METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XI/2, 2010, str. 25–38

WPŁYW SPOSOBU NORMALIZACJI ZMIENNYCH NA OCENĘ REGIONALNEGO ZRÓŻNICOWANIA ROLNICTWA

Agata Binderman
Katedra Ekonometrii i Statystyki SGGW
e-mail: agata_binderman@sggw.pl

Streszczenie: W pracy zbadano problem dotyczący oceny regionalnego zróżnicowania rolnictwa w Polsce, ze względu na sposób normalizacji zmiennych i wybór klasyfikatora. Do analizy posłużono się wzorcowymi i bezwzorcowymi miernikami syntetycznymi, wykorzystując trzy różne sposoby normalizacji cech. Na podstawie syntetycznych mierników dokonano uporządkowania i klasyfikacji poszczególnych województw. Wyniki badań autorki dotyczące polskiego rolnictwa wykazują, że klasyfikacje i grupowania województw zależą w istotny sposób od metody normowania zmiennych. Praca stanowi kontynuację badań autorki na ten temat. Materiałem empirycznym w pracy były dane regionalne dotyczące stanu polskiego rolnictwa w roku 2008.

Słowa kluczowe: metody normalizacji zmiennych, mierniki syntetyczne, funkcje użyteczności, poziom rozwoju rolnictwa, klasyfikacja, grupowanie

WSTĘP

Do analizy zjawisk złożonych takich jak poziom rozwoju czy potencjał rolnictwa oraz oceny pod tym kątem województw, konieczne jest rozważenie wielu zmiennych diagnostycznych. Zmienne te, objaśniające dane zjawisko są zarówno mierzalne jak i niemierzalne. Analiza na podstawie tych danych może mieć charakter niejednoznaczny. Do sumarycznego charakteryzowania zjawisk złożonych stosuje się zmienne syntetyczne (agregatowe). Zastąpienie zestawu wielu zmiennych objaśniających badanego zjawiska przez zmienną syntetyczną daje pewną ocenę badanego zjawiska. Wadą zmiennych syntetycznych jest niejednoznaczność oraz to, że nie zawsze można im nadać interpretację merytoryczną. Istnieje wiele metod tworzenia zmiennych syntetycznych. Metody wzorcowe zakładają istnienie pewnego hipotetycznego obiektu wzorcowego, uporządkowania badanych obiektów dokonuje się w zależności od osiągniętych przez nie odległości od obiektu wzorcowego. Metody te wykorzystują odpowiednio wybrane zmienne diagnostyczne (objaśniające), charakteryzujące badane zjawisko, różnią się miedzy sobą, co do sposobu normalizacji zmiennych oraz postaci funkcji je agregujących [Kukuła 2000, Malina 2004, Młodak 2006, Nowak 1990, Zeliaś 2000].

Badania autorki dotyczące polskiego rolnictwa [Binderman 2005a,b 2006a,b, 2008] wykazały, że klasyfikacje i grupowania województw, uzyskiwane za pomocą metod wzorcowych w istotny sposób zależą zarówno od wybranej metryki, jak i od wyboru wzorca. W niniejszej pracy natomiast autorka podjęła temat zależności klasyfikacji obiektów opartej na miernikach syntetycznych od wyboru metody normalizacji cech.

METODYKA BADAŃ

Przyjmijmy założenie, że dane zjawisko jest opisane przez zmienne, które są symulantami, mającymi wartości nieujemne. Przy takim podejściu dany obiekt (obserwacja) badanego zjawiska jest opisany za pomocą wektora, będącego elementem przestrzeni $\Re^n_+ := \{\mathbf{x}=(x_1,x_2,...,x_n):\ x_i \geq 0,\ i=1,2,...,n\}$ gdzie $n \geq 1$ jest liczbą zmiennych zakwalifikowanych do oceny zjawiska.

Rozważmy problem polegający na klasyfikacji $m \in N$ obiektów $\mathbf{Q}_1,\mathbf{Q}_2,...,\mathbf{Q}_m$ badanego zjawiska za pomocą $n \in N$ zmiennych. Zgodnie z przyjętymi założeniami każdy taki obiekt daje się przedstawić za pomocą wektora należącego do przestrzeni $\Re^n_+$. Niech wektor $\mathbf{x}_i=(x_{i1},x_{i2},...,x_{in})$, $i=1,2,...,m$, opisuje i-ty obiekt $\mathbf{Q}_i$.

Jeżeli $x_{ik}>x_{jk}$ ($x_{ik} \geq x_{jk}$) dla $k=1,2,...,n$, to pisać będziemy:

$$\mathbf{x}_i > \mathbf{x}_j,\ (\mathbf{x}_i \geq \mathbf{x}_j),\ \text{gdzie } i, j \in [1,m].$$

Jeżeli $\mathbf{x}_i \geq \mathbf{x}_j$ i $\mathbf{x}_i \neq \mathbf{x}_j$, to naturalnym jest nazywać obiekt $\mathbf{x}_i$ lepszym (wyżej ocenianym) od obiektu $\mathbf{x}_j$. Oznacza to, że żadna z wartości składowych wektora $\mathbf{x}_i$ nie jest mniejsza od odpowiednich wartości składowych wektora $\mathbf{x}_j$, a przynajmniej jedna z nich ma wartość większą, tj. istnieje takie $k \in [1,n]$, że $x_{ik}>x_{jk}$.

Przyjmijmy następujące oznaczenia:

$$x_{m+1,k} := \max_{1 \leq i \leq m} x_{ik}, \quad x_{0,k} := \min_{1 \leq i \leq m} x_{ik}, \quad k = 1, 2, ..., n,$$

oraz $\mathbf{x}_0 := (x_{0,1},x_{0,2},...,x_{0,n})$, $\mathbf{x}_{m+1} := (x_{m+1,1},x_{m+1,2},...,x_{m+1,n})$.

