Iwona ROESKE-SŁOMKA

Rozkłady cech a rozkłady liczebności na przykładzie gospodarstw domowych

Pochodzące ze spisów ludności dane, dotyczące np. liczebności gospodarstw domowych według: liczby osób, liczby czynnych zawodowo oraz łącznie liczby osób i liczby czynnych zawodowo, publikowane są w postaci otwartych szeregów statystycznych. Fakt ten uniemożliwia wyznaczenie podstawowych parametrów charakteryzujących rozkład cech w zbiorowości statystycznej takich, jak: wartość średnia i odchylenie standardowe. Możliwe jest jednak badanie koncentracji np. za pomocą współczynnika:

$$(1) \qquad k=\frac{5000-M}{5000},$$

gdzie M jest sumą iloczynów odsetka jednostek statystycznych przypadających na określone wartości zmiennej i kumulowanych odsetek wartości zmiennej odpowiadających tym jednostkom. Współczynnik ten, jak wiadomo, przyjmuje wartości z przedziału domkniętego [0, 1]. Jeżeli $k=0$, to rozkład jest równomierny, jeżeli $k=1$ to koncentracja jest zupełna, co oznacza, iż całą sumą wartości cechy dysponuje tylko jedna jednostka populacji.

Do analizy koncentracji, a także symetrii rozkładu, można również wykorzystać pojęcie entropii. Entropię rzeczywistą (empiryczną) wyznaczamy na podstawie następującego wzoru:

$$(2) \qquad H=-\sum_i p_i \log_2 p_i,$$

w którym p_i oznacza częstość występowania i-tego stanu danej zmiennej, czyli p_i jest po prostu stosunkiem liczby jednostek statystycznych o i-tym wariancie rozpatrywanej cechy do ogólnej liczby jednostek.

Entropia równa zeru oznacza, iż wszystkie jednostki statystyczne przyjmują ten sam wariant wartości zmiennej. Im zaś jest wyższa wartość entropii empirycznej, tym wyższy jest stopień nieuporządkowania sytuacji czy elementów. Górną granicę wartości, jaką może przyjąć entropia empiryczna, stanowi tzw. entropia maksymalna, którą wyznacza wzór:

$$(3) \qquad H_{max} = log_2 n,$$

w którym n jest liczbą wariantów danej cechy[1].

Równość entropii empirycznej z maksymalną oznacza, że wszystkie możliwe stany danej zmiennej losowej są jednakowo częste, tzn. że wartości, jakie może ta zmienna przyjąć, występują w zbiorowości statystycznej taką samą ilość razy. W takiej sytuacji nieodparcie nasuwa się stwierdzenie o równomiernym rozkładzie jednostek zbiorowości statystycznej na warianty badanej zmiennej.

Z tego względu, iż wartość entropii zależy także od liczby wariantów, co na ogół oznacza, iż entropia jest wyższa, gdy liczba wariantów danej cechy jest większa, porównywanie rozkładu różnych cech poprzez porównywanie ich entropii poprawne jest jedynie wówczas, gdy liczby wariantów cech są jednakowe. W przypadku, gdy warunek ten nie jest spełniony, porównywanie rozkładu różnych cech należy oprzeć na stosunku entropii empirycznej do maksymalnej. Określony w poniższy sposób współczynnik entropii:

$$(4) \qquad WH = 1 - \frac{H}{H_{max}}$$

posiada pewne podobieństwo interpretacyjne do współczynnika koncentrcji k, jakkolwiek występują tutaj istotne różnice.

Jeżeli mianowicie $WH = 0$, podobnie gdy $k = 0$, rozkład jest równomierny z tym, że w ujęciu tradycyjnym oznacza to, iż wszystkie jednostki mają taką samą wartość badanej cechy, a w ujęciu entropijnym, że każdy wariant cechy występuje jednakową ilość razy. Jeżeli natomiast $WH = 1$, podobnie gdy $k = 1$, to znaczy, iż w zbiorowości statystycznej występuje całkowita koncentracja. Jednakże $k = 1$ oznacza, iż cała suma wartości danej zmiennej przypada na jedną jednostkę zbiorowości, a wszystkie pozostałe są jej całkowicie pozbawione.

