

Alexander Kowarik
Statistik Austria
Mai, 2016

Mikrozensus - Stichprobe, Hochrechnung, Fehlerrechnung

- ▶ “Kleine Volkszählung”
- ▶ Die größte Stichprobenerhebung Österreichs
- ▶ Haushaltsbefragung in Privathaushalten mit zumindest einer im ZMR mit Hauptwohnsitz gemeldeten Person
- ▶ Auskunftspflicht (Fremdauskunft möglich)
- ▶ Pro Quartal ca. 20.000 Haushalte bzw. 44.000 Personen

- ▶ Fünftel-Rotation der Quartalerhebung (20% neue Haushalte pro Quartal)
- ▶ Erstbefragung Face-to-Face (CAPI), 4 Folgebefragungen (wenn möglich) telefonisch (CATI)
- ▶ Zuweisung einer Referenzwoche bei der Stichprobenziehung
- ▶ Referenzwochen sind gleichmäßig über das Quartal verteilt
- ▶ Befragung muss spätestens 5 Wochen nach Ende der Referenzwoche erfolgen

- ▶ **Wohnungserhebung**
 - ▶ Rechtsverhältnis, Größe, Baualter, Ausstattung, Kosten, ...
 - ▶ → Jährliche Wohnungsstatistik und Wohnungsaufwand im VPI
- ▶ **Familien- und Haushaltsstatistik**
 - ▶ Haushaltsgröße, Familientyp, Anzahl der Kinder, ...
 - ▶ → Jährliche Familien- und Haushaltsstatistik
- ▶ **Arbeitskräfteerhebung**
 - ▶ Erwerbsstatus nach ILO Definition, berufliche Stellung, Beruf, Branche, Arbeitszeit, ...
 - ▶ → Monatliche und Quartalsweise Arbeitsmarktstatistik

- ▶ 2010: Vereinbarkeit von Beruf und Familie
- ▶ 2011: Beschäftigung behinderter Personen
- ▶ 2012: Übergang vom Berufsleben in den Ruhestand
- ▶ 2013: Arbeitsunfälle und berufsbedingte Gesundheitsprobleme
- ▶ 2014: Arbeitsmarktsituation von Zuwanderern und ihren direkte Nachkommen
- ▶ 2015: Arbeitsorganisation und Arbeitszeitgestaltung
- ▶ 2016: Junge Menschen auf dem Arbeitsmarkt

- ▶ Grundgesamtheit: Österreichische Wohnbevölkerung in Privathaushalten
- ▶ Auswahlrahmen: Haushalte (bzw. Adressen) im ZMR (Anstalten werden ausgeschlossen)
- ▶ Erhoben werden alle im HH lebenden Personen (unabhängig von Melderealität)
- ▶ ...

EU-Verordnung:

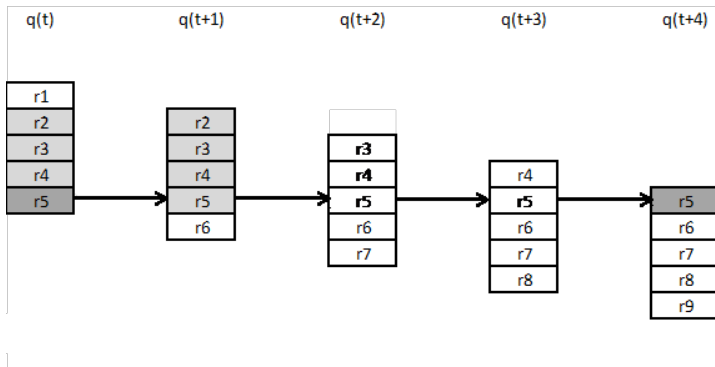
- ▶ maximal 8% relativer Standardfehler für Jahresdurchschnittswerte für Bundesländer (Für eine Bevölkerungsgruppe >5%, z.B. Arbeitslosigkeit)
- ▶ Veränderung einer Bevölkerungsgruppe soll maximal 3% der betreffenden Bevölkerungsgruppen betragen

⇒

- ▶ Stichprobenumfänge für alle Bundesländer ähnlich gleich groß
- ▶ Rotation zur Erhöhung der Genauigkeit für die Schätzung von Veränderungen

- ▶ Geschichtete Zufallsstichprobe von Adressen
- ▶ Disproportionale Allokation auf die Bundesländer
- ▶ Für Personenmerkmale → Clustereffekt
- ▶ 20.000 Haushalte od. 44.000 Personen (0.5 % der Bevölkerung)
- ▶ Jedes Quartal werden 4.000 Adressen als Auffrischungstichprobe gezogen

- Die Jahrestichprobe besteht aus ca. 80.000 Haushalten, wobei 32.000 nicht mehrfach vorkommen.



