

Tomasz Jurkiewicz
Ewa Wycinka
Katedra Statystyki
Uniwersytet Gdański

PROPOZYCJA WYKORZYSTANIA METOD ANALIZY WIELOWYMIAROWEJ DO DOBORU ZMIENNYCH W BADANIU STOPNIA INTEGRACJI RYNKÓW UBEZPIECZENIOWYCH

1. Geneza europejskiego rynku ubezpieczeń

Utworzenie wspólnego rynku ubezpieczeń wewnątrz Wspólnoty Europejskiej przewidziane zostało już w Traktacie o utworzeniu Wspólnoty Europejskiej. Rozwój wspólnotowych unormowań prawnych dotyczących ubezpieczeń postępował w sposób ewolucyjny. Ewolucja ta prowadziła do coraz większej liberalizacji obrotu ubezpieczeniowego na terenie Wspólnoty Europejskiej¹. W efekcie jednolity rynek ubezpieczeń funkcjonuje we Wspólnocie Europejskiej od 1993 roku. Jego działanie oparte jest na trzech podstawowych zasadach tworzących Unię Europejską:

- zasadzie swobody tworzenia podmiotów gospodarczych (osiedlenia się),
- zasadzie swobody świadczenia usług,
- zasadzie przepływu kapitału między krajami.

W dziedzinie ubezpieczeń celem rynku wewnętrznego jest zagwarantowanie wszystkim mieszkańcom Wspólnoty dostępu do możliwie najszerszej gamy wysokiej jakości produktów ubezpieczeniowych oferowanych przez zakłady ubezpieczeń z obszaru całej Wspólnoty². Zawierane na tej podstawie transakcje ubezpieczeniowe mają być objęte jednakową ochroną prawną i finansową. Przepływy środków finansowych jakiegokolwiek ze stron transakcji nie mogą być ograniczane. Ubezpieczyciele upoważnieni do działania w jakimkolwiek państwie członkowskim mogą prowadzić swoją działalność na terenie całej Wspólnoty i podlegają takim samym zasadom nadzoru. Rynek wewnętrzny ubezpieczeń realizowany jest również poprzez umożliwienie agentom ubezpieczeniowym i innym pośrednikom działania na jednakowych zasadach w obrębie całego rynku³.

Gwarantowane dyrektywami jednakowe warunki rozwoju sektora ubezpieczeń nie znajdują pełnego odzwierciedlenia w regulacjach wewnętrznych państw członkowskich. Część z nich, poprzez sprzyjające regulacje podatkowe i administracyjne, jest zdecydowanie częściej wybierana przez zakłady ubezpieczeń jako państwo siedziby. Pozostaje więc otwarte pytanie, czy europejski rynek ubezpieczeń jest organizmem jednolitym? Próba odpowiedzi na to pytanie wiąże się z oceną stopnia integracji rynków ubezpieczeniowych w Unii Europejskiej. W celu dokonania takiej oceny posłużyc się

¹ R. Śwital, Wspólnotowe prawo ubezpieczeń gospodarczych – ewolucja unormowań dyrektywalnych, orzecznictwo Trybunału Sprawiedliwości, „Wiadomości Ubezpieczeniowe” 1997 nr 11/12, s.13

² Na podstawie: E. Wierzbicka, Rynek ubezpieczeń w okresie transformacji w Polsce (w:) A. Wernik, Polityka finansowa w okresie transformacji – doświadczenia lat dziewięćdziesiątych, Instytut Finansów, WSUiB, Warszawa 1999, s.141

³ Zob. Biała Księga. Przygotowanie krajów stowarzyszonych Europy Środkowej i Wschodniej do integracji z Rynkiem Wewnętrznym Unii Europejskiej, *Prawo, Ubezpieczenia, Reasekuracja* 1997, nr 1, s.3-4. Na ten temat również: C. Parsons, (De)regulation of insurance in the EC & The single insurance market, Warszawa 1996, s.1; K. Kruczalak, Dostosowanie polskiego prawa ubezpieczeniowego do prawa europejskiego (prawa wspólnotowego), (w:) A. Wąsiewicz (red.), *Ubezpieczenia w gospodarce rynkowej*, tom 3, Bydgoszcz 1997, s.413

można metodami statystycznymi.