---

[1] Zob. **G. Ja. Portnow, A. I. Ujomow:** *Isledowanije zawisimostiej mieżdu sistemnymi parametrami s pomoszczju EWM.* [W:] *Sistemnyje Isledowanija* 1971, s. 123 oraz **T. G. Uoterman:** *Teorja sistem i biologia, toczka zrenja biologa.* [W:] *Sistemnyje Isledowanija* 1970, s. 173.

Natomiast $WH = 1$ oznacza, iż wszystkie jednostki statystyczne wykazują się taką samą wartością badanej zmiennej. Pierwszy przypadek odnosi się do koncentracji sum wartości przy pewnych jednostkach, drugi zaś do koncentracji jednostek przy pewnych wartościach[2].

Dużą zaletą entropii jest możliwość jej wykorzystania w graficznej analizie rozkładu zmiennych. Temu celowi może służyć kwadrat o boku równym 1. Na boku reprezentującym oś X-ów oznaczamy kumulowane częstości występowania określonych wariantów danej cechy, zaś na boku reprezentującym oś Y, oznaczamy kumulowane cząstkowe współczynniki entropii. Cząstkowe współczynniki entropii wyznaczamy według wzoru:

$$(5) \qquad w_i h = \frac{|p_i \log_2 p_i|}{H_{max}}.$$

Połączenie współrzędnych punktów, kumulowanych częstości 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 ogólnego współczynnika entropii WH, charakteryzującego stopień koncentracji. Połączenie zaś punktu WH' z początkiem układu osi XY daje linię, którą nazywamy linią empirycznej symetrii rozkładu. Jeżeli linia entropii przebiega poniżej linii empirycznej symetrii rozkładu, to rozkład badanej zmiennej jest rozkładem asymetrycznym prawostronnym, czyli tzw. rozkładem o asymetrii dodatniej (wówczas wartość średnia jest większa od dominanty). Jeżeli natomiast linia entropii przebiega ponad linią empiryczną symetrii rozkładu, to rozkład badanej zmiennej jest rozkładem asymetrycznym lewostronnym, czyli tzw. rozkładem o asymetrii ujemnej (charakteryzującym się niższą wartością średniej od dominanty). Równomierny przebieg linii entropii po obu stronach empirycznej linii symetrii oznacza, iż rozkład badanej zmiennej jest symetryczny, a punkt przecięcia obu linii, jest punktem symetryczności rozkładu.

Jeżeli chodzi o pomiar stopnia asymetrii rozkładu za pomocą entropii, to miarę taką proponujemy oprzeć na różnicy sum cząstkowych współczynników entropii między pierwszymi i ostatnimi klasami szeregu statystycznego. Jeżeli ta różnica jest dodatnia, to rozkład danej zmiennej charakteryzuje się asymetrią prawostronną (co oznacza, iż dominująca liczba jednostek statystycznych cechuje się wartościami badanej zmiennej niższymi od przeciętnej), jeżeli zaś ujemna, to występuje asymetria lewostronna (czyli wartości zmiennej dominują-

[2] W kwestii rozróżnienia tych przypadków koncentracji zob. **B. Szulc:** *Statystyka dla ekonomistów*. Cz. 1. PWE, Warszawa 1972, s. 250.

cej liczby jednostek są wyższe od średniej). Dla szeregu statystycznego o nieparzystej liczbie klas stopień asymetrii ustalamy za pomocą wyrażenia:

$$(6) \qquad As = \sum_{i=1}^{k-1} w_i h - \sum_{i=k+1}^{n} w_i h,$$

w którym k oznacza numer klasy środkowej.
Natomiast dla szeregu o parzystej liczbie klas stopień asymetrii określa wzór:

$$(7) \qquad As = \sum_{i=1}^{k_1} w_i h - \sum_{i=k_2}^{n} w_i h,$$

w którym k_1 i k_2: pierwsza i druga klasa środkowa.

Wnioskowanie o rozkładach za pomocą współczynnika koncentracji k oraz za pomocą entropii zilustrujemy na przykładzie danych o liczbach gospodarstw domowych według liczby osób, pochodzących ze spisów ludności w latach 1960 i 1988.

