

Metody wyliczania błędów standardowych przy złożonych schematach doboru próby, na przykładzie badania „Nasza dalsza nauka i praca”

Tomasz Żółtak
IFiS PAN, IS UW



INSTYTUT FILOZOFII I SOCJOLOGII
POLSKIEJ AKADEMII NAUK



Badania na (reprezentatywnych) próbach losowych

- Klasyczne metody statystyczne zakładają, że próba dobierana była w sposób prosty, niezależny (ze zwracaniem).
- W praktyce badań sondażowych próby bardzo często dobierane są w ramach schematów złożonych:
 - losowanie bez zwracania;
 - warstwowanie (w tym poststratyfikacja);
 - dobór wielostopniowy;
 - dobór zespołowy.

Analiza danych dobranych w oparciu o złożone schematy doboru próby

- Stosowanie metod klasycznych do danych dobranych w oparciu o złożone schematy prowadzi do błędnych oszacowań.
- Poprawna estymacja wymaga:
 - Dla **oszacowań punktowych** – zastosowania wag o wartościach odwrotnie proporcjonalnych do prawdopodobieństwa trafienia do próby lub wag poststratyfikacyjnych.
 - Dla **błędów standardowych** – wykorzystania specjalistycznych technik statystycznych, omawianych w dalszej części prezentacji.

Analiza danych dobranych w oparciu o złożone schematy doboru próby

- Poprawna estymacja wymaga:
 - Dla **oszacowań punktowych** – zastosowania wag o wartościach odwrotnie proporcjonalnych do prawdopodobieństwa trafienia do próby lub wag poststratyfikacyjnych.
 - ✓ Wagi tego rodzaju są co do zasady dostępne w ramach ogólnodostępnych zbiorów danych z dużych badań sondażowych (ESS, PGSS, PGSW, DS).
 - Dla **błędów standardowych** – wykorzystania specjalistycznych technik statystycznych, omawianych w dalszej części prezentacji.
 - ✗ W zbiorach danych z badań o tematyce socjologicznej i politologicznej co do zasady brak informacji, które pozwoliłyby wykorzystać te metody.

Metody szacowania błędów standardowych dla danych dobranych w sposób złożony

- Metody „analityczne” (linearyzacja Taylora):
 - Wybierz estymator sumy wartości zmiennej w populacji (*population total*) o szerokim zakresie stosowalności i dobrych własnościach, a następnie analitycznie (linearyzacja Taylora) wyprowadź z wzoru na oszacowanie jego wariancji wzór na oszacowanie wariancji parametru będącego obiektem zainteresowania.
- Metody „replikacyjne”:
 - Potraktuj swoją wylosowaną próbę jako populację i symuluj losowanie z niej prób w celu oszacowania rozkładu danego parametru lub wariancji takiego rozkładu (pierwiastek z tej wariancji to oszacowanie błędu standardowego tego parametru).

Linearyzacja Taylora

- Dwa najczęściej wykorzystywane estymatory sumy:
 - estymator Horvitz-Thompsona (dowolne schematy doboru);
 - estymator Hansena-Hurvitza (schematy z losowaniem niezależnym);
- ✓ Zalety podejścia:
 - bardzo mocne podstawy formalne, w pełnym (HT) lub bardzo szerokim (HH) zakresie możliwych schematów doboru prób losowych;
 - ogólnie dobre własności oszacowań błędów standardowych (wariancji) estymatorów uzyskiwanych tą metodą;
- ✗ Wady podejścia:
 - skomplikowana procedura wywodzenia estymatorów wariancji dla nowych (rzadko używanych) parametrów;
 - skomplikowana implementacja w programach statystycznych;
 - wymaga podania bardzo wyczerpującego opisu sposobu doboru próby;
 - trudne w stosowaniu dla „przeciętnego użytkownika” wyników badań sondażowych – konieczność zrozumienia sposobu doboru próby do badania i opisanie go w odpowiedni sposób wykorzystywanemu oprogramowaniu;

Estymatory wariancji estymatorów sum

- estymator (Yatesa, Grundy'ego i Sena) wariancji estymatora Horvitz-Thompsona:

$$\hat{D}_{YGS}^2(\hat{T}_{HT}) = \sum_{i=1}^n \sum_{j>i}^n \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{x_i}{\pi_i} - \frac{x_j}{\pi_j} \right)^2$$

gdzie:

π_i - prawdopodobieństwo znalezienia się i -tej obserwacji w próbie,
 π_{ij} - prawdopodobieństwo jednoczesnego znalezienia się w próbie i -tej i j -tej obserwacji

- estymator wariancji estymatora Hansena-Hurvitza:

$$\hat{D}^2(\hat{T}_{HH}) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\frac{x_i}{p_i} - \frac{1}{n} \sum_{j=1}^n \frac{x_j}{p_j} \right)^2$$

gdzie: p_i – prawdopodobieństwo wybrania do próby w pojedynczym losowaniu zwrotnym

Metody „replikacyjne”

Ogólna idea: Potraktuj swoją wylosowaną próbę jako populację i symuluj losowanie z niej prób (replikacji) w celu oszacowania rozkładu danego parametru. Stosowany schemat losowania powinien odpowiadać (naśladować) schematowi, w wyniku którego dobrana została analizowana próba.

