

gesis

Leibniz-Institut
für Sozialwissenschaften



Varianzschätzung von Nettoveränderungen mit dem Mikrozensus ab 2012

9. Nutzerkonferenz "Forschen mit dem Mikrozensus"
Mannheim, 28.11.2018

Bernhard Schimpl-Neimanns

















Gliederung

- 1 Einleitung
- 2 Mikrozensus Scientific Use Files ab 2012
- 3 Schätzung von Nettoveränderungen
 - 3.1 Varianzschätzung für Querschnittsdaten
 - 3.2 Varianzschätzung für Nettoveränderungen
- 4 Ergebnisse für 3 ausgewählte Indikatoren
- 5 Schluss

1 Einleitung

- Vorteile des Mikrozensus (MZ) für Sozialberichterstattung
 - Große Stichprobe; MZ-Scientific Use-Files (SUF): 0,7 %
 - Jährlicher Erhebungszyklus - Trendstudien
 - Standardfehler als Gradmesser statistischer Unsicherheit auch für Längsschnitt wichtig.
- MZ-Stichprobendesign
 - Geschichtete einstufige Flächen-/Klumpenstichprobe
 - Partielle Rotation der Auswahlbezirke (AB)
- Scientific-Use-Files ab 2012 – neue Analysemöglichkeiten
 - Auswahleinheiten: Auswahlbezirke
 - Längsschnittkonsistente Ordnungsnummern (IDs)
 - Nettoveränderungen und Erstellung von Rotationspanels

Rotationschema im Mikrozensus

Stichproben- nummer	Rotations- gruppe	Erhebungsjahr			
		2012	2013	2014	2015
7	3				
7	4				
8	1				
8	2				
8	3				
8	4				
9	1				

Abhängige Stichproben

Nettoveränderung:

$$\hat{\theta}_{\Delta} = \hat{\theta}_2 - \hat{\theta}_1 \quad \text{Var}(\hat{\theta}_{\Delta}) = \text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2) - 2 \text{Cov}(\hat{\theta}_1, \hat{\theta}_2)$$

Kumulation:

$$\hat{\theta}_{\Sigma} = \hat{\theta}_1 + \hat{\theta}_2 \quad \text{Var}(\hat{\theta}_{\Sigma}) = \text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2) + 2 \text{Cov}(\hat{\theta}_1, \hat{\theta}_2)$$

2 MZ Scientific Use Files ab 2012

Substichprobenziehung (siehe [readme](#) zum MZ 2012)

- Weitestgehend analog zum MZ-Stichprobenplan und in Anlehnung an MZ-Panel 1996-1999, 2001-2004.
- Getrennt für Privathaushalte und Gemeinschaftsunterkünfte.
- Auswahlatz 70 % in allen Schichten.
 - In t ausgewählte AB sind auch in $t+1$.
 - Aktualisierung: AB aus Neubausauswahl und neue Rotationsviertel.

Hochrechnung

- Poststratifizierte Anpassungsgewichte
 - Anpassungsklassen: Bundesland \times Geschlecht \times Altersgruppe \times Erwerbsstatus.

3 Schätzung von Nettoveränderungen

- Schichtung: Bundesland, Gebäudegrößenklasse, Quartal.
 - MZ-Stichprobenplan: Keine unterjährige Rotation, Rotation der Auswahlbezirke innerhalb der Schichten.
 - Praxis bei quartalsweiser Hochrechnung: Berichtsquartal realisierter Interviews (Afentakis & Bihler 2005).
 - Für Jahresdurchschnitte: Verzicht auf Berichtsquartal zur Vermeidung von Schichtwechseln in einem AB bei verspäteten Antworten (u. a. Jahresüberhängen).
- Hochrechnung mit Anpassungsgewichten
 - Regressionsschätzung (Deville 1999; Särndal et al. 1997)
→ Linearisierung von Verhältniswerten.