Tabela 1. Współczynniki koncentracji k dla rozkładu gospodarstw domowych według liczby osób w gospodarstwie w latach: 1960 i 1988.

Liczba osób w gospodarstwie	1960		1988	
	% gospodarstw domowych	% ludności	% gospodarstw domowych	% ludności
1	16,10	4,66	18,28	5,90
2	18,52	10,72	22,33	14,40
3	18,98	16,48	20,28	19,62
4	19,79	22,90	21,99	28,37
5 i więcej	26,61	45,24	17,12	31,71
k	0,0581		0,0831	

Źródło: obliczenia własne na podstawie danych GUS.

W tabeli 1. podano procent gospodarstw domowych, a także procent ludności według liczby osób w gospodarstwie domowym oraz odpowiadające im współczynniki koncentracji k.

Tabela 2. zawiera obliczenia pomocnicze dla wyznaczenia empirycznej entropii gospodarstw domowych według liczby osób w latach: 1960 i 1988. Entropia maksymalna gospodarstw domowych według liczby osób w obu rozpatrywanych latach jest oczywiście jednakowa i wynosi: $H_{max} = log_2 5 = 2{,}3219$ wobec czego współczynnik entropii dla 1960 r.

Tabela 2. Obliczenia pomocnicze do określenia koncentracji i asymetrii – za pomocą entropii – gospodarstw domowych według liczby osób w 1960 i 1980 r.

Liczba osób w gospodarstwie domowym	1960					1988				
	p_i	$\lvert p_i log_2 p_i \rvert$	$w_i h$	$cum\ p_i$	$cum\ w_i h$	p_i	$\lvert p_i log_2 p_i \rvert$	$w_i h$	$cum\ p_i$	$cum\ w_i h$
1	0,1610	0,4242	0,1827	0,1610	0,1827	0,1828	0,4482	0,1930	0,1828	0,1930
2	0,1852	0,4506	0,1940	0,3462	0,3767	0,2233	0,4830	0,2080	0,4061	0,4010
3	0,1898	0,4550	0,1960	0,5360	0,5727	0,2028	0,4668	0,2011	0,6089	0,6021
4	0,1997	0,4625	0,1992	0,7339	0,7719	0,2199	0,4805	0,2069	0,8288	0,8090
5 i więcej	0,2661	0,5082	0,2189	1,0000	0,9908	0,1712	0,4359	0,1877	1,0000	0,9967
H		2,3005					2,3144			

Źródło: Obliczenia własne na podstawie danych GUS.

Tabela 3. Obliczenia pomocnicze do określenia koncentracji i asymetrii – za pomocą entropii – gospodarstw domowych według liczby czynnych zawodowo w 1960 i 1988 r.

Liczba osób w gospodarstwie domowym	1960					1988				
	p_i	$\lvert p_i log_2 p_i \rvert$	$w_i h$	$cum\ p_i$	$cum\ w_i h$	p_i	$\lvert p_i log_2 p_i \rvert$	$w_i h$	$cum\ p_i$	$cum\ w_i h$
0	0,0718	0,2728	0,1175	0,0718	0,1175	0,1704	0,4350	0,1874	0,1704	0,1874
1	0,4155	0,5265	0,2267	0,4873	0,3442	0,3227	0,5266	0,2268	0,4931	0,4142
2	0,3648	0,5307	0,2286	0,8521	0,5728	0,3782	0,5305	0,2285	0,8713	0,6427
3 i więcej	0,1479	0,4078	0,1756	1,0000	0,7484	0,1287	0,3807	0,1640	1,0000	0,8067
H		1,7378					1,8728			

Źródło: Obliczenia własne na podstawie danych GUS.