- **Jackknife:**
 - rozpatrz wszystkie możliwe replikacje powstające w wyniku usunięcia z próby (warstwy) pojedynczej obserwacji.
- **Bootstrapping:**
 - wielokrotnie losuj z próby (warstwy) ze zwracaniem replikacje równoliczne z próbą.
- **Balanced Repeated Replication (BRR):**
 - dobierz jednostki obserwacji (zespoły) w pary (pseudowarstwy) i rozpatrz wszystkie $2^{n/2}$ replikacje powstające w wyniku niezależnego wyboru jednej obserwacji w ramach każdej pary (pseudowarstwy);
 - w praktyce daje się pokazać, że identyczne oszacowanie wariancji można uzyskać rozpatrując nie więcej niż $n/2+4$ odpowiednio dobranych replikacji (schemat doboru wyznacza się przy pomocy tzw. macierzy Hadamarda);

Balanced Repeated Replication (BRR)

- Metoda popularna w badaniach edukacyjnych i badaniach związanych ze zdrowiem (w Ameryce Północnej).
 - W przypadku doboru wielostopniowego zwykle stosuje się wersję uproszczoną, w której na potrzeby tworzenia replikacji jednostki losowania z pierwszego poziomu traktowane są jakby były zespołami (pozwala to ograniczyć liczbę rozpatrywanych replikacji, choć kosztem precyzji szacowania błędów standardowych).
 - Zestaw replikacji do rozpatrzenia zadawany jest w postaci zmiennych z wagami – po jednej zmiennej na replikację – w których jednostkom nie wybranym do replikacji przypisywana jest waga 0, a jednostkom wybranym do replikacji waga 2.
 - Zestaw takich wag przygotowywany jest przez instytucję prowadzącą badanie i dołączany jest do zbioru danych z wynikami badania.
 - Wagi te mogą (i powinny) być w analizie łączone z wagami odwrotnie proporcjonalnymi do prawdopodobieństwa wylosowania obserwacji do próby, lub wagami poststratyfikacyjnymi.

Balanced Repeated Replication (BRR)

- Modyfikacja Fay'a:
 - W każdej replikacji zamiast przypisywać połowie obserwacji (zespołów) wagi 2 razy większe, a połowie 0, niech przypisane zostaną wagi o wartościach $1 \pm \varepsilon$ razy dotychczasowa waga ($0 < \varepsilon < 1$).
 - Estymatory Fay'a zachowują się lepiej w sytuacji, gdy wagi poststratyfikacyjne wykorzystywane w analizie mają dużą rozpiętość oraz przy analizach w podzbiorowościach.
 - Często wybieraną wartością jest $\varepsilon = 0,5$.
 - Rao i Shao (1999) pokazali, że dla $\varepsilon \rightarrow 0$ oszacowania wariancji estymatorów metodą Fay'a zbiegają do wyników uzyskiwanych w wyniku zastosowania linearyzacji Taylora.

Balanced Repeated Replication (BRR)

- Oszacowanie błędu standardowego parametru wyliczane jest na podstawie replikacji:

$$\hat{D}_{\varepsilon}(\hat{\theta}) = \sqrt{\frac{1}{G \varepsilon^2} \sum_{i=1}^G (\hat{\theta}_i - \hat{\theta})^2}$$

gdzie:

G – liczba replikacji,

θ_i – oszacowanie parametru w ramach i -tej replikacji,

θ – oszacowanie parametru w ramach całej próby

ε – parametr Fay'a.

- Jednocześnie w testach statystycznych liczba stopni swobody określana jest w odniesieniu do G , a nie w odniesieniu do liczby obserwacji.