Celem niniejszego referatu jest analiza zmiennych, które mogą być zastosowane do oceny stopnia integracji rynków ubezpieczeniowych, oraz metod statystycznych używanych do doboru zmiennych diagnostycznych.

2. Zestaw potencjalnych zmiennych

Rynki ubezpieczeniowe opisywane są przez szereg zmiennych o charakterze miar bezwzględnych i względnych. Wykorzystanie tylko miar względnych (wskaźniki struktury i natężenia) umożliwi porównywanie rynków różnej wielkości, a wynik analizy nie jest ukierunkowany na preferowanie największych rynków. Ponadto, aby wyeliminować wpływ różnych rzędów wielkości zmiennych, zmienne powinny zostać poddane standaryzacji.

Selekcja merytoryczna doprowadziła do wyboru następujących zmiennych⁴:

A1 – składka przypisana brutto / PNB – najwyższy poziom tego wskaźnika w Luksemburgu, Wielkiej Brytanii i Holandii jest spowodowany tym, że krajowe firmy przyjmują zagraniczne ryzyka korzystając ze swobody świadczenia usług, nie jest to więc odzwierciedleniem realiów ekonomicznych danych państw, a efektem korzystania przez zakłady z swobody świadczenia usług;

A2 – inwestycje zakładów ubezpieczeń / PNB;

A3 – akcje posiadane przez zakłady ubezpieczeń⁵ / kapitalizacja rynku – jest to miara udziału zakładów ubezpieczeń jako inwestorów instytucjonalnych na rynku kapitałowym;

A4 – liczba pracowników zakładów ubezpieczeń / liczba zatrudnionych w usługach × 100%;

A5 – wielkość składki przypisanej brutto / liczba mieszkańców – wysoka wartość tego wskaźnika w Luksemburgu i Szwajcarii wynika z tego, że część umów jest zawierana na rzecz nierezydentów;

A6 – liczba pracowników zakładów ubezpieczeń na 100 tys. ludności;

B1 – wielkość rezerw techniczno - ubezpieczeniowych / składka przypisana brutto w ubezpieczeniach na życie;

B2 - liczba zakładów ubezpieczeń na 100 tys. ludności;

B3 – koncentracja rynku ubezpieczeń na życie mierzona udziałem pięciu największych zakładów;

B4 - koncentracja rynku ubezpieczeń majątkowych mierzona udziałem pięciu największych zakładów;

B5 – udział składek ubezpieczeń na życie w składkach ogółem;

B6 – inwestycje ubezpieczeń na życie / inwestycje ubezpieczeniowych ogółem;

B7 – inwestycje ubezpieczeń na życie / składka w ubezpieczeniach na życie – wskaźnik ten informuje o długości kontraktów i wielkościach kapitałów inwestowanych przez zakłady;

B8 - inwestycje ubezpieczeń majątkowych / składka w ubezpieczeniach majątkowych;

B9 – udział ubezpieczeń komunikacyjnych w ubezpieczeniach ogółem;

⁴ Do obliczeń wykorzystano dane liczbowe z roku 1999 zaczerpnięte z European Insurance in Figures, Basic Data 2000/Complete 1999 Data, CEA 2001. Zmienne oznaczone literą A charakteryzują znaczenie sektora ubezpieczeń w gospodarce, oznaczone literą B strukturę rynków ubezpieczeniowych.

