

Gewichtung und Validierung der Nettostichprobe (Welle 1 des Projekts „Lebensziele und Lebensverläufe in Ostdeutschland“)

Gewichtung

Da es sich bei dieser Studie um eine Einwohnermeldeamtsstudie handelt, bei der alle Mitglieder der Grundgesamtheit dieselbe Auswahlwahrscheinlichkeit hatten, ist eine Designgewichtung, bei der ungleiche Auswahlwahrscheinlichkeiten aufgrund des Stichprobendesigns ausgeglichen werden müssen, nicht notwendig gewesen.

Durchgeführt wurde jedoch eine Ausfallgewichtung (Kalibration). Hierbei werden Stichprobenausfälle ausgeglichen, um den Bias zu korrigieren. Die Gewichtungsfaktoren werden so berechnet, dass die Daten in den Gewichtungsvariablen nach der Gewichtung der Grundgesamtheit entsprechen. Diese Strukturen werden meist der offiziellen Statistik entnommen. Durch die Anpassung der Gewichtungsvariablen soll die Präzision der Ergebnisschätzer der interessierenden Variablen gesteigert werden. Dies wird dann erreicht, wenn die Gewichtungsvariablen mit den interessierenden Variablen korrelieren. Daher ist genau zu überlegen, welche Variablen sich als Gewichtungsvariablen eignen.

Für die Ausfallgewichtung gibt es verschiedene Gewichtungsverfahren. Hier wurde das iterative Randsummenverfahren (Iterative Proportional Fitting) verwendet. Zusätzlich wurden die Gewichtungsfaktoren so ermittelt, dass sie im Intervall von 0,2 bis 4,9 liegen (TNS-Standard).

Es wurden zwei Gewichtungen durchgeführt:

- (1) nur die Stichprobe der Welle 1 des Projekts „Lebensziele und Lebensverläufe in Ostdeutschland“ (N = 1.489)
- (2) die Stichprobe der Welle 1 des Projekts „Lebensziele und Lebensverläufe in Ostdeutschland“ gemeinsam mit der Welle 1 des Projekts „Beziehungen und Familienleben in Deutschland“ (N = 13.891)

Für beide Gewichtungen entstammen die Gewichtungsstrukturen dem Mikrozensus 2009. Die Kohorten wurden jeweils getrennt gewichtet.

(1) Gewichtung der Welle 1 des Projekts „Lebensziele und Lebensverläufe in Ostdeutschland“

Die Kalibration erfolgte anhand der Variablen Bundesland, Alter, Geschlecht, BIK-Typ (zehn Gruppen), Familienstand (in den Ausprägungen ledig, verheiratet, geschieden, verwitwet). Gewichtet wurde dabei nach sieben Rändern:

- Bundesland x Geschlecht x Alter (36 Zellen)
- Bundesland x BIK (30 Zellen)
- BIK (10 Zellen)

- Bundesland (6 Zellen)
- Familienstand x Geschlecht (7 Zellen)
- Familienstand x BIK (24 Zellen)
- Geschlecht x Alter (6 Zellen)

Aufgrund der nicht ausgewiesenen Werte im Mikrozensus wurden in beiden Kohorten die Kategorien ‚geschieden‘ und ‚verwitwet‘ als eine Kategorie betrachtet.

Aufgrund der Datenschutzbestimmungen des Mikrozensus dürfen nur Werte über 5.000 ausgewiesen werden – dies hat zur Folge, dass einige Bundesland-BIK-Kombinationen nicht ausgewiesen werden konnten. In diesen Fällen wurden die Kombinationen mit naheliegenden Kombinationen zusammengefasst (wobei grundsätzlich zwei oder mehr BIK-Typen im selben Bundesland zusammengefasst wurden). Dasselbe gilt auch für den Rand Familienstand x BIK.