3.1 Varianzschätzung für Querschnittsdaten

Regressionsschätzung für Verhältniswerte

$$\hat{R} = \hat{Y} / \hat{Z} = \sum_{j \in S} w_j y_j / \sum_{j \in S} w_j z_j \quad (1)$$

Regression des linearisierten Verhältniswertes $(y_{hij} - \hat{R}z_{hij}) * 1/\hat{Z}$ auf Anpassungsmerkmale $\mathbf{x}_{hij} \rightarrow$ Residuum $e_{hij} \rightarrow$ Hilfsmerkmal für Varianzschätzung: $u_{hij} = e_{hij} * g_{hij}$.

$$\hat{V}(\hat{R}) = \sum_{h=1}^L n_h / (n_h - 1) \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)^2 \quad (2)$$

h Schicht $h = 1, \dots, L$

i Primäreinheit (PSU; Auswahlbezirk)

j Sekundäreinheit (Person) in PSU i

w Endgewicht

y, z Analysevariablen

f Auswahlatz (0,7 %)

$d_j = 1/f$ Designgewicht

$g_j = w_j/d_j$ Korrekturfaktor

$u_{hi} = \sum_{j=1}^{m_{hi}} d_{hij} u_{hij}$ PSU-Gesamtwert

$\bar{u}_h = 1/n_h \sum_{i=1}^{n_h} u_{hi}$ PSU-Mittelwert

\mathbf{x}_j Anpassungsmerkmale (Poststratifik.)

3.2 Varianzschätzung für Nettoveränderungen

Schätzung der Kovarianz mittels multivariater Regression.

Annahmen: Positive Korrelation der Schätzer, einfache Zufallsauswahl der PSUs (Berger & Priam 2016).

$$\hat{\Delta} = \hat{\theta}_2 - \hat{\theta}_1 \tag{3}$$

$$\begin{pmatrix} \hat{t}_{1i} \\ \hat{t}_{2i} \end{pmatrix} = \begin{pmatrix} \beta_1^{(1)} z_{1i} + \beta_2^{(1)} z_{2i} + \beta_{12}^{(1)} z_{1i} z_{2i} \\ \beta_1^{(2)} z_{1i} + \beta_2^{(2)} z_{2i} + \beta_{12}^{(2)} z_{1i} z_{2i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix}$$

mit $\hat{t}_{1i} = \sum_{i \in S_1} u_{ij}$, $\hat{t}_{2i} = \sum_{i \in S_2} u_{ij}$

$z_{1i}, z_{2i}, z_{1i}z_{2i}$

Interaktion Schichtung × Rotation

- Korrelation der Residuen der multivariaten Regression $\hat{\rho}_\Sigma \rightarrow$ Schätzer für $\rho(\hat{t}_{1i}, \hat{t}_{2i})$.

$$V(\hat{\Delta}) = V(\hat{\theta}_1) + V(\hat{\theta}_2) - 2\sqrt{V(\hat{\theta}_1)}\sqrt{V(\hat{\theta}_2)}\hat{\rho}_\Sigma \tag{4}$$

4 Ergebnisse für ausgewählte Indikatoren

1 Erwerbstätigenquote für 60- bis 64-Jährige

- *Erhöhung des Alters des effektiven Renteneintritts, alterungsbedingter Fachkräftemangel.*
- *Erwerbstätigenquote für 20- bis 64-Jährige ist Leitindikator der Strategie Europa 2020.*

2 Anteil frühzeitiger Schul- und Ausbildungsabgänger unter 18- bis 24-Jährigen

- *Gruppe mit überdurchschnittlichen Schwierigkeiten auf dem Arbeitsmarkt.*
- *Leitindikator der Strategie Europa 2020.*

3 Teilzeitquote abhängig beschäftigter Frauen im Alter von 15 bis 64 Jahren

- *Definition Teilzeit: <32 Normale Arbeitsstd./Woche.*
- *Teilzeitarbeit als Chance, Familie und Beruf besser zu vereinbaren. Zunahme im Zeitverlauf.*

