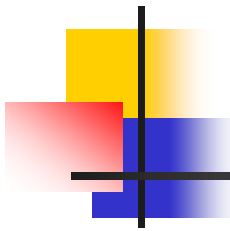


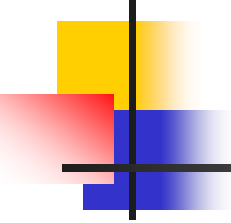
Re-estimating Weights for IPUMS-Greece Samples

Dr. Stefanos G. Giakoumatos
Technological Educational
Institute of Kalamatas



Introduction

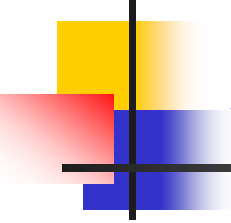
- Weights are commonly used to survey data file in order to make the weighted results to represent the population of inference as closely as possible.
- In recent years have seen theoretical developments and increased use of Weighting methods
- These methods take account of substantial amounts of auxiliary information (official statistics, etc)



Preliminaries and notations

- A target population U of size N
- A study variable y
- A sample s of size n drawn from U
- Let $\{d_k, k \in s\}$ the design weights (i.e. the inverse of the selection probability)
- « Natural » estimator for the total Y of y :

$$\hat{Y} = \sum_{k=1}^n d_k y_k$$

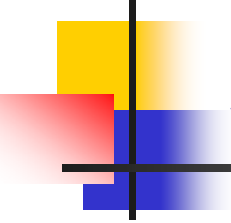


Aim of Weighting Methods (1)

- In many cases the total Y of the study variable is known (Based on the official Statistics or other sources)
- However the estimator is not equal to the known total

$$Y \neq \hat{Y}$$

- This discrepancy is caused by
 - the **randomness** of the sample
 - Non-response
 - Non-coverage

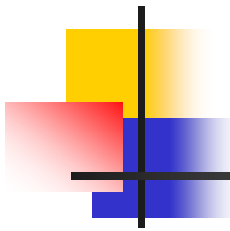


Aim of Weighting Methods (2)

- The aim of the Weighting methods is to adjust the design weights in order

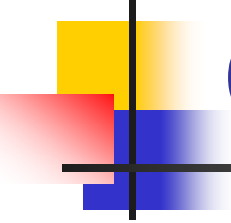
$$Y = \hat{Y}$$

- The weighting methods could be applied not only to totals but also to any sub-total of our variable
- In addition, they reduce the bias and the variance of the estimator.



Review of Weighting Methods

- Cell weighting (limiting usage)
- Raking Weighting (Ireland and Kullback 1968)
- Linear Weighting (Deville, Sarndal, and Sautory 1993)
- GREG (Logit) Weighting (Deville and Sarndal 1992; Fuller, McLoughlin, and Baker 1994; Fuller 2002).



General Framework

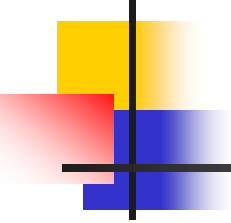
- THE CALIBRATION PROBLEM:

- We seek weights $\omega = \{\omega_k, k \in r\}$ satisfying the following problem of optimisation :

$$\omega = \arg \min_{\{s_k\}} \left[\sum_{k \in r} G_k(s_k, d_k) \right]$$

so that

$$\sum_{k \in r} \omega_k \cdot x_k = X$$



Linear

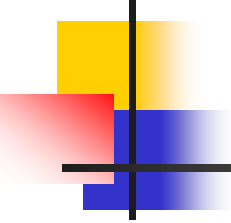
$$G_k(\omega_k, d_k) = \frac{(\omega_k - d_k)^2}{2d_k}$$

Advantages:

- Always convergent
- The convergence is fast

Drawbacks:

- The weights can take negative values



Raking ratio

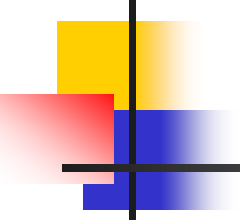
$$G_k(\omega_k, d_k) = \omega_k \cdot \log\left(\frac{\omega_k}{d_k}\right) - \omega_k + d_k$$

Advantages:

- The calibrated weights always take positive values

Drawbacks:

- The final weights are not bounded



Logit

$$G_k(\omega_k, d_k) = K \cdot \left[\left(\frac{\omega_k}{d_k} - L \right) \cdot \log \left(\frac{\frac{\omega_k}{d_k} - L}{1 - L} \right) + \left(U - \frac{\omega_k}{d_k} \right) \cdot \log \left(\frac{U - \frac{\omega_k}{d_k}}{U - 1} \right) \right]$$

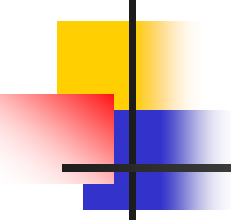
$$\text{if } L \leq \frac{\omega_k}{d_k} \leq U$$

Advantages:

- The weights are bounded (by L and U)

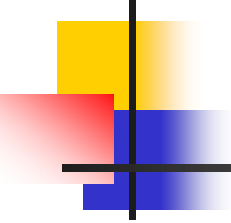
Drawbacks:

- If L and U are chosen too close, the problem can have no solution



IPUMS Data (Greek Case)

- Random Sample from the censuses 1971-2001 (random systematic sample, easy to construct design weights)
- Most of the Totals and the Subtotals are known from the National Statistical Service of Greece



Application of Weighting the Greek IPUMS Data

- Focus on
 - Census 2001
 - The sample from the Households
 - Use SAS
 - CALMAR (CALage sur MARges) = a SAS macro for calibrating a sample