Tak określone obiekty $\mathbf{x}_0$, $\mathbf{x}_{m+1}$ (być może fikcyjne) są nie gorsze oraz nie lepsze od pozostałych $\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_m$, tj. $\mathbf{x}_{m+1} \geq \mathbf{x}_i$ oraz $\mathbf{x}_i \geq \mathbf{x}_0$ dla każdego i: $m \geq i \geq 1$.

W dalszej części rozważań przyjmijmy założenie, że $\mathbf{Q}_{m+1}$, $\mathbf{Q}_0$ są różnymi obiektami wzorcowymi, reprezentowanymi przez odpowiednio wektory $\mathbf{x}_{m+1}$ i $\mathbf{x}_0$, ($\mathbf{x}_{m+1} \neq \mathbf{x}_0$).

W celu uporządkowania zbioru obiektów $\mathcal{Q} := \{\mathbf{Q}_0, \mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_m, \mathbf{Q}_{m+1}\}$ rozważmy dwie następujące funkcje [por. Binderman 2009, Kukuła 2000]:

$$u(\mathbf{x}) = \frac{<\mathbf{x}_{m+1} - \mathbf{x}_0, \mathbf{x} - \mathbf{x}_0>}{\left\| \mathbf{x}_{m+1} - \mathbf{x}_0 \right\|^2}, \tag{1}$$

$$v(\mathbf{x}) = \frac{1}{2} + \frac{d(\mathbf{x}_0, \mathbf{x}) - d(\mathbf{x}, \mathbf{x}_{m+1})}{2d(\mathbf{x}_0, \mathbf{x}_{m+1})}, \tag{2}$$

gdzie $<\mathbf{x},\mathbf{y}>$ - oznacza iloczyn skalarny wektorów $\mathbf{x}=(x_1,x_2,...,x_n)$; $\mathbf{y}=(y_1,y_2,...,y_n)\in\Re^n$, norma wektora $\|\mathbf{x}\|=(<\mathbf{x},\mathbf{x}>)^{1/2}$, natomiast $d_p(\mathbf{x},\mathbf{y})$ oznacza metrykę Minkowskiego, tj.:

$$d_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{k=1}^{n} \left(x_k - y_k \right)^p \right)^{1/p}, \tag{3}$$

gdzie $p \in [1, \infty)$.

Zauważmy, że powyższe funkcje mają następujące własności [zob. Binderman 2009]:

$$u(\mathbf{x}_0) = v(\mathbf{x}_0) = 0,\ u(\mathbf{x}_{m+1}) = v(\mathbf{x}_{m+1}) = 1,$$

$$0 \le u(\mathbf{x}_i) \le 1,\ 0 \le v(\mathbf{x}_i) \le 1 \quad \text{dla każdego } i=1,2,...,m.$$

Jeżeli $u(\mathbf{x}_i)=u(\mathbf{x}_j)$ to mówić będziemy, że obiekty $\mathbf{Q}_i$ i $\mathbf{Q}_j$ są obojętne (indyferentne) względem siebie. Jeżeli $u(\mathbf{x}_i)>u(\mathbf{x}_j)$ ($u(\mathbf{x}_i)\geq u(\mathbf{x})$) to obiekt $\mathbf{Q}_i$ uważać będziemy za lepszy (nie gorszy) od obiektu $\mathbf{Q}_j$, Fakty te zapisywać będziemy odpowiednio w następujący sposób:

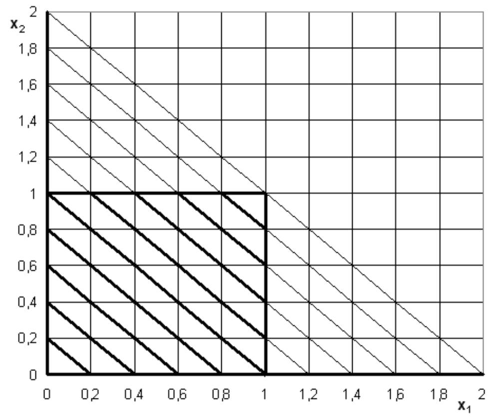
$$\mathbf{Q}_i \sim_u \mathbf{Q}_j,\ \mathbf{Q}_i \succ_u \mathbf{Q}_j,\ \mathbf{Q}_i \succeq_u \mathbf{Q}_j,\ i,j\in[0,1, ..., m, m+1],$$

gdzie funkcja $u(\mathbf{x})$ jest określona za pomocą wzoru (1).

Uporządkowane pary $(Q\sim_u)$, $(Q\succ_u)$, $(Q\succeq_u)$ nazywamy odpowiednio relacją indyferencji, słabej i silnej preferencji [Allen 1964, Panek 2000]. W podobny sposób określić można relacje preferencji $(Q\sim_v)$, $(Q\succ_v)$, $(Q\succeq_v)$, generowane przez funkcje $v(\mathbf{x})$, określoną za pomocą wzoru (2). Można pokazać, że jeżeli $\mathbf{x}_i\geq\mathbf{x}_j$ i $\mathbf{x}_i\neq\mathbf{x}_j$, gdzie $i, j \in [0, m+1]$, to $u(\mathbf{x}_i)>u(\mathbf{x}_j)$ oraz $v(\mathbf{x}_i)>v(\mathbf{x}_j)$. Oznacza to, że w polach słabej preferencji $(Q\succeq_u)$, $(Q\succeq_v)$ zachodzi *zjawisko niedosytu*. Funkcje u, v są *funkcjami użyteczności* dla relacji $\succeq_u$, $\succeq_v$, odpowiednio [Panek 2000].

