

Wagi poststratyfikacyjne w Europejskim Sondażu Społecznym: możliwości i ograniczenia prawidłowego wykorzystania w analizach statystycznych

Tomasz Żółtak

Instytut Badań Edukacyjnych,
IFiS PAN

Motywacje

- Publikacja oficjalnych wag poststratyfikacyjnych do badania ESS w kwietniu 2014 r.
 - Dosyć uboga dokumentacja.
 - Brak informacji o tym, że zastosowanie poststratyfikacji wpływa na wielkość błędów standardowych.
 - Nie opublikowano danych koniecznych, aby móc uwzględnić wpływ poststratyfikacji na wielkości błędów standardowych.
 - Dane te można uzyskać po zwróceniu się do autorów badania.
 - Podziękowania dla Any Slavec i Vasji Vehovara za udostępnienie danych.
- Publikacja przez IFiS informacji pozwalających uwzględnić złożony dobór polskiej próby ESS przy szacowaniu błędów standardowych we wrześniu 2014 r.

Plan wystąpienia

- Reprezentatywność proceduralna i przedmiotowa
- Różne techniki poststratyfikacji
- Na co wpływa poststratyfikacja?
- Wpływ poststratyfikacji na oszacowania błędów standardowych w badaniu ESS 6
- Wpływ uwzględnienia złożonego schematu doboru próby na oszacowania błędów standardowych w polskiej próbie badania ESS 6
- Podsumowanie

Reprezentatywność proceduralna i przedmiotowa

Teoria reprezentatywności proceduralnej:

- Reprezentatywność jako zachowanie pewnych formalne własności przez metodę **losowego** doboru próby.
- Odchylenia statystyk z próby od wartości populacyjnych są tu niejako naturalne – są konsekwencją randomizacji.
- Stanowi podstawę do zastosowania odpowiednich metod **wnioskowania statystycznego**.

Teoria reprezentatywności przedmiotowej:

- Reprezentatywność jako zgodność pewnych statystyk (rozkładów) z próby ze znanymi rozkładami populacyjnymi.
- **Intuicyjnie** zgodna z oczekiwaniami wielu badaczy i odbiorców badań.
- Nie operuje pojęciem błędu pomiaru – **brak podstaw do prowadzenia wnioskowania statystycznego**.

Reprezentatywność proceduralna i przedmiotowa

Co do zasady chcemy to połączyć oczekując, że:

- Zachowana będzie zgodność pewnych statystyk (rozkładów) z próby ze znanymi rozkładami populacyjnymi.
- Będziemy mieć podstawy do prowadzenia wnioskowania statystycznego.

Czy to się daje zrobić?

- Na gruncie *teorii* reprezentatywności proceduralnej możemy znaleźć dwa problemy zbieżne z teorią reprezentatywności przedmiotowej:
 - Wykorzystanie znanych informacji o populacji do zmniejszenia wariancji estymatorów.
 - Korekta ew. obciążenia estymatorów w sytuacji występowania braków danych (niepełnej realizacji próby).

Reprezentatywność proceduralna i przedmiotowa

Rozbieżności, o których zwykle nie pamiętamy

Teoria reprezentatywności przedmiotowej

- Zakładamy, że uzgodnienie pewnych statystyk z próby ze znanymi wartościami populacyjnymi automatycznie podnosi jakość całej prób (również innych zmiennych).
- Skupiona na konkretnej próbie (pojedynczej realizacji).

Teoria reprezentatywności proceduralnej

- Wpływ korygowania statystyk z próby jest bardzo *lokalny* – może być bardzo różny dla różnych zmiennych i nie zawsze jest jednoznacznie pozytywny.
- Skupiona na własnościach estymatorów w ramach zbioru *wszystkich możliwych prób*.

Reprezentatywność proceduralna i przedmiotowa

Jak rozumiemy słowo **obciążenie** (*bias*)?