⁵ Bez akcji własnych, chyba, że zostały zakupione w celach komercyjnych

Tabela 1

Wielkości zmiennych wybranych do badania									
kraj		A1	A2	A3	A4	A5	A6		
Austria	AT	5,6%	23,7%	43,4%	1,2%	1356	3,42		
Belgia	BE	7,3%	41,2%	16,6%	0,9%	1661	2,47		
Niemcy	DE	6,6%	41,2%	19,1%	1,0%	1587	2,92		
Dania	DK	6,3%	79,6%	54,1%	0,7%	1942	2,56		
Hiszpania	ES	5,7%	13,1%	1,1%	0,6%	820	1,19		
Finlandia	FI	8,4%	60,0%	8,8%	0,7%	1985	2,15		
Francja	FR	8,5%	58,5%	14,7%	0,8%	1892	2,20		
Wielka Brytania	GB	13,7%	109,9%	31,1%	1,2%	3306	3,87		
Grecja	GR	2,2%	4,2%	0,9%	0,4%	236	0,88		
Irlandia	IE	9,7%	48,1%	34,6%	1,1%	2267	3,02		
Włochy	IT	5,6%	18,5%	5,6%	0,3%	1073	0,74		
Luksemburg	LU	33,2%	119,5%	22,3%	1,4%	13188	4,26		
Holandia	NL	10,2%	66,5%	11,3%	0,8%	2413	2,98		
Portugalia	PT	6,5%	19,8%	7,6%	0,6%	669	1,41		
Szwecja	SE	6,9%	88,7%	29,5%	0,6%	1797	1,91		
kraj	B1	B2	B3	B4	B5	B6	B7	B8	B9
AT	5,71%	0,95	49,4%	55,8%	44,1%	68,9%	6,6	2,3	33,0%
BE	6,53%	2,14	73,2%	56,7%	59,3%	71,7%	6,9	3,9	35,6%
DE	8,51%	0,94	31,0%	26,8%	45,1%	61,8%	8,6	4,4	27,8%
DK	15,15%	4,94	55,5%	71,5%	61,9%	89,3%	18,2	3,5	28,9%
ES	3,75%	0,94	35,2%	27,8%	53,5%	76,0%	3,3	1,2	43,8%
FI	5,84%	1,28	99,7%	88,9%	79,2%	75,3%	6,8	8,4	35,0%
FR	8,22%	0,87	56,3%	55,1%	65,5%	85,9%	9	2,8	35,9%
GB	9,34%	1,40	42,3%	41,8%	75,2%	90,4%	9,7	3,1	30,0%
GR	2,90%	1,08	70,4%	46,4%	55,6%				65,2%
IE		4,35	64,3%	65,2%	71,7%	86,7%	6	2,3	52,4%
IT	3,91%	0,43	28,7%	35,4%	57,5%	76,0%	4,3	1,9	60,4%
LU	3,67%	21,46	44,3%	0,0%	87,6%	91,6%	3,8	2,4	29,5%
NL	6,96%	3,27	62,6%	48,5%	59,2%	88,7%	9,7	1,8	22,2%
PT	3,71%	0,88	78,5%	55,4%	56,2%	76,5%	4,1	1,6	50,5%
SE	14,56%	5,34	69,6%	81,6%	71,7%	81,0%	14,5	8,6	31,7%

Źródło: opracowanie własne na podstawie European Insurance in Figures, Basic Data 2000/ Complete 1999 Data, CEA 2001.

3. Konwencjonalne metody doboru zmiennych diagnostycznych

Zmienne wybrane do badania powinny charakteryzować się następującymi właściwościami⁶:

- duża zmienność – usunięcie cech, które nie różnicują badanych obiektów (państw),
- niezależność zmiennych - wyeliminowanie zjawiska powtarzania tych samych informacji niesionych przez różne zmienne.

Zmienność cech badać można m.in. współczynnikiem zmienności. Wśród wytypowanych zmiennych żadna nie ma charakteru cechy stałej, odchylenie standardowe stanowi od kilkunastu do stu kilkudziesięciu procent średniej.

