Część 2

ZASTOSOWANIE METOD STATYSTYCZNYCH W EKONOMII I ZARZĄDZANIU

Iwona Roeske-Słomka

ROZKŁADY GOSPODARSTW DOMOWYCH PRACOWNICZYCH WEDŁUG DOCHODU NA PODSTAWIE MIAR ENTROPIJNYCH

Problem rozkładu dochodów należy w ekonomii do najważniejszych poznawczo zagadnień badawczych. Stanowi bowiem podstawę charakterystyki i oceny warunków bytu ludności.

W empirycznym badaniu rozkładu danej cechy (dochodu) można wyróżnić dwa podejścia:[1]

– analizę rozkładu tej cechy na poszczególne jednostki statystyczne, a więc jakoby bezpośrednie rozpatrywanie rozłożenia ogólnej sumy wartości badanej cechy pomiędzy poszczególne jednostki zbiorowości,

– analizę rozkładu jednostek statystycznych (np. osób bądź gospodarstw domowych) na – wyodrębnione – warianty badanej cechy, w wyniku której

---

[1] Na rozróżnienie problemu koncentracji sum wartości przy pewnych jednostkach od skupiania się jednostek przy pewnych wartościach zwraca uwagę B. Szulc. Zob. **B. Szulc,** Statystyka dla ekonomistów, t. 1, PWE, Warszawa 1972, s. 250.

istnieje możliwość pośredniego wnioskowania o rozkładzie badanej cechy w zbiorowości statystycznej.

W procesie postępowania empirycznego pierwszy sposób podejścia – z formalnego punktu widzenia – możliwy jest w przypadku dysponowania danymi o wartościach badanej zmiennej każdej jednostki statystycznej (tzn. dane występują w postaci szeregów statystycznych prostych) bądź przynajmniej pewnych grup jednostek (dane empiryczne w formie szeregów rozdzielczych; w przypadku badania dochodów na ogół w postaci szeregów wielowariantowych). W sytuacji posiadania informacji o dochodach każdej jednostki statystycznej do badania zróżnicowania dochodów w analizie ich rozkładów można stosować tzw. wskaźnik Giniego[2]:

$$G = \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} |x_i - x_j|}{2n^2 m}, \qquad (1)$$

gdzie x – dochód
m – dochód średni
n – liczebność zbiorowości.

Popularnymi miernikami analizy rozkładu dochodów, zwłaszcza w sytuacji dysponowania danymi w formie wielowariantowych szeregów rozdzielczych, są parametry pozycyjne. Ich znaczenie jest szczególnie ważne w tych przypadkach, w których informacja statystyczna ujęta jest w postaci szeregów otwartych, jednakże pod tym warunkiem, że wartości tych parametrów (dotyczy to zwłaszcza kwartyli) nie przypadają na otwarte przedziały klasowe. W przeciwnym bowiem przypadku niemożliwa jest ocena zróżnicowania cechy na podstawie odchylenia ćwiartkowego. Ponadto należy wskazać na mankamenty wynikające z wyłącznego posługiwania się w analizie parametrami pozycyjnymi, wynikającymi stąd, iż ich wartości opierają się nie na wszystkich wartościach zmiennej – w odróżnieniu od parametrów klasycznych, które jednakże mogą być stosowane tylko w przypadku szeregów zamkniętych.

Drugie z wymienionych podejść analitycznych, z formalnego punktu widzenia – w świetle propozycji, jaką podaje się w niniejszym artykule – możliwe jest do zastosowania praktycznie w każdej sytuacji, gdyż opiera się na funkcji entropii:

[2] Zob. **H. Góralska**, Zmiany rozpiętości i zróżnicowania dochodów w Polsce w latach 1973-1978, Instytut Pracy i Spraw Socjalnych, Warszawa 1982.

$$H = -\sum_i p_i \log_2 p_i, \tag{2}$$

w której p_i jest stosunkiem liczby jednostek statystycznych o i-tym wariancie badanej cechy do ogólnej liczebności zbiorowości. Zatem w odniesieniu do rozkładu dochodu p_i jest wynikiem stosunku liczby jednostek statystycznych o dochodach określonych przez daną klasę i do ogólnej liczby jednostek statystycznych badanej zbiorowości.

