ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 156, 2002

Dorota Pekasiewicz , Krystyna Pruska***

ANALIZA ROZKŁADÓW WYBRANYCH ESTYMATORÓW W STATYSTYCE MAŁYCH OBSZARÓW

Streszczenie. Statystyka małych obszarów dostarcza metod badania podpopulacji wyróżnionych w skończonych populacjach. Do głównych problemów w analizach małego obszaru należy szacowanie parametrów rozkładu rozpatrywanych zmiennych, przy czym wnioskowanie prowadzone jest na podstawie prób losowanych według różnych schematów.

W pracy rozpatrywane są cztery estymatory wartości globalnej dla małego obszaru. Ich własności analizowane są w przypadku losowania warstwowego i warstwowania po wylosowaniu próby dla różnych rozmiarów próby z populacji. Badanie przeprowadzone jest za pomocą eksperymentów Monte Carlo. Dla populacji o ustalonych rozkładach normalnych i Poissona na podstawie wygenerowanych danych wyznaczone są histogramy rozkładów estymatorów wartości globalnej oraz ich odchylenia standardowe, które można uznać za miarę efektywności estymatorów ze względu na nieznaczne ich obciążenie. Otrzymane wyniki dla skonstruowanych populacji świadczą o dużej efektywności analizowanych estymatorów, przy czym największą efektywnością charakteryzuje się estymator syntetyczny, wykorzystujący rzeczywiste wartości globalne cechy pomocniczej, a nie ich oszacowania.

Słowa kluczowe: statystyka małych obszarów, estymator bezpośredni, estymator pośredni, rozkład estymatora.

1. WPROWADZENIE

Statystyka małych obszarów jest działem statystyki zajmującym się badaniem podpopulacji wyróżnionych w skończonych populacjach. Do głównych problemów w analizach małego obszaru należy szacowanie parametrów rozkładu rozpatrywanych zmiennych na podstawie prób losowanych według różnych schematów. Charakterystyka rozkładu estymatorów parametrów jest istotnym zagadnieniem wykorzystywanym zarówno przy estymacji,

* Dr, Katedra Metod Statystycznych, Uniwersytet Łódzki.
** Dr hab., Katedra Metod Statystycznych, Uniwersytet Łódzki.

jak i weryfikacji hipotez statystycznych. Skomplikowane schematy losowania próby utrudniają dokładne wyznaczenie rozkładu odpowiednich statystyk i dlatego w literaturze (por. np. Bracha (1996)) często spotyka się jedynie wzory na wartość oczekiwaną i wariancję estymatorów, co jest i tak wynikiem wielu żmudnych przeliczeń.

W pracy tej rozpatrywane są wybrane estymatory wartości globalnej dla małego obszaru w przypadku losowania warstwowego i warstwowania po wylosowaniu próby. Badanie rozkładu tych estymatorów przeprowadzone zostało za pomocą eksperymentów Monte Carlo. Dotyczyło ono miar średnich, rozproszenia i asymetrii rozkładu estymatora oraz zgodności tego rozkładu z rozkładem normalnym.

2. WYBRANE ESTYMATORY WARTOŚCI GLOBALNEJ DLA MAŁEGO OBSZARU

Znane są różne estymatory wartości globalnej dla małego obszaru. Wyróżniamy wśród nich dwie główne grupy: estymatory bezpośrednie i estymatory pośrednie, do których należą estymatory syntetyczne (por. np. Dol (1991), Bracha (1994), Schaible i Casady (1994), Gołata (1996), Kordos (1999)).

Estymatory bezpośrednie są skonstruowane na podstawie tych elementów próby wylosowanej z całej populacji, które należą do małego obszaru. Przy budowie estymatorów pośrednich wykorzystywane są jeszcze dodatkowe informacje o populacji i małym obszarze. Stosowanie ich ma na celu uzyskanie estymatorów o mniejszej wariancji niż estymatory bezpośrednie. Cechą charakterystyczną konstrukcji estymatorów syntetycznych jest założenie, że parametry rozkładu zmiennej, ze względu na którą badany jest mały obszar, są bardzo zbliżone do parametrów rozkładu populacji. Ostateczne oceny parametrów dla małego obszaru otrzymuje się po korektach, uwzględniających dodatkowe informacje o badanej podpopulacji. Przy estymacji syntetycznej często przyjmuje się, że stosunek wartości globalnych badanej zmiennej i zmiennej pomocniczej dla każdej warstwy, na jakie została podzielona populacja, jest taki sam w małym obszarze, jak i w populacji.

W pracy tej analizowane są cztery estymatory wartości globalnej. Badana zmienna oznaczona jest przez Y, a zmienna pomocnicza przez X. Zakładamy, że populacja podzielona jest na G warstw i wyróżnia się w niej H małych obszarów. Ponadto przyjęte są następujące oznaczenia dla $h = 1, ..., H$ i $g = 1, ..., G$:

TX_{hg} – wartość globalna zmiennej X w h-tym małym obszarze i g-tej warstwie,

$TX_{.g}$ – wartość globalna zmiennej X w g-tej warstwie w populacji,

$TY_{h.}$ – wartość globalna zmiennej Y w h-tym małym obszarze,

X_{hg} – wartość globalna zmiennej X w h-tym małym obszarze i g-tej warstwie dla próby wylosowanej z populacji,

$X_{.g}$ – wartość globalna zmiennej X w g-tej warstwie dla próby wylosowanej z populacji,

Y_{hg} – wartość globalna zmiennej Y w g-tej warstwie i h-tym małym obszarze dla próby wylosowanej z populacji,

$Y_{.g}$ – wartość globalna zmiennej Y w g-tej warstwie dla próby wylosowanej z populacji,

N_{hg} – liczba elementów populacji, które należą do h-tego małego obszaru i g-tej warstwy,

$N_{.g}$ – liczebność g-tej warstwy w populacji,

n_{hg} – liczba elementów próby wylosowanej z populacji, które należą do h-tego małego obszaru i g-tej warstwy.

$n_{.g}$ – liczba elementów z g-tej warstwy w próbie wylosowanej z populacji.

Rozpatrywane estymatory parametru TY_h. określone są następującymi wzorami:

$$T_1 = \sum_{g=1}^{G} \frac{N_{hg}}{n_{hg}} \cdot Y_{hg} \tag{1}$$

$$T_2 = \sum_{g=1}^{G} \frac{N_{.g}}{n_{.g}} \cdot \frac{TX_{hg}}{TX_{.g}} \cdot Y_{.g} \tag{2}$$

$$T_3 = \sum_{g=1}^{G} \frac{N_{.g}}{n_{.g}} \cdot \frac{X_{hg}}{X_{.g}} \cdot Y_{.g} \tag{3}$$

$$T_4 = \sum_{g=1}^{G} \frac{N_{hg}}{n_{hg}} \cdot \frac{X_{hg}}{X_{.g}} \cdot Y_{.g} \tag{4}$$

przy czym estymatory te rozważane są dla dwóch schematów losowania próby: dla losowania warstwowego i dla warstwowania po wylosowaniu próby.