Variables

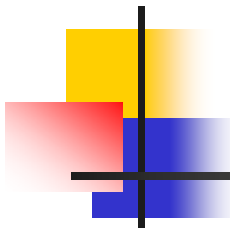
SAS Variable	Μεταβλητή	Variables
V3	Θέση κατοικίας στον οικισμό	Position of the HH in the Area
V4	Είδος κατοικίας	Type of HH
V6	Τύπος κτιρίου	Type of Building
V7	Περίοδος κατασκευής	Construction Period
V11	Έχει κουζίνα?	Kitchen or cooking facilities
V13	Έχει ηλεκτρισμό?	Electricity
V14	Τύπος ύδρευσης	Water
V15	Λουτρό	Bath
V16	Αποχέτευση	Sewage
V17	Αποχωρητήριο	Toilet
V18	θέρμανση	Central Heat
V19	Φορέας Ιδιοκτησίας	Ownership of dwelling

Comparison and Results (1)

	Values	Design Weights		Results from NSSG		Calibration	
		Count	%	Count	%	Count	%
V3	1	3,648,774	99.30	3,629,721	98.78	3,629,721	98.78
	2	25,607	0.70	44,660	1.22	44,660	1.22
V4	1	3,667,079	99.80	3,667,471	99.81	3,667,471	99.81
	2	7,302	0.20	6,910	0.19	6,910	0.19
V6	1	1,401,785	38.15	1,505,134	40.96	1,505,134	40.96
	2	688,174	18.73	663,198	18.05	663,198	18.05
	3	1,570,129	42.73	1,490,574	40.57	1,490,574	40.57
	4	14,293	0.39	15,475	0.42	15,475	0.42
V7	1	90,572	2.46	112,603	3.06	112,603	3.06
	2	239,027	6.51	267,892	7.29	267,892	7.29
	3	460,581	12.53	448,731	12.21	448,731	12.21
	4	784,493	21.35	718,637	19.56	718,637	19.56
	5	952,738	25.93	900,904	24.52	900,904	24.52
	6	397,562	10.82	406,404	11.06	406,404	11.06
	7	278,405	7.58	295,132	8.03	295,132	8.03
	8	250,120	6.81	261,814	7.13	261,814	7.13
	9	220,883	6.01	262,264	7.14	262,264	7.14
V11	1	3,656,086	99.50	3,631,178	98.82	3,631,178	98.82
	2	18,295	0.50	43,203	1.18	43,203	1.18
V13	1	3,649,627	99.33	3,590,361	97.71	3,590,361	97.71
	2	24,754	0.67	84,020	2.29	84,020	2.29

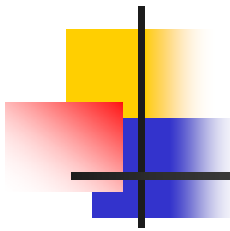
Comparison and Results (2)

	Values	Design Weights		Results from NSSG		Calibration	
		Count	%	Count	%	Count	%
V14	1	3,595,434	97.85	3,525,851	95.96	3,525,851	95.96
	2	34,714	0.94	70,674	1.92	70,674	1.92
	3	30,030	0.82	42,829	1.17	42,829	1.17
	4	5,105	0.14	11,167	0.30	11,167	0.30
	5	5,236	0.14	9,651	0.26	9,651	0.26
	6	3,862	0.11	14,209	0.39	14,209	0.39
V15	1	3,450,650	93.91	3,384,227	92.10	3,384,227	92.10
	2	139,037	3.78	144,361	3.93	144,361	3.93
	3	4,183	0.11	4,021	0.11	4,021	0.11
	4	80,511	2.19	141,772	3.86	141,772	3.86
V16	1	2,396,478	65.22	2,220,887	60.44	2,220,887	60.44
	2	1,264,021	34.40	1,445,565	39.34	1,445,565	39.34
	3	13,882	0.38	7,929	0.22	7,929	0.22
V17	1	3,451,904	93.95	3,423,410	93.17	3,423,410	93.17
	2	162,046	4.41	181,627	4.94	181,627	4.94
	3	4,323	0.12	8,286	0.23	8,286	0.23
	4	31,294	0.85	51,762	1.41	51,762	1.41
	5	24,814	0.68	9,296	0.25	9,296	0.25
V18	1	2,274,602	61.90	2,066,040	56.23	2,066,040	56.23
	2	1,217,271	33.13	1,103,510	30.03	1,103,510	30.03
	3	182,507	4.97	504,831	13.74	504,831	13.74
V19	1	3,620,600	98.54	3,626,308	98.69	3,626,308	98.69
	2	16,851	0.46	16,286	0.44	16,286	0.44
	3	36,931	1.01	31,787	0.87	31,787	0.87



Conclusion

- With calibration methods we can use and incorporate the results of the censuses in the IPUMS data, therefore
- to produce weights that will provide “accurate” results (for qualitative and quantitative variables)



References

- Ireland, C.T. and Kullback, S. (1968). Contingency Tables With Given Marginals. *Biometrika*, 55, 179-188.
- Deville, J.-C. and Sarndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87, 376-382.
- Deville, J.-C, Sarndal, C.-E., and Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, 88, 1013-1020.
- Fuller, W.A. (2002). Regression Estimation for Survey Samples. *Survey Methodology*, 28, 5-23.
- Fuller, W.A., McLoughlin, M.M., and Baker, H.D. (1994). Regression Weighting in the Presence of Nonresponse With Application to the 1987-1988 Nationwide Food Consumption Survey. *Survey Methodology*, 20, 75-85.