Andrzej MŁODAK

Zastosowanie metody bootstrapowej do oceny precyzji szacunków danych z badania reprezentacyjnego NSP 2011 na przykładzie dojazdów do pracy

Jednym z istotnych zadań realizowanych w ramach analizy danych uzyskanych z Narodowego Spisu Ludności i Mieszkań 2011 (NSP 2011), mającym na celu dostarczenie informacji o stanie i strukturze demograficznej oraz społeczno-ekonomicznej ludności Polski, jest estymacja określonych danych dla całej populacji na podstawie informacji z badania reprezentacyjnego, które stanowiło integralną część tegoż spisu. Estymacji takiej dokonuje się zazwyczaj wykorzystując wagi kalibracyjne umożliwiające oszacowanie wartości wyrażanych w liczbach bezwzględnych (np. liczba osób aktywnych zawodowo czy liczba osób z wyższym wykształceniem itp.) dla całej rozpatrywanej populacji przy użyciu posiadanych danych dla wylosowanej próby reprezentacyjnej.

Każde tego rodzaju oszacowanie obarczone jest jednak błędem wynikającym m.in. z zakłóceń reprezentatywności próby, błędów systematycznych itp. Aby móc efektywnie ocenić jakość uzyskanych szacunków i ustalić ewentualny zakres ulepszeń, trzeba także mieć wiedzę na temat wielkości i rozkładu owych błędów. W artykule pokazujemy przykład zastosowania nieparametrycznej metody szacowania rozkładu błędów estymacji za pomocą wielokrotnego losowania ze zwracaniem danych z próby (*bootstrap*), która umożliwia ocenę precyzji szacunków uzyskanych w przypadku braku informacji o otrzymanych lub pokrewnych parametrach dla populacji generalnej. Dzięki przeprowadzonemu eksperymentowi symulacyjnemu możliwe stało się oszacowanie względnego pierwiastkowego średniego kwadratu błędu, obciążenia względnego oraz empirycznego obciążenia względnego szacunków danych dla populacji charakteryzujących różnorodne aspekty dojazdów do pracy. Oczywiście tego rodzaju wskaźniki jakościowe mieszczą w sobie również ładunek błędów nielosowych. Ich wielkość można ocenić poprzez badanie kontrolne (*Post Enumeration…*, 2010) lub korektę logiczną.

Artykuł składa się z czterech zasadniczych części. W pierwszej podajemy szczegółową charakterystykę idei i cech metody bootstrapowej. Część druga poświęcona została omówieniu założeń przeprowadzonego eksperymentu. Część trzecia zawiera prezentację i analizę jego rezultatów. Wnioski płynące z tego doświadczenia, jak również sugestie dotyczące możliwej poprawy jakości szacunków zebrano w części czwartej.

IDEA METODY BOOTSTRAPOWEJ

Punktem wyjścia idei bootstrappingu jest ocena natężenia zjawisk społeczno-gospodarczych na podstawie badania reprezentacyjnego dokonana na wylosowanej próbie członków danej populacji. Inaczej mówiąc, chcemy oszacować wartość parametrów używając odpowiedniej statystyki próbkowej. Mogą to być zarówno proste miary tendencji centralnej (np. średnia arytmetyczna czy mediana), jak również statystyka charakteryzująca jakość uzyskanych oszacowań tych miar (odchylenie standardowe, średni kwadrat błędu — *MSE*, granice przedziałów ufności itp.). Podstawę bootstrappingu stanowi obserwacja, że wartości tych samych statystyk próbkowych, uzyskiwane z kolejnych próbek losowanych w ten sam sposób z danej populacji, mogą się do pewnego stopnia różnić. Stąd też wynika potrzeba oceny wielkości owych odchyleń czy dyspersji rzeczonych wartości za pomocą odpowiednich parametrów rozkładu próbkowego, czyli rozkładu wszystkich możliwych wartości szukanej statystyki uzyskanej z owych próbek. Jedną z metod przeprowadzania takiej oceny, niewymagającą istotnego angażowania zaawansowanych narzędzi teoretycznych, jest metoda bootstrapowa.

Podstawowe założenie omawianej metody polega na wielokrotnie powtarzanym losowaniu próbek jednakowego rozmiaru z danej populacji. Ta duża zazwyczaj liczba losowań umożliwia uzyskanie obrazu rozkładu potencjalnych wartości statystyki próbkowej. Jednakże w praktyce dokonywanie znacznej liczby losowań z liczącej wiele elementów populacji (najlepszym tego przykładem jest właśnie ogół osób spisywanych podczas NSP 2011) pociąga za sobą bardzo duże koszty i wymaga ogromnej wydajności obliczeniowej. Z tego powodu jako „populacji zastępczej" używa się wylosowanej jednokrotnie próbki, a z niej losuje się wielokrotnie, z powtórzeniami odpowiedniego (i jednakowego) rozmiaru próbki zwane „próbkami bootstrapowymi". Dla każdej z takich próbek obliczana jest wartość szukanej statystyki. Dzięki temu uzyskuje się aproksymację rozkładów owej statystyki próbkowej. Większość wylosowanych w ten sposób próbek bootstrapowych dość wiernie odzwierciedla w swej strukturze próbę wyjściową (a więc w znacznym stopniu i całą populację). Odrzucając zatem niewielką część potencjalnych obserwacji odstających (czyli wartości skrajnych, ekstremalnych), w tym rozkładzie możemy uzyskać stosowne przedziały ufności dla szukanej statystyki. K. Singh i M. Xie (2008) zaproponowali odrzucenie 5%, czyli po 2,5% krańcowych obserwacji z każdej strony.

Warto przy tym zauważyć, że metoda bootstrapowa nie wymaga stosowania żadnych specyficznych założeń odnośnie rozkładów (np. rozkładów błędów), a nawet informacji na temat sposobu doboru „populacji zastępczej", wykorzystuje się bowiem centralne twierdzenie graniczne. Jego dwa główne założenia są spełnione, gdyż próbki bootstrapowe losuje się niezależnie, zaś wewnętrzny rozkład w każdej próbce odzwierciedla rozkład próby wyjściowej. Metoda ta może być bardziej efektywna w sytuacjach, gdy rozkład obserwacji w próbie jest nieznany lub nieregularny (a więc trudny do identyfikacji, nawet asymptotycznie) bądź też rozmiar wylosowanej próby okazuje się stosunkowo niewielki.

Oprócz tego metoda bootstrapowa daje się łatwo zastosować do złożonych planów badań reprezentacyjnych (obejmujących np. losowanie warstwowe z grupowaniem czy wielostopniowe). J. Fox (2008) oprócz tych spostrzeżeń zauważa także, że bootstrapping może być użyty do estymacji wartości statystyki nieliniowej, dla której nie ma możliwości wyznaczenia błędu standardowego lub dla której dostępne są tylko asymptotyczne informacje o błędach standardowych. Metodę bootstrapową wybrano właśnie ze względu na te walory. Dodatkowo łatwość obliczeniowa i programistyczna czyni ją lepszą od pokrewnego podejścia typu *jackknife* czy metody grup losowych bądź półprób zrównoważonych (Valliant i in., 2000).

Od strony formalnej metodę można opisać następująco. Niech $\theta \in \mathbb{R}$ będzie parametrem dla badanej populacji, który pragniemy oszacować. Załóżmy, że celem jego oszacowania z owej populacji wylosowano próbę rozmiaru $n \in \mathbb{N}$, czyli (X_1, X_2, …, X_n), którą nazywać będziemy „próbą wyjściową" oraz że na podstawie tej próby uzyskano oszacowanie $\hat{\theta} \in \mathbb{R}$ parametru θ. Z centralnego twierdzenia granicznego wiemy, że dla dużych prób[1] próbkowe wartości statystyki $\hat{\theta}$ rozłożone są asymptotycznie w otoczeniu punktu θ (czyli scentrowane wokół θ), z wyznaczającym typowe granice rozrzutu odchyleniem standardowym $s/\sqrt{n}$, gdzie s jest liczbą dodatnią (parametrem), której wartość zależy od własności populacji i rodzaju statystyki $\hat{\theta}$. K. Singh i M. Xie (2008) zauważyli, że bootstrapping szczególnie dobrze sprawdza się w przypadku szacowania odchylenia standardowego, gdzie tradycyjne postępowanie stwarzałoby najwięcej trudności technicznych.