Najprostszą miarą badania skorelowania cech jest współczynnik korelacji

⁶ W literaturze zaleca się pozostawienie jak najmniejszej liczby zmiennych reprezentatywnych dla zmiennych odrzuconych, czyli najsilniej z nimi związanych. Jednakże gdy celem badania jest wszechstronna analiza integracji rynków ubezpieczeń celowe jest pozostawienie jak największej liczby zmiennych niosących szeroki zakres informacji, pod warunkiem, że spełniają one statystyczne wymogi doboru cech diagnostycznych (por. Z. Hellwig, Wielowymiarowa analiza porównawcza i jej zastosowanie w badaniach wielocechowych obiektów gospodarczych, PWE, Warszawa 1981, E. Nowak, Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych, PWE, Warszawa 1990).

liniowej Pearsona. Mierzy on siłę i kierunek zależności liniowej pomiędzy parami zmiennych.

Tabela 2

Macierz współczynników korelacji															
	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6	B7	B8	B9
A1	1														
A2	0,72	1													
A3	0,13	0,48	1												
A4	0,67	0,56	0,52	1											
A5	0,99	0,69	0,16	0,63	1										
A6	0,68	0,72	0,60	0,95	0,65	1									
B1	-0,13	0,53	0,72	0,06	-0,11	0,23	1								
B2	0,91	0,65	0,22	0,51	0,95	0,54	-0,01	1							
B3	-0,17	-0,01	-0,11	-0,19	-0,18	-0,15	0,03	-0,10	1						
B4	-0,54	-0,04	0,27	-0,31	-0,55	-0,22	0,51	-0,45	0,73	1					
B5	0,70	0,76	0,10	0,31	0,66	0,38	0,14	0,63	0,30	0,07	1				
B6	0,53	0,70	0,26	0,20	0,48	0,34	0,26	0,50	0,06	-0,03	0,69	1			
B7	-0,23	0,46	0,64	-0,10	-0,18	0,13	0,98	-0,06	0,13	0,51	0,05	0,29	1		
B8	-0,12	0,33	0,10	-0,14	-0,08	-0,04	0,51	-0,02	0,53	0,61	0,36	-0,17	0,44	1	
B9	-0,37	-0,68	-0,46	-0,57	-0,35	-0,71	-0,56	-0,29	0,05	-0,02	-0,18	-0,16	-0,54	-0,31	1

Źródło: opracowanie własne.

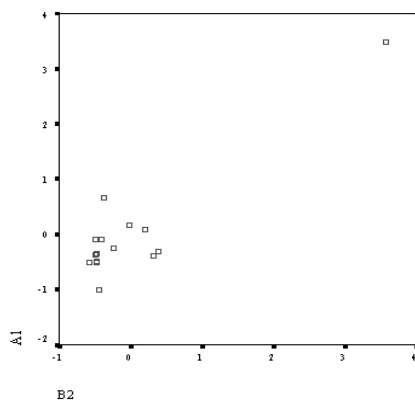
Z tablicy 2. wynika, że parą najbardziej skorelowanych ze sobą zmiennych jest para A1, A5 ($r = 0,99$). Jedna zmienna zależy od drugiej w 97% (r^2), co oznacza, że zaledwie 3% niesionych przez nie informacji jest różna. Do badania należy wziąć jedną z nich, mniej skorelowaną z pozostałymi zmiennymi. Obydwie cechy są w podobnym stopniu powiązane z innymi cechami, jednakże w przypadku cechy A5 występuje obserwacja mocno odstająca od pozostałych (ang. outlier). Ponieważ wiele metod analizy wielowymiarowej jest nieodporna na wartości odstające, tym samym wpływają negatywnie na jakość oszacowań, z dalszego badania usunięto zmienną A5.

