Ludwig-Maximilians-Universität München
Institut für Statistik

Gewichtungsverfahren Vorbereitungsmaterial

Betreuer: Prof. Dr. Thomas Augustin
Verfasser: Lia Gegeshidze

München, den 11.06.2016

Inhaltsverzeichnis

1 Einleitung

Das Hauptthema der vorliegenden Seminararbeit ist der Vergleich und die Analyse der verschiedenen Gewichtungsmethoden.

- Was ist Gewichtung und warum wird die angewendet?

Unter Gewichtung versteht man die Bewertung einzelner Untersuchungseinheiten hinsichtlich ihrer Wichtigkeit für die Stichprobe und später für die Beantwortung der Fragestellung. Das Ziel der Gewichtung besteht darin die Repräsentativität der Stichproben zu verbessern und in den Stichproben auftretender Verzerrung entgegenzuwirken. Gewichte, die in den Stichproben für die Auswahlwahrscheinlichkeiten der einzelnen Untersuchungseinheiten verwendet werden, werden als Designgewichte benannt. Die Berechnung von Designgewichten ergibt sich logisch aus dem Plan der Stichprobenziehung und ist deshalb unproblematisch und im Grunde unumstritten. Vielmehr umstritten ist die sogenannte anpassungsbasierte Gewichtung. Bei dieser Gewichtungsprozedur werden die Probleme, die fehlende Daten mit sich bringen, behandelt. Meistens werden die Gewichtungsverfahren bei den vollständigen Antwortausfällen (Unit Non-Response) angesetzt. Bei dem Item Non-Response wiederum werden die Imputationsverfahren in Betracht gezogen.
Über die Verbesserung der Repräsentativität durch die Anwendung der anpassungsbasierte Gewichtung wird oft skeptisch geäußert. Christian Alt und Walter Bien kommen nach dem Vergleich der ungewichteten und unterschiedlich gewichteten Daten des Familiensurveys mit der Referenzdatei des Mikrozensus zu dem Schluss, dass „die schlechte Daten durch Gewichtung noch schlechter werden und gute Daten durch Gewichtung nicht verbessert werden können“.(Gabler et al. 1994, S. 138)
Außerdem wird es oft kritisiert, dass durch die anpassungsbasierte Gewichtung die Ergebnisse manipuiliert werden und sich jedes Ergebnis reproduzieren lassen können. (Gabler et al. 1994, S. 2)

2 Designbasierte Gewichtung

2.1 Der Horvitz-Thompson-Schätzer (HT)

Kernpunkt einer jeden designbasierten Gewichtung bildet die Logik des sog. „Horvitz-Thompson-Schätzers“. Der Schätzer kann als gewichtetes Mittel der Beobachtungen interpretiert werden.

Gegeben sei eine finite Population U=$\{1,..,N\}$, sowie die Variablenwerte $\{y_1, y_2, ..., y_N\}$ der interessierenden Variable Y. Zwei Populationsparameter, die üblicherweise durch Stichprobenuntersuchungen geschätzt werden sollen, sind der Gesamtwert

$$\theta = \sum_{k \in U} y_k$$

sowie das arithmetische Mittel aller Variablenwerte

$$\overline{\theta}_U = \frac{y}{N}$$

Bei einer Stichprobenauswahl wird ein Sample S vom Umfang n aus der Population U gezogen. Beim Vorliegen einer einfachen Zufallsauswahl sind die Auswahlwahrscheinlichkeiten für alle Elemente gleich und ergeben sich mit $\pi_k = \pi = \begin{pmatrix} N-1 \\ n-1 \end{pmatrix} / \begin{pmatrix} N \\ n \end{pmatrix} = n/N$, wobei $\begin{pmatrix} N \\ n \end{pmatrix}$ Anzahl der möglichen Stichproben beim Ziehen ohne Zurücklegen beträgt. Der Kehrwert der Auswahlwahrscheinlichkeit bildet nach Logik des HT-Schätzers das Gewicht der Merkmalswerte.

Der Horvitz-Thompson-Schätzer ist für die Schätzung des Totalwertes folgendermaßen

$$\hat{\theta}_{HT} = \sum_{k \in S} y_k \cdot \frac{1}{\pi_k} \tag{1}$$

und für die Schätzung des arithmetischen Mittelwertes folgendermaßen definiert

$$\hat{\overline{\theta}}_{HT} = \frac{1}{N} \sum_{k \in S} y_k \cdot \frac{1}{\pi_k} \tag{2}$$

Die Eigenschaften des Stichprobenverfahrens, worauf HT-Schätzer beruht:

- Jedes Element kann nur einmal in die Stichprobe gelangen.

- Die Wahrscheinlichkeit für jedes Element der Grundgesamtheit, in die Stichprobe zu gelangen, ist größer als 0 ($\pi_k > 0$).

ist in vielen Stichproben erfüllt. Daher kann der HT-Schätzer für beliebige Stichprobendesigns mit diesen Eigenschaften verwendet werden. (Kauermann et al. 2011, S. 100)
Wie es an dem Beispiel von Basu's Elefanten zu sehen war, kann bei der Verwendung des Hilfsmerkmals X ($\{x_1, x_2, ..., x_N\}$) (wird auch als Sekundärinformation benannt) die Über- oder Unterrepräsentierung vermieden werden. Es kann eine Stichprobe so gezogen werden, dass die einzelne Auswahlwahrscheinlichkeiten proportional zur Größe der Sekundärinformation sind. Die zu X proportionalen Auswahlwahrscheinlichkeiten ergeben sich dann zu

$$\pi_k = n \cdot \frac{x_k}{\sum_{k \in U} x_k}$$

Wird die beschriebenen größenproportionalen Auswahlwahrscheinlichkeiten für die Berechnung des HT-Schätzers verwendet, wird ersichtlich, dass die Stärke der Gewichtung von der reziproken Auswahlchance abhängt. Je größer die Auswahlwahrscheinlichkeit π für ein Element ist, desto geringer fällt sein Variablenwert bei der Berechnung eines Kennwertes ins Gewicht und Umgekehrt, je geringer dessen Chance, desto größer ist seine relative Wichtigkeit. Die Designbasierte Gewichtung sorgt dafür, dass die Unter- und Überrepräsentierung ausgeglichen wird.
Man kann zeigen, dass der HT-Schätzer im Falle einer einfachen Zufallsstichprobe, also wenn $\pi_k = n/N$ gesetzt wird, gleich der Stichprobenmittelwert ist

$$\hat{\bar{\theta}}_{HT} = \frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k} = \frac{1}{N} \frac{N}{n} \sum_{k \in S} y_k = \bar{\theta}_S$$

Daraus folgt, dass HTS ein erwartungstreuer und unverzerrter Schätzer ist

$$Bias(\hat{\bar{\theta}}_{HT}) = \mathrm{E}(\hat{\bar{\theta}}_{HT}) - \bar{\theta}_U = \mathrm{E}(\frac{1}{N} \sum_{k \in S} \frac{y_k}{\pi_k}) - \bar{\theta}_U = \frac{1}{N} \frac{N}{n} \sum_{k \in S} \mathrm{E}(y_k) - \bar{\theta}_U = 0$$

Neben der Erwartungstreue ist die Präzision eines Schätzers in Statistik auch sehr relevant. Diese Präzision wird mit Hilfe von Varianz gemessen. Sie ist allgemein definiert als die Summe der quadratischen Abstände aller möglichen Schätzer vom Erwartungswert, geteilt durch ihre Anzahl. Für eine präzise Schätzung wird erwartet, dass sie möglichst klein ist. Man kann durch verschiedene Maßnahmen die Verringerung der Varianz erreichen, aber darauf wird in diesem Seminararbeit nicht eingegangen.

