

Metodologiczne Inspiracje 2021

Badania ilościowe w naukach społecznych - wyzwania i problemy



UNIwersytet
Warszawski

Projektowanie procedur w porównawczych badaniach sondażowych

Przypadek piątej edycji European Values Study

Krzysztof Bulkowski

Katedra Modeli Formalnych i Metod Ilościowych w Socjologii

Wydział Socjologii UW

Sondażowe badania porównawcze

Cel nadrzędny: możliwość porównań między krajami

- Osiągnięcie możliwie najbardziej miarodajnych wyników w każdym z krajów
- Minimalizacja czynników zaburzających porównywalność
- Konieczność ustalania procedur standaryzujących proces badawczy i kontroli ich przestrzegania
- Rola instytucji zarządzającej projektem
 - Wewnętrznie – planowanie i spójność poszczególnych etapów procesu badawczego
 - Zewnętrznie – dobra komunikacja z zespołami krajowymi, jasne zasady, reguły, wytyczne

Efekt: dobra jakość zebranych danych, pełna dokumentacja, możliwość bezpośredniego porównywania wyników między krajami

Plan prezentacji

- Kilka słów o EVS
- Motywacja do przygotowania wystąpienia
- Założenia doboru próby
- Wagi w udostępnionych zbiorach danych
- Opis doboru próby w dokumentacji
- Harmonogram badania
- Diagnoza problemów i rekomendacje

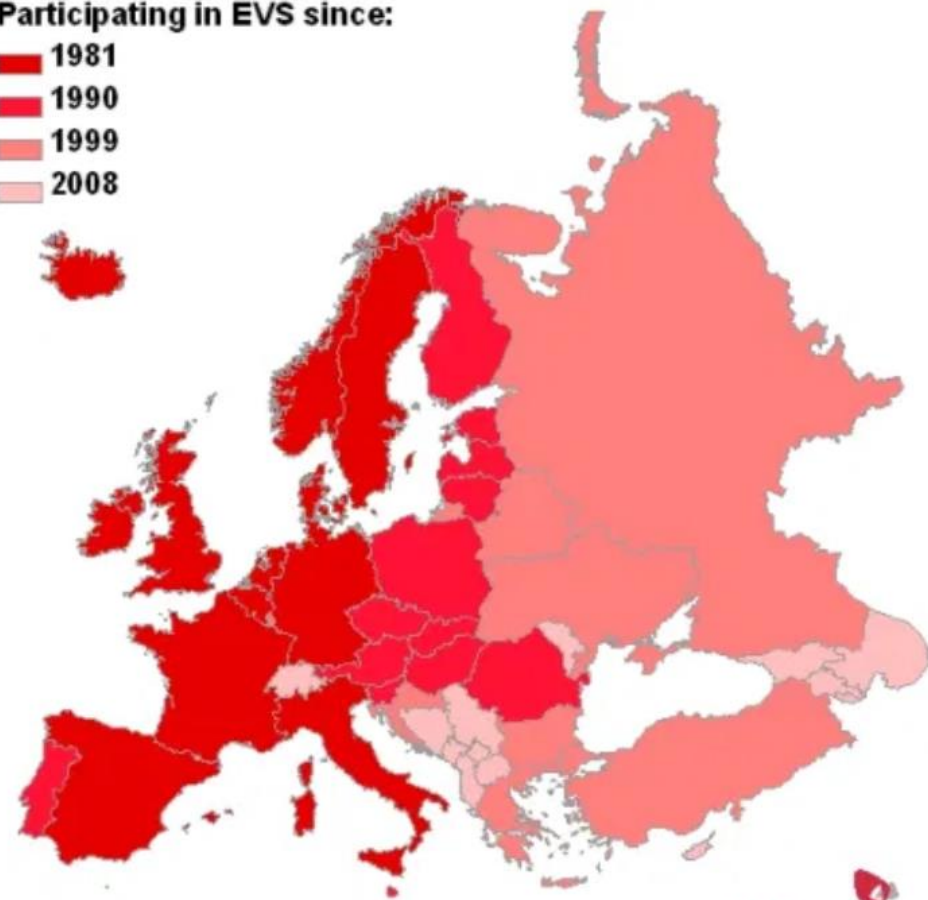
European Values Study

- Europejskie Badanie Wartości
- Populacja: 18+
- Główne tematy: rodzina, praca, środowisko, polityka, społeczeństwo, religia, moralność, tożsamość narodowa
- Realizacja od 1981 roku co 9 lat
- W 2008 roku w badaniu wzięło udział 47 krajów
- W 2017 roku – ścisła kooperacja z WVS

Kraje uczestniczące w EVS

Participating in EVS since:

- 1981
- 1990
- 1999
- 2008



European *Values* Study

Participating countries in EVS 2017-2020



Integrated dataset Country dataset

Źródło: <https://europeanvaluesstudy.eu/about-evs/participating-countries/>

Motywacja

Jabłonna 2018:

Piotr Jabkowski, Instytut Socjologii UAM, *O zróżnicowaniu praktyk sondażowych w międzykrajowych badaniach porównawczych na podstawie metodologicznej archiwizacji 1537 surveyów zrealizowanych w latach 1981-2017 w krajach europejskich.*

Artykuł 2020:

Jabkowski, P., Kołczyńska, M. (2020). Sampling and Fieldwork Practices in Europe: Analysis of Methodological Documentation From 1,537 Surveys in Five Cross-National Projects, 1981-2017. *Methodology*. 16. 186-207.

Podstawowe informacje o metodologii

The main features of the EVS Methodology are the following:

- *Mode of data collection: F2F interview. In 2017 mixed-mode has been introduced as an experiment*
- *Sample size: effective size=1000 in 1981; increased up to 1500 in 2008; 1200 in 2017*
- *Sampling: the use of quotas was admitted in the first waves, but since 2008 **only probabilistic representative sample***
- *Target: Resident population older than 18 years old*
- *Translation and monitoring: centrally coordinated since 2008*
- *Data Access: data and documentation of all the waves are available free of charge.*

Plan doboru próby (*sampling plan*)

- Populacja
- Opis procedury doboru próby
- Operat losowania
- Warstwy
- Schemat losowania na poszczególnych poziomach
- Planowanie wielkości próby:
 - wielkość próby (effective sample size), $N = 1200$
 - *design effect* (wielkość wiązki, współczynnik korelacji wewnątrz-klasowej (ICC))
 - przewidywany *response rate*
 - przewidywany *ineligible rate*
- Prawdopodobieństwa inkluzji (wzory)
- Alokacja próby w poszczególnych warstwach