Kolejną parą silnie skorelowanych zmiennych była para B1, B7 ($r = 0,98$). Z uwagi na silniejsze skorelowanie cechy B1 z pozostałymi zmiennymi, została ona usunięta ze zbioru danych. W przypadku kolejnej pary silnie skorelowanych zmiennych A4, A6 ($r = 0,95$) nieznacznie silniej związana z pozostałymi była zmienna A6. Ponieważ jednak liczba zatrudnionych w usługach zależy głównie od struktury sektorowej gospodarki kraju, racjonalnym wydawało się pozostawienie w analizie zmiennej A6, czyli liczby zatrudnionych w sektorze ubezpieczeń odniesionej do liczby mieszkańców.

Kolejną parę powiązanych zmiennych tworzyły cechy A1 i B2 ($r = 0,91$), jednakże merytoryczna analiza wskazywała na brak logicznego związku pomiędzy oboma wskaźnikami. Wyjaśnieniem pozornej zależności okazało się występowanie w przypadku obu zmiennych wartości odstających⁷. Usunięcie obserwacji dla Luksemburga powodowało zmniejszenie wartości współczynnika korelacji Pearsona do poziomu 0,19 co świadczy praktycznie o braku zależności (por. wykres 1). Stąd też pozostawiono do dalszej analizy obydwie cechy⁸.

⁷ Jeżeli w zbiorze danych występuje wartość odstająca, to m. in. współczynnik korelacji będzie tym silniejszy, im bardziej odstaje ta obserwacja od pozostałych.

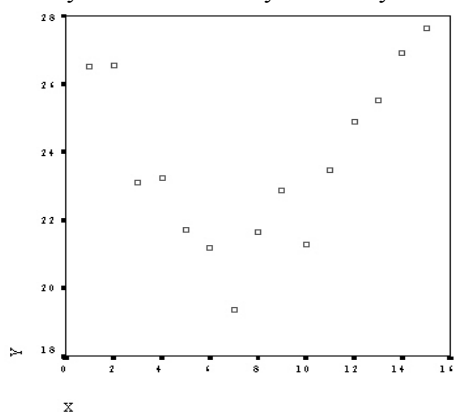
⁸ Z uwagi na nieodporność niektórych metod na wartości skrajne dalsze analizy wykonywano w dwóch wariantach, zarówno na pełnym zbiorze danych oraz po odrzuceniu wartości skrajnych i następnie porównywano ich rezultaty. Proste odrzucenie wartości skrajnych byłoby niewłaściwe (por. K. Jajuga, Statystyczna analiza wielowymiarowa, PWN, Warszawa 1993, N. R. Draper, H. Smith, Analiza regresji stosowana, PWN, Warszawa 1973).



Wykres 1. Korelacja między zmiennymi A1 i B2
 Źródło: opracowanie własne.

Pozostałe pary cech skorelowane były ze sobą w znacznie mniejszym stopniu ($r < 0,77$), tak więc kolejny etap procesu redukcji liczby zmiennych został wstępnie zakończony.

Współczynnik korelacji liniowej nie pozwala na ustalenie występowania zależności innych niż liniowe. Przykładem takiej zależności, może być zależność występująca między zmiennymi o charakterze stymulanta i nominanta⁹. Przykład takiej zależności można zaobserwować na wykresie 2. Obliczony współczynnik korelacji liniowej ma wartość 0,24, co oznaczałoby brak zależności. Tymczasem indeks korelacji krzywoliniowej dla funkcji postaci $y = |x - 7,16| + 19,94$ wynosi 0,95 zaś dla funkcji kwadratowej 0,93. Analiza wykresów korelacyjnych dla wszystkich par zmiennych nie wykazała występowania istotnych zależności krzywoliniowych.