Górną granicę wartości, jaką może przyjąć entropia danego rozkładu liczebności, stanowi tzw. entropia maksymalna:

$$H_{\max} = \log_2 n, \tag{3}$$

gdzie n jest liczbą wariantów danej cechy, a więc liczbą wyodrębnionych klas dochodowych.

Zauważmy, że funkcja (2) osiąga wartość maksymalną, gdy liczby jednostek statystycznych przypadające na poszczególne wyodrębnione warianty badanej cechy (wyodrębnione klasy dochodu) są sobie równe. Występuje wówczas (całkowicie) równomierny rozkład jednostek statystycznych na wyodrębnione warianty danej cechy. Ponieważ warianty badanej cechy (przedziały klasowe) są uporządkowane (od najniższych do najwyższych klas wartości zmiennej), przeto rozkład równomierny w ujęciu entropijnym oznacza maksymalne zróżnicowanie badanej cechy w zbiorowości. I podobnie – największa koncentracja, czyli najmniejsze zróżnicowanie jednostek statystycznych z punktu widzenia badanej cechy, występuje wówczas gdy wszystkie jednostki statystyczne koncentrują się na tym samym wariancie badanej cechy (należą do tego samego przedziału klasowego). Zatem następujący współczynnik może być miarą zróżnicowania:

$$WH = 1 - \frac{H}{H_{\max}}. \tag{4}$$

Jeżeli zatem $WH = 0$, to rozkład jednostek statystycznych na warianty badanej cechy jest równomierny, co oznacza – biorąc pod uwagę fakt uporządkowania zmiennej od najniższej do najwyższej wartości – maksymalne zróżnicowanie tej cechy w zbiorowości. Natomiast przypadek: $WH = 1$ oznacza, iż wszystkie jednostki statystyczne wykazują się tym samym wyodrębnionym wariantem cechy (należą do tego samego przedziału klasowego). Tak więc współczynnik WH, podobnie jak współczynnik G, jest także miarą unormowaną, a więc łatwą w ocenie wartości, jaką przyjmuje. Jednakże pojmowanie określenia

równomierności w kontekście obu tych współczynników jest odmienne. Równomierność rozkładu w ujęciu wskaźnika Gini oznacza równomierny rozkład dochodu na poszczególne jednostki statystyczne, natomiast równomierność rozkładu w ujęciu współczynnika entropijnego oznacza równomierność rozkładu jednostek statystycznych na poszczególne, wyodrębnione warianty (klasy) cechy, co ze względu na jej rosnące uporządkowanie oznacza maksymalne zróżnicowanie tej cechy w zbiorowości.

Podobne problemy jak przy badaniu zróżnicowania zmiennej ujętej w szeregach rozdzielczych (otwartych) pojawiają się przy zagadnieniu badania asymetrii rozkładu. W tym względzie funkcja entropii może także znaleźć zastosowanie. Ze względu na to, że funkcja entropii charakteryzuje się tym, iż przy niektórych różnych wartościach występowania wariantów badanej cechy składniki ogólnej funkcji entropii (tj. $p_i \log_2 p_i$) mogą przyjąć te same wartości[3], przeto we wskaźnikach asymetrii (których idea budowy polega na tym, że stanowią one różnice między udziałami liczebności pierwszych i ostatnich klas szeregu statystycznego w jego ogólnej funkcji entropii) tzw. cząstkowe współczynniki entropii:

$$w_i h = \frac{\left| p_i \log_2 p_i \right|}{H_{\max}} \tag{5}$$

mnoży się przez stosunki liczebności klas do liczebności ogólnej (p_i).

