Weights for the Hellenic Voter Study 2012

Ioannis Andreadis

The 2012 Hellenic (Greek) National Election Voter Study was conducted as a mixed-mode survey in the period between 19 October 2012 and 5 January 2013. Half of the sample was designed to be collected online. The sample was selected randomly (using RDD) by area proportional to total population. The selected respondents were called on the phone and they were asked to provide their email address if they wanted to participate in a web survey conducted by the Laboratory of Applied Political Research, Aristotle University of Thessaloniki. These email addresses have been used into the epolls.gr web survey system (Andreadis 2010). The other half of the sample was designed to be collected by face-to-face interviews. During the design phase of the survey we expected that the 55+ age group would be greatly under-represented in the web survey sample. We also expected that the regions of Attiki (region of the capital Athens) and Thessaloniki would be over-represented and the rural areas would be significantly under-represented in the web survey sample (taking into consideration the limited access and use of the Internet of these groups). The face to face sample was collected using the following method: Firstly, geographical cluster sampling was used, but due to the aforementioned expectations, the regions of Attiki and Thessaloniki have been excluded from the selection of the clusters. Then in each selected cluster, systematic sampling was used, i.e. the interviewers selected the first house randomly and they continued in the cluster selecting every kth house. Secondly, they have used quota sampling for age (i.e. when people aged <55 reached 30% of the sample they have stopped collecting any more responses from people of this age group). The final dataset consists of 529 respondents to the web survey and 500 respondents to the face-to-face interviews. This document describes how the Hellenic Voter Study 2012 weights have been constructed.

Gender

The gender distribution in the unweighted sample of the Hellenic Voter Study 2012 is presented in Table 1.

Table 1. Gender distribution in ELNES 2012 unweighted

	Frequency	Relative
MALE	611	0.59
FEMALE	413	0.40

According to the Hellenic Statistical Authority - ELSTAT (2013) during the 2011 population census, there were found 5302703 males and 5512494 females. Thus, the gender (which has 5 missing values in the sample), should be distributed close to the expected distribution presented in Table 2.

Table 2. Expected gender distribution according to Greek census 2011

	D2	Freq
1	MALE	501.76
2	FEMALE	522.24
3	MISSING	5.00

With poststratification the population is partitioned into subgroups that are called poststrata. The original weights (in our case they are all equal to one) are multiplied by a ratio which is formed by the corresponding population poststratum size in the nominator and the corresponding sample poststratum size in the denominator (see Lehtonen and Pahkinen 2004, p.88-92; Holt and Smith 1979). For instance, this ratio for the male group is: 501.76/611=0.8212. These adjustments to the sampling weights make the estimated gender distribution to match the known population gender distribution, making the sample more representative of the population. Thus, after the poststratification adjustment on gender the Hellenic Voter Study 2012 is weighted according to the variable that is summarized in Table 3.

Table 3. Summary of weights after adjusting for gender

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.821   0.821   0.821   1.000   1.260   1.260
```

and the distribution of the gender variable in the weighted sample is presented in Table 4:

Table 4: Gender distribution in ELNES 2012 after weighting

	Frequency	Relative
MALE	502	0.49
FEMALE	522	0.51

Age

The age distribution in the unweighted ELNES 2012 sample is displayed in Table 5.

Table 5. Age distribution in ELNES 2012 unweighted

	Frequency	Relative
18-25	72	0.073
26-40	220	0.223
41-64	523	0.529
65+	173	0.175

Using data from Table 2. Permanent population by age, gender and marital status available at: http://www.statistics.gr/portal/page/portal/ESYE/PAGE-cencus2011tables (http://www.statistics.gr/portal/page/portal /ESYE/PAGE-cencus2011tables) the age distribution for the voting population is:

Table 6. Age distribution of voting age population

Age	Frequency	Relative
18-25	991178	0.111
26-40	2391855	0.268
41-64	3433578	0.385
65+	2108670	0.236

Post-stratification using more than one variable requires the groups to be constructed as a complete cross-classification of the variables, but often the population values of the inner cells of the cross-classified table are not available (i.e. only the marginal values are known). Even when the values of the inner cells are known, the number of cross-classified categories can become so large that the values of inner cells become small and unstable (Holt and Elliot 1991). Raking allows multiple grouping variables to be used by post-stratifying on each variable in turn, and repeating this process until the weights stop changing (Lumley 2011).

Table 7: Gender distribution in ELNES 2012 after weighting for gender and age

	Frequency	Relative
MALE	502	0.49
FEMALE	522	0.51

Table 8: Age distribution in ELNES 2012 after weighting for gender and age

	Frequency	Relative
18-25	110	0.111
26-40	265	0.268
41-64	380	0.385
65+	233	0.236

From the previous two tables it is obvious the both age and gender in the weighted sample follow a distribution that is similar to the corresponding population distribution. Table 9 displays the summary of weights after adjusting for gender and age.

Table 9. Summary of weights after adjusting for gender and age

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.639   0.639   0.909   1.000   1.220   1.740
```

