

WSPÓŁCZYNNIK DWUMODALNOŚCI BC I JEGO ZASTOSOWANIE W ANALIZACH ROZKŁADÓW ZMIENNYCH LOSOWYCH

Aleksandra Baszczyńska, Dorota Pekasiewicz
Katedra Metod Statystycznych Uniwersytet Łódzki
e-mail: albasz@uni.lodz.pl, pekasiewicz@uni.lodz.pl

Streszczenie: W analizach zjawisk ekonomicznych, przyrodniczych i psychologicznych istotne znaczenie ma badanie modalności rozkładu zmiennej losowej utożsamianej z analizowaną cechą statystyczną. Jedną z metod polega na wyznaczeniu na podstawie próby losowej współczynnika dwumodalności BC . Jego wartość pozwala podjąć decyzję o jednomodalności lub dwumodalności rozkładu analizowanej zmiennej. W pracy zaprezentowane zostały wyniki badań symulacyjnych dotyczących własności współczynnika BC . Ponadto zastosowano analizowany współczynnik do badania modalności rozkładu cen surowców na giełdach towarowych.

Słowa kluczowe: modalność, współczynnik dwumodalności BC , mieszaniny rozkładów

WSTĘP

Dwumodalność jest przedmiotem analiz w obszarze zagadnień nie tylko ekonomicznych. Badanie dwumodalności ma zastosowanie w analizach zagadnień przyrodniczych, lingwistycznych oraz dotyczących psychologii i socjologii [Knapp 2007].

Jedną z pierwszych prac, w której rozważane były procedury określające czy badany rozkład może być traktowany jako dwa rozkłady normalne jest praca Karola Pearsona z 1894 roku [Pearson 1894]. Jeżeli badany rozkład charakteryzował się dwoma „szczytami” oddalonymi od siebie to taki rozkład określany był jako dwumodalny. W badaniach dwumodalności wykorzystywane były, m.in., następujące miary: czwarty moment centralny, kwadrat momentu trzeciego centralnego (metody Pearsona, Darlinga), kurtoza (metoda Chissoma).

Rozwój metod wnioskowania statystycznego sprawił, że skonstruowano testy weryfikujące hipotezę o jednomodalności wobec hipotezy o wielomodalności, w tym test Dip Hartigana [Hardigan, Hartigan 1985] oraz test Silvermana [Silverman 1996]. Zastosowanie wymienionych testów statystycznych jest możliwe w praktyce dzięki istniejącemu oprogramowaniu (np. Matlab, R). Wyniki badania własności testu Dip i jego wykorzystania w analizach modalności rozkładu populacji przedstawione są w pracach Baszczyńska, Pekasiewicz 2014a, 2014b.

Alternatywną metodą określania modalności rozkładu jest wyznaczanie współczynnika dwumodalności BC . Współczynnik ten skonstruowany jest w oparciu o miarę asymetrii i kurtozę. W pracy przedstawiono wyniki analiz jego własności dla grup rozkładów jedno- i dwumodalnych. Badania przeprowadzone zostały metodą Monte Carlo i miały na celu oszacowanie odsetka błędnych decyzji o dwumodalności rozkładu mimo, że rozkład był jednomodalny i odwrotnie. Analizowano również charakterystyki liczbowe współczynnika dwumodalności BC i badano zgodność jego rozkładu z rozkładem normalnym.

WSPÓŁCZYNNIK DWUMODALNOŚCI BC

Współczynnik dwumodalności BC określony jest następującym wzorem:

$$BC = \frac{m_3^2 + 1}{m_4 + 3 \cdot \frac{(n-1)^2}{(n-2)(n-3)}}, \quad (1)$$

gdzie: m_3 jest współczynnikiem asymetrii opartym na trzecim momencie centralnym, natomiast m_4 jest kurtozą rozkładu.

Decyzję o dwumodalności rozkładu podejmuje się porównując obliczony współczynnik BC z wartością krytyczną: $BC_{kryt.} = \frac{5}{9} \approx 0,555$. Jeśli $BC > BC_{kryt.}$, to przyjmuje się, że rozkład analizowanej zmiennej losowej jest dwumodalny.

ANALIZA MONTE CARLO WŁASNOŚCI WSPÓŁCZYNNIKA BC

W celu zbadania własności współczynnika dwumodalności rozważano następujące grupy rozkładów:

- rozkłady normalne $N(\mu; \sigma)$,
- rozkłady gamma $G(k; \theta)$,
- mieszaniny rozkładów o funkcji gęstości:

$$f(x) = wf_1(x) + (1-w)f_2(x), \quad (2)$$

gdzie $w \in (0, 1)$ oraz f_1, f_2 to funkcje gęstości rozkładu normalnego lub rozkładu gamma i normalnego.

