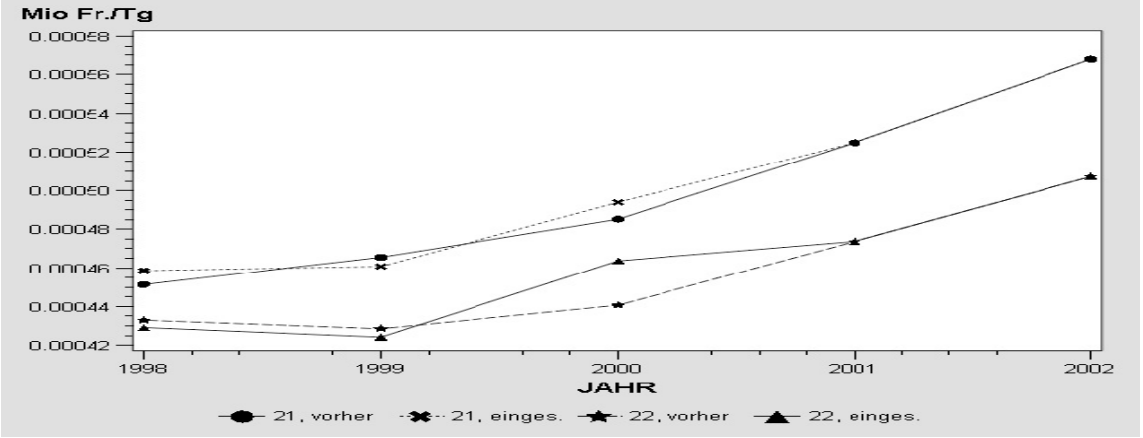


Abbildung 31 Durchschnittliche Kosten (o. Invest. & Amb.) pro Tag 1998-2002 Typ 23.

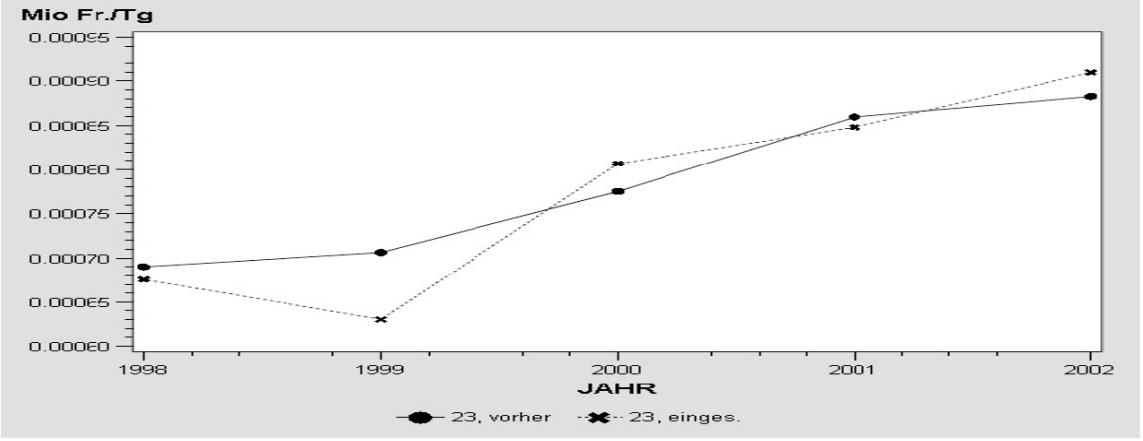


Abbildung 32 Vzae. Pflegepersonal pro 1'000 Pflegetage 1998-2002 Typ 11 und 12.

