

Weighting European Social Survey Data

1. Do tables run on the ESS website need to be weighted?

Almost certainly yes. Most users will want results based on samples which represent the country (or group of countries) they are interested in as closely as possible – which means that the most accurate estimates will be obtained only after first weighting the data.

When attempting to tabulate two (or more) variables on the website, the first tables that appear will be *unweighted* and therefore potentially misleading. So the straightforward procedures for generating *weighted* tables will need to be followed (**see para 6 below**).

2. What weights are there to apply?

There are two weights that may and often should be 'switched on':

- a) The first is the *Design* weight (DWEIGHT)
Several of the sample designs used by countries participating in the ESS were not able to give all individuals in the population aged 15+ precisely the same chance of selection. Thus, for instance, the unweighted samples in some countries over- or under-represent people in certain types of address or household, such as those in larger households. The design weight corrects for these slightly different probabilities of selection, thereby making the sample more representative of a 'true' sample of individuals aged 15+ in each country. The *Design* weights are computed as normed inverse of the inclusion probabilities. For more information, see the ESS Documentation Report.
- b) The second weight is the *Population size* weight (PWEIGHT). This is used when examining data for two or more countries combined.
This weight corrects for the fact that most countries taking part in the ESS have very similar sample sizes, no matter how large or small their population. Without weighting, any figures combining two or more country's data would be incorrect, over-representing smaller countries at the expense of larger ones. So the *Population size* weight makes an adjustment to ensure that each country is represented in proportion to its population size. The *Population size* weight is calculated as $PWEIGHT = \frac{\text{Population size aged 15 years and above}}{[(\text{Net sample size in data file}) * 10\ 000]}$.

Users should note that the two weights currently available do not adjust for non-response in the sample. (Weighting for non-response is sometimes done to reduce differences between the sample and original population, caused by variations in response between different groups in the sample, e.g. age groups.)

3. Which weights should be 'switched on' for different purposes?

Looking at data only from *one* round of the ESS:

When computing any tables or percentages, we recommend that you use *weighted* data. However, different types of tables require different *combinations* of weights. The general rules of thumb are:

- when looking at data for one country alone, only the design weight need be applied
- when comparing data from two or more countries but without reference to the average (or combined total) of those countries, only the design weight need be applied
- when comparing data of two or more countries and *with* reference to the average (or combined total) of those countries, both design and population size weights should be applied
- when combining countries into a group, such as 'accession countries' or 'EU member states', both design and population size weights should be applied.

Examples of different types of analyses and the weights recommended for them are given

	<i>Example – voter turnout (% of respondents voting in the last election)</i>	Weights to be used	
		Design weight	Population weight
a) To examine data from a single country – whether a single variable or a cross-tabulation	Voter turnout in Germany	X	
	Voter turnout in Germany by age and gender	X	
b) To compare results for two or more countries separately – without using totals or averages	Compare voter turnout in France, Germany and the UK.	X	
c) To combine countries – whether on a single variable or via a cross-tabulation	i) Voter turnout in Scandinavia.	X	X
	ii) Voter turnout in the EU.	X	X
	iii) Voter turnout across all countries participating in the ESS.	X	X
	iv) Compare voter turnout between EU member states and accession countries.	X	X
	v) Voter turnout by age group across all ESS participating countries.	X	X

Looking at data from *two* rounds of the ESS:

Again, if any tables or percentages are computed, we recommend to use *weighted* data. However, different types of tables require different *combinations* of weights. The general rules of thumb are:

- if looking at data for one single country in two rounds, only the design weights need be applied

- if comparing data from two or more countries but without reference to the average (or combined total) of those countries, only the design weights need be applied
- if comparing data of two or more countries and *with* reference to the average (or combined total) of those countries, both design and population size weights of both rounds should be applied
- if combining countries into a group, such as ‘accession countries’ or ‘EU member states’, both design and population size weights of both rounds should be applied.

