

Weights for the Hellenic Panel study of EES 2014

Ioannis Andreadis

ELNES 2014 was conducted as a web survey on a non-probability sample. As a result, using ELNES 2014 without any post-stratification adjustments would give biased estimates. This document describes how ELNES 2014 weights have been constructed.

Participants in ELNES 2014 are volunteers who have indicated that they wish to participate in web surveys conducted by the Laboratory of Applied Political Research, Aristotle University Thessaloniki (Andreadis, 2010), by registering here (<http://get.epolls.gr/index.php?sid=14476>). Most of the volunteers have registered after using the Greek Voting Advice Application *HelpMeVote* (Andreadis, 2013b).

Gender

There are more male than female HelpMeVote users (Andreadis, 2013a). A similar gender bias is observed in most of the VAAs (Andreadis, Wall and Krouwel, 2014). The gender distribution in the unweighted sample of ELNES 2014 is presented in Table 1.

Table 1. Gender distribution in ELNES 2014 unweighted

##	Frequency	Relative
## Male	984	0.6959
## Female	430	0.3041

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 3 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	692.9
## 2	Female	721.1
## 3	NA	3.0

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, R., and Pahkinen, E. (2004) p.88-92; Holt and Smith (1979)) For instance, this ratio for the male group is: $692.86/984=0.7041$, 1.6113, 230.9533. These adjustments to the sampling weights makes 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 ELNES 2014 using a weight variable that is summarized in Table 3.

Table 3. Summary of weights after adjusting for gender

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.704	0.704	0.704	1.000	1.680	1.680

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

Table 4: Gender distribution in ELNES 2014 after weighting

##	Frequency	Relative
## Male	693	0.49
## Female	721	0.51
## NA	3	0.00

Age

Table 5. Age distribution in ELNES 2014 unweighted

##	Frequency	Relative
## 18-25	76	0.05860
## 26-40	584	0.45027
## 41-64	596	0.45952
## 65+	41	0.03161

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 ditribution 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). 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, 2010).

Table 7: Gender distribution in ELNES 2014 after weighting for gender and age

##	Frequency	Relative
## Male	693	0.49
## Female	721	0.51

Table 8: Age distribution in ELNES 2014 after weighting for gender and age

##	Frequency	Relative
## 18-25	144	0.111
## 26-40	348	0.268
## 41-64	499	0.385
## 65+	306	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, but after the poststratification adjustment on these variables ELNES 2014 includes a weight variable that has a maximum value of 18.2 (Table 9).

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

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.228	0.348	0.554	1.000	1.040	18.200

There is a trade-off between the reduction of estimation bias and the increase in the sample variance arising from due to 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.

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 initially trimmed the weights to the value of 5 (item 9d) but because all capped cases were from the age group 65+ I have used the value of 8 (item 10a). After trimming the distributions of age and gender are not exactly the same with the corresponding population distributions, but they are close (see tables 10 and 11).

Table 10: Gender distribution in ELNES 2014 after trimming

##	Frequency	Relative
## Male	728	0.52
## Female	686	0.48

Table 11: Age distribution in ELNES 2014 after trimming

##	Frequency	Relative
## 18-25	147	0.113
## 26-40	369	0.285
## 41-64	521	0.403
## 65+	257	0.198

Education

The education levels of the unweighted sample are presented in Table 12.

Table 12. Education distribution in ELNES 2014 unweighted

##	Frequency	Relative
## ISCED97 (0-2)	12	0.008469
## ISCED97 (3-4)	336	0.237121
## ISCED97 (5-6)	1069	0.754411

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 34 250 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 (1 000) (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) is:

Table 13. Distribution of the population education levels

Education	Frequency	Relative
ISCED97(0-2)	2928.5	0.367
ISCED97(3-4)	3184.4	0.399
ISCED97(5-6)	1864.6	0.234

After the poststratification adjustment on gender, age and education ELNES 2014 includes a weight variable that has a maximum value of 241 (Table 14).