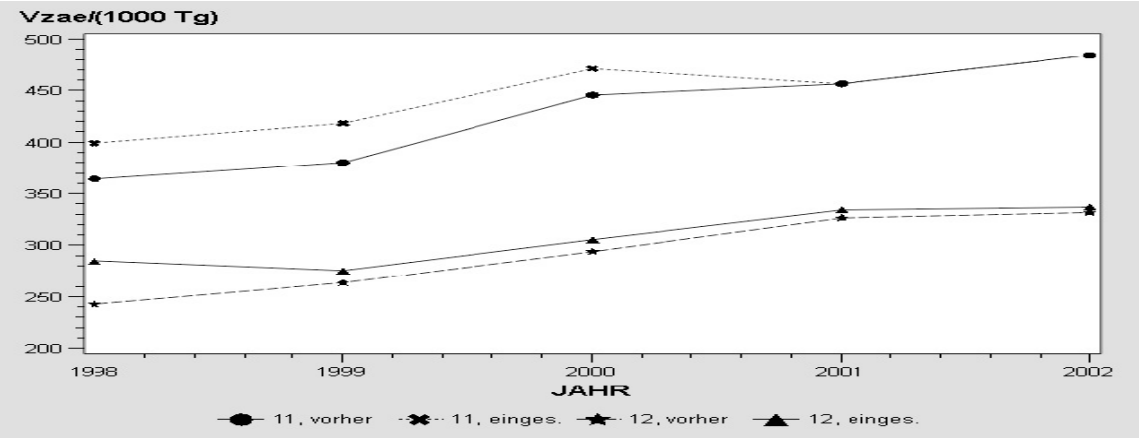


Abbildung 33 Vzae. Pflegepersonal pro 1'000 Pflegetage 1998-2002 Typ 21 und 22.

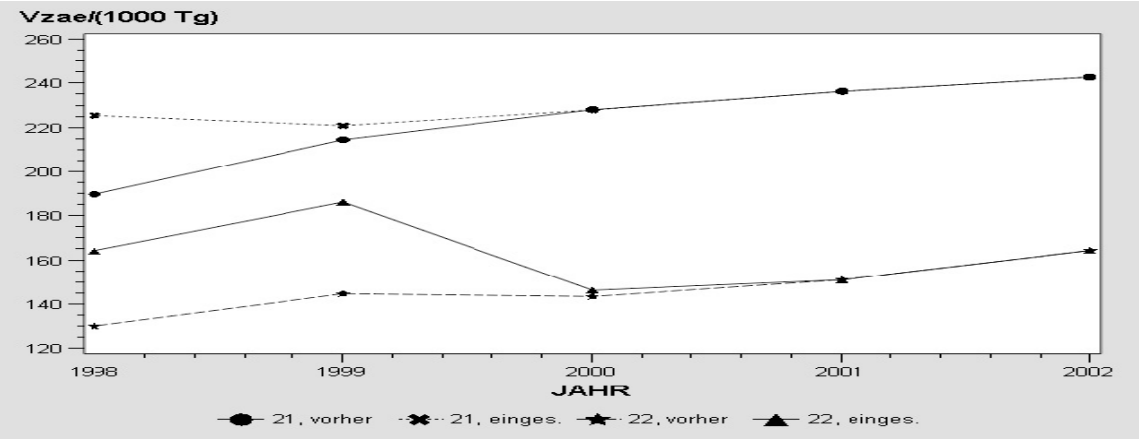


Abbildung 34 Vzae. Pflegepersonal pro 1'000 Pflegetage 1998-2002 Typ 23.

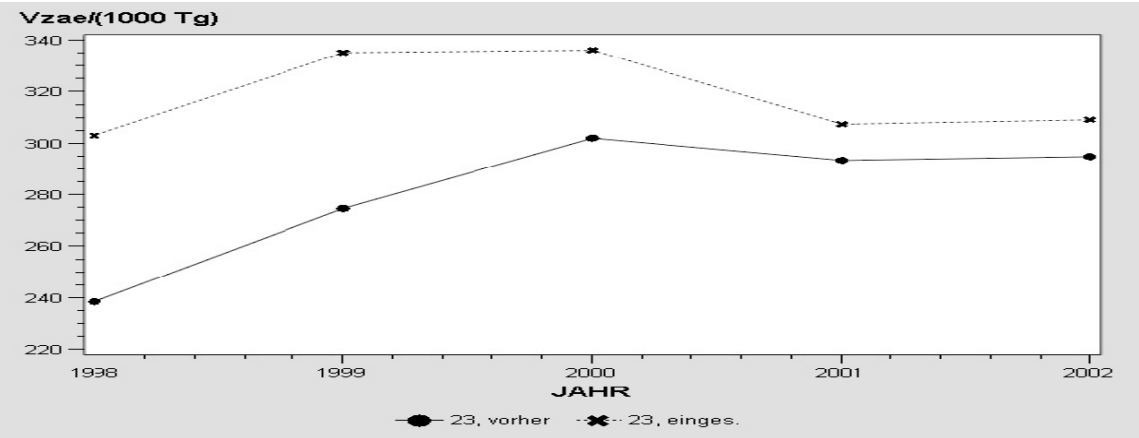


Abbildung 35 Total Vzae. pro 1'000 Pflegetage 1998-2002 Typ 11 und 12.