Verteilung der Quartalsstichprobe

Bundesland	Stp-HH	GGH-HH	HH-%	Stp-P	GGH-P	P-%
Burgenland	1,600	122,000	1.3	3,500	287,000	1.2
Niederösterreich	2,300	708,000	0.3	5,300	1,629,000	0.3
Wien	2,800	893,000	0.3	5,600	179,300	0.3
Kärnten	2,100	249,000	0.9	4,500	552,000	0.8
Steiermark	2,300	534,000	0.4	5,100	1,210,000	0.4
Oberösterreich	2,300	619,000	0.4	5,300	1,428,000	0.4
Salzburg	2,100	235,000	0.9	4,600	535,000	0.9
Tirol	2,100	316,000	0.7	4,800	724,000	0.7
Vorarlberg	2,000	162,000	1.2	4,600	378,000	1.2

Designgewicht

Bundesland	Stp-HH	GGH-HH	HH-%	Gewicht
Burgenland	1,600	122,000	1.3	78
Niederösterreich	2,300	708,000	0.3	304
Wien	2,800	893,000	0.3	321
Kärnten	2,100	249,000	0.9	117
Steiermark	2,300	534,000	0.4	230
Oberösterreich	2,300	619,000	0.4	268
Salzburg	2,100	235,000	0.9	114
Tirol	2,100	316,000	0.7	152
Vorarlberg	2,000	162,000	1.2	81

- ▶ Auswertung ohne Gewichtung
- ▶ \Rightarrow eigentlich verwendetes Gewicht $\frac{19,600}{3,838,000} = 196$
- ▶ \Rightarrow B, K, S, T und V überrepräsentiert
- ▶ \Rightarrow NÖ, W, ST und OÖ unterrepräsentiert

- ▶ Anpassung an externe Information (z.b.:
 - ▶ dass die geschätzte Anzahl an Personen zwischen 20 und 25 den exakt bekannten Werten entspricht.)
- ▶ Verbesserung der Schätzung
- ▶ Verkleinerung des Fehlers
- ▶ Minimierung des Non-Response-Bias

	y_1	y_2	geschätzte Eckzahl	Exakte Eckzahl
x_1	10	20	30	20
x_2	20	10	30	40
geschätzte Eckzahl	30	30	60	
exakte Eckzahl	40	20		

Wir suchen neue Gewichte, die folgende Kalibrierungsgleichungen erfüllen:

$$\sum_{k|x_k=x_1} w_k = N_{1*} = 20 \qquad \sum_{k|x_k=x_2} w_k = N_{2*} = 40$$

$$\sum_{k|y_k=y_1} w_k = N_{*1} = 40 \qquad \sum_{k|y_k=y_2} w_k = N_{*2} = 20$$

Rechenbeispiel mit Raking-Ratio/Iterative Proportional Fitting (Updating):

Schritt 1:

$\frac{Y}{X}$			gesch.EZ
	10	20	30
	20	10	30
gesch.EZ	30	30	

Multipliziere mit

$$\times \frac{2}{3} \rightarrow 20$$

$$\times \frac{4}{3} \rightarrow 40$$

Schritt 2:

$\frac{Y}{X}$			gesch.EZ
	$\frac{20}{3}$	$\frac{40}{3}$	20
	$\frac{80}{3}$	$\frac{40}{3}$	40
gesch.EZ	33.33	26.67	

Multipliziere mit:

$$\times \frac{6}{5} \rightarrow 40$$

$$\times \frac{3}{4} \rightarrow 20$$

Schritt 3:

$\frac{Y}{X}$			gesch.EZ
	8	10	18
	32	10	42
gesch.EZ	40	20	

Multipliziere mit

$$\times \frac{20}{18} \rightarrow 20$$

$$\times \frac{40}{42} \rightarrow 40$$

... Schritt 9:

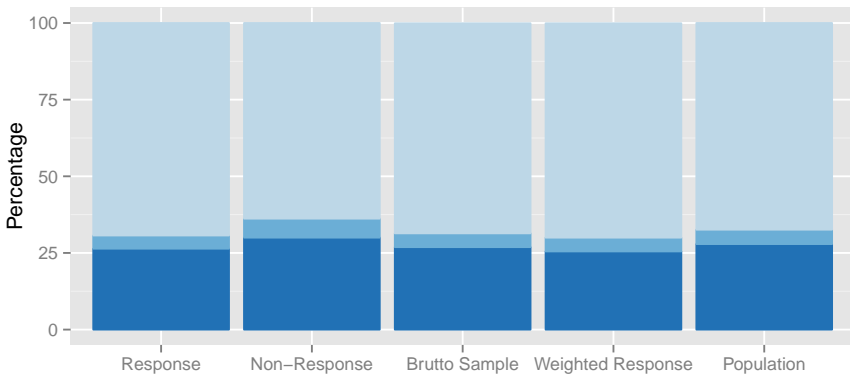
$\frac{Y}{X}$			gesch.EZ
	9.1485	10.8515	20
	30.8515	9.1485	40
gesch.EZ	40	20	

Nach 9 Schritten der Anpassung sind die exakten Eckzahlen erreicht, wir haben also die Stichprobe auf diese Eckzahlen kalibriert.

