

**Elżbieta Sobczak**

## **STATYSTYKA MAŁYCH OBSZARÓW – ASPEKT METODOLOGICZNY**

### **1. Wstęp**

Zdecydowany wzrost znaczenia statystyki małych obszarów obserwowany w Polsce jest ściśle związany z przeobrażeniami, jakie nastąpiły w ostatnich latach. Należą do nich przede wszystkim:

- transformacja gospodarki centralnie planowanej w gospodarkę rynkową,
- zmiany w systemie funkcjonowania administracji państwowej (wyodrębnienie gmin i samorządów lokalnych) i związany z tym wzrost znaczenia lokalnych jednostek terytorialnych.

Wymienione przeobrażenia są zarazem determinantami wzrostu zapotrzebowania na dane dla małych obszarów. Można wyodrębnić następujące grupy użytkowników zainteresowanych rozwojem statystyki małych obszarów:

- władze centralne dokonujące dystrybucji funduszy państwowych pomiędzy jednostki lokalne,
- władze lokalne wykorzystujące tego typu informacje statystyczne do skutecznego zarządzania małymi obszarami,
- deputowani oceniający prowadzoną politykę lub dokonujący porównań obszaru przez nich reprezentowanego z innymi obszarami,
- przedstawiciele sektora prywatnego analizujący lokalne warunki ekonomiczne, społeczne i środowiskowe [3; 7].

Na świecie funkcjonuje wiele definicji pojęcia „mały obszar”, gdyż w poszczególnych krajach w ten sposób definiuje się różnej wielkości obszary geograficzne lub administracyjne. Wydaje się, że za poprawne można przyjąć sformułowanie L. Nowak, mówiące, że: „małe obszary występują wówczas, gdy bieżąca statystyka nie zapewnia informacji w podstawowych przekrojach terytorialnych kraju, wyniki zaś badań reprezentacyjnych są mało wiarygodne i nie mogą być prezentowane w pożądanym przekrojach” [9]. J. Kordos w swojej

pracy [6] określa statystykę małych obszarów jako „dane statystyczne gromadzone, opracowywane i publikowane dla takich jednostek administracyjnych, jak: gromady, gminy, osiedla miasta i powiaty”, a jednocześnie zaznacza, że zakres małych obszarów nie musi być ograniczony wyłącznie do jednostek administracyjnych [7].

## 2. Źródła danych statystycznych dla małych obszarów

Istnieją trzy podstawowe źródła informacji dla małych obszarów [1; 9]:

- 1) spisy,
- 2) statystyka bieżąca,
- 3) rejestry urzędowe.

Powszechne spisy ludności przeprowadzane okresowo stanowią najstarsze i najobszerniejsze źródło informacji dla małych obszarów. Przeprowadzane na ogół co dziesięć lat, dostarczają w latach spisowych danych statystycznych dla wszystkich jednostek terytorialnych. Powszechne spisy ludności zakresem swym obejmują [9]:

- 1) stan i strukturę ludności według:
  - podstawowych cech demograficznych (płci, wieku, stanu cywilnego),
  - podstawowych cech społeczno-zawodowych (źródła utrzymania, aktywności zawodowej, wykonywanego zawodu itd.),
- 2) liczbę i strukturę gospodarstw domowych i rodzin.

Podstawową zaletą spisów powszechnych jest rejestracja danych dla wszystkich jednostek, nawet tych ujętych w mniejszej skali przestrzennej, oraz umożliwienie analizy nie tylko stanów liczebnych badanych zbiorowości, ale również ich struktur wewnętrznych [11]. Zasadniczymi wadami tego źródła informacji są rzadka częstotliwość przeprowadzania spisów powszechnych, ogromne koszty pozyskiwania informacji i długi czas opracowywania tak wielkiej liczby danych. W większości krajów nie funkcjonują źródła informacji mogące zastąpić spis powszechny i zapewnić rejestrację szerokiego zakresu danych w przekroju najmniejszych jednostek terytorialnych.

W naszym kraju przeprowadza się również inne masowe spisy mające charakter badań pełnych, jednak cechuje je wąski zakres danych. Należą do nich spisy rolne i badania infrastruktury [11].