Teoria reprezentatywności przedmiotowej

- **Obciążenie jako rozbieżność pomiędzy statystyką z próby i jej wartością populacyjną.**
- Obciążenie jako różnica pomiędzy respondentami chętnymi a niechętnymi do współpracy.

Teoria reprezentatywności proceduralnej

- Obciążenie jako różnica pomiędzy wartością oczekiwaną estymatora (tj. statystyką określoną w zbiorze *wszystkich możliwych prób*) a wartością statystyki w populacji.

(za: Jaak Billiet, Hideko Matsuo, Koen Beullens, Vasja Vehovar. (2009). „Non-Response Bias in Cross-National Surveys: Designs for Detection and Adjustment in the ESS”. *ASK Research & Methods* Vol. 18(1): 3–43)

Reprezentatywność proceduralna i przedmiotowa

Jak rozumiemy słowo **obciążenie** (*bias*)?

Teoria reprezentatywności przedmiotowej

- **Obciążenie jako rozbieżność pomiędzy statystyką z próby i jej wartością populacyjną.**

Czy wartość statystyki w pojedynczej próbie jest dobrą podstawą do przewidywania wartości oczekiwanej estymatora?

Teoria reprezentatywności proceduralnej

- Obciążenie jako różnica pomiędzy wartością oczekiwaną estymatora (tj. statystyką określoną w zbiorze *wszystkich możliwych prób*) a wartością statystyki w populacji.

Reprezentatywność proceduralna i przedmiotowa

Jak rozumiemy słowo **obciążenie** (*bias*)?

Teoria reprezentatywności przedmiotowej

- **Obciążenie jako rozbieżność pomiędzy statystyką z próby i jej wartością populacyjną.**

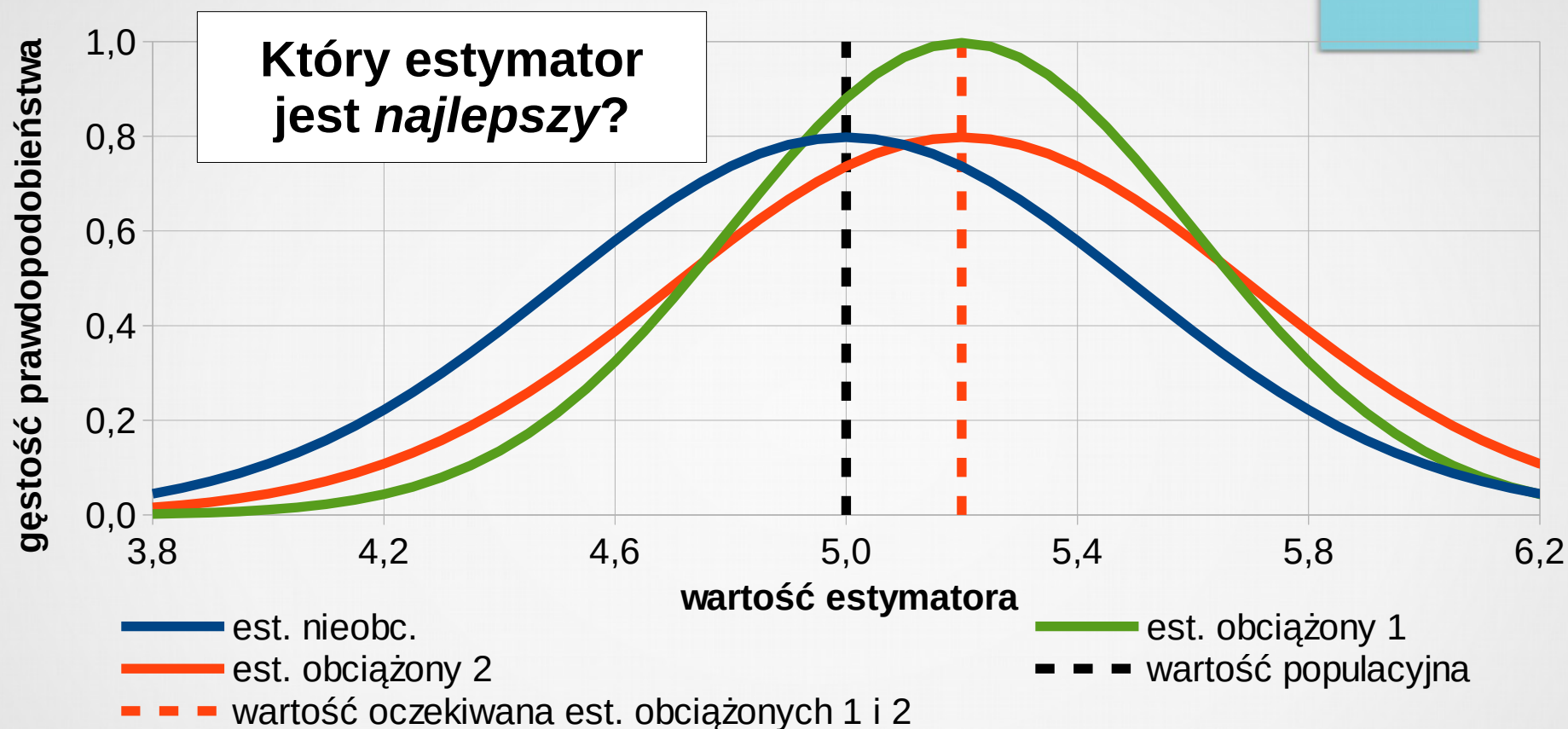
Czy wartość statystyki w pojedynczej próbie jest dobrą podstawą do przewidywania wartości oczekiwanej estymatora?