3. ANALIZA MONTE CARLO ROZKŁADÓW ESTYMATORÓW WARTOŚCI GLOBALNEJ DLA MAŁEGO OBSZARU

Analiza Monte Carlo badanego zjawiska polega na skonstruowaniu modelu probabilistycznego tego zjawiska i przeprowadzeniu na tej podstawie eksperymentów w celu poznania jego własności.

3.1. ORGANIZACJA BADAŃ MONTE CARLO

Metody symulacyjne stwarzają możliwość porównywania rzeczywistych parametrów rozkładu badanej zmiennej z ich oszacowaniami, a więc dokładnego określenia wielkości błędu oszacowania. Pozwalają też na zapewnienie takiej wielkości próby, aby można było stosować metody statystyczne odpowiednie do przeprowadzenia badania.

W celu zbadania własności estymatorów T_1, T_2, T_3 i T_4 wartości globalnej TY dla małego obszaru utworzona została populacja 60 000 wartości otrzymanych według następującego algorytmu:

I. Generujemy 60 000 wartości według rozkładu o funkcji prawdopodobieństwa postaci:

$$P(Z = k) = \frac{1}{3} \quad \text{dla } k = 1, 2, 3$$

i oznaczamy je $z_1, \ldots, z_{60\,000}$.

II. Tworzymy szereg $\tilde{y}_1, \ldots, \tilde{y}_{60\,000}$ taki, że dla $k = 1, \ldots, 60\,000$ liczba $\tilde{y}_k$ jest wartością wygenerowaną według rozkładu normalnego $N\left(3; \frac{1}{2}\right)$, gdy $z_k = 1$; $\tilde{y}_k$ jest wartością wygenerowaną według rozkładu $N\left(5; \frac{1}{2}\right)$, gdy $z_k = 2$ oraz $\tilde{y}_k$ jest wartością z rozkładu $N(10; 1)$, gdy $z_k = 3$.

III. Tworzymy trzy zbiory liczb, które będą stanowiły trzy warstwy w populacji; pierwsza warstwa to liczby $\tilde{\tilde{y}}_k = \tilde{y}_k$ dla $k = 1, \ldots, 20\,000$, druga warstwa to liczby $\tilde{\tilde{y}}_k = \tilde{y}_k + 1{,}5$, dla $k = 20\,001, \ldots, 40\,000$, trzecia warstwa to liczby $\tilde{\tilde{y}}_k = \tilde{y}_k + 2{,}5$ dla $k = 40\,001, \ldots, 60\,000$.

IV. Pierwszych 1000 wartości z każdej warstwy zostaje przekształconych według wzoru:

$$y_l = \tilde{\tilde{y}}_l + \varepsilon_l, \quad l = 1, \ldots, 1000, 20\,001, \ldots, 21\,000, 40\,001, \ldots, 41\,000$$

gdzie ε_l jest liczbą wygenerowaną według rozkładu $N(1; 0{,}3)$, a $\tilde{\tilde{y}}_l$ odpowiednią wartością z szeregu $\tilde{\tilde{y}}_1, \ldots, \tilde{\tilde{y}}_{60\,000}$; wartości y_l stanowią mały obszar (jest ich 3 000).

VI. Ostatecznie populację stanowią elementy małego obszaru i nie przekształcone pozostałe wartości spośród $\tilde{\tilde{y}}_1, \ldots, \tilde{\tilde{y}}_{60\,000}$; elementy populacji będziemy oznaczać przez $y_1, \ldots, y_{60\,000}$.

VII. Z kolei tworzymy dodatkowy szereg $x_1, \ldots, x_{60\,000}$ według wzoru:

$$x_k = 2y + \xi_k, \quad k = 1, \ldots, 60\,000$$

gdzie ξ_k są liczbami wygenerowanymi według rozkładu $N\left(0; \frac{1}{2}\right)$.

Otrzymaliśmy w ten sposób 60 000 par liczbowych (x_1, y_1), ..., $(x_{60\,000}, y_{60\,000})$. Możemy je potraktować jako wartości dwuwymiarowej zmiennej (X, Y), przy czym Y jest zmienną badaną w populacji, a X zmienną pomocniczą.

Z wyznaczonej populacji wylosowano zależnie po 10 000 razy próby o liczebnościach: 900, 2100, 3000, 6000. Próby te stanowiły podstawę do obliczenia wartości estymatorów T_1, T_2, T_3 i T_4 po dokonaniu warstwowania próby, czyli po określeniu, do której warstwy należy dany element próby. Był to I wariant eksperymentów. Wariant II eksperymentów różnił się od pierwszego jedynie schematem losowania. Zastosowano teraz zależne losowanie warstwowe, przy czym liczebności próby z populacji pozostały na tym samym poziomie, tzn. 900, 2100, 3000, 6000, a liczby losowań z poszczególnych warstw były równe $\frac{1}{3}$ liczebności próby z populacji. Warianty III i IV polegały na zastąpieniu w VI etapie tworzenia populacji rozkładu $N\left(0; \frac{1}{2}\right)$ zmiennej ξ_k, rozkładem $N(0; 2)$, przy czym w wariancie III zastosowano warstwowanie po wylosowaniu próby, a w wariancie IV losowanie warstwowe. Wszystkie cztery warianty, w których w II etapie tworzenia populacji wystąpił rozkład normalny, nazwane zostały grupą A eksperymentów. Prowadzoną analizę rozszerzono o jeszcze jedną grupę eksperymentów, oznaczoną grupą B, w której rozkład normalny w II etapie tworzenia populacji zastąpiono rozkładem Poissona z wartościami oczekiwanymi odpowiednio: 3, 5, 10.