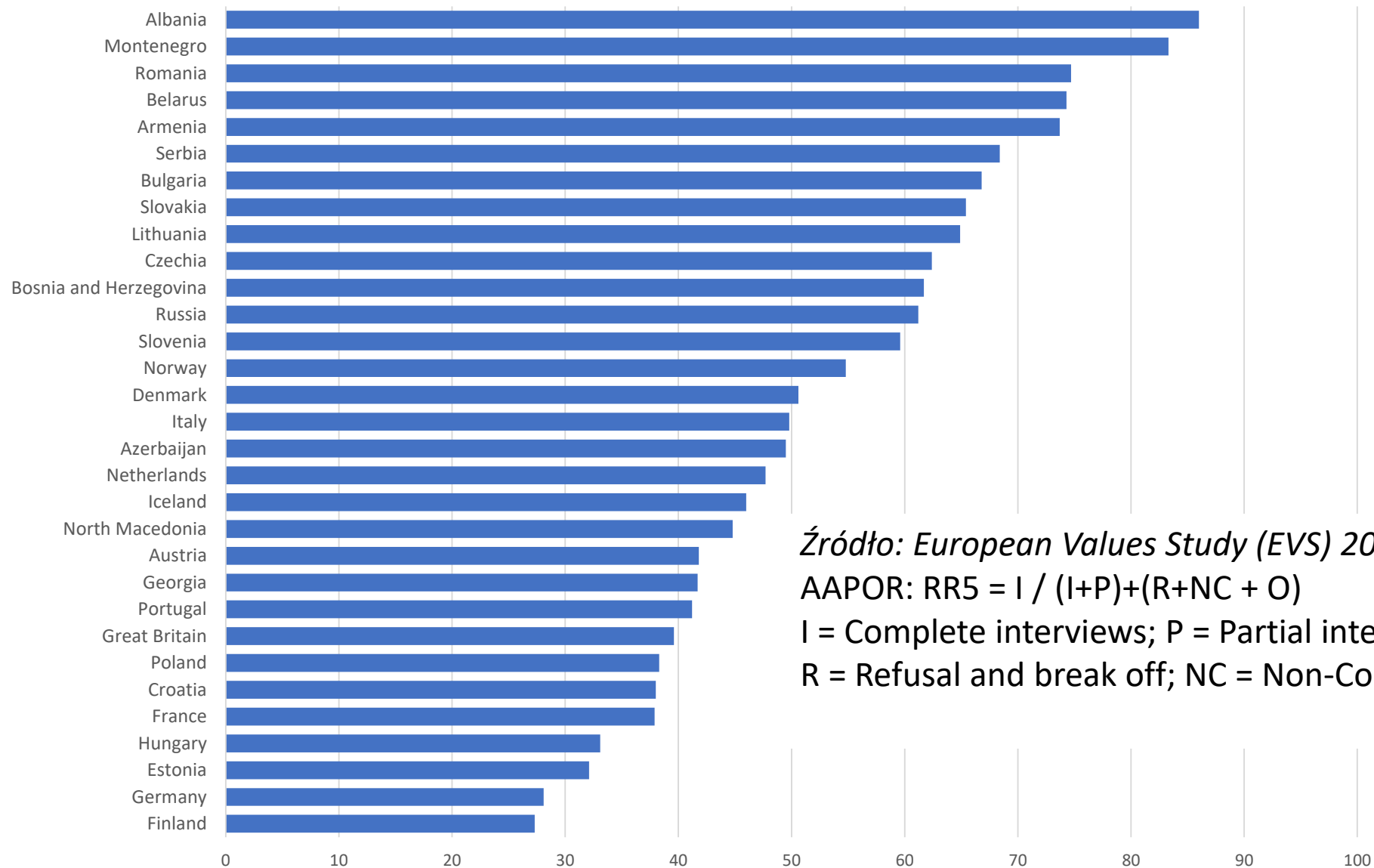
Realizacja terenowa

Polska:

- 12.2017-01-2018
- N (wylosowane) = 4113 (wiązka 9)
- N (zrealizowane) = 1352
- Technika CAPI

- Zbiór danych (N=1352) zawierający dane zebrane za pomocą kwestionariusza
- Zbiór danych (N=4113) zawierający dla każdego respondenta informację o:
 - każdej wizycie ankietera (na podstawie zestandaryzowanego formularza kontaktowego)
 - prawdopodobieństwie inkluzji na każdym etapie losowania

Poziom realizacji próby (RR5)



Źródło: *European Values Study (EVS) 2017: Method Report*

AAPOR: $RR5 = I / (I+P)+(R+NC + O)$

I = Complete interviews; P = Partial interviews;

R = Refusal and break off; NC = Non-Contact; O = Other

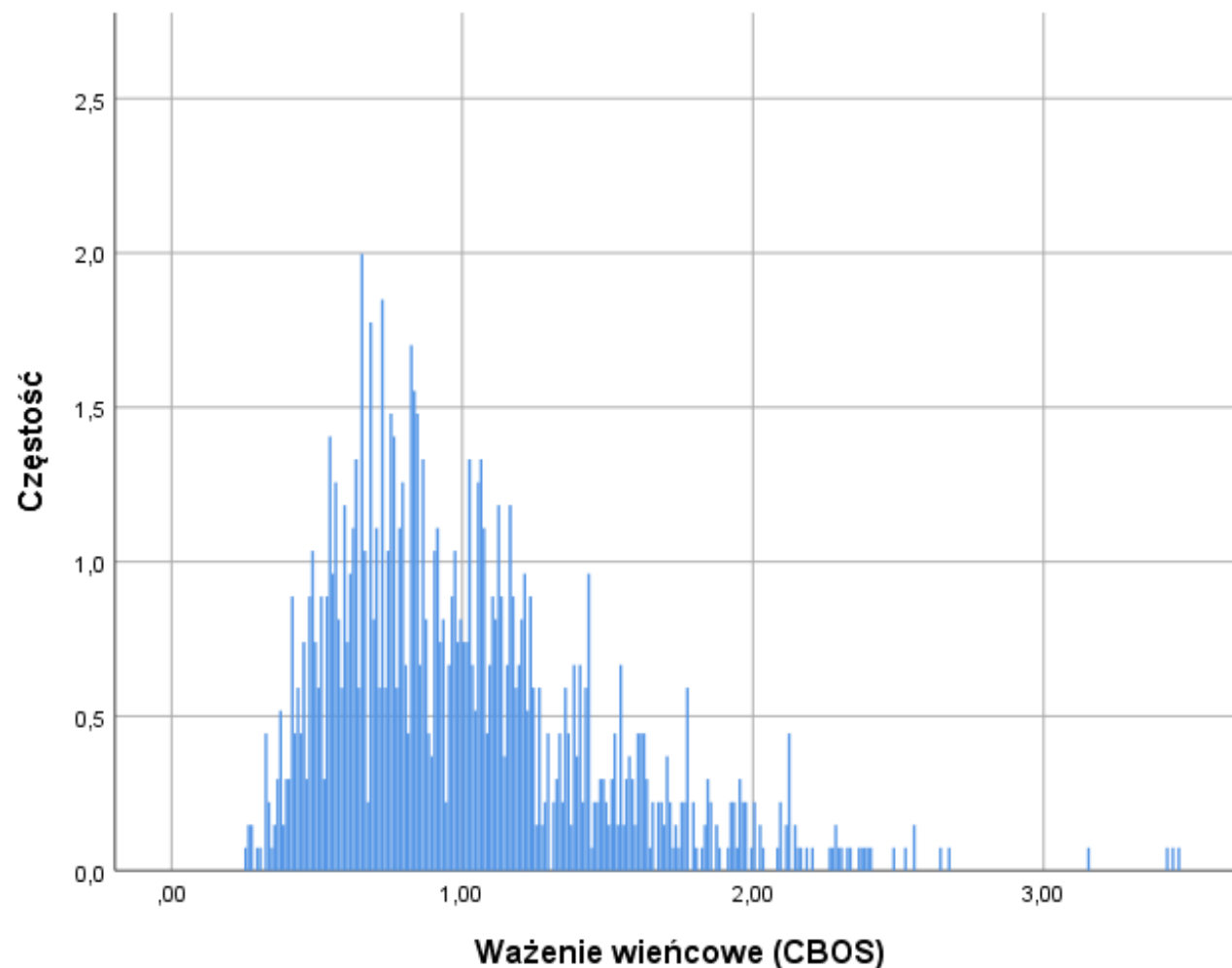
Ważenie CBOS

Ważenie wieńcowe (*rim weighting*) – iteracyjne korygowanie rozkładów brzegowych do założonych rozkładów populacyjnych.