Niech $m \in \mathbb{N}$ będzie liczbą losowanych prób bootstrapowych, zaś $n_\ell \in \mathbb{N}$, $n_\ell \leq n$ — liczebnością każdej próbki bootstrapowej. Niech $\hat{\theta}_b \in \mathbb{R}$ oznacza oszacowanie parametru θ w b-tej próbce bootstrapowej wylosowanej z próby (X_1, X_2, …, X_n). Otóż istnieje tzw. bootstrapowe centralne twierdzenie graniczne (Bickel, Freedman, 1981; Singh, 1981), które orzeka, że dla dużych prób (a konkretnie, gdy $n \to \infty$) rozkład oszacowań $\hat{\theta}_b$ jest scentrowany względem $\hat{\theta}$ i z, tym samym co dla prób losowanych z populacji, odchyleniem standardowym $s/\sqrt{n}$. Stąd rozkład różnic $\hat{\theta}_b - \hat{\theta}$ bardzo dobrze aproksymuje rozkład różnic $\hat{\theta} - \hat{\theta}$. K. Singh i M. Xie (2008) zauważyli, że jeśli w rozkładzie próbkowym nie bierzemy pod uwagę nieznanych parametrów populacyjnych, bootstrapping jest lepszą metodą estymacji rozkładu szacowanego parametru aniżeli narzędzia oparte na klasycznym centralnym twierdzeniu granicznym. Ma to miejsce w przypadku, gdy badamy rozkład statystyki $\left(\hat{\theta}_b - \hat{\theta}\right)/SE$, gdzie SE to prawdziwa wartość błędu standardowego dla $\hat{\theta}$. Jest on wówczas zazwyczaj asympto-

[1] W statystyce matematycznej termin „próba rozmiaru n" oznacza n niezależnych zmiennych losowych o jednakowym rozkładzie. Singh i Xie (2008) za „duże próby" uznali takie, których rozmiar jest nie mniejszy od 30 (tzn. $n \geq 30$).

tycznie normalny (i to w formie standaryzowanej). To zjawisko nosi nazwę „korekcji drugiego rzędu dla bootstrapu". Według K. Singha i M. Xie (2008), jeśli θ jest średnią z populacji, $\hat{\theta} = \overline{X}$ — średnią z próby, σ — odchyleniem standardowym z populacji (stąd $SE = \sigma / \sqrt{n}$), s — odchyleniem standardowym z próby wyjściowej (a zatem estymator błędu standardowego ma postać $\widehat{SE} = s / \sqrt{n}$), zaś $\hat{\theta}_b = \overline{X}_b$ i s_b — odpowiednio średnią i odchyleniem standardowym z *b*-tej próby bootstrapowej (skąd $\widehat{SE}_b = s_b / \sqrt{n_\ell}$), to rozkład próbkowy statystyki $(\overline{X} - \theta)/SE$ może być efektywnie aproksymowany rozkładem bootstrapowym wartości statystyki $(\overline{X}_b - \overline{X})/\widehat{SE}$, a z kolei rozkład statystyki $(\overline{X} - \theta)/\widehat{SE}$ może być skutecznie aproksymowany rozkładem bootstrapowym wartości statystyki $(\overline{X}_b - \overline{X})/\widehat{SE}_b$, b=1, 2, …, m.

Na tej podstawie konstruuje się estymatory bootstrapowe parametrów. W najbardziej klasycznym przypadku średnia arytmetyczna wartości szukanego parametru θ wyestymowanych z kolejnych prób bootstrapowych (czyli $\sum_{b=1}^{m} \hat{\theta}_b / m$) jest bootstrapowym estymatorem parametru θ. Obciążenie estymatora szacuje się jako $Bias_\ell\left(\hat{\theta}\right) \overset{\text{def}}{=} \left(\sum_{b=1}^{m} \hat{\theta}_b / m \right) - \hat{\theta}$. Jest to bootstrapowe oszacowanie obciążenia estymacji parametru θ. Na tej podstawie czasami koryguje się wyjściowy estymator o obciążenie bootstrapowe — $\hat{\theta}_{cr} \overset{\text{def}}{=} \hat{\theta} - Bias_\ell\left(\hat{\theta}\right)$.

Jak wspomnieliśmy, bootstrapping nie eliminuje obciążenia nielosowego, tkwiącego w obserwowanych wartościach próbek. Dlatego, żeby w jak największym stopniu zredukować tę niedogodność, wskazane jest dokonanie wcześniejszych korekt logicznych, a w razie potrzeby także imputacji. Estymatorami bootstrapowymi odpowiedniej statystyki populacyjnej będą też estymatory błędów, o czym piszę dalej. Skądinąd metoda bootstrapowa ma znacznie szersze zastosowanie. Za jej pomocą można bowiem estymować parametry funkcji regresji, jak również testować hipotezy statystyczne dla parametrów zmiennej losowej. Podejście to znalazło także zastosowanie w estymacji bayesowskiej oraz nieparametrycznej analizie rozkładów obserwacji. Szczegółowy syntetyczny opis takiego zastosowania podali np. Cz. Domański i K. Pruska (2000).

ZASTOSOWANIE METODY BOOTSTRAPOWEJ DO OCENY PRECYZJI SZACUNKÓW W NSP 2011

Przedmiotem analizy były szacunki informacji statystycznych dla populacji na podstawie danych z badania reprezentacyjnego w ramach NSP 2011. W tym przy-

padku chodziło o szacunki dotyczące różnorodnych aspektów dojazdów do pracy z wykorzystaniem tzw. „złotego rekordu" z Analitycznej Bazy Mikrodanych (ABM). Badaniem tym objęto 7906717 osób, czyli 20,3% badanej populacji, którą stanowiła ludność Polski. Analiza ograniczyła się do 1702262 osób deklarujących, że dojeżdżają do pracy (co stanowiło 21,5% próby reprezentacyjnej). Była to wyjściowa próba użyta do estymacji danych dla populacji i losowania próbek bootstrapowych. Analizie precyzji poddano estymację następujących danych dotyczących faktycznego miejsca zamieszkania, tj. liczby dojeżdżających do pracy według:

- lokalizacji miejsca pracy i województw;
- środka transportu do głównego miejsca pracy i województw;
- czasu dojazdu do głównego miejsca pracy, płci i województw;
- częstotliwości dojazdu do głównego miejsca pracy i województw;
- odległości w kilometrach od miejsca zamieszkania do głównego miejsca pracy i województw.

Mieliśmy zatem do czynienia z informacjami wyrażonymi w liczbach bezwzględnych (liczba osób). Estymacji poddano jednak także związane z tym zmienne wskaźnikowe, czyli udziały poszczególnych kategorii w ogólnej wielkości (np. w przypadku dojeżdżających według lokalizacji miejsca pracy badano udziały dojeżdżających do pracy poza gminę zamieszkania i w gminie zamieszkania w liczbie ludności danego województwa dojeżdżającej do pracy ogółem) oraz udziały województw w odpowiednich wielkościach ogółem (np. udział liczby dojeżdżających do pracy pociągiem w woj. mazowieckim w liczbie dojeżdżających tym środkiem lokomocji ogółem).

Warto nadmienić, że określenie „lokalizacja miejsca pracy" oznacza ustalenie, czy dana osoba dojeżdżała do pracy w obrębie gminy, w której zamieszkuje czy też jej miejsce pracy było poza tą gminą. W przypadku odległości od miejsca zamieszkania do miejsca pracy wprowadzono przedziały: 0—5 km, 6—20, 21—50 oraz 51 km i więcej.