3.2. WYNIKI ANALIZY MONTE CARLO

Celem analizy Monte Carlo było poznanie własności estymatorów T_1, T_2, T_3 i T_4. Na podstawie przeprowadzonych eksperymentów wyznaczono wartości globalne dla zmiennych X i Y w populacji i małym obszarze oraz w poszczególnych warstwach populacji i małego obszaru. Wyniki obliczeń przedstawione są w tabl. 1 i 2. Tablica 3 zawiera informacje o liczebności prób dla małego obszaru. Następnie wyznaczono oszacowania parametrów rozkładów statystyk T_1, T_2, T_3 i T_4, takie jak: średnia, odchylenie standardowe i współczynniki asymetrii dla każdej grupy i każdego wariantu eksperymentów. Oszacowania te podane są w tabl. 4 i 5. Wynika z nich, że rozkłady rozpatrywanych estymatorów charakteryzują się bardzo słabą asymetrią (można ją uznać za nieistotną). Średnie oszacowań uzyskanych za pomocą estymatorów T_1, T_2, T_3 i T_4 są zbliżone do siebie i do rzeczywistej wartości globalnej (por. tabl. 1) dla małego obszaru. Świadczy to o nieznacznym obciążeniu estymatorów. Zróżnicowanie ich wartości również nie jest wielkie,

Tablica 1

Wartości globalne zmiennych Y oraz X w populacji i jej małym obszarze w przeprowadzonych eksperymentach

Grupa eksperymentów	Wariant eksperymentów	Wartości globalne zmiennej Y		Wartości globalne zmiennej X	
		w populacji	w małym obszarze	w populacji	w małym obszarze
A	I, II	443 435,005	25 079,122	886 735,933	50 124,117
	III, IV	443 435,005	25 079,122	886 333,704	50 021,740
B	I, II	443 612,740	25 357,740	887 301,680	50 754,828
	III, IV	443 612,740	25 357,740	887 530,279	50 872,870

Źródło: obliczenia własne.

Tablica 2

Wartości globalne zmiennych Y oraz X w poszczególnych warstwach populacji w przeprowadzonych eksperymentach

Grupa eksperymentów	Wariant eksperymentów	Numer warstwy	Wartości globalne zmiennej Y w warstwach populacji	Wartości globalne zmiennej X w warstwach populacji
A	I, II	1	121 167,892	242 332,706
		2	151 513,567	302 987,599
		3	170 753,546	341 415,629
	III, IV	1	121 167,892	242 323,473
		2	151 513,567	302 868,992
		3	170 753,546	341 141,239
B	I, II	1	121 489,000	243 006,861
		2	150 872,765	301 723,648
		3	171 241,975	342 571,171
	III, IV	1	121 489,000	243 039,445
		2	150 872,765	301 658,001
		3	171 241,975	342 832,833

Źródło: obliczenia własne.

Tablica 3

Charakterystyka rozkładu liczebności próby dla małego obszaru w przeprowadzonych eksperymentach

Grupa eksperymentów	Wariant	Liczbność próby z populacji	Charakterystyki rozkładu liczebności n próby dla małego obszaru			
			przeciętna liczebność $\bar{n}$	odchylenie standardowe d(n)	minimalna liczebność	maksymalna liczebność
A	I, III	900	44,5542	6,4421	19	71
		2 100	103,0546	9,7741	68	142
		3 000	146,2478	11,4742	106	197
		6 000	285,4022	15,6464	223	348
	II, IV	900	44,5697	6,5323	22	74
		2 100	103,2634	9,7011	68	142
		3 000	146,2619	11,4195	99	194
		6 000	285,6164	15,4337	233	343
B	I, III	900	44,5694	6,4671	24	71
		2 100	103,0813	9,7665	67	141
		3 000	146,2526	11,3934	100	190
		6 000	285,3620	15,7010	228	353
	II, IV	900	44,5641	6,5919	21	79
		2 100	103,2471	9,7771	67	139
		3 000	146,2504	11,3797	107	202
		6 000	285,6061	15,6956	225	353

Źródło: obliczenia własne.

Tablica 4

Oszacowania parametrów rozkładu ekstymatorów T_1, T_2, T_3 i T_4 wartości globalnej TY zmiennej Y w małym obszarze w grupie eksperymentów A

Wariant	Liczebność próby z populacji	Estymator wartości globalnej TY w małym obszarze	Wartość średnia estymatora	Odchylenie standardowe estymatora	Współczynnik asymetrii estymatora
1	2	3	4	5	6
I	900	T_1	25 089,293	1 511,158	0,050
		T_2	25 063,701	363,448	–0,046
		T_3	25 011,232	3 937,787	0,161
		T_4	25 076,164	1 512,496	0,045
	2 100	T_1	25 094,805	975,849	0,048
		T_2	25 064,056	238,335	0,025
		T_3	25 049,757	2 570,732	0,119
		T_4	25 082,595	978,050	0,046

Tablica 4 (cd.)

1	2	3	4	5	6
	3 000	T_1	25 084,815	807,006	0,050
		T_2	25 063,139	199,808	0,014
		T_3	25 060,055	2 131,239	0,091
		T_4	25 072,705	809,472	0,048
	6 000	T_1	25 080,051	546,864	0,029
		T_2	25 063,667	139,173	0,021
		T_3	25 060,421	1 488,527	0,069
		T_4	25 067,157	548,451	0,025
II	900	T_1	25 106,919	1 491,572	0,032
		T_2	25 069,057	362,716	0,006
		T_3	25 045,474	3 974,196	0,197
		T_4	25 093,100	1 491,872	0,035
	2 100	T_1	25 081,670	950,677	0,018
		T_2	25 065,103	238,258	0,003
		T_3	25 085,473	2 559,010	0,094
		T_4	25 067,518	952,077	0,020
	3 000	T_1	25 078,555	799,385	0,017
		T_2	25 062,281	197,845	0,052
		T_3	25 059,046	2 119,144	0,105
		T_4	25 064,404	800,861	0,018
	6 000	T_1	25 083,277	554,918	0,032
		T_2	25 067,166	139,659	0,023
		T_3	25 082,979	1 470,861	0,064
		T_4	25 069,354	556,145	0,036
III	900	T_1	25 089,293	1 511,158	0,050
		T_2	25 023,747	362,919	–0,046
		T_3	24 974,067	3 955,376	0,164
		T_4	25 036,893	1 565,404	0,031
	2 100	T_1	25 094,805	975,849	0,048
		T_2	25 024,102	237,990	0,025
		T_3	25 013,235	2 582,916	0,124
		T_4	25 045,992	1 015,207	0,041
	3 000	T_1	25 081,415	807,006	0,050
		T_2	25 023,186	199,516	0,014
		T_3	25 023,420	2 139,613	0,089
		T_4	25 036,394	842,489	0,037
	6 000	T_1	25 080,051	546,864	0,029
		T_2	25 023,715	138,971	0,021
		T_3	25 021,723	1 496,483	0,063
		T_4	25 028,468	571,124	0,014

Tablica 4 (cd.)