4 Ergebnisse für ausgewählte Indikatoren

Tab. 1 Nettoveränderungen (%) – Annahme: abhängige / *unabhängige* Stichproben

Indikator (Kurzbez.)	2013	2012	Δ	SE	Diff. (%)	95 %-KI	sig.
1 Erwerbstätige 60-64 J.	49,7	46,7	3,0	0,2420		2,5 3,5	*
				<i>0,3272</i>	<i>35,2</i>	<i>2,4 3,6</i>	*
2 Frühe Schulabgänger	9,9	10,4	-0,5	0,2147		-0,9 -0,1	*
				<i>0,2656</i>	<i>23,7</i>	<i>-1,0 -0,0</i>	*
3 Teilzeit Frauen	46,2	45,7	0,5	0,1734		0,1 0,8	*
				<i>0,2410</i>	<i>39,0</i>	<i>-0,0 0,9</i>	—

Quelle: FDZ der Statistischen Ämter des Bundes und der Länder, MZ 2012 und 2013, eigene Berechnungen. Bevölkerung am Ort der Hauptwohnung, hochgerechnete Werte. SE: Standardfehler, Diff.: Relative Differenz SE unabh. vs. abh. Stichpr. ; KI: Konfidenzintervall.

Schluss

- Berücksichtigung des Stichproben- und Rotationsdesigns erlaubt effiziente Schätzung für Nettoveränderungen.
- Bei nicht zutreffender Annahme statistischer Unabhängigkeit:
 - Größere Standardfehler.
 - Ggf. falsche Schlussfolgerungen zur statistischen Signifikanz.
- Umsetzung des Verfahrens von (Berger & Priam 2016) ist ohne großen Mehraufwand möglich; auch für Gesamt- und Mittelwerte.
- Verbesserungsmöglichkeiten
 - Bereitstellung des nach dem Stichprobenplan vorgesehenen Quartals.

Literatur

- Afentakis A, Bihler W (2005) Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. [Wirtschaft und Statistik \(10\): 1039-1048](#).
- Berger Y G, Priam R (2016) A simple variance estimator of change for rotating repeated surveys: an application to the European Union Statistics on Income and Living Conditions household surveys. *Journal of the Royal Statistical Society: Series A*, 179: 251–272. doi:10.1111/rssa.12116.
- Bihler W, Zimmermann D (2016) Die neue Mikrozensusstichprobe ab 2016. [Wirtschaft und Statistik \(6\): 20-29](#).
- Deville, J C (1999) Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25: 193-203.
- Herter-Eschweiler R, Schimpl-Neimanns B (2018) [Möglichkeiten der Verknüpfung von Mikrozensus-Querschnitterhebungen ab 2012 zu Panels](#). Bonn/Mannheim.
- Särndal, C-E, Swensson B, Wretman J (1997) *Model Assisted Survey Sampling*. New York: Springer.
- Schimpl-Neimanns B (2011) Schätzung des Stichprobenfehlers in Mikrozensus Scientific Use Files ab 2005. *Wirtsch Sozialstatist Arch* 5 (1):19-38.
- Statistisches Bundesamt, GESIS (Hrsg.) (2016) [Wichtige Informationen zur Nutzung des Mikrozensus Scientific Use Files 2012](#). Bonn/Mannheim.

Vielen Dank für Ihre Aufmerksamkeit.

Für hilfreiche Anmerkungen danke ich Robert Herter-Eschweiler und Heike Wirth.

gesis

Leibniz-Institut
für Sozialwissenschaften



Bernhard Schimpl-Neimanns
GESIS – Leibniz-Institut für Sozialwissenschaften
German Microdata Lab (GML)
Quadrat B2, 1
D-68159 Mannheim
E-Mail: bernhard.schimpl-neimanns@gesis.org