Da es meistens die Ziehung einer einfachen Zufallsstichprobe wegen der umfangreichen Population als schwierig erweist, wird oft die Schichtung der Population nach unterschiedlichen Kriterien durchgeführt. Die Population wird in nicht überlappende $h = \{1, \cdots H\}$ Schichten mit Hilfe der Sekundärinformatio aufgeteilt, so dass die Anzahl der Population $N = \sum_{h=1}^{H} N_h$ entspricht. Als Sekundärinformation kann zum Beispiel vorhandene Information über Region, Bundesland, Stadtbezirk betrachtet werden und wird auch als Schichtungsmerkmal genannt. Der Stichprobenumfang wird für jede Schicht einzeln festgelegt, aus den einzelnen Schichten wird jeweils eine Stichprobe gezogen und die erhobenen Daten dann erst zur Auswertung zusammengeführt. Bei der Schichtung soll darauf geachtet werden, dass die Merkmalsträger innerhalb einer Schicht so ähnlich wie möglich sind. Die einzelne Schichte sollten sich aber untereinander so weit wie möglich unterschieden. (Kauermann et al. 2011, S. 144) Die geschichtete Stichprobe ist in der Praxis am häufigsten verwendetes Design, da

- die getrennte Auswertung der Daten innerhalb der Schichten möglich ist.
- bei der Schätzung des Gesamtmittelwertes einen erheblichen Effizienzgewinn erreicht wird.
- bei einer Wahl der schichtenproportionalen Stichprobenumfänge, die Stichprobe repräsentativ bezüglich des Schichtmerkmals ist. (Kauermann et al. 2011, S. 138)

Die Auswahlwahrscheinlichkeit, dass ein Individuum k aus der Schicht h gezogen wird, bei der einfachen Zufallsauswahl beträgt

$$\pi_{h,k} = \frac{n_h}{N_h}$$

Für die Schätzung des Totalwertes aus den Schichten gezogenen Daten kann Horvitz-Thompson-Schätzer angewendet werden.

$$\hat{\theta}_{HT,GS} = \sum_{h \in H} \sum_{k \in n_h} y_{hk} \cdot \frac{1}{\pi_{hk}} \tag{3}$$

Die Schätzung des arithmetischen Mittelwertes sieht folgendermassen aus

$$\hat{\bar{\theta}}_{HT,GS} = \frac{1}{N} \sum_{h \in H} \sum_{k \in n_h} y_{hk} \cdot \frac{1}{\pi_{hk}} \tag{4}$$

2.2 Der verallgemeinerte Regressionsschätzer (GREG)

Im Unterschied zum Horvitz-Thompson-Schätzer wird beim verallgemeinerten Regressionsschätzer statistisches Modell verwendet und die vorhandene Hilfsvariablen in das Modell aufgenommen. Damit wird die Präzison erzielt. Der Präzisionsgewinn des GREG gegenüber dem HT-Schätzer kann beim Vorliegen nur einer Hilfsvariable als $1 - \rho^2$ angegeben werden. (Münnich et al. 2012, S. 42) ρ^2 bezeichnet die Korrelationskoeffizient zwischen Untersuchungs- Y und Hilfsmerkmal X.

Im Regressionsschätzer wird das Designgewicht $d_k = 1/\pi_k$ für die Gewichtung der erklärenden Variablen und der abhängigen Variable angewendet. Bei der Schätzung des verallgemeinerten Regressionsschätzers wird zwischen

- Kombinierten Schätzung: Es wird über alle Schichten hinweg ein einziges Regressionsmodell betrachtet.

$$\hat{\theta}_{GREG} = \sum_{k \in S} d_k y_k + (\sum_{k \in U} x_k - \sum_{k \in S} d_k x_k) \cdot \hat{\beta} \tag{5}$$

$$\text{wobei} \quad \hat{\beta} = (\sum_{k \in S} d_k x_k x_k^{'})^{-1} (\sum_{k \in S} d_k x_k y_k)$$

- und separaten Schätzung: es wird separat in jeder Schicht ein eigenes Regressionsmodell betrachtet.

$$\hat{\theta}_{GREG} = \sum_{k \in S_h} d_{kh} y_{kh} + (\sum_{k \in U_h} x_{kh} - \sum_{k \in S_h} d_{kh} x_{kh}) \cdot \hat{\beta}_h \tag{6}$$

$$\text{wobei} \quad \hat{\beta}_h = (\sum_{k \in S_h} d_{hk} x_{hk} x'_{kh})^{-1} (\sum_{k \in S_h} d_{kh} x_{kh} y_{kh})$$

unterschieden.

Der GREG-Schätzer weist eine geringe Verzerrung auf, solange die Teilstichprobenumfänge nicht zu klein sind. (Münnich et al. 2012, S. 42) Der Regressionsschätzer kann auch als Produkt des Anpassungsgewichts und des Designgewichts formuliert werden:

$$\hat{\theta}_{GREG} = \sum_{k \in S} d_k g_k y_k \tag{7}$$

wobei $g_k = 1 + \lambda'_s \cdot x_k$ und $\lambda'_s = (\sum_{k \in U} X_k - \sum_{k \in S} d_k \cdot X_k)'(\sum_{k \in S} d_k \cdot X_k \cdot X'_k)^{-1}$ entspricht. (Särndal et al. 2005, S. 35) Diese Schreibweise des verallgemeinerten Regressionsschätzers weist eine wünschenswerte Eigenschaft auf und zwar die Aggregation von Schätzungen auf Untergruppen stets der Schätzung auf der übergeordneten Ebene entspricht. (Münnich et al. 2012, S. 42)