Balanced Repeated Replication (BRR)

✓ Zalety podejścia:

- bardzo duża elastyczność w stosowaniu, łatwość użycia w odniesieniu do niemal zupełnie dowolnych parametrów;
- łatwość implementacji w oprogramowaniu (np. w badaniu PISA przygotowano makra pozwalające dokonać obliczeń w programach nie posiadających wsparcia dla złożonych schematów doboru próby);
- możliwość zdjęcia z użytkownika danych niemal w całości konieczności pamiętania, według jakiego schematu dobierana była próba;
- w typowych zastosowaniach daje wyniki zbliżone do linearyzacji Taylora;

✗ Wady podejścia:

- nie we wszystkich zastosowaniach posiada dobre uzasadnienie formalne;
- w dużych sondażach musi być stosowana w formie uproszczonej;
- konieczność przygotowania zestawu wag przez instytucję prowadzącą badanie (dodatkowa praca);
- przy typowo stosowanych schematach doboru przygotowanie zestawu wag wymaga dokonania arbitralnych decyzji o przypisaniu obserwacji (zespołów) do pseudowarstw – będą one mieć pewien wpływ na uzyskiwane oszacowania;

Wykorzystywane dane

- Dane potrzebne do prowadzenia wnioskowania statystycznego metodami uwzględniającymi złożony schemat doboru próby dostępne są w przypadku części badań edukacyjnych.
 - W wystąpieniu wykorzystane zostaną z badania „Nasza dalsza nauka i praca” realizowanego przez IFiS PAN na zlecenie CKE.
 - Jest to badanie panelowe, oparte na próbie uczniów wylosowanej do tzw. „opcji narodowej” badania PISA 2009. Badanie objęło łącznie 4387 uczniów, którzy w 2009 roku uczęszczali do klas pierwszych szkół ponadgimnazjalnych.
 - Schemat doboru próby:
 - technicznie – warstwowy (rodzaj szkoły), w ramach warstw zespołowy, dwustopniowy, z losowaniem szkół z prawdopodobieństwem proporcjonalnym do liczby oddziałów klas pierwszych na 1. stopniu i prostym losowaniem jednego oddziału w ramach szkoły na 2. stopniu;
 - formalnie (w praktyce) – warstwowy, zespołowy z równym prawdopodobieństwem wyboru każdego zespołu w ramach warstwy;
 - po zebraniu danych przeprowadzono poststratyfikację z korektą na różnice w poziomie realizacji w ramach szkół.

Przykładowe wyniki z badania „Nasza dalsza nauka i praca”

	Skala samooceny (self-esteem) Rosenberga		Inwentarz cechy lęku		Test czytania PISA	
średnia (ważona)	29,36		40,43		513,26	
odch. stand. (ważone)	4,53		8,49		94,58	
błędy standardowe	wartość	bs/bs.pr	wartość	bs/bs.pr	wartość	bs/bs.pr
dobór prosty	0,07		0,13		1,48	
dobór prosty z wagami	0,07	1,00	0,14	1,02	1,51	1,02
poststratyfikacja	0,08	1,04	0,15	1,09	1,65	1,11
linearyzacja Taylora	0,09	1,26	0,18	1,34	3,80	2,57
BRR – Fay 0,5	0,08	1,17	0,17	1,26	3,76	2,54

wszystkie pomiary zostały dokonane w 2010 r, w klasach II szkół ponadgimnazjalnych

Przykładowe wyniki dla I rundy ESS (2002)

za: Gabler, Häder, Lynn. 2006. Design Effects for Multiple Design Samples.
Statistics Canada 32, s. 115-120.

	Polska	Wielka Brytania	Niemcy
	bł.stand./bł. stand. przy doborze prostym		
liczba lat nauki	1,33	1,68	1,80
czas poświęcany na oglądanie TV	1,14	1,43	1,44
zaufanie do policji	1,11	1,11	1,39
aktywność polityczna	1,47	1,39	1,81

- Również w socjologicznych badaniach sondażowych efekt złożonego schematu doboru próby na wielkość błędów standardowych może być bardzo duży.

Podsumowanie

- Stosowanie złożonych schematów doboru próby może mieć bardzo znaczący wpływ na wyniki analiz danych sondażowych.
- Obecnie w badaniach socjologicznych i politologicznych właściwie zaniedbuje się wpływ stosowania złożonych schematów doboru próby na zwiększanie się wariancji estymatorów, a więc i błędów standardowych uzyskiwanych oszacowań.
 - Dla niektórych zmiennych jest to nawet niemal dwukrotny wzrost wartości błędów standardowych.
- Prowadzi to do częstszego popełniania błędów I rodzaju w procedurze weryfikacji hipotez statystycznych.

Podsumowanie

- Istnieją metody pozwalające szacować błędy standardowe z uwzględnieniem wpływu złożonych schematów doboru próby.
- W obecnych warunkach nie ma jednak możliwości ich zastosowania w odniesieniu do ogólnie dostępnych zbiorów danych z badań sondażowych, gdyż nie zawierają one koniecznych w tym celu informacji (choćby o przydziale obserwacji do warstw).
- Należy mieć nadzieję, że w przyszłości temat ten spotka się ze znacznie większym zainteresowaniem ze strony badaczy społecznych.

Dziękuję za uwagę!

Tomasz Żółtak
tzoltak@ifispan.waw.pl



INSTYTUT FILOZOFII I SOCJOLOGII
POLSKIEJ AKADEMII NAUK