Od strony formalnej estymacja wygląda następująco: załóżmy, że dotyczy ona wartości zmiennej Y wyrażonych w liczbach bezwzględnych. Niech $y_1, y_2, \ldots, y_n$ oznaczają jej obserwacje dla wyjściowej próby $\not{s}$, p_b — b-tą próbę bootstrapową (b=1, 2, 3, …, m), zaś w_i — wagę kalibracyjną dla i-tej jednostki. Załóżmy, że populacja jest podzielona na h warstw $H_1, H_2, \ldots, H_h$ oraz na k domen (w naszym przypadku — województw) $G_1, G_2, \ldots, G_k$ (h i k są liczbami naturalnymi nie większymi niż rozmiar populacji). Wtedy estymator wartości zmiennej Y dla populacji w j-tej domenie na podstawie wyjściowej próby ma postać bezpośrednią według formuły Horvitza-Thompsona (Bracha, 1996):

$$\hat{\theta}_j = \sum_{i \in G_j \cap \not{s}} w_i y_i = \sum_{l=1}^{h} \sum_{i \in G_j \cap H_l \cap \not{s}} w_i y_i \; j = 1, 2, ..., k \qquad (1)$$

gdzie w_i — waga kalibracyjna dla i-tej jednostki, $i = 1, 2, ..., n$.

Jeśli natomiast estymujemy udział domeny G_j w wartości ogółem zmiennej Y dla populacji, to estymator ma postać:

$$\hat{\theta}_j = \sum_{i \in G_j \cap \not{p}} w_i \frac{y_i}{\sum_{i=1}^{n} w_i y_i} = \sum_{l=1}^{h} \sum_{i \in G_j \cap H_l \cap \not{p}} w_i \frac{y_i}{\sum_{i=1}^{n} w_i y_i} \tag{2}$$

Dla próbek bootstrapowych estymację przeprowadzono wykorzystując również wzory (1) i (2), z tym że wagi poddano korekcie uwzględniającej alokację próby w poszczególnych warstwach:

$$w_i^* = w_i \cdot \frac{n_{H_i}}{n_{H_i b}} \cdot m_{ib} \tag{3}$$

gdzie:

n_{H_i} — liczba rekordów w warstwie H_i w próbie podstawowej — warstwie, do której należy rekord i,

$n_{H_i b}$ — liczba rekordów warstwy H_i w b-tej próbie bootstrapowej,

m_{ib} — liczba powtórzeń rekordu i w b-tej próbie bootstrapowej, i=1, 2, …, $n_{\not{p}}$.

Wtedy odpowiednie estymatory bootstrapowe mają postać:

$$\hat{\theta}_{jb} = \sum_{i \in G_j \cap p_b} w_i^* y_i \quad \text{oraz} \quad \hat{\theta}_{jb} = \sum_{i \in G_j \cap p_b} w_i^* \frac{y_i}{\sum_{i=1}^{n_{\not{p}}} w_i^* y_i}$$

$j = 1, 2, \ldots, k, \qquad b = 1, 2, \ldots, m.$

Wagi kalibracyjne konstruuje się w celu umożliwienia efektywnej estymacji dla populacji uwzględniającej i redukującej negatywny wpływ braku odpowiedzi. Ten efekt uzyskuje się poprzez skorygowanie wag wynikających ze schematu losowania próby wyjściowej z wykorzystaniem zmiennych pomocniczych, tak aby spełnione były odpowiednie równania kalibracyjne wyrażające postulat, by suma ważonych obserwacji dla każdej z owych zmiennych równała się wartości tejże zmiennej dla całej populacji. Dokładną charakterystykę techniki kalibracji opisali T. Józefowski i M. Szymkowiak (2012). W naszym eksperymencie wykorzystano wagi kalibracyjne obliczone przez członków podgrupy roboczej do spraw metod statystyczno-matematycznych na rzecz spisów w ramach projektu NSP 2011. Zmienną pomocniczą była tutaj liczba ludności lub liczba mieszkań (Szymkowiak, 2014).

Wszystkie obliczenia przeprowadzono za pomocą oryginalnego algorytmu napisanego w środowisku SAS Enterprise Guide 4.3 (język SAS 4GL).

Do właściwego zaprojektowania analizy bootstrapowej konieczne jest sprecyzowanie trzech najistotniejszych założeń badawczych:

- przyjęcie sposobu losowania,
- określenie rozmiaru próby bootstrapowej,
- ustalenie liczby losowanych próbek bootstrapowych.

Tym, co w znacznej mierze decyduje o efektywności metody bootstrapowej jest wybór metody losowania próbek[2]. W rozpatrywanym przypadku — z uwagi na przyjęte założenia odnośnie tego rodzaju metody oraz stosowalność centralnego twierdzenia granicznego — najpowszechniejszym rozwiązaniem wydaje się być losowanie niezależne. Innymi słowy, próbki bootstrapowe losujemy w sposób nieograniczony, z powtórzeniami i z jednakowym prawdopodobieństwem wyboru (Fox, 2008). W celu zapewnienia odpowiedniej jakości szacunków należy uwzględnić także podział populacji na warstwy (jeśli takowy istnieje). W naszym przypadku warstwowanie wprowadziła metodologia spisowa. Ostatecznie więc zastosowano losowanie nieograniczone warstwowe z jednakowym prawdopodobieństwem wyboru i powtórzeniami oraz z alokacją próby proporcjonalną do liczebności poszczególnych warstw stosowanych podczas spisu.

Zastanawiając się nad kwestią rozmiaru próbki bootstrapowej należy wspomnieć, że wielu autorów, np. wspomniany J. Fox (2008), jako naturalną wielkość wybiera rozmiar próby wyjściowej (tzn. przyjmuje, że $n_\ell = n$). Taki wybór nie jest jednak efektywny z punktu widzenia estymacji błędów, co zauważył już twórca metody bootstrapowej B. Efron (1983), a potwierdził J. Shao (1996). Prawdopodobieństwo wyboru prawidłowego modelu może bowiem nie dążyć do jedności wraz ze wzrostem liczby n. R. R. Wilcox (2012), na podstawie wyników uzyskanych przez J. Shao (1996), proponuje zastosować wielkość $n_\ell = 5\log(n)$ (a dokładniej — $n_\ell = [5\log(n)]$, gdzie $[a]$ oznacza część całkowitą liczby rzeczywistej a, czyli największą liczbę całkowitą nie większą od a). Dostrzega on jednakże użyteczność wprowadzonego przez B. Efrona (1983) tzw. estymatora 0,632. Chodzi o to, że najefektywniejszym oszacowaniem błędu predykcji z wykorzystaniem bootstrappingu jest estymator:

$$\hat{\varepsilon} = 0{,}368\hat{\varepsilon}_0 + 0{,}632\hat{\varepsilon}_\ell$$

gdzie $\hat{\varepsilon}_0$ — tzw. „błąd oczywisty" (*apparent error*), czyli klasyczny średni kwadrat błędu predykcji dany wzorem:

[2] Valliant i in. (2000) wskazali nawet, że przy tych samych wartościach oszacowania dla różnych estymatorów to plan próbkowania decyduje o jakości estymacji.

$$\hat{\varepsilon}_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

gdzie:

y_i — wartość rzeczywista cechy Y dla i-tej jednostki,
$\hat{y}_i$ — predykcja tej wartości z wyjściowej próby, i=1, 2, …, n dokonana za pomocą odpowiedniej funkcji regresji,
$\hat{\varepsilon}_{\ell}$ — tzw. „oczyszczony estymator błędu bootstrapowego" (*out-of-bag estimator*) dany wzorem:

$$\hat{\varepsilon}_{\ell} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{b\in C_i}(y_i - \hat{y}_{i(b)})^2$$

gdzie:

$\hat{y}_{i(b)}$ — predykcja i-tej obserwacji uzyskanej z b-tej próby bootstrapowej p_b,
C_i — indeksy tych prób bootstrapowych, które nie zawierają obserwacji i, czyli $C_i = \{b: b=1, 2, \ldots, m, i \in p_b\}$, $i=1, 2, \ldots, n$, $b=1, 2, \ldots, m$.