	<i>Example – voter turnout (% of respondents voting in the last election)</i>	Design weight Round 1	Weights to be used		Population weight Round 2
			Population weight Round 1	Design weight Round 2	
a) To examine data from a single country – whether a single variable or a cross-tabulation	Voter turnout in Germany	X		X	
	Voter turnout in Germany by age and gender	X		X	
b) To compare results for two or more countries separately – without using totals or averages	Compare voter turnout in France, Germany and the UK.	X		X	
c) To combine countries – whether on a single variable or via a cross-tabulation	i) Voter turnout in Scandinavia.	X	X	X	X
	ii) Voter turnout in the EU.	X	X	X	X
	iii) Voter turnout across all countries participating in the ESS.	X	X	X	X
	iv) Compare voter turnout between EU member states and accession countries.	X	X	X	X
	v) Voter turnout by age group across all ESS participating countries.	X	X	X	X

4. Some considerations when analysing groups of countries

WARNING Users who are interested in combining data for groups of countries (e.g. EU member states or accession countries) should note that this was not the primary aim of the ESS design (which instead aimed to facilitate comparisons *between* individual countries). The population size weights enable the estimation of these ‘combined totals’, but users should note that these estimates may have relatively high margins of error. We would generally advise checking the margins of error associated with such estimates before drawing firm conclusions.


Another point to note is that the ‘total’ column of any table is merely an aggregate of those countries included in that table. It cannot be automatically assumed to be a figure for “Europe” or

“the EU”. For example, not all the EU member states are included in the first release of the data. Conversely, some countries participating in the ESS would be outside some users’ definition of “Europe”. Users who wish to look at combined country tables must satisfy themselves that (a) all the countries they require, and (b) only those they require, are present in the tables they run. (When running tables, it is possible to select a sub-set of countries.)

5. Describing sample sizes

The sample size is contained in the N (Total) row of the **unweighted tables, NOT in the weighted tables**. So, even when quoting weighted rather than unweighted *percentages*, avoid quoting weighted **Ns** (*numbers of cases*), which are not meaningful in themselves.

6. How to weight tables on the ESS website?

Having first selected the table you wish to run, the *unweighted* table will appear as a default. Then click on the Weight icon . The top left-hand box (weighting variables defined in the dataset) then lists the pre-defined weights, the *Design weight* and the *Population size weight*. Highlight the weighting variable you require, and then click on the right-pointing arrow to move it to the right-hand box. You can do this sequentially to select more than one weighting factor for a table. Then press the OK button, and the weighted table will appear. A footnote will indicate that weights have been applied.

7. How does weighting affect the data?

Applying weights to raw data will change any measures based on the data, such as:

- the distribution of variables (descriptive statistics);
- percentages in cross-tabulations (tabulations);
- average measures such as the mean or mode;
- measures of spread and variance, such as the standard deviation or standard error of a variable, or the ‘margin of error’;
- tests of statistical significance for differences between e.g. countries or years.

If you analyse data from complex designs, please bear in mind: The confidence intervals computed in Nesstar WebView 3.0 are only correct for data collected by simple random sampling. For more complex sample designs, design effects should be taken into consideration. A detailed description of the ESS sample designs can be found in the ESS Documentation Report.

8. Do weights need to be applied for more sophisticated analyses?

For multivariate analyses the same weighting rules apply. The data should always be weighted: the decision between using design weights alone or design weights plus population weights depends upon the purpose of the analysis. Analysis that aims to compare countries should use the design weight; analysis that is based on combining data from countries should use the design weight plus the population weight.

References

- Chambers, Ray and Chris Skinner, 2003: *Analysis of Survey Data*. Chichester: Wiley
- Korn, Edward and Barry Graubard, 1999: *Analysis of Health Surveys*. Chichester: Wiley
- Lohr, Sharon L., 1999: *Sampling: Design and Analysis*. Duxbury Press
- Särndal, Carl-Erik, Bengt Swensson and Jan Wretman, 1992: *Model Assisted Survey Sampling*. New York: Springer.

9. Running tables using other statistical packages?

If downloading data for use by another statistical package, it will be necessary to check first how that package handles weights.

Please note that each weight provided in the data file corrects for one factor at a time. On the ESS web site, it is possible to combine the weights by specifying more than one weight to be used. However, other packages may allow the use of only one weight at a time. In this case, if combined weights are required, a new weight variable will need to be created by multiplying together the existing weights in the data file required.