3 Anpassungsbasierte Gewichtung

3.1 Einführung in die Problematik

Da die meisten sozialwissenschaftliche Erhebungen die Personen als Untersuchungseinheiten verwenden, ist naheliegend dass nicht alle in die Stichprobe vorgesehenen Fälle tatsächlich in die Stichprobe gelangen. Dafür gibt es zahlreiche Gründe. Ein Grund dafür könnte sein, dass die Untersuchungseinheiten die Antwort verweigern und dadurch völlig ausfallen. Anderer Grund könnte aber sein, dass die Untersuchungseinheiten die Antworten auf einzelnen Fragen verweigern. Das führt zu zwei Typen von fehlenden Werten in Nettostichprobe:

- Unit Non-Response → Ausfall ganzer Befragung und
- Item Non-Response → Verweigerung der Antwort auf einzelnen Fragen

Fehlende Werte führen häufig zu systematischen Verzerrungen in der Stichprobe. Dadurch sind die erhobenen Daten nicht mehr repräsentativ, d.h. kein strukturgleiches Abbild der Grundgesamtheit, für die Wohnbevölkerung. Das kann dazu führen, dass über den gesamten Zustand der Grundgesamtheit falsche Rückschlüsse gezogen werden.

Ohne das Auftreten von Non-Response liefert der Horvitz-Thompson-Schätzer unverzerrte Schätzwerte. Da aber eine Stichprobe ohne Non-Response kaum realisierbar ist, und in letzter Zeit immer zurückgehende Responseraten, wirft das Problem auf, dass nur Designgewichtung alleine nicht ausreichend ist, um unverzerrte Schätzergebnisse zu bekommen. (Schupp et al. 2015, S. 13-14)

Sowohl Unit Non-Response als auch Item Non-Response können demnach die Unter- oder Überrepräsentation einzelner Bevölkerungsgruppen in der Nettostichprobe bewirken.

Eigenes Beispiel zur Illustration des Problems (inspiereirt nach „Stressreport Deutschland 2012 “von A. Lohmann-Haislah): Man möchte Arbeitsbelastung der Arbeiter in einem Unternehmen untersuchen und man führt eine Stichprobe durch. Im Unternehmen gibt es eine Gruppe von Mitarbeiter, die aufgrund ihrer hohen Arbeitsbelastung keine Gelegenheit besitzen Fragebogen auszufüllen. Die andere Gruppe aber, die weniger Arbeitsbelastung haben, haben Zeit den Fragebogen auszufüllen. Dadurch werden die Mitarbeiter mit weniger Arbeitsbelastung in der Stichprobe überrepräsentiert und die Mitarbeiter mit höhere Arbeitsbelastung unterrepräsentiert. Dadurch zieht man falsche Schlussfolgerungen aus der Stichprobe und zwar, dass es eine geringe Arbeitsbelastung im Unternehmen gibt. An diesem Beispiel wird klar, dass Non-Response gefährdet nicht nur die Repräsentativität von Untersuchungen sondern auch die Validität der Ergebnisse.

3.2 Definition des Non-Response Bias

Oben beschriebenes Problem führt zu dem sogenannten Non-Response Bias. Der Non-response Bias bezeichnet eine Verzerrung des Stichprobenergebnisses, die durch den Antwortausfall der Untersuchungseinheit entsteht. Non-Response Bias ist definiert als Verhältnis der Kovarianz einer Variable y und der Antwortwahrscheinlichkeit δ zu der mittleren Antwortwahrscheinlichkeit in der Stichprobe (Schupp et al. 2015, S. 15):

$$bias_{\overline{y}} \sim \frac{\sigma_{y,\delta}}{\overline{\delta}} \tag{8}$$

Kovarianz zwischen Zielvariable und der Antwortwahrscheinlichkeit in der Stichprobe ist definiert als

$$\sigma_{y,\delta} = \frac{1}{n-1} \sum_{k \in S} (y_k - \overline{y}) \cdot (\delta_k - \overline{\delta})$$

Die Berechnung der Antwortwahrscheinlichkeit wird in den nachfolgenden Kapiteln diskutiert.

Wie es der Gleichung 8 zu entnehmen ist, einerseits kann Bias verringert werden, wenn die durchschnittliche Teilnahmewahrscheinlichkeit höher ist oder die Kovarianz im Zähler der Formel klein wird. Die erste Möglichkeit ist mit zusätzlichen Kosten in den Stichproben verbunden und wird auch nicht als geeignetes Mittel zur Reduktion des Bias betrachtet. Eher wird die Reduktion der Kovarianz angestrebt. (Schupp et al. 2015, S. 13-15)

Bei der Gewichtung der Unit Non-Response Daten müssen zwei Voraussetzungen erfüllt sein, um den gewünschten Erfolg zu haben.

- Erstens, die zur Gewichtung verwendeten Variablen müssen in einem engen

Zusammenhang zur Teilnahmebereitschaft stehen und möglichst viele Variablen, die die Teilnahmewahrscheinlichkeit beeinflussen können, berücksichtigt werden sollten.

- Zweitens, dürfen sich innerhalb der durch die Gewichtungsvariablen definierten Gruppen der Teilnehmer und Interviewverweigerer nicht systematisch unterscheiden.

3.3 Verfahren zur Korrigierung des Non-Response Bias

3.3.1 Straight Expansion Estimator (EXP)

Wie schon im Abschnitt 2.1 zu sehen war Horvitz-Thompson-Schätzer ist eine erwartungstreue Schätzfunktion. Aber was bei dieser Schätzfunktion nicht berücksichtigt ist, ist der Antwortausfall. In diesem Fall ist die der Schätzer verzerrt. Die einfachste Methode, der Non-Response-Bereinigung durchzuführen, besteht in der Multiplikation des HTS mit der inversen Antwortwahrscheinlichkeit. Dadurch werden durch den Antwortausfall verlorene Designgewichte kompensiert.

$$\hat{\theta}_{EXP} = \frac{1}{N} \sum_{k \in r} \frac{d_k}{\hat{\delta}_k} y_k \tag{9}$$

wobei $\hat{\delta}_k = \frac{m}{n}$ die Antwortwahrscheinlichkeit und $\hat{d}_k = \frac{1}{\hat{\pi}}$ die inverse Auswahlwahrscheinlichkeit darstellt.
Die Annahme, die bei dieser Anpassung getroffen wird, zwar die konstante Antwortwahrscheinlichkeit für alle Stichprobeneinheiten, ist kaum realisierbar. Dadurch ist der Straight Expansion Estimator kein geeignetes Verfahren um Non-Response Bias zu reduzieren. Dieses Verfahren wird aber als Startpunkt im Vergleich immer komplexerer Gewichtungsmethoden verwendet.

