

STUDIA METODOLOGICZNE

Marcin SZYMKOWIAK

Estymatory kalibracyjne stosowane w Narodowym Spisie Powszechnym w 2011 r.¹

Narodowy Spis Powszechny Ludności i Mieszkań 2011 (NSP 2011) po raz pierwszy w historii polskiej statystyki publicznej został przeprowadzony metodą mieszaną. Oznaczało to wykorzystanie danych pochodzących ze źródeł administracyjnych (rejestrów i sytemów administracyjnych), a także danych zebranych w badaniu reprezentacyjnym, którego próba liczyła 20% mieszkań w Polsce².

Wykorzystanie danych z różnych źródeł wymagało, na etapie uogólniania wyników badania reprezentacyjnego, odpowiedniego skorygowania wag dla wszystkich osób biorących w nim udział, zarówno w odniesieniu do ludności faktycznej jak i rezydującej. Było to spowodowane koniecznością odpowiedniego zintegrowania i zachowania spójności pomiędzy wynikami badania reprezentacyjnego z danymi ze spisu pełnego w zakresie podstawowych zmiennych demograficznych odnoszących się do płci, wieku oraz miejsca zamieszkania — poziom powiatu z wyodrębnieniem części miejskiej i wiejskiej.

¹ Oryginalną wersję referatu, na podstawie którego powstał artykuł, wygłoszono w Łodzi na konferencji naukowej *Statystyka—wiedza—rozwój* (październik 2013 r.).

² Szczegółowy opis zasad tworzenia operatu losowania mieszkań, zmiennych uwzględnionych w operacie losowania, schematu losowania próby oraz sposobu tworzenia wag można znaleźć w opracowaniu: *Raport...* (2012).

Artykuł zawiera najważniejsze wnioski z przeprowadzonego badania, polegającego na wyznaczeniu wag kalibracyjnych dla ludności rezydującej w NSP 2011³.

TEORETYCZNE PODSTAWY KALIBRACJI

Na potrzeby wyznaczenia wag kalibracyjnych w NSP 2011, celem dostosowania odpowiednich struktur z badania reprezentacyjnego do danych ze spisu pełnego w odniesieniu do podstawowych zmiennych demograficznych, wykorzystano kalibrację. Jej podstawy teoretyczne zostały sformułowane w pracy Deville'a i Särndala (1992). Szeroko technika ta została opisana również w monografii, której autorami są Särndal i Lundström (2005), także w kontekście możliwości stosowania jej w badaniach statystycznych z brakami odpowiedzi oraz w artykule Särndala (2007).

W tym opracowaniu sformułowano niezbędne pojęcia i definicje, które istotne są z punktu widzenia prowadzonych rozważań. Zgodnie z ideą zaproponowaną przez Deville'a i Särndala kalibracja to metoda korygowania wyjściowych wag, wynikających ze schematu losowania próby z wykorzystaniem zmiennych pomocniczych, tak aby odtworzyć znane wartości globalne tych zmiennych dla całej populacji. Informacje o zmiennych pomocniczych, jak i ich wartościach globalnych pochodzą przy tym najczęściej ze spisów bądź rejestrów administracyjnych.

W ujęciu matematycznym kalibrację można zatem opisać w następujący sposób: niech dana będzie N -elementowa populacja $U = \{1, \dots, N\}$. Z populacji tej losujemy zgodnie z określonym schematem losowania n -elementową próbę $s \subseteq U$. Niech π_i oznacza prawdopodobieństwo inkluzji i -tej jednostki do próby, tzn. $\pi_i = P(i \in s)$ dla $i = 1, \dots, N$, a $d_i = \frac{1}{\pi_i}$ będzie wagą odpowiadającą jednostce i .

Założmy, że celem badania jest oszacowanie wartości globalnej pewnej zmiennej y , określonej wzorem:

$$Y = \sum_{i=1}^N y_i \quad (1)$$

gdzie y_i — wartość zmiennej y dla i -tej jednostki badania, $i = 1, \dots, N$.

Klasycznym estymatorem wartości globalnej (1) jest znany z metody reprezentacyjnej estymator Horvitz-Thompsona, który wyraża się wzorem:

$$\hat{Y}_{HT} = \sum_s d_i y_i \quad (2)$$

³ Wyznaczenie wag kalibracyjnych przeprowadzono w Ośrodku Statystyki Małych Obszarów w Urzędzie Statystycznym w Poznaniu. Autor składa podziękowania Tomaszowi Józefowskiemu, Tomaszowi Klimankowi, Andrzejowi Młodakowi, Janowi Paradyszowi i Łukaszowi Wawrowskiemu, bez udziału których wyznaczenie wag kalibracyjnych w NSP 2011 nie byłoby możliwe.