1	2	3	4	5	6
IV	900	T_1	25 106,919	1 491,572	0,032
		T_2	25 029,098	362,186	0,006
		T_3	25 004,597	3 985,546	0,200
		T_4	25 051,738	1 541,918	0,042
	2 100	T_1	25 081,670	950,677	0,018
		T_2	25 025,145	237,909	0,003
		T_3	25 043,110	2 567,765	0,094
		T_4	25 025,102	986,972	0,025
	3 000	T_1	25 078,555	799,385	0,017
		T_2	25 022,331	197,558	0,052
		T_3	25 016,417	2 125,160	0,096
		T_4	25 021,953	830,478	0,015
	6 000	T_1	25 083,277	554,918	0,032
		T_2	25 027,206	139,454	0,023
		T_3	25 041,085	1 476,466	0,064
		T_4	25 027,579	577,532	0,043

Źródło: obliczenia własne.

Tablica 5

Oszacowania parametrów rozkładu ekstymatorów T_1, T_2, T_3 i T_4 wartości globalnej TY zmiennej Y w małym obszarze w grupie eksperymentów B

Wariant	Liczebność próby z populacji	Estymator wartości globalnej TY w małym obszarze	Wartość średnia estymatora	Odchylenie standardowe estymatora	Współczynnik asymetrii estymatora
1	2	3	4	5	6
I	900	T_1	25 388,934	1 784,560	0,101
		T_2	25 368,212	440,282	0,005
		T_3	25 351,198	4 075,545	0,149
		T_4	25 406,414	1 791,269	0,096
	2 100	T_1	25 388,508	1 138,550	0,045
		T_2	25 371,254	288,567	0,003
		T_3	25 377,793	2 645,671	0,092
		T_4	25 406,001	1 142,569	0,041
	3 000	T_1	25 375,528	938,997	0,020
		T_2	25 370,670	238,920	–0,005
		T_3	25 382,174	2 198,038	0,190
		T_4	25 394,215	940,949	0,018
	6 000	T_1	25 369,609	649,177	0,072
		T_2	25 373,316	165,899	–0,022
		T_3	25 374,292	1 335,113	0,075
		T_4	25 388,625	651,371	0,069

Tablica 5 (cd.)

1	2	3	4	5	6
II	900	T_1	25 372,087	1 781,330	0,162
		T_2	25 374,275	440,348	0,018
		T_3	25 316,995	4 140,508	0,205
		T_4	25 388,986	1 785,000	0,146
	2 100	T_1	25 355,022	1 134,160	0,062
		T_2	25 374,292	286,317	–0,029
		T_3	25 385,898	2 636,433	0,096
		T_4	25 373,050	1 137,360	0,066
	3 000	T_1	25 345,726	939,779	0,042
		T_2	25 374,396	240,007	0,026
		T_3	25 351,013	2 486,597	0,102
		T_4	25 166,171	941,241	0,045
	6 000	T_1	25 364,890	650,155	0,060
		T_2	25 374,893	170,631	0,041
		T_3	25 392,432	1 551,233	0,047
		T_4	25 384,488	651,831	0,058
III	900	T_1	25 388,934	1 784,560	0,110
		T_2	25 420,799	441,395	0,100
		T_3	25 402,414	4 106,241	0,146
		T_4	25 457,824	1 847,014	0,088
	2 100	T_1	25 388,508	1 138,550	0,045
		T_2	25 423,857	289,293	0,005
		T_3	25 429,959	2 664,865	0,091
		T_4	25 458,519	1 181,608	0,027
	3 000	T_1	25 375,528	938,997	0,020
		T_2	25 423,277	239,516	–0,005
		T_3	25 435,026	2 214,828	0,144
		T_4	25 449,853	970,052	0,004
	6 000	T_1	25 369,609	649,177	0,072
		T_2	25 425,929	166,320	–0,022
		T_3	25 426,016	1 545,686	0,070
		T_4	25 445,354	672,163	0,064
IV	900	T_1	25 372,087	1 781,330	0,162
		T_2	25 426,895	441,463	0,018
		T_3	25 358,293	4 171,769	0,201
		T_4	25 443,726	1 835,765	0,131
	2 100	T_1	25 355,022	1 134,160	0,062
		T_2	25 426,925	287,052	–0,029
		T_3	25 440,213	2 655,744	0,100
		T_4	25 427,183	1 173,412	0,071

Tablica 5 (cd.)

1	2	3	4	5	6
	3 000	T_1	25 343,726	939,779	0,042
		T_2	25 427,029	240,605	0,026
		T_3	25 404,983	2 203,651	0,103
		T_4	25 427,172	970,390	0,035
	6 000	T_1	25 364,890	650,155	0,060
		T_2	25 427,515	171,052	0,041
		T_3	25 444,682	1 562,012	0,048
		T_4	25 441,774	671,915	0,042

Źródło: obliczenia własne.

Tablica 6

Wartości statystyki testu zgodności χ^2 oraz wskaźniki podobieństwa struktur estymatora wartości globalnej zmiennej w małym obszarze w grupie eksperymentów A

Wariant	Liczebność próby z populacji	Estymator wartości globalnej TY w małym obszarze	Wartość statystyki testu χ^2	Wskaźniki podobieństwa struktur
1	2	3	4	5
I	900	T_1	22,994	0,986
		T_2	27,542	0,980
		T_3	67,235	0,976
		T_4	25,776	0,985
	2 100	T_1	34,807	0,985
		T_2	29,878	0,980
		T_3	52,022	0,982
		T_4	27,481	0,982
	3 000	T_1	43,742	0,976
		T_2	18,490	0,987
		T_3	39,638	0,979
		T_4	44,199	0,977
	6 000	T_1	27,150	0,986
		T_2	19,378	0,985
		T_3	35,189	0,984
		T_4	23,615	0,983
	900	T_1	27,721	0,984
		T_2	59,754	0,979
		T_3	112,319	0,970
		T_4	22,804	0,986

Tablica 6 (cd.)

1	2	3	4	5
II	2 100	T_1	26,235	0,988
		T_2	13,641	0,988
		T_3	74,070	0,978
		T_4	47,087	0,980
	3 000	T_1	22,111	0,983
		T_2	24,627	0,985
		T_3	37,603	0,979
		T_4	28,168	0,983
	6 000	T_1	23,043	0,986
		T_2	25,262	0,987
		T_3	23,984	0,985
		T_4	30,348	0,982
III	900	T_1	22,994	0,986
		T_2	26,408	0,980
		T_3	69,713	0,974
		T_4	27,554	0,985
	2 100	T_1	34,807	0,985
		T_2	31,171	0,979
		T_3	42,114	0,982
		T_4	38,464	0,978
	3 000	T_1	43,742	0,976
		T_2	19,165	0,987
		T_3	38,476	0,979
		T_4	36,411	0,978
	6 000	T_1	27,150	0,986
		T_2	20,976	0,984
		T_3	43,156	0,980
		T_4	24,671	0,983
IV	900	T_1	27,721	0,984
		T_2	59,472	0,980
		T_3	106,085	0,975
		T_4	32,869	0,983
	2 100	T_1	26,235	0,988
		T_2	14,639	0,987
		T_3	46,055	0,979
		T_4	41,843	0,986
	3 000	T_1	22,111	0,983
		T_2	25,016	0,985
		T_3	42,234	0,976
		T_4	26,479	0,981
	6 000	T_1	23,043	0,986
		T_2	25,004	0,987
		T_3	22,793	0,986
		T_4	29,915	0,979

Źródło: obliczenia własne.