Education

The education levels of the unweighted sample are presented in Table 10.

Table 10. Education distribution in ELNES 2012 unweighted

	Frequency	Relative
ISCED97 (0-2)	239	0.232
ISCED97 (3-4)	330	0.321
ISCED97 (5-6)	450	0.437
Missing	10	0.010

The Hellenic Statistical Authority has not published the education level frequencies from the 2011 census. Thus, I have used education data from the EU Labour Force Survey (EU-LFS) which is the largest European household sample survey (1.8 million interviews are conducted each quarter). For Greece, the theoretical quarterly sample size is approximately 34250 households, corresponding to a sampling rate of about 0.85% (Eurostat 2013). Educational level attained in EU-LFS is measured on the International standard classification of education (ISCED 1997) scale (UNESCO 2006). Using data from "Population by educational attainment level, sex and age (1000) (edat_lfs_9901)" downloaded from the Eurostat database (http://epp.eurostat.ec.europa.eu/portal/page/portal /statistics/search_database) the education level distribution of the population (ages 18-74) in 2012 was:

Table 11. Distribution of the population education levels

Education	Frequency	Relative
ISCED97(0-2)	3038.4	0.381
ISCED97(3-4)	3168.2	0.398
ISCED97(5-6)	1762.5	0.221

Table 12 includes the summary of weights after the poststratification adjustment on gender, age and education.

Table 12. Summary of weights after adjusting for gender, age and education

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.184   0.612   0.928   1.000   1.260   4.190
```

The distributions of age and gender and education are presented in tables 13, 14 and 15.

Table 13: Gender distribution

	Frequency	Relative
MALE	501	0.49
FEMALE	523	0.51

Table 14: Age distribution

	Frequency	Relative
18-25	109	0.110
26-40	263	0.266
41-64	381	0.386
65+	235	0.238

Table 15: Education distribution

	Frequency	Relative
ISCED97 (0-2)	388	0.381
ISCED97 (3-4)	406	0.398
ISCED97 (5-6)	225	0.221

Region

There is a trade-off between the reduction of estimation bias and the increase in the sample variance arising from the variation in the weights. The increase of sample variance is not large when the variation in weights is modest, but as the variation of weights increases the variance in the sample can become very large.

According to Kalton, and Maligalig (1991, p.413), "it may be preferable to collapse two cells if the variance is reduced sufficiently, even though this may create a bias". They show that if a quantity of interest has the same value in two subgroups of respondents, it is always preferable to collapse the two subgroups for estimating the quantity. In other cases, whether to collapse the subgroups depends on the sample sizes. If they are small, collapsing may be preferred.

The distribution of the regions in the unweighted sample is

Table 16. Region distribution in Hellenic Voter Study 2012

	Frequency	Relative
Anatoliki Makedonia, Thraki	135	0.131
Kentriki Makedonia	204	0.198
Dytiki Makedonia	29	0.028
Thessalia	43	0.042
Ipeiros	103	0.100
Ionia Nisia	10	0.010
Dytiki Ellada	207	0.201
Sterea Ellada	49	0.048
Peloponnisos	91	0.089
Attiki	119	0.116
Voreio Aigaio	0	0.000
Notio Aigaio	23	0.022
Kriti	15	0.015

Since some relative frequencies are very small, I combine Kentriki with Dytiki Makedonia, Ipeiros with Ionia Nisia and Aigaio with Kriti.

The distribution of the modified regions in the unweighted sample is

Table 17. Modified region distribution in Hellenic Voter Study 2012

	Frequency	Relative
Anatoliki Makedonia, Thraki	135	0.131
Kentriki & Dytiki Makedonia	233	0.227
Thessalia	43	0.042
Ipeiros & Ionia Nisia	113	0.110
Dytiki Ellada	207	0.201
Sterea Ellada	49	0.048
Peloponnisos	91	0.089
Attiki	119	0.116
Aigaio & Kriti	38	0.037

According to the Hellenic Statistical Authority - ELSTAT (2013) publication of the 2011 population census, the sample regions should be distributed close to the expected distribution presented in Table 18 .

Table 18. Modified region expected distribution (Census 2011)

Regions	Frequency	Relative
Anatoliki Makedonia, Thraki	58	0.056
Kentriki & Dytiki Makedonia	206	0.200
Thessalia	70	0.068
Ipeiros & Ionia Nisia	51	0.050

Dytiki Ellada	65	0.063
Sterea Ellada	52	0.051
Peloponnisos	54	0.053
Attiki	364	0.354
Aigaio & Kriti	108	0.105
	1	0.001

After the poststratification adjustment on gender, age, recoded education and modified regions the Hellenic Voter Study 2012 includes a weight variable that has a maximum value: 10.5 (Table 19).

Table 19. Summary of weights after adjusting for gender, age and recoded education and modified regions

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.045   0.207   0.540   1.000   1.150  10.500
```