Table 14. Summary of weights after adjusting for gender, age and education

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.03	0.07	0.20	1.00	0.65	241.00

After trimming the weights the distributions of age and gender and education are far from with the corresponding population distributions (see tables 15, 16 and 17).

Table 15: Gender distribution in ELNES 2014 after trimming

##	Frequency	Relative
## Male	574	0.41
## Female	839	0.59

Table 15: Age distribution in ELNES 2014 after trimming

##	Frequency	Relative
## 18-25	153	0.117
## 26-40	468	0.358
## 41-64	599	0.459
## 65+	87	0.066

Table 16: Education distribution in ELNES 2014 after trimming

##	Frequency	Relative
## ISCED97 (0-2)	96	0.068
## ISCED97 (3-4)	656	0.463
## ISCED97 (5-6)	665	0.469

Collapsing levels

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 redpondents, 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.

ISCED97 levels 0-2 with a relative frequency of 0.008 should be combined with the next category “ISCED97 levels 3-4”

After the poststratification adjustment on gender, age and recoded education ELNES 2014 includes a weight variable that has a lower maximum value: 21.7 (Table 17).

Table 17. Summary of weights after adjusting for gender, age and education

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.110	0.122	0.438	1.000	1.230	21.700

After trimming the weights the distributions of age and gender and recoded education are far from with the corresponding population distributions (see tables 18, 19 and 20).

Table 18. Gender distribution in ELNES 2014 after trimming

##	Frequency	Relative
## Male	651	0.46
## Female	763	0.54

Table 19. Age distribution in ELNES 2014 after trimming

##	Frequency	Relative
## 18-25	151	0.118
## 26-40	405	0.315
## 41-64	558	0.434
## 65+	172	0.134

Table 20. Education distribution in ELNES 2014 after trimming

##	Frequency	Relative
## ISCED97 (0-4)	981	0.692
## ISCED97 (5-6)	436	0.308

Region

The distribution of the regions in the unweighted sample is

Table 21. Region distribution in ELNES 2014

##	Frequency	Relative
## Anatoliki Makedonia, Thraki	54	0.038
## Kentriki Makedonia	323	0.229
## Dytiki Makedonia	31	0.022
## Thessalia	75	0.053
## Ipeiros	26	0.018
## Ionia Nisia	12	0.008
## Dytiki Ellada	62	0.044
## Sterea Ellada	39	0.028
## Peloponnisos	46	0.033
## Attiki	660	0.467
## Voreio Aigaio	8	0.006
## Notio Aigaio	16	0.011
## Kriti	46	0.033
## Living Abroad	14	0.010

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 21. Modified region distribution in ELNES 2014

##		Frequency	Relative
##	Anatoliki Makedonia, Thraki	54	0.039
##	Kentriki & Dytiki Makedonia	354	0.253
##	Thessalia	75	0.054
##	Ipeiros & Ionia Nisia	38	0.027
##	Dytiki Ellada	62	0.044
##	Stereia Ellada	39	0.028
##	Peloponnisos	46	0.033
##	Attiki	660	0.472
##	Aigaio & Kriti	70	0.050

The sample regions should be distributed close to the expected distribution presented in Table 22 (According to the Hellenic Statistical Authority - ELSTAT (2013) publication of the 2011 population census).

Table 22. Modified region expected distribution (Census 2011)

##	Regions	Frequency	Relative
## 1	Anatoliki Makedonia, Thraki	78	0.055
## 2	Kentriki & Dytiki Makedonia	280	0.197
## 3	Thessalia	95	0.067
## 4	Ipeiros & Ionia Nisia	70	0.049
## 5	Dytiki Ellada	88	0.062
## 6	Stereia Ellada	71	0.050
## 7	Peloponnisos	74	0.052
## 8	Attiki	495	0.349
## 9	Aigaio & Kriti	147	0.104
## 10	Living Abroad	14	0.010
## 11	<NA>	5	0.004

After the poststratification adjustment on gender, age, recoded education and modified regions ELNES 2014 includes a weight variable that has a maximum value: 73.9 (Table 23).