- województwo zamieszkania (16)
- wykształcenie (4)
- klasa wielkości miejscowości zamieszkania (6)
- grupy wiekowe (6)
- płeć (2)

Dane o populacji pochodziły z dwóch źródeł:

- rejestr PESEL - stan na 18 lipca 2017 r.
- Badanie Aktywność Ekonomiczna Ludności (BAEL) III kwartał 2017 r.



Publikacja zbiorów danych z wagami (lipiec 2019)

spodziewane zmienne na podstawie ESS

- Design weight
- Post-stratification weight including design weight
- Population size weight
- Analysis weight

Jabłonna 2014:

Tomasz Żółtak, Instytut Badań Edukacyjnych oraz Instytut Filozofii i Socjologii PAN,
Wagi poststratyfikacyjne w Europejskim Sondażu Społecznym — możliwości i ograniczenia prawidłowego wykorzystania w analizach statystycznych

Publikacja zbiorów danych z wagami (lipiec 2019)

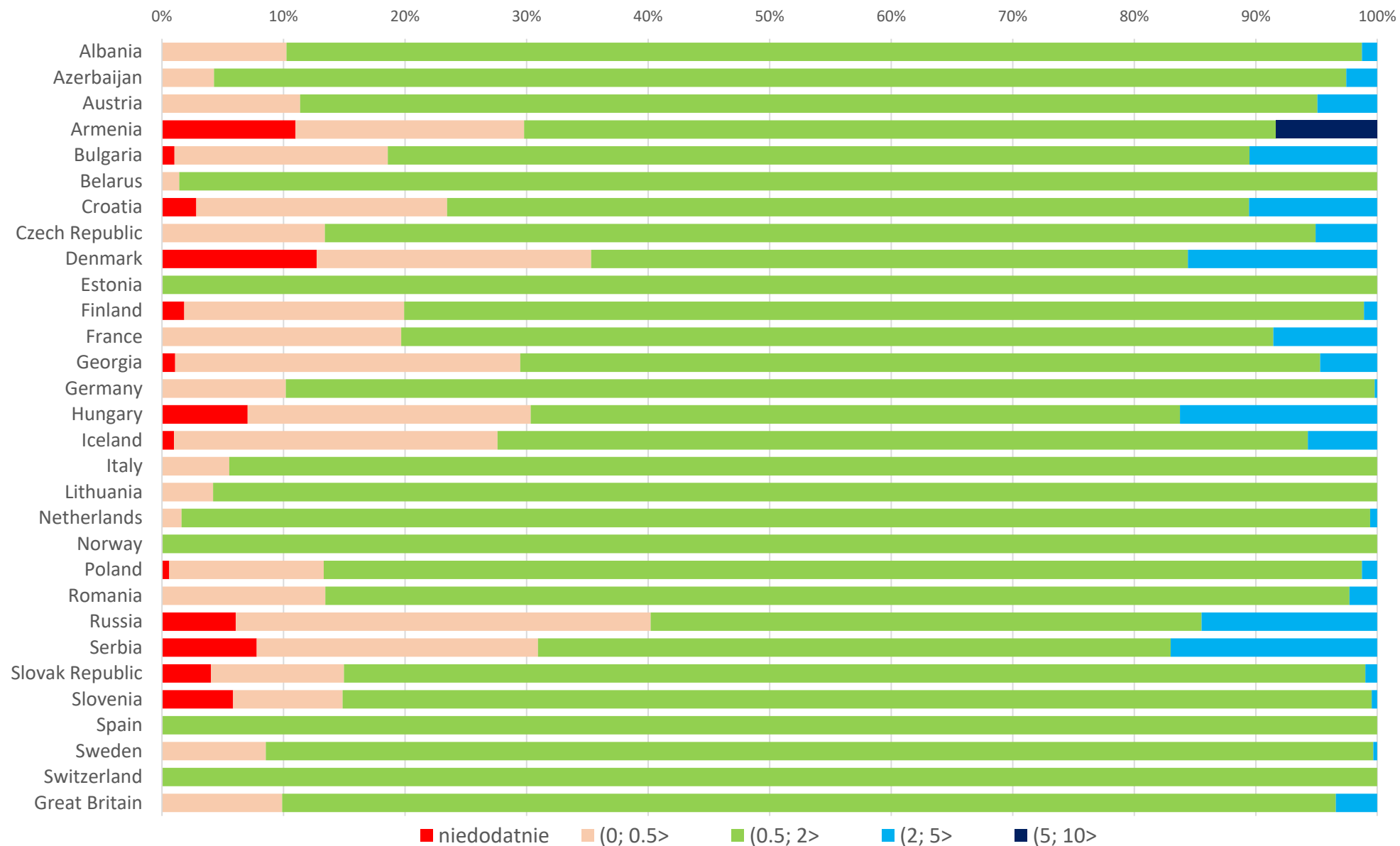
zawartość zbioru danych

- Zbiór danych zawierający dane z 30 krajów wraz z dokumentacją ważenia (algorytm)
- Zmienne:
 - waga kalibracyjna (gweight)
 - population size weights (pweight)
- Wartości trzech zmiennych (płeć, kategorie wiekowe, poziom wykształcenia), na podstawie których została przeprowadzona kalibracja

Parametry wag

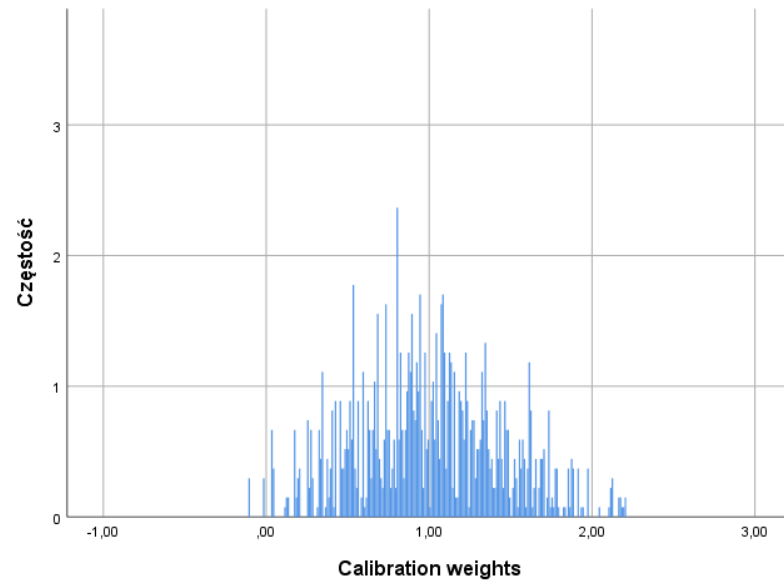
Country	Minimum	Maksimum	Średnia	Odchylenie standardowe	N
Albania	0,03	2,55	0,9994	0,44053	1435
Azerbaijan	0,29	2,98	1,0000	0,41201	1800
Austria	0,30	2,99	1,0002	0,53644	1644
Armenia	-0,87	7,33	1,0131	1,70479	1500
Bulgaria	-0,04	3,15	1,0045	0,63517	1560
Belarus	0,46	1,29	0,9999	0,15563	1548
Croatia	-0,08	3,02	0,9969	0,62020	1488
Czech Republic	-0,10	3,09	0,9995	0,52284	1812
Denmark	-1,05	3,87	0,9978	0,90312	3362
Estonia	0,52	1,73	1,0028	0,32784	1304
Finland	-0,07	2,73	1,0025	0,50869	1199
France	0,08	2,46	1,0000	0,59805	1870
Georgia	-0,18	2,50	1,0014	0,61898	2194
Germany	0,05	2,24	0,9995	0,37504	5407
Hungary	-0,64	4,11	0,9996	0,95611	1516
Iceland	-0,03	2,67	0,9989	0,60792	2506
Italy	0,15	1,93	1,0027	0,37505	2277
Lithuania	0,48	1,82	0,9997	0,27652	1448
Netherlands	0,39	2,18	1,0000	0,28695	686
Norway	0,62	1,68	1,0003	0,29394	1123
Poland	-0,11	2,20	1,0000	0,44224	1352
Romania	0,06	2,23	0,9982	0,43537	1614
Russia	-0,93	3,98	0,9987	0,97039	1825
Serbia	-1,19	4,38	1,0000	1,11622	1500
Slovak Republic	-0,37	2,10	1,0003	0,46307	1435
Slovenia	-0,68	2,11	1,0008	0,51290	1076
Spain	0,53	1,60	1,0001	0,22977	1212
Sweden	0,08	2,13	1,0008	0,35899	1194
Switzerland	0,62	1,72	0,9999	0,26955	3660
Great Britain	0,12	2,76	1,0031	0,44493	1788