W latach pospisowych głównym źródłem informacji dla jednostek lokalnych jest statystyka bieżąca. Do 1988 r. zbiory danych dla małych obszarów były dość bogate (dotyczyły ok. 270 grup tematycznych), jednak w następnych latach ograniczono ich gromadzenie i opracowywanie. Spadek znaczenia sprawozdawczości statystycznej przystosowanej do warunków gospodarki centralnie planowanej i dotyczącej wszystkich jednostek badania jest jednym z przejawów transformacji polskiej statystyki. Jej celem jest adaptacja do wymagań gospodarki rynkowej i zapewnienie porównywalności danych z innymi krajami.

Sprawozdawczość bieżąca dostarczy obecnie w miarę zadowalającą liczbę danych dla małych obszarów w wielu dziedzinach statystyki. Jednakże gdy chodzi o takie sfery jak przemysł, handel wewnętrzny czy rzemiosło, jest odczuwalny dotkliwy brak informacji statystycznych dla małych obszarów.

Kolejnym źródłem danych o małych obszarach są rejestry urzędowe. Można wyodrębnić rejestry ludności, podatkowe, administracyjne, finansowe, a także lokalne rejestry pojazdów, małych zakładów usługowych itd. Użyteczność rejestrów urzędowych jest zróżnicowana [1].

Przedstawione dotychczas źródła informacji o małych obszarach bazują na pełnym wykazie jednostek. Dane są pozyskiwane na podstawie badań kompleksowych, pełnych i wyczerpujących, uwzględniających wszystkie jednostki.

### **3. Znaczenie badań reprezentacyjnych dla statystyki małych obszarów**

W okresie powojennym nastąpił w wielu krajach o gospodarce rynkowej dynamiczny rozwój badań reprezentacyjnych. Natomiast zdecydowana większość badań statystycznych realizowanych w naszym kraju w warunkach gospodarki centralnie planowanej miała charakter badań pełnych, czyli dotyczyła wszystkich jednostek sprawozdawczych. W związku z przemianami systemowymi, jakie nastąpiły w naszym kraju w ostatnich latach, statystyka polska stanęła przed koniecznością adaptacji do wymogów gospodarki rynkowej. Jednym z jej elementów jest stopniowe poszerzanie zakresu stosowania badań reprezentacyjnych i zmniejszanie zakresu badań wyczerpujących.

S. Szulc w pracy [10] określa metodę reprezentacyjną jako: „jedną z form badania częściowego, tj. badania, w którym na podstawie części zbiorowości wnioskujemy o całości”. Cechą wyróżniającą tę formę badań częściowych jest wybór pewnej liczby jednostek drogą losowania. Wylosowaną część zbiorowości nazywamy próbą.

Metodę badań reprezentacyjnych cechuje wiele zalet. Należą do nich [8]:

- 1) zmniejszenie kosztów badań,
- 2) ograniczenie pracochłonności,
- 3) zwiększenie dokładności danych przy tych samych kosztach badania jak w badaniach wyczerpujących,
- 4) przyspieszenie opracowania danych.

Występują też następujące ograniczenia tej metody badań [7]:

- 1) zmniejszenie stopnia szczegółowości opracowań (liczby przekrojów terytorialnych i przedmiotowych) przy danej liczebności próbki,
- 2) większe nakłady pracy wysoko kwalifikowanej kadry w fazie planowania i organizacji badania, a także przy analizie i interpretacji wyników:

Zalety badań reprezentacyjnych przyczyniły się do wzrostu ich zastosowań w krajach o gospodarce rynkowej, gdzie są częściej stosowane niż badania pełne

i dostarczają szerokiego zakresu danych na poziomie krajowym oraz w podstawowych przekrojach terytorialnych i przedmiotowych.

Pierwsza z wymienionych wad badań metodą reprezentacyjną stanowi istotny problem dla statystyki małych obszarów. Metodę tę bowiem cechuje niezdolność do pozyskania wiarygodnych danych dla małych obszarów. Przy badaniu pełnym nie występują tego rodzaju ograniczenia.

Do czynników uniemożliwiających uzyskiwanie wiarygodnych danych dla małych obszarów należą [9; 12]:

- wielkość próby,
- stosowany schemat losowania,
- pożądana precyzja wyników,
- przewidywane koszty badania.