Jeżeli w odniesieniu do zmiennej y znane są z badania reprezentacyjnego również wartości zmiennych pomocniczych $x_1, x_2, \dots, x_k$ oraz ich wartości globalne $X_j = \sum_{i=1}^N x_{ij}$, gdzie $j=1, 2, \dots, k$, może się zdarzyć, że:

$$\sum_s d_i x_{ij} \neq X_j \quad (3)$$

gdzie x_{ij} — wartość j -tej zmiennej pomocniczej dla i -tej jednostki badania. Oznacza to, że wagi wynikające ze schematu losowania próby d_i nie odtwarzają znanych struktur w odniesieniu do podstawowych zmiennych pomocniczych. Inaczej mówiąc, wagi nie odtwarzają znanych wartości globalnych X_j zmiennych pomocniczych z badania pełnego.

W praktyce dokonuje się zatem korekty wag d_i tak, aby odtworzone zostały z wykorzystaniem danych pochodzących z badania reprezentacyjnego wartości globalne X_j dla wszystkich kluczowych zmiennych pomocniczych. Ma to zagwarantować swego rodzaju „równowagę” rozumianą w ten sposób, że po zastosowaniu kalibracji próba „wyglądem” jest zbliżona do całej populacji. Należy rozumieć to w ten sposób, że jeżeli zastosujemy wagi kalibracyjne do zmiennych pomocniczych $x_1, x_2, \dots, x_k$, to odtworzymy dokładnie ich wartości globalne dla całej populacji. Wyznaczone wagi kalibracyjne powinny przy tym być bliskie wagom oryginalnym. Celem jest zatem znalezienie nowych wag $\mathbf{w} = (w_1, \dots, w_n)^T$ tzw. wag kalibracyjnych.

W dalszej części artykułu będziemy zakładać, że $\mathbf{d} = (d_1, \dots, d_n)^T$ jest wektorem wag wynikających ze schematu losowania próby.

Ponadto niech $\mathbf{w} = (w_1, \dots, w_n)^T$ oznacza poszukiwany wektor wag kalibracyjnych i G będzie dowolną funkcją spełniającą następujące warunki na zbiorze liczb rzeczywistych dodatnich:

- G jest ściśle wypukła i dwukrotnie różniczkowalna,
- $G \geq 0$,
- $G(1) = 0$,
- $G'(1) = 0$,
- $G''(1) = 1$.

Niech ponadto $x_1, \dots, x_k$ oznaczają zmienne pomocnicze, które wykorzystane zostaną do wyznaczania wag kalibracyjnych, a X_j oznacza wartość globalną zmiennej x_j , $j=1, \dots, k$, tj.:

$$X_j = \sum_{i=1}^N x_{ij} \quad (4)$$

gdzie x_{ij} — wartość j -tej zmiennej pomocniczej dla i -tej jednostki badania.

Problem poszukiwania wag kalibracyjnych w ujęciu matematycznym można przedstawić następująco:

- (W1) minimalizacja funkcji odległości:

$$D(\mathbf{w}, \mathbf{d}) = \sum_{i \in S} d_i G\left(\frac{w_i}{d_i}\right) \rightarrow \min \quad (5)$$

- (W2) równania kalibracyjne:

$$\sum_{i \in S} w_i x_{ij} = X_j \quad j = 1, \dots, k \quad (6)$$

- (W3) warunki ograniczające:

$$L \leq \frac{w_i}{d_i} \leq U, \quad \text{gdzie } 0 < L < 1 \text{ i } U > 1, i = 1, \dots, n \quad (7)$$

Pierwszy z warunków (W1) orzeka, że wagi kalibracyjne powinny być w taki sposób wyznaczone, aby były możliwie bliskie — w sensie przyjętej funkcji odległości — wogom wynikającym ze schematu losowania próby. Funkcja G mierzy odległość między ilorazem wag w_i/d_i a 1 i jest wykorzystywana w konstrukcji funkcji odległości D , dla której szuka się lokalnego minimum warunkowego.

Warunek (W2) stanowi istotę teorii kalibracji i orzeka, że wagi powinny być tak dobrane, aby po ich zastosowaniu do wszystkich zmiennych pomocniczych można było uzyskać ich wartości globalne. Jeżeli ten warunek będzie spełniony, to również po wykorzystaniu tych wag do zmiennej y powinniśmy dostać ocenę wartości globalnej bliską jej prawdziwej wartości. Z kolei (W3) jest tzw. warunkiem ograniczającym, który zapobiegać ma sytuacjom, w których uzyskane wagi kalibracyjne przyjmują wartości ujemne bądź ekstremalne. Występowanie wag ujemnych jest sprzeczne z definicją wagi, która będąc odwrotnością prawdopodobieństwa dostania się jednostki do próby przyjmuje wartości dodatnie. Występowanie wag ekstremalnych, znacznie odbiegających od wag wynikających ze schematu losowania próby, wpłynąć może na obciążenie i wariancję stosowanych estymatorów.

