

*Halina Klepacz\**, *Elżbieta Żółtowska\*\**

## WPLYW AGREGACJI DANYCH NA WARTOŚCI WSPÓLCZYNNIKA GINIEGO

**Streszczenie.** Artykuł dotyczy badania wpływu stopnia agregacji danych statystycznych na wielkość i poziom miar oraz wskaźników ubóstwa. Przeprowadzono eksperymenty numeryczne, których celem była analiza wartości współczynnika Giniego przed i po agregacji danych. Przygotowanie eksperymentu wymagało przesłania wyników różnych prac empirycznych.

**Słowa kluczowe:** ubóstwo, agregacja, współczynnik Giniego.

### 1. WPROWADZENIE

Jak wynika z raportów ONZ, ponad 50% ludności świata żyje w krajach biednych, w których dochód na jednego mieszkańca jest czterdziestokrotnie mniejszy niż w krajach bogatych.

W Polsce być może sytuacja nie jest tak drastyczna, ale oprócz rosnącej stale biedy zwiększa się systematycznie liczba osób żyjących na granicy ubóstwa.

Dlatego problem badania tego zjawiska jest niezwykle ważki, mimo że takie badania ani nie zmniejszają poziomu ubóstwa, ani go nie wyeliminują, ale poznanie jego rozmiaru i zasięgu być może przyczyni się do podjęcia przez rząd odpowiednich decyzji i środków zapobiegających temu zjawisku.

Wśród wielu mierników ubóstwa, takich jak: procent ubogich, stosunek luk dochodowych, indeksy: Sena, Kakwaniego i Thona (por. Biernacki 1999), krzywa Lorenza, uogólniona krzywa Lorenza, jest także współczynnik Giniego, definiowany jako iloraz przeciętnej absolutnej różnicy między dochodami poszczególnych gospodarstw i dochodu przeciętnego.

---

\* Dr, Katedra Ekonometrii Uniwersytetu Łódzkiego.

\*\* Prof. nadzw. dr hab., Katedra Ekonometrii Uniwersytetu Łódzkiego.

## 2. WSPÓŁCZYNNIK GINIEGO I JEGO WARTOŚCI W NIEKTÓRYCH KRAJACH

Z twierdzenia Atkinsona (1970) wynika, iż warunkiem koniecznym i wystarczającym do tego, by dobrobyt społeczny  $D_F$  był nie mniejszy niż dobrobyt społeczny  $D_G$  jest, aby rozkład  $F$  dominował w sensie Lorenza nad rozkładem  $G$ . Niekiedy takie porównanie nierówności rozkładów dochodów dla różnych grup społecznych czy różnych krajów jest wystarczające. Dlatego o dominacji w sensie Lorenza można wnioskować na podstawie współczynników Giniego (Kasprzyk 2001). Współczynnik ten odzwierciedla wtedy, w pewnym sensie, dobrobyt społeczny porównywanych grup społecznych czy krajów.

Współczynnik Giniego oznacza przeciętną absolutną różnicę między dochodami poszczególnych gospodarstw (danej grupy gospodarstw czy kraju) i jest definiowany następująco:

$$G = \frac{E(|X - Y|)}{2m_1},$$

gdzie:

$E(|X - Y|)$  – przeciętna absolutna różnica między dochodami poszczególnych gospodarstw;

$m_1$  – dochód przeciętny.

Jak można byłoby jeszcze inaczej zinterpretować ten współczynnik?

Otóż, z następujących przeliczeń:

$$\begin{aligned} E(X - Y) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |(x_i - \bar{x}) - (x_j - \bar{x})| \leq \\ &\leq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (|x_i - \bar{x}| + |x_j - \bar{x}|) = \\ &= \frac{1}{n^2} \left( \sum_{i=1}^n n |x_i - \bar{x}| + \sum_{i=1}^n \sum_{j=1}^n |x_j - \bar{x}| \right) = \frac{2}{n} \sum_{i=1}^n |x_i - \bar{x}| \end{aligned}$$

wynika, że

$$G = \frac{\sum(|X - Y|)}{m_1} \leq \frac{\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|}{\bar{x}} = \frac{V_x}{\bar{x}},$$

co oznacza, że współczynnik Giniego można interpretować również jako miarę zróżnicowania dochodów.