W przypadku populacji o rozkładzie normalnym $N(\mu; \sigma)$ zastosowanie współczynnika dwumodalności BC prowadziło do podjęcia decyzji o jednomodalności rozkładu populacji nawet dla prób małych, np. dla prób 20-elementowych tylko 1% decyzji była błędnych.

Drugi z rozważanych rozkładów, rozkład gamma $G(k; \theta)$ o funkcji gęstości:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ \frac{\theta^{-k} x^{k-1} \exp\left(-\frac{x}{\theta}\right)}{\Gamma(k)} & \text{dla } x > 0 \end{cases}, \quad (3)$$

jest rozkładem jednomodalnym, ale odpowiednio dobierając parametr k można uzyskać różną asymetrię i kurtozę (współczynnik asymetrii $\gamma = \frac{2}{\sqrt{k}}$, kurtoza

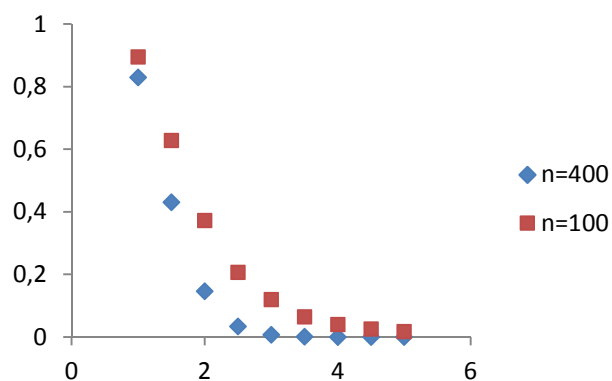
$\kappa = \frac{6}{k}$). W przypadku tego rozkładu badano zależność liczby błędnych decyzji od wielkości parametru k . Dla wygenerowanych populacji weryfikację hipotez powtarzano 20000 razy. W Tabeli 1 podano odsetek błędnych decyzji związanych z odrzuceniem hipotezy o jednomodalności populacji o rozkładzie gamma w oparciu o próbę 100-elementową i 400-elementową, a na Rysunku 1 graficznie zaprezentowane są odsetki niepoprawnych decyzji dla różnych wielkości parametru k . Na podstawie otrzymanych rezultatów można sformułować wniosek, iż przy dużej asymetrii rozkładu liczebności prób muszą być znacznie większe niż przy małej asymetrii i kurtozie, aby prawdopodobieństwo błędu związanego z przyjęciem hipotezy o dwumodalności było na poziomie 0,05.

Tabela 1. Odsetek błędnych decyzji związanych z odrzuceniem hipotezy o jednomodalności populacji o rozkładzie $G(k; 2)$

n	k								
	1	1,5	2	2,5	3	3,5	4	4,5	5
100	0,895	0,628	0,372	0,207	0,119	0,065	0,040	0,025	0,017
400	0,829	0,430	0,146	0,034	0,007	0,001	0,000	0,000	0,000

Źródło: obliczenia własne

W Tabeli 2 przedstawione są podstawowe charakterystyki liczbowe współczynnika dwumodalności. Choć współczynnik asymetrii jest bliski 0, a kurtoza 3, zastosowanie testu Shapiro-Wilka prowadziło do podjęcia decyzji o odrzuceniu hipotezy zerowej o normalności rozkładu BC . W Tabeli 3 podane są wartości statystyk testu Shapiro-Wilka i p -value dla współczynnika BC wyznaczonego w oparciu o 400 obserwacji.

Rysunek 1. Graficzna prezentacja zależności odsetka błędnych decyzji od parametru k 

Źródło: opracowanie własne

Tabela 2. Oszacowane wartości charakterystyk rozkładu współczynnika BC dla wybranych rozkładów jednomodalnych

f	n	Charakterystyki liczbowe rozkładu BC					
		min.	max.	wartość średnia	odchylenie standardowe	współczynnik asymetrii	kurtoza
$N(4; 2)$	100	0,1397	0,5401	0,3526	0,0443	-0,0425	3,0178
	400	0,2338	0,4243	0,3395	0,0256	-0,1354	3,0591
$N(2; 2)$	100	0,1699	0,5351	0,3538	0,0440	-0,0373	2,9823
	400	0,2227	0,4417	0,3374	0,0260	-0,1301	3,0693
$G(4; 2)$	100	0,2658	0,6699	0,4703	0,0484	-0,0317	2,9879
	400	0,3189	0,5653	0,4574	0,0321	-0,4018	3,4460
$G(2; 2)$	100	0,3455	0,7547	0,5400	0,0494	-0,0595	2,9438
	400	0,3680	0,6354	0,5230	0,0347	-0,4548	3,3714

