postweight or calibrate? Survey post-adjustments in Stata

Pablo Cabrera Álvarez (USAL)
Modesto Escobar Mercado (USAL)

2017 Spanish Stata Users Group meeting

Madrid, 19 octubre 2017

Survey post-adjustments

- Non-response and coverage errors affect survey estimates
- Some population members are not in the sample frame (e.g. households with no internet connection in CAWI)
- Respondents may differ from non-respondents on key characteristics
- Weighting (e.g. postestratification, calibration) is a form of post-adjustment that aims to rebalance the sample using auxilary information (Biemer & Christ, 2008)

Postestratification and calibration

- Both methods use auxilary information in order to force the sample distribution to match the population on key auxilary variables
- "Postestratification could refer to any method of data analysis which involves forming units into homogeneous groups after observation of the sample" (Smith, 1991)
- Calibration doesn't include any or some of the interactions; It is a relaxed form of postestratification (Zhang, 2000; Särndal, 2007)

Postestratification in Stata

- *svyset* is the Stata command for complex survey analysis and it allows to compute postestratification weights in the background:

```
svyset, poststrata(varname) postweight(varname)
```

- *poststrata* refers to a variable which identifies the cases in the same poststrata
- *postweight* refers to the population size of each poststratum
- The poststratification weight is computed for each estimate based on the sample of valid cases

Calibration in Stata

- *calibrate* (D'Souza, 2010) is a command to compute calibration weights:

```
calibrate, marginals(varlist) poptot(matrix)
           entrywt(varname) exitwt(varname)
```

- The *calibrate* command generates a calibration weight which forces the sample to match the population distribution on the *marginals* variables
- The survey estimate is computed each time applying the weight *svyset*, [*pweight* = *calibwt*]

Data and methods

- Barometer 3183 (July 2017) from Centre for Sociological Research. Multistage clustered sample. Age and sex quotas to select final respondents
- Information about past vote and voting intention available; those with no valid past vote were excluded
- Three steps: 1) Rebalancing the sample to match the 2016 election results; 2) Estimating voting intention variables using calibration weights and svy poststratification; 3) Comparing estimates

Postweight vs *calibrate* on auxilary variable

Past vote estimate using postweight and calibrate (%):

Past vote	2016 elec.	Raw	*Postwt.*	*Calib.*	wt
PP	22.95	23.54	22.95	22.95	0.98
PSOE	15.74	22.08	15.74	15.74	0.71
UP	14.71	16.97	14.71	14.71	0.87
Cs	9.08	9.49	9.08	9.08	0.96
ERC	1.83	2.92	1.83	1.83	0.63
Another party	5.53	8.62	5.53	5.53	0.64
Didn't vote	30.17	16.38	30.17	30.17	1.84

Postweight vs *calibrate* on target variables (I)

	Raw (%)	*Postweight* (%)	*Calibrate* (%)
Would vote	84.95	79.64	79.64
Wouldn't vote	15.05	20.36	20.36

Postweight vs *calibrate* on target variables (II)

	Raw (%)	*Postweight* (%)	*Calibrate* (%)	Diff.
PP	25.60	26.64	28.37	-1.73
PSOE	29.69	27.46	26.90	0.56
UP	20.71	20.42	20.91	-0.49
Cs	14.16	16.82	16.12	0.70
ERC	3.87	2.77	2.94	-0.17
PDCat	0.85	0.83	0.66	0.18
PNV	1.37	1.28	1.02	0.26
Another party	3.75	3.78	3.09	0.69

Comparing weights

Past vote	*Calibrate* wt (full sample)	*Postweight* wt (if vote)
PP	0.98	0.87
PSOE	0.71	0.64
UP	0.87	0.76
Cs	0.96	0.90
ERC	0.63	0.53
Another party	0.64	0.84
Didn't vote	1.84	3.12

Wrap-up

- Both *Postweight* and *Calibrate* are tools for rebalancing the sample
- *Postweight* works in the background recalculating the weights for each estimate based on the valid sample but the same population totals; *Calibrate* computes a weight to force the sample to match the population, this weight is used for estimates
- *Postweight* can only be used for general sample estimates while the weight produced by *Calibrate* can be used for general and subsample estimates
- The estimation procedures using *Postweight* [*svy poststrata* and *postweight* options] or *Calibrate* [*svy pweight* option] lead to different standard errors

Syntax (I)

```
** Preliminar
scalar Population = 34596892
matrix Weights = (7941236, 5443846, 5087538, ///
          3141570, 632234, 1911558, 10438910)
matrix colnames Weights= PP PSOE UP Cs ERC ///
                              Others DontVote
matrix rownames Weights=Population
```

Syntax (II)

```
** Postweight
gen poptotal = 0
foreach num of numlist 1/7 {
replace poptotal=Weights[1,`num'] if VarPost==`num'
}
```

```
svyset _n, poststrata(VarPost) postweight(poptotal)
svy: prop Variable
```

Syntax (III)

```
** Calibrate (I)
quietly: sum VarPost
gen start = Population/r(N)
tab VarPost, gen(ValuesPost)
calibrate, marginals(ValuesPost1-ValuesPost7) ///
  poptot(Weights) ent(start) exit(wtcal) method(logistic)
svyset cues [pweight=wtcal]
svy: prop Variable
```

Syntax (IV)

```
** Calibrate (II)
quietly: sum VarPost if demost == 1
gen start2 = Population/r(N) if demost == 1
calibrate, marginals(ValuesPost1-ValuesPost7) ///
  poptot(Weights) ent(start2) exit(wtcal2) method(logistic)
quietly: sum wtcal2, d
replace wtcal2=wtcal2/r(mean)
svyset cues [pweight=wtcal2]
svy: prop Variable
```

Bibliography

Biemer, P. P., & Christ, S. L. (2008). Weighting survey data. International handbook of survey methodology, 317-341. Chicago

D'Souza, J. (2010). Calibrate: a Stata Program for Calibration Weighting. London: Stata User Group.

Särndal, C. E. (2007). The calibration approach in survey theory and practice. Survey Methodology, 33(2), 99-119.

Smith, T. M. (1991). Post-stratification. The Statistician, 315-323.

StataCorp. (2015). Stata 14 Base Reference Manual. College Station, TX: Stata Press.

Zhang, L. C. (2000). Post-stratification and calibration a synthesis. The American Statistician, 54(3), 178-184.