Table 23. Summary of weights after adjusting for gender, age and recoded education and modified regions

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.00	0.01	0.06	1.00	0.22	73.90

After trimming the weights the distributions of age and gender and recoded education are far from with the corresponding population distributions (see tables 24, 25, 26 and 27).

Table 24. Gender distribution in ELNES 2014 after trimming

##		Frequency	Relative
##	Male	680	0.48
##	Female	733	0.52

Table 25. Age distribution in ELNES 2014 after trimming

##		Frequency	Relative
##	18-25	128	0.098
##	26-40	479	0.366
##	41-64	560	0.428
##	65+	142	0.108

Table 26. Education distribution in ELNES 2014 after trimming

##		Frequency	Relative
##	ISCED97 (0-4)	640	0.451
##	ISCED97 (5-6)	777	0.549

Table 27. Modified region distribution in ELNES 2014 after trimming

##		Frequency	Relative
##	Anatoliki Makedonia, Thraki	72	0.052
##	Kentriki & Dytiki Makedonia	327	0.236
##	Thessalia	101	0.073
##	Ipeiros & Ionia Nisia	26	0.019
##	Dytiki Ellada	98	0.071
##	Stereia Ellada	56	0.041
##	Peloponnisos	76	0.055
##	Attiki	523	0.376
##	Aigaio & Kriti	110	0.079

Valid votes

The distribution of valid votes in the unweighted sample is

Table 28. Valid votes distribution in ELNES 2014

##		Frequency	Relative
##	Coalition of Radical Left (SYRIZA)	376	0.335
##	New Democracy (ND)	131	0.117
##	Popular Association- Golden Dawn (XA)	71	0.063
##	Olive Tree- Democratic Alignment	65	0.058
##	The River	169	0.151
##	Communist Party of Greece (KKE)	51	0.045
##	Independent Greeks (AN.EL)	58	0.052
##	Popular Orthodox Rally (LAOS)	19	0.017
##	Other	181	0.161

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 29. Expected vote distribution

##	Parties	Frequency	Relative
## 1	Coalition of Radical Left (SYRIZA)	298	0.266
## 2	New Democracy (ND)	254	0.227
## 3	Popular Association- Golden Dawn (XA)	105	0.094
## 4	Olive Tree- Democratic Alignment	90	0.080
## 5	The River	74	0.066
## 6	Communist Party of Greece (KKE)	68	0.061
## 7	Independent Greeks (AN.EL)	39	0.035
## 8	Popular Orthodox Rally (LAOS)	30	0.027
## 9	Other	161	0.144

After the poststratification adjustment on gender, age, recoded education, modified regions and valid votes ELNES

2014 includes a weight variable that has a maximum value: 115 (Table 30).

Table 30. Summary of weights after adjusting for gender, age, recoded education, modified regions and valid votes

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.00	0.00	0.01	1.00	0.11	115.00

After trimming the weights the distributions of age and gender and recoded education are far from with the corresponding population distributions (see tables 31, 32, 33, 34, and 35).

Table 31. Gender distribution in ELNES 2014 after trimming

##	Frequency	Relative
## Male	734	0.52
## Female	678	0.48

Table 32. Age distribution in ELNES 2014 after trimming

##	Frequency	Relative
## 18-25	100	0.077
## 26-40	491	0.375
## 41-64	583	0.445
## 65+	135	0.103

Table 33. Education distribution in ELNES 2014 after trimming

##	Frequency	Relative
## ISCED97 (0-4)	582	0.41
## ISCED97 (5-6)	835	0.59

Table 34. Modified region distribution in ELNES 2014 after trimming

##	Frequency	Relative
## Anatoliki Makedonia, Thraki	68	0.049
## Kentriki & Dytiki Makedonia	318	0.229
## Thessalia	77	0.055
## Ipeiros & Ionia Nisia	30	0.021
## Dytiki Ellada	98	0.071
## Sterea Ellada	52	0.037
## Peloponnisos	77	0.056
## Attiki	560	0.404
## Aigaio & Kriti	108	0.078