W krajach rozwiniętych gospodarczo współczynnik Giniego, dochodów po opodatkowaniu, jest rzędu 0,26–0,33. W krajach rozwijających się współczynnik ten osiąga wartości co najmniej równe 0,5, np. w Brazylii 0,60 (Sundrum 1990).

Tabela 1 przedstawia współczynnik Giniego dla krajów zachodnich, natomiast tabele 2 i 3 wartości tych współczynników dla krajów postkomunistycznych (podajemy odpowiednio za Kasprzyk 2001; Kwan S. Kim 1997), tabela 4 dla krajów Afryki bez Sahary i tabela 5 dla krajów Ameryki Łacińskiej.

Tabela 1. Współczynnik Giniego w niektórych krajach zachodnich

Kraj	Wielkość współczynnika	Ranga nierówności rozkładu dochodów (od max (G))
Stany Zjednoczone Ameryki Płn.	0,330	1
Australia	0,314	2
Kanada	0,306	3
Wielka Brytania	0,303	4
Niderlandy	0,302	5
Szwajcaria	0,292	6
Izrael	0,292	7
Niemcy	0,280	8
Szwecja	0,264	9
Norwegia	0,255	10

Źródło: Buhmann, Rainwater, Schmaus i Smeeding (1988); por. Kasprzyk (2001).

Tabela 2. Współczynniki Giniego w niektórych krajach Centralnej i Wschodniej Europy

Kraj	1989	1997
1	2	3
Białoruś	0,229	0,249
Bułgaria	0,331 (1992)	0,336
Estonia	0,280	0,361
Litwa	0,262	0,309
Łotwa	0,260	0,336 <sup>a</sup>
Mołdawia	0,250	0,390
Polska	0,275	0,334
Republika Czeska	0,198	0,239
Słowacja		0,249

Tabela 2 (cd.)

1	2	3
Rosja	0,265	0,385 <sup>a</sup> 0,483 <sup>b</sup>
Rumunia	0,237	0,305
Ukraina	0,229	0,413*
Węgry	0,225	0,254

<sup>a</sup> Na podstawie danych z roku 1995.

<sup>b</sup> Według danych wcześniejszych.

Źródło: UNICEF ICDC (1999).

Tabela 3. Zróżnicowanie dochodów i stopa ubóstwa w krajach Centralnej i Wschodniej Europy

Kraj	Współczynnik Giniego	Wzrost od 1987/1988	Stopa ubóstwa 1993	Stopa ubóstwa 1987/1988
Rosja	0,48	0,14-0,24 <sup>a</sup>	0,38	–
Polska	0,30	0,05	0,12	0,06
Węgry	0,23	0,02	0,02	0,01
Czechy	0,27	0,08	0,01	0,00
Słowenia	0,28	0,04	0,01	0,00
Estonia	0,39	0,16	0,23	–
Bułgaria	0,34	0,11	0,33	0,02

<sup>a</sup> Przedział wzrostu szacowany na podstawie różnych źródeł.

Źródło: dane Banku Światowego (por. Kwan S. Kim 1997).

Tabela 4. Standard życia i zróżnicowanie dochodów w krajach Afryki (bez Sahary)

Kraj	Parytet siły nabywczej produktu krajowego brutto na głowę (US = 100)		Współczynnik Giniego <sup>a</sup>
	1987	1994	
1	2	3	4
Ruanda	3,8	1,3	0,29
Tanzania	2,6	2,4	0,38
Uganda	5,0	5,4	0,41
Madagaskar	3,1	2,5	0,43
Niger	3,8	3,0	0,36
Kenia	5,7	5,1	0,58

Tabela 4 (cd.)