A common practice to reduce the variance of the weights is to truncate the weights (Potter 1990; Little 1993). By trimming large weights we also reduce the influence of outlying observations. The total amount trimmed is divided among the observations that were not trimmed, so that the total weight remains the same. Following DeBell and Krosnick (2009), I have trimmed the weights to the value of 5.

After trimming the weights the distributions of age and gender and recoded education are far from with the corresponding population distributions (see tables 20, 21, 22 and 23).

Table 20. Gender distribution in ELNES 2012 after trimming

	Frequency	Relative
MALE	518	0.51
FEMALE	505	0.49

Table 21. Age distribution in ELNES 2012 after trimming

	Frequency	Relative
18-25	99	0.101
26-40	238	0.242
41-64	403	0.409
65+	245	0.249

Table 22. Education distribution in ELNES 2012 after trimming

	Frequency	Relative
ISCED97 (0-2)	391	0.384
ISCED97 (3-4)	374	0.367
ISCED97 (5-6)	253	0.248

Table 23. Modified region distribution in ELNES 2012 after trimming

	Frequency	Relative
Anatoliki Makedonia, Thraki	66	0.064
Kentriki & Dytiki Makedonia	220	0.214
Thessalia	72	0.070
Ipeiros & Ionia Nisia	59	0.057
Dytiki Ellada	78	0.076
Sterea Ellada	56	0.054
Peloponnisos	60	0.059
Attiki	309	0.301
Aigaio & Kriti	108	0.105

Valid votes

The distribution of valid votes in the unweighted sample is

Table 24. Valid votes distribution in ELNES 2012

	Frequency	Relative
ND	207	0.289
SYRIZA-EKM	190	0.265
PASOK	79	0.110
ANEL	53	0.074
LS-XA	38	0.053
DIMAR	88	0.123
KKE	30	0.042
Other	31	0.043

The sample votes should be distributed close to the expected distribution presented in Table 29 (according to the election results available at: http://ekloges.ypes.gr/may2014/e/public/index.html (http://ekloges.ypes.gr/may2014/e/public/index.html)).

Table 25. Expected vote distribution

Parties	Frequency	Relative
ND	213	0.297
SYRIZA-EKM	193	0.269
PASOK	88	0.123
ANEL	54	0.075
LS-XA	49	0.069
DIMAR	45	0.063
KKE	32	0.045
Other	42	0.059

After the poststratification adjustment on gender, age, recoded education, modified regions and valid votes ELNES 2012 includes a weight variable that has a maximum value: 13.2 (Table 26).