Obszary obojętności dla wybranego obiektu, generowane przez funkcję użyteczności $u(\mathbf{x})$, indukującą relację preferencji leżą na hiperpłaszczyznach (u(x)=constans). W szczególności, na płaszczyznach dla n=3 oraz na prostych dla n=2 [Binderman 2007]. Dla n=2 krzywe obojętności są odcinkami leżącymi na prostych równoległych do prostej o równaniu $x_2=1-x_1$, patrz rysunek 1 (przy założeniu, że dla i=1,2,...,m: $\mathbf{0}\leq\mathbf{x}_i\leq\mathbf{1}$).

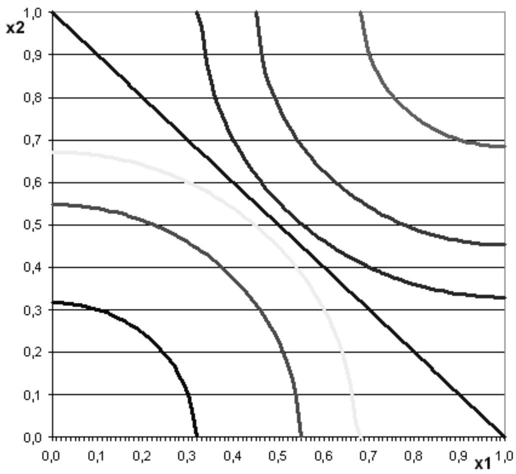
Rysunek 1. Krzywe obojętności liniowej funkcji użyteczności u: x_2=c-x_1



Źródło: opracowanie własne

Obszary obojętności generowane przez równanie $v(\mathbf{x})$=constans nie będą, jak w przypadku funkcji liniowej u, hiperpłaszczyznami. Jednak kiedy n=2, p=2 (tj. metryka Euklidesa [Bartosiewicz 1976]) krzywa obojętności wyznaczona przez równanie $v(\mathbf{x})=c$, $c\in[0,1]$ jest odcinkiem prostej o równaniu $x_2=1-x_1$ (przekątną kwadratu) dla c=0,5 oraz hiperbolą dla pozostałych c. Wynika to z faktu, że hiperbola jest miejscem geometrycznym punktów, których różnica odległości od dwóch stałych punktów zwanych ogniskami jest stała. Krzywe obojętności przedstawia rysunek 2 (przy założeniu, że dla i=1,2,...,m: $\mathbf{0}\leq\mathbf{x}_i\leq\mathbf{1}$).

Rysunek 2. Krzywe obojętności funkcji nieliniowej v



Źródło: opracowanie własne

Niniejsze rozważania pokazują, że rozpatrywane funkcje $u(\mathbf{x})$ i $v(\mathbf{x})$, określone za pomocą wzorów (1), (2) służyć mogą do liniowego uporządkowania obiektów $\mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_m$.

Złożone zjawiska ekonomiczne, określone są najczęściej za pomocą wielu zmiennych o różnych mianach oraz rzędach wielkości. Powstaje problem przekształcenia tych cech w taki sposób, aby rozpatrywane łącznie, spełniały warunek porównywalności. Transformację zmiennych diagnostycznych, która prowadzi do porównywalności oraz addytywności wartości tych cech nazywamy normalizacją lub normowaniem zmiennych [Borkowski, Dudek, Szczesny 2006, Kukuła 2000, Zeliaś 2002].

Problem normowania różnoimiennych zmiennych od strony metodologicznej ma zasadnicze znaczenie. Prawidłowo dobrana metoda normalizacyjna wpływa bezpośrednio na wyniki badań, dlatego normowanie cech diagnostycznych pełni często kluczową rolę w procedurach badawczych szczególnie, jeśli chodzi o analizę przestrzenną zjawisk ekonomiczno – społecznych. W trakcie badań empirycznych, po dokonaniu wyboru cech diagnostycznych, należy zdecydować, którą z wielu przedstawionych w literaturze przedmiotu metodę normowania użyć. Za pomocą analizy własności poszczególnych formuł normalizacyjnych trzeba zbadać, która metoda w danym przypadku będzie najlepsza.

W literaturze przedmiotu można wyróżnić cztery grupy metod normalizacji:
1. standaryzację,
2. unitaryzację,
3. przekształcenia ilorazowe względem punktu odniesienia,
4. rangowanie wartości zmiennych.

Wyczerpująca lista sposobów normalizacji zmiennych podana jest w książkach [Kukuła 2000, Młodak 2006, Zeliaś 2002].

Niech będzie dany układ wektorów $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m, \mathbf{x}_{m+1}$ opisujących obiekty $\mathbf{Q}_0, \mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_m, \mathbf{Q}_{m+1}$. Rozważmy trzy często stosowane sposoby normalizacji zmiennych [Kukuła 2000, Zeliaś 2002]:
1. unitaryzacji zerowanej

$$z_{ij} = \frac{x_{ij} - x_{0j}}{x_{m+1j} - x_{0j}}, \ 0 \le i \le m+1, \ 1 \le j \le n; \tag{4}$$

2. standaryzacji

$$t_{ij} = \frac{x_{ij} - x_{0j}}{S_j}, \tag{5}$$

gdzie $S_j^2 = \sum_{k=1}^{m} \left(x_{kj} - \overline{x}_j\right)^2, \ \overline{x}_j = \frac{1}{m}\sum_{k=1}^{m} x_{kj}, \ \ 0 \le i \le m+1, \ 1 \le j \le n;$

3. przekształcenia ilorazowego

$$r_{ij}=\frac{x_{ij}-x_{0j}}{\sum_{k=1}^{m}x_{kj}},\quad 0\le i\le m+1;1\le j\le n. \tag{6}$$

W wyniku normalizacji zmiennych przy pomocy wspomnianych wyżej reprezentantów, trzech sposobów normalizacji zmiennych, otrzymujemy trzy różne układy wektorów:

- $\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m, \mathbf{z}_{m+1}$ – dla unitaryzacji zerowanej,
- $\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_m, \mathbf{t}_{m+1}$ – dla standaryzacji,
- $\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m, \mathbf{r}_{m+1}$ – dla przekształcenia ilorazowego.