Table 35. Vote distribution in ELNES 2014 after trimming

##	Frequency	Relative
## Coalition of Radical Left (SYRIZA)	312	0.282
## New Democracy (ND)	207	0.187
## Popular Association- Golden Dawn (XA)	87	0.079
## Olive Tree- Democratic Alignment	91	0.082
## The River	133	0.121
## Communist Party of Greece (KKE)	38	0.035
## Independent Greeks (AN.EL)	50	0.045
## Popular Orthodox Rally (LAOS)	30	0.027
## Other	159	0.144

And here is the summary of the final weight variable:

Table 36. Summary of final trimmed weights

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.535	0.536	0.545	1.000	0.642	8.500

Acknowledgements

This report reflects some of the knowledge I have acquired while I was a Fulbright Visiting Scholar at the University of Michigan working on the project: “Establishing the Hellenic (Greek) National Election Studies”. I would like to thank my host Dave Howell (<https://mcommunity.umich.edu/#profile:dahowell>) at the Center for Political Studies (<http://www.isr.umich.edu/cps/index.html>) and James Wagner (<http://psm.isr.umich.edu/wagner>) for introducing me to raking.

References

- Andreadis, I. (2010). Διαδικτυακές Πολιτικές Έρευνες [Web-based Political Surveys]. Proceedings of the 23rd Panhellenic Statistics Conference: “Statistics and Internet”. Veroia, 7-11 April 2010, <http://invenio.lib.auth.gr/record/126916/files/Web-based-surveys.pdf> (<http://invenio.lib.auth.gr/record/126916/files/Web-based-surveys.pdf>)
- Andreadis, I. (2013a) Who responds to website visitor satisfaction surveys? General Online Research Conference GOR 13 March 04-06, 2013, DHBW Mannheim, Germany <http://www.polres.gr/en/sites/default/files/GOR13.pdf> (<http://www.polres.gr/en/sites/default/files/GOR13.pdf>)
- Andreadis, I. (2013b) Voting Advice Applications: a successful nexus between informatics and political science. BCI '13, September 19 - 21 2013, Thessaloniki, Greece <http://www.polres.gr/en/sites/default/files/BCI-2013.pdf> (<http://www.polres.gr/en/sites/default/files/BCI-2013.pdf>)
- Andreadis, I., Wall, M. and Krouwel, A. (2014). Who are the users of voting advice applications? 2014 MPSA conference, Chicago April 2-6, 2014 <http://www.polres.gr/en/sites/default/files/MPSA2014.pdf> (<http://www.polres.gr/en/sites/default/files/MPSA2014.pdf>)
- DeBell, M., & Krosnick, J. A. (2009). Computing weights for American national election study survey data. ANES Technical Report series, no. nes012427. Ann Arbor, MI, and Palo Alto, CA: American National Election Studies. Available at <http://www.electionstudies.org>.
- Eurostat (2013). Labour force survey in the EU, candidate and EFTA countries - Main characteristics of national surveys, 2012. doi:10.2785/44503 (doi:10.2785/44503)
- Hellenic Statistical Authority - 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)
- Holt, D., & Smith, T. M. F. (1979). Post stratification. Journal of the Royal Statistical Society. Series A (General), 33-46.
- Kalton, G., & Maligalig, D. S. (1991). A comparison of methods of weighting adjustment for nonresponse. In

Proceedings of the US Bureau of the Census 1991 annual research conference (pp. 409-428).

Lehtonen, R., & Pahkinen, E. (2004). Practical methods for design and analysis of complex surveys. John Wiley & Sons.

Lumley, T. (2010). Complex surveys: A guide to analysis using R (Vol. 565). John Wiley & Sons.

Little, R. J. (1993). Post-stratification: a modeler's perspective. *Journal of the American Statistical Association*, 88(423), 1001-1012.

Potter, F. (1990). A study of Procedures to Identify and Trim Extreme Sampling Weights, *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 225-230.

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