**Co jest obciążone/*obciążone*:
estymator
czy
próba?**

Teoria reprezentatywności proceduralnej

- Obciążenie jako różnica pomiędzy wartością oczekiwaną estymatora (tj. statystyką określoną w zbiorze *wszystkich możliwych prób*) a wartością statystyki w populacji.

Nieobciążoność a wariancja estymatora

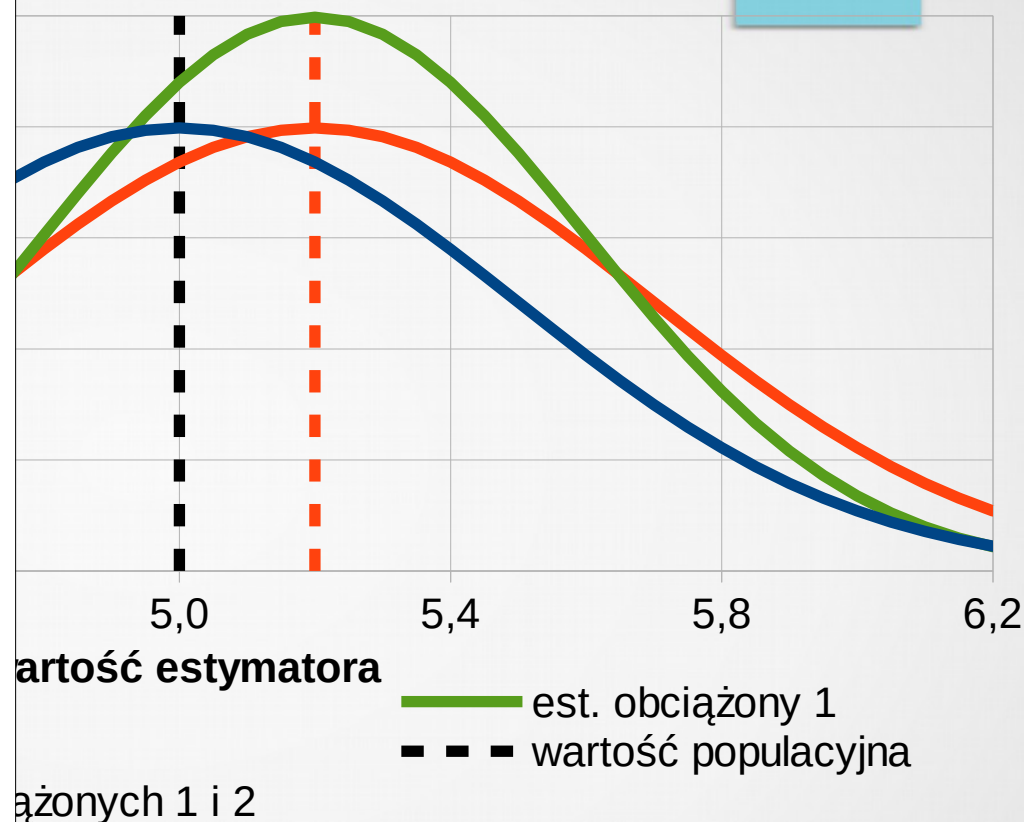


estymator	obciążenie	wariancja	błąd średniokwadratowy
nieobciążony	0	0,25	0,25
obciążony 1	0,2	0,16	0,20
obciążony 2	0,2	0,25	0,29

Nieobciążoność a wariancja estymatora

Spośród dwóch estymatorów obciążonego i nieobciążonego ten drugi jest z pewnością *lepszy* tylko wtedy, gdy ma nie większą wariancję, niż estymator obciążony, z którym jest porównywany.

(oczywiście zakładając, że dla obu estymatorów działają twierdzenia graniczne)



estymator	obciążenie	wariancja	błąd średniokwadratowy
nieobciążony	0	0,25	0,25
obciążony 1	0,2	0,16	0,20
obciążony 2	0,2	0,25	0,29

Poststratyfikacja i pokrewne techniki

- **Poststratyfikacja**

Przypisz obserwacjom wagi w ten sposób, aby ważony rozkład częstości zmiennej w próbie był taki sam, jak znany rozkład populacyjny (wszystkie wartości, jakie przyjmuje zmienna w populacji powinny pojawić się również w próbie).

- **Wagi wieńcowe (raking)**

- **Kalibracja (calibration)**