Wynik ten świadczy o tym, że — przeciętnie rzecz biorąc — bootstrapping z efektywnością 63,2% może wyjaśnić błąd estymacji danych statystycznych. Biorąc to pod uwagę J. Khan i in. (2010) zaproponowali jako efektywny rozmiar próby bootstrapowej przyjąć przeciętnie $n_{\ell} = [0{,}632n] = 1074466$. I tak też uczyniono w tym eksperymencie.

Inny problem to liczba prób bootstrapowych. Naturalnym rozwiązaniem są wszystkie możliwe wariacje n_{ℓ} elementów ze zbioru n-elementowego[3], a zatem przyjęcie $m = n^{n_{\ell}}$. I ten właśnie punkt wyjścia obrał J. Fox (2008). Jednak nietrudno zauważyć, że w przypadkach znacznego rozmiaru próby (a z taką w tym przypadku mieliśmy do czynienia) realizacja takiego losowania może być bardzo trudna lub nawet niemożliwa, z uwagi na brak odpowiedniej wydajności obliczeniowej. W przypadku testowania hipotez, opartego na metodzie bootstrapowej, istnieją nawet dość skomplikowane algorytmy do wyznaczenia optymalnej liczby prób (Davidson, MacKinnon, 2000). Jednak K. Singh i M. Xie (2008) zaproponowali wykorzystanie pierwiastka z liczebności wyjściowej próby, a zatem $m = \lceil\sqrt{n}\rceil$. Zgodnie z tym wzorem $m = 1303$. I taką właśnie liczbę prób bootstrapowych wykonano.

[3] Co wówczas *de facto* przestaje być losowaniem, a staje się wyliczaniem.

A zatem podczas eksperymentu losowano próbę bootstrapową, dokonywano estymacji odpowiedniej wartości dla populacji (wykorzystując wspomniane wagi kalibracyjne), a następnie porównywano ten wynik ze stosowną wartością dla próby. Te czynności powtarzano *m* razy.

Oto rodzaje błędów w ocenie parametrów populacji, jakie badano (zachowujemy przy tym wszystkie oznaczenia z części pierwszej opracowania). Analizie poddano błędy szacunków wartości opisanych zmiennych dla populacji. Dla każdego z tych estymatorów parametrów statystycznych wyznaczono:

— względny pierwiastek średniego kwadratu błędu (*RMSE — Relative Root Mean Squared Error*)[4]:

$$\widehat{RMSE}\left(\hat{\theta}_j\right)=\frac{1}{\left|\hat{\theta}_j\right|}\sqrt{\frac{1}{m}\sum_{b=1}^{m}\left(\hat{\theta}_{jb}-\hat{\theta}_j\right)^2}$$

— obciążenie względne (*RB — Relative Bias*):

$$\widehat{RB}\left(\hat{\theta}_j\right)=\frac{\left(\sum_{b=1}^{m}\hat{\theta}_{jb}/m\right)-\hat{\theta}_j}{\hat{\theta}_j}$$

— empiryczne obciążenie względne (*ERB — Empirical Relative Bias*):

$$\widehat{ERB}\left(\hat{\theta}_j\right)=\frac{\widehat{RB}\left(\hat{\theta}_j\right)}{\widehat{RMSE}\left(\hat{\theta}_j\right)}$$

$j=1, 2, \ldots, k$.

RMSE wyraża względny błąd uwzględniający wszystkie możliwe cząstkowe odchylenia od wartości szacowanego parametru dla próby wyjściowej w poszczególnych próbach bootstrapowych. Dlatego też jest to najsilniejszy i najefektywniejszy wskaźnik jakościowy. Wartości obciążenia względnego (*RB*) po pomnożeniu przez 100 można wyrażać w procentach. Ze względu na zastosowaną formułę, *RB* może przyjmować wartości ujemne. Ponadto cząstkowe odchylenia poszczególnych prób o przeciwnych znakach mogą się w tej formule wzajemnie redukować. Empiryczne obciążenie względne uwidacznia relacje pomiędzy *RB* i *RMSE*. Wskaźnik *ERB* może pokazywać, czy niskie obciążenie

[4] *Mean Squared Error* często tłumaczy się jako „błąd średniokwadratowy" lub „średni błąd kwadratowy", co jest kwestionowane przez wielu badaczy. Argumentują oni, że nie ma czegoś takiego, jak „błąd kwadratowy", tak samo jak nie istnieje błąd sześcienny, kulisty, stożkowy itp. Poza tym przymiotnik *squared* oznacza raczej „podniesiony do kwadratu" niż „kwadratowy".

kompensuje odchylenie wynikające z *RMSE* czy też jest na odwrót. Optymalne byłyby tutaj wartości *ERB* bliskie jedności (lub –1), przy równoczesnym niskim poziomie *RB* i *RMSE*. Wtedy mamy pewność, że autoredukcja w *RB* jest także minimalna. Z kolei przy takich wartościach *ERB*, ale przy wysokim *RMSE* i *RB*, znak empirycznego obciążenia względnego wskazuje, czy odchylenie wynika z oczekiwanego niedoszacowania czy też z przeszacowania estymowanych wartości.

REZULTATY PRZEPROWADZONYCH OBLICZEŃ

Przeprowadzone obliczenia pozwalają na dokonanie wszechstronnych analiz jakości oszacowań. Analizy owe muszą jednak także dotyczyć czynników nieuwzględnionych przez bootstrap. Choć należy pamiętać, że są to jedynie badania symulacyjne, to jednak zastosowana technika pokazuje obraz sytuacji bardzo zbliżony do rzeczywistości, szczególnie gdy inne obciążenia są mało istotne. Omówmy rozkład wysokości poszczególnych rodzajów błędów[5] według województw w kontekście podstawowych cech ich rozkładów (średniej arytmetycznej, wartości maksymalnych i minimalnych, mediany, trzeciego kwartyla oraz współczynnika zmienności). Przedstawmy kształtowanie się tych wielkości dla oszacowań ogółem, odnosząc się również do rezultatów uzyskanych według płci. Przyjrzymy się lokalizacji miejsca pracy (z uwagi na incydentalność przypadków odpowiedzi „nieustalony" pomijamy je). Wskaźniki użyte po symbolu błędu oznaczają:

- *liczb.* — bezwzględną liczbę dojeżdżających do pracy w danym układzie;
- *kat.* — estymację odsetka dojeżdżających do pracy według kategorii;
- *woj.* — odsetek dojeżdżających do pracy według kategorii określonej daną zmienną według województw.

Wyjściowy obraz sytuacji pokazuje tabl. 1.

TABL. 1. DOJEŻDŻAJĄCY DO PRACY WEDŁUG LOKALIZACJI MIEJSCA PRACY (faktyczne miejsce zamieszkania)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
			W gminie			
RMSE liczb.	0,006	0,003	0,006	0,007	0,009	26,890
RMSE kat.	0,005	0,003	0,005	0,005	0,007	26,494
RMSE woj.	0,006	0,003	0,006	0,007	0,009	28,195
RB liczb.	0,000	–0,000	0,000	0,000	0,000	156,943
RB kat.	0,000	–0,000	0,000	0,000	0,000	154,804
RB woj.	0,000	–0,000	0,000	0,000	0,000	645,272
ERB liczb.	0,014	–0,032	0,013	0,034	0,046	165,684
ERB kat.	0,017	–0,025	0,017	0,030	0,064	138,085
ERB woj.	0,001	–0,049	–0,000	0,021	0,032	1625,845

[5] Szczegółowe tablice zawierające oszacowania błędów według województw i płci dostępne są na życzenie czytelników u autora niniejszego artykułu.