Funkcje użyteczności $u(\mathbf{x})$ i $v(\mathbf{x})$ określone za pomocą wzorów (1) i (2) dla wektorów $\mathbf{z}_0,\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_m,\mathbf{z}_{m+1}$ przyjmują (w zależności od przyjętej metryki typu Minkowskiego) następujące postacie:

$$u(\mathbf{z}_i)=\frac{1}{n}\sum_{k=1}^{n}z_{ik}, \tag{7}$$

$$v(\mathbf{z}_i)=\frac{1}{2n^{1/p}}\left[n^{1/p}+\left(\sum_{j=1}^{n}z_{ij}^{p}\right)^{1/p}-\left(\sum_{j=1}^{n}\left(1-z_{ij}\right)^{p}\right)^{1/p}\right],\ 1\le p<\infty, \tag{8}$$

W celu otrzymania prostych wzorów na użyteczności obiektów po normalizacji zmiennych typu standaryzacja i przekształcenia ilorazowe, dokonajmy następujących transformacji:

$$t'_{ij}=\frac{x_{ij}-x_{0j}}{\alpha S_j}, \tag{9}$$

$$\text{gdzie}: S_j^2=\sum_{k=1}^{m}\left(x_{kj}-\bar{x}_j\right)^2, \bar{x}_j=\frac{1}{m}\sum_{k=1}^{m}x_{kj}, 0\le i\le m+1;1\le j\le n, \alpha^p=\sum_{k=1}^{n}\left(\frac{x_{m+1k}-x_{0k}}{S_k}\right)^p,$$

oraz

$$r'_{ij}=\frac{x_{ij}-x_{0j}}{\beta s_j}, \tag{10}$$

$$\text{gdzie: } s_j=\sum_{k=1}^{m}x_{kj},\ 0\le i\le m+1;\ 1\le j\le n,\ \beta^p=\sum_{k=1}^{n}\left(\frac{x_{m+1k}-x_{0k}}{s_k}\right)^p\ \ 1\le p<\infty.$$

Po takiej zamianie zmiennych transformacje wektorów wzorcowych mają następujące własności:

$$\mathbf{t}'_0=0,\quad d(\mathbf{t}'_0,\mathbf{t}'_{m+1})=1,\quad \mathbf{r}'_0=0,\quad d(\mathbf{r}'_0,\mathbf{r}'_{m+1})=1,$$

gdzie metryka d_p określona jest za pomocą wzoru (3).

Znormalizowane funkcje użyteczności $u(\mathbf{x})$ i $v(\mathbf{x})$ według wzorów (1) i (2) przyjmują dla wektorów $\mathbf{t}'_i$, $\mathbf{r}'_i$ ($i = 0,1,\ldots m+1$) następujące postacie:

$$u(\mathbf{t}_i^{'}) = \frac{1}{\left\|\mathbf{t}_{m+1}^{'}\right\|^2}\sum_{k=1}^{n} t_{m+1k}^{'} t_{ik}^{'}, \tag{11}$$

$$v(\mathbf{t}_i^{'}) = \frac{1}{2}\left[1 + \left(\sum_{j=1}^{n}(t_{ij}^{'})^p\right)^{1/p} - \left(\sum_{j=1}^{n}\left(t_{m+1j}^{'} - t_{ij}^{'}\right)^p\right)^{1/p}\right], \tag{12}$$

$$u(\mathbf{r}_i^{'}) = \frac{1}{\left\|\mathbf{r}_{m+1}^{'}\right\|^2}\sum_{k=1}^{n} r_{m+1k}^{'} r_{ik}^{'}, \tag{13}$$

$$v(\mathbf{r}_i^{'}) = \frac{1}{2}\left[1 + \left(\sum_{j=1}^{n}(r_{ij}^{'})^p\right)^{1/p} - \left(\sum_{j=1}^{n}\left(r_{m+1j}^{'} - r_{ij}^{'}\right)^p\right)^{1/p}\right]. \tag{14}$$

W pracy dokonano porównania wyników, otrzymanych za pomocą rozważanych wyżej mierników syntetycznych i sposobów normalizacji zmiennych, z jednowzorcową metodą Zdzisława Hellwiga [Hellwig 1968]. W tym celu przekształcono zmienne za pomocą wzoru:

$$w_{ij} = \frac{x_{ij} - x_{0j}}{\frac{1}{\sqrt{m}} S_j} = \sqrt{m} \cdot t_{ij}, \tag{15}$$

gdzie: $i = 0,1,...,m+1;\ j = 1,2,...,n.$

Na tej podstawie otrzymano znormalizowaną funkcję Hellwiga [Binderman 2009], określoną za pomocą wzoru:

$$H(\mathbf{w}_i) := \frac{h(\mathbf{w}_i) - h(\mathbf{w}_0)}{1 - h(\mathbf{w}_0)},\ i = 0,1,...,17; \tag{16}$$

gdzie: $h(\mathbf{w}_i) = 1 - \frac{\sqrt{m}\cdot c_i}{\sqrt{m}\cdot \overline{c} + 2d_2(\mathbf{c}, \overline{\mathbf{c}})}$, $c_i = d_2(w_i, w_{m+1})$, $\mathbf{c} = (c_1, c_2, ..., c_m)$,

$\overline{c} = \frac{1}{m}\sum_{k=1}^{m} c_k$, $\overline{\mathbf{c}} = (\overline{c}, \overline{c}, ..., \overline{c}) \in \Re_+^m$, $i = 0,1,2,...,m+1;$

d_2 – jest metryką Euklidesa, określoną za pomocą wzoru (3), dla p=2.