Poststratyfikacja i pokrewne techniki

- **Poststratyfikacja**
- **Wagi wieńcowe (raking)**

Procedura poststratyfikacji kolejno ze względu na rozkład brzegowy każdej z kilku zmiennych – „zapętłana” i kontynuowana do osiągnięcia zbieżności (zmian w wagach nie większych niż założony próg).

Pozwala uniknąć problemów z zerowymi liczebnościami w rozkładach łącznych kilku zmiennych w próbie.

- **Kalibracja (calibration)**

Poststratyfikacja i pokrewne techniki

- **Poststratyfikacja**
- **Wagi wieńcowe (raking)**
- **Kalibracja (calibration)**

Przypisz obserwacjom wagi w ten sposób, aby wartość statystyki z próby (ważona) była równa znanej wartości parametru danej zmiennej w populacji, jednocześnie minimalizując funkcję opisującą różnice wag przed i po ważeniu.

Może być postrzegana jako uogólnienie poststratyfikacji.

Różne warianty – w zależności od wyboru funkcji opisującej różnice wag.

Cele poststratyfikacji (i technik pokrewnych)

- **Redukcja obciążenia estymatorów związanego z odmowami odpowiedzi.**

Osiągana, jeśli zmienna wykorzystywana do poststratyfikacji jest w związku (im silniejszym, tym lepiej):

- z prawdopodobieństwem odmowy odpowiedzi,
- lub ze zmienną, której parametry estymujemy (zwiększenie precyzji – choć stricte formalnie nie jest to redukcja obciążenia).