TABL. 1. DOJEŻDŻAJĄCY DO PRACY WEDŁUG LOKALIZACJI MIEJSCA PRACY (faktyczne miejsce zamieszkania) (dok.)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
Poza gminą						
RMSE liczb.	0,007	0,004	0,007	0,009	0,010	23,314
RMSE kat.	0,005	0,003	0,005	0,006	0,008	27,092
RMSE woj.	0,007	0,004	0,007	0,009	0,010	24,828
RB liczb.	–0,000	–0,000	–0,000	0,000	0,000	–349,195
RB kat.	–0,000	–0,000	–0,000	–0,000	0,000	–153,261
RB woj.	0,000	–0,000	0,000	0,000	0,000	747,606
ERB liczb.	–0,009	–0,048	–0,010	0,013	0,032	–277,113
ERB kat.	–0,017	–0,064	–0,017	–0,005	0,025	–138,085
ERB woj.	0,002	–0,039	0,001	0,023	0,043	1116,276

Ź r ó d ł o: opracowanie własne na podstawie danych z ABM przy użyciu algorytmu napisanego w SAS Enterprise Guide 4.3.

Widać, że szacowane błędy są niezbyt duże, nie przekraczają 1%. Największe błędy *RMSE* oraz obciążenia względne (*RB*) występują dla liczb bezwzględnych. Lepiej prezentuje się jakość szacunków odsetek, szczególnie w przypadku struktury według kategorii. Warto zauważyć, że *RMSE* przewyższa *RB*, stąd empiryczne obciążenie względne jest (co do modułu) niskie. W ekstremalnych przypadkach przekracza ono 4%, co oznacza, że źródło błędów tkwi przede wszystkim w wariancji estymatora. Potwierdza to analiza według płci. Na przykład w przypadku kobiet maksymalne wartości *RMSE* dla liczb bezwzględnych osiągają 0,014 — dotyczy dojeżdżających do pracy poza gminę faktycznego zamieszkania w woj. podlaskim; podobnie dla mężczyzn — tyle, że dotyczy to dojeżdżających do pracy w gminach zamieszkania na Opolszczyźnie. Tabl. 2 uwidacznia symulację jakości szacunków według środka transportu dojazdu do głównego miejsca pracy.

TABL. 2. DOJEŻDŻAJĄCY DO PRACY WEDŁUG ŚRODKA TRANSPORTU DOJAZDU DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
Komunikacja publiczna (w tym miejska)						
RMSE liczb.	0,011	0,005	0,011	0,014	0,018	31,297
RMSE kat.	0,010	0,004	0,009	0,013	0,018	34,757
RMSE woj.	0,011	0,005	0,010	0,014	0,018	33,085
RB liczb.	–0,000	–0,001	–0,000	0,000	0,001	–378,460
RB kat.	–0,000	–0,001	–0,000	0,000	0,001	–275,984
RB woj.	0,000	–0,000	–0,000	0,000	0,001	22022,048
ERB liczb.	–0,009	–0,054	–0,018	0,007	0,058	–316,037
ERB kat.	–0,012	–0,051	–0,017	0,010	0,057	–242,928
ERB woj.	0,000	–0,042	–0,004	0,020	0,064	9546,631

TABL. 2. DOJEŻDŻAJĄCY DO PRACY WEDŁUG ŚRODKA TRANSPORTU DOJAZDU DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania) (cd.)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
Pociąg						
RMSE liczb.	0,036	0,012	0,033	0,043	0,071	42,533
RMSE kat.	0,036	0,011	0,033	0,042	0,071	42,655
RMSE woj.	0,035	0,009	0,033	0,042	0,071	44,476
RB liczb.	0,000	–0,002	0,000	0,001	0,002	575,706
RB kat.	0,000	–0,002	0,000	0,001	0,002	608,604
RB woj.	0,000	–0,002	0,000	0,001	0,002	376,322
ERB liczb.	0,002	–0,059	0,004	0,032	0,065	1494,693
ERB kat.	0,002	–0,052	0,001	0,029	0,063	1548,992
ERB woj.	0,005	–0,058	0,007	0,034	0,070	630,677
Prywatny przewoźnik						
RMSE liczb.	0,024	0,010	0,021	0,028	0,056	48,057
RMSE kat.	0,023	0,010	0,020	0,028	0,055	48,340
RMSE woj.	0,023	0,009	0,020	0,027	0,055	50,041
RB liczb.	–0,000	–0,002	0,000	0,000	0,001	–388,638
RB kat.	–0,000	–0,002	0,000	0,000	0,001	–380,421
RB woj.	–0,000	–0,002	–0,000	0,000	0,001	–283,039
ERB liczb.	–0,002	–0,065	0,001	0,020	0,086	–2294,733
ERB kat.	–0,001	–0,070	0,004	0,021	0,076	–3440,356
ERB woj.	–0,006	0,069	–0,005	0,016	0,084	–648,899
Samochód osobowy Jako kierowca						
RMSE liczb.	0,006	0,004	0,006	0,007	0,008	23,340
RMSE kat.	0,004	0,002	0,004	0,005	0,006	21,496
RMSE woj.	0,006	0,003	0,006	0,007	0,008	24,474
RB liczb.	0,000	–0,000	–0,000	0,000	0,000	291,184
RB kat.	0,000	–0,000	0,000	0,000	0,000	597,999
RB woj.	0,000	–0,000	–0,000	0,000	0,000	414,900
ERB liczb.	0,005	–0,020	–0,002	0,023	0,034	385,170
ERB kat.	0,005	–0,043	0,009	0,027	0,049	580,428
ERB woj.	0,003	–0,024	–0,004	0,021	0,033	731,799
Samochód osobowy Jako pasażer						
RMSE liczb.	0,015	0,010	0,015	0,016	0,021	20,483
RMSE kat.	0,015	0,010	0,015	0,015	0,020	20,021
RMSE woj.	0,015	0,010	0,015	0,016	0,021	21,877
RB liczb.	0,000	–0,001	0,000	0,000	0,001	268,919
RB kat.	0,000	–0,001	0,000	0,000	0,001	272,553
RB woj.	–0,000	–0,001	–0,000	0,000	0,001	–4361,029
ERB liczb.	0,013	–0,061	0,012	0,027	0,055	215,848
ERB kat.	0,012	–0,062	0,015	0,029	0,047	212,188
ERB woj.	–0,000	–0,072	–0,001	0,015	0,042	–6233,977

TABL. 2. DOJEŻDŻAJĄCY DO PRACY WEDŁUG ŚRODKA TRANSPORTU DOJAZDU DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania) (dok.)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
			Inny			
RMSE liczb.	0,020	0,012	0,020	0,022	0,033	27,127
RMSE kat.	0,019	0,012	0,019	0,022	0,032	27,309
RMSE woj.	0,019	0,011	0,019	0,022	0,032	28,411
RB liczb.	0,000	–0,001	–0,000	0,000	0,001	1536,171
RB kat.	0,000	–0,001	–0,000	0,000	0,001	3541,847
RB woj.	0,000	–0,001	–0,000	0,000	0,001	514,677
ERB liczb.	–0,001	–0,045	–0,005	0,016	0,049	–2428,464
ERB kat.	–0,002	–0,045	–0,004	0,012	0,046	–1456,567
ERB woj.	0,002	–0,042	–0,001	0,018	0,054	997,857

Ź r ó d ł o: jak przy tabl. 1.

Analiza tych danych prowadzi do podobnych wniosków, jak w przypadku tabl. 1. Najlepsze wyniki *RMSE* szacunku liczb bezwzględnych osiągnięte były dla kategorii „samochód osobowy — jako kierowca". Estymacje bywają tu bardzo precyzyjne (np. dla kobiet w woj. mazowieckim — 0,006), ale zdarzają się też wartości sięgające 7% (np. dla mężczyzn dojeżdżających pociągiem w woj. świętokrzyskim — 0,0682). Takie sytuacje występują jednak sporadycznie. Generalnie zróżnicowanie błędów wydaje się jednak większe niż w poprzednim przypadku (nawet jeśli wziąć pod uwagę sztuczne zawyżanie zmienności przez wartości ujemne niektórych współczynników). Najniższa zmienność szacunków *RMSE* wystąpiła dla kategorii „samochód osobowy — jako pasażer", choć były one nieco większe niż w przypadku najliczniej reprezentowanych kategorii, takich jak dojazdy „komunikacją publiczną (w tym miejską)". Warto także zauważyć, że w niektórych przypadkach także obciążenie odgrywało pewną rolę (*ERB* ok. 4—5%). Wybór środka lokomocji może mieć wpływ na czas dojazdu do pracy. Ten aspekt przedstawiono w tabl. 3.