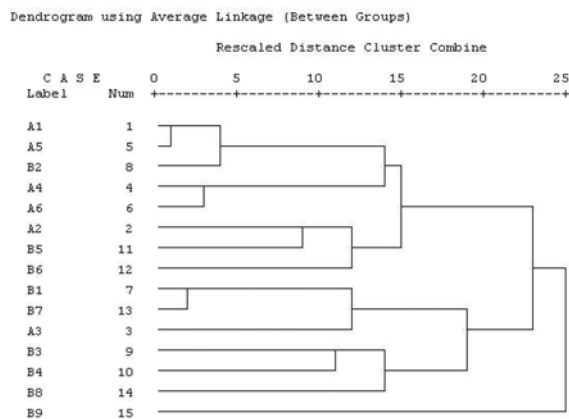
Wykres 2. Przykład zależności między stymulantą (X) a nominantą (Y)
 Źródło: opracowanie własne na podstawie danych umownych.

⁹ Stymulanta jest to zmienna, której wysokie wartości świadczą o wyższym poziomie rozwoju obiektu badania. Przeciwnieństwem stymulanta jest destymulanta, dla której wysokie wartości świadczą o niższym poziomie rozwoju. Nominanta jest cechą diagnostyczną, dla której wartości mieszczące się w pewnym ustalonym przedziale, lub równe pewnej wartości, świadczą o wyższym poziomie rozwoju obiektu, zaś wartości niższe bądź wyższe są właściwe dla obiektów o niższym poziomie rozwoju (por. T. Michalski, Ubezpieczenia gospodarcze w Polsce i w Unii Europejskiej, Difin, Warszawa 2001).

4. Wielowymiarowe metody doboru zmiennych diagnostycznych

Obok klasycznych metod, opartych na analizie macierzy korelacji, w doborze zmiennych diagnostycznych można wykorzystać także szereg metod analizy wielowymiarowej. Często stosowaną metodą jest analiza skupisk (ang. cluster). Metoda ta pozwala na łączenie obiektów (w prezentowanym badaniu – zmiennych) w coraz większe skupiska z zastosowaniem miar odległości. Miary te wyznaczają odległość między zmiennymi w n-wymiarowej przestrzeni utworzonej z n obiektów (krajów) objętych badaniem. Do poniższego badania wykorzystano odległości euklidesowe pomiędzy cechami. Na wykresie 3. przedstawiono wynik grupowania zmiennych. Im szybciej dochodzi do połączenia zmiennych na wykresie tym większe jest między nimi podobieństwo. Wyniki są zbieżne z analizą macierzy korelacji i sugerują redukcję zbioru zmiennych. Jednakże pomimo, że metoda ta jest szybsza i łatwiejsza w odczytywaniu wyników, to podobnie jak przy analizie współczynników korelacji nie uwzględnia ona zależności krzywoliniowych oraz nie podpowiada, którą ze zmiennych tworzących parę usunąć. W wyniku grupowania hierarchicznego oprócz dendrogramu otrzymujemy również macierz odległości między zmiennymi. Suma odległości od pozostałych zmiennych, podobnie jak suma współczynników korelacji, może być kryterium decydującym o odrzuceniu bądź pozostawieniu zmiennej.

Drugą z zastosowanych metod analizy skupień było grupowanie metodą k-średnich. Metoda ta, dla założonej z góry ilości skupień, grupuje zmienne tak, by skupiska były jak najbardziej jednorodne i jednocześnie różniły się między sobą maksymalnie. Tym samym metoda ta dąży do uzyskania najbardziej istotnych wyników analizy wariacji. Metoda ta wymaga założenia z góry ilości skupień, przy czym optymalną ich liczbę można ustalić stosując indeks Davisa-Bouldinga¹⁰.



Wykres 3. Wyniki grupowania hierarchicznego zmiennych uwzględnionych w badaniu
Źródło: opracowanie własne.