Durch die Kalibrationsrechnung konnten die Daten pro Kohorte wie folgt an die Vorgaben angepasst werden:

	Kohorte 2	Kohorte 3
Bundesland x Geschlecht x Alter	99,00%	99,00%
Bundesland x BIK ¹	98,92%	97,96%
BIK	98,80%	99,16%
Bundesland	99,93%	99,68%
Familienstand x Geschlecht	99,08%	99,57%
Familienstand x BIK	99,79%	99,85%
Geschlecht x Alter	100,00%	99,99%
Effektivität ²	70%	63%

TNS Infratest Sozialforschung 2011

¹ Die relativ niedrige Anpassung im Rand Bundesland x BIK in allen beiden Kohorten lässt sich durch die oben beschriebenen Schwierigkeiten in den Strukturzahlen erklären.

² Die Effektivität einer Gewichtung ist ein Maß für die Abweichung von 1,0 – d.h. für die Abweichung von den ungewichteten Strukturen. Je höher die Effektivität ist, umso besser ist die Qualität der Netto-Stichprobe. Effektivitäten über 70% sind als sehr gut anzusehen. Die Effektivität berechnet sich wie folgt:

$$E = \frac{(\sum_1^n g_n)^2}{(\sum_1^n (g_n^2)) * n}$$

mit $E = \text{Effektivität}$ ($0 \leq E \leq 1$)

$g_n = \text{Gewicht / Gewichtungsfaktor für n-ten Fall}$

$n = \text{Fallzahl (ungewichtet)}$

(2) Gemeinsame Gewichtung mit der Stichprobe der Welle 1 des Projekts „Beziehungen und Familienleben in Deutschland“

Die Kalibration erfolgte anhand derselben Variablen wie in der separaten Gewichtung. Gewichtet wurde dabei ebenso nach sieben Rändern:

- Bundesland x Geschlecht x Alter (86 Zellen)
- Bundesland x BIK (91 Zellen)
- BIK (10 Zellen)
- Bundesland (16 Zellen)
- Familienstand x Geschlecht (8 Zellen³)
- Familienstand x BIK (31 Zellen)
- Geschlecht x Alter (6 Zellen)

Bei der jüngsten Kohorte wurden die beiden Ränder, die den Familienstand beinhalten, nicht in die Gewichtung einbezogen, da die 15-17jährigen fast ausschließlich ledig sind.

Aufgrund der Datenschutzbestimmungen des Mikrozensus dürfen nur Werte über 5.000 ausgewiesen werden – dies hat zur Folge, dass einige Bundesland-BIK-Kombinationen nicht ausgewiesen werden konnten. In diesen Fällen wurden die Kombinationen mit naheliegenden Kombinationen zusammengefasst (wobei grundsätzlich zwei oder mehr BIK-Typen im selben Bundesland zusammengefasst wurden).

Durch die Kalibrationsrechnung konnten die Daten pro Kohorte wie folgt an die Vorgaben angepasst werden:

	Kohorte 1	Kohorte 2	Kohorte 3
Bundesland x Geschlecht x Alter	99,51%	99,54%	99,47%
Bundesland x BIK	98,90%	99,26%	99,06%
BIK	99,93%	98,84%	99,93%
Bundesland	99,99%	99,97%	99,96%
Familienstand x Geschlecht	-	99,86%	99,94%
Familienstand x BIK	-	99,98%	99,96%
Geschlecht x Alter	100,00%	100,00%	100,00%
Effektivität	85%	73%	74%

TNS Infratest Sozialforschung 2011

Die gemeinsame Gewichtung mit den Pairfam-Daten ist insofern problematisch, als die Daten zu unterschiedlichen Zeitpunkten erhoben wurden, als Referenz aber den Mikrozensus 2009 aufweisen. Vor allem in Hinsicht auf den Familienstand, der sich besonders in den Kohorten 2 und 3 auch innerhalb kurzer Zeiträume ändern kann, sollte dies beachtet werden. Zudem ist