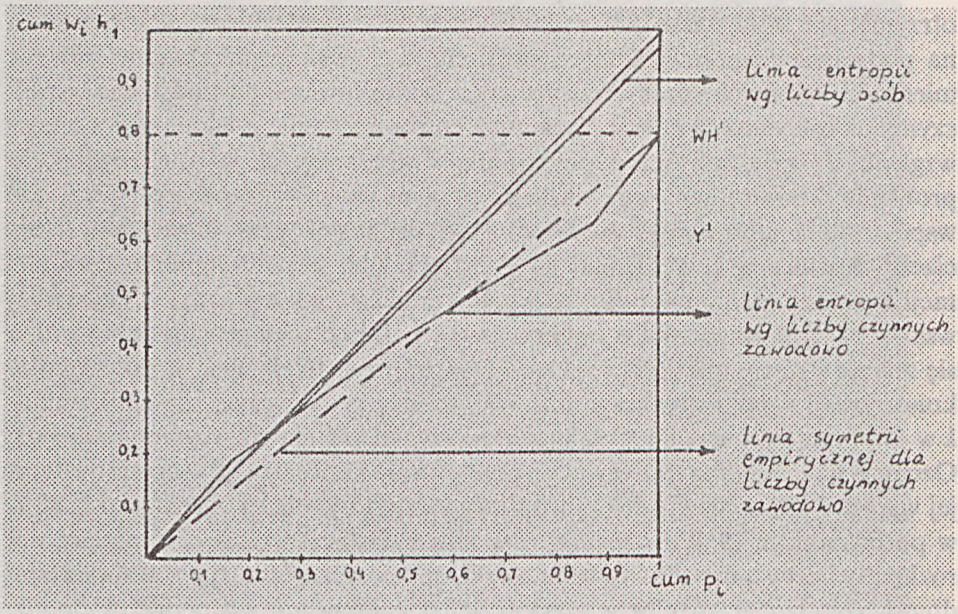
otrzymano: 0,0092, a dla 1988 r.: 0,0032. Pierwszym wnioskiem, jaki się nasuwa na podstawie przeprowadzonych obliczeń, jest stwierdzenie znacznej równomierności badanych rozkładów. Jednakże wnioskowanie na podstawie współczynnika koncentracji *k* i współczannika entropijnego nie jest identyczne z tego względu, iż pierwszy z tych współczynników odnosi się, jak już stwierdzono, do problemu koncentracji sum wartości przy pewnych jednostkach, drugi zaś współczynnik (entropijny) do koncentracji jednostek przy pewnych wartościach[3]. Różnice w stwierdzonej na podstawie obu współczynników koncentracji znacznej równomierności rozkładów można , jak sądzimy, najwyraźniej wskazać w tym przypadku w ujęciu czasowym. Jakkolwiek generalnie stwierdza się znaczną równomierność rozkładów zarówno na podstawie współczynnika koncentracji *k*, jak i współczynnika *WH*, to można zauważyć, że współczynnik *k* w 1960 r. w porównaniu z 1988 r. był niższy, odwrotnie niż współczynnik *WH*. Oznacza to, iż, według tradycyjnego pojmowania koncentracji, w czasie nastąpił jej wzrost. Natomiast spadek entropijnego współczynnika koncentracji 1988 r. w porównaniu z 1960 r. oznacza, iż w czasie nastąpił wzrost równomierności rozkładu liczby gospodarstw domowych na poszczególne typy gospodarstw domowych, tzn. że różnice w częstościach występowania gospodarstw domowych różnej wielkości w 1988 r. były mniejsze niż w 1960 r. (zauważmy, iż suma bezwzględnych różnic w odsetkach gospodarstw domowych między poszczególnymi typami gospodarstw domowych w 1960 r. wynosiła 44,58, a w 1988 r. 28,26). Rozkład liczby gospodarstw domowych na poszczególne typy gospodarstw domowych (czyli na warianty liczby osób w gospodarstwie domowym) był więc bardziej równomierny w 1988 r.

Współczynniki entropii, jakie otrzymano dla gospodarstw domowych według liczby czynnych zawodowo dla rozpatrywanych lat, są wyraźnie wyższe niż współczynniki entropii gospodarstw domowych według liczby osób, co świadczy o tym, iż koncentrcja gospodarstw domowych ze względu na liczbę czynnych zawodowo jest większa niż z punktu widzenia liczby osób w gospodarstwie domowym. Zatem rozkład gospodarstw domowych według liczby osób jest bardziej równomierny niż rozkład gospodarstw domowych według liczby czynnych zawodowo (zob. rys. 1 i 2.).

