

Sampling Procedure and Weighting for ESS Round 10 (Ukraine)

As a sampling frame, electoral statistics from the 2019 Extraordinary Parliamentary Elections was used to select primary and secondary territorial units. This provided the most recent official information about Ukrainian adults (aged 18+).

The frame is stratified by 25 administrative regions (24 oblasts and the capital, Kyiv) and by area type (urban or rural), resulting in 49 strata (24 oblasts by two urbanity categories, plus Kyiv as a separate stratum). The sampling design is three-stage. In the first stage, 300 voting precincts were randomly selected from each stratum. In the second stage, within each selected precinct, households were chosen using the “Random Walk” method. In the final stage, within each selected household, one individual aged 15 or older was randomly chosen for an interview. If the selected person was not at home, the interviewer scheduled a follow-up visit, with up to four contact attempts per dwelling. The total sample size is 1531¹.

Once the sampling was completed, the design weight, post-stratification weight, population weight, and analysis weight were calculated (*Weighting | European Social Survey*, n.d.).

The *design weight* ($dweight_i$ for each respondent i) was calculated as the inverse of the inclusion probability (p_i for each respondent i):

$$dweight_i = \frac{1}{p_i}$$

Then the design weight was scaled ($dweight_i^{scaled}$ for each respondent i). The sum of all weights should be equal to the number of respondents:

$$dweight_i^{scaled} = \frac{1531 \cdot dweight_i}{\sum dweight_i}$$

The inclusion probability p_i is the multiplication of the inclusion probabilities at each stage of selection:

$$p_i = p_j^{s1} p_i^{s2} p_i^{s3}$$

For the first sampling stage, the probability p_j^{s1} of each voting precinct j being selected within the given stratum was computed:

$$p_j^{s1} = \frac{n_j c_j}{N_j^{strat}}$$

n_j – the size of the voting precinct j (number of voters); c_j – the number of voting precincts selected within the stratum to which the voting precinct j belongs; N_j^{strat} – the size of the stratum (number of voters in the stratum) to which the voting precinct j belongs.

¹In formulas below 1531 is a sample size.

At the second sampling stage, the probability p_i^{s2} of each household i being included in the selected voting precincts was calculated:

$$p_i^{s2} = \frac{k_i}{N_i^{prec}}$$

k_i – the number of selected households within the voting precinct to which the household i belongs; N_i^{prec} – the estimated total number of households in the precinct to which the household i belongs.

The probability p_i^{s3} of each respondent i being included in the selected households was calculated in the third sampling stage:

$$p_i^{s3} = \frac{1}{N_i^{15+}}$$

N_i^{15+} – the total number of people aged 15 and older living in the household i .

The *post-stratification weight* is based on the design weight ($dweight_i^{scaled}$) with raking applied using the “anesrake” package in R (Pasek, 2018). The control variables were used such as gender (1=Male, 2=Female), age (1=15-34 years old, 2=35-54 years old, and 3=55+ years old), and region² (1=Kyiv city, 2=North, 3=Center, 4=North East, 5=North West, 6=South East, 7=West, 8=South West, 9=South, 10=East). A cross-classified variable was created for gender and age (*Documentation of ESS Post-Stratification Weights*, 2014; *ESS6 - 2012 Documentation Report Edition 2.4*, 2018). To calculate the proportions for age, gender, and region in the Ukrainian population aged 15 and older, data from the State Statistics Service of Ukraine (beginning of 2022) were used (*State Statistics Service of Ukraine*, n.d.). Trimming procedures were applied in raking, and weights were capped at 4 (*Documentation of ESS Post-Stratification Weights*, 2014).

The code in R for raking is the following:

```
anesrake(target, data, caseid = data$id, weightvec = data$dweight, cap = 4)
```

The `anesrake` function performs the raking procedure. The `target` parameter is a list of control variables, including the cross-classified gender and age variable and region. The `data` parameter is a dataframe containing the control variables, design weights, and respondent IDs. The `caseid` parameter is a vector of respondent IDs, while `weightvec` is a vector of design weights. The `cap` parameter sets a maximum value for the post-stratification weights (they cannot exceed 4). The output of this function is the post-stratification weights (p_{spwght_i} for each respondent i).

The *population weight* ($pweight$) was computed as follows:

$$pweight = \frac{p^{15+}}{1531 \cdot 10000}$$

² Previously, the regions of Ukraine were grouped into 11 regions. However, in the beginning of 2022, due to the occupation of the Autonomous Republic of Crimea and parts of the Donetsk and Luhansk regions, 10 regions were identified.

p^{15+} – the number of people in the Ukrainian population aged 15 and older; based on data from the State Statistics Service of Ukraine at the beginning of 2022 (*State Statistics Service of Ukraine*, n.d.).

Finally, the *analysis weight* ($anweight_i$) is a product of the post-stratification ($pspwght_i$) and the population weight ($pweight$):

$$anweight_i = pspwght_i \cdot pweight$$

References

1. *Documentation of ESS Post-Stratification Weights*. (2014). https://www.europeansocialsurvey.org/sites/default/files/2023-06/ESS_post_stratification_weights_documentation.pdf
2. *ESS6 - 2012 Documentation Report Edition 2.4*. (2018). https://stessrelpubprodwe.blob.core.windows.net/data/round6/survey/ESS6_data_documentation_report_e02_4.pdf
3. Pasek, J. (2018). *anesrake: ANES Raking Implementation*. R package version 0.80. <https://CRAN.R-project.org/package=anesrake>
4. *State Statistics Service of Ukraine*. (n.d.). <https://www.ukrstat.gov.ua/>
5. *Weighting | European Social Survey*. (n.d.). <https://www.europeansocialsurvey.org/methodology/ess-methodology/data-processing-and-archiving/weighting>