

Angelika Meraner  
Qualitätsmanagement und  
Methodik  
Wien  
11. Mai 2016

## mzR: -Paket zur schnellen und selbstständigen Berechnung von Stichprobenfehlern beim Mikrozensus

- ▶ Fehlerrechnungsmethode im Zuge der neuen Hochrechnung der MZ-AKE ab dem 4. Quartal 2014 umgestellt
  - ▶ Vorher: Taylorreihen-Linearisierung und Approximation mit Hilfe des Binomialansatzes
  - ▶ NEU: Bootstrapverfahren

- ▶ 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

⇒ Hausinternes R-Paket **mzR** auch für Mitarbeiter/-innen ohne vorherige R-Kenntnisse

- ▶ Mini-Einführung Bootstrap-Fehlerrechnung
- ▶ Erste Schritte mit **mzR**
- ▶ MZ-Files einlesen
- ▶ Schätzer und zugehörige Fehler berechnen
- ▶ Labels abfragen, Ergebnisse exportieren

- ▶ Ziehung von  $b$  Bootstrapstichproben aus der ursprünglichen MZ-Stichprobe (derzeit  $b=500$ )
- ▶ Österreichischer MZ: rotierende Haushaltsstichprobe ohne Zurücklegen aus einer endlichen Grundgesamtheit  
→ Rescaled Bootstrap Procedure (Rao und Wu, 1988),  
Vorkommenshäufigkeiten in Bootstrapstichprobe dabei nicht mehr ganzzahlig
- ▶ Vorkommenshäufigkeiten \* Ursprüngliche Stichprobengewichte = Bootstrap-Basisgewichte
- ▶ Rotationsstichprobe → Nur neuer Teil der Stichprobe erhält neue Vorkommenshäufigkeiten
- ▶ Kalibrierung der Bootstrap-Basisgewichte mit identischem Verfahren (IPU) wie für ursprüngliche Stichprobengewichte liefert die endgültigen Bootstrapgewichte

- ▶ Berechnung des Schätzers mit jedem der  $b$  Bootstrapgewichte  
→  $b$  Schätzer
- ▶ Stichprobenfehler = Standardabweichung dieser  $b$  Schätzer
- ▶ Untere und obere Grenze des 95%-KI = 2.5% und 97.5% Quantil der  $b$  Schätzer

## Grundfunktionalität des Pakets

- ▶ `ImportData()`, `IndivImportData()`, `ImportAndMerge()`
- ▶ `GetLabels()`
- ▶ `GroupSize()`, `GroupRate()`
- ▶ `Total()`, `Mean()`
- ▶ `export()`

## Automatische Erzeugung von Dokumentationen und Berichten: Prototyp

- ▶ `MakeTable()`, `MakeAKETimeInstantsTable()`
- ▶ `FillExcelTemplate()`

- ▶ Installation
- ▶ **mzR**-Paket laden

```
> library("mzR")
```
- ▶ Hilfe zu **mzR**-Paket aufrufen (Package-Vignette zu finden unter "User guides, package vignettes and other documentation")

```
> help(package=mzR)
```
- ▶ Hilfe zu einzelner Funktion aufrufen, z.B. zur Funktion `ImportData()`

```
> ?ImportData
```



## Funktion für hausinterne Nutzer

```
> ImportData(year = NULL, quarter = NULL, comp_diff_lag = NULL,  
+   from = NULL, to = NULL, hh = FALSE, families = FALSE,  
+   whichVar = NULL, nbw = NULL)
```

## Beispiele

- ▶ Daten einlesen für Jahresergebnisse, z.B. für 2014  

```
> dat <- ImportData(year=2014)
```
- ▶ Daten einlesen für Schätzer von Veränderungen, z.B. zwischen 2013q4 and 2014q4  

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

## Funktion für externe Nutzer

```
> IndivImportData(curr_inFile, curr_inFile_bw, prev_inFile = NULL,  
+   prev_inFile_bw = NULL, whichVar = NULL, mergeBy = "asbhh",  
+   nbw = NULL, bwNames = NULL, gewName = "gew1")
```

Nachfolgende Beispiele enthalten Ergebnisse für 2013q4 und 2014q4 sowie die Veränderungen zwischen diesen beiden Zeiträumen

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

## ► Arbeitslosenzahlen 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%
2014q4	134,136.28	5,169.98	0.04	124,435.44	145,230.91
2013q4	118,029.02	5,172.23	0.04	107,584.92	127,989.34
Absolute change	16,107.26	7,194.81	0.45	3,895.09	30,467.08
Relative change	13.65	6.61	0.48	3.16	27.91

\$bsex\_2

	est	sd	cv	cil_2.5%	ciu_97.5%
2014q4	108,136.31	4,773.78	0.04	99,350.46	117,784.88
2013q4	113,565.06	4,465.39	0.04	104,689.67	121,277.15
Absolute change	-5,428.75	6,465.38	-1.19	-17,066.98	7,050.19
Relative change	-4.78	5.62	-1.18	-14.35	6.71

## ► 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

## ► Geleistete Arbeitsstunden der 15-74-Jährigen in Millionen

```
> Total(dat, TFstring="xerwstat==1 & balt>=15 & balt<=74",  
+       var="estund*13/10^6 + dtstd*13/10^6")
```

	est	sd	cv	cil_2.5%	ciu_97.5%
2014q4	1,743.18	8.28	0.00	1,728.44	1,758.64
2013q4	1,727.83	8.10	0.00	1,713.36	1,744.18
Absolute change	15.35	11.04	0.72	-5.80	38.69
Relative change	0.89	0.64	0.72	-0.33	2.25

- Durchschnittlich geleistete Arbeitsstunden der 15-74-Jährigen nach Geschlecht

```
> Mean(dat, TFstring="xerwstat==1 & balt>=15 & balt<=74",  
+       var="estund + dtstd", each="bsex")
```

\$bsex\_1

	est	sd	cv	cil_2.5%	ciu_97.5%
2014q4	37.27	0.18	0	36.94	37.64
2013q4	37.02	0.18	0	36.68	37.34
Absolute change	0.26	0.26	1	-0.20	0.75
Relative change	0.69	0.69	1	-0.54	2.03

\$bsex\_2

	est	sd	cv	cil_2.5%	ciu_97.5%
2014q4	27.42	0.17	0.01	27.11	27.77
2013q4	27.39	0.17	0.01	27.08	27.72
Absolute change	0.03	0.24	7.30	-0.43	0.49
Relative change	0.12	0.88	7.31	-1.54	1.81

- Labels von Variablen abfragen, z.B. der Variable Geschlecht

```
> GetLabels(dat, "bsex")
```

```
$bsex
```

```
$bsex$label
```

```
      bsex
```

```
"Geschlecht"
```

```
$bsex$value.labels
```

```
Weiblich Männlich  Filter
```

```
      2      1      -3
```

- Ergebnis in ein CSV-File exportieren, z.B. als "D:/MeinTestFile.csv"

```
> will_raus <- GroupRate(dat,
```

```
+   TFstring="xerwstat==2 & balt>=15 & balt<=74",
```

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

```
> export(will_raus, outFilePath="D:",
```

```
+   outFileName="MeinTestFile")
```



- ▶ 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), 227-234
- ▶ Rao JN, Wu C (1988). Resampling Inference with Complex Survey Data. Journal of the american statistical association, 83(401), 231-241