Table 26. Summary of weights after adjusting for gender, age, recoded education, modified regions and valid votes

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.020   0.197   0.460   1.000   1.130  13.200
```

After trimming the weights the distributions of age, gender and recoded education, modified regions and valid votes are as follows:.

Table 27. Gender distribution in ELNES 2012 after trimming

	Frequency	Relative
MALE	521	0.51
FEMALE	503	0.49

Table 28. Age distribution in ELNES 2012 after trimming

	Frequency	Relative
18-25	97	0.098
26-40	241	0.244
41-64	404	0.409
65+	246	0.249

Table 29. Education distribution in ELNES 2012 after trimming

	Frequency	Relative

	Frequency	Relative
ISCED97 (0-2)	389	0.382
ISCED97 (3-4)	374	0.367
ISCED97 (5-6)	256	0.251

Table 30. Modified region distribution in ELNES 2012 after trimming

	Frequency	Relative
Anatoliki Makedonia, Thraki	67	0.065
Kentriki & Dytiki Makedonia	222	0.216
Thessalia	70	0.068
Ipeiros & Ionia Nisia	59	0.058
Dytiki Ellada	79	0.077
Sterea Ellada	56	0.054
Peloponnisos	61	0.059
Attiki	308	0.300
Aigaio & Kriti	106	0.103

Table 31. Vote distribution in ELNES 2012 after trimming

	Frequency	Relative
ND	213	0.299
SYRIZA-EKM	183	0.256
PASOK	89	0.125
ANEL	52	0.073
LS-XA	49	0.069
DIMAR	51	0.072
KKE	32	0.045
Other	44	0.061

And here is the summary of the final weight variable:

Table 32. Summary of final trimmed weights

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.090   0.267   0.530   1.000   1.200   5.000
```

Acknowledgements

References

Andreadis, Ioannis. 2010. "Διαδικτυακές Πολιτικές Έρευνες [Web Based Political Surveys]." In *Proceedings of the 23rd Conference of the Hellenic Statistical Institute*, 44–51. http://invenio.lib.auth.gr/record/126916 (http://invenio.lib.auth.gr/record/126916).

DeBell, Matthew, and Jon A Krosnick. 2009. "Computing Weights for American National Election Study Survey Data." ANES Technical Report series, no. nes012427. Ann Arbor, MI,; Palo Alto, CA: American National Election Studies. Available at http://www. electionstidies. org.

ELSTAT. 2013. "Greece in Figures 2013." http://www.statistics.gr/portal/page/portal/ESYE/BUCKET/General /ELLAS_IN_NUMBERS_EN.pdf (http://www.statistics.gr/portal/page/portal/ESYE/BUCKET/General /ELLAS_IN_NUMBERS_EN.pdf).

Eurostat. 2013. "Labour Force Survey in the EU, Candidate and EFTA Countries - Main Characteristics of National Surveys, 2012." Eurostat. doi:10.2785/44503 (http://dx.doi.org/10.2785/44503).

Holt, D, and David Elliot. 1991. "Methods of Weighting for Unit Non-Response." *The Statistician*, 333–42.

Holt, D, and TMF Smith. 1979. "Post Stratification." *Journal of the Royal Statistical Society. Series A (General)*, 33–46.

Kalton, Graham, and Dalisay S Maligalig. 1991. "A Comparison of Methods of Weighting Adjustment for Nonresponse." In *Proceedings of the US Bureau of the Census 1991 Annual Research Conference*, 409–28.

Lehtonen, Risto, and Erkki Pahkinen. 2004. *Practical Methods for Design and Analysis of Complex Surveys*. John Wiley & Sons.

Little, Roderick JA. 1993. "Post-Stratification: a Modeler's Perspective." *Journal of the American Statistical Association* 88 (423). Taylor & Francis Group: 1001–12.

Lumley, Thomas. 2011. *Complex Surveys: A Guide to Analysis Using R*. Vol. 565. John Wiley & Sons.

Potter, Frank. 1990. "A Study of Procedures to Identify and Trim Extreme Sampling Weights." In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 758–63.

UNESCO. 2006. "International Standard Classification of Education (ISCED 1997)." UNESCO. http://www.uis.unesco.org/Library/Documents/isced97-en.pdf (http://www.uis.unesco.org/Library/Documents/isced97-en.pdf).