3.3.2 Response Homogeneity Groups (RHG)

Eine weitere Möglichkeit ist es die Grundgesamtheit in h ∈{1,...,H} Schichten aufzuteilen und innerhalb der Schichten die konstante Antwortwahrscheinlichkeit anzunehmen. Die Antwortwahrscheinlichkeit kann als relative Häufigkeit der Respondenten innerhalb jeder der h Gruppen beschrieben werden.

$$\hat{\delta}_{hk} = P(k \in h \wedge k \in r) = \frac{m_{hk}}{n_{hk}}$$

Das Anpassungsgewicht ergibt sich somit als inverse Antwortwahrscheinlichkeit pro Gruppe.

$$\hat{\theta}_{RHG} = \sum_{k \in r} \frac{\hat{d}_{hk}}{\hat{\delta}_{hk}} \cdot y_{hk} \tag{10}$$

Dieser beschriebene Schätzer entspricht dem Horvitz-Thompson Schätzer für eine Stichprobe mit zwei Phasen. Falls die Poststratifizierungsgewichte innerhalb der Schichten tatsächlich pro Stichprobenelement gleich sind, ist der Schätzer unverzerrt.

3.3.3 Logistische Regression (LOG)

In den in vorigen Abschnitten beschriebenen Methoden werden angenommen, dass die Antwortwahrscheinlichkeiten für alle Stichprobeneinheiten oder für die Stichprobeneinheiten innerhalb der Gruppen konstant sind. Diese Annahmen sind nicht immer gerechtfertigt. In diesem Abschnitt wird wiederum gezeigt, wie die individuelle Antwortwahrscheinlichkeiten für jedes Element in der Bruttostichprobe geschätzt werden können. Eine geeignete Methode um dies zu schätzen bietet logistische Regression. Es wird der Response-Indikator $R \in \{0;1\}$ definiert, der angibt, ob ein Element der Bruttostichprobe an der Erhebung teilnimmt oder nicht. Mit Hilfe von geeigneten erklärenden Variablen $x = (x_1, \cdots, x_I)$ kann die Teilnahmewahrscheinlichkeit $P(R = 1|x)$ geschätzt werden. Es liegt nahe, dass die Wahl der externen Hilfsvariablen die Qualität der geschätzten Antwortwahrscheinlichkeit beeinflusst.

Logistisches Regressionsmodell für die Schätzung der Antwortwahrscheinlichkeiten ist folgendermaßen definiert:

$$ln\left(\frac{P(R=1|x)}{P(R=0|x)}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_I x_I$$

$$P(R=1|x) = \frac{exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_I x_I)}{1 + exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_I x_I)}$$

Es wird ersichtlich, dass die Werte der geschätzten Antwortwahrscheinlichkeiten gleich für Gruppen mit gleichen Werten in den erklärenden Variablen $x = (x_1, \cdots, x_I)$ sind.
Nach dem die Antwortwahrscheinlichkeiten für jedes Element $P(R = 1|x) = \hat{\delta}_k^{LOG}$ geschätzt wird, kann der erklärende Variable folgendermaßen gewichtet werden

$$\hat{\theta}_{LOG} = \sum_{k \in r} \frac{d_k}{\hat{\delta}_k^{(LOG)}} y_k \tag{11}$$

Wie der Formel zu entnehmen ist die Designgewichte werden an die Non-Response angepasst. Dafür wird der Kehrwert der geschätzten Antwortwahrscheinlichkeit $\nu_k = 1/\hat{\delta}_k$ mit dem Designgewicht multipliziert. Bei der Verwendung des Kehrwerts der geschätzten Antwortwahrscheinlichkeit als Non-Response Anpassungsgewicht kann sehr große Non-Response Gewichte zur Folge haben. Das ist der Fall, wenn die geschätzten Antwortwahrscheinlichkeiten nahe Null sind, da der Kehrwert folglich sehr groß wird. Umgekehrt gilt das gleiche für geschätzte Antwortwahrscheinlichkeiten nahe 1.
Eine Strategie, um dieses Problem umzugehen besteht darin, die auftretende extreme Gewichte zu normieren, was aktuell in EU-SILC Stichproben angewendet wird (Osier et al. 2006, S. 5).

$$\frac{1}{C} \leq \frac{\nu_k/\overline{\nu}}{d_k/\overline{d}} \leq C$$

Bei der Unterschreitung $1/C$ oder Überschreitung C der Grenze wird ν_k folgendermaßen normiert

$$\nu_k^{low} = \frac{1}{C} d_k \frac{\overline{\nu}}{\overline{d}}$$

und

$$\nu_k^{high} = C d_k \frac{\overline{\nu}}{\overline{d}}$$

C wird im Rahmen von EU-SILC länderabhängig definiert. In Österreich wird es gleich 2 gesetzt, in Italien wiederum gleich 3.
Um zu sehen, welchen Einfluss diese korrigierende Maßnahme auf die Ergebnisse hat, habe ich die Daten aus Stichprobenbefragungen des Statistischen Amts

Anpassungsbasierte Gewichtung	Schätzer	Rel. Abweichung von dem ungewichteten Median in %	Bias
nach Trimming (C=3)	149.42	1.38	-3.87
nach Trimming (C=2)	147.79	0,27	-3.90
ohne Trimming	149.42	1,38	-3.88
nur mit designbasierter Gewichtung	148.47	0,73	

Tabelle 1: Vergleich der Trimming Effekte

Georgiens aus dem Jahr 2009 für die Analyse verwendet. Das Ziel der Stichprobenbefragungen ist die Schätzung des Lebensniveaus in Georgien. Das verwendete Stichprobenverfahren für die Ziehung der Daten basiert auf dem zweistufigen Verfahren. Die Daten beinhalten Designgewichte, welche ich für die spätere Berechnungen mitberücksichtige. Aus den Daten habe ich Einpersonenhaushalte ausgewählt und für die Schätzung der Antwortwahrscheinlichkeit der Response-Indikator definiert. Bei der Definition des Response-Indikators habe ich folgende Annahme getroffen: Haushalte, deren Einkommen unten 100 Lari oder über 1.000 lag, bekommen den Wert 0, andere Haushalte aber den Wert 1. In das Modell sind die erklärende Variablen Alter und Gebiet nach Rückwärtsselektion aufgenommen. Mit Hilfe von Response-Indikator und verbleibenden erklärende Variablen habe ich die Antwortwahrscheinlichkeit geschätzt. Es sind einige geschätzte Antwortwahrscheinlichkeiten nahe Null und Eins aufgetreten. Die aufgetretene extreme Gewichte habe ich mit der von EU-SILC verwendete Trimming-Methode korrigiert und danach der gewichtete Schätzer berechnet. Für die Berechnung des gewichteten Schätzers habe ich die in Daten vorhandenen Designgewichte miteinbezogen. Für die Berechnung der relativen Abweichung habe ich den Median in Betracht gezogen. Der Median der ungewichteten Einkommenswerten beträgt 147,38 Lari. Betrachtet man die Tabelle 1, wird ersichtlich, dass sich die Schätzer voneinander nur gering unterscheiden. Bei der Anwendung der korrigierenden Maßnahme mit dem Trimming Wert C=2 nährt sich der Schätzer dem Median. Trimming mit dem Wert C=3 hat aber kaum Einfluss. Grund dafür ist, dass bei der Anwendung des Trimming Wertes C=3 nur sehr wenige Werte korrigiert werden. Bei der Betrachtung des Bias fällt auf, dass er sich bei der mit dem Trimming-Wert C=2 korrigierten anpassungsbasierten Gewichtung leicht verschlächtert hat.