Źródło: obliczenia własne

Tabela 3. Test Shapiro-Wilka – normalność rozkładu BC

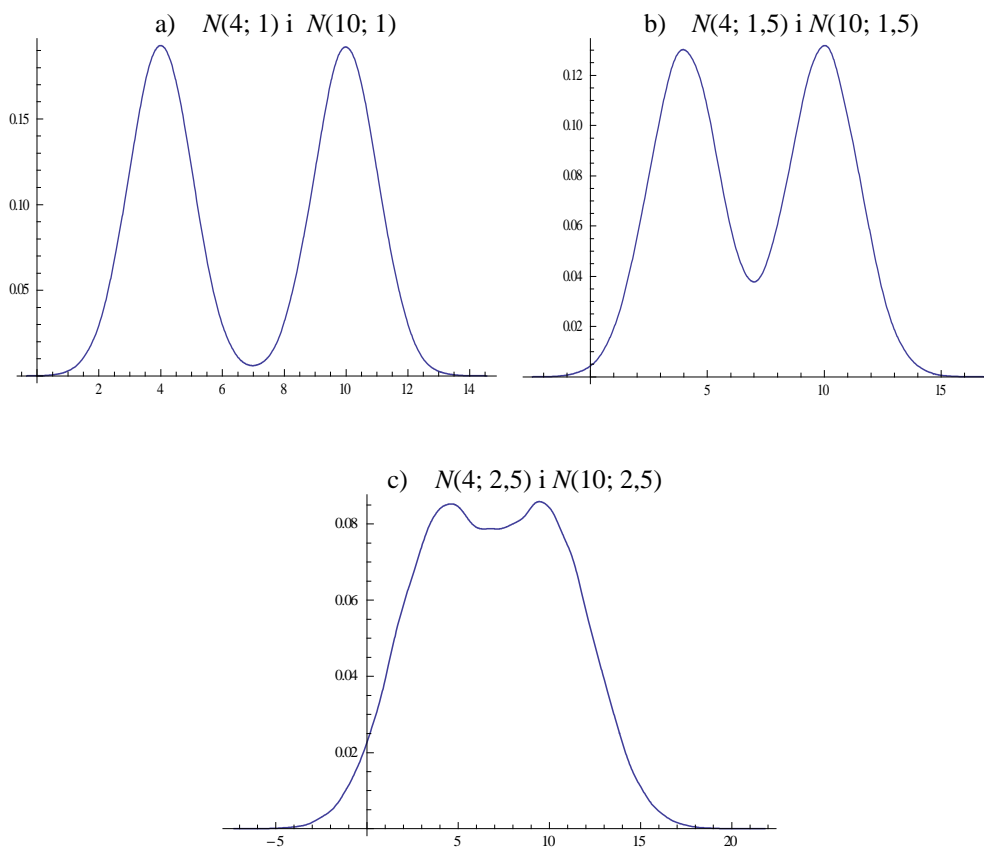
Rozkład	Statystyka testu	p -value
$N(4; 2)$	0,9977	0,1904
$N(2; 2)$	0,9957	0,0069
$G(4; 2)$	0,9832	0,0000
$G(2; 2)$	0,9853	0,0000