WYNIKI BADAŃ

Materiałem empirycznym w pracy były dane Głównego Urzędu Statystycznego, dotyczące czynników opisujących polskie rolnictwo w 2008 roku, w podziale na szesnaście województw. Czynniki zostały odpowiednio wybrane, a dane przeliczone tak, że wszystkie zmienne diagnostyczne miały charakter stymulant. Oznacza to, że wyższe wartości zmiennej wskazywały na wyższy poziom badanego zjawiska (uzasadnienie wyboru cech opisano w pracy [Binderman 2007]). Tymi zmiennymi były poniższe czynniki.

X_1 Udział użytków rolnych w % powierzchni ogółem.
X_2 Wskaźnik waloryzacji rolniczej przestrzeni produkcyjnej (w punktach).
X_3 Plony zbóż w tonach z 1 hektara.
X_4 Obsada bydła w sztukach dużych na 100 hektarów użytków rolnych.
X_5 Skup owoców z drzew w kg na 1 hektar powierzchni upraw.
X_6 Wartość skupu produktów rolnych ogółem w zł na 1 ha użytków rolnych.
X_7 Nakłady inwestycyjne w rolnictwie w zł na 1 hektar użytków rolnych.
X_8 Wskaźnik zatrudnienia w rolnictwie w %.
X_9 Średnia powierzchnia gospodarstwa rolnego w hektarach.
X_{10} Produkt krajowy brutto w zł na 1 mieszkańca.

W ten sposób otrzymano m=16 obiektów $\mathbf{Q}_1,\mathbf{Q}_2,...,\mathbf{Q}_{16}$, gdzie każdy był opisany przez $n = 10$ cech: $X_1, X_2,..., X_{10}$. Wartości przyjętych zmiennych diagnostycznych dla 16 województw w 2008 roku utworzyły następnie macierz $\mathbf{X}$ o wymiarach 16×10. Na podstawie wartości zmiennych w 16 województwach, stworzono dwa hipotetyczne (wzorcowe) obiekty, minimalne $\mathbf{Q}_0$ oraz maksymalne $\mathbf{Q}_{17}$, określone za pomocą najmniej i najbardziej korzystnego zestawu wartości cech (odpowiednio). Hipotetyczne województwa reprezentowane były przez wektory $\mathbf{x}_0$ oraz $\mathbf{x}_{17}$, każdy o 10 składowych.

Ze względu na różne miana i rzędy wielkości, zmienne poddano normalizacji, oczywiście dla celów pracy wykorzystując trzy różne metody: unitaryzację zerowaną - wg wzoru (4), standaryzację - wzory (5) i (15) oraz przekształcenie ilorazowe – wg wzoru (6).

Do konstrukcji mierników syntetycznych województw w 2008 roku, określających poziom polskiego rolnictwa wykorzystano rozważane wcześniej dwie znormalizowane funkcje użyteczności $u(\mathbf{x})$ i $v(\mathbf{x})$, dla trzech różnych sposobów normalizacji, tj. $u(\mathbf{z}_i)$ wg wzoru (7), $v(\mathbf{z}_i)$ wzór - (8), po unitaryzacji, $u(\mathbf{t}_i')$ wg wzoru (11), $v(\mathbf{t}_i')$ - wzór (12), po standaryzacji, $u(\mathbf{r}_i')$ - wzór (13), $v(\mathbf{r}_i')$ - wzór (14), po przekształceniu ilorazowym oraz znormalizowaną funkcję Hellwiga $H(\mathbf{w}_i)$ wzór (16).

Wszystkie w ten sposób wyznaczone mierniki były znormalizowanymi funkcjami użyteczności, przyjmującymi wartości z przedziału [0,1].

Na podstawie wartości mierników syntetycznych województw w następnym kroku dokonano uporządkowania obiektów i na tej podstawie otrzymano klasyfikację województw ze względu na poziom rolnictwa w 2008 roku. Pojedynczy klasyfikator (miernik syntetyczny) może być daleki od „optymalnego”, natomiast kombinacje wielu, dają na ogół klasyfikator bliski „optymalnemu” [por. Breiman 1966, 1968]. Z tego względu, jako ostateczny (wynikowy) miernik poziomu rozwoju rolnictwa dla poszczególnych województw, w badaniach posłużono się dodatkowym miernikiem, określonym za pomocą wzoru:

$$\bar{\mathbf{m}}_i := \bar{\mathbf{m}}(\mathbf{x}_i) = \frac{H(\mathbf{w}_i)+u(\mathbf{t}_i)+v(\mathbf{t}_i)+u(\mathbf{z}_i)+v(\mathbf{z}_i)+u(\mathbf{r}_i)+v(\mathbf{r}_i)}{7}, \tag{17}$$

gdzie: $i = 1,2,...,16$.

W tabeli 1 podane zostały wartości mierników polskich województw w roku 2008, otrzymanych za pomocą rozważanych ośmiu funkcji.