Tablica 7

Wartości statystyki testu zgodności χ^2 oraz wskaźniki podobieństwa struktur estymatora wartości globalnej zmiennej w małym obszarze w grupie eksperymentów B

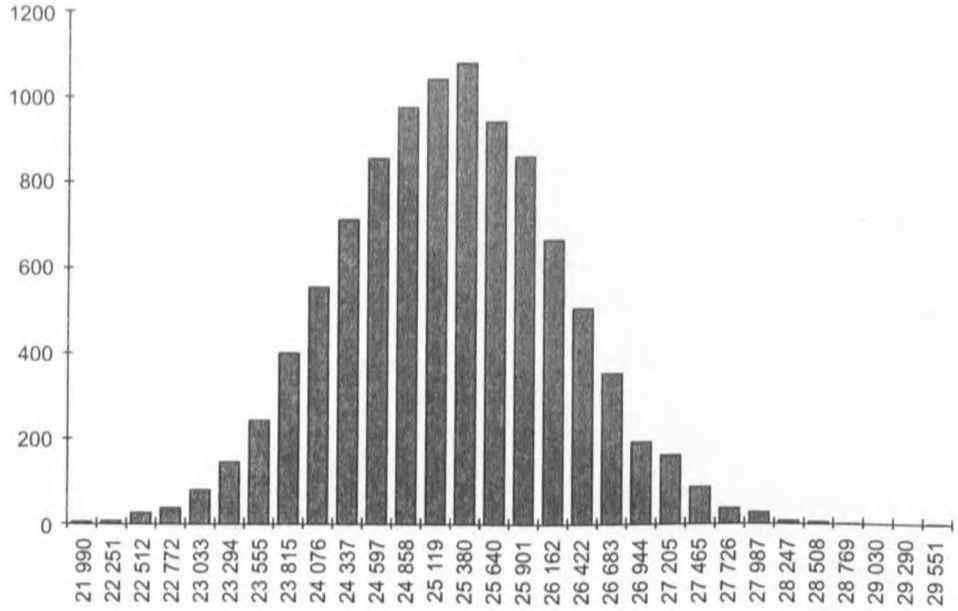
Wariant	Liczebność próby z populacji	Estymator wartości globalnej TY w małym obszarze	Wartość statystyki testu χ^2	Wskaźniki podobieństwa struktur
1	2	3	4	5
I	900	T_1	53,481	0,983
		T_2	20,099	0,986
		T_3	67,486	0,981
		T_4	50,278	0,983
	2 100	T_1	34,745	0,981
		T_2	33,529	0,975
		T_3	52,762	0,975
		T_4	30,563	0,980
	3 000	T_1	31,610	0,983
		T_2	31,698	0,977
		T_3	69,023	0,978
		T_4	25,667	0,983
	6 000	T_1	35,962	0,980
		T_2	26,311	0,981
		T_3	46,693	0,975
		T_4	43,157	0,980
II	900	T_1	207,712	0,978
		T_2	39,878	0,979
		T_3	487,843	0,973
		T_4	161,537	0,980
	2 100	T_1	33,624	0,984
		T_2	25,793	0,983
		T_3	44,086	0,983
		T_4	42,038	0,982
	3 000	T_1	38,190	0,979
		T_2	26,663	0,980
		T_3	44,270	0,980
		T_4	27,529	0,981
	6 000	T_1	37,104	0,979
		T_2	27,772	0,979
		T_3	19,187	0,983
		T_4	32,259	0,979
	900	T_1	53,481	0,983
		T_2	21,367	0,987
		T_3	56,613	0,981
		T_4	61,160	0,984

Tablica 7 (cd.)

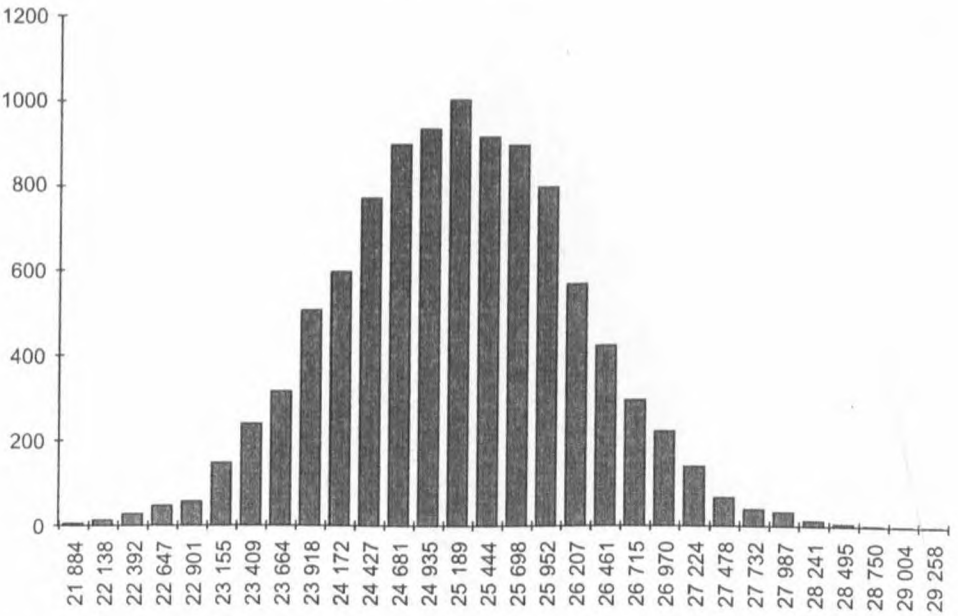
1	2	3	4	5
III	2 100	T_1	34,745	0,981
		T_2	33,320	0,975
		T_3	56,952	0,977
		T_4	42,038	0,982
	3 000	T_1	31,610	0,983
		T_2	33,034	0,977
		T_3	64,248	0,978
		T_4	22,672	0,984
	6 000	T_1	35,962	0,980
		T_2	26,707	0,981
		T_3	37,892	0,981
		T_4	30,498	0,980
IV	900	T_1	207,712	0,978
		T_2	39,870	0,979
		T_3	520,491	0,973
		T_4	240,768	0,980
	2 100	T_1	33,624	0,984
		T_2	27,109	0,983
		T_3	43,018	0,981
		T_4	41,039	0,981
	3 000	T_1	38,190	0,979
		T_2	25,200	0,981
		T_3	36,881	0,980
		T_4	24,845	0,986
	6 000	T_1	37,104	0,979
		T_2	29,290	0,978
		T_3	25,736	0,981
		T_4	28,892	0,981

Źródło: obliczenia własne.