Źródło: obliczenia własne

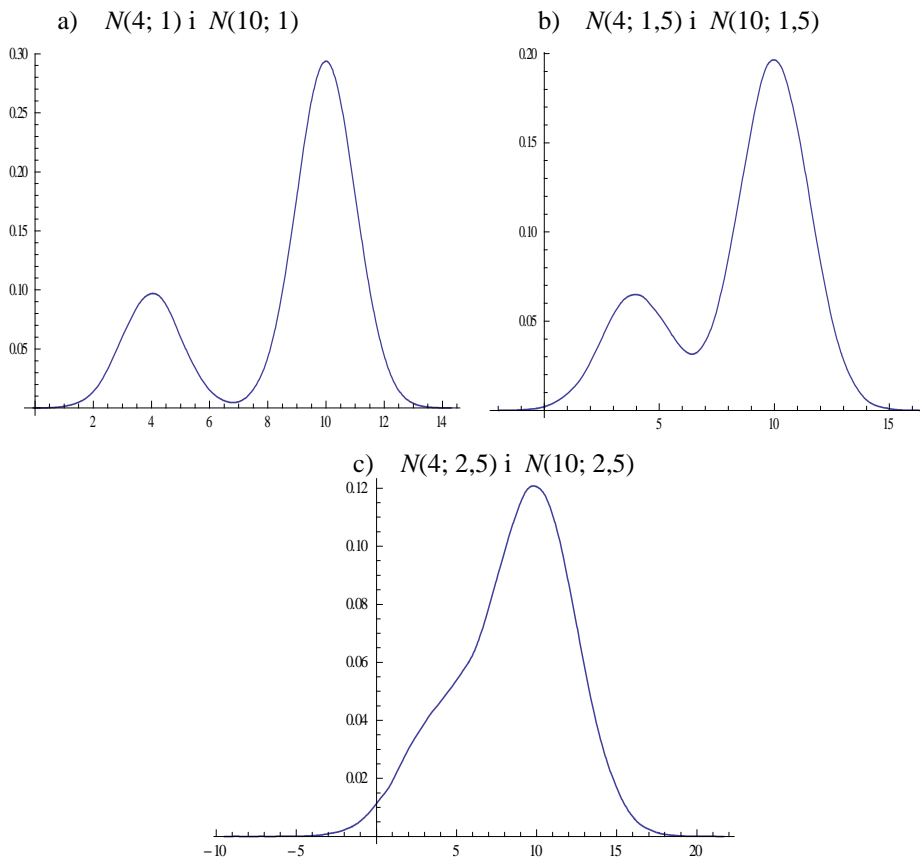
Dla mieszanin rozkładów przeprowadzono analogiczne rozważania. Odpowiednio ustalając parametry i udziały rozkładów wchodzących w skład mieszanin otrzymuje się rozkłady dwumodalne lub jednomodalne. Funkcje gęstości dla mieszanin dwóch wybranych rozkładów normalnych, dla $w=0,5$ oraz $w=0,25$, przedstawione są na Rysunkach 2 i 3.

Odsetki decyzji o dwumodalności populacji obliczone w wyniku powtórzenia 20000 razy procedury wyznaczania BC przedstawia Tabela 4. Uzyskane rezultaty są zgodne z rozkładami populacji, z których pochodzą próby (por. Rys. 2 i 3). Większe wartości odchyłeń standardowych rozkładów wchodzących w skład mieszanin sprawiają, że rozkłady populacji są praktycznie rozkładami jednomodalnymi. W przypadku małych wartości σ rozkłady populacji są dwumodalne i współczynnik BC różnicuje te rozkłady.

Rysunek 2. Funkcje gęstości wybranych populacji będących mieszaninami dwóch rozkładów dla $w=0,5$



Źródło: opracowanie własne

Rysunek 3. Funkcje gęstości wybranych populacji będących mieszaninami dwóch rozkładów dla $w=0,25$ 

Źródło: opracowanie własne

Tabela 4. Odsetek decyzji o dwumodalności rozkładów populacji będących mieszaniną rozkładów $N(4; \sigma)$ i $N(10; \sigma)$

w	n	σ						
		0,5	1,0	1,25	1,5	1,75	2,0	2,5
0,5	100	1,0000	1,0000	0,9771	0,5993	0,1702	0,0241	0,0008
	400	1,0000	1,0000	1,0000	0,8409	0,0917	0,0006	0,0000
0,25	100	1,0000	1,0000	0,9841	0,7316	0,2771	0,0617	0,0029
	400	1,0000	1,0000	1,0000	0,9330	0,1940	0,0033	0,0000

Źródło: obliczenia własne

Dla populacji będących mieszaninami rozkładów, podobnie jak w przypadku populacji o rozkładach normalnym i gamma, szacowano charakterystyki liczbowe rozkładu współczynnika BC i sprawdzano, czy jego rozkład można uznać za

normalny stosując test Shapiro-Wilka. Uzyskane rezultaty dla wybranych mieszanin rozkładów zawarte są w Tabelach 5 i 6, przy czym wyniki testu Shapiro-Wilka dotyczą badania normalności współczynnika BC wyznaczanego w oparciu o 400 obserwacji.