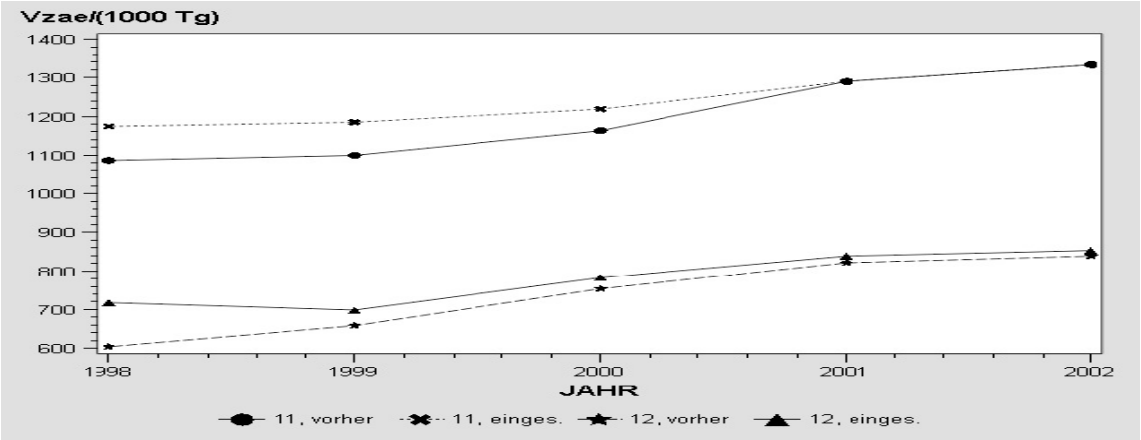


Abbildung 36 Total Vzae. pro 1'000 Pflegetage 1998-2002 Typ 21 und 22.

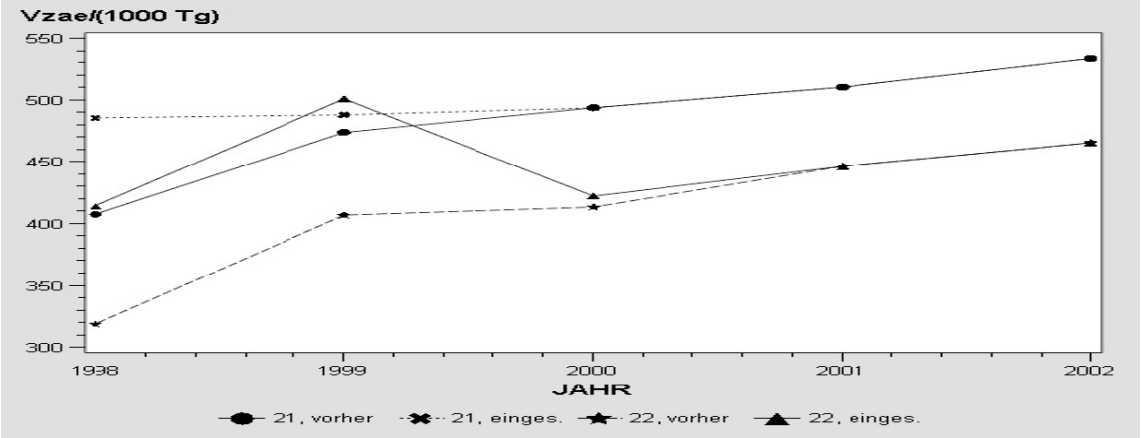
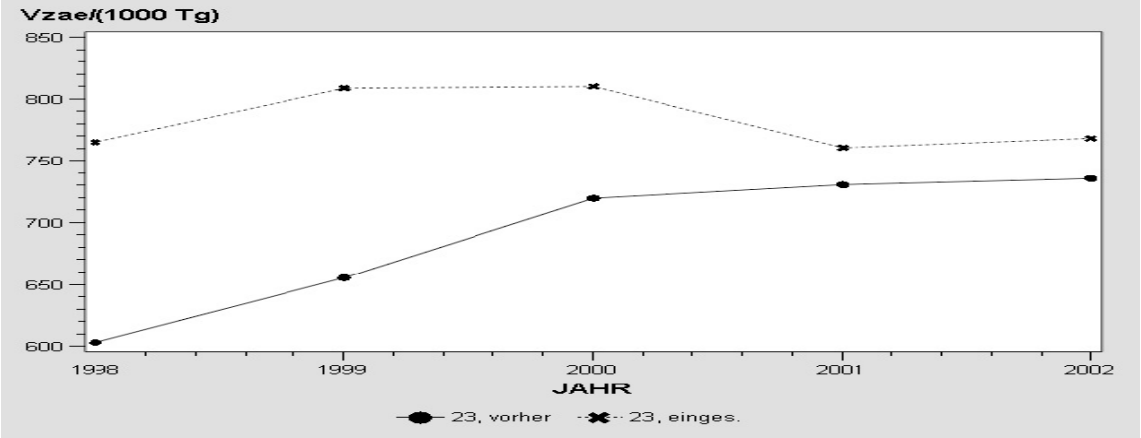