Zmienne wchodzące w skład utworzonych skupisk są najbardziej podobne do siebie, i tym samym można je w ramach skupienia redukować. Obok numeru skupienia podawana jest także odległość danej zmiennej od środka skupienia. Miarę tą można wykorzystać jako

¹⁰ D. L. Davis, D. W. Boulding, A Cluster Separation Measure, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, 1979, s. 224-227

kryterium wyboru zmiennej charakterystycznej dla skupiska, przy czym kryterium to można zastosować wyłącznie w przypadku skupisk liczących więcej niż dwie zmienne. Jednocześnie odległość zmiennych od środka skupiska pozwala na ocenę stopnia podobieństwa zmiennych. Porównując skupisko 4 i 6 (patrz tabela 3) można zaobserwować, że zmienne A4 i A6 są zdecydowanie bardziej podobne niż B4 i B8.

Tabela 3

Wyniki grupowania zmiennych metodą k-średnich

Zmienna	Skupienie	Odległość
B3	1	,000
A2	2	1,564
B5	2	1,627
B6	2	1,759
A3	3	1,980
B1	3	1,033
B7	3	1,224
A4	4	,582
A6	4	,582
B9	5	,000
B4	6	1,685
B8	6	1,685
A1	7	,740
A5	7	,331
B2	7	,941

Źródło: opracowanie własne.

Obok odległości zmiennych od środka skupienia analizować można także macierz odległości pomiędzy skupieniami. Pozwala ona ocenić, które grupy zmiennych wyraźnie różnią się od siebie, a które są do siebie podobne (patrz tabela 4). Przykładowo skupienie drugie jest stosunkowo blisko siódmego, zaś najbardziej odległe są od siebie czwarte i piąte.

Tabela 4

Odległości pomiędzy centrami skupień

Skupienie	1	2	3	4	5	6	7
1		4,928	5,259	5,888	5,333	2,876	5,830
2	4,928		3,938	3,705	6,212	4,618	2,829
3	5,259	3,938		4,341	6,576	3,606	5,210
4	5,888	3,705	4,341		6,986	5,594	3,280
5	5,333	6,212	6,576	6,986		5,574	6,296
6	2,876	4,618	3,606	5,594	5,574		5,894
7	5,830	2,829	5,210	3,280	6,296	5,894	

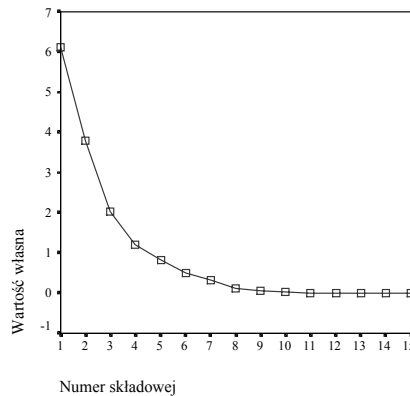
Źródło: opracowanie własne.

Odmianą metodą grupowania zmiennych, możliwą do zastosowania, jest wykorzystanie sieci neuronowej typu SOM (self-organizing map)¹¹. Zaletą tej metody jest zdecydowanie mniejsza wrażliwość na braki danych, obserwacje odstające czy zależności nieliniowe. Problem stwarza jednak konieczność starannego, empirycznego doboru parametrów sieci, co wymaga przeprowadzenia dużej ilości obliczeń dla poszczególnych

¹¹ T. Kohonen, Self-Organizing Maps, Springer Verlag, 1997

wariantów. Postać wyników tej metody jest analogiczna do metody grupowania k-średnich¹².

Metodą pozwalającą także na ocenę zmiennych w badaniu jest analiza czynnikowa (ang. factor analysis). Analiza czynnikowa służy identyfikacji zmiennych zwanych czynnikami, które wyjaśniają korelacje występujące w ramach zbiorów obserwowanych zmiennych. Metoda ta jest często wykorzystywana w redukcji danych w celu identyfikacji niewielkiej liczby czynników, wyjaśniających większą część wariacji zmiennych. Może być również wykorzystywana do ustalania hipotez dotyczących mechanizmów przyczynowo-skutkowych lub klasyfikowania zmiennych do dalszych analiz.



Wykres 4. Wykres osypiska
Źródło: opracowanie własne.