Tabela 5. Oszacowane wartości charakterystyk rozkładu współczynnika BC dla wybranych mieszanin rozkładów

$f_1,$ f_2	w	n	Charakterystyki liczbowe rozkładu BC					
			min.	max.	wartość średnia	odchylenie standardowe	współczynnik asymetrii	kurtoza
$N(4;1), (10;1)$	0,5	100	0,5615	0,8088	0,6868	0,0269	-0,1768	3,1103
		400	0,6573	0,7699	0,7151	0,0149	-0,1141	2,9921
	0,25	100	0,5739	0,8446	0,7281	0,0318	-0,2282	3,0754
		400	0,6782	0,8111	0,7487	0,0161	-0,1234	3,0414
$N(4;1,5), N(10;1,5)$	0,5	100	0,4058	0,6878	0,5632	0,0351	-0,1256	3,0234
		400	0,4927	0,6494	0,5784	0,0193	0,0870	3,0251
	0,25	100	0,3617	0,7358	0,5786	0,0465	-0,2372	3,1298
		400	0,4798	0,6875	0,5942	0,0243	-0,1534	3,1036
$N(4;2,5), N(10;2,5)$	0,5	100	0,2680	0,5830	0,4352	0,0402	-0,0881	3,0211
		400	0,3415	0,5163	0,4370	0,0222	-0,0656	3,0183
	0,25	100	0,2320	0,6143	0,4237	0,0490	-0,0222	2,9927
		400	0,3046	0,5234	0,4228	0,0272	-0,0334	3,0467

Źródło: obliczenia własne

Tabela 6. Test Shapiro-Wilka – normalność rozkładu BC

Rozkłady	w	Statystyka testu	p -value
$N(4;1)$ i $N(10;1)$	0,5	0,9958	0,0080
	0,25	0,9969	0,0474
$N(4;1,5)$ i $N(10;1,5)$	0,5	0,9965	0,0266
	0,25	0,9975	0,1391
$N(4;2,5)$ i $N(10;2,5)$	0,5	0,9963	0,0181
	0,25	0,9977	0,1818

Źródło: obliczenia własne

ZASTOSOWANIE WSPÓŁCZYNNIKA BC W BADANIACH MODALNOŚCI ROZKŁADÓW CEN SUROWCÓW Z GIEŁD TOWAROWYCH

Współczynnik dwumodalności BC zastosowano do badania modalności rozkładów cen pszenicy konsumpcyjnej w Polsce i cen złota na giełdzie nowojorskiej. Korzystając z danych zawartych na stronach internetowych: http://agrolok.pl/matif/wykres_okno.php?id_menu=291 oraz <http://gieluda.onet.pl/zloto,41920,430,1,8,1,historyczne,profile-surowce-archiwum-s?dateRadio=radioDate3&fromDate=2012-06-01&toDate=2014-05-31> badano modalność rozkładów cen rozważanych artykułów w oparciu o dzienne ceny w następujących okresach:

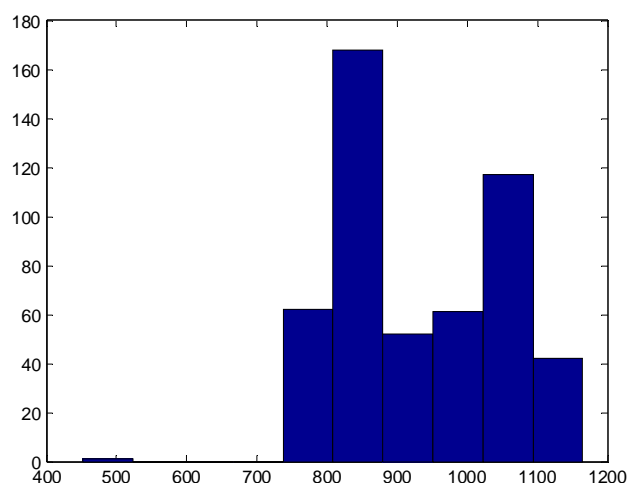
- 3.06.2013 - 30.05.2014,
- 26.10.2012 - 30.05.2014,
- 1.06.2012 - 30.05.2014 .

W badaniu dotyczącym rozkładów cen pszenicy otrzymano następujące współczynniki dwumodalności:

$$BC_{p,a} = 0,3128; \quad BC_{p,b} = 0,4176; \quad BC_{p,c} = 0,4698,$$

świadczące o jednomodalności cen pszenicy konsumpcyjnej w Polsce. Rysunek 4 przedstawia histogram rozkładu cen pszenicy w ostatnich dwóch latach.