3.3.4 Feed Forward Neural Networks (NNW)

Zwar im vorherigen Abschnitt wird gezeigt wie die Gewichtung mithilfe von logistischen Regression geschätzt werden kann, aber das Problem dieses Gewichtungsverfahrens besteht darin, dass es meistens unbekannt ist, welche erklärende Variablen den Response-Indikator am stärksten beeinflussen. Außerdem in logistische Regression wird eine direkte Verbindung zwischen erklärende und abhängige Variablen geschätzt und dadurch werden die Antwortwahrscheinlichkeiten verzerrt. Eine Methode, welche diese Schwierigkeiten umgeht, ist Feed Forward Neural Networks (NNW). Das ist eine englische Bezeichnung für künstliche Neuronale Netze ohne Rückkopplung. Bei dieser Methode handelt es sich um Mustererkennungsverfahren, aber diese Methode kann auch zu der Schätzung der Wahrscheinlichkeiten von Gruppenzugehörigkeiten dienen und dadurch auch Antwortwahrscheinlichkeiten geschätzt werden können. Im Folgenden wird gezeigt wie mit Hilfe der von feed forward Neural Networks die Antwortwahrscheinlichkeiten geschätzt werden kann. Dafür werden Hilfsvariablen X als Input Variablen und hidden units h_j $(j = 1, \cdots, J)$ verwendet, welche eine lineare Funktion in sogenannten hidden layer bilden.

$$f(X) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \cdot h_j \tag{12}$$

$$\text{wobei} \quad h_j = \psi(\beta_{oj} + \sum_{i=1}^{I} \beta_{ij} \cdot X_i) \quad \text{und} \quad \psi(t) = \frac{1}{1 + exp(-t)}$$

$$f(X) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \cdot \psi(\beta_{0j} + \sum_{i=1}^{I} \beta_{ij} \cdot X_i)$$

Einsetzen der obigen Formel in Andere, ergibt folgende Funktion (Günther et al. 2010, S. 31)

$$f(X) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \cdot \frac{1}{1 + exp(-(\beta_{0j} + \sum_{i=1}^{I} \beta_{ij} \cdot X_i))}$$

Wie das Modell des künstlichen Neurons zeigt, besteht ein wesentlicher Teil des Neurons aus der Summenfunktion. In dem Schritt werden die eingangswerte mit Gewichtungen multipliziert und dann summiert. Summierte Werte werden an Aktivierungsfunktion weitergeleitet. Die weitergeleiteten Werte werden in Akti-

vierungsfunktion als Endergebnisse bearbeitet. Die geschätzten Werte für den Output entsprechen den geschätzten Antwortwahrscheinlichkeiten δ_k.
Das „Wissen“eines künstlichen neuronalen Netzes steckt in seinen Gewichten. Dieses Wissen muss zuvor erlernt werden. Daher werden neuronale Netze zuerst mit Hilfe des so genannten Backpropagation-Algorithmus trainiert. Dafür werden Anpassungen der Gewichte an das Modell betrachtet. Es wird die Abweichung zwischen der gewünschten Ausgabe y_k und der tatsächlichen Ausgabe $f_k(X)$ betrachtet.

$$R(\theta) = \sum_{k=1}^{K} (y_k - f_k(X))^2$$

Um eine gute Anpassung zu bekommen wird die Abweichung minimiert. Für die Minimierung des Fehlers wird Backpropagation-Algorithmus verwendet und besteht aus folgenden Schritten:

- Im ersten Schritt wird Startwerte für alle Gewichte festgelegt.Außerdem wird ein Eingabemuster angelegt, welches anschließend vorwärts durch das Netz propagiert wird. Daraus erhält man die Ausgabe $f_k(X)$.
- Im zweiten Schritt wird die Ausgabe mit den gewünschten Werten verglichen und die Differenz als Fehler aufgefasst. Daraus erhält man die Summe der Fehlerquadrate.
- Im dritten Schritt, im so genannten backward pass, wird der Fehler über die Ausgabeschicht zurück zur Eingabeschicht propagiert. Dabei werden die Gewichtungen der Verbindungen zwischen den Neuronen geändert.

Wie stark die Gewichte verändert werden, hängt von der Lernrate ab. Die Lernrate wird meistens zwischen 0,01 und 0,5 gewählt. Wählt man einen hohen Wert für die Lernrate, führt dies zu einer großen Änderung der Gewichte, wodurch das Minimum von $R(\theta)$ leicht verfehlt werden kann. Bei einem sehr niedrigen Wert wird dagegen das Einlernen verlangsamt. Oft wird 0,1 als Lernrate verendet.

$$\beta_{ij}^{neu} = \beta_{ij}^{alt} + \Delta\beta_{ij} \quad \mathrm{u}nd \quad \alpha_j^{neu} = \alpha_j^{alt} + \Delta\alpha_j$$

Die Wahl der hidden layer ist auch bei der Anpassung entscheidend, wie es im Abschnitt 3.4 auch zu sehen sein wird. Je mehr verdeckte Neuronen ein neuronales Netz besitzt, desto flexibler ist es. Zu viel Flexibilität hat jedoch zur Folge, dass das Modell zu Überanpassung neigt. Die Wahl weniger verdeckten Neuronen kann zur Folge haben, dass das neuronales Netz aufgrund der geringen Anzahl an Parametern nicht komplex genug für die Problemstellung ist und daher nicht richtig trainiert werden kann. Allgemein gilt es, ein möglichst sparsames neuronales Netz zu finden, das dennoch genug Raum für Komplexität lässt.