Istnieje dowolność przy wyborze funkcji G . Najczęściej rozważa się następujące jej postaci:

$$G_1(x) = \frac{1}{2}(x-1)^2 \quad (8)$$

$$G_2(x) = \frac{(x-1)^2}{x} \quad (9)$$

$$G_3(x) = x(\log x - 1) + 1 \quad (10)$$

$$G_4(x) = 2x - 4\sqrt{x} + 2 \quad (11)$$

$$G_5(x) = \frac{1}{2\alpha} \int_1^x \sinh\left[\alpha\left(t - \frac{1}{t}\right)\right] dt \quad (12)$$

gdzie:

α — dodatni parametr, pozwalający sterować stopniem rozrzutu wag kalibracyjnych w stosunku do wag wynikających ze schematu losowania próby (domyślnie parametr przyjmuje wartość 1),

$\sinh$ — funkcja sinusa hiperbolicznego zdefiniowanego jako $\sinh(x) = \frac{e^x - e^{-x}}{2}$.

Przykładowo, dla funkcji G zdefiniowanej wzorem (8) funkcja odległości D , dla której poszukuje się lokalnego minimum warunkowego, przyjmuje postać:

$$D(\mathbf{w}, \mathbf{d}) = \sum_{i \in S} d_i G\left(\frac{w_i}{d_i}\right) = \sum_{i \in S} d_i \frac{1}{2} \left(\frac{w_i}{d_i} - 1\right)^2 = \frac{1}{2} \sum_{i \in S} \frac{(w_i - d_i)^2}{d_i} \quad (13)$$

Mając wybraną funkcję G , która stanowi podstawę w konstrukcji funkcji odległości D , można zdefiniować estymator kalibracyjny wartości globalnej zmiennej y w następujący sposób:

Definicja. Estymatorem kalibracyjnym wartości globalnej (1), przy znanym wektorze wartości globalnych $\mathbf{x}$ zmiennych pomocniczych, jest:

$$\hat{Y}_X = \sum_{i \in S} w_i y_i \quad (14)$$

gdzie wektor wag $\mathbf{w} = (w_1, \dots, w_n)^T$ jest rozwiązaniem zadania optymalizacyjnego, tj.:

$$\mathbf{w} = \arg\min_{\mathbf{v}} D(\mathbf{v}, \mathbf{d}) \quad (15)$$

przy warunku:

$$\mathbf{x} = \tilde{\mathbf{x}} \quad (16)$$

gdzie:

$$\tilde{\mathbf{x}} = \left(\sum_{i \in S} w_i x_{i1}, \sum_{i \in S} w_i x_{i2}, \dots, \sum_{i \in S} w_i x_{ik} \right)^T \quad (17)$$

a wektor wartości globalnych $\mathbf{x}$ zdefiniowany jest wzorem:

$$\mathbf{x} = \left(\sum_{i=1}^N x_{i1}, \sum_{i=1}^N x_{i2}, \dots, \sum_{i=1}^N x_{ik} \right)^T \quad (18)$$

W praktyce bardzo często wykorzystuje się funkcję G zdefiniowaną wzorem (8), która umożliwia konstrukcję funkcji odległości D zgodnie ze wzorem (13). Wynika to z faktu, że dla tak zdefiniowanej funkcji wagi kalibracyjne można uzyskać wprost ze wzoru, co pokazuje twierdzenie.

Twierdzenie. Rozwiązaniem zadania minimalizacji funkcji odległości (15), przy warunku (16) dla funkcji odległości (13), jest wektor wag kalibracyjnych postaci $\mathbf{w} = (w_1, \dots, w_n)^T$, którego składowe spełniają równanie:

$$w_i = d_i + d_i (\mathbf{x} - \hat{\mathbf{x}})^T \left(\sum_{i \in S} d_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \mathbf{x}_i \quad (19)$$

przy czym:

$$\hat{\mathbf{x}} = \left(\sum_{i \in S} d_i x_{i1}, \sum_{i \in S} d_i x_{i2}, \dots, \sum_{i \in S} d_i x_{ik} \right)^T \quad (20)$$

a

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^T \quad (21)$$

jest wektorem złożonym z wartości wszystkich k zmiennych pomocniczych dla i -tej jednostki, $i = 1, \dots, n$.

Ponadto wzór (19) umożliwia wyznaczenie wag kalibracyjnych w jawnej postaci i nie jest wymagane korzystanie z iteracyjnych metod wyznaczania wag, które nie zawsze gwarantują osiągnięcie zbieżności algorytmu ich poszukiwania.