1	2	3	4
Nigeria	4,3	4,6	0,38
Zambia	4,1	3,3	0,46
Ghana	7,9	7,9	0,34
Mauretania	6,4	6,1	0,42
Zimbabwe	8,7	7,9	0,57
Senegal	7,3	6,1	0,54
Wybrzeże Kości Słoniowej	8,3	5,3	0,37
Lesoto	14,4	14,4	0,56

<sup>a</sup> Różne lata badań pomiędzy 1990–1993.

Źródło: dane Banku Światowego (por. Kwan S. Kim 1997).

Tabela 5. Dochód na głowę według dochodów na warstwę społeczną<sup>a</sup> w krajach Ameryki Łacińskiej

Kraj	Niski dochód 1992	Współczynnik wysokiego do niskiego dochodu 1992	Współczynnik Giniego	
			okres	zmiana
Argentyna (miejska)	1,40	13,11	1980–1989	0,410–0,478
Boliwia	0,52	18,37	1986–1989	0,518–0,528
Brazylia <sup>b</sup>	0,56	27,45	1979–1989	0,599–0,638
Chile	0,81	15,77	–	–
Kolumbia	0,58	18,14	1988–1991	0,531–0,513
Costa Rica	0,84	10,18	1981–1989	0,475–0,461
Honduras	0,28	17,71	1986–1989	0,553–0,595
Meksyk	0,79	14,47	1984–1989	0,508–0,591
Panama	0,64	18,19	1979–1989	0,409–0,568
Paragwaj (miejski)	0,65	11,25	1983–1990	0,453–0,399
Urugwaj	1,51	7,83	1981–1989	0,436–0,424
Wenezuela	0,71	11,89	1981–1989	0,430–0,443

<sup>a</sup> Dochód mierzony jest jako iloczyn odpowiedniej linii ubóstwa na głowę. Niskodochodowe warstwy społeczne reprezentują dolne 40%; a wysokodochodowe, górne 10%; <sup>b</sup> dla roku 1990.

Źródło: ECLAC (1994); dane Banku Światowego (por. Kwan S. Kim 1997).

## 3. ZAŁOŻENIA I DANE DO EKSPERYMENTÓW

Badanie wpływu stopnia agregacji danych statystycznych na wielkość i poziom miar oraz wskaźników ubóstwa można przeprowadzić na podstawie eksperymentu numerycznego. Jego przygotowanie wymagało prześledzenia stosownych miar i metod oraz analizy wyników różnych prac empirycznych. Wyniki tych studiów przedstawiono w opracowaniu Klepacz i Żółtowska (2002).

Podjęto decyzję, aby w pierwszej kolejności przeanalizować wpływ agregacji danych na wartości współczynnika Giniego (Klepacz i Żółtowska 2003).

Do eksperymentu przyjęto dane statystyczne, pochodzące z badań GUS dla roku 2001. Obejmują one:

a) średni dochód na osobę w gospodarstwach domowych, należących do danej grupy społeczno-zawodowej (pracowników na stanowiskach robotniczych, pracowników na stanowiskach nierobotniczych, pracowników użytkujących gospodarstwo rolne, rolników, pracujących na własny rachunek, emerytów i rencistów);

b) średnie wydatki na żywność i mieszkanie;

c) średnią liczbę osób w gospodarstwie domowym;

d) liczbę badanych budżetów w odpowiednich grupach społeczno-zawodowych.

Dane zawiera tabela 6.

Tabela 6. Dane do eksperymentu

Numer grupy	Średni dochód pomniejszony o wydatki na żywność i mieszkanie ( <i>per capita</i> )	Liczba zbadanych budżetów w roku 2001	Średnia liczba osób w gospodarstwie domowym	Średni dochód pomniejszony o wydatki na żywność i mieszkanie (na gospodarstwo)	% zbadanych budżetów
1	285,8	7098	3,68	1051,6	23,5
2	596,5	5513	3,11	1855,2	18,2
3	278,4	2518	4,52	1258,5	8,3
4	251,6	1318	4,21	1059,2	4,4
5	514,1	2058	3,64	1871,2	6,8
6	311,0	11711	2,26	702,8	38,8

Źródło: *Rocznik statystyczny* (2002).