Entropię, jak powiedziano, można również wykorzystać w analizie symetrii szeregu statystycznego, przy czym ze względu na uporządkowanie szeregu statystycznego od najniższej do najwyższej wartości cechy wnioskowanie na podstawie proponowanego entropijnego wskaźnika asymetrii jest takie samo jak na podstawie tradycyjnych wskaźników asymetrii. Należy jednakże zauważyć, iż entropijny wskaźnik asymetrii opierający się wyłącznie na częstościach

---

[3] W tym też upatrujemy przyczyny stwierdzanej rozbieżności wyników uzyskanych na podstawie entropijnego i innych współczynników koncentracji. Zob. *Badania przestrzenne rynku i konsumpcji*. PWE, Warszawa 1992, s. 86 oraz *Ekonometria przestrzenna*. PWE, Warszawa 1991, s.53.



Rys. 1. Entropia gospodarstw domowych w 1988 r. według liczby osób w gospodarstwie domowym oraz według liczby czynnych zawodowo.

występowania cech może być stosowany również w tych przypadkach, w których tzw. tradycyjne parametry opisowe zbiorowości, na których zbudowane są tradycyjne wskaźniki asymetrii, nie mogą zostać wykorzystane ze względu , na przykład, na otwarte przedziały klasowe.

Zastosowanie wzoru (6) w przypadku badania symetrii rozkładu gospodarstw domowych według liczby osób daje następujący wskaźnik asymetrii:

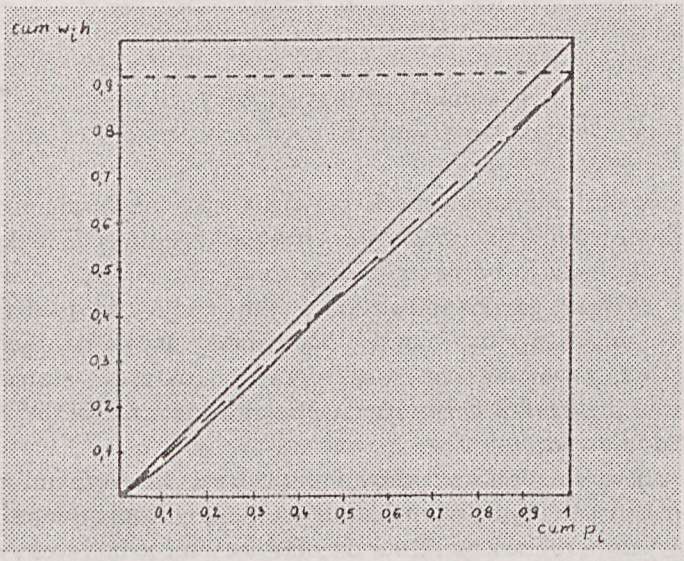
dla 1960 r.:

$$As = \sum_{i=1}^{2} w_i h - \sum_{i=4}^{5} w_i h = 0{,}3767 - 0{,}4181 = -0{,}0414$$

dla 1988 r.:

$$As = \sum_{i=1}^{2} w_i h - \sum_{i=4}^{5} w_i h = 0{,}4010 - 0{,}3946 = 0{,}0064$$

Zauważmy, iż wyznaczenie asymetrii na podstawie różnicy między średnią arytmetyczną a dominantą w rozpatrywanych przez nas przypadkach jest niemożliwe ze względu na otwarte szeregi statystyczne. Wyznaczenie wskaźnika asymetrii dla rozkładu gospodarstw domowych z punktu widzenia liczby osób



Rys. 2. Entropia gospodarstw domowych w 1988 r. według łącznie liczby osób i liczby czynnych zawodowo w gospodarstwie domowym.

w 1960 r. wyłącznie na podstawie parametrów pozycyjnych jest także niemożliwe, ponieważ niemożliwe jest wyznaczenie dominanty, która przypada na ostatni, otwarty przedział klasowy. W tym przypadku jedyną miarą asymetrii jest wskaźnik entropijny. Natomiast współczynnik asymetrii wyznaczony według wzoru: $AS=\frac{Me-D}{Q}$ wynosi dla 1988 r. 0,68. Podobnie jak entropijny wskaźnik asymetrii jest on dla tego przypadku dodatni, ale wskazuje na stosunkowo duży stopień asymetrii (współczynnik: $AS=\frac{Me-D}{Q}$ jest unormowany w przedziale 0,1) w rozkładzie gospodarstw domowych według liczby osób w 1988 r., co w konfrontacji z danymi wyjściowymi (zob. tabela 2.) nie wydaje się w pełni przekonywające. Sądzimy, iż lepszą miarą asymetrii jest wskaźnik entropijny.