Funkcja (13) pozwala również wyznaczyć wagi kalibracyjne stosunkowo szybko, co przy bardzo wielu przekrojach do odtworzenia jest szczególnie ważną i pożądaną własnością. Należy jednak również pamiętać o słabych stronach funkcji odległości (13):

- 1) w przypadku źle dobranego wektora zmiennych pomocniczych mogą pojawić się ujemne wagi kalibracyjne, co jest sprzeczne z ideą wag;
- 2) dla niektórych przekrojów reprezentowanych w próbie przez bardzo małą liczbę jednostek mogą pojawić się wagi ekstremalne. Dotyczy to sytuacji, gdy dla danego przekroju w próbie istnieje niewielka liczba jednostek, podczas gdy w odpowiadającym mu przekroju w badaniu pełnym (np. spisie) liczba jednostek jest znacznie większa.

Pierwszą z tych niedogodności można zniwelować wykorzystując funkcje odległości z ograniczeniami nałożonymi na wagi. Nie zawsze jest to jednak w praktyce możliwe do wykonania ze względu na brak zbieżności odpowiednich algorytmów iteracyjnych poszukiwania wag kalibracyjnych.

Druga z niedogodności jest charakterystyczna dla wszystkich funkcji odległości, a nie tylko dla funkcji (13). Powstanie zatem wag ekstremalnych związane jest w takich sytuacjach najczęściej ze zbytnią szczegółowością danego przekroju⁴ bądź z niedoreprezentowaniem w próbie takiego przekroju w stosunku do odpowiadającego mu przekroju w badaniu pełnym.

MAKRO CALMAR

Na potrzeby wyznaczenia wag kalibracyjnych w NSP 2011 dla ludności rezydującej wykorzystano makro CALMAR (Sautory, 2003) (nazwa programu bierze się od zwrotu **Cal**ibration on **Marg**ins), które:

- jest napisane w języku 4GL w środowisku SAS i służy do wyznaczania wag kalibracyjnych,
- jest opracowane na potrzeby prac francuskiego urzędu statystycznego⁵,
- oferuje 4 sposoby wyznaczania wag kalibracyjnych w zależności od przyjętej postaci funkcji G .

CALMAR⁶ oferuje cztery sposoby wyznaczania wag kalibracyjnych, w zależności od postaci funkcji G :

- podejście liniowe:

⁴ Na przykład w badaniu reprezentacyjnym dla danego powiatu, w grupie 35-letnich mężczyzn w części miejskiej powiatu może być tylko dwóch reprezentantów.

⁵ Dokumentacja techniczna dostępna tylko w języku francuskim.

⁶ Na potrzeby prac polegających na wyznaczaniu wag kalibracyjnych w NSP 2011 dokonano spolszczenia wyników generowanych przez makro CALMAR. Należy również podkreślić, że uruchomienie makro wymaga wcześniejszego przygotowania odpowiednich zbiorów wejściowych do CALMAR-a i naliczenia wartości globalnych w ściśle określonym porządku. Wymaga to z reguły czasochłonnych prac programistycznych.

$$G(x) = \frac{1}{2}(x-1)^2 \quad (22)$$

- raking ratio:

$$G(x) = x(\log x - 1) + 1 \quad (23)$$

- podejście logitowe:

$$G(x) = \left[(x-L) \log \frac{x-L}{1-L} + (U-x) \log \frac{U-x}{U-1} \right] \frac{1}{A} \quad (24)$$

gdzie:

$$A = \frac{U-L}{(1-L)(U-1)} \quad x \in (L, U) \quad (25)$$

- podejście liniowe z warunkami ograniczającymi:

$$G(x) = \frac{1}{2}(x-1)^2 \quad L \leq \frac{w_i}{d_i} \leq U \quad (26)$$

W przypadku pierwszej funkcji mogą pojawić się wagi ujemne. Dla drugiej funkcji eliminowana jest niedogodność związana z występowaniem wag ujemnych, mogą jednak wystąpić znaczne odchylenia od wag wynikających ze schematu losowania próby.

Jedynie wykorzystując ostatnie dwie postacie funkcji G jesteśmy w stanie zapewnić, aby wagi kalibracyjne znajdowały się w z góry określonym przedziale, czyli narzucamy warunek, aby nie różniły się one zbyt od wag wynikających ze schematu losowania próby i aby były dodatnie. Zbyt restrykcyjne przyjęcie parametrów L i U może jednak powodować brak zbieżności algorytmu iteracyjnego służącego do wyznaczania wag i w konsekwencji brak rozwiązań. Dotyczy to przede wszystkim sytuacji bardzo dużej liczby przekrojów, dla których należy zapewnić zgodność struktur z badania reprezentacyjnego i spisu pełnego. Dla niektórych bowiem przekrojów może wystąpić sytuacja już opisana, a więc mała liczebność próby w badaniu reprezentacyjnym w danym przekroju, co rzutuje bądź na występowanie ekstremalnych wag kalibracyjnych, bądź na niemożliwość osiągnięcia zbieżności algorytmu ich wyznaczania.