Celem eksperymentów numerycznych jest odpowiedź na następujące pytania:

a) jakich wartości tych mierników można oczekiwać, gdy do ich obliczenia wykorzystana są dane niezagregowane?

b) jakie wartości mogą przyjąć te mierniki dla Polski, gdy do ich obliczenia wykorzystana są dane dotyczące całej populacji gospodarstw domowych?

W przeprowadzonych eksperymentach (Klepacz i Żółtowska 2003) przyjęto, że dochody gospodarstw występujących w wyróżnionych grupach będą losowane z rozkładu logarytmiczno-normalnego o takich średnich, jakie podano w tabeli 6. Przyjęcie takiego rozkładu wynikało z założenia, iż rozkład ten zależy od dwóch parametrów i jest rozkładem prawostronnie asymetrycznym, a ponadto, co jest niesłychanie ważne, istnieją skończone formuły podające zależności między parametrami rozkładu i momentami zmiennej losowej.

Zakładamy, że dochody gospodarstw domowych dla każdej z grup społeczno-zawodowych mają rozkłady logarytmiczno-normalne o wartościach oczekiwanych równych średnim zaobserwowanym w badaniach GUS-u dla roku 2001, czyli  $\bar{x} = E(X)$ .

Niestety nie są publikowane informacje o wielkościach miar rozrzutu dla dochodów w grupach społeczno-zawodowych, wobec czego dla celów eksperymentu zostały przyjęte współczynniki zmienności dochodów dla wyróżnionych grup gospodarstw domowych. Przyjmując zatem ustalone wartości dla współczynników zmienności  $V = \frac{D(X)}{\bar{x}}$ , wyznaczono parametry zmiennej losowej  $X \sim LN(\alpha, \sigma)$  ze wzorów:

$$\alpha = \ln \bar{x} - \frac{1}{2} \ln(V^2 + 1),$$

$$\sigma = \sqrt{\ln(V^2 + 1)}.$$

Wzory na dominantę i współczynnik asymetrii dla tego rozkładu mają postać:

$$M_o = \frac{\bar{x}}{(V^2 + 1)\sqrt{(V^2 + 1)}},$$

$$A_s = \frac{\bar{x} - M_o}{V\bar{x}} = \frac{1}{V} \left( 1 - \frac{1}{(V^2 + 1)\sqrt{(V^2 + 1)}} \right).$$

W eksperymentach przyjęto także, aby losować próby 10 000-elementowe z zachowaniem struktury dla poszczególnych grup, takiej jak w badaniach GUS-u z roku 2001. Na podstawie danych z tabeli 6 oraz podanych wzorów wyznaczono parametry dla funkcji gęstości (por. tabela 7) dla sześciu grup gospodarstw domowych.

Tabela 7. Parametry funkcji gęstości dla eksperymentu numerycznego

Grupa	Dochody średnie (setki zł)	Liczebność próby	V(%)	$\alpha$	$\sigma$	$M_0$	$A_1$
1	10,516	2350	30	2,3098	0,2936	9,241	0,404
2	18,552	1820	70	2,7212	0,6315	10,200	0,643
3	12,585	830	20	2,5129	0,1980	11,866	0,286
4	10,592	440	20	2,3405	0,1980	9,987	0,286
5	18,712	680	70	2,7298	0,6315	10,288	0,643
6	7,028	3880	60	1,7962	0,5545	4,431	0,616

Źródło: opracowanie własne.

Według założonych parametrów wylosowano z rozkładu logarytmiczno-normalnego sześć podpopulacji, a następnie połączono je w jedną populację. Dla każdej z nich wyznaczano interesujące nas wartości współczynników Giniego.

#### 4. WYNIKI EKSPERYMENTÓW

W tabelach 8–10 podane są wyniki, z wybranych dziesięciu 10 000-elementowych prób. Eksperymenty zostały przeprowadzone z wykorzystaniem pakietu *Mathematica*.

Dla każdej z podpopulacji wyznaczono wartości współczynników Giniego (por. tabela 9), a następnie wyznaczono dwa współczynniki dla każdej z populacji: współczynnik przed agregacją  $G$  oraz współczynnik  $G_a$  po agregacji dochodów w populacjach, na podstawie łącznych dochodów w sześciu wyróżnionych podpopulacjach. Zgodnie z oczekiwaniem wartości zagregowanych współczynników  $G_a$  są niższe niż współczynnika  $G$ , ponieważ w wyniku agregacji tracimy część informacji o zmienności dochodów. W przeprowadzonych eksperymentach zagregowany współczynnik Giniego był niższy od rzeczywistego współczynnika Giniego dla populacji od 6,8% do 13%, średnia wartość tego współczynnika w przeprowadzonych eksperymentach była niższa o 9,3% od wartości  $G$ .



Tabela 8. Średnie dochody w wybranych próbach

Numer eksperymentu	Średni dochód z próby (w setkach złotych) dla						
	populacji	podpopulacji					
		1	2	3	4	5	6
1	11,216	10,273	18,456	12,570	10,058	19,448	6,792
2	11,341	10,553	18,720	12,694	10,318	17,418	7,120
3	11,105	10,551	17,730	12,933	10,362	18,029	6,812
4	11,355	10,510	19,594	12,308	10,194	18,141	6,741
5	11,055	10,420	17,835	12,299	10,610	17,825	6,793
6	11,214	10,596	17,793	12,696	10,411	17,480	7,179
7	11,559	10,408	17,908	12,956	11,209	19,884	7,559
8	11,180	10,543	16,845	13,005	10,261	18,900	7,270
9	11,012	10,722	18,756	12,365	10,487	15,144	6,602
10	11,034	10,285	16,822	12,402	10,919	19,972	6,929
założona wartość	<b>11,216</b>	<b>10,516</b>	<b>18,552</b>	<b>12,585</b>	<b>10,592</b>	<b>18,712</b>	<b>7,028</b>

Źródło: opracowanie własne.

Tablica 9. Wartości współczynników Giniego w przykładowych eksperymentach

Numer eksperymentu	Wartości współczynników Giniego dla							
	populacji		podpopulacji					
	$G$	$G_a$	1	2	3	4	5	6
1	0,3385	0,3039	0,1646	0,3538	0,1076	0,0420	0,3173	0,2964
2	0,3322	0,3114	0,1593	0,3450	0,1123	0,1009	0,3426	0,2983
3	0,3312	0,2974	0,1643	0,3582	0,1102	0,1016	0,2958	0,2884
4	0,3482	0,3171	0,1653	0,3592	0,1100	0,1221	0,3043	0,3040
5	0,3300	0,3001	0,1505	0,3414	0,1182	0,1069	0,3412	0,2845
6	0,3219	0,3032	0,1707	0,3383	0,1115	0,1113	0,3333	0,2850
7	0,3350	0,2955	0,1611	0,3216	0,1069	0,1108	0,3650	0,3330
8	0,3308	0,2905	0,1624	0,3451	0,1174	0,0938	0,3581	0,3114
9	0,3324	0,3156	0,1507	0,3493	0,1025	0,1154	0,2729	0,2978
10	0,3289	0,2861	0,1501	0,3541	0,1029	0,1025	0,2837	0,3035
Średnia	<b>0,3330</b>	<b>0,3021</b>	<b>0,1599</b>	<b>0,3466</b>	<b>0,1100</b>	<b>0,1007</b>	<b>0,3214</b>	<b>0,2992</b>

Źródło: opracowanie własne.