Wielkość prób w małych obszarach jest niewystarczająca. W wielu obszarach próbka w ogóle nie jest reprezentowana, więc na takiej podstawie nie można opierać wiarygodnych ocen. Prowadzenie badań reprezentacyjnych zapewniających uzyskanie precyzyjnych danych dla małych obszarów jest niemożliwe chociażby ze względu na wysokie koszty. W takiej sytuacji jedynym rozwiązaniem jest wykorzystanie procedur matematyczno-statystycznych, umożliwiających szacowanie danych dla małych obszarów, wykorzystujące różne źródła danych. Tego rodzaju metody tworzenia statystyki małych obszarów potwierdziły swą użyteczność w innych krajach.

W Polsce metoda reprezentacyjna jako technika częściowych badań statystycznych była początkowo stosowana w celu przyspieszenia opracowania spisów powszechnych ludności w latach 1950 i 1960. Później znalazła zastosowanie do badań rolniczych, budżetów rodzinnych, warunków życia ludności i innych badań społecznych [6].

W początkowym okresie stosowania metody badań reprezentacyjnych niewiele uwagi poświęcano metodom pozyskiwania danych dla małych obszarów. Najważniejsze znaczenie przypisywano uzyskaniu właściwego operatu do losowania i prezentacji wyników dla całego kraju. Stosowano proporcjonalną lokalizację próbki w warstwach i otrzymywano mniejszą precyzję ocen dla małych województw. Takie wyniki były uznawane za zadowalające.

Nawet w następnych latach metoda reprezentacyjna stosowana była przede wszystkim do badań społecznych i rolniczych, znacznie rzadziej do badań handlu, przemysłu czy budownictwa, gdzie bazowano głównie na pełnej sprawozdawczości statystycznej. Taki system badań funkcjonował w Polsce do 1989 r. W niewystarczającym stopniu zajmowano się wówczas metodologią badań reprezentacyjnych, co niestety spowodowało jej stagnację. Również szkolenie statystyków z tego zakresu badań było niewystarczające. Obecnie problemom tym poświęca się więcej uwagi ze względu na wzrastające znaczenie badań reprezentacyjnych, a także duży popyt na dane statystyczne dla małych obszarów.

#### 4. Metody szacowania danych dla małych obszarów

Prace nad rozwojem metodologii szacowania danych dla małych obszarów są prowadzone w krajach o gospodarce rynkowej dopiero około dwadzieścia pięć lat, dlatego też wiele aspektów metodologicznych nie zostało jeszcze dostatecznie dopracowanych. Przegląd ważniejszych prac dotyczących statystyki małych obszarów prezentują w swoich pracach m.in. J. Kordos [8] i Cz. Bracha [2].

W kraju problematyka uzyskiwania ocen dla małych obszarów nabrała większego znaczenia dopiero w latach dziewięćdziesiątych, przy opracowywaniu programów transformacji polskiej statystyki. Obecnie dąży się do praktycznego wykorzystania metod statystyki małych obszarów przez zapoznanie się z międzynarodowymi osiągnięciami w tej dziedzinie, a także przez prowadzenie własnych prac badawczych i podejmowanie prób adaptacji różnych metod do warunków krajowych.

Punktem wyjścia rozwoju metodologii statystyki małych obszarów była estymacja bezpośrednia. Ten rodzaj estymacji wykorzystuje jedynie informacje dotyczące wartości badanej cechy, pochodzące z próby.

Wyróżnia się następujące metody estymacji bezpośredniej [2; 3]:

- 1) estymacja typu POS,
- 2) estymacja syntetyczna.

Estymacja typu POS wykorzystuje efekt warstwowania próby po losowaniu (por. [12]). Warunkiem stosowania tej metody jest znajomość liczebności populacji dla danego małego obszaru.

Do zastosowania metody estymacji syntetycznej jest z kolei konieczne przyjęcie założenia, że struktura populacji (czyli wszystkie parametry populacji) jest tożsama ze strukturami poszczególnych małych obszarów lub przynajmniej tylko nieznacznie różni się od nich.

Gdy to założenie jest spełnione, można wówczas zastosować ideę tej metody i podzielić ocenę wartości globalnej dla całej populacji na części odpowiadające poszczególnym małym obszarom. Podział ten powinien być proporcjonalny do liczebności populacji w  $i$ -tym małym obszarze.