³ Je nachdem, was der Mikrozensus jeweils an Strukturdaten ausgewiesen hat, musste der Familienstand unterschiedlich zusammengefasst werden. In manchen Fällen musste geschieden und verwitwet zusammengefasst werden, in manchen wiederum nicht.

die Gesamtstichprobe durch den höheren Anteil an ostdeutschen Befragten nun disproportional.⁴

Nettovalidierung

Die Validierung beinhaltet den Abgleich der Merkmalsverteilung in der Stichprobe mit einer Referenzstatistik, von der angenommen wird, dass sie die Grundgesamtheit richtig abbildet. Ein solcher Vergleich sollte berücksichtigen, ob in beiden Datenquellen bei den Merkmalen Effekte unterschiedlicher Fragen- und Messkonzepte auf die Verteilung der Merkmale weitgehend ausgeschlossen werden können. Diese Bedingung ist, soweit man die Ergebnisse der amtlichen Statistik als Referenz heranzieht, in der Regel nur für wenige Merkmale der Sozialstruktur erfüllt.

Folgende soziodemographische Merkmale sind hier Gegenstand der Betrachtung:

- Geschlecht
- Familienstand
- Gemeindegrößenklasse (BIK)
- Bundesland

Der Abgleich mit der amtlichen Statistik ist für diesen Methodenbericht dergestalt möglich, als dass die Daten des Mikrozensus 2009 Grundlage dieser Analysen sind.

Tabelle 1: Anteil männlicher Bevölkerung nach Bundesland (in %)

	MPI (ungew.)		MPI (gew.)		Mikrozensus 2009	
	1981-1983	1971-1973	1981-1983	1971-1973	1981-1983	1971-1973
Ostdeutschland gesamt	49,3	49,2	53,5	51,8	53,4	51,8
Ost-Berlin	45,6	47,3	54,0	52,7	53,5	52,2
Brandenburg	57,6	43,0	54,6	51,2	54,5	50,5
Mecklenburg-Vorpommern	46,9	52,4	51,9	56,0	50,9	56,5
Sachsen	48,6	46,4	51,9	51,9	52,1	51,7
Sachsen-Anhalt	50,0	52,5	54,6	48,7	55,3	50,0
Thüringen	44,8	56	55,2	51,8	55,4	50,6

Quelle: Mikrozensus 2009, MPI 2010

TNS Infratest Sozialforschung 2011

Was sich zeigt, ist, dass je nach Bundesland und Kohorte der Anteil der Männer in der ungewichteten Stichprobe teilweise deutlich von der Verteilung laut Mikrozensus abweicht. Diese Verzerrungen können aber durch die Gewichtung weitgehend ausgeglichen werden.

⁴ Die Ost-West-Disproportionalität wird dadurch, dass u.a. auf den Rand Bundesland gewichtet wurde, weitestgehend automatisch ausgeglichen.

Tabelle 2: Familienstand und BIK (in %)