TABL. 3. DOJEŻDŻAJĄCY DO PRACY WEDŁUG CZASU DOJAZDU DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
			Do 30 minut			
RMSE liczb.	0,005	0,003	0,005	0,006	0,007	22,363
RMSE kat.	0,003	0,002	0,003	0,003	0,003	17,085
RMSE woj.	0,005	0,003	0,005	0,006	0,007	23,094
RB liczb.	–0,000	–0,000	–0,000	0,000	0,000	–650,609
RB kat.	–0,000	–0,000	–0,000	0,000	0,000	–160,841
RB woj.	0,000	–0,000	0,000	0,000	0,000	526,030
ERB liczb.	–0,006	–0,061	–0,008	0,009	0,043	–427,933
ERB kat.	–0,015	–0,061	–0,019	0,005	0,019	–159,379
ERB woj.	0,003	–0,054	0,001	0,017	0,054	906,029

TABL. 3. DOJEŻDŻAJĄCY DO PRACY WEDŁUG CZASU DOJAZDU DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania) (dok.)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
Od 31 minut do 1 godziny						
RMSE liczb.	0,011	0,005	0,010	0,014	0,017	30,902
RMSE kat.	0,010	0,004	0,009	0,013	0,016	33,247
RMSE woj.	0,011	0,005	0,010	0,014	0,017	32,794
RB liczb.	0,000	–0,000	0,000	0,000	0,001	192,679
RB kat.	0,000	–0,000	0,000	0,000	0,001	223,398
RB woj.	0,000	–0,000	0,000	0,000	0,001	620,724
ERB liczb.	0,013	–0,027	0,012	0,028	0,059	168,395
ERB kat.	0,014	–0,023	0,010	0,040	0,054	193,898
ERB woj.	0,002	–0,038	0,003	0,016	0,053	981,441
Powyżej 1 godziny do 2 godzin						
RMSE liczb.	0,023	0,007	0,022	0,031	0,041	38,523
RMSE kat.	0,023	0,007	0,022	0,031	0,041	39,345
RMSE woj.	0,023	0,006	0,022	0,031	0,041	40,676
RB liczb.	0,000	–0,001	0,000	0,001	0,001	202,189
RB kat.	0,000	–0,001	0,000	0,001	0,001	203,404
RB woj.	0,000	–0,001	0,000	0,000	0,001	357,949
ERB liczb.	0,011	–0,043	0,012	0,024	0,053	229,651
ERB kat.	0,011	–0,029	0,012	0,025	0,051	210,432
ERB woj.	0,004	–0,054	0,007	0,019	0,049	615,336
Powyżej 2 godzin						
RMSE liczb.	0,036	0,017	0,035	0,039	0,056	28,605
RMSE kat.	0,035	0,017	0,035	0,039	0,056	28,694
RMSE woj.	0,035	0,016	0,034	0,039	0,055	29,851
RB liczb.	–0,000	–0,003	–0,000	0,000	0,001	–357,178
RB kat.	–0,000	–0,003	–0,000	0,000	0,001	–322,488
RB woj.	–0,000	–0,002	0,000	0,000	0,001	–1122,098
ERB liczb.	–0,008	–0,073	–0,002	0,007	0,033	–384,189
ERB kat.	–0,008	–0,072	–0,002	0,009	0,029	–359,909
ERB woj.	–0,002	–0,068	0,003	0,018	0,039	–1835,226

Ź r ó d ł o: jak przy tabl. 1.

Uwagę przyciągają stosunkowo duże wartości błędów *RMSE* dla liczb bezwzględnych i odsetków w przypadku kategorii „powyżej 2 godzin", które w skrajnych przypadkach przekraczają 5,5%. Podobne obserwacje można poczynić w układzie według płci. Na przykład dla kobiet w tej kategorii w woj. lubuskim *RMSE* liczb bezwzględnych przekracza 11%, zaś dla mężczyzn w woj. lubelskim w przypadku czasu dojazdu wynoszącego od 31 minut od 1 godziny — nawet 14%. Wartości bezwzględne współczynnika *ERB* wskazują niezbyt duże, ale niemożliwe do pominięcia znaczenie obciążenia (np. dla mężczyzn „powyżej 2 godzin" na Podkarpaciu wynosi on nawet, co do modułu, ponad 7%, a dla odsetków — ponad 8%).

Jakość oszacowań częstotliwości dojazdów ilustruje tabl. 4.

TABL. 4. DOJEŻDŻAJĄCY DO PRACY WEDŁUG CZĘSTOTLIWOŚCI DOJAZDU DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
Codziennie						
RMSE liczb.	0,005	0,003	0,005	0,006	0,007	25,628
RMSE kat.	0,002	0,001	0,002	0,003	0,003	25,811
RMSE woj.	0,005	0,003	0,005	0,006	0,007	26,627
RB liczb.	0,000	–0,000	0,000	0,000	0,000	375,627
RB kat.	0,000	–0,000	0,000	0,000	0,000	529,855
RB woj.	0,000	–0,000	0,000	0,000	0,000	360,139
ERB liczb.	0,004	–0,038	0,005	0,022	0,041	637,912
ERB kat.	0,003	–0,040	–0,000	0,029	0,042	777,160
ERB woj.	0,004	–0,039	0,005	0,022	0,041	612,741
Częściej niż raz w tygodniu, ale nie codziennie						
RMSE liczb.	0,012	0,007	0,011	0,013	0,016	21,361
RMSE kat.	0,011	0,007	0,011	0,012	0,014	21,279
RMSE woj.	0,011	0,006	0,011	0,013	0,015	23,231
RB liczb.	–0,000	–0,001	0,000	0,000	0,000	–572,572
RB kat.	–0,000	–0,001	–0,000	0,000	0,000	–378,444
RB woj.	–0,000	–0,000	0,000	0,000	0,000	–1890,576
ERB liczb.	–0,004	–0,051	0,003	0,020	0,032	–785,928
ERB kat.	–0,005	–0,044	0,000	0,016	0,041	–548,538
ERB woj.	–0,000	–0,050	0,007	0,025	0,036	–17262,107
Raz w tygodniu						
RMSE liczb.	0,036	0,020	0,036	0,041	0,053	24,148
RMSE kat.	0,036	0,020	0,036	0,041	0,053	24,183
RMSE woj.	0,035	0,018	0,036	0,041	0,053	26,110
RB liczb.	0,000	–0,002	0,000	0,001	0,002	796,350
RB kat.	0,000	–0,002	0,000	0,001	0,002	915,333
RB woj.	–0,000	–0,002	0,000	0,001	0,002	–12812,383
ERB liczb.	0,004	–0,048	0,013	0,024	0,056	665,957
ERB kat.	0,004	–0,050	0,012	0,025	0,053	671,928
ERB woj.	–0,000	–0,051	0,010	0,019	0,053	–13004,900
Raz na 2 tygodnie						
RMSE liczb.	0,067	0,047	0,069	0,072	0,097	20,085
RMSE kat.	0,067	0,047	0,069	0,072	0,097	20,080
RMSE woj.	0,065	0,043	0,067	0,071	0,096	21,681
RB liczb.	0,000	–0,002	–0,000	0,001	0,003	4448,517
RB kat.	0,000	–0,002	–0,000	0,001	0,003	9504,248
RB woj.	0,000	–0,002	–0,000	0,002	0,003	725,983
ERB liczb.	–0,000	–0,033	–0,008	0,020	0,042	–6261,106
ERB kat.	–0,000	–0,032	–0,006	0,020	0,041	–4839,651
ERB woj.	0,002	–0,033	–0,005	0,023	0,046	1016,333