Przykładowy estymator bezpośredni jest zaprezentowany w poniższej formule [3]:

$$\hat{Y}_i^{(1)} = \frac{N}{n} \sum_{k=1}^n y_{ik} = N_i \bar{y}_i, \quad (1)$$

gdzie:  $\hat{Y}_i^{(1)}$  – oszacowana wartość globalna badanej cechy  $y$  w  $i$ -tym małym obszarze,

$N$  – liczebność całej populacji,

$n$  – liczebność całej próby,

$k = 1, \dots, n$  – elementy próby,

- $y_{ik}$  – wartość cechy  $y$  dla  $k$ -tej jednostki wylosowanej do próby w  $i$ -tym małym obszarze,  
 $N_i$  – liczebność populacji w  $i$ -tym małym obszarze,  
 $\bar{y}_i$  – średnia arytmetyczna cechy  $y_i$  w  $i$ -tym małym obszarze wyznaczona na podstawie próby.

Metody estymacji bezpośredniej nie zawsze umożliwiają uzyskanie wiarygodnych ocen badanych zmiennych dla małych obszarów. Stało się to podstawowym stymulatorem rozwoju bardziej złożonych metod, wykorzystujących informacje statystyczne o innych, dodatkowych zmiennych dla małych obszarów. Rozwinęły się również techniki będące kombinacją zarówno bezpośrednich, jak i pośrednich metod estymacji.

Podstawowym warunkiem stosowania metod estymacji pośredniej jest dysponowanie danymi o wartościach, jakie przyjmuje dla małego obszaru co najmniej jedna cecha dodatkowa (tzw. pomocnicza), silnie skorelowana z cechą badaną. Informacje statystyczne o takich cechach mogą pochodzić zarówno z badania reprezentacyjnego, jak i innych źródeł, takich jak spis, sprawozdawczość bieżąca czy rejestry urzędowe.

Aby możliwe było efektywne wykorzystanie danych pochodzących z innego źródła niż próba losowa, powinny być spełnione następujące warunki [2]:

- zgodność określenia jednostki badania w badaniu źródłowym i docelowym,
- duża przewaga liczebności próby w badaniu źródłowym nad liczebnością próby w badaniu docelowym, umożliwiającą uznanie danych z badania źródłowego za dane z populacji, a nie za realizacje zmiennych losowych.

Do metod estymacji pośredniej zalicza się:

- 1) estymację ilorazową,
- 2) estymację regresyjną.

Technika estymacji regresyjnej znajduje dość szerokie zastosowanie ze względu na nie tylko swą prostotę, ale również duże możliwości adaptacyjne. Dysponując dla małych obszarów wartościami cech dodatkowych, można – wykorzystując ich związek z cechami badanymi – konstruować modele regresji i na ich podstawie uzyskiwać szacunkowe wartości badanych zmiennych.

Precyzja szacunków uzależniona jest od tego, w jakim stopniu są spełnione warunki stosowalności estymatorów skonstruowanych specjalnie dla małych obszarów. Na szczególną uwagę zasługują następujące warunki [2; 3]:

- 1) aktualność wartości cech dodatkowych,
- 2) silny stopień skorelowania cech badanych z cechami dodatkowymi,
- 3) poprawność modelu,
- 4) stabilność współczynnika regresji.

Poniżej przedstawiono ogólny model estymatora regresyjnego [3]:

$$\hat{Y}_i^{(2)} = N_i \left[ \bar{y}_i + \sum_{j=1}^J \hat{B}_j (\bar{X}_{ij} - \bar{x}_{ij}) \right], \quad (2)$$

gdzie:  $\hat{B}_j$  – współczynnik regresji dla  $j = 1, \dots, J$  zmiennych dodatkowych,  
 $N_i$  – liczebność populacji w  $i$ -tym małym obszarze,  
 $\bar{y}_i, \bar{x}_{ij}$  – średnie arytmetyczne, wyznaczone na podstawie próby, odpowiednio dla cechy  $y$  i  $x_j$  w  $i$ -tym obszarze,  
 $\bar{X}_{ij}$  – średnia arytmetyczna, wyznaczona na podstawie całej populacji w  $i$ -tym małym obszarze dla  $j$ -tej zmiennej dodatkowej.