Natomiast zastosowanie wzoru (7) w przypadku badania symetrii rozkładu gospodarstw domowych ze względu na liczbę czynnych zawodowo w gospodarstwie domowym daje:

dla 1960 r.:

$$As=\sum_{i=0}^{1} w_i h-\sum_{i=2}^{3} w_i h=0{,}3442-0{,}4042=-0{,}0600$$

dla 1988 r.:

$$As = \sum_{i=0}^{1} w_i h - \sum_{i=2}^{3} w_i h = 0{,}4142 - 0{,}3925 = 0{,}0217$$

Jak widać, w obu rozpatrywanych przypadkach i okresach, rozkłady gospodarstw domowych były prawie symetryczne. Niemniej jednak, w 1988 r. zarówno jeżeli chodzi o liczbę osób w gospodarstwie , jak i liczbę czynnych zawodowo, rozkład gospodarstw domowych cechowała (słaba) asymetria dodatnia, tj. prawostronna, zaś w 1960 (słaba) asymetria ujemna, tj. lewostronna, przy czym dla obu rozpatrywanych okresów, stopień asymetrii był większy w przypadku liczby czynnych zawodowo niż liczby osób w gospodarstwie domowym.

Cenną zaletą entropii, w porównaniu z tzw. klasycznymi narzędziami badania koncentracji i symetrii rozkładu, jest to, że entropia pozwala na analizę tych zjawisk jednocześnie z punktu widzenia więcej niż jednej cechy dzięki temu, iż entropia jako miara opiera się wyłącznie na częstościach występowania zmiennej (bez uwzględniania wartości tej zmiennej). Przykładem w tym względzie może być wykorzystanie entropii do analizy koncentracji i symetrii rozkładu gospodarstw domowych jednocześnie ze względu na liczbę osób w gospodarstwie domowym i liczbę czynnych zawodowo. Tabela 4. zawiera dane o częstościach występowania gospodarstw domowych według liczby osób i liczby czynnych zawodowo oraz obliczenia pomocnicze dla określenia koncentracji i symetrii rozkładu gospodarstw domowych tych cech w latach: 1960 i 1988. Łatwo sprawdzić, iż przy równej dla obu rozpatrywanych lat entropii maksymalnej wynoszącej 4,0875 ogólny współczynnik entropii w 1960 r. wynosił 0,0856, a w 1988 r. 0,0783, zaś wskaźnik asymetrii dla rozkładu gospodarstw domowych według liczby czynnych zawodowo względem liczby osób wynosił w 1960 r. 0,0404, a w 1988 r. 0,1497.

Wypada zauważyć, że jakkolwiek różnice w poziomach entropii gospodarstw domowych między 1960 a 1988 r. nie są znaczące, to jednak w 1988 r. w porównaniu z 1960 r. nastąpił (nieduży) wzrost zróżnicowania gospodarstw domowych z punktu widzenia obu rozpatrywanych cech, tj. liczby osób i liczby czynnych zawodowo w gospodarstwie domowym.

Należy podkreślić, iż w przypadku zmiany kolejności kryteriów klasyfikacyjnych zbiorowości statystycznej wartość współczynnika entropii, jako miary koncentracji, nie ulega zmianie, zmienia się natomiast wskaźnik asymetrii. Jest on bowiem miarą porządkową, gdyż opiera się na kumulowanych cząstkowych współczynnikach entropii. W przypadku badania symetrii rozkładu ze względu na więcej niż jedną cechę, ważne jest określenie kolejności cech, tj. jakby ustalenie stosunku nadrzędności między cechami. W naszym przypadku możemy rozpatrywać symetrię rozkładu gospodarstw domowych według liczby

Tabela 4. Obliczenia pomocnicze do określenia koncentracji i asymetrii – za pomocą entropii – gospodarstw domowych, z uwzględnieniem liczby osób i liczby czynnych zawodowo w gospodarstwie.