Ładunki czynnikowe (po rotacji Varimax z normalizacją Kaisera)

Tabela 5

Zmienna	Składowa				
	1	2	3	4	5
A1	,880	-,173	,387	-,199	-,011
A2	,762	,479	,368	,033	-,082
A3	,009	,652	,607	,030	,234
A4	,331	-,055	,918	-,109	,032
A5	,874	-,145	,359	-,228	-,051
A6	,401	,150	,885	-,083	,021
B1	,010	,966	,122	,112	-,104
B2	,862	-,044	,251	-,177	-,022
B3	,038	-,061	-,078	,948	-,058
B4	-,312	,433	-,090	,825	,000
B5	,885	,075	,036	,369	,032
B6	,696	,319	-,045	,074	,593
B7	-,028	,960	-,007	,136	-,021
B8	,122	,373	-,099	,576	-,667
B9	-,222	-,470	-,615	,072	,372

Źródło: opracowanie własne.

¹² Wyniki uzyskane z nałożenia siatki neuronów na obiekty są ostatecznie grupowane metodą k-średnich dla optymalnej (ustalonej indeksem Davisa-Bouldinga) ilości klas.

Pierwszym etapem analizy jest określenie ilości czynników na podstawie wykresu osypiska (wykres 4). Test osypiska polega na znalezieniu miejsca, na prawo od którego regularny spadek wartości własnych staje się wolniejszy. Zatem, nie powinno się wyodrębniać więcej czynników niż znajduje się po lewej stronie tego punktu.

Analiza wskazuje, że istotne znaczenie ma pierwszych pięć czynników, wyjaśniających łącznie 93% zmienności całkowitej. W tabeli 5 przedstawione są ładunki czynnikowe, których analiza jest kolejnym etapem oceny zmiennych. Wyróżnione zostały najwyższe wartości ładunków, odpowiadające im zmienne składają się na dany czynnik. Przykładowo zmienne A1, A2, A5, B2, B5, B6 mieszczą się w pierwszym czynniku, co jednocześnie świadczy o występujących między nimi zależnościach.

5. Wnioski

Prawidłowy dobór zmiennych jest warunkiem uzyskania poprawnych rezultatów dalszych analiz¹³. Przedstawione powyżej wybrane metody statystycznego doboru cech dają zbliżone rezultaty. Jednocześnie wyniki analizy wskazują, że metody wielowymiarowe, jakkolwiek także nieodporne na występowanie wartości odstających, zależności nieliniowych itp. pozwalają na pełniejszą i szybszą redukcję liczby zmiennych.

Stosując statystyczne metody doboru zmiennych nie należy poprzestawać na pobieżnej analizie wyników, ale szukać uzasadnienia merytorycznego dla znalezionych zależności statystycznych. W przeciwnym przypadku wyniki badania mogą być obciążone przez zależności pozorne. Cenna jest również graficzna analiza diagramów korelacyjnych, pozwalająca na wykrycie zależności nieliniowych, które nie są ujawniane przez większość metod. Analiza wykresów pozwala również na szybkie ujawnienie wartości skrajnych w zbiorach danych.

Ostateczny zestaw zmiennych zależy jednak zawsze od badacza. Należy pamiętać, że to samo zagadnienie może zostać opisane odmiennie, jeżeli wybrane zostaną dwa różniące się zestawy zmiennych diagnostycznych. Tym większą wagę należy przykładac do prawidłowego ich doboru.

¹³ Rezultaty badania stopnia integracji rynków ubezpieczeniowych były prezentowane m.in. w: Jurkiewicz T., Wycinka E. (2003) *Analiza podobieństw rynków ubezpieczeniowych w krajach Unii Europejskiej i krajach kandydujących z wykorzystaniem metod analizy wielowymiarowej* w: Inwestycje finansowe i ubezpieczenia - tendencje światowe a polski rynek. Prace Naukowe Akademii Ekonomicznej we Wrocławiu 991/2003 s. 209-218.