W statystyce małych obszarów zastosowanie znajduje również wiele innych estymatorów, będących kombinacją wymienionych wcześniej metod estymacji bezpośredniej i pośredniej. Można tu wymienić m.in. następujące estymatory złożone [2; 3]:

1) estymatory POS:

- ilorazowy,
- regresyjny,

2) estymatory syntetyczne:

- ilorazowy,
- regresyjny,

3) uogólnione estymatory:

- ilorazowy,
- regresyjny,

4) estymatory złożone, których budowa jest uzależniona od wielkości próby.

Ze względu na dość szerokie zastosowanie modeli regresyjnych zaprezentowane zostaną jeszcze trzy ich rodzaje.

Syntetyczne estymatory regresyjne znajdują zastosowanie wówczas, gdy znane są wartości cech dodatkowych i gdy struktura populacji nie różni się od struktury w małych obszarach lub różni się od niej nieznacznie.

Poniżej przedstawiono syntetyczny estymator regresyjny, dla którego współczynnik regresji wyznaczono na podstawie wartości cechy badanej i dodatkowej we wszystkich jednostkach w próbie:

$$\hat{Y}_i^{(3)} = N_i \left[ \bar{y}_i + \frac{\bar{y}}{\bar{x}} (\bar{X}_i - \bar{x}_i) \right], \quad (3)$$

gdzie:  $\bar{y}, \bar{x}$  – średnie arytmetyczne odpowiednio dla cechy  $y$  i  $x$ , obliczone na podstawie próby we wszystkich małych obszarach,

$\frac{\bar{y}}{\bar{x}}$  – współczynnik regresji,

$\bar{y}_i, \bar{x}_i$  – średnie arytmetyczne odpowiednio dla cechy  $y$  i  $x$ , obliczone na podstawie próby w  $i$ -tym małym obszarze.

Stosowane są również estymatory regresyjne uwzględniające zróżnicowaną strukturę małych obszarów przez wyznaczenie odrębnego współczynnika regresji dla każdego z nich.

Tego typu estymator może przybierać następującą postać:

$$\hat{Y}_i^{(4)} = N_i[\bar{y}_i + B_i(\bar{X}_i - \bar{x}_i)], \quad (4)$$

gdzie:

$$B_i = \frac{\sum_{k=1}^n (x_{ik} - \bar{x}_i)(y_{ik} - \bar{y}_i)}{\sum_{k=1}^n (x_{ik} - \bar{x}_i)^2} \quad (5)$$

$B_i$  – współczynnik regresji dla  $i$ -tego małego obszaru,

$k = 1, \dots, n$  elementy próby,

$x_{ik}$  – wartość cechy  $x$  dla  $k$ -tej jednostki z  $i$ -tego małego obszaru.

Ostatnim prezentowanym rodzajem estymatorów regresyjnych są tzw. uogólnione estymatory regresyjne. Estymatory te stanowią pewną modyfikację estymatorów regresyjnych znanych z teorii metody reprezentacyjnej [2].

Składają się one z dwóch elementów:

1) odpowiedniego estymatora regresyjnego,

2) składnika korygującego.

Składnik korygujący tworzy się na podstawie reszty modelu będącej odchyleniem wartości rzeczywistej cechy  $y$  (dla  $k$ -tej jednostki) od jej wartości oszacowanej przy zastosowaniu jednego z estymatorów regresyjnych.

Jeżeli zastosowany został estymator  $\hat{Y}_i^{(4)}$  (por. formuła (4)), to wówczas uogólniony estymator regresyjny przyjmie następującą postać:

$$\hat{Y}_i^{(5)} = \hat{Y}_i^{(4)} + \sum_{k=1}^n \frac{e_{ki}}{\pi_k}, \quad (6)$$

gdzie:

$$e_{ki} = y_k - \hat{Y}_k \quad (7)$$

$y_k, \hat{Y}_k$  – odpowiednio rzeczywista i oszacowana wartość cechy  $y$  dla  $k$ -tej jednostki z próby dla  $i$ -tego małego obszaru,

$\pi_k$  – prawdopodobieństwo wylosowania do próby  $k$ -tego elementu.

Jak wynika z tych rozważań, wiele metod szacunku może znaleźć zastosowanie w statystyce małych obszarów. Jednak nie można jednoznacznie określić, który z estymatorów jest w danych warunkach optymalny.