Dla szeregu statystycznego o nieparzystej liczbie klas asymetrię określamy na podstawie wzoru:

$$_H As = \sum_{i=1}^{k-1} p_i w_i h - \sum_{i=k+1}^{n} p_i w_i h, \tag{6}$$

w którym k oznacza numer klasy środkowej.

Natomiast w przypadku szeregu o parzystej liczbie klas, asymetrię określa wzór:

$$_H As = \sum_{i=1}^{k_1} p_i w_i h - \sum_{i=k_2}^{n} p_i w_i h, \tag{7}$$

w którym k_1 i k_2 oznaczają pierwszą i drugą klasę środkową.

[3] Zob. **I. Roeske-Słomka**, Rozkłady cech a rozkłady liczebności – propozycja analizy. Rozwój metodologii badań statystycznych w Polsce, PTS, Warszawa 1995, s. 144.

Ze względu na to, iż w szeregu statystycznym warianty danej cechy uporządkowane są rosnąco, tj. od najniższej do najwyższej jej wartości, interpretacja tzw. entropijnych wskaźników asymetrii jest taka sama jak tzw. wskaźników klasycznych czy pozycyjnych.[4] I tak, dodatnią wartość wskaźników asymetrii (6) lub (7) nazywamy asymetrią prawostronną i oznacza ona, iż dominująca liczba jednostek zbiorowości statystycznej charakteryzuje się niższymi od przeciętnej wartościami badanej zmiennej. W przypadku ujemnych wartości wskaźników asymetrii (6) lub (7) mówimy o występowaniu asymetrii lewostronnej, która oznacza, iż dominująca liczba jednostek zbiorowości statystycznej charakteryzuje się wyższymi od przeciętnej wartościami badanej zmiennej. Jeżeli wskaźniki (6) lub (7) są równe zero, wówczas mówimy, że rozkład badanej zmiennej jest symetryczny, co oznacza, że dominująca liczba jednostek zbiorowości statystycznej charakteryzuje się średnimi wartościami badanej zmiennej.

Charakterystykę rozkładu jednostek statystycznych na warianty danej cechy, z punktu widzenia jej zróżnicowania i asymetrii, w ujęciu entropijnym można także przedstawić graficznie. Do tego celu służy kwadrat o boku równym jeden, w którym na osi poziomej odpowiadającej osi X, zaznaczamy kumulowane częstości występowania wyodrębnionych wariantów danej cechy, a na osi pionowej Y, kumulowane cząstkowe współczynniki entropii ($w_i h$). Połączenie punktów o współrzędnych: kumulowanych częstości występowania wariantów badanej cechy i kumulowanych cząstkowych współczynników entropii, daje wykres linii entropii, która – wyprowadzona z początku układu – kończy się na boku pionowym kwadratu (y') równoległym do boku reprezentującego oś Y. Punkt przecięcia linii entropii z bokiem y' wyznacza wartość stosunku entropii empirycznej do maksymalnej, który jest podstawą wyznaczenia ogólnego współczynnika entropii WH, charakteryzującego zróżnicowanie danej cechy w zbiorowości. Połączenie punktu $\frac{H}{H_{max}}$ z początkiem układu osi XY daje linię, którą nazywamy linią empirycznej symetrii rozkładu. Jeżeli linia entropii przebiega poniżej linii empirycznej symetrii rozkładu, to rozkład danej cechy statystycznej w zbiorowości jest prawostronny (o asymetrii dodatniej); jeżeli zaś linia entropii przebiega ponad linią empirycznej symetrii rozkładu, to rozkład tej cechy jest lewostronny (o asymetrii ujemnej). Równomierny przebieg linii entropii po obu stronach empirycznej linii symetrii oznacza, iż rozkład danej cechy w zbiorowości statystycznej jest symetryczny.

[4] Zob. **I. Roeske-Słomka**, Wielkość i entropia gospodarstw domowych w Polsce (w świetle powojennych spisów ludności), AE, Poznań 1994.