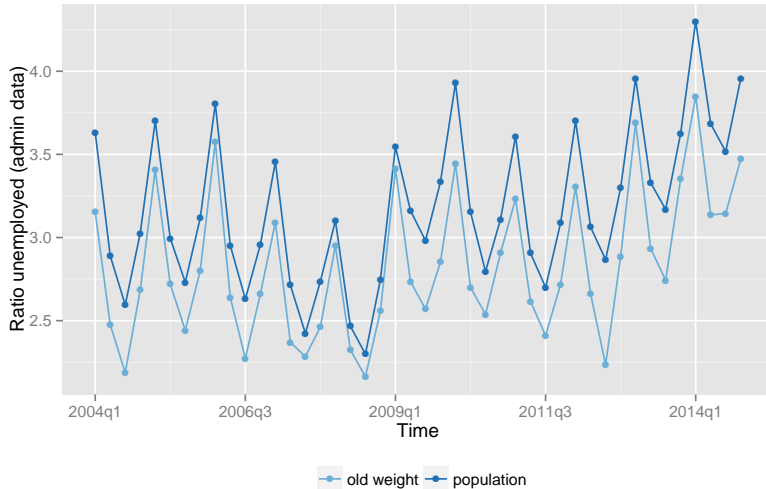
- ▶ “Alte” Kalibrierung (bis 2014)
- ▶ Eckzahlen:
 - ▶ NUTS-2 X Geschlecht X Altersklassen
 - ▶ NUTS-2 X Nationalität(sgruppen)
 - ▶ NUTS-2 X Haushaltsgröße
- ▶ Alle Personen eines HH haben dasselbe Gewicht

- ▶ Registerzählung 2011 \Rightarrow Neue Bevölkerungs- und Haushaltszahlen
- ▶ Veraltete Definition der Nationalitätsgruppen z.B., Kroatien \in EU
- ▶ Möglichkeit über BPK mit Administrativdaten zu mergen
- ▶ Kleiner aber selektiver Non-response \Rightarrow bias correction
- ▶ Reduzierung des Stichprobenfehlers

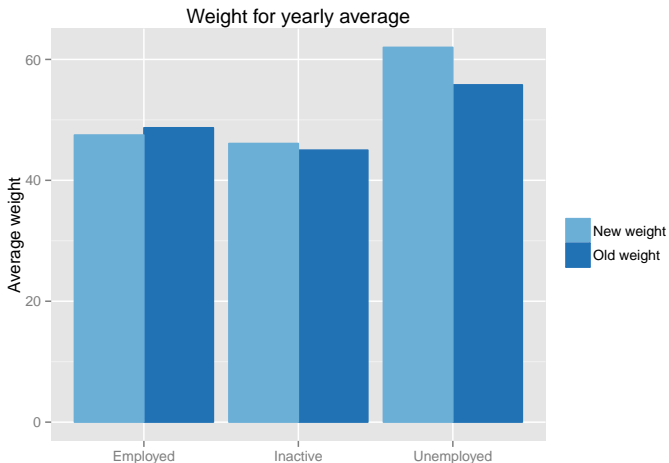
- ▶ Non-Response 5.7%
- ▶ Kleine Gruppe von Non-Respondents, aber selektiv
- ▶ (Source: LFS 2012, ZMR, HV, AMS)



Austria



- ▶ Administrativdaten über den Erwerbsstatus \Rightarrow Neue Eckzahlen
 - ▶ NUTS-2 X Geschlecht X Verwaltungsdatenerwerbsstatus