Analiza efektywności estymatorów powinna obejmować ustalenie ich precyzji i obciążenia. Badacz dokonujący wyboru określonego estymatora chciałby mieć pewność, że całkowity błąd szacunku (pierwiastek kwadratowy z oceny wariancji powiększony o kwadrat przewidywanego obciążenia) nie przekroczy z góry zadanej wielkości [2]. Konieczna jest zatem znajomość nie tylko efektywności i obciążeń poszczególnych estymatorów, ale również czynników, od których zależą te wielkości. Jeśli chodzi o estymatory syntetyczne i uogólnione estymatory regresyjne, możliwe jest ustalenie jedynie przybliżonych wartości ich obciążenia. Dokładność przybliżenia zależy od liczby informacji o badanych cechach.



Analizy teoretyczne wykazują, że najwyższą efektywność posiadają uogólnione estymatory regresyjne, nieznacznym obciążeniem są obarczone estymatory POS, nieco większym estymatory syntetyczne. Natomiast uogólnione estymatory regresyjne są mniej obciążone niż estymatory syntetyczne [2]. Znajomość tych faktów nie wystarcza jednak do konstruowania estymatorów przydatnych w badaniach praktycznych.

Badania symulacyjne są szczególnie użyteczne do oceny i analizy stochastycznych własności poszczególnych estymatorów. Ich wyniki potwierdzają wyższą efektywność estymatorów regresyjnych. Mimo to nie są one tak często stosowane, jakby to wynikało z ich własności. Ich złożona postać analityczna wydłuża znacznie czas ustalania szacunków [3].

Z badań symulacyjnych Cz. Brachy [2] wynika, że w ogóle nie należy stosować estymatorów złożonych, jeżeli współczynnik korelacji cechy badanej z cechą dodatkową jest mniejszy niż 0,5, ponieważ nie zapewniają one wiarygodnych szacunków. Podjęcie racjonalnej decyzji o zastosowaniu do badań określonego estymatora jest bardzo trudne, możliwe jedynie po przeprowadzeniu wnikliwej analizy warunków realizowanego badania.

Wybór optymalnego estymatora jest niewątpliwie podstawowym, lecz nie jedynym problemem statystyki małych obszarów. Ostatnio są podejmowane również próby znalezienia rozwiązań umożliwiających uzyskanie próby reprezentacyjnej dla małego obszaru. Polegają one na pewnych modyfikacjach próby uzyskanej dla małego obszaru na podstawie badania reprezentacyjnego dla całej populacji. Tego rodzaju rozważania teoretyczne poparte wynikami badań symulacyjnych prezentują w swoich pracach Cz. Domański i K. Pruska [4; 5]. W pracy [5] autorzy podejmują próbę doprowadzenia do reprezentatywności próby dla małego obszaru, zakładając, że jest ona wystarczająco liczna. Algorytm zaproponowanej metody polega na modyfikacji próby przez losowe odrzucenie lub dołosowanie pewnych elementów próby, w sposób zapewniający zachowanie struktury małego obszaru w próbie.

Autorzy przyjmują, że próba jest reprezentatywna, jeżeli zachodzi poniższa relacja:

$$\frac{n_{ig}}{n_i} = \frac{N_{ig}}{N_i}, \quad (8)$$

gdzie:  $n_{ig}$  – liczba elementów próby z całej populacji należących do  $i$ -tego małego obszaru i  $g$ -tej warstwy,

$n_i$  – liczba elementów próby z całej populacji należących do  $i$ -tego małego obszaru,

$N_{ig}$  – liczebność populacji w  $i$ -tym małym obszarze i  $g$ -tej warstwie,

$N_i$  – liczebność populacji w  $i$ -tym małym obszarze,

$g = 1, \dots, G$  – liczba warstw.

Jeżeli relacja (8) nie zachodzi, to należy znaleźć taką liczbę całkowitą  $n_{ig}$ ,

dla której próba będzie mogła być uznana za reprezentatywną. Jeżeli  $n'_{ig} < n_{ig}$ , to należy odrzucić losowo  $n_{ig} - n'_{ig}$  elementów, wykorzystując schemat losowania, jaki był stosowany dla całej populacji.