Jak stwierdzono, podstawowym materiałem źródłowym pozwalającym na analizę rozkładu dochodu, są badania budżetów gospodarstw domowych. Badania budżetów gospodarstw domowych prowadzone w Polsce przed 1973 rokiem miały charakter incydentalny i dotyczyły tylko wybranych grup ludności.

Można przyjąć, iż dopiero od 1973 r. badania budżetów gospodarstw domowych mają charakter badań reprezentacyjnych, przy czym w ciągu kolejnych lat, w celu zwiększenia reprezentatywności próby dla ogółu gospodarstw domowych w Polsce, do obserwacji budżetów gospodarstw domowych wprowadzano zmiany, z których najważniejszą było odejście od ciągłej techniki obserwacji do rotacyjnej. Badania budżetów gospodarstw domowych prowadzone w latach 1977 i 1978 zostały uznane przez Główny Urząd Statystyczny za nieudane i z tego powodu ich wyników w zasadzie nie opublikowano.

Analizę rozkładu tzw. pracowniczych gospodarstw domowych (tj. utrzymujących się głównie z pracy w gospodarce uspołecznionej) według wysokości dochodu przypadającego na jedną osobę w gospodarstwie domowym, z punktu widzenia zróżnicowania i asymetrii rozkładu, przeprowadzono na podstawie opublikowanych danych pochodzących z ogólnopolskich badań budżetów gospodarstw domowych w latach 1973 - 1992.[5]

Podstawą obliczeń entropijnych współczynników koncentracji i asymetrii dla poszczególnych lat były liczby pracowniczych gospodarstw domowych należących do wyodrębnionych klas dochodu przypadającego na osobę. Do roku 1981 wyodrębniano w publikacjach siedem grup dochodowych, a począwszy od 1982 r. – osiem. W kolejnych latach dolne i górne granice wszystkich przedziałów dochodowych określał oczywiście coraz wyższy dochód nominalny.

Generalnie można zaobserwować niskie wartości entropijnych współczynników koncentracji (zob. tab 1), co oznacza dość równomierny rozkład pracowniczych gospodarstw domowych na poszczególne grupy dochodowe, a więc znaczne zróżnicowanie tych gospodarstw pod względem dochodu przypadającego na osobę.

Stosunkowo najwyższe współczynniki koncentracji odnotowano dla roku 1979 i 1988, co oznacza, iż w tych latach zróżnicowanie dochodów przypadających na osobę w pracowniczych gospodarstwach domowych było stosunkowo najniższe. Natomiast najwyższe zróżnicowanie wysokości dochodów przypadających na osobę w pracowniczych gospodarstwach domowych wystąpiło w latach 1982 i 1986.

[5] Od 1993 r. informacje o dochodach gospodarstw domowych pochodzące z ogólnopolskich badań budżetów publikowane są w zupełnie innym ujęciu.

Znamienne jest to, iż w roku 1979 – o najwyższym współczynniku koncentracji, wystąpiła asymetria lewostronna (krzywa przebiega ponad linią empirycznej symetrii rozkładu) – zob. rys. 1 oraz tab. 1, co oznacza, że dominująca liczba pracowniczych gospodarstw domowych miała dochody przypadające na osobę wyższe od przeciętnego dochodu. Natomiast w roku 1988, o takim samym, a więc najwyższym współczynniku koncentracji, wystąpiła asymetria prawostronna (krzywa entropii przebiega poniżej empirycznej linii symetrii) – zob. rys. 4 i tab. 1, która oznacza pojawienie się zjawiska odmiennego, polegającego na tym, że dominująca liczba pracowniczych gospodarstw domowych miała w tym roku dochód przypadający na osobę niższy od przeciętnego.