- **Zwiększenie precyzji szacowania (zmniejszenie wariancji estymatorów).**

Osiągana, jeśli zmienna wykorzystywana do poststratyfikacji jest w związku (im silniejszym, tym lepiej):

- ze zmienną, której parametry estymujemy.

Redukcja obciążenia estymatorów związanego z odmowami odpowiedzi

Łatwo jest:

- Znaleźć zmienne, które pozwalają uzyskać redukcję obciążenia (choć niekoniecznie dużą).
- Uwzględnić ważenie w analizach (wystarczy estymując parametry przypisać jednostkom wyliczone wagi).

Problemy:

- Redukcja obciążenia nie musi iść w parze ze zwiększeniem precyzji szacowania (a czasem może wręcz przyczyniać się do zmniejszenia precyzji).
- W praktyce trudno ocenić, jaka jest skala redukcji obciążenia.

Zwiększenie precyzji szacowania

Pozytywy:

- Można dosyć dokładnie ocenić, jaki jest wpływ ważenia na precyzję szacowania.

Problemy:

- Wpływ ważenia bardzo zróżnicowany w zależności od zmiennej, której parametry są szacowane.
- Ważenie ze względu na zmienne słabo powiązane ze zmienną, której parametry są szacowane może wręcz zmniejszać precyzję szacowania.
- Aby uwzględnić wpływ ważenia na precyzję trzeba zastosować specjalne techniki szacowania błędów standardowych.

Zwiększenie precyzji szacowania

Uwzględnienie wpływu ważenia na precyzję szacowania wymaga:

- Informacji o wartościach wag przed (jeśli zastosowano złożony schemat doboru próby) i po poststratyfikacji.
- Zastosowania odpowiednich technik statystycznych (oprogramowania).

Wpływ poststratyfikacji na oszacowania błędów standardowych w badaniu ESS 6

- Dane z ESS 6 (2012).
- Uwzględniono dwa kraje (Dania, Szwecja), w których próba losowa dobrana została w sposób prosty.
 - Brak problemów z uwzględnieniem złożonego schematu doboru próby.
- Poziom realizacji:
 - **Dania:** 49,06%.
 - **Szwecja:** 52,44%.
- Oficjalna poststratyfikacja (*raking*) ESS ze względu na:
 - **Dania:** łączny rozkład płci, kategorii wieku i edukacji oraz podział na 5 regionów.
 - **Szwecja:** łączny rozkład płci, kategorii wieku i edukacji oraz podział na 8 regionów.

Wpływ poststratyfikacji na oszacowania błędów standardowych w badaniu ESS 6

Analizowany był wpływ na precyzję oszacowań następujących sposobów analizy:

- bez poststratyfikacji,
- *naiwna* poststratyfikacja, uwzględniająca wagi przy wyliczaniu wartości estymatorów punktowych, ale nieuwzględniająca wpływu zróżnicowania wag na wariancję estymatorów,
- *pełna* poststratyfikacja, uwzględniająca również wpływ na wariancję estymatorów (przy użyciu pakietu *survey*, w programie R).

Wpływ poststratyfikacji na oszacowania błędów standardowych w badaniu ESS 6

zmienna	poststratyfikacja	Dania			Szwecja		
		średnia	bł. std.	(1)	średnia	bł. std.	(1)
wiek	bez	48,70	0,47		47,83	0,44	
	<i>naiwna</i>	50,07	0,47	1,00	46,94	0,44	1,00
	<i>pełna</i>		0,18	0,39		0,15	0,34
lata nauki	bez	13,01	0,128		12,85	0,079	
	<i>naiwna</i>	12,61	0,126	0,98	13,01	0,082	1,05
	<i>pełna</i>		0,114	0,89		0,048	0,62
zaufanie do parlamentu (0-10)	bez	6,10	0,055		5,93	0,054	
	<i>naiwna</i>	6,01	0,055	1,01	5,96	0,054	1,00
	<i>pełna</i>		0,057	1,03		0,053	0,99
samookreślenie na skali lewica-prawica (0-10)	bez	5,34	0,060		5,31	0,053	
	<i>naiwna</i>	5,40	0,060	1,00	5,33	0,053	1,00
	<i>pełna</i>		0,063	1,06		0,055	1,04
głosowanie na socjaldemokrację (0 – nie, 1 -tak)	bez	0,22	0,010		0,25	0,010	
	<i>naiwna</i>	0,23	0,010	1,02	0,23	0,010	0,98
	<i>pełna</i>		0,011	1,10		0,010	0,97