Tabela 1. Mierniki syntetyczne województw w 2008 roku

Województwo	$H(\mathbf{w}_i)$	$u(\mathbf{t}_i)$	$v(\mathbf{t}_i)$	$u(\mathbf{z}_i)$	$v(\mathbf{z}_i)$	$u(\mathbf{r}_i)$	$v(\mathbf{r}_i)$	$\bar{m}_i$
Dolnośląskie	0,359	0,402	0,411	0,390	0,400	0,248	0,268	**0,354**
Kujawsko-pomorskie	0,400	0,442	0,447	0,448	0,453	0,457	0,459	**0,444**
Lubelskie	0,314	0,394	0,412	0,419	0,433	0,477	0,481	**0,418**
Lubuskie	0,167	0,182	0,204	0,187	0,211	0,178	0,211	**0,191**
Łódzkie	0,348	0,395	0,405	0,408	0,418	0,406	0,410	**0,399**
Małopolskie	0,301	0,334	0,348	0,340	0,355	0,295	0,309	**0,326**
Mazowieckie	0,426	0,518	0,516	0,531	0,526	0,626	0,610	**0,536**
Opolskie	0,430	0,541	0,534	0,522	0,518	0,296	0,326	**0,452**
Podkarpackie	0,166	0,190	0,221	0,189	0,219	0,151	0,169	**0,186**
Podlaskie	0,325	0,406	0,421	0,437	0,446	0,603	0,591	**0,461**
Pomorskie	0,374	0,386	0,390	0,400	0,403	0,427	0,431	**0,402**
Śląskie	0,374	0,425	0,433	0,410	0,421	0,272	0,306	**0,377**
Świętokrzyskie	0,242	0,272	0,294	0,277	0,299	0,236	0,246	**0,267**
Warmińsko-mazurskie	0,353	0,407	0,418	0,434	0,442	0,600	0,586	**0,463**
Wielkopolskie	0,512	0,601	0,588	0,608	0,595	0,583	0,577	**0,581**
Zachodnio-pomorskie	0,256	0,296	0,318	0,309	0,334	0,312	0,344	**0,310**

Źródło: obliczenia własne

W tabeli 2 podane zostały podstawowe charakterystyki opisowe mierników syntetycznych opisujących poziom rolnictwa polskich województw w roku 2008. Natomiast w tabeli 3 współczynniki korelacji Pearsona między wektorami wyników.

Tabela 2. Charakterystyki opisowe dla syntetycznych mierników województw

Charakterystyki opisowe	$H(\mathbf{w}_i)$	$u(\mathbf{t}_i)$	$v(\mathbf{t}_i)$	$u(\mathbf{z}_i)$	$v(\mathbf{z}_i)$	$u(\mathbf{r}_i)$	$v(\mathbf{r}_i)$	$\bar{m}_i$
Średnia arytmet.	0,334	0,387	0,394	0,397	0,412	0,394	0,399	0,385
Mediana	0,351	0,398	0,406	0,411	0,419	0,409	0,413	0,400
Rozstęp	0,347	0,419	0,397	0,384	0,363	0,421	0,397	0,394
Odchylenie stand.	0,090	0,111	0,105	0,101	0,094	0,111	0,105	0,107
Wsp. zmienności	27%	29%	27%	25%	23%	28%	26%	28%
Wsp. asymetrii	-0,287	-0,135	-0,212	-0,215	-0,216	-0,265	-0,301	-0,295

Źródło: obliczenia własne

Tabela 3. Współczynniki korelacji Pearsona

	$H(\mathbf{w}_i)$	$u(\mathbf{t}_i)$	$v(\mathbf{t}_i)$	$u(\mathbf{z}_i)$	$v(\mathbf{z}_i)$	$u(\mathbf{r}_i)$	$v(\mathbf{r}_i)$	$\bar{\mathfrak{m}}_i$
$H(\mathbf{t}_i)$	1							
$u(\mathbf{t}_i)$	0,981	1						
$v(\mathbf{t}_i)$	0,976	0,999	1					
$u(\mathbf{z}_i)$	0,970	0,993	0,994	1				
$v(\mathbf{z}_i)$	0,965	0,991	0,993	0,999	1			
$u(\mathbf{r}_i)$	0,648	0,674	0,684	0,756	0,760	1		
$v(\mathbf{r}_i)$	0,662	0,691	0,701	0,770	0,775	0,998	1	
$\bar{\mathfrak{m}}_i$	0,926	0,947	0,951	0,978	0,979	0,875	0,885	1

Źródło: obliczenia własne

Otrzymane wyniki (patrz tabela 2) wskazują, że w Polsce, niezależnie od zastosowanych mierników, poziom rolnictwa w województwach był mocno zróżnicowany. Wskazuje na to między innymi wielkość rozstępu. Dla danego miernika syntetycznego, przy ustalonym sposobie normalizacji zmiennych, różnica pomiędzy maksymalnymi a minimalnymi wartościami mierników była większa od średniej arytmetycznej. Potwierdzają to również wielkości współczynników zmienności, które wskazują, że odchylenie standardowe stanowi (zależnie od funkcji) od 23% do 29% średniej arytmetycznej.

Ocena asymetrii rozkładów wartości mierników województw pokazuje, że w badanym okresie rozkłady charakteryzują się słabą asymetrią lewostronną (ujemną), tj. skupienie badanych województw przy wyższych niż średnie wartościach cechy. Lewostronny asymetryczny rozkład wartości mierników wskazuje, że poziom rolnictwa większości województw w 2008 roku był wyższy od średniej. Świadczą o tym również wartości mediany, które dla wszystkich funkcji, a także dla miernika $\bar{\mathfrak{m}}_i$, miały wartości powyżej średniej arytmetycznej.