Analizowano także wpływ zmienności dochodów w wybranej grupie 6, przyjmując różne poziomy współczynnika zmienności dochodów dla tej grupy. Była to grupa najliczniejsza, jej dochody były najniższe, zatem analiza tego przypadku może pokazać, ile tracimy na wartości współczynnika Giniego, agregując dane. Wyznaczono wartości współczynników: bez agregacji  $G$ , po agregacji  $G_a$  i dla wyróżnionej grupy  $G_6$  (por. tabela 10).

Tabela 10. Wartości współczynników Giniego

V	G	$G_a$	$G_6$
0,4	0,3051	0,2883	0,2248
0,5	0,3199	0,2869	0,2669
0,6	0,3289	0,2861	0,3035
0,7	0,3414	0,2886	0,3596
0,8	0,3597	0,2813	0,3813
0,9	0,3524	0,2883	0,4033

Źródło: opracowanie własne.

Zauważmy przede wszystkim, że im wyższa zmienność dochodów w wyróżnionej grupie, tym wyższe wartości mają współczynniki  $G$ ,  $G_6$ . Natomiast wartości współczynnika po agregacji danych  $G_a$  praktycznie nie wykazują zmian. Oznacza to, że posługując się zagregowanymi danymi do wyznaczania współczynnika Giniego, musimy być bardzo ostrożni w wyciąganiu wniosków dotyczących poziomu ubóstwa przy porównywaniu różnych populacji. Zatem miara ta nie może być stosowana jako jedyna przy badaniu tego zjawiska.

#### LITERATURA

- Atkinson S. N. (1970), *On the Measurement of Inequality*, „Journal of Economic Theory”, 2.
- Biernacki M. (1999), *Pomiar ubóstwa*, „Ekonomia Matematyczna”, 3, Wydawnictwo Akademii Ekonomicznej, Wrocław.
- Buhmann B., Rainwater L., Schmaus G., Smeeding T. M. (1988), *Equivalence Scales, Well-being, Inequality and Poverty: Sensitive Estimates across Ten Countries Using the Luxembourg Income Study (LIS) Database*, „Review of Income and Wealth”, 2, 121.
- Kasprzyk B. (2001), *Zróżnicowanie nierówności rozkładu dochodów grup społeczno-ekonomicznych*, „Wiadomości Statystyczne”, 2.
- Klepacz H., Żółtowska E. (2002), *Zastosowanie metod ilościowych do badania pomiaru ubóstwa. Wpływ stopnia zagregowania danych statystycznych na poziom mierników ubóstwa*, opracowanie niepublikowane, Łódź.

- Klepacz H., Żółtowska E. (2003), *Zastosowanie metod ilościowych do badania pomiaru ubóstwa. Poziom mierników ubóstwa a agregacja danych statystycznych – eksperymenty numeryczne i badania empiryczne*, opracowanie niepublikowane, Łódź.
- Klepacz H., Żółtowska E. (2003), *Zróżnicowanie nierówności dochodowych w grupach społeczno-zawodowych*, „Zeszyt Naukowy, Instytut Ekonomii Filii Akademii Świętokrzyskiej w Piotrkowie Trybunalskim”, 4.
- Kwan S. Kim (1997), *Income Distribution and Poverty: An Interregional Comparison*, „World Development”, 25 (11).
- Orłowski L. T. (2001), *Transition and Growth in Post-Communist Countries*, Edward Elgar Publishing, Cheltenham–Northampton.
- Rocznik statystyczny* (2002), GUS, Warszawa.
- Sundrum R. M. (1980), *Income Distribution in Less Developed Countries*, Routledge, London–New York.
- UNICEF ICDC (1999), (International Child Development Center UC DC), *Women in Transition*, MONEE Project CEE/CIS/Baltics, „Regional Monitoring Report”, 6, Florence.

Halina Klepacz, Elżbieta Żółtowska

#### AN INFLUENCE OF DATA AGGREGATION ON GINI RATIO VALUE

##### Summary

We study an influence of statistical data aggregation rate on value and level of poverty factors. Numerical experiments were conducted to analyze Gini ratio values prior and posterior to data aggregation. An experiment required a thorough study of various empirical results.