Na potrzeby prac nad wyznaczaniem wag kalibracyjnych wykorzystano ostatecznie pierwszą z funkcji G , którą oferuje makro CALMAR (wzory (8)

i (22). W przypadku pozostałych funkcji jakie oferuje makro CALMAR, podczas narzucania warunków ograniczających na wagi pojawiał się problem ze zbieżnością algorytmu i dla wielu przekrojów nie było możliwości wyznaczenia wag kalibracyjnych.

Po przeprowadzeniu kalibracji dokonano analizy uzyskanych wag kalibracyjnych z punktu widzenia ich własności statystycznych. Skupiono się przy tym na analizie występowania wag ujemnych i ekstremalnych, korelacyjnych diagramów rozrzutu wag, wynikających ze schematu losowania próby i wag kalibracyjnych oraz na sprawdzeniu, czy we wszystkich przekrojach odtworzono dla wybranych zmiennych pomocniczych znane struktury z badania pełnego.

KALIBRACJA W NSP 2011

Jak to już nadmieniono, wykorzystanie danych z różnych źródeł wymagało, na etapie uogólniania wyników badania reprezentacyjnego, odpowiedniego skorygowania wag dla ludności rezydującej. Było to spowodowane koniecznością odpowiedniego zintegrowania i zachowania spójności pomiędzy wynikami badania reprezentacyjnego z danymi pochodzącymi ze spisu pełnego w zakresie podstawowych zmiennych.

W charakterze zmiennych pomocniczych wykorzystanych przy wyznaczaniu wag kalibracyjnych uwzględniono najważniejsze zmienne demograficzne odnoszące się do płci, wieku oraz miejsca zamieszkania dla powiatu z wyodrębnieniem części miejskiej i wiejskiej. Wybór ten podyktowany był faktem, że informacje dla tych zmiennych dostępne były dla wszystkich osób i nie zawierały braków informacji. Ponadto, jak wskazuje literatura przedmiotu, zmienne demograficzne wykorzystywane są najczęściej przy poszukiwaniu wag kalibracyjnych w różnego rodzaju badaniach społecznych, w tym i w spisach (Särndal, Lundström, 2005). Przyjęto przy tym założenie, że kalibracja będzie przeprowadzona w każdym z przekrojów wyznaczonych przez kategorię zmiennej płeć (mężczyzna i kobieta), miejsce zamieszkania (część miejska i wiejska) oraz wiek.

W odniesieniu do wieku przyjęto następujące założenia: dla województwa na potrzeby kalibracji przyjęto roczniki wieku, tj. 0, 1, ..., 84, 85+, a dla powiatu grupy wieku: 0—4, 5—9, 10—12, 13—14, 15—17, 18—19, 20—24, 25—29, 30—34, 35—39, 40—44, 45—49, 50—54, 55—59, 60—64, 65—69, 70—74, 75—79, 80—84, 85+. Oznaczało to, że wagi kalibracyjne wyznaczone były w taki sposób, aby odtwarzać znane struktury demograficzne w województwach ze spisu pełnego w każdym z przekrojów generowanych przez kategorię płci, miejsca zamieszkania i rocznika wieku. Zdecydowano przy tym, że dla roczników wieku 85, 86 itd. utworzona zostanie jedna kategoria 85+. Wagi te były wyznaczone również w taki sposób, aby oprócz zapewnienia zgodności dla województw zachowana została zgodność na poziomie powiatów. W odniesieniu do powiatów przyjęto przy tym założenie, aby odtwarzać znane struktury demograficzne ze spisu pełnego w każdym z przekrojów generowanych przez

kategorię płci, miejsce zamieszkania i grupę wieku. Przyjęto przy tym w większości przypadków 5-letnie grupy wieku. Wyjątek stanowiły roczniki 10, 11, ..., 19, w ramach których utworzono grupy wieku 10—12, 13—14, 15—17, 18—19 oraz grupa 85+.

Wagi kalibracyjne konstruowane były również w taki sposób, aby odtwarzane były znane struktury ze spisu pełnego dla pięciu największych miast w kraju (Poznania, Wrocławia, Łodzi, Krakowa i Warszawy). Uwzględniono przy tym osobno delegatury dla Poznania, Wrocławia, Łodzi i Krakowa przyjmując w każdej delegaturze i kategorii płci roczniki wieku 0, 1, ..., 84, 85+. Dla Warszawy, w odniesieniu do dzielnic, przyjęto podobne założenie. Wagi były jednak w taki sposób wyznaczane, aby dla Warszawy jako całości odtwarzane były struktury w każdym z roczników wieku 0, 1, ..., 99, 100+, z uwzględnieniem płci (tabl. 1).