(1) = bł. std.(przy poststratyfikacji) / bł. std. (bez poststratyfikacji)

Wpływ uwzględnienia złożonego schematu doboru próby w polskiej próbie badania ESS 6

Dobór próby ESS w Polsce:

- Warstwowo – łącznie 150 warst:
 - Alokacja do warstw proporcjonalna do liczby ludności i odwrotnie proporcjonalna do przewidywanego poziomu realizacji.
 - W 86 warstwach – miastach powyżej 50 tys. mieszkańców dobór prosty, bez zwracania.
 - W 64 warstwach wyznaczanych przez przecięcie województwa z rodzajem gminy (miejska, wiejska, część wiejska gminy m-w, część wiejska gminy m-w) dobór dwustopniowy z losowaniem:
 - miejscowości bezzwrtonie z prawdopodobieństwem proporcjonalnym do liczby ludności
 - w ramach miejscowości proste losowanie bez zwracania wiązek czteroosobowych.
- Oszacowany wskaźnik efektywności schematu $deffp = 1,02$.
- Poziom realizacji: 74,87%.

Wpływ uwzględnienia złożonego schematu doboru próby w polskiej próbie badania ESS 6

zmienna	metoda estymacji	Polska			
		średnia	bł. std.	(1)	(2)
wiek	jak dla doboru prostego	46,10	0,44		
	z uwzgl. schematu		0,42	0,96	
	ze schematem i poststrat.	47,29	0,18	0,40	0,42
lata nauki	jak dla doboru prostego	12,23	0,081		
	z uwzgl. schematu		0,078	0,96	
	ze schematem i poststrat.	12,16	0,075	0,92	0,96
zaufanie do parlamentu (0-10)	jak dla doboru prostego	2,95	0,055		
	z uwzgl. schematu		0,056	1,03	
	ze schematem i poststrat.	2,95	0,057	1,04	1,01
samookreślenie na skali lewica-prawica (0-10)	jak dla doboru prostego	5,67	0,060		
	z uwzgl. schematu		0,060	1,01	
	ze schematem i poststrat.	5,67	0,062	1,03	1,03
głosowanie na socjaldemokrację	jak dla doboru prostego	0,04	0,005		
	z uwzgl. schematu		0,004	0,95	
	ze schematem i poststrat.	0,04	0,004	0,97	1,02

(1) = bł. std.(~) / bł. std. (jak dla doboru prostego)

(2) = bł. std.(z uwzgl. schematu) / bł. std. (z uwzgl. schematu i poststrat.)

Podsumowanie

- Poststratyfikacja nie zawsze pozwala poprawić jakość przewidywania wartości populacyjnych.
- Czasem należy wręcz podejrzewać, że pogarsza (wzrost wariancji estymatorów).
- Ogólnie rzecz biorąc wpływ poststratyfikacji na wyniki analiz danych ESS jest niewielki – zarówno w kategoriach zmiany oszacowań punktowych jak i ich błędów standardowych - jeśli pominąć zmienne bezpośrednio związane ze zmiennymi wykorzystanymi do poststratyfikacji.
 - Czy warto uwzględniać poststratyfikację w analizach ESS?
- Potwierdza się wysoka efektywność polskiego schematu doboru próby do badania ESS.



Dziękuję za uwagę!

t.zoltak@ibe.edu.pl