Weight Calibration across Packages

Stas Kolenikov
9/23/2019
Weight calibration

- Last step in creating analysis weights in survey data files
- Adjusting the weights so that they sum to known population totals in different subgroups (age, sex, race, ethnicity, geography, etc.)
- Desirable to minimize changes from the input weights (probability of selection, nonresponse adjustments, frame integration, etc.)

Deville & Sarndal (1992)
Contenders

Stata

- `ipfraking` (Kolenikov 2014, 2019)
- `svycal` (official Stata)
- `survwgt` (Winter 2002)
- `sreweight` (Pacifico 2014)

R

- `survey::calibrate()` (Lumley 2010)

SAS

- `rake_and_trim()` (Izrael, Battaglia, Hoaglin, Frankel, Ball, 2017)
Out of scope

- SUDAAN PROC WGTADJ, PROC WGTADJX
- Stata ipfweight (Bergmann 2011)
- R library(ReGenesees) (Zardetto 2015)
- R library(ipfr) (Ward, Macfarlane 2019)
Expectations

- Produce usable results
- Provide weight diagnostics
- Speed
- Fool proof
Running example
Running example

CPS 2018 March ASEC data

- estimate control totals based on 13353 adults in CA
- calibrate 8403 adults in TX on
  - sex
  - age (14 categories)
  - race/ethnicity (6 categories)
  - education (5 categories)
  - HH income (9 categories)
  - nativity (3 categories)
  - marital status (6 categories)
  - own vs. rent
  - metro area of TX (23 categories)
Tasks and tests

1. Straight raking
2. Raking with divergent population control totals
3. Raking with bounded weight adjustment ratios [0.3,3]
4. Raking with bounded weight values (2nd and 98th percentile of unrestricted distribution)
5. Linear calibration
6. Linear calibration with trimming
7. (Informative error expected) incorrect specification of control totals
## Performance summary

<table>
<thead>
<tr>
<th>package</th>
<th>TOT</th>
<th>DIV</th>
<th>REL</th>
<th>ABS</th>
<th>LIN</th>
<th>LIN+TR</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ipfraking</td>
<td>name</td>
<td>+W</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>N</td>
<td>7.14 sec</td>
</tr>
<tr>
<td>svycal</td>
<td>name/=</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>+</td>
<td>+</td>
<td>0.18 sec</td>
</tr>
<tr>
<td>survwgt</td>
<td>order</td>
<td>NW</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>0.80 sec</td>
</tr>
<tr>
<td>sreweight</td>
<td>order</td>
<td>F</td>
<td>F</td>
<td>N</td>
<td>+</td>
<td>N</td>
<td>0.19 sec</td>
</tr>
<tr>
<td>calibrate</td>
<td>name</td>
<td>-W</td>
<td>..</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>0.35 sec</td>
</tr>
<tr>
<td>rake_and_trim</td>
<td>name+magic</td>
<td>F</td>
<td>-W</td>
<td>+</td>
<td>N</td>
<td>N</td>
<td>61 sec</td>
</tr>
</tbody>
</table>

- **N**: no documented functionality exists
- **W**: issued reasonable warnings
- **F**: failed with cryptic error message / no message
Stata ipfrraking

The worst relative discrepancy of 2.7e-07 is observed for age_cat = 18

Target value = 76457.48, achieved value = 76457.48
Truncated due to the upper absolute limit: 121 weights.
Truncated due to the lower absolute limit: 74 weights.

Summary of the weight changes

<table>
<thead>
<tr>
<th>Orig weights</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min</th>
<th>Max</th>
<th>CV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2467</td>
<td>943</td>
<td>1133</td>
<td>9670</td>
<td>.3822</td>
</tr>
<tr>
<td>Raked weights</td>
<td>2059</td>
<td>3935</td>
<td>699</td>
<td>15600</td>
<td>.6419</td>
</tr>
<tr>
<td>Weighth factor</td>
<td>1.0734</td>
<td>0.1528</td>
<td>1.5615</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Stata `survwgt`
R survey::calibrate()
*** Program terminated at iteration 11 because raking converged ***

The FREQ Procedure

 Weighted Distribution After Raking

<table>
<thead>
<tr>
<th>Sex</th>
<th>Output Weight Sum of Weights</th>
<th>Target Total</th>
<th>Sum of Weights Difference</th>
<th>% of Output Weights</th>
<th>Target % of Weights</th>
<th>Difference in %</th>
<th>Marginal Category Difference in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14813715.70</td>
<td>14813331</td>
<td>385.02</td>
<td>49.012</td>
<td>49.011</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>2</td>
<td>15410993.62</td>
<td>15411379</td>
<td>-385.02</td>
<td>50.988</td>
<td>50.989</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
</tbody>
</table>
SAS `rake_and_trim()`
Weight trimming ≠ methodology
## Misspecified control totals

<table>
<thead>
<tr>
<th>package</th>
<th>Extra in pop</th>
<th>Extra in data</th>
<th>Wrong order</th>
</tr>
</thead>
<tbody>
<tr>
<td>ipfraking</td>
<td>E</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>svycal</td>
<td>E</td>
<td>!!!</td>
<td></td>
</tr>
<tr>
<td>survwgt</td>
<td>N/A</td>
<td>N/A</td>
<td>!!!</td>
</tr>
<tr>
<td>sreweight</td>
<td>E</td>
<td>E</td>
<td>!!!</td>
</tr>
<tr>
<td>calibrate</td>
<td>E</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>rake_and_trim</td>
<td>!?!?</td>
<td>!!!</td>
<td></td>
</tr>
</tbody>
</table>

E: issued an error and stopped

!!!: did not issue an error – results highly suspect!
Thanks and out

Questions?

- stas_kolenikov@abtassoc.com
- @StatStas on Twitter