	MPI (ungew.)		MPI (gew.)		Mikrozensus 2009 ⁵	
	25-27	35-37	25-27	35-37	25-27	35-37
Gesamt						
Ledig	78,2	34,0	83,4	42,9	83,7	42,8
Verheiratet	19,2	55,3	15,3	49,4	15,7	49,5
Verwitwet	0,3	0,4	0,1	0,3	-	-
Geschieden	2,1	10,3	0,9	7,5	-	7,3
BIK 0 (Kern 500.000+)						
Ledig	76,1	37,2	85,6	53,3	87,2	52,9
Verheiratet	20,1	51,1	11,9	40,0	12,1	39,7
Verwitwet	-	-	-	-	-	-
Geschieden	3,0	11,7	1,5	6,7	-	7,4
BIK 1 (Rand 500.000+)						
Ledig	80,6	27,0	82,5	30,9	82,8	32,5
Verheiratet	19,4	62,2	17,5	60,0	17,2	60,0
Verwitwet	-	2,7	-	1,8	-	-
Geschieden	-	8,1	-	7,3	-	-
BIK 2 (Kern 100.000-u. 500.000)						
Ledig	81,9	34,1	83,6	48,2	85,6	48,1
Verheiratet	17,1	52,9	15,6	42,9	15,6	43,2
Verwitwet	-	-	-	-	-	-
Geschieden	1,0	12,9	0,8	8,9	-	7,4
BIK 3 (Rand 100.000-u. 500.000)						
Ledig	78,3	27,6	82,8	36,8	82,6	36,0
Verheiratet	21,7	67,2	17,2	58,8	17,4	58,0
Verwitwet	-	1,7	-	1,5	-	-
Geschieden	-	3,4	-	2,9	-	-
BIK 4 (Kern 50.000-u. 100.000)						
Ledig	80,4	35,3	79,5	50,0	75,0	50,0
Verheiratet	15,7	54,9	20,5	46,7	21,4	45,5
Verwitwet	2,0	-	0,0	-	-	-
Geschieden	2,0	9,8	0,0	3,3	-	-

⁵ Aufgrund der unterschiedlichen Zellenbesetzung liegen nicht für jede Zelle die entsprechenden Strukturdaten vor.

	MPI (ungew.)		MPI (gew.)		Mikrozensus 2009	
	25-27	35-37	25-27	35-37	25-27	35-37
BIK 5 (Rand 50.000- u. 100.000)						
Ledig	74,5	34,6	82,7	33,7	81,5	33,3
Verheiratet	16,4	57,7	16,0	59,3	16,7	58,7
Verwitwet	-	-	-	-	-	-
Geschieden	9,1	7,7	1,3	7,0	-	-
BIK 6 (20.000-u. 50.000)						
Ledig	76,6	30,0	82,7	37,4	83,1	37,5
Verheiratet	21,3	58,5	15,3	52,5	15,5	52,8
Verwitwet	0,7	0,8	1,0	0,0	-	-
Geschieden	1,4	10,8	1,0	10,1	-	8,3
BIK 7 (5.000-u. 20.000)						
Ledig	73,1	39,3	79,6	43,9	81,1	43,9
Verheiratet	23,1	51,8	18,5	49,1	18,9	48,8
Verwitwet	-	-	-	-	-	-
Geschieden	2,8	8,9	1,9	7,0	-	-
BIK 8 (2.000-u. 5.000)						
Ledig	93,3	41,4	81,3	46,2	82,6	40,7
Verheiratet	6,7	43,1	18,8	46,2	-	48,1
Verwitwet	-	-	-	-	-	-
Geschieden	-	15,5	-	7,7	-	-
BIK 9 (u. 2.000)						
Ledig	83,9	11,1	85,0	34,6	85,7	40,0
Verheiratet	16,1	77,8	15,0	57,7	-	55,0
Verwitwet	-	-	-	-	-	-
Geschieden	-	11,1	-	7,7	-	-

Quelle: Mikrozensus 2009, MPI 2009

TNS Infratest Sozialforschung 2011

Diese Tabelle zeigt, dass sich – ähnlich wie auch bei Pairfam – vor allem in den Großstädten (BIK 0⁶ und BIK 1) besonders Personen, die bereits in festen Partnerschaften leben, vom Thema der Befragung angesprochen fühlen. Ledige sind hier unterrepräsentiert. Aber auch dies kann durch die Gewichtung mit dem Mikrozensus als Referenz gut korrigiert werden.

Da der Mikrozensus 2009 keine Angaben zur Kinderlosigkeit von Frauen ausweist, konnte dies nicht als zusätzlicher Gewichtungparameter berücksichtigt werden.

⁶ In der Stichprobe des Projekts „Lebensziele und Lebensverläufe in Ostdeutschland“ ist jedoch Berlin die einzige Stadt mit BIK=0.