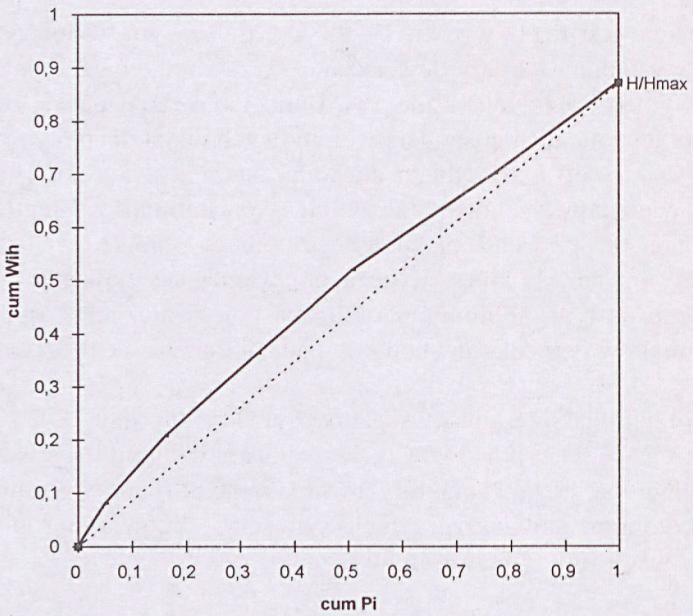
Najbardziej bliski zeru jest wskaźnik asymetrii dla roku 1992 – zob. tab 1. Oznacza to, iż rozkład pracowniczych gospodarstw domowych według wysokości dochodu na głowę charakteryzował się w tym roku, w porównaniu z latami poprzednimi, największą symetrycznością – krzywa entropii przebiega po obu stronach linii symetrii empirycznej – rys. 5.

Tab. 1. Entropijna koncentracja i asymetria rozkładu gospodarstw domowych pracowniczych według dochodu w latach 1973 - 1992

Rok	Współczynnik koncentracji	Wskaźnik asymetrii	Rok	Współczynnik koncentracji	Wskaźnik asymetrii
1973	0,04	0,0106	1984	0,04	–0,0629
1974	0,05	–0,0324	1985	0,07	–0,0381
1975	0,08	–0,0701	1986	0,03	–0,0338
1976	0,10	–0,0951	1987	0,06	0,0414
1979	0,13	–0,0977	1988	0,13	0,0797
1980	0,06	–0,0656	1989	0,12	0,0294
1981	0,06	0,0350	1990	0,05	–0,0246
1982	0,03	0,0264	1991	0,05	0,0310
1983	0,04	–0,0504	1992	0,06	0,0095

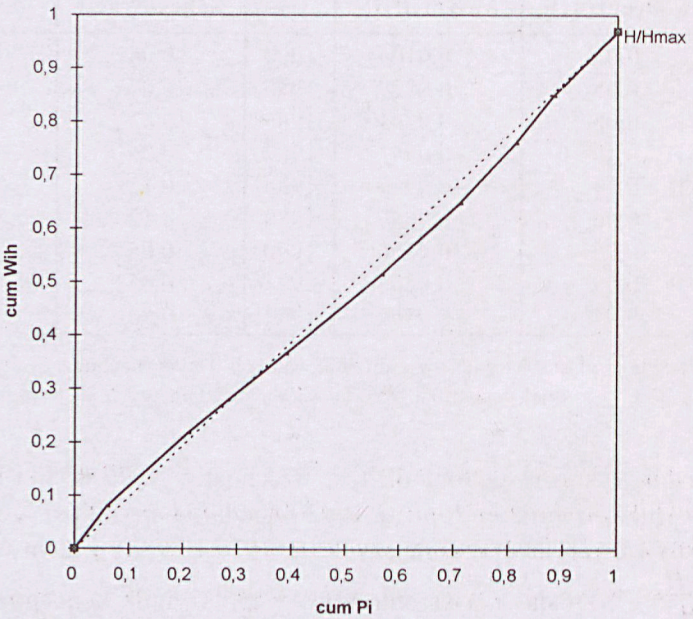
Źródło: Obliczenia własne na podstawie opublikowanych danych pochodzących z ogólnopolskich badań budżetów gospodarstw domowych.