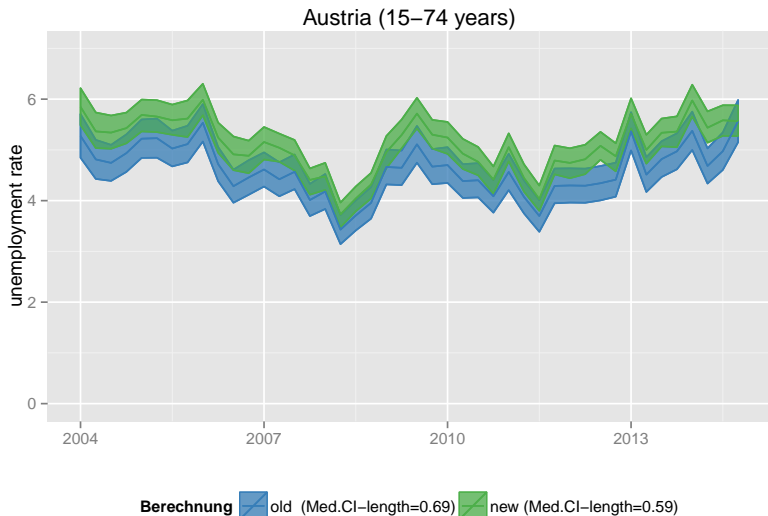
Auswirkung der Anpassung - Verschiebung der Arbeitslosenrate

Austria



Durchschnitt 2014

group	estimator	OldWeight	NewWeight	Difference
Total	Employment Rate	72.60	71.10	-1.54
Male	Employment Rate	77.00	75.20	-1.80
Female	Employment Rate	68.20	66.90	-1.28
Total	Unemployment Rate	5.00	5.60	0.66
Male	Unemployment Rate	5.20	5.90	0.62
Female	Unemployment Rate	4.70	5.40	0.70



Was man bei der Fehlerrechnung beachten muss

- ▶ Clustereffekt
- ▶ Kalibrierung
- ▶ Rotation

⇒

- ▶ Vorher: Taylorreihen-Linearisierung und Approximation mit Hilfe des Binomialansatzes
- ▶ Aktuell: Bootstrapverfahren

- ▶ Ziehe b Bootstrap Samples ($b=500$ im Moment)
- ▶ Haushalte werden in die Stichprobe gezogen \Rightarrow gleiches Bootstrap Gewicht für alle Personen im Haushalt
- ▶ Sampling without replacement from finite population \Rightarrow Rescaled Bootstrap verwendet \Rightarrow non-integer bootstrap weights
- ▶ Kalibrierung jedes Bootstrapgewichts mit dem selben Schema (Eckzahlen) wie Original
- ▶ Rotation: Nur der neue Teil der Stichprobe erhält neue Stichprobengewichte

As usual:

- ▶ Berechne den Schätzer mit allen b Bootstrap Gewichten
- ▶ Standardabweichung der b Schätzer \Rightarrow Standardfehler des Schätzers
- ▶ Untere/Obere Grenze des KI definiert durch 2.5%/97.5% Quantil der 500 Schätzwerte

- ▶ Schnelle und für jedermann/-frau durchführbare Möglichkeit der Fehlerrechnung zu jedem beliebigen MZ-bezogenen Schätzer - insbesondere zu Schätzwerten von Veränderungen
- ▶ Erster Schritt zur automatisierten Tabellenerstellung (Dynamic Reporting) unter Miteinbeziehung von Stichprobenfehlern

⇒

- ▶ Einsatz von R
- ▶ R-Paket mzR für alle MZ Datensatznutzer

Ergebnisse für 2014

```
> dat <- ImportData(year=2014)
```

Anzahl Arbeitslose der Altersgruppe 15-74 nach Geschlecht

```
> GroupSize(dat, TFstring="xerwstat==2 & balt>=15 & balt<=74",  
+           each="bsex")
```

\$bsex_1

	est	sd	cv	cil_2.5%	ciu_97.5%
2014	134,655.5	2,900.84	0.02	129,417.4	140,089.9

\$bsex_2

	est	sd	cv	cil_2.5%	ciu_97.5%
2014	110,227.3	2,792.34	0.03	104,545.1	115,476.2

Veränderungen zwischen 2014q4 und 2013q4

```
> dat <- ImportData(year=2014, quarter=4, comp_diff_lag=4)
```

Arbeitslosenrate der Altersgruppe 15-74

```
> GroupRate(dat, TFstring="xerwstat==2 & balt>=15 & balt<=74",  
+   TFstring2="xerwstat%in%c(1,2) & balt>=15 & balt<=74")
```

	est	sd	cv	cil_2.5%	ciu_97.5%
2014q4	5.57	0.16	0.03	5.27	5.88
2013q4	5.36	0.15	0.03	5.05	5.66
Absolute change	0.21	0.22	1.09	-0.22	0.65
Relative change	3.85	4.29	1.11	-3.84	12.46

- ▶ Demnächst auf <https://github.com/statistik/mzR>
- ▶ oder direkt installieren mit:
`devtools::install_github("statistik/mzR")`

- ▶ Preston J (2009). Rescaled bootstrap for stratified multistage sampling. *Survey Methodology*, 35(2),
- ▶ Rao J, Wu C (1984). Bootstrap inference for sample surveys. *Proceeding Section on Survey Methods Research, Journal of the American Statistical Association*, 106, 112.