Jeżeli  $n'_{ig} > n_{ig}$ , to należy odpowiednio wygenerować  $n'_{ig} - n_{ig}$  liczb (por. [5]). Wyniki eksperymentów symulacyjnych pozwoliły autorom sformułować następujące hipotezy badawcze, wytyczające dalsze kierunki prac nad metodą wyznaczania próby reprezentatywnej dla małego obszaru:

1) przypuszcza się, że proponowana metoda umożliwi poprawę oceny średniej dla małego obszaru przy stosunkowo małych próbach wylosowanych z całej populacji i względnie niewielkim liczebnie małym obszarze,

2) prawdopodobnie zaobserwowana prawidłowość jest bardziej widoczna dla populacji charakteryzującej się większym zróżnicowaniem.

Cz. Domański i K. Pruska w pracy [4] zaproponowali również metodę symulacyjnego zwiększania rozmiaru próby dla małego obszaru w wypadku rozkładu zero-jedynkowego. Jest to metoda tworzenia populacji o tym samym rozkładzie, jaki ma rzeczywista zbiorowość. Autorzy nazwali ją pseudopopulacją i wykorzystali do generowania danych symulujących wartości badanych zmiennych. Badania symulacyjne dotyczące zwiększania rozmiaru próby dla małego obszaru mogą być w przyszłości rozszerzone na zmienne cechujące się innymi rozkładami niż zero-jedynkowy.

Metodologia statystyki małych obszarów nie jest jeszcze w pełni rozwinięta. Istnieje zatem stała potrzeba podejmowania szeroko zakrojonych prac badawczych kontynuujących dotychczasowe osiągnięcia w tej dziedzinie.

## LITERATURA

- [1] Balicki A., Szreder M.: *Użyteczności rejestrów urzędowych jako operatów losowania (wyniki badań firm marketingowych)*. W: *Badania sondażowe w zintegrowanym systemie statystyki regionalnej*. Poznań: Centrum Statystyki Regionalnej AE 1996 (maszynopis powielony).
- [2] Bracha Cz.: *Metodologiczne aspekty badania małych obszarów*. Warszawa: GUS 1994. Z Prac Zakładu Badań Statystyczno-Ekonomicznych nr 43.
- [3] Dehnel G.: *Estymacja wskaźników rozwoju gospodarczego regionów za pomocą statystyki małych obszarów*. W: *Badania sondażowe w zintegrowanym systemie statystyki regionalnej*. Poznań: Centrum Statystyki Regionalnej AE 1996 (maszynopis powielony).
- [4] Domański Cz., Pruska K.: *Metoda symulacyjna zwiększania rozmiaru próby dla małego obszaru*. W: *Badania sondażowe w zintegrowanym systemie statystyki regionalnej*. Poznań: Centrum Statystyki Regionalnej AE 1996 (maszynopis powielony).
- [5] Domański Cz., Pruska K.: *Reprezentatywność próby w statystyce małych obszarów* „Wiadomości Statystyczne” 1996 nr 5, s. 11–16.
- [6] Kordos J.: *Podejścia do statystyki małych obszarów w Polsce*. „Wiadomości Statystyczne” 1992 nr 10, s. 1–5.

- 
- [7] Kordos J.: *Statystyka małych obszarów a badania reprezentacyjne*. „Wiadomości Statystyczne” 1991 nr 4, s. 1–5.
- [8] Kordos J.: *Warunki efektywnego wykorzystywania postępów w zakresie statystyki małych obszarów*. W: *Badania sondażowe w zintegrowanym systemie statystyki regionalnej*. Poznań: Centrum Statystyki Regionalnej AE 1996 (maszynopis powielony).
- [9] Nowak L.: *Źródła danych i estymacja statystyki ludności dla małych obszarów*. „Wiadomości Statystyczne” 1991 nr 11, s. 5–9.
- [10] Szulc S.: *Metody statystyczne*. Warszawa: PWE 1968.
- [11] Szwalek S., Zaremba H.: *Źródło danych statystycznych oraz szacunków dla małych obszarów*. „Wiadomości Statystyczne” 1992 nr 11, s. 15–19.
- [12] Zasępa R.: *Metoda reprezentacyjna*. Warszawa: PWE 1972.

## STATISTICS OF SMALL REGIONS – METHODOLOGICAL ASPECT

### Summary

The paper presents some statistical methods applied to estimate of information for small regions. The article is of a methodological character.