Liczba osób/ liczba czynnych	1960			1988								
	p_i	$\lvert p_i log_2 p_i \rvert$	$w_i h$	p_i	$\lvert p_i log_2 p_i \rvert$	$w_i h$	$cum\ p_i$	$cum\ w_i h$	liczba czynnych/ /liczba osób	p_i	$\lvert p_i log_2 p_i \rvert$	$w_i h$
1/0	0,0412	0,1898	0,0464	0,0898	0,3123	0,0764	0,0898	0,0764	0|1	0,0898	0,3123	0,0764
1/1	0,1199	0,3669	0,0898	0,0930	0,3187	0,0780	0,1828	0,1544	0/2	0,0658	0,2583	0,0632
2/0	0,0213	0,1183	0,0289	0,0658	0,2583	0,0632	0,2486	0,2176	0/3	0,0102	0,0675	0,0165
2/1	0,0805	0,2926	0,0716	0,0815	0,2948	0,0721	0,3301	0,2897	0/4	0,0030	0,0251	0,0062
2/2	0,0835	0,2991	0,0732	0,0760	0,2826	0,0691	0,4061	0,3588	0/5 i w.	0,0015	0,0141	0,0034
3/0	0,0050	0,0382	0,0094	0,0102	0,0675	0,0165	0,4163	0,3753	1/1	0,0930	0,3187	0,0780
3/1	0,0704	0,2695	0,0659	0,0628	0,2508	0,0614	0,4791	0,4367	1/2	0,0815	0,2948	0,0721
3/2	0,0848	0,3019	0,0739	0,1007	0,3335	0,0816	0,5798	0,5183	1/3	0,0628	0,2508	0,0614
3/3	0,0295	0,1500	0,0367	0,0291	0,1485	0,0363	0,6089	0,5546	1/4	0,0533	0,2254	0,0552
4/0	0,0024	0,0209	0,0051	0,0030	0,0251	0,0062	0,6119	0,5607	1/5 i w.	0,0322	0,1596	0,0391
4/1	0,0732	0,2761	0,0676	0,0533	0,2254	0,0552	0,6652	0,6159	2/2	0,0760	0,2826	0,0691
4/2	0,0861	0,3046	0,0745	0,1277	0,3792	0,0928	0,7929	0,7086	2/3	0,1007	0,3335	0,0816
4/3 i w.	0,0361	0,1730	0,0423	0,0359	0,1723	0,0422	0,8288	0,7508	2/4	0,1277	0,3792	0,0928
5 i w./0	0,0019	0,0172	0,0042	0,0015	0,0141	0,0034	0,8303	0,7542	2/5 i w.	0,0738	0,2775	0,0679
5/1	0,0716	0,2724	0,0666	0,0322	0,1596	0,0391	0,8625	0,7933	3/3	0,0291	0,1485	0,0363
5/2	0,1104	0,3510	0,0859	0,0738	0,2775	0,0679	0,9393	0,8612	3/4	0,0359	0,1723	0,0422
5/3 i w.	0,0822	0,2963	0,0725	0,0637	0,2531	0,0619	1,0000	0,9231	3/5 i w.	0,0637	0,2531	0,0619
H		3,7376			3,7675						3,7675	

Źródło: Obliczenia własne na podstawie danych GUS.

czynnych zawodowo w gospodarstwach domowych w określonej liczbie osób, czyli symetrię rozkładu gospodarstw domowych według liczby czynnych zawodowo względem liczby osób w gospodarstwie domowym, bądź odwrotnie: symetrię rozkładu gospodarstw domowych według liczby osób względem liczby czynnych zawodowo w gospodarstwie domowym – zob. tabela 4.

Najważniejszą zaletą entropii jako miary koncentracji i asymetrii jest to, iż pozwala ona na analizę i porównywanie tych zjawisk nie tylko w przypadku cech ilościowych, ale także jakościowych. Wynika to również z tego faktu, że jak już stwierdzono, entropia opiera się wyłącznie na częstościach występowania w zbiorowości statystycznej wariantów danej zmiennej. Nieważne jest zatem to, czy zmienna ta ma charakter cechy ilościowej czy jakościowej.