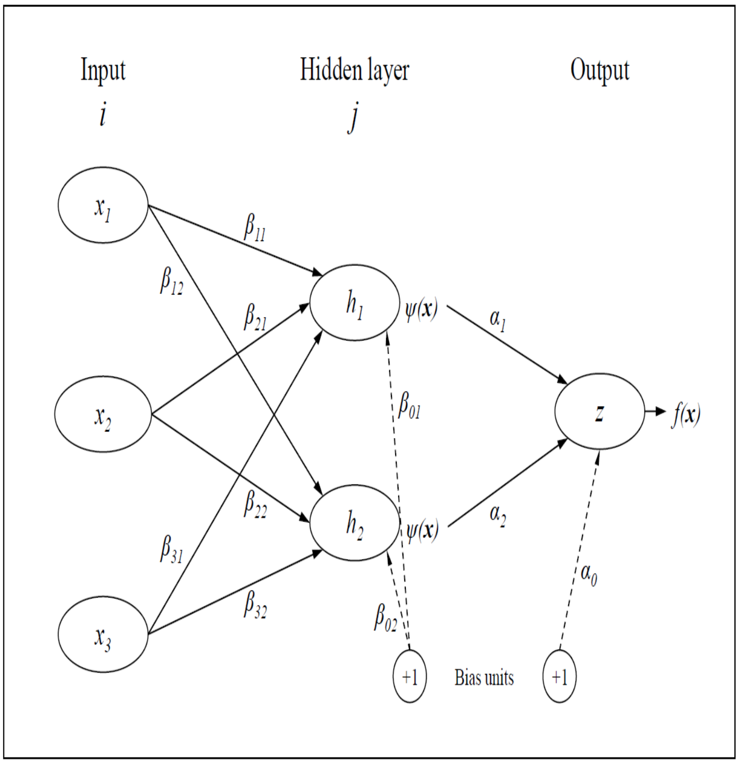

Abbildung 1: Feed-forward neural network mit zwei hidden und einer Output-Einheit; Quelle der Abbildung: Schupp et al. 2015, S. 466

3.3.5 Gewichtung nach „Soll durch Ist“ Prinzip

In dieser Methode wird die vorhandene Information über der Bevölkerungsstruktur hinsichtlich bestimmter Merkmale berücksichtigt. Diese Methode ist in der Marktforschungspraxis ziemlich verbreitet. Die Struktur der Befragten wird mit der Bevölkerungsstruktur hinsichtlich bestimmter Merkmale verglichen und aufgrund dieses Vergleichs die Gewichtungsfaktoren berechnet. Dafür werden einige soziodemografische Merkmale ausgewählt und deren Randverteilungen in der Population und in der Stichprobe miteinander verglichen.

Diese Methode ermöglicht eine Gewichtung anhand eines oder mehreren Merkmale. Um dies zu erreichen, müsste man die Verteilung der wichtigsten soziodemographischen Variablen in der Bevölkerung kennen. Natürlich in der Population sind nicht alle Merkmalsverteilungen bekannt. Außerdem muss man darauf achten, dass die Zellenaufteilung in der Stichprobe Sinn macht. Denn es könnte sein, dass einige Zellen sehr spärliche oder gar keine Besetzungen aufweisen. In solchen Fällen sollen gut überlegte Zellfusionen stattfinden und erst dann die Gewichtung durchführen.

Um einen Bild über die Verzerrung der Gesamtstichprobe zu bekommen, sollte man erst die relative Abweichung der einzelnen soziodemographischen Merkmale untersuchen und die relative Abweichungen schätzen. Die relative Abweichung wird folgendermaßen berechnet

$$ABW_{relij} = (Q_{sij} - Q_{uij}) \cdot \frac{Q_{uij}}{100}$$

wobei i für Merkmal, j für die Ausprägung des Merkmals, Q_{sij} für den Anteil der Ausprägung eines Merkmals in der Stichrobe und Q_{uij} für den Anteil der Ausprägung eines Merkmals in der Bevölkerung steht.

Ein Beispiel aus der Stichprobe der östereichischen Bevölkerung (Kurz 1987, S. 81). Wie es der Tabelle 3 zu entnehmen ist, gibt es Abweichungen in dr Struktur der Bevölkerung und der Stichprobe. Der negative Wert der relativen Abweichung in der Tabelle heißt, dass die zum Beispiel ledige Personen in der Stichprobe unterrepräsentiert sind und der positive Wert der relativen Abweichung heißt, dass die zum Beispiel verheiratete oder in Lebensgemeinschaft lebenden Personen in der Stichprobe überrepräsentiert sind. Die Summierte relative Abweichung pro

	Soziodemographisches Merkmal	Ausprägung in der		Relative
		Stichprobe	Bevölkerung	Abweichung
I	Familienstand	%	%	
	ledig	19,2	29,1	-2,8809
	verheiratet oder in Lebensgemeinschaft	64,1	55,8	+4,6314
	geschieden oder getrennt lebend	3,9	4,3	-0,0172
	verwitwet	12,7	10,8	+0,2052
	Summe	100	100	7,7347
II	Geschlecht	%	%	
	weiblich	56,9	53,5	+1,8190
	mänlich	43,1	46,5	-1,5810
	Summe	100	100	3,4000

Tabelle 2: Struktur der ungewichteten Stichprobe; Quelle der Tabelle: Kurz 1987, S. 81

Merkmal erhält die Auskunft über den Beitrag jedes einzelnen soziodemographischen Merkmals zur Gesamtverzerrung der Stichprobenstruktur.
Die Gewichtung für die einzelne Merkmale oder Merkmalskombinationen erfolgt folgendermaßen (Gabler et al. 1994, S. 67):

$$\hat{q}_i = \frac{Q_{ui}}{Q_{si}} \tag{13}$$

Ein beispielhafte Berechnung des Gewichts für einen Merkmal oder Merkmalskombinationen der Untersuchungseinheiten in der Stichprobe:

$$\hat{q}_{weiblich} = \frac{Weiblich\ in\ der\ Bevölkerung}{Weiblich\ in\ der\ Stichprobe} = \frac{53,5}{56,9} = 0,94$$

$$\hat{q}_{weiblich,ledig} = \frac{Ledig \wedge Weiblich\ in\ der Bevölkerung}{Ledig \wedge Weiblich\ in\ der\ Stichprobe}$$

Es ist sinnvoll die Gewichtung nach unterschiedlichen Merkmalskombinationen durchzuspielen und die gewichteten Stichprobenergebnisse (mit dem geeigneten Test, zum Beispiel Chi quadrat Test) mit dem ungewichteten Stichprobenergebnisse zu vergleichen.

3.3.6 Kalibrierung an externen Randverteilungen

Im vorherigen Kapitel erwähntes Problem der spärlichen oder sogar Null Besetzungen der Zellen in der Stichprobe wird mit anderem Gewichtungsverfahren umgegangen. Das Kalibrierungsverfahren an externen Randverteilungen basiert zwar auch auf „Soll durch Ist“ Prinzip, berücksichtigt jedoch zusätzliche Kriterien mit.

In diesem Verfahren werden mit Hilfe der Anpassungsgewichte die Designgewichte so verändert, dass die bekannte Randverteilungen aus externen Quellen mit Hilfe dieser so kalibrierten Gewichte reproduzierbar sind.