Z porównania wykresów entropii dla tych trzech lat, tj. 1979, 1988 i 1992 – rys. 1, 4 i 5 wynika, iż linia entropii w 1992 r. oddaliła się od osi X w stosunku do porównywanych lat wcześniejszych (tj. 1979 i 1988). Zatem wartość stosunku $\frac{H}{H_{max}}$ wzrosła, a więc entropijny współczynnik koncentracji spadł – zob. także tab. 1. Oznacza to, iż rozkład pracowniczych gospodarstw domowych na poszczególne klasy dochodu na głowę był bardziej równomierny



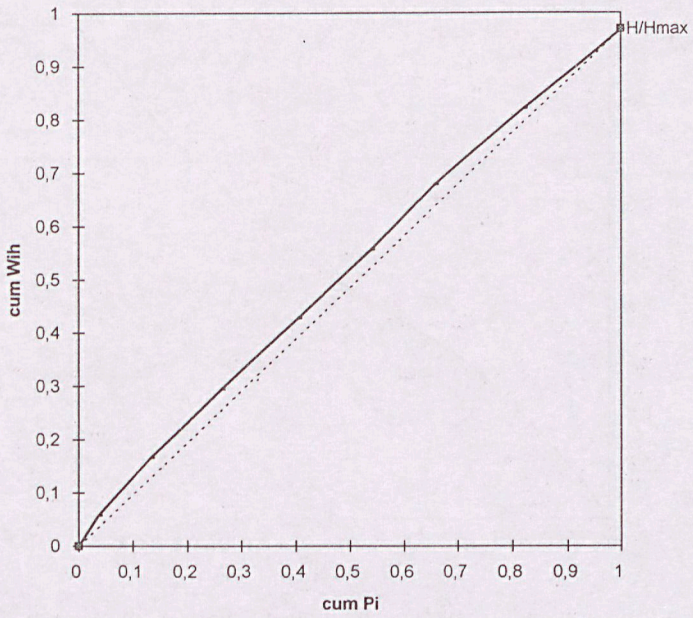
Rys. 1. Entropia rozkładu pracowniczych gospodarstw domowych w 1979 r.

Źródło: Opracowanie własne na podstawie opublikowanych danych pochodzących z ogólnopolskich badań budżetów gospodarstw domowych



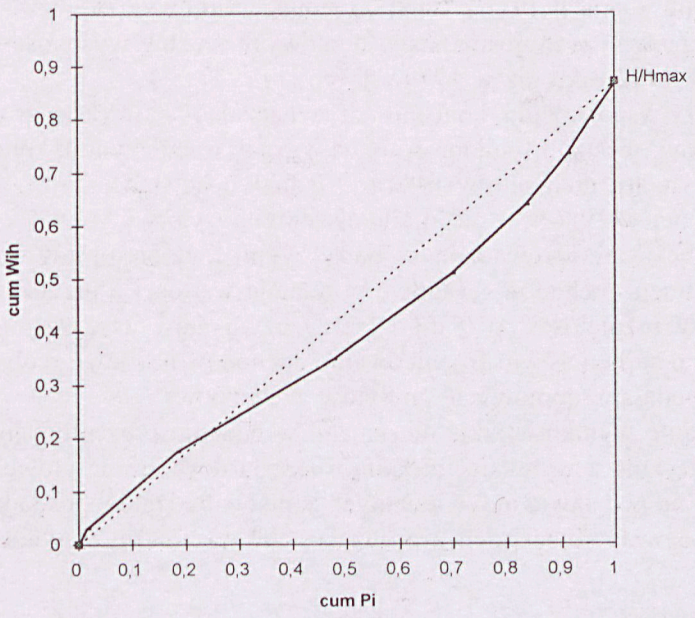
Rys. 2. Entropia rozkładu pracowniczych gospodarstw domowych w 1982 r.