Rozkład wagi podzielonej na kategorie



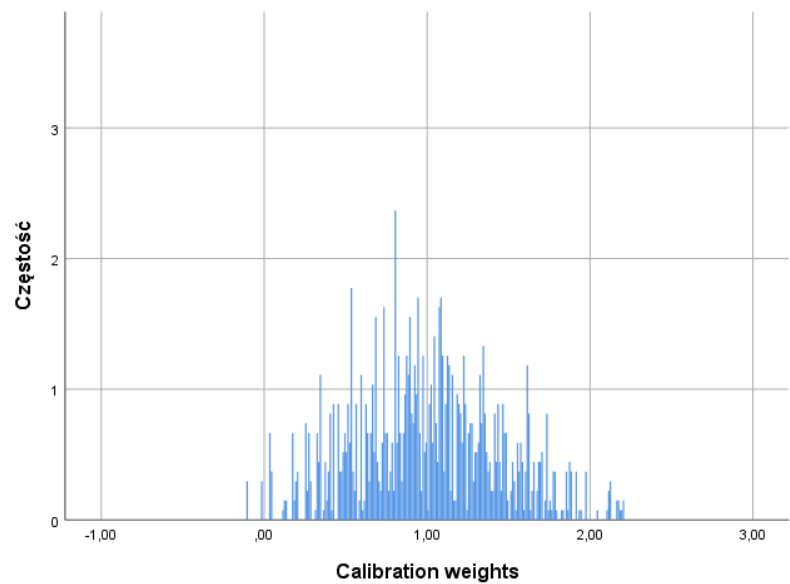
Porównanie rozkładu wartości wag (1/2)

Polska

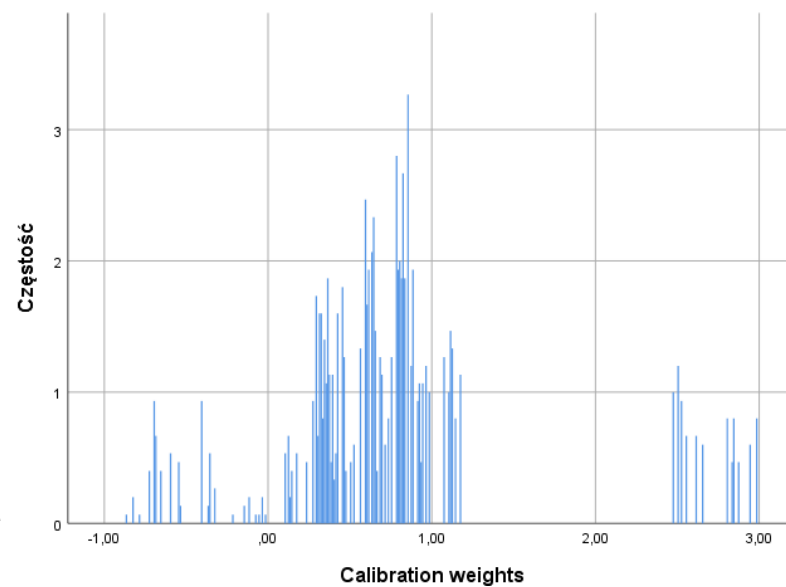


Porównanie rozkładu wartości wag (1/2)