Mit Hilfe der kalibrierten Designgewichte gewichtete Stichprobe soll folgende Bedingung erfüllen:

$$\sum_{k \in S} d_k \cdot g_k \cdot x_k = \sum_{k \in U} x_k$$

$$\text{wobei} \quad w_k = d_k \cdot g_k \quad \text{das kalibrierte Gewicht darstellt.}$$

Das heißt, dass alle Variablen der Stichprobe, welche mit Hilfe der kalibrierten Designgewichte gewichtet werden, die extern vorhanden Randverteilungen der gleichen Variablen reproduzieren.

Allerdings gilt dies nur für die Randverteilungen der Variablen. Die beliebige Verkreuzungen der entsprechenden Variable werden durch die Gewichtung nicht an die gleiche Verteilung aus der externen Quelle angepasst.

Anpassungsgewicht kann geschrieben werden als

$$g_k = 1 + \lambda_s' \cdot x_k \tag{14}$$

wobei

$$\lambda_s' = (\sum_{k \in U} x_k - \sum_{k \in S} d_k \cdot x_k)'(\sum_{k \in S} d_k \cdot x_k \cdot x_k')^{-1}$$

Kalibrierte Gewichte ermöglichen damit einerseits die Reduktion der Varianz und andererseits die Konsistenz mit externen Randverteilungen. Bei der Anwendung der kalibrierten Gewichte muss beachtet werden, dass die kalibrierten Gewichte größer als die Designgewichte sein müssen, um durch die Non-Response verlorene Gewichte kompensiert werden zu können. Außerdem kalibrierte Gewichte sollen so nah wie möglich am Designgewicht sein, was eine wichtige Eigenschaft zur

Minimierung des Non-Response Bias ist. Kalibrierungsschätzer kann geschrieben werden als

$$\hat{\theta}_C = \sum_{k \in s} w_k y_k$$

$$\hat{\theta}_C = \sum_{k \in s} w_k y_k + \underbrace{\sum_{k \in s} d_k y_k}_{\hat{\theta}_{HT}} - \underbrace{\sum_{k \in s} d_k y_k}_{\hat{\theta}_{HT}} = \sum_{k \in s} d_k y_k + \sum_{k \in s} (w_k - d_k) y_k$$

$$Bias_C = E(\hat{\theta}_C - \theta) = E(\sum_{k \in s} d_k y_k + \sum_{k \in s} (w_k - d_k) y_k) - \theta = E(\sum_{k \in s} (w_k - d_k) y_k)$$

Die Minimierung der Distanz $\triangle(w_k, d_k)$ erfolgt über die Minimierung verschiedener Distanzfunktionen. Die sind:

- Lineare Methode

  $$\triangle(w_k, d_k) = \frac{(w_k - d_k)^2}{d_k}$$

  Bei der linearen Methode kann negative Anpassungsgewichte auftreten, was zur Folge hätte, dass die kalibrierte Gewichte auch negativ sind, weil die Designgewichte immer positiv sind. Die Anpassungsgewichte sollen aber strickt positiv sein, um durch das Non-Response verlorene Information zu kompensieren.

- Exponentielle Methode

  $$\triangle(w_k, d_k) = w_k \cdot ln(\frac{w_k}{d_k}) - w_k + d_k$$

  Bei der Benutzung der exponentiellen Methode kann die negative Werte vermieden werden da zur Bestimmung der Anpassungsgewicht die Exponentialfunktion angewendet wird. Der Nachteil dieser Methode liegt aber daran, dass extrem hohe Anpassungsgewichte produziert werden können. Das kann die Ursache extremer kalibrierter Gewichte sein.

- Logit Methode

Die Distanzfunktion der Logit Methode ist gegeben bei

$$\triangle(w_k, d_k) = (\frac{w_k}{d_k} - L) \cdot ln\left(\frac{\frac{w_k}{d_k} - L}{1 - L}\right) + (U - \frac{w_k}{d_k}) \cdot ln\left(\frac{U - \frac{w_k}{d_k}}{U - 1}\right)$$

wobei L eine untere Grenze und U eine obere Grenze für die Anpassungsgewicht darstellt. Es gilt $L < \frac{w_k}{d_k} < U$, welche als Teil der Distanzfunktion implementiert wird. Dadurch kann oben beschriebenes Problem der extremerer Gewichte vermiden werden.

- Trunkierte lineare Methode
  Die ober beschriebene Methode der Begrenzung der Anpasungsgewicht kann auch für die lineare Methode angewendet werden.

3.3.7 Die Politz-Simmons Methode

Diese Methode wurde von Politz und Simmons bereits Ende vierziger Jahren vorgestellt. Dieses Verfahren findet kaum Anwendung in Stichprobenerhebung, aber meiner Ansicht nach ist eine sinnvolle Antwortgewichtung der Untersuchungseinheiten, die schwer oder nicht erreichbar sind.
Für die Gewichtung nach dieser Methode wird die Antwort auf die Frage „an wie vielen der letzten fünf Abende waren Sie zuhause? “benötigt. (Meine Version dieser Frage „An wie vielen der letzten fünf Werktagen waren Sie zuhause?“)
Die Antworten der Untersuchungseinheiten, die an alle fünf Abenden zuhause gewesen wären, werden in der Stichprobe mit geringen Gewicht eingehen. Aber die Antworten der Untersuchungseinheiten, die an keinen von fünf Abenden zuhause gewesen wären, werden höher gewichtet. Entsprechend der angegebenen Antwort werden die Auskunftspersonen in sechs Gruppen t (0, 1, 2, 3, 4 oder 5) aufgeteilt und Die Anzahl dieser Personen in den jeweiligen Gruppen n_t ermittelt. Für die interessierende metrische Variable wird Gruppenmittelwert y_t berechnet.
Schätzformel für nominalskalierte Daten

$$\hat{\hat{\theta}}_{PS} = \frac{\sum_{t=0}^{5} \frac{6}{t+1} \cdot y_t}{\sum_{t=0}^{5} \frac{6}{t+1} \cdot n_t} \tag{15}$$

und für metrische Daten (Kurz 1987, S. 25)

$$\hat{\hat{\theta}}_{PS} = \frac{\sum_{t=0}^{5} \frac{6}{t+1} \cdot n_t \cdot y_t}{\sum_{t=0}^{5} \frac{6}{t+1} \cdot n_t} \tag{16}$$

Die Anwendbarkeit dieser Methode fordert nach Politz und Simmons die Erfüllung folgender Bedingungen:

- Die Befragten müssen die Gewichtungsfrage exakt und richtig beantworten.
- Die Befragten sollten zu zufällig ausgewählten Zeiten besucht werden.