Rysunek 4. Ceny pszenicy konsumpcyjnej w Polsce (1.06.2012 - 30.05.2014)



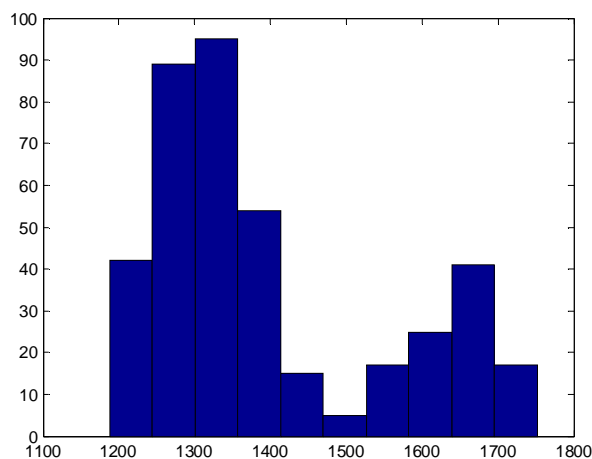
Źródło: opracowanie własne

W przypadku badania modalności cen złota w wyszczególnionych okresach otrzymano następujące wartości współczynnika BC :

$$BC_{z,a} = 0,3986 ; BC_{z,b} = 0,7415 ; BC_{z,c} = 0,7012 .$$

Rysunek 5 przedstawia histogram rozkładu cen złota w ostatnich dwóch latach, który wskazuje na wyraźną dwumodalność.

Rysunek 5. Ceny złota (1.06.2012 - 30.05.2014)



Źródło: opracowanie własne

UWAGI KOŃCOWE

Współczynnik dwumodalności BC stosuje się do badania modalności rozkładów zmiennych losowych. Dość prosta jego konstrukcja wykorzystująca tylko współczynnik asymetrii oparty na trzecim momencie centralnym i kurtozę zachęca do praktycznego stosowania w różnego rodzaju analizach statystycznych. Może on być traktowany jako podstawowe narzędzie w badaniu modalności rozkładów zmiennych losowych oraz jako metoda inicjująca lub wspomagająca bardziej zaawansowane metody statystyczne w analizach dotyczących zmiennych losowych. Na podstawie uzyskanych wyników można stwierdzić, że wartość współczynnika BC dość dobrze klasyfikowała rozkłady. Różnice pojawiały się jedynie w przypadku bardzo silnej asymetrii rozkładów jednomodalnych i braku wyraźnych „szczytów” w rozkładach dwumodalnych. Uogólnienie wyników na inne klasy rozkładów wymaga dalszych badań symulacyjnych, ale otrzymane rezultaty wskazują na możliwość zastosowania tego współczynnika do badania modalności rozkładów.

BIBLIOGRAFIA

- Baszczyńska A., Pekasiewicz D. (2014a) Statistical Inference About Modality of Random Variable, Proceedings of the 32nd International Conference Mathematical Methods in Economics 2014.
- Baszczyńska A., Pekasiewicz D. (2014b) Biavarage and Multimodality in Investigating Distribution of Electricity Prices, Folia Oeconomica, Acta Universitatis Lodziensis.
- Hartigan J., Hartigan P. (1985) The Dip Test of Unimodality, The Annals of Statistics, 13, 1, 70-84.
- Knapp T.R. (2007) Bimodality Revisited, Journal of Modern Applied Statistical Methods, 6, 1, 8 – 20.
- Pearson K. (1894) Contributions to the Mathematical Theory of Evolution: On the Dissection of Asymmetrical Frequency Curves, Philosophical Transaction of the Royal Society, A, 1, 185, 71-90.
- Silverman B. W. (1996) Density Estimation for Statistics and Data Analysis, Chapman & Hall, London.

**BIMODALITY COEFFICIENT BC AND ITS APPLICATION
IN ANALYSES OF RANDOM VARIABLES DISTRIBUTIONS**

Abstract: In the analyses of economic, natural and psychological phenomena, it is important to study the modality of distribution of a random variable identified with the analyzed statistical feature. One of the methods involves the determination of the coefficient of bimodality BC based on a random sample. Its value can decide about unimodality or bimodality of the analyzed variable distribution. The paper presents the results of simulation studies on the properties of the coefficient BC . Furthermore, the bimodality coefficient was used in the analyses of distributions of prices on commodity exchanges.

Keywords: modality, bimodality coefficient BC , mixtures of distributions