TABL. 1. SPOSÓB TWORZENIA PRZEKROJÓW W PODEJŚCIU KALIBRACYJNYM W NSP 2011 — LUDNOŚĆ REZYDUJĄCA

Poziom agregacji/ /Zmienna pomocnicza	Część miejska/ /wiejska — 1, 2	Płeć — 1, 2	Grupy wieku — 0—4, 5—9, ..., 80—84, 85+	Pojedyncze roczniki wieku	
				0, 1, ..., 83, 84, 85+	0, 1, ..., 98, 99, 100+
Polska	1	1	1	1	0
Województwa	1	1	1	1	0
Powiaty (bez 5 największych miast)	1	1	1	0	0
Największe miasta	x	1	1	1	0
Warszawa	x	1	1	1	1
Dzielnice Warszawy	x	1	1	1	0
Delegatury 4 największych miast	x	1	1	1	0

U w a g a. 1 — kalibracja możliwa, 0 — kalibracja niemożliwa⁷, x — przekrój nieadekwatny.

Ź r ó d ł o: opracowanie własne.

Tabl. 1 pokazuje przekroje, dla których możliwe było wyznaczenie wag kalibracyjnych. Należy podkreślić, że kalibracja wag wyznaczana była dla kombinacji przekrojów wyznaczanych przez warianty zmiennych płeć, miejsce zamieszkania i wiek, a nie dla każdego przekroju osobno. Wszystkich przekrojów, dla których należało wyznaczyć wagi kalibracyjne, tak aby odtworzone zostały wartości globalne znane ze spisu pełnego, łącznie było ok. 43 tys.

⁷ Oznacza to, że w niektórych przekrojach nie było reprezentanta w próbie i nie było możliwe w takich sytuacjach kalibrowanie wag wynikających ze schematu losowania. Przykładowo, w jakimś powiecie w próbie mogła nie znaleźć się osoba 40-letnia i nie można było skalibrować wag tak, aby odtworzyć znaną ze spisu pełnego liczbę takich osób w analizowanym przekroju.

STATYSTYCZNA OCENA WAG KALIBRACYJNYCH

Na potrzeby wyznaczania wag kalibracyjnych dla ludności rezydującej w NSP 2011, jak już zasygnalizowano wcześniej, skorzystano z funkcji (8), która umożliwia wyznaczenie wag kalibracyjnych wprost ze wzoru (19). Wybór tej funkcji podyktowany był tym, że dla pozostałych funkcji odległości algorytmy iteracyjne w niektórych przekrojach nie osiągały zbieżności i nie było możliwości wyznaczenia wag kalibracyjnych.

Choć funkcja ta najczęściej jest wykorzystywana na potrzeby kalibracji w praktycznych zastosowaniach, to należy jednak dokonać oceny wag kalibracyjnych mając na uwadze jej ograniczenia dotyczące m.in. możliwości wystąpienia wag ujemnych, co byłoby zjawiskiem niepożądanym z punktu widzenia ich wykorzystywania w uogólnianiu wyników. Dokonano zatem analizy wag kalibracyjnych uwzględniając następujące aspekty:

- 1) występowanie ujemnych wag kalibracyjnych,
- 2) stopień odtworzenia znanych ze spisu pełnego struktur we wszystkich założonych przekrojach,
- 3) stopień skorelowania wag wynikających ze schematu losowania próby d_i i wag kalibracyjnych w_i oraz analiza histogramów wag,
- 4) występowanie wag ekstremalnych.

Ad 1. Dokonując analizy uzyskanych wag kalibracyjnych we wszystkich założonych przekrojach nie stwierdzono występowania wag ujemnych. Jest to szczególnie pożądana własność wag po kalibracji, gdyż nie stoi w sprzeczności z definicją wagi, która jest odwrotnością prawdopodobieństwa dostania się jednostki do próby i przyjmuje wartości dodatnie.

Ad 2. Dla wszystkich założonych przekrojów wyznaczono ze spisu pełnego odpowiednie tablice z liczbą ludności rezydującej i porównano je z naliczonymi tablicami z wykorzystaniem wag kalibracyjnych i danych ze spisu reprezentacyjnego. Wykorzystując procedurę *proc compare* w systemie SAS stwierdzono zgodność wszystkich porównywanych ze sobą tablic. Oznacza to, że wyznaczone wagi kalibracyjne odtwarzały dokładnie, znane ze spisu pełnego, struktury demograficzne dla wszystkich założonych przekrojów na dowolnym poziomie agregacji, czyli z uwzględnieniem województwa, powiatu oraz pięciu największych miast w Polsce.

Ad 3. Dla wszystkich założonych przekrojów dokonano analizy wykresów korelacyjnych i histogramów rozkładów wag d_i i w_i . Wykr. 1 przedstawia przykładowy korelacyjny diagram rozrzutu i histogram wag w pow. poznańskim. Jak pokazuje analiza rozrzutu istnieje silna dodatnia zależność między wagami wynikającymi ze schematu losowania próby d_i a wagami kalibracyjnymi w_i . Jest to pożądana własność, gdyż oznacza, że wyższym wagom d_i odpowiadają po kalibracji również wyższe wartości wag w_i i na odwrót, niższym wagom d_i odpowiadają po kalibracji niższe wartości wag kalibracyjnych w_i . Przy analizie rozrzutu ważna jest nie tylko ocena stopnia

korelacji między dwoma rodzajami wag, ale również sposób ułożenia się punktów. Pożądane jest, aby punkty układały się wzdłuż prostej $y=x$, gdyż oznaczać to będzie, że po kalibracji wagi nie będą odbiegać zbyt od wag wynikających ze schematu losowania próby.

Analizując histogram wag należy zwrócić uwagę na dwie rzeczy: po pierwsze na „spłaszczanie” się histogramu dla wag kalibracyjnych w porównaniu z histogramem wag wejściowych d_i . Wynikało to z faktu, że suma wag d_i wynikających ze schematu losowania próby w pow. poznańskim i wynosząca 348324,57 była wyższa od rzeczywistej liczby ludności rezydującej w tym powiecie, która wynosiła 329826. Oznaczało to, że w kalibracji wagi d_i musiały na ogół zmaleć, tak aby ich suma po korekcie wynosiła 329826. W konsekwencji histogram wag kalibracyjnych był mniej „stromy” w porównaniu z wejściowym histogramem wag. Zjawisko to zaobserwowano w bardzo wielu przekrojach, co oznaczało, że suma wag d_i była zawyżona w stosunku do prawdziwej liczby ludności rezydującej ze spisu pełnego. Po drugie, należy zwrócić uwagę analizując histogramy wag, jakie były ich kształty przed i po kalibracji. Jak pokazuje analiza histogramów w pow. poznańskim, ich kształt był „bardzo podobny”, co jest zjawiskiem pożądanym. Podobnych analiz dokonano dla pozostałych przekrojów i nie stwierdzono znaczących różnic w kształtach obydwu histogramów.

Analiza wykresu rozrzutu pozwala również wychwycić występowanie wag ekstremalnych, tzn. wag, które po kalibracji najczęściej znacznie wzrosły. W pow.

poznańskim dokonując oceny korelacyjnego diagramu rozrzutu nie zaobserwowano wag ekstremalnych. Potwierdzeniem tego jest statystyka opisowa wag (tabl. 2) oraz kształt histogramu przedstawiającego rozkład ilorazu wag kalibracyjnych i wynikających ze schematu losowania próby (wykr. 2). Ilorazy wag koncentrowały się najczęściej w pobliżu 1. Oznacza to, że wagi kalibracyjne nie odbiegały od wag wynikających ze schematu losowania próby.

TABL. 2. STATYSTYKA OPISOWA WAG W POW. POZNAŃSKIM

W a g a	Minimum	Maksimum	Suma	Mediana	Odchylenie standardowe
d_i	1,3919308	13,8937500	348324,57	7,9692308	1,8659472
w_i	1,0942524	14,6566282	329826,00	7,5879680	1,8179145

Ź r ó d ł o: opracowanie własne.

Nie dla wszystkich jednak poddanych analizie przekrojów uzyskano takie wykresy rozrzutu, co było konsekwencją występowania wśród wag kalibracyjnych wag ekstremalnych, nieraz znacznie odbiegających od wag d_i .

Ad 4. Analizy wag kalibracyjnych dokonano również z punktu widzenia występowania wag ekstremalnych. Wartości odstające można było wychwycić dokonując graficznej oceny korelacyjnego diagramu rozrzutu bądź tablic ze statystyką opisową wag. Wykr. 3 przedstawia przykładowy wykres rozrzutu wag dla Warszawy. Nie ma on charakteru jednolitej „smugi” położonej blisko prostej $y=x$ i występują na nich punkty leżące z „dala” od niej. Oznacza to, że po kalibracji wagi zmieniły znacznie swoją wartość i odbiegają od wag wejściowych d_i wynikających ze schematu losowania próby.

W takich sytuacjach należy zachować szczególną uwagę przy estymacji, zwłaszcza gdyby wagi ekstremalne występowały na dużą skalę. W przypadku Warszawy można było jednak zaobserwować, że wśród wag kalibracyjnych nie ma zbyt dużo wartości ekstremalnych. Co więcej, dotyczą one starszych

roczników wieku. Wzrost wag miał bowiem miejsce w sytuacji, gdy w danym przekroju (np. mężczyzn w wieku 98 lat) występował tylko jeden reprezentant w próbie, podczas gdy w spisie pełnym było ich kilku. W takim przypadku naturalne było, że waga wejściowa d_i musiała znacznie wzrosnąć, tak aby reprezentować wszystkie takie jednostki w spisie pełnym.

Dokonując podsumowania dotychczasowych rozważań należy podkreślić, że szczególną uwagę podczas uogólniania wyników należy zwrócić w starszych grupach wiekowych (85+) czy w starszych rocznikach wieku (85, 86...). Zaobserwowano tam bowiem występowanie ekstremalnych wag kalibracyjnych, które w estymacji mogą wpłynąć na uzyskiwane wyniki. W przypadku uogólniania wyników na dość dużym stopniu agregacji (np. województwo czy nawet powiat) wpływ ekstremalnych wag kalibracyjnych będzie pomijany. Ich znaczenie może jednak wzrastać w sytuacji, gdy estymacja wyników z wykorzystaniem wag kalibracyjnych odbywać się będzie dla szczegółowo zdefiniowanych przekrojów czy domen. Ze względu jednak na fakt, że występowanie ekstremalnych wag kalibracyjnych odnotowano w starszych rocznikach i grupach wieku dla wielu tablic spisowych z różnych tematów (np. rynek pracy) nie będą one miały większego znaczenia.

Podsumowanie

Opisana technika korygowania wag wynikających ze schematu losowania próby była wykorzystana z powodzeniem w NSP 2011. Umożliwiła ona dostosowanie struktur w odniesieniu do najważniejszych zmiennych demograficznych z badania reprezentacyjnego do znanych struktur ze spisu pełnego. Wyznaczone zatem wagi kalibracyjne mogą być wykorzystane przy uogólnianiu wyników w odniesieniu do wszystkich zmiennych z części reprezentacyjnej spisu. Jak bowiem wykazała analiza wyznaczonych wag kalibracyjnych nie stwierdzono występowania wag ujemnych, a wagi ekstremalne występowały w niektórych przekrojach i dotyczyły w zasadzie starszych roczników wieku.

Kalibrację, jako technikę korygowania wag wynikającą ze schematu losowania próby, z powodzeniem zastosowano w NSP 2011, a zatem może stanowić ona cenną metodę wykorzystywaną w innych badaniach prowadzonych przez GUS⁸. Może być ona zatem wykorzystywana w spisach nowej generacji, w których dane uzyskuje się z różnych źródeł metodą mieszaną.

dr Marcin Szymkowiak — *Uniwersytet Ekonomiczny w Poznaniu, Urząd Statystyczny w Poznaniu*

⁸ Warto zaznaczyć, że technika ta jest już stosowana przez GUS w badaniu EU-SILC (Kordos i in., 2011, s. 81–96).

LITERATURA

- Deville J-C., Särndal C-E. (1992), *Calibration Estimators in Survey Sampling*, „Journal of the American Statistical Association”, Vol. 87
- Kordos J., Zięba-Pietrzak A., Wieczorkowski R. (2011), *Bootstrap method with calibration for standard error estimation of income poverty measures*, „Statistics in Transition — new series”, Vol. 12, No.1
- Raport z wyników Narodowy Spis Powszechny Ludności i Mieszkań 2011 (2012), GUS
- Sautory O. (2003), *Calmar 2: A New Version of the Calmar Calibration Adjustment Program*, Statistics Canada International Symposium Series — Proceedings, Symposium 2003: Challenges in Survey Taking for the Next Decade
- Särndal C-E., Lundström S. (2005), *Estimation in Surveys with Nonresponse*, John Wiley & Sons, Ltd.
- Särndal C-E. (2007), *The Calibration Approach in Survey Theory and Practice*, „Survey Methodology”, Vol. 33, No. 2

SUMMARY

National Census of Population and Housing 2011 (NSP 2011) for the first time in the history of Polish public statistics was conducted by mixed method. This meant the use of data from administrative sources and sample survey, covering 20% of apartments in Poland. During generalizing the results of a sample survey, the data required corresponding correct weights for all persons involved therein. The purpose of this article is to present the theoretical basis of calibration and its practical application to the needs of the census. It discusses the idea of calibration, as well as stages of its use in adapting the structures of gender, age and place of residence from the representative survey data to the data of the full Census (2002), which was treated as a reference population.

РЕЗЮМЕ

Всеобщая национальная перепись населения и квартир 2011 (ВНП 2011) впервые в истории польской официальной статистики проводилась так называемым смешанным методом. Этот метод заключался в использовании данных из административных источников и из выборочного обследования, охватывающего 20% квартир в Польше. Эти данные требовали во время обобщения результатов выборочного обследования, соответствующей корректировки весов для всех лиц принимающих в нем участие.

Целью статьи является представление теоретических основ калибровки вместе с ее практическим использованием для целей переписи. Статья обсуждает идею калибровки, а также этапы её использования в приспособливании структур пола, возраста и места проживания из выборочного обследования к данным из полной переписи, которая составляла образцовую (модельную) популяцию.