Abbildung 37 Total Vzae. pro 1'000 Pflegetage 1998-2002 Typ 23.



Literatur

Béguin, C. and B. Hulliger (2004). Multivariate oulier detection in incomplete survey data: The epidemic algorithm and transformed rank correlations. *J.R.Statist.Soc.A 167*(Part 2.), 275–294.

Chambers, R. (1986). Outlier robust finite population estimation. *Journal of the American Statistical Association 81*(396), 1063–1069.

EUREDIT Project (2004). *Towards Effective Statistical Editing and Imputation Strategies - Findings of the Euredit project*, Volume 1. http://www.cs.york.ac.uk/euredit/results/results.html.

Haziza, D. (2002). Imputation Classes. *The Imputation Bulletin of Statistics Canada 2*(1), 7–11.

Haziza, D. (2003). *Comment traiter la non-réponse* (Kursunterlagen ed.). Neuchâtel: Statistics Canada.

Hidiroglou, M. and J. Berthelot (1986). Statistical editing and imputation for periodic business surveys. *Survey Methodology 12*(1), 73–83. Statistics Canada.

Hulliger, B. (1999). Simple and robust estimators for sampling. In *Proceedings of the Section on Survey Research Methods*, pp. 54–63. American Statistical Association.

Ruckstuhl, A. (2004). *Einführung in die robusten Schätzmethoden* (Unterlage Nachdiplomkurs ed.). Winterthur: Zürcher Hochschule Winterthur.

Schafer, J. L. (2000). *Analysis of Incomplete Multivariate Data*. New York: Chapmann and Hall/CRC.

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

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. Bestellnummer: 338-0050

Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro de commande: 338-0049

Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. Bestellnummer: 338-0048

Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. Bestellnummer: 338-0047

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. Bestellnummer: 338-0046

Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro de commande: 338-0045

Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro de commande: 338-0044

Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0043

Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. Bestellnummer: 338-0042

Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0041

Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro de commande: 338-0040

Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro de commande: 338-0039

Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. Order number: 338-0038

Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0037

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). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. 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. 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
Methodology Reports by 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 der Statistik der sozialmedizinischen Institutionen und der Krankenhausstatistik traten in den Jahren 1999 bis 2004, respektive 1998 bis 2002, wenige Antwortausfälle von Institutionen oder Krankenhäusern auf. Zusätzlich ergaben sich aber auch Merkmalsausfälle aufgrund von unvollständig gelieferten Daten. Beide Ausfallstypen wurden mittels einer Nächster-Nachbar-Einsetzungsmethode behandelt. Das Ziel, die multivariaten Zusammenhänge soweit möglich zu berücksichtigen und konsistente Entwicklungen abzubilden, gab den Ausschlag für die Anwendung der Nächster-Nachbar-Einsetzungsmethode. Diese Ziele konnten denn auch grösstenteils erreicht werden. Ausreisser wurden in der Krankenhausstatistik mit einer robustifizierten Version der Nächster-Nachbar-Einsetzungsmethode berücksichtigt. In beiden Datensätzen sind aber auch nach den Einsetzungen Ausreisser und Inkonsistenzen vorhanden. Grund dafür ist die grosse Anzahl sozialmedizinischer Institutionen, resp. Krankenhäuser, die über all die Jahre mindestens einmal unvollständige Daten geliefert hatten und somit nicht als Datenspender verwendet werden konnten. Deshalb wurden nicht alle Variablen eingesetzt, um die Anzahl Datenspender zu vergrössern.

Bestellnummer: 338-0050
Bestellungen: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Preis: gratis
ISBN 978-3-303-00402-9