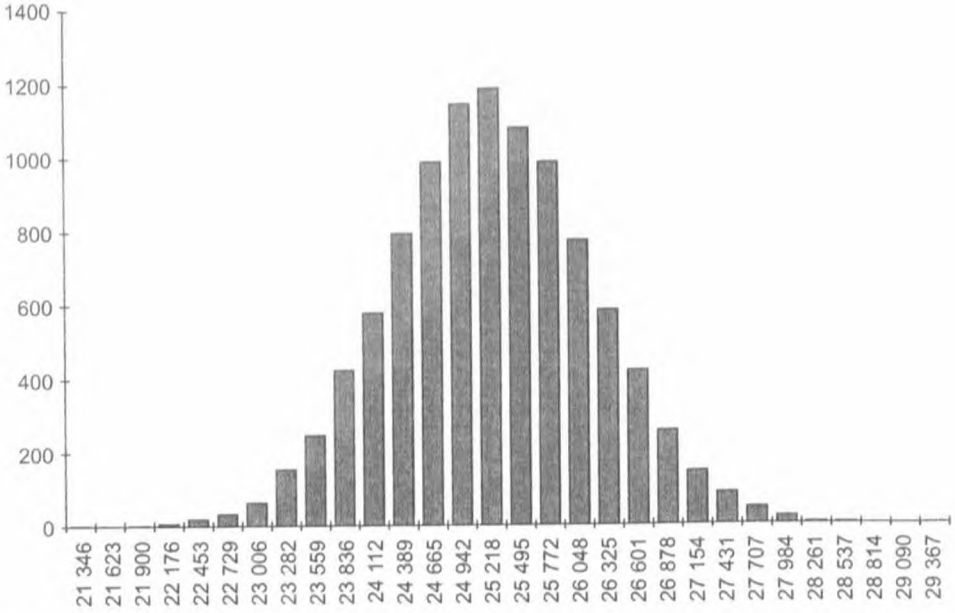
względnie najmniejszym charakteryzuje się estymator T_2, a największym estymator T_3. Przy konstrukcji estymatora T_2 wykorzystywane są informacje z badania całkowitego dla cechy X, natomiast przy obliczaniu wartości estymatorów T_3 i T_4 – informacje z próby o tej zmiennej. Wartości estymatora T_1 są wyznaczone jedynie na podstawie danych z próby dla zmiennej Y. Przy konstrukcji estymatora T_3 przyjęto, że stosunek wartości globalnych cechy X w g-tej warstwie ($g = 1, ..., G$) h-tego małego obszaru ($h = 1, ..., H$) oraz w g-tej warstwie populacji jest w przybliżeniu równy stosunkowi odpowiednich wartości globalnych w próbie, czyli $TX_{hg} / TX_{.g} \approx X_{hg} / X_{.g}$. Budowa estymatora T_4 opiera się na podobnym



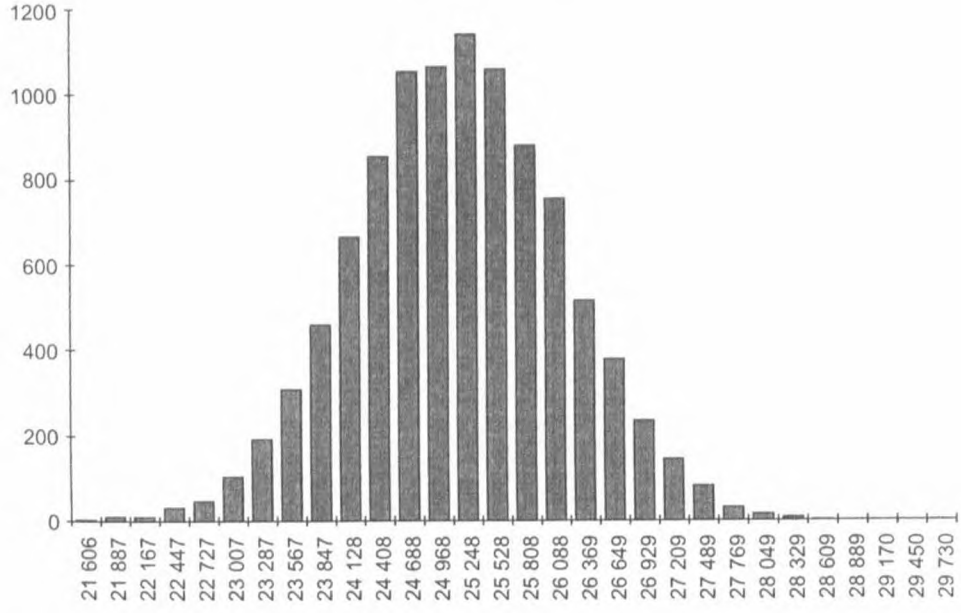
Rys. 1. Rozkład estymatora T_1 w przypadku warstwowania po wylosowaniu próby – wariant III w grupie eksperymentów A



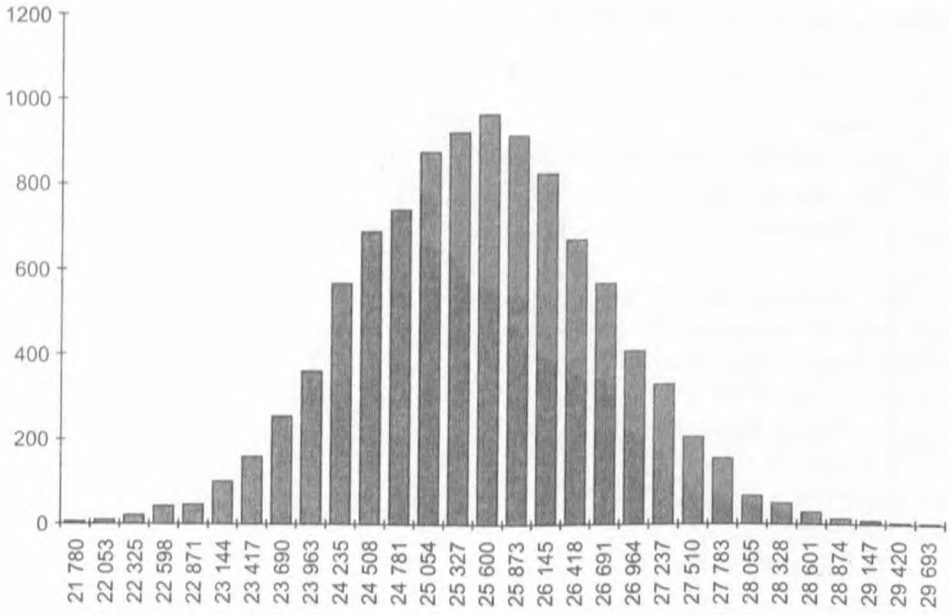
Rys. 2. Rozkład estymatora T_4 w przypadku warstwowania po wylosowaniu próby – wariant III w grupie eksperymentów A