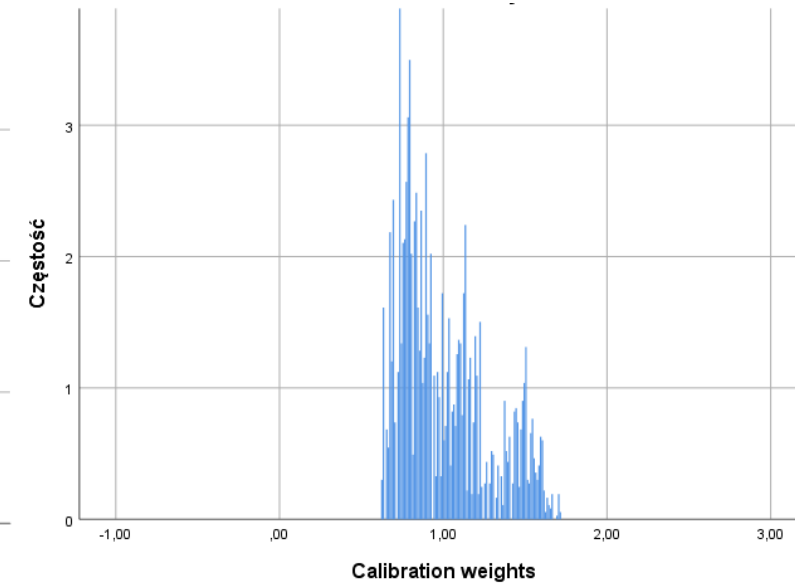
Polska



Serbia

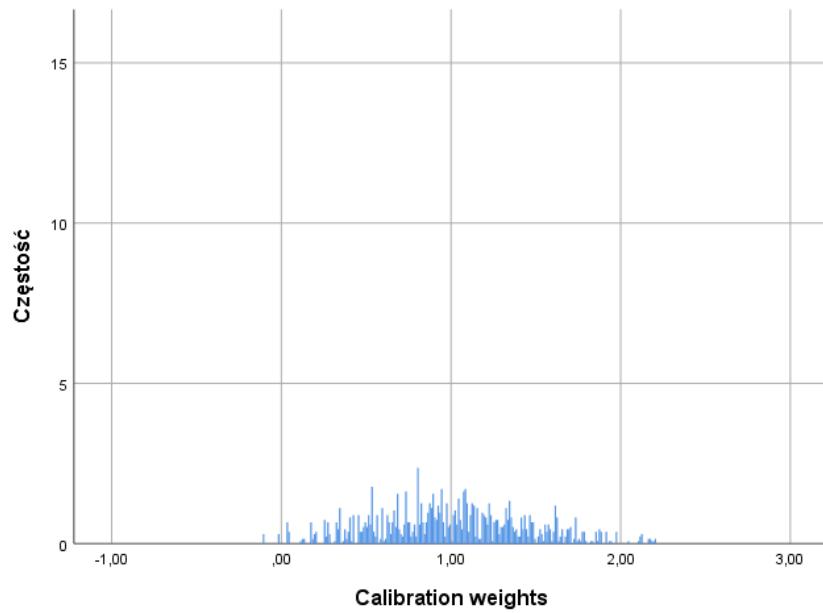


Szwajcaria



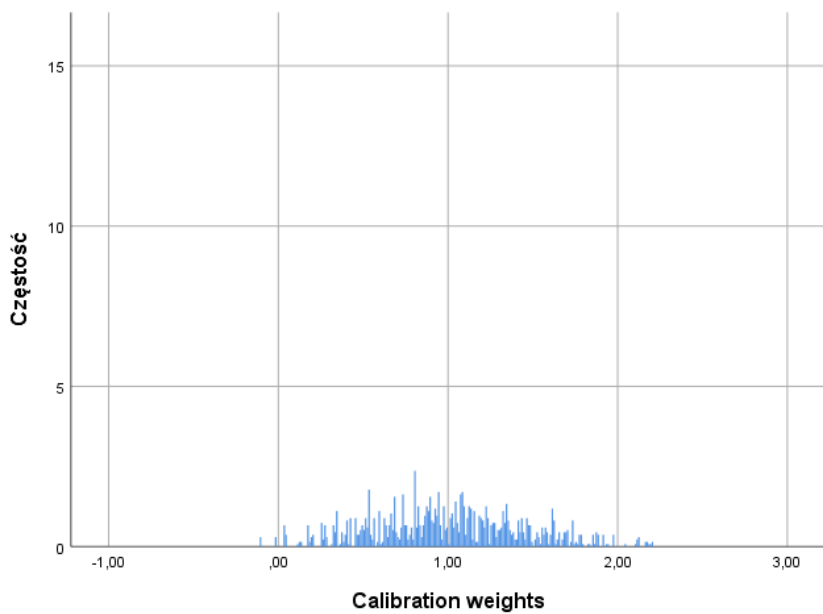
Porównanie rozkładu wartości wag (2/2)

Polska

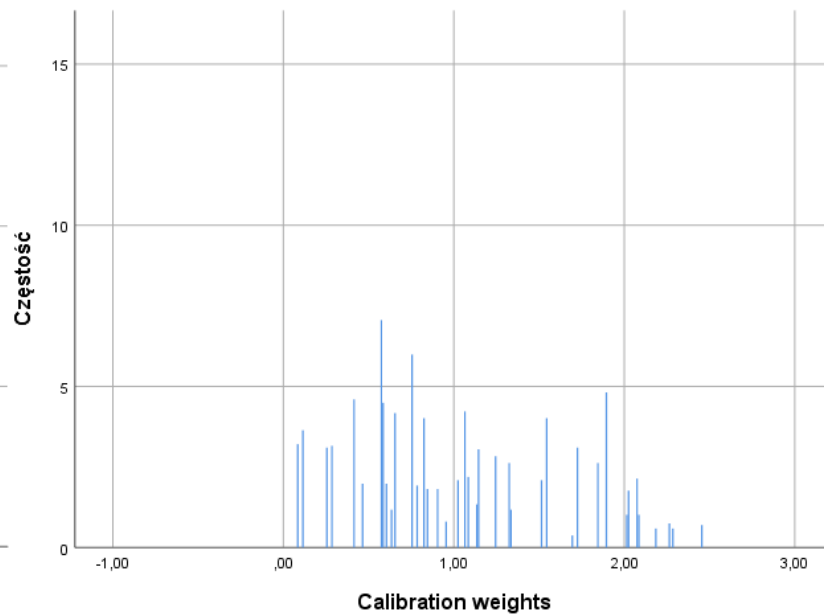


Porównanie rozkładu wartości wag (2/2)

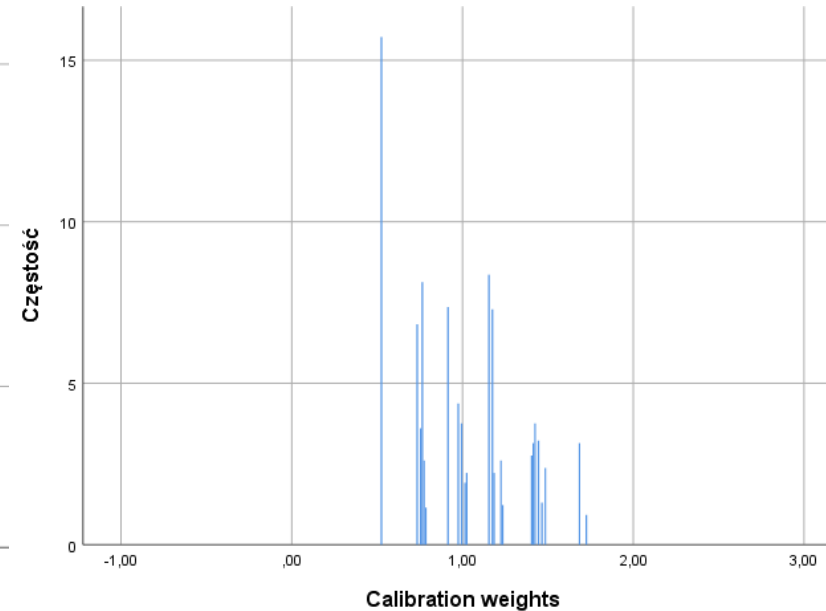
Polska



Francja



Estonia



Zastosowana kalibracja

- Bardzo dobra dokumentacja procedury ważenia (podany pełen algorytm)
- Równanie do rozkładów dostarczonych przez kraj biorący udział w badaniu:
 - wieku (7 kategorii: 18-24; 25-34; 35-44; 45-54; 55-64; 65-75; 75+)
 - płci (2 kategorie)
 - poziomu wykształcenia (3 kategorie: 0-2; 3-4; 5-8 według ISCED 2011)
- Taka sama procedura zastosowana we wszystkich krajach
 - Dokumentacja: wymaganie określenia źródła podanych rozkładów. **Dla 12 krajów: brak danych.**
- Imputacja braków danych w rozkładach:
 - Czechy – 66 wiek, 17 wykształcenie

Algorytm R: funkcja *calib {sampling}*

```
gweight <- calib(as.matrix(Xs),  
d = rep(p, nrow(Xs) ),  
mar_pop,  
method = "linear")
```

Te same wartości dla każdego respondenta w wektorze wejściowym. Według założeń przed badaniem powinny być to wagi probabilistyczne.

Model liniowy skutkujący wartościami ujemnymi zmiennej wynikowej.

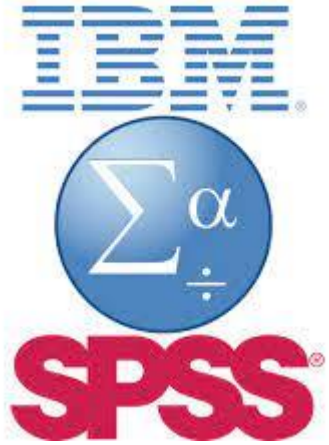
The *Linear* method has been employed, which results also in negative weights.