Na podstawie obliczeń przedstawionych w tabeli 3 zauważamy, że współczynniki korelacji dla wszystkich mierników są bardzo wysokie. Można wnioskować, że w ramach tej samej metody normalizacji wyniki nie różnią się od siebie znacząco. Warto zauważyć, że wyniki otrzymane przy pomocy metody unitaryzacji zerowanej oraz funkcji nieliniowej, opartej na dwóch wzorcach, wykazują największą sumaryczną korelację w stosunku do pozostałych mierników. Najmniejszą korelację z pozostałymi wynikami wykazują rezultaty otrzymane poprzez ilorazową normalizację zmiennych.

W celu analizy różnic pomiędzy różnymi sposobami liczenia mierników zbadano również, jak zmienia się uporządkowanie województw. W tabeli 4 podane zostały pozycje poszczególnych województw ze względu na wartości rozważanych mierników.

Tabela 4. Uporządkowanie województw w 2008 roku

Województwo	$H(\mathbf{w}_i)$	$u(\mathbf{t}_i)$	$v(\mathbf{t}_i)$	$u(\mathbf{z}_i)$	$v(\mathbf{z}_i)$	$u(\mathbf{r}_i)$	$v(\mathbf{r}_i)$	$\overline{\mathbf{m}}_i$
Wielkopolskie	1	1	1	1	1	4	4	**1**
Mazowieckie	3	3	3	2	2	1	1	**2**
Warmińsko-mazurskie	8	6	7	6	6	3	3	**3**
Podlaskie	10	7	6	5	5	2	2	**4**
Opolskie	2	2	2	3	3	10	10	**5**
Kujawsko-pomorskie	4	4	4	4	4	6	6	**6**
Lubelskie	11	10	8	7	7	5	5	**7**
Pomorskie	6	11	11	10	10	7	7	**8**
Łódzkie	9	9	10	9	9	8	8	**9**
Śląskie	5	5	5	8	8	12	12	**10**
Dolnośląskie	7	8	9	11	11	13	13	**11**
Małopolskie	12	12	12	12	12	11	11	**12**
Zachodnio-pomorskie	13	13	13	13	13	9	9	**13**
Świętokrzyskie	14	14	14	14	14	14	14	**14**
Lubuskie	15	16	16	16	16	15	15	**15**
Podkarpackie	16	15	15	15	15	16	16	**16**

Źródło: obliczenia własne

Przy użyciu wyznaczonych mierników dla województw, dokonano następnie podziału województw na 4 klasy, charakteryzujące się (w obrębie klasy) zbliżonym poziomem rozwoju rolnictwa. Podstawą podziału województw na klasy były wartości syntetycznych mierników przy wykorzystaniu tzw. metody E. Nowaka [Nowak 1990]. Grupowanie województw przy wszystkich sposobach obliczania mierników przedstawia tabela 5.

Tabela 5. Klasyfikacja województw w 2008 roku

Województwo	$H(\mathbf{w}_i)$	$u(\mathbf{t}_i)$	$v(\mathbf{t}_i)$	$u(\mathbf{z}_i)$	$v(\mathbf{z}_i)$	$u(\mathbf{r}_i)$	$v(\mathbf{r}_i)$	$\overline{\mathbf{m}}_i$
Wielkopolskie	I	I	I	I	I	I	I	**I**
Mazowieckie	I	I	I	I	I	I	I	**I**
Warmińsko-mazurskie	II	II	II	II	II	II	II	**II**
Podlaskie	III	II	II	II	II	II	II	**II**
Opolskie	I	I	I	I	I	I	I	**II**
Kujawsko-pomorskie	II	II	II	II	II	II	II	**II**
Lubelskie	III	II	II	II	II	II	II	**II**
Pomorskie	II	III	III	III	III	II	II	**II**
Łódzkie	II	II	II	II	II	II	II	**II**
Śląskie	II	II	II	II	II	II	II	**III**
Dolnośląskie	II	II	II	II	II	III	III	**III**
Małopolskie	III	III	III	III	III	III	III	**III**
Zachodnio-pomorskie	III	III	III	III	III	III	III	**III**
Świętokrzyskie	IV	IV	IV	IV	III	IV	IV	**IV**
Lubuskie	IV	IV	IV	IV	IV	IV	IV	**IV**
Podkarpackie	IV	IV	IV	IV	IV	IV	IV	**IV**

Źródło: obliczenia własne

Na podstawie uzyskanych wyników można sądzić, że dla rozpatrywanego zestawu zmiennych diagnostycznych wybór sposobu normalizacji, miał większy wpływ na uporządkowanie województw niż wybór miernika. Natomiast na rezultat grupowania, nie miał znaczącego wpływu ani wybór sposobu normalizacji, ani miernika syntetycznego. Jedyne niewielkie różnice w przyporządkowaniu do danej grupy występują w przypadku zastosowania metody Hellwiga oraz mierników wykorzystujących normalizację metodą przekształcenia ilorazowego.

Uporządkowanie województw według wartości funkcji porządkującej $\overline{m}_i$ (patrz tabela 4) wskazuje, że najlepszym regionem pod względem poziomu rolnictwa w 2008 roku było województwo wielkopolskie. Wysoką drugą pozycje zajęło następnie województwo mazowieckie. Wspomniane wyżej województwa pod względem poziomu rolnictwa odstają „na plus" od pozostałych województw. Dwa ostatnie miejsca pod względem poziomu rolnictwa w 2008 roku, zajęły województwa: lubuskie i podkarpackie, które miały dużo niższe od pozostałych wartości mierników syntetycznych.