Für die Verdeutlichung des Effekts, den diese Gewichtung auf die Schätzung des interessierenden Parameters hat, werde ich mit folgenden Beispiel verdeutlichen: In dem Beispiel bin ich davon ausgegangen, dass die Untersuchungseinheiten, die schwer erreichbar sind, einen Job haben und höheres Einkommen besitzen als die Untersuchungseinheiten die an allen Abenden erreichbar sind.

Anzahl der Abende	Gewicht	Anzahl der Personen	Anzahl Personen ohne Job	Durchschnittliches Einkommen in Euro
0	6	200	0	3.000
1	3	250	10	2.500
2	2	270	50	2.000
3	1,5	180	90	2.000
4	1,2	60	40	1.000
5	1	40	40	900

Tabelle 3: Konstruiertes Beispiel zur Illustration

Berechnet man den Anteil der Personen ohne Job ohne die Daten zu Gewichten, bekommt man, dass 23% der Population keinen Job hat.

$$\frac{230}{1000} = 23,00\%$$

Bei der Verwendung der Gewichtung sind es aber nur 12,29%.

$$\frac{6 \cdot 0 + 3 \cdot 10 + 2 \cdot 50 + 1,5 \cdot 90 + 1,2 \cdot 40 + 1 \cdot 40}{6 \cdot 200 + 3 \cdot 250 + 2 \cdot 270 + 1,5 \cdot 180 + 1,2 \cdot 60 + 1 \cdot 40} = 12,29\%$$

Bei der Berechnung des durchschnittlichen Einkommens mit und ohne Verwendung der Gewichtung bekommt man den gleichen Ergebnis. Ohne Berücksichtigung der Gewichte wird durchschnittliches Einkommen der Population unterschätzt und beträgt nur 2.221 €.

$$3.000 \cdot \frac{200}{1.000} + 2.500 \cdot \frac{250}{1.000} + \cdots + 1.000 \cdot \frac{60}{1.000} + 900 \cdot \frac{40}{1.000} = 2.221$$

Während bei der Verwendung der Gewichte es bei 2.508 € liegt.

$$\frac{6 \cdot 200 \cdot 3.000 + 3 \cdot 250 \cdot 2.500 + \cdots + 1,2 \cdot 60 \cdot 1.000 + 1 \cdot 40 \cdot 900}{6 \cdot 200 + 3 \cdot 250 + 2 \cdot 270 + 1,5 \cdot 180 + 1,2 \cdot 60 + 1 \cdot 40} = 2.508$$

Mit diesem Beispiel ist deutlich zu sehen, dass die höhere Gewichtung der Untersuchungseinheiten nach Politz-Simons Methode, die schwer oder kaum erreichbar sind, eine sinnvolle Korrigierung liefert.

4 Zusammenfassung

Designgewichtung ist in der Regel unverzichtbarer Bestandteil qualitativ anspruchsvoller Stichproben in der Umfrageforschung. Sonst ist die Bezeichnung „repräsentativ“jedenfalls nicht gerechtfertigt. Das ist bei der Anwendung der anpassungsbasierten Gewichtung aber nicht der Fall.

Vor der Gewichtung der Datensätze für die Reduktion des Non-Response Bias sollte man sich überlegen, ob die Gewichtung Sinn macht und welches Gewichtungsverfahren am Besten geeignet ist. Bei dem Vergleich der anpassungsbasierte Gewichtungen fällt auf, dass eine vollständige Reduzierung des Bias unrealistisch, aber die Erzielung besserer Ergebnisse durchaus realistisch ist. Bei der Anwendung der anpassungbasierte Gewichtung sollte die Gewichtungseffekte ständig kontrolliert, die Gewichtungsschritte detailliert beschrieben und offen dargelegt werden. Man sollte aber auch nicht vergessen, dass „die einzig wirklich repräsentative Stichprobe einer Grundgesamtheit, die Grundgesamtheit selber ist.“(Gabler et al. 1994, S. 128)

Literatur

[1] Gabler, S., J.H.P. Hoffmeyer-Zlotnik, D. Krebs, 1994, *Gewichtung in der Umfragepraxis*: Westdeutscher Verlag

[2] Kauermann, G. und H. Küchenhoff, 2011, *Stichproben*: Springer-Verlag Berlin Heidelberg

[3] Kurz, H., 1987, *Die Genauigkeit von Umfrageergebnissen*: Service-Fachverlag, Wien

[4] Schupp, J. und Ch. Wolf , 2015, *Nonresponse Bias*: Springer Fachmedien Wiesbaden

[5] Särndal, C.E. und Lundström, 2005, *Estimation in Surveys with Nonresponse*: John Wiley and Sons, West Sussex, England

[6] Einbeck, J., T.Augustin und J. M. Singer, 2007, *Smoothing, sampling, and Basu's elephants;* Available at: `http://www.maths.dur.ac.uk/~dma0je/Papers/einbeck_augustin_singer_iwsm07.pdf`, Accessed 21 May 2016

[7] Frick, J. und K. Krell, 2009, *Einkommensmessungen in Haushaltspanelstudien für Deutschland: Ein Vergleich von EU-SILC und SOEP;* Available at: `https://www.diw.de/documents/publikationen/73/diw_01.c.343744.de/diw_sp0237.pdf`, Accessed 29 April 2016

[8] Günther, F. und St. Fritsch, 2010, *Neuralnet: Training of Neural Networks;* Available at: `https://journal.r-project.org/archive/2010-1/RJournal_2010-1_Guenther+Fritsch.pdf`, Accessed 27 April 2016

[9] *Visualizing neural networks in R;* Available at: `https://beckmw.wordpress.com/tag/nnet/`, Accessed 30 May 2016

[10] Lohmann-Haislah A., 2012, *Stressreport Deutschland 2012;* Available at: `http://www.baua.de/de/Publikationen/Fachbeitraege/Gd68.pdf?__blob=publicationFile`, Accessed 01 June 2016

[11] Münnich, R. und S. Gabler, 2012, *Stichprobenoptimierung und Schätzung im Zensus 2011;* Available at: `https://www.zensus2011.de/SharedDocs/`

Downloads/DE/Publikationen/Monografien_Archiv/2012_07_Destatis_Stochprobenoptimierung_und_Schaetzung_im_Zensus.pdf?__blob=publicationFile&v=12, Accessed 29 April 2016

[12] Osier G., J. Museux und P. Seoane, 2006, *Cross-sectional and longitudinal weighting for the EU- SILC rotational design;* Available at: https://www.iser.essex.ac.uk/files/survey/ulsc/methodological-research/mols-2006/scientific-social-programme/papers/Osier.pdf, Accessed 01 May 2016

[13] Rosenbaum, Paul R. 1983, *The central role of the propensity score in observational studies for causal effects;* Available at: http://biomet.oxfordjournals.org/content/70/1/41.full.pdf+html, Accessed 25 April 2016