Such a method requires the specification of each individual inclusion probability. Due to the lack of this information, we employed a proxy of the inclusion probability computed in each country, assuming the sample to be drawn randomly.

$$\text{Inclusion probability} = \frac{\text{Sample size}}{\text{Target Population size}}$$

The inverse of the resulting value has been employed as a vector of initial weights.

„Reakcja” pakietów statystycznych na ujemne wagi (1/3)



- Liczy
- Nie bierze pod uwagę jednostek z ujemnymi lub zerowymi wagami
- „Ostrzega” pod wynikami analizy

Ostrzeżenie nr 3211

Przynajmniej w jednym przypadku wartość zmiennej ważącej była zerowa, ujemna lub nieobecna. Takie obserwacje są niewidoczne dla procedur statystycznych i wykresów, które wymagają obserwacji z dodatnią wagą, ale pozostają w pliku i są przetwarzane przez funkcje niestatystyczne, takie jak LIST i SAVE.

„Reakcja” pakietów statystycznych na ujemne wagi (2/3)



- Nie liczy
- Wypowiada się na temat sensowności komendy wprowadzonej przez użytkownika 😊

```
negative weights encountered;
```

```
negative weights not allowed;
```

```
You specified a variable that contains negative values as  
the weighting variable, so your request made no sense.
```

```
Perhaps you meant to specify another variable.
```

„Reakcja” pakietów statystycznych na ujemne wagi (3/3)



- (Czasem) liczy
- (Czasem) bierze pod uwagę jednostki z ujemnymi lub zerowymi wagami
- (Czasem) nie „ostrzega” pod wynikami analizy

```
aggregate(gweight ~ v1, data = df1, FUN = sum)
```

```
library(stats)  
xtabs(gweight ~ v1, data=df)
```

```
library(srvyr)  
df %>%  
  as_survey(weights = c(gweight)) %>%  
  group_by(v1) %>%  
  summarize(n = survey_total())
```

```
library(dplyr)  
count(x = df1, v1, wt = gweight)
```

```
lm(df$v2 ~ df2$v1, weights=df$gweight)
```

**Błąd w poleceniu 'lm.wfit(x, y, w, offset =
offset, singular.ok = singular.ok, ':
brakujące lub ujemne wagi nie są dozwolone**

Interwencja (lipiec 2019)

- Problem z interpretacją ujemnych wartości wag
- Nieuwzględnienie informacji o prawdopodobieństwach to brak spójności pomiędzy etapami procesu badawczego i marnowanie czasu zespołów krajowych
- Dlaczego informacje przesłane przez krajowy zespół nie zostały uwzględnione w algorytmie jako wektor *initial weights* używany w kalibracji?
- Dlaczego nie ma zmiennej reprezentującej wagę probabilistyczną (*design weight*)?
- Każdy kraj powinien być brany pod uwagę oddzielnie:
 - może mieć inny schemat doboru próby
 - można użyć innych („najlepszych”) zmiennych do przygotowania wag poststratyfikacyjnych.
- Problem z wiarygodnością wyników badania wśród odbiorców
- Są różne metody ważenia i jeżeli metoda liniowa daje niepożądane własności zmiennej wynikowej, to warto rozważyć inne metody lub przynajmniej jakiś sposób trymowania wag.

Odpowiedź – wrzesień 2019 (1/2)

Sorry for the somewhat belated answer, but we took the issue seriously and there was quite some consultation between the sampling experts and the methodology group and we came to a solution only last week. Unfortunately **not all countries were able to deliver inclusion probabilities**, so we had to find a solution that is **applicable across all the countries** to maintain the maximum of consistency. We will indeed compute design weights where possible and include them in the next version of the dataset; **limited resources** did not allow us to do so until now. We are finalizing at this moment the procedures to check the delivered information (we are using Poland as a test country so you will hear from us soon) and will get in touch with the countries in coming period.

Odpowiedź – wrzesień 2019 (2/2)

We are aware of the issue of the negative weights and, albeit the sampling experts assured us that **statistically speaking it is a legitimate outcome**, we are trying to see how we can improve. As a short-term fix, we will deliver in the coming week a patch including a syntax for substituting negative weights in the dataset. **Negative weights will be truncated** to a threshold of 0.1. For the coming final data release, we are considering different options for truncating/trimming/re-computing the weights with different calibration variables (e.g. by not including education or using NUTS3 regions).

Tydzień później:

- Variable *gweight_new* 'Truncated calibration weights' was generated to correct for negative weight values.
- **This is a temporary solution!** New weighting variables will be delivered in the final data release scheduled for the end of **April 2020**.

Full release (październik 2020)

The weights included in the full release of EVS data are two versions of calibration weights, a population size weight and – **for a limited number of countries** – a design weight. For the countries where it is provided the **design weight has not been factored in the computation of calibration weights.**