TABL. 4. DOJEŻDŻAJĄCY DO PRACY WEDŁUG CZĘSTOTLIWOŚCI DOJAZDU DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania) (dok.)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
Raz w miesiącu						
RMSE liczb.	0,089	0,060	0,079	0,096	0,145	27,165
RMSE kat.	0,089	0,060	0,079	0,096	0,145	27,219
RMSE woj.	0,087	0,059	0,077	0,093	0,144	28,556
RB liczb.	0,000	–0,005	0,000	0,002	0,004	3271,298
RB kat.	0,000	–0,005	0,000	0,002	0,004	4069,689
RB woj.	–0,000	–0,005	0,000	0,002	0,003	–3639,129
ERB liczb.	0,001	–0,041	0,002	0,020	0,046	2708,344
ERB kat.	0,001	–0,040	0,002	0,019	0,046	2886,809
ERB woj.	–0,001	–0,046	0,001	0,019	0,046	–4395,761
Rzadziej niż raz w miesiącu						
RMSE liczb.	0,140	0,101	0,132	0,156	0,208	23,715
RMSE kat.	0,140	0,101	0,132	0,156	0,208	23,736
RMSE woj.	0,137	0,093	0,127	0,154	0,208	25,077
RB liczb.	0,000	–0,006	0,000	0,003	0,004	31308,704
RB kat.	–0,000	–0,006	0,000	0,003	0,004	–19695,038
RB woj.	–0,000	–0,006	0,000	0,002	0,004	–1029,610
ERB liczb.	0,001	–0,051	0,001	0,023	0,030	2362,925
ERB kat.	0,001	–0,051	0,001	0,022	0,030	2698,549
ERB woj.	–0,001	–0,053	0,001	0,020	0,026	–1970,364

Ź r ó d ł o: jak przy tabl. 1.

Najliczniej „obsadzone" kategorie („codziennie" i „raz w tygodniu") wykazały najlepszą jakość. Szczególnie duże i niepokojące wartości *RMSE* dla liczb bezwzględnych obserwuje się w przypadku kategorii „rzadziej niż raz w miesiącu" — tutaj błędy sięgają kilkunastu, a czasami nawet 20%. Wynika to z ich bardzo dużych wartości strukturalnych, np. w przypadku kobiet w woj. świętokrzyskim *RMSE* dla tej opcji wyniósł 387,616, a dla mężczyzn na Podlasiu — 120,888. Ciekawostką jest też to, że duże wartości błędów dla najrzadziej dojeżdżających nie pokrywają się, np. w woj. świętokrzyskim u mężczyzn w tej kategorii *RMSE* wyniósł 0,1857. Jest to prawdopodobnie wynikiem sporadycznej niewystarczającej reprezentatywności niektórych próbek w warstwach, szczególnie tych najmniej licznych. Jedynie w przypadku odsetków według województw dla dojeżdżających do pracy „rzadziej niż raz w miesiącu" nie ma ekstremalnych wartości — tu wartość *RMSE* tylko incydentalnie przekracza 0,30 dla kobiet i 0,26 dla mężczyzn. W ostatnim przypadku rzeczony błąd jest na ogół niższy niż dla płci pięknej.

Na częstotliwość dojazdów pewien wpływ ma odległość miejsca pracy od miejsca zamieszkania. Jakość szacunków tego zjawiska przedstawia tabl. 5.

TABL. 5. DOJEŻDŻAJĄCY DO PRACY WEDŁUG ODLEGŁOŚCI OD MIEJSCA ZAMIESZKANIA DO GŁÓWNEGO MIEJSCA PRACY (faktyczne miejsce zamieszkania)

Wskaźniki	Średnia arytmetyczna	Minimum	Mediana	Kwartyl górny	Maksimum	Współczynnik zmienności w %
0—5 km						
RMSE liczb.	0,008	0,005	0,008	0,010	0,011	20,707
RMSE kat.	0,007	0,005	0,007	0,008	0,009	18,411
RMSE woj.	0,008	0,005	0,008	0,010	0,011	22,058
RB liczb.	–0,000	–0,000	–0,000	0,000	0,000	–355,406
RB kat.	–0,000	–0,000	–0,000	–0,000	0,000	–201,251
RB woj.	0,000	–0,000	–0,000	0,000	0,000	724,715
ERB liczb.	–0,009	–0,046	–0,013	0,006	0,043	–283,087
ERB kat.	–0,013	–0,039	–0,018	–0,002	0,033	–176,904
ERB woj.	0,002	–0,036	–0,002	0,018	0,055	1135,594
6—20						
RMSE liczb.	0,007	0,004	0,007	0,008	0,010	26,533
RMSE kat.	0,005	0,002	0,004	0,006	0,007	29,761
RMSE woj.	0,007	0,003	0,006	0,008	0,010	27,836
RB liczb.	0,000	–0,000	0,000	0,000	0,000	465,521
RB kat.	0,000	–0,000	0,000	0,000	0,000	949,106
RB woj.	0,000	–0,000	0,000	0,000	0,000	633,478
ERB liczb.	0,003	–0,037	0,005	0,022	0,037	650,315
ERB kat.	0,003	–0,045	0,006	0,025	0,048	986,121
ERB woj.	0,002	–0,039	0,003	0,021	0,037	1191,127
21—50						
RMSE liczb.	0,011	0,006	0,011	0,014	0,017	27,225
RMSE kat.	0,011	0,006	0,010	0,013	0,017	27,309
RMSE woj.	0,011	0,005	0,011	0,014	0,017	29,043
RB liczb.	0,000	–0,000	0,000	0,000	0,000	120,699
RB kat.	0,000	–0,000	0,000	0,000	0,000	138,426
RB woj.	0,000	–0,000	0,000	0,000	0,000	908,342
ERB liczb.	0,014	–0,031	0,017	0,023	0,053	127,099
ERB kat.	0,015	–0,026	0,019	0,028	0,056	141,433
ERB woj.	0,001	–0,044	0,005	0,012	0,037	2127,565
51 km i więcej						
RMSE liczb.	0,021	0,009	0,021	0,024	0,034	27,678
RMSE kat.	0,020	0,009	0,021	0,023	0,033	27,805
RMSE woj.	0,020	0,008	0,020	0,023	0,033	29,229
RB liczb.	0,000	–0,001	0,000	0,000	0,001	3380,986
RB kat.	0,000	–0,001	0,000	0,000	0,001	71726,830
RB woj.	0,000	–0,001	0,000	0,001	0,001	552,196
ERB liczb.	–0,001	–0,067	0,003	0,018	0,060	–2436,496
ERB kat.	–0,002	–0,054	0,002	0,016	0,057	–1923,761
ERB woj.	0,003	–0,065	0,008	0,023	0,065	1024,429

Ź r ó d ł o: jak przy tabl. 1.

Jakość danych wydaje się być bardzo dobra (szczególnie w kontekście mediany). Nawet dla liczb bezwzględnych wartości *RMSE* i *RB* są na ogół niskie. Również w ujęciu według płci tylko w bardzo sporadycznych sytuacjach dla

najniższych poziomów odległości wartości *RMSE* osiągnęły niespełna 4% dla mężczyzn (tak było np. w woj. podlaskim dla dojeżdżających do pracy na odległość 51 i więcej km) oraz nieco ponad 6% dla kobiet (np. ta sama kategoria na Podlasiu). Ogólnie precyzja dla szacunków w przypadku płci pięknej jest nieco gorsza. Wartości górnego kwartyla potwierdzają, że na zmienność wyników mają wpływ nieliczne obserwacje odstające.