Analizując podział województw na grupy (patrz tabela 5), ze względu na wartości funkcji porządkującej $\overline{m}_i$ można wnioskować, że do I grupy zaliczają się dwa województwa – wielkopolskie oraz mazowieckie, co potwierdza, że odznaczają się najwyższym poziomem rozwoju rolnictwa w Polsce. Do grupy II zaliczonych zostało siedem województw: warmińsko-mazurskie, podlaskie, opolskie, kujawsko-pomorskie, lubelskie, pomorskie, łódzkie, które charakteryzują się stosunkowo wysokim (powyżej średniej) poziomem rozwoju rolnictwa. W III grupie znalazły się cztery województwa o niskim, poniżej przeciętnego poziomie rolnictwa. Było to województwo śląskie, dolnośląskie, małopolskie i zachodnio–pomorskie. Grupa IV obejmuje trzy województwa: świętokrzyskie, lubuskie i podkarpackie, co wskazuje na to, że w 2008 roku były to regiony, które miały najniższy poziom rozwoju rolnictwa w Polsce.

WNIOSKI

Na podstawie przeprowadzonych badań można wnioskować, że przy przyjętym zestawie zmiennych diagnostycznych wybór metody normalizacyjnej miał większy wpływ na uporządkowanie województw niż wybór miernika syntetycznego, natomiast wybór sposobu normalizacji i miernika nie miał znaczącego wpływu na wynik grupowania. Wyniki otrzymane przy tym samym sposobie normalizacji nie różnią się znacząco między sobą.

Opierając się na otrzymanych rezultatach można twierdzić, że najwyższym poziomem rolnictwa w 2008 roku charakteryzowało się województwo wielkopolskie. Wraz z województwem mazowieckim wyraźnie odbiegało swoim poziomem od pozostałych województw. Najniższym poziomem rolnictwa charakteryzowały się województwa podkarpackie i lubuskie. One również wyraźnie odbiegały swoim poziomem od pozostałych województw.

Otrzymane wyniki pokazują duże zróżnicowanie polskiego rolnictwa w 2008 roku. Dla potwierdzenia tego zjawiska wskazana jest dalsza analiza występujących zależności.

LITERATURA

Allen R. G. D. (1964) Ekonomia matematyczna, PWN, Warszawa.

Bartosiewicz S. (1976) Propozycja metody tworzenia zmiennych syntetycznych, Prace Naukowe AE we Wrocławiu, nr 84, Wrocław.

Binderman A. (2005a) Klasyfikacja polskich województw według poziomu rozwoju rolnictwa, RNR, Seria G., T.92, Z.1, str.42, Warszawa.

Binderman A. (2005b) O problemie wyboru wzorca przy badaniu przestrzennego zróżnicowania potencjału rolnictwa w Polsce, MIBE V, Warszawa, str. 46.

Binderman A. (2006a) Wykorzystanie funkcji użyteczności do badania przestrzennego zróżnicowania rolnictwa, RNR SERiA, T.VIII, Z.5, Warszawa, s.5.

Binderman A. (2006b) Klasyfikacja obiektów oparta na dwóch wzorcach, EiOGŻ, Zeszyty Naukowe SGGW, nr 60, Warszawa, str. 25.

Binderman A. (2007) Wielowymiarowa analiza regionalnego zróżnicowania rolnictwa w Polsce, praca doktorska, SGGW, Warszawa.

Binderman A. (2008) Zastosowanie liniowej i nieliniowej funkcji użyteczności do badania poziomu rolnictwa w Polsce, MIBE IX, wyd. SGGW, Warszawa, str. 29.

Binderman A. (2009) Dynamika rozwoju rolnictwa w Polsce po akcesji do Unii Europejskiej, Roczniki Nauk Rolniczych, SERiA, Tom XI, Zeszyt 3.

Borkowski B, Dudek H., Szczesny W. (2004) Ekonometria. Wybrane zagadnienia, PWN, Warszawa.

Borkowski B., Dudek H., Szczesny W. (2006) O pewnych problemach przekształcania wartości cech, ACTA AGRARIA ET SILVESTRIA, vol. XLVII.

Breiman L. 1994, Bagging predictors, Technical Report 420, Departament of Statistics, University of California, CA, USA.

Breiman L. 1996, Arcing Classifiers, Technical Report 460, Departament of Statistics, University of California, CA, USA.

Hellwig Z. (1968) Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę kwalifikowanych kadr, „Przegląd Statystyczny", z. 4.

Kukuła K. (2000) Metoda unitaryzacji zerowanej, PWN, Warszawa.

Malina A. (2004) Wielowymiarowa analiza przestrzennego zróżnicowania struktury gospodarki Polski według województw, AE, Seria Monografie nr 162, Kraków.

Młodak A. (2006) Analiza taksonomiczna w analizie regionalnej, Difin, Warszawa.

Nowak E. (1990) Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych, PWE, Warszawa.

Panek E. (2000) Ekonomia matematyczna, Akademia Ekonomiczna, Poznań.

Zeliaś A. (2000) Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym, Wyd. Akademii Ekonomicznej, Kraków.

Zeliaś A. (2002) Some notes on the selection of normalization of diagnostic variables, Statistics in transition, Vol 5, No. 5, pp 787 – 802.

Dependence of development level analysis of Polish agriculture on choice of ways of normalization

Abstract:: The paper studied the problem of stability of evaluations of regional diversification of Polish agriculture with regard to the way of variables normalization. The methods applied use three types of normalization. The analysis is based on synthetic measures that use two patterns of objects. On the basis of synthetic measures there were made arrangement and classification of different voivodeships. The subject of this study was regional variability of Polish voivodeships in terms of level of agricultural development in the 2008 year. As diagnostic variables, factors affecting the development of agriculture were used. The paper is a continuation of the author's research in this field.

Key words: methods of normalization, synthetic measures, utility functions, agriculture development level, classification, class division