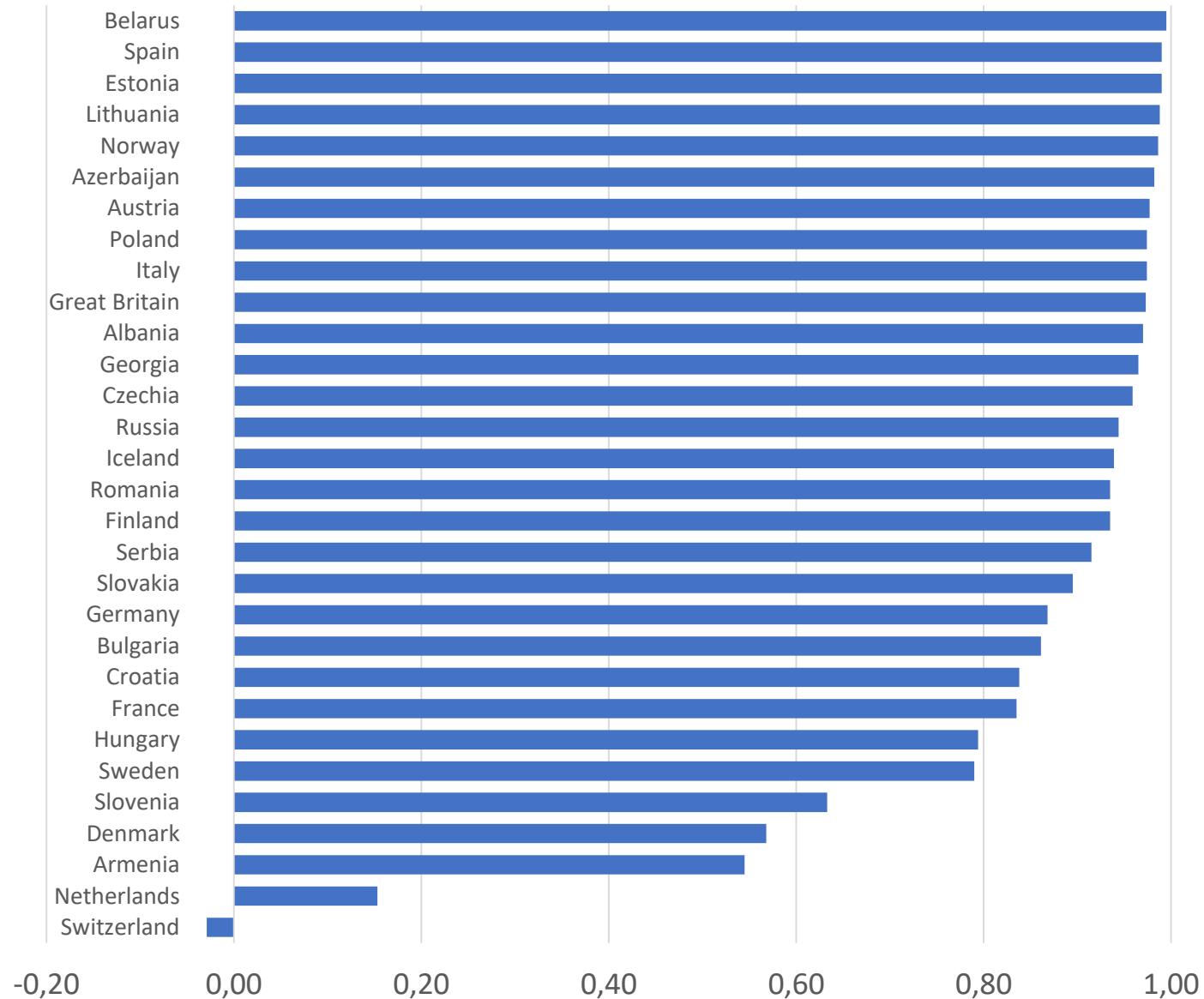
Wagi kalibracyjne

Two versions of calibration weights have been computed with different combinations of calibration variables:

- **'gweight'** has been computed using the marginal distribution of age, sex, educational attainment and region. This weight is provided as a standard version for consistency with previous releases.
- **'gweight_no_edu'** has been computed using the marginal distribution of age, sex and region. This weight is provided for researchers that reject the assumption of consistent measuring of educational level across countries (Ortmanns V. & Schneider S.L. 2016) and/or assume that non-responses have the same distribution across educational levels.

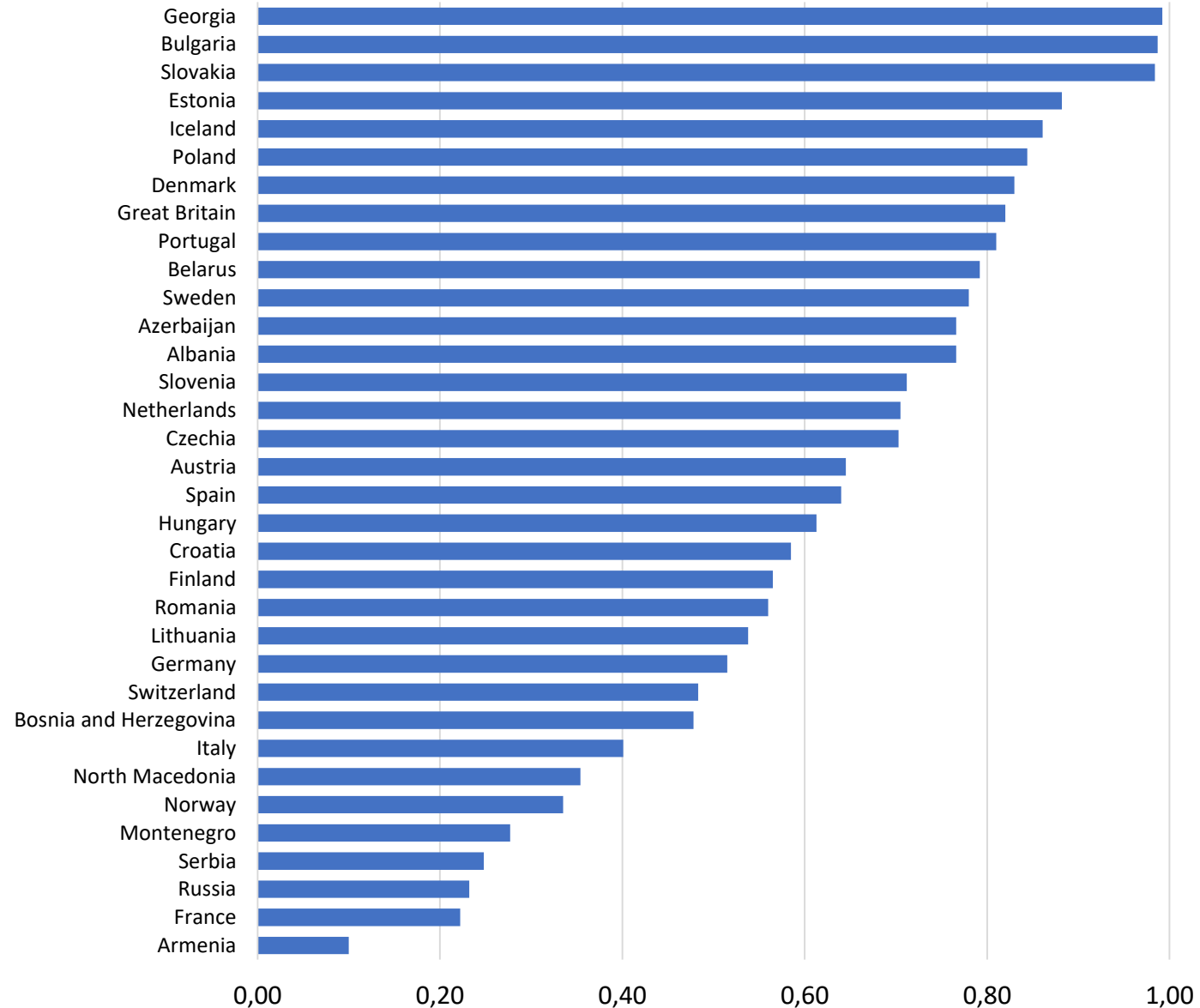
The two weight variables have been computed employing a raking algorithm specifically prepared for this procedure. A template of the algorithm is provided in Appendix A. Ultimately, all calibration weights have been trimmed at the 97.5th percentile for avoiding extreme values. Hence, the mean of calibration weights in each country is slightly different from 1.

Korelacje wag kalibracyjnych (2019.07 vs 2020.10)



Źródło: obliczenia własne na podstawie: EVS (2020): European Values Study 2017: Integrated Dataset (EVS2017). GESIS Data Archive, Cologne. ZA7500 Data file Version 2.0.0 and 4.0.0.

Korelacje wag kalibracyjnych 2020.10 (gweight vs. gweight_no_edu)



Źródło: obliczenia własne na podstawie: EVS (2020): European Values Study 2017: Integrated Dataset (EVS2017). GESIS Data Archive, Cologne. ZA7500 Data file Version 4.0.0.

Design weight

- For example, in a sample design with three stages, $PROB = \text{PROB_PSU} * \text{PROB_SSU} * \text{PROB_USU}$. These weights are then rescaled to the sample size in a way by which their mean is 1 and their sum equals the sample size. Consequently, the scaled weight takes the form of:

$$dweight = \frac{1}{PROB} / \sum \frac{1}{PROB} * n$$

- The accuracy of design weights is assessed by comparing the target population and the sum of the unscaled design weights divided by the response rate. Design weights will be provided for **Azerbaijan, Croatia, Poland, Russia and Germany**.

Sampling method - brief description (1/2)

AZERBAIJAN

- Sampling of voting precincts, households selected with **random walk** and individuals in each household sampled with last birthday method.

BOSNIA AND HERZEGOVINA

- Three stage sampling design with cumulative size method used for drawing settlements; households are selected with **random route** and the selection of individuals follows the last birthday method.