Źródło: Zob. rys. 1



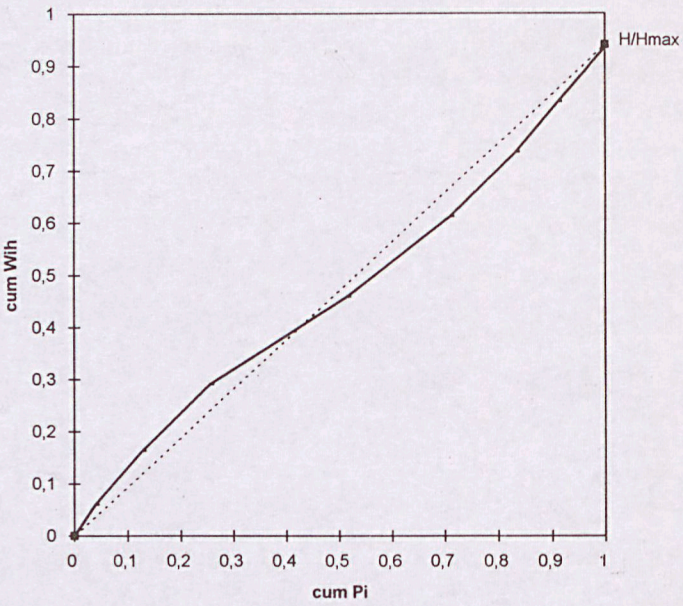
Rys. 3. Entropia rozkładu pracowniczych gospodarstw domowych w 1986 r.

Źródło: Zob. rys. 1



Rys. 4. Entropia rozkładu pracowniczych gospodarstw domowych w 1988 r.

Źródło: Zob. rys. 1



Rys. 5. Entropia rozkładu pracowniczych gospodarstw domowych w 1992 r.

Źródło: Zob. rys. 1

w 1992 r. niż w latach 1973 i 1988, co jednocześnie wyraża większe zróżnicowanie pracowniczych gospodarstw domowych według wysokości dochodu na głowę w 1992 roku niż w 1979 i 1988.

Najniższy współczynnik koncentracji, oznaczający najwyższe zróżnicowanie dochodów, wystąpił (podobnie jak najwyższy współczynnik koncentracji) raz przy asymetrii dodatniej w 1982 r. – jednak dość słabej – i raz przy asymetrii ujemnej w 1986 r. – nieco silniejszej, zob. rys. 4 i 5.

Z dotychczas przeprowadzonych badań wynika, że entropijny współczynnik koncentracji dochodów na głowę przyjmuje wartości z przedziału 0,03 - - 0,30 (przy rozpiętości od 0 do 1), przy czym najwyższe współczynniki, świadczące o najmniejszym zróżnicowaniu dochodów na głowę zaobserwowano dla gospodarstw domowych emerytów i rencistów.[6]

Porównanie wyników, jakie otrzymano w badaniu prezentowanym w niniejszym artykule z wynikami badania rozkładu dochodu na głowę przeprowadzonego na podstawie indywidualnych zapisów budżetów gospodarstw domowych, pozwala stwierdzić, iż prowadzą one do zbieżnych wniosków. Po

[6] Wyniki prezentowane w niniejszym artykule stanowią fragment szerszych badań prowadzonych także dla pozostałych typów gospodarstw domowych.

pierwsze: wskaźnik Giniego obliczony przez H. Góralską[7] dla wysokości dochodu przypadającego na osobę w gospodarstwach domowych pracowniczych był wyższy niż dla gospodarstw domowych emerytów i rencistów; po drugie: na podstawie miernika entropijnego stwierdzono wyższe zróżnicowanie pod względem wysokości dochodu na głowę gospodarstw domowych pracowniczych niż gospodarstw domowych emerytów i rencistów.

[7] Zob. **H. Góralska**, Zmiany rozpiętości i zróżnicowanie dochodów w Polsce w latach 1973 - 1978, Instytut Pracy i Spraw Socjalnych, Warszawa 1982, z. 4 (168), s. 26 i 35.