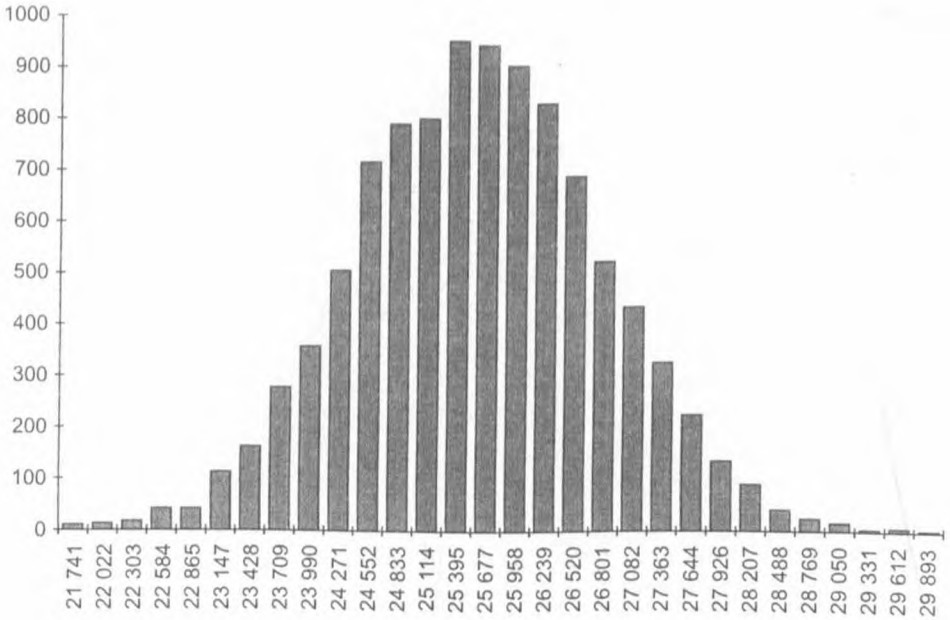
Rys. 3. Rozkład estymatora T_1 w przypadku losowania warstwowego zależnego – wariant IV w grupie eksperymentów A



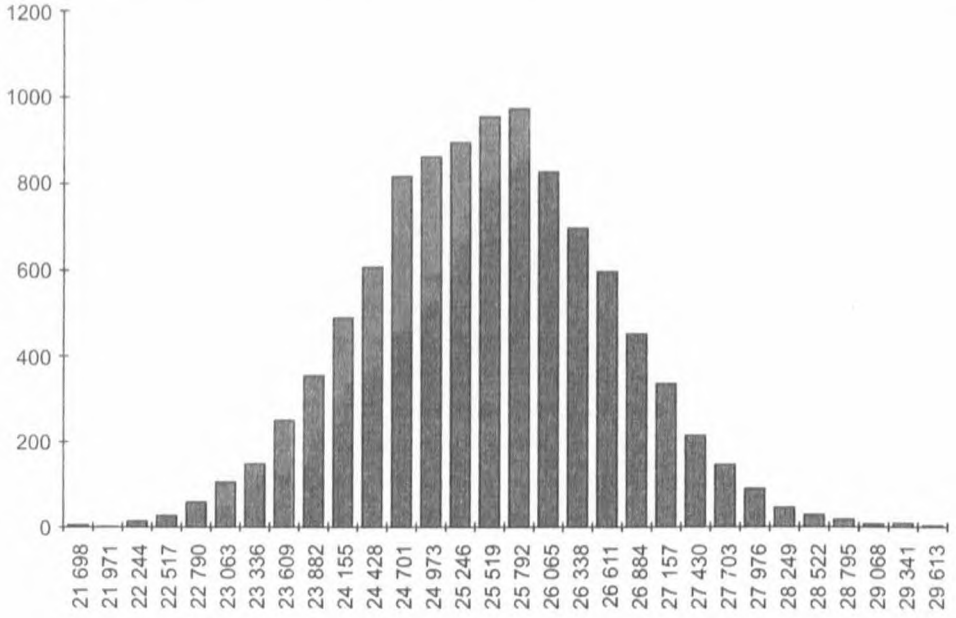
Rys. 4. Rozkład estymatora T_4 w przypadku losowania warstwowego zależnego – wariant IV w grupie eksperymentów A



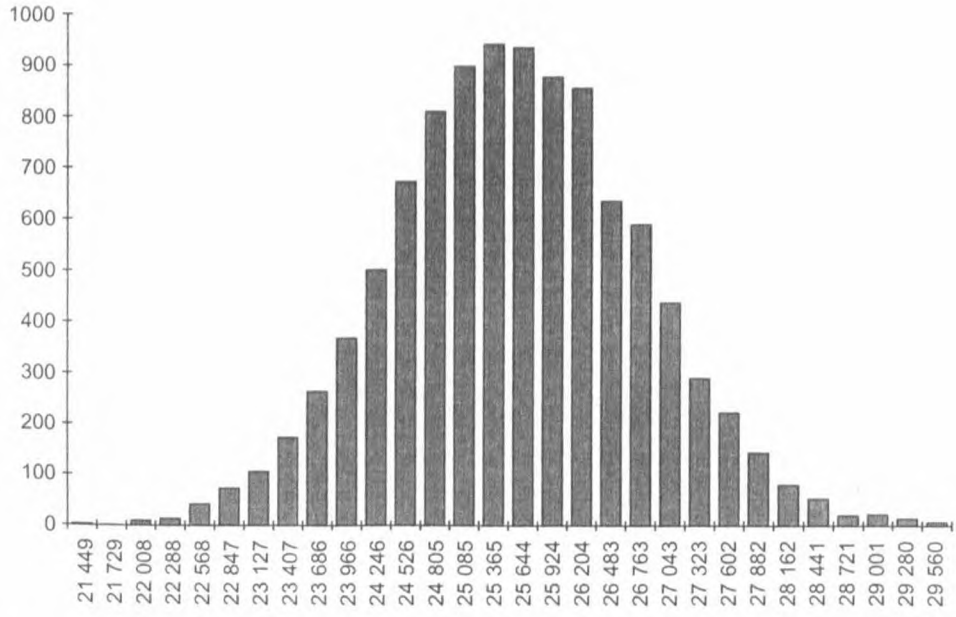
Rys. 5. Rozkład estymatora T_1 w przypadku warstwowania po wylosowaniu próby – wariant III w grupie eksperymentów B