CROATIA

- Stratified sampling of municipalities with probability proportional to their population size; households are sampled with **random walk**, and the last birthday method is used for sampling individuals within each household.

GEORGIA

- Primary sampling units stratified by settlement type and quadrant; then, selected proportionally to their population size. Households selected with **random walk** method and Kish selection is used for individuals within the household.

Sampling method - brief description (2/2)

GEORGIA

- Primary sampling units stratified by settlement type and quadrant; then, selected proportionally to their population size. Households selected with **random walk** method and Kish selection is used for individuals within the household.

MONTENEGRO

- Simple random selection of polling stations within strata based on probability proportional to size, **random walk**, and individuals selected with the Kish method.

RUSSIA

- In the first domain settlements are selected by a stratified sample with probability proportional to the population of settlements. Then, electoral districts are randomly sampled and households are selected with **random walk** method. Individuals are drawn with a systematic sample with equal probability. In the second domain, electoral districts are sampled.

SERBIA

- Polling stations stratified by region selected with Probability Proportional to Size, households are chosen with the **STOP-AND-GO procedure**; individuals within households are selected with the Kish grid method.

Sampling method - brief description (2/2)

CZECHIA

- Multistage stratified random sample.

DENMARK

- Stratified simple random selection.

FINLAND

- Two staged stratified random sampling without clustering.

FRANCE

- Random sampling of households within municipalities. Individuals in each household are sampled with the Kish method.

NETHERLANDS

- Stratified simple random sample (STSI), with strata formed by municipalities.

Terminy realizacji

The recommended fieldwork period was set for **September 2017 to December 2017**.

- **CZ, RU, SK** and **SI** were able to complete their fieldwork within this time frame;
- BG, CH, HR, DK, DE, FI, IS, LT, NL, PL, ES, and CH carried the fieldwork over into 2018;
- AL, AM, AT, AZ, BY, EE, FR, GB, HU, GE, NO, RO and RS conducted the fieldwork in 2018;
- **IT** and **MK** carried the fieldwork over into 2019;
- **BA** and **ME** conducted the fieldwork in 2019;
- **PT** conducted the fieldwork in 2020.

Źródło: European Values Study (EVS). (2020). European Values Study (EVS) 2017: Method Report. (GESIS Papers, 2020/16). Köln. <https://doi.org/10.21241/ssoar.70109>

Harmonogram (z perspektywy polskiego zespołu krajowego)

- 2017 – przygotowania
- 2017.12-2018.01 – realizacja terenowa
- 2018.03.12 – zbiory danych przekazane do konsorcjum
- 2019.07 – upublicznienie zbioru z wagami (30 krajów) – pre-release
- 2019.09 – „Przycięcie” wag – oczekiwanie na ostateczne
- 2020.10 – upublicznienie ostatecznych zbiorów (34 kraje) – full-release

Diagnoza problemów i rekomendacje

- Warunki wzięcia udziału w badaniu ustalone na długo przed rozpoczęciem pracy w zespołach krajowych
 - 03.2017 – *Methodological guidelines* (10 stron)
 - 09.2017 – sugerowany termin rozpoczęcia realizacji.
- Zestaw nieprzekraczalnych warunków wzięcia udziału w badaniu
 - „16 countries out of 32 deviated from the guidelines and planned with an effective sample size below the set threshold (unknown for 2 out of 34 countries)”
- Informacja o zestawie kompetencji, które powinien mieć zespół krajowy
- Zwiększenie częstotliwości pomiaru skutkuje:
 - lepszym planowaniem budżetu na badanie (zarówno po stronie organizatora jak i w zespołach krajowych)
 - doświadczonym, dedykowanym zespołem do realizacji badania (zarówno po stronie organizatora jak i w zespołach krajowych)
- Większa kontrola:
 - na etapie projektowania realizacji
 - na etapie tworzenia dokumentacji badawczej (spójność dokumentacji pomiędzy krajami vs. *copy-paste*)
- Odpowiednio duży zespół projektowy po stronie organizatora
- Koszty: „nie ma badań szybkich, tanich i dobrej jakości”

Bibliografia

- European Values Study (EVS). (2020): European Values Study 2017: Integrated Dataset (EVS2017). GESIS Data Archive, Cologne. ZA7500 Data file Version 4.0.0, doi:10.4232/1.13560.
- European Values Study (EVS). (2020). European Values Study (EVS) 2017: Method Report. (GESIS Papers, 2020/16). Köln. <https://doi.org/10.21241/ssoar.70109>
- European Values Study (EVS). (2020). European Values Study (EVS) 2017: Methodological Guidelines. (GESIS Papers, 2020/13). Köln. <https://doi.org/10.21241/ssoar.70110>
- European Values Study (EVS). (2020). European Values Study (EVS) 2017: Weighting Data. (GESIS Papers, 2020/15). Köln. <https://doi.org/10.21241/ssoar.70113>
- Cassel, C.-M., Särndal, C.-E., and Wretman, J. (1976). Some results on generalized difference estimation and generalized regression estimation for finite population. *Biometrika*, 63:615–620.
- Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382.
- Deville, J.-C., Särndal, C.-E., and Sautory, O. (1993). Generalized raking procedure in survey sampling. *Journal of the American Statistical Association*, 88:1013–1020.
- Jabkowski, P., O zróżnicowaniu praktyk sondażowych w międzykrajowych badaniach porównawczych na podstawie metodologicznej archiwizacji 1537 surveyów zrealizowanych w latach 1981-2017 w krajach europejskich.
- Jabkowski, P., Kołczyńska, M. (2020). Sampling and Fieldwork Practices in Europe: Analysis of Methodological Documentation From 1,537 Surveys in Five Cross-National Projects, 1981-2017. *Methodology*. 16. 186-207.
- Ortmanns, V. and S.L. Schneider. 2016a. "Can We Assess Survey Representativeness of Cross-National Surveys Using the Education Variable?". *Survey Research Methods* 10(3): 189–210. DOI: <https://doi.org/10.18148/srm/2016.v10i3.6608>.
- Ortmanns, V. and S.L. Schneider. 2016b. "Harmonization Still Failing? Inconsistency of Education Variables in Cross-National Public Opinion Surveys." *International Journal of Public Opinion Research* 28(4): 562–582. DOI: <https://doi.org/10.1093/ijpor/edv025>.
- Ortmanns, V. 2020. Education distributions and survey characteristics of ten cross-national surveys. GESIS, SowiDataNet|datorium. Data file Version: 1.0.0. DOI: <https://doi.org/10.7802/1.2002>.
- Ortmanns, V. 2020. "Explaining Inconsistencies in the Education Distributions of Ten Cross-National Surveys – the Role of Methodological Survey Characteristics" *Journal of Official Statistics*, vol.36, no.2, pp.379-409. <https://doi.org/10.2478/jos-2020-0020>
- Żółtak T., Instytut Badań Edukacyjnych oraz Instytut Filozofii i Socjologii PAN, *Wagi poststratyfikacyjne w Europejskim Sondażu Społecznym — możliwości i ograniczenia prawidłowego wykorzystania w analizach statystycznych*

Metodologiczne Inspiracje 2021
Badania ilościowe w naukach społecznych - wyzwania i problemy



UNIwersytet
Warszawski

Dziękuję za uwagę

Krzysztof Bulkowski

e-mail: bulkowskik@is.uw.edu.pl

Katedra Modeli Formalnych i Metod Ilościowych w Socjologii

Wydział Socjologii UW