Wnioski

Przeprowadzony eksperyment pokazał użyteczność metody bootstrapowej w ocenie jakości szacunków dotyczących dojazdów do pracy uzyskiwanych z danych badania reprezentacyjnego zrealizowanego w ramach NSP 2011. Dzięki wykorzystywaniu własności nieznanego literalnie rozkładu danej wielkości statystycznej, metoda ta pozwala efektywnie uzyskiwać estymację błędów, zarówno w ujęciu kwadratów odchyleń jak i w kontekście obciążenia. Mimo wynikającej z definicji nieobciążoności estymatora bezpośredniego Horvitza-Thompsona (co skutkuje niskimi wartościami empirycznych odchyleń względnych), niezerowe obciążenia względne w niektórych przypadkach mogą sygnalizować skalę błędów różnych typów w szacunkach odpowiednich wielkości dla populacji. W bardziej szczegółowych kategoriach mogą wystąpić incydentalnie znaczne wartości błędów, co zdaje się sugerować wrażliwość tej metody (a także estymatora bezpośredniego) na niską liczebność próby wyjściowej na niektórych poziomach agregacji i w warstwach (a taka tu wystąpiła). Warto jednakże zauważyć, że w niektórych przypadkach sugeruje ona lepszą jakość szacunków aniżeli klasyczny współczynnik zmienności (*Wybrane…*, 2014).

Można domniemywać, że oprócz odchyleń wynikających z samej specyfiki metody bootstrapowej, na wartości *RMSE* i *RB* ma wpływ fakt, iż użyte wagi kalibracyjne zostały skalibrowane jedynie na poziomie wielkości podstawowej statystyki dla populacji, a nie wyjściowej próby (korekcja (3) tego nie zapewnia). Można byłoby w tym kontekście rozważyć stworzenie mechanizmu autokalibracji wag dla startowej próby na poziomie losowania próbek bootstrapowych. Nie wiadomo wszakże, czy nie prowadziłoby to jednak do powstania innego rodzaju zniekształceń szacunków, wynikających z odmienności wag stosowanych w każdym losowaniu.

Nie sposób pominąć także problemu wpływu błędów nielosowych na jakość oszacowań. Warto wspomnieć, że w trakcie prac po spisie poczyniono w tym zakresie stosowne działania. Mianowicie, trzeba było skorygować zapisy, które były nielogiczne, np. gdy odległość dojazdu do pracy okazała się znaczna, podczas gdy respondent dojeżdżał do pracy w gminie zamieszkania lub gdy deklarowany czas dojazdu do pracy był dalece niewspółmierny do pokonywanej odległości (np. odległość ponad 300 km pokonywano w czasie poniżej 30 minut, co w naszych warunkach jest niemożliwe itp.). Są to typowe błędy nielosowe, które często występują w badaniach tego typu, a wynikają np. z pomyłek respondenta lub rachmistrza.

Na poprawę jakości estymacji mogłoby wpłynąć także wykorzystanie rejestrów administracyjnych. Z jednej strony, opierając się na tych zasobach można z dużą dozą precyzji wskazać główne miejsce pracy większości pracujących. Z drugiej zaś strony, przy ich użyciu dałoby się też np. oszacować skalę odchyleń rzeczywistej oraz wynikającej z zasobów rejestrowych lokalizacji głównego miejsca pracy, a potem wykorzystać to do skorygowania oszacowań. Zabieg ten mógłby także poprawić w przyszłości jakość ewentualnych oszacowań na niższych poziomach agregacji (np. dla powiatów), gdyż obecne wyniki skłaniają do obaw w tym zakresie, a zapotrzebowanie na takie informacje wielokrotnie sygnalizowali potencjalni użytkownicy. Zagadnienia te winny stać się przedmiotem dalszej dyskusji prowadzonej w gronie specjalistów.

dr hab. Andrzej Młodak — *Urząd Statystyczny w Poznaniu*

LITERATURA

Bickel P. J., Freedman D. (1981), *Some Asymptotic Theory for the Bootstrap*, „Annals of Statistics", Vol. 9

Bracha Cz. (1996), *Teoretyczne podstawy metody reprezentacyjnej*, Wydawnictwo Naukowe PWN, Warszawa

Davidson R., MacKinnon J. (2000), *Bootstrap tests: how many bootstraps?*, „Econometric Reviews, Taylor and Francis Journals", Vol. 19(1)

Domański Cz., Pruska K. (2000), *Nieklasyczne metody statystyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa

Efron B. (1983), *Estimating the Error Rate of a Prediction Rule: Improvement on Cross-validation*, „Journal of the American Statistical Association", Vol. 78

Fox J. (2008), *Applied Regression Analysis and Generalized Linear Models*, Second Edition, SAGE Publications, Inc., Thousand Oaks, California, U.S.A.

Józefowski T., Szymkowiak M. (2012), *Estymatory kalibracyjne w badaniach statystycznych*, „Wiadomości Statystyczne", nr 1

Khan J., van Aelst S., Zamar R. (2010), *Fast Robust Estimation of Prediction Error based on Resampling*, „Computational Statistics and Data Analysis", Vol. 54

Post Enumeration Surveys, Operational Guidelines (2010), UN Statistics Division, Technical Report, New York, April

Shao J. (1996), *Resampling Methods in Sample Surveys*, Invited paper, Statistics, Vol. 27, with discussion

Singh K., Xie M. (2008), *Bootstrap: A Statistical Method*, Rutgers University, http://www.stat..rutgers.edu/home/mxie/RCPapers/bootstrap.pdf

Singh K. (1981), *On Asymptotic Accuracy of Efron's Bootstrap*, „Annals of Statistics", Vol. 9

Szymkowiak M. (2014), *Estymatory kalibracyjne w NSP 2011*, „Wiadomości Statystyczne", nr 11

Valliant R., Dorfman A. H., Royall R. M. (2000), *Finite Population Sampling and Inference. A Prediction Approach*, Wiley Series in Probability and Statistics, John Wiley & Sons Inc., New York, Chichester, Weinheim, Brisbane, Singapore, Toronto

Wilcox R. R. (2012), *Introduction to Robust Estimation and Hypothesis Testing*, Third Edition Elsevier Inc., Waltham-San Diego-Oxford-Amsterdam

Wybrane aspekty aktywności ekonomicznej ludności. Narodowy Spis Powszechny Ludności i Mieszkań 2011 (2014), GUS, Warszawa

SUMMARY

This paper is devoted to a presentation of results of simulation experiment aimed at assessment of estimation quality of data on commuting to work on the basis of information collected during sample survey conducted within the National Population and Housing Census 2011. The analysis uses the bootstrap method consisting in multiple sampling with replacement of samples from a given 'starting' sample and studying distribution of values of parameter of interest estimated for them. The generalization was conducted using Horvitz--Thompson direct estimator with calibration weights computed by working subgroup for statistical and mathematical methods for censuses within the 2011 Census project and strata defined for such census. On the basis of suggestions contained in subject matter literature sizes and number of bootstrap samples were established and computation of relative mean squared error, relative bias and empirical relative bias for estimation of various values expressed in absolute numbers and percentage by categories and voivodships for each analyzed sample as total and by sex. Using the results relevant practical conclusions were formulated.

РЕЗЮМЕ

Статья представляет результаты моделированного эксперимента использованного для оценки качества данных (предварительного рассчета данных) о проездах на работу на основе информаций полученных в выборочном обследовании проводимым во время Всеобщей переписи населения и квартир в 2011 г. Для анализа был использован метод Бутстрап заключающийся в многократых случайных выборках с повторением выборок из исходной выборки и в анализе распределения полученных для них значений оцененых параметров. Оценки были проведены с использованием прямой оценки Horvitza-Thompsona с калибровочными весами и определенными слоями в этой же переписи. На основе предложений в литературе были проведены и установлены оптимальные размеры и объем выборок, а также были сделаны расчеты относительного среднего квадрата ошибки, относительной нагрузки и эмпирической относительной нагрузки для оценки различных значений выраженных в абсолютных числах по категориям и воеводствам в общем подходе и по полу. На этой основе были сформулированы практические выводы.