Jako przykład materiału empirycznego służącego do ilustracji zastosowania entropii jako miary koncentracji i asymetrii rozkładu w przypadku cech jakościowych mogą posłużyć dane o liczbie gospodarstw domowych według poziomu wykształcenia małżonków, pochodzące ze spisu ludności w 1988 r. – tabela 5.

Tabela 5. Małżeńskie gospodarstwa domowe według poziomu wykształcenia małżonków, w woj. poznańskim w 1988 r.

Lp.	Poziom wykształcenia	Liczba gospodarstw	P_i	$\lvert p_i \log_2 p_i \rvert$	$w_i h$
1	oboje z wyższym	18 740	0,0624	0,2497	0,0788
2	jedno z wyższym	26 809	0,0892	0,3110	0,0981
3	oboje ze średnim	42 492	0,1414	0,3991	0,1259
4	jedno ze średnim	59 115	0,1968	0,4615	0,1456
5	oboje z zasad. zawod.	44 453	0,1480	0,4079	0,1287
6	jedno z zasad. zawod.	48 542	0,1616	0,4249	0,1340
7	oboje z podstawowym	56 993	0,1897	0,4549	0,1435
8	jedno z podstawowym	2 042	0,0068	0,0490	0,0155
9	oboje bez podstawowego	1 242	0,0041	0,0325	0,0103

Źródło: dane GUS oraz obliczenia własne.

Na podstawie zawartych w tej tabeli obliczeń pomocniczych ustalamy następującą wartość współczynnika entropii:

$$WH = 1 - \frac{H}{H_{max}} = 1 - \frac{2,7905}{3,1699} = 0,1197,$$

która wskazuje na dość znaczną równomierność rozkładu gospodarstw domowych na wyodrębnione warianty poziomu wykształcenia małżonków, a więc na dość duże zróżnicowanie poziomu wykształcenia małżonków w gospodarstwach domowych.

Współczynniki asymetrii rozkładu:

$$As = \sum_{i=1}^{4} w_i h - \sum_{i=6}^{9} w_i h = 0{,}4484 - 0{,}3033 = 0{,}1451$$

świadczy o tym, iż w rozpatrywanej zbiorowości gospodarstw domowych przeważają gospodarstwa o poziomie wykształcenia małżonków nieco niższym od przeciętnego.

W zakończeniu pragniemy jeszcze zwrócić uwagę na to, że entropia, jako miara koncentracji i asymetrii rozkładu, posiada jeszcze tę zaletę, iż może być wykorzystana do analizy i porównań tych zjawisk, w różnych kombinacjach cech, tj. ilościowych, jakościowach i ilościowo-jakościowych, należy jednak pamiętać o różnicach występujących w tradycyjnym i entropijnym pojmowaniu koncentracji.

Iwona Roeske-Słomka

DISTRIBUTION OF FEATURES VS DISTRIBUTION OF FREQUENCY ON THE EXAMPLE OF HOUSEHOLDS

Summary

As a measure of concentration the author assumes entropy coefficient: $WH = 1 - H/H_{max}$, *in which H denotes real entropy* ($H = -\Sigma p_i \log_2 p_i$, *where* p_i *– the frequency of occurrence of the i-variant of a given feature) and* H_{max} *– maximum entropy* ($H_{max} = \log_2 n$, *where n is a number of variants of the a given feature). Entropic approach to concentration differs from the traditional approach.* $WH = 1$ *means a complete concentration where all the statistical units have the same value of the studied variable.* $WH = 0$ *means that the distribution is absolutely even which is expressed in the fact that each variant of the feature appears the same number of times in the statistical population. As a measure of asymmetry of distribution the author assumes the coefficient – for statistical series of the uneven number of classes:*

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

and for the series of even number of classes

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

where $w_i h$ – a partial coefficient of entropy for the i-variant of the studied feature, i.e. $w_i h = | p_i \log_2 p_i | / H_{max}$, while k, k_1, k_2 – number of median classes. The article proves that entropy as a measure of concentration and asymmetry has a number of advantages because it facilitates not only the analysis of these phenomena for the quantitative and qualitative features but also a simultaneous analysis from the viewpoint of more than one feature.