Rys. 6. Rozkład estymatora T_4 w przypadku warstwowania po wylosowaniu próby – wariant III w grupie eksperymentów B



Rys. 7. Rozkład estymatora T_1 w przypadku losowania warstwowego zależnego – wariant IV w grupie eksperymentów B



Rys. 8. Rozkład estymatora T_4 w przypadku losowania warstwowego zależnego – wariant IV w grupie eksperymentów A

założeniu, dotyczącym oszacowań wartości globalnych cechy X na podstawie próby, tzn. $TX_{hg}/TX_{.g} \approx \hat{T}X_{hg}/\hat{T}X_{.g}$, gdzie $\hat{T}X_{hg} = \frac{N_{hg}}{n_{hg}} X_{hg}$, $\hat{T}X_{.g} = \frac{N_{.g}}{n_{.g}} X_{.g}$ (estymatory i ich wartości zostały oznaczone tymi samymi symbolami). Histogramy rozkładów estymatorów T_1 i T_4 dla III i IV wariantu uzyskane na podstawie próby o liczebności 2100 przedstawione są na rys. 1–4 w przypadku grupy eksperymentów A oraz na rys. 5–8 w przypadku grupy eksperymentów B.

Prowadzoną analizę rozszerzono o badanie zgodności rozkładu rozpatrywanych estymatorów z rozkładem normalnym. Wyznaczono w tym celu wskaźnik podobieństwa struktur oraz zastosowano test zgodności chi-kwadrat (wartość krytyczna dla poziomu istotności $\alpha = 0{,}05$ i 27 stopni swobody wynosi $\chi^2_\alpha = 40{,}113$). Wskaźnik podobieństwa struktur osiągał bardzo wysoką wartość (tzn. bliską jedności). W grupie eksperymentów A hipoteza, że rozkład estymatora jest normalny została odrzucona dwa razy w przypadku estymatora T_1, dwa razy w przypadku T_2, 10 razy dla T_3 i trzy razy dla T_4 na 16 przypadków dla każdego estymatora. W grupie eksperymentów B hipoteza ta została odrzucona cztery razy dla estymatora T_1, 12 razy dla estymatora T_3 i siedem razy dla estymatora T_4. W przypadku estymatora T_2 hipoteza nie była odrzucona ani razu (również na 16 przypadków). Wyniki te pozwalają sformułować przypuszczenie, że rozkład estymatora T_2 w przeprowadzonych eksperymentach jest zbliżony do normalnego, natomiast estymatora T_3 różni się istotnie od normalnego.

4. WNIOSKI KOŃCOWE

Analizy Monte Carlo dostarczają bardzo ciekawych informacji o własnościach estymatorów, które nie zawsze byłyby poznane na drodze teoretycznych rozważań lub badań empirycznych. Przeprowadzone eksperymenty prowadzą do wniosku, że dla skonstruowanych populacji rozważane estymatory charakteryzują się nieznacznym obciążeniem i niezbyt dużymi odchyleniami standardowymi, co świadczy o dużej efektywności tych estymatorów. Najbardziej efektywnym estymatorem okazał się estymator syntetyczny T_2, zaś najmniej efektywnym estymator T_3, różniący się od T_2 tym, że stosunek rzeczywistych wartości globalnych cechy pomocniczej X w wyróżnionym podzbiorze populacji zastąpiono stosunkiem odpowiednich wielkości globalnych w próbie. Estymator T_4 został skonstruowany na podstawie T_2 poprzez wprowadzenie w miejsce rzeczywistych wartości globalnych ich oszacowań z próby. Oceny wariancji estymatora T_4 okazały się nieco wyższe niż oceny wariancji estymatora bezpośredniego T_1. Wyniki

te pozwalają sformułować wniosek, że stosowanie estymatorów pośrednich, w których zastępujemy rzeczywiste wartości parametrów cechy pomocniczej ich oszacowaniami, nie zawsze prowadzi do zwiększenia efektywności estymacji. Spowodowane to może być małą precyzją oszacowań parametrów rozkładu zmiennej pomocniczej (parametry te można estymować różnymi metodami, a ponadto mimo dużej efektywności wybranych estymatorów błąd oszacowania uzyskanego na podstawie otrzymanych wyników próby może okazać się duży). Stosowanie estymatorów pośrednich, w których wykorzystywane są rzeczywiste wartości parametrów zmiennej pomocniczej, zgodnie np. z pracą Schaible i Casady (1994), prowadzi do uzyskania ocen parametrów z większą dokładnością niż przy zastosowaniu estymatorów bezpośrednich. Jeżeli nie mamy odpowiednich wartości parametrów cechy pomocniczej, to należałoby zastanowić się, czy nie lepiej wykorzystywać estymatory bezpośrednie, tym bardziej że koszty badania są wówczas mniejsze (nie trzeba zbierać informacji o dodatkowej zmiennej). Przy podejmowaniu tego typu decyzji istotna jest wielkość próby i własności estymatorów, za pomocą których szacowane są parametry zmiennej pomocniczej.

LITERATURA

Bracha Cz. (1994), *Metodologiczne aspekty badania małych obszarów*, Zakład Badań Statystyczno-Ekonomicznych GUS i PAN, 43, Warszawa.

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

Dol W. (1991), *Small Area Estimation. A Synthesis Between Sampling Theory and Econometrics*, Wolters Noordhoff, Groningen.

Golata E. (1996), *Statystyka małych obszarów w analizie pracy*, „Wiadomości Statystyczne", **3**, 45–60.

Kordos J. (1999), *Problemy estymacji danych dla małych obszarów*, „Wiadomości Statystyczne", **1**, 85–101.

Schaible W. L., Casady R. J. (1994), *The Development, Application and Evaluation of Small Area Estimators*, „Statistics of Transition", **1**, 6, 727–746.

Dorota Pekasiewicz, Krystyna Pruska

ANALYSIS OF DISTRIBUTIONS OF SOME ESTIMATORS IN SMALL AREA STATISTICS

(Summary)

Small area statistics is the part of statistics which deals with inference about subpopulations distinguished in the whole population.

In the paper four estimators of the total value for small area are considered. Their properties are analysed in the case of the stratified sampling and the stratification of sample. The investigation is conducted by Monte Carlo methods. The histograms and some characteristics of the estimator distributions (mean, standard deviation, coefficient of skewness) are determined.