



O KORZYŚCIACH WYNIKAJĄCYCH Z ZASTOSOWANIA ANALIZY PROFIT

Piotr Jabkowski, Uniwersytet im. A. Mickiewicza w Poznaniu

Celem referatu jest ukazanie korzyści z wykorzystania analizy PROFIT (PROperty FITting) w badaniach rynkowych (segmentacja i percepcja rynku), marketingowych (segmentacja klientów oraz potencjalnych grup odbiorców produktów i usług), jak też w badaniach opinii publicznej (typologie, klasyfikacje, podobieństwa). Procedura analizy PROFIT łączy niemetryczne skalowanie wielowymiarowe MDS oraz analizę regresji wielorakiej, ułatwiając interpretację uzyskanego rozrzutu obiektów. O ile bowiem celem skalowania wielowymiarowego jest graficzna prezentacja struktury podobieństwa (lub odmienności) pomiędzy analizowanymi obiektami względem wybranego zbioru zmiennych (skalowanie wielowymiarowe dąży do takiego uporządkowania obiektów, przy danej liczbie wymiarów, by najlepiej odtworzyć pierwotne obserwowane odległości pomiędzy obiektami), to celem analizy PROFIT jest interpretacja uzyskanego rozrzutu obiektów oraz wyjaśnienie wymiarów mapy podobieństw MDS.

Poza statystycznymi podstawami analizy PROFIT przedstawiony zostanie także przykład praktycznej implementacji omawianej procedury. Analiza ta zostanie wykorzystana do segmentacji rynku usług komunalnych w Poznaniu, prowadzonej w oparciu o oceny wyrażane przez odbiorców tego rodzaju usług, tj. mieszkańców miasta. Poszczególne firmy i przedsiębiorstwa oceniane były z punktu widzenia ich: (a) nowoczesności, (b) konkurencyjności, (c) jakości świadczonych usług, (d) dynamiki rozwoju, (e) dbania o klienta oraz (f) wiarygodności.

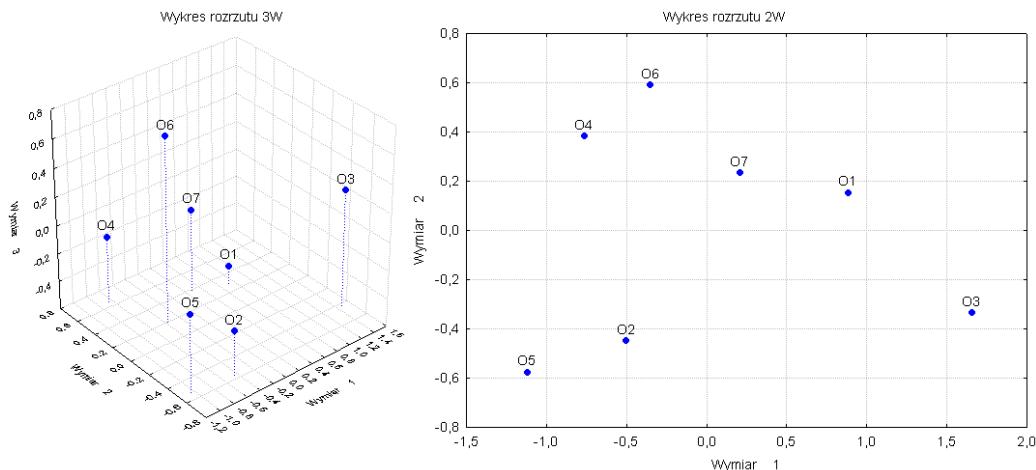
Wprowadzenie do analizy PROFIT

Procedurę *PROFIT* podzielić można na dwa etapy. W etapie pierwszym wykonywane jest klasyczne skalowanie wielowymiarowe, natomiast w etapie drugim wykorzystuje się równania regresyjne i odtwarza na osiach skalowania wielowymiarowego wejściowy zbiór zmiennych (cech obiektów).

Etap I – skalowanie wielowymiarowe

Skalowanie wielowymiarowe jest w swojej naturze podobne do analizy czynnikowej, z tą różnicą, że powiązania pomiędzy poszczególnymi obiektami mogą być wyrażone nie tylko poprzez macierze korelacji, ale także jako macierze dowolnych miar odległości, np. euklidesowej, kwadratu odległości euklidesowej, miejskiej Manhattan, Czebyszewa, niezgodności procentowej, potęgowej i innych. Jest to o tyle istotne, że skalowanie wielowymiarowe pozwala porównywać obiekty nie tylko względem cech ilościowych, ale także tych jakościowych. Zakres wykorzystania skalowania wielowymiarowego (analizy *PROFIT*) jest więc szerszy, niż zakres zastosowań analizy czynnikowej.

Głównym celem skalowania wielowymiarowego jest graficzna prezentacja struktury podobieństwa (lub odmienności) pomiędzy analizowanymi obiektami względem wybranego zbioru zmiennych (cech). Generalnie zatem rzecz ujmując, skalowanie wielowymiarowe dąży do takiego uporządkowania obiektów, by jednocześnie zredukować liczbę wymiarów i możliwie najlepiej odtworzyć pierwotne obserwowane odległości (różnice) pomiędzy obiektami. Graficzna prezentacja wyników przyjmuje dobrze znaną postać wykresu rozrzutu obiektów, nazywaną „mapą skalowania wielowymiarowego”. Mapa taka, najczęściej 2- lub 3-wymiarowa, ma bardzo prostą interpretację. Przyjmuje się bowiem, że im mniejsza odległość pomiędzy badanymi obiektami, tym są one bardziej do siebie podobne. W ten sposób można wyznaczać grupy (klastry) obiektów, obiekty izolowane itd. Przykładowe mapy skalowania wielowymiarowego, dla dwóch oraz trzech wymiarów, zaprezentowane zostały na poniższym rysunku.



Rys. 1. Przykładowe mapy MDS.

Prowadząc analizę skalowania wielowymiarowego, należy zawsze sprawdzić, czy wybrana n - k -wymiarowa ($k = 1, 2, \dots, n-1$) przestrzeń (mapa) w sposób dostatecznie dobry odzwierciedla pierwotny układ n -wymiarowy. W praktyce dąży się najczęściej do dwu lub trójwymiarowej przestrzeni, wtedy bowiem wyniki skalowania wielowymiarowego interpretuje się w sposób zupełnie intuicyjny.

Jakość dopasowania danych odtworzonych do danych wejściowych mierzy się przy pomocy tzw. funkcji *STRESSu*, przy czym im mniejsza jej wartość, tym lepsze dopasowanie macierzy odległości odtworzonych do macierzy odległości obserwowanych. Funkcję *STRESSu* definiuje się przy tym najczęściej jako pierwiastek standaryzowanej sumy kwadratów reszt pomiędzy odległościami wejściowymi a odległościami odtworzonymi w wyniku skalowania wielowymiarowego. Przyjmuje ona postać:

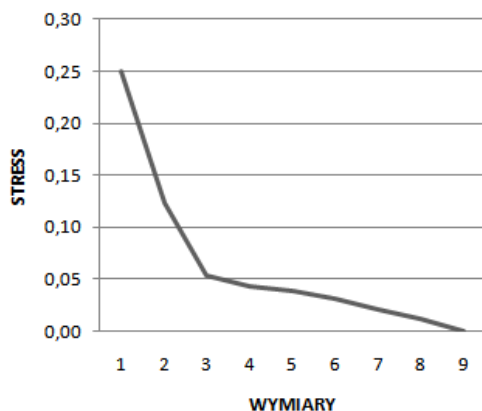
$$\phi = \sqrt{\frac{\sum \sum (d_{ij} - f(\delta_{ij}))^2}{\sum \sum d_{ij}^2}}$$

gdzie:

- ♦ d_{ij} oznacza odtworzoną odległość pomiędzy punktami i oraz j na mapie *MDS* przy danej liczbie $n-k$ wymiarów, przy czym $k=1, 2, \dots, n-1$;
- ♦ δ_{ij} oznacza odległość pomiędzy punktami i oraz j dla danych wejściowych (odległości obserwowane);
- ♦ $f(\delta_{ij})$ jest funkcją na danych wejściowych, przy czym zależy ona od tego, czy ma się do czynienia z metrycznym czy też niemetrycznym skalowaniem wielowymiarowym. W metrycznym skalowaniu wielowymiarowym $f(\delta_{ij}) = \delta_{ij}$, natomiast przy skalowaniu niemetrycznym funkcja ta jest monotoniczną transformacją obserwowanych danych wejściowych.

Jeżeli hipotetycznie przyjmiemy, że mapa *MDS* w sposób idealny odzwierciedla odległości obserwowane, wtedy $f(\delta_{ij}) - d_{ij} = 0$ i stąd funkcja *STRESSu* będzie też równa 0. Wartość funkcji *STRESSu* jest zatem wskaźnikiem dopasowania do danych wejściowych.

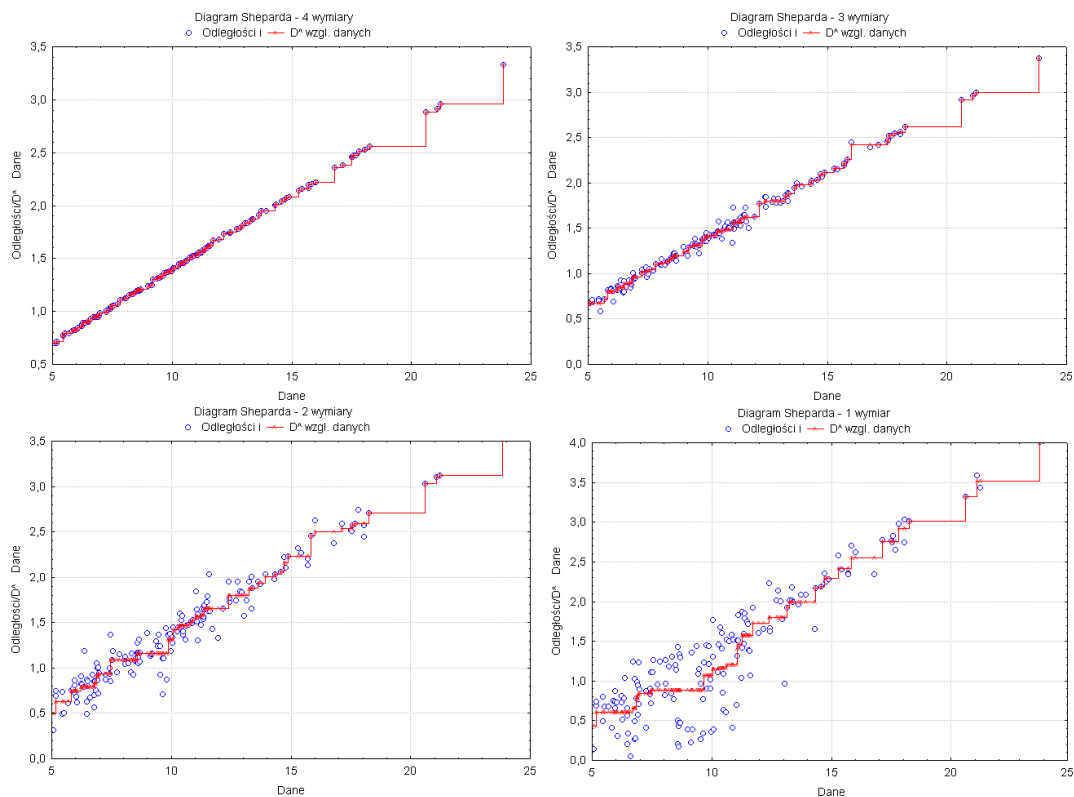
Najczęściej wykorzystywanym sposobem ustalenia liczby wymiarów koniecznych do prawidłowego skalowania wielowymiarowego jest tzw. test osypiska *STRESSu*, który polega na wyznaczeniu wartości *STRESSu* względem różnych liczb wymiarów – od jednego wymiaru, do liczby wymiarów pomniejszonej o jeden względem liczby zmiennych wejściowych. Powszechnie przyjmuje się przy tym, że liczba wymiarów konieczna do prawidłowego odwzorowania obserwowanych odległości odpowiada takiemu punktowi na wykresie osypiska, w którym wartości *STRESSów* przestają wyraźnie maleć (tworzą linię „w miarę” poziomą względem osi odciętych). Przykładowy test osypiska przedstawiony został na poniższym rysunku. Można z niego odczytać, że dla symulowanych danych pożądana liczba wymiarów skalowania wynosi 3.



Rys. 2. Wykres osypiska STRESSu.

Do określenia liczby wymiarów pomocny jest również tzw. diagram Sheparda, który przedstawia zależność pomiędzy miarą odległości z pierwotnego (źródłowego) zbioru danych a odległością odtworzoną na podstawie skalowania wielowymiarowego. Odpowiednie punkty na diagramie przedstawione są w postaci okręgów. Łamana (tzw. funkcja schodkowa) pokazuje monotonicznie przekształcone dane wejściowe. Im bardziej punkty skupiają się względem funkcji schodkowej, tym lepsze dopasowanie całego modelu, tzn. lepsze odtworzenie pierwotnych odległości pomiędzy obiektami przy danej liczbie wymiarów. Oczywiście jest, że najlepsze odtworzenie pierwotnych odległości otrzyma się przy większej liczbie wymiarów (wtedy punkty opisujące zależność odległości pierwotnych względem odległości odtworzonych na podstawie skalowania wielowymiarowego będą praktycznie pokrywać się z funkcją schodkową), natomiast im bardziej będzie się redukowało liczbę wymiarów, tym rozbieżności te będą większe (tzn. ze względu na większą ogólność model będzie gorzej dopasowany). Jasne jest również to, że mniejsza liczba wymiarów oznacza prostszą i bardziej intuicyjną analizę struktury odmienności pomiędzy obiektami, poza tym celem skalowania wielowymiarowego jest przecież redukcja obserwowanej złożoności, to znaczy wyjaśnienie wejściowej macierzy odległości przy pomocy mniejszej liczby wymiarów.

Należy zatem wyważyć zysk z minimalizacji liczby wymiarów względem kosztów maksymalizacji błędów dopasowania modelu. Generalnie zatem rzecz ujmując, im więcej wymiarów zastosuje się do odtworzenia pierwotnych odległości, tym lepsze będzie dopasowanie macierzy odległości odtworzonych do macierzy odległości obserwowanych (tzn. mniejszy będzie *STRESS*). Dobrą ilustracją takich prawidłowości są poniższe diagramy Sheparda, na których dla różnej liczby wymiarów zaprezentowano symulowane dopasowanie macierzy odległości odtworzonych w wyniku skalowania wielowymiarowego do macierzy odległości obserwowanych. Wyraźnie widać, że im mniejsza liczba wymiarów, tym gorsze dopasowanie danych.



Rys. 3. Diagram Sheparda – symulacja dla różnej liczby wymiarów.

W sposób naturalny dążyć będziemy do tego (lub algorytm analizy *PROFIT* dążyć będzie za nas), by odwzorowanie odległości obserwowanych (mapę *MDS*) udało się poprawnie wykonać na płaszczyźnie lub w przestrzeni trójwymiarowej. Przy każdej innej liczbie wymiarów większej od trzech pozbawilibyśmy się możliwości intuicyjnej interpretacji wyników skalowania wielowymiarowego. Bogarti (1997a) proponuje przy tym, by uznać, że 2- lub 3-wymiarowa przestrzeń euklidesowa jest odpowiednia dla prezentacji danych, o ile wartość *STRESSu* dla danej liczby wymiarów wyniosła poniżej wartości 0,15, oznaczającej akceptowalny poziom odtworzenia odległości zaobserwowanych.

Użytkownik pakietów statystycznych powinien zawsze (przy pomocy testu osypiska *STRESSu* lub diagramu Sheparda) skontrolować wykonane przez siebie analizy *MDS*, bowiem tylko wtedy będzie miał pewność, że utworzona mapa podobieństw nie jest zbyt daleko posuniętym uogólnieniem danych wejściowych. Z drugiej jednak strony „odbiorców” map *MDS* nie interesuje sposób tworzenia mapy oraz teoretyczno-metodologiczne zawiłości procedur statystycznych. Zainteresowanie wzbudza raczej rezultat skalowania wielowymiarowego, a więc wzajemne odniesienie obiektów do siebie (podobieństwa i różnice) oraz ustalenie, względem jakich cech owe obiekty (klastry obiektów) są do siebie podobne, a względem jakich cech odmienne. Niestety w klasycznym skalowaniu wielowymiarowym interpretacja wymiarów jest dosyć karkołomna. Wynika to z faktu, iż liczba



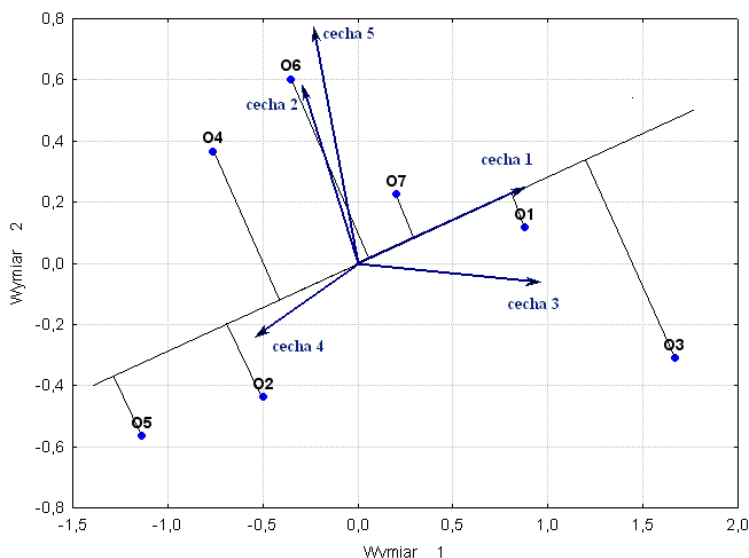
i kierunki osi tworzących przestrzeń liniową map MDS (osie będą zawsze ortogonalne) nie będą najlepiej odtwarzać liczby i kierunku wejściowych wymiarów, stosowanych do generowania podobieństw obiektów. Innymi słowy, zmienne, które były podstawą generowania podobieństw pomiędzy poszczególnymi obiektami, nie sposób jednoznacznie odnieść do układu współrzędnych mapy skalowania wielowymiarowego. Na szczęście pomocny jest tutaj ów drugi etap analizy PROFIT, oparty na równaniach regresyjnych.

Etap II – modele regresyjne

Zakładamy zatem, że w wyniku skalowania wielowymiarowego (w metodzie *PROFIT* stosuje się zazwyczaj niemetryczne skalowanie wielowymiarowe) otrzymano przestrzeń, na której rozlokowane są interesujące nas obiekty. Po utworzeniu takiej mapy każdemu obiektowi przyporządkować można wartości, które odpowiadają współrzędnym na mapie *MDS*. Co oczywiste zatem, jeżeli ma się do czynienia z płaszczyzną, to każdemu obiektowi przyporządkowane będą dwie wartości, jeżeli natomiast z rozwiązaniem w przestrzeni, to obiektom przyporządkowane będą po trzy wartości. Ponieważ celem analizy *PROFIT* jest interpretacja tego, w jaki sposób wejściowy zbiór cech obiektów odwzorowany jest na osiach skalowania wielowymiarowego, to te współrzędne (koordynaty) przypisane obiektom będzie się traktować jako zmienne wyjaśniające (niezależne), a wartości poszczególnych cech obiektów jako zmienne zależne (wyjaśniane). Potrzeba nam bowiem informacji o tym, w jaki sposób (w jakim kierunku) na płaszczyźnie (w przestrzeni) ułożone są obiekty, ze względu na natężenie każdej z tych wejściowych cech. Sposobem na to jest estymacja parametrów modelu, poprzez odniesie każdej cechy do pozycji obiektów na mapie *MDS*.

Algorytm analizy *PROFIT* wykorzysta zatem informacje o współrzędnych (jako zmiennych niezależnych) oraz wartościach obiektów względem każdej z poszczególnych cech (jako zmiennych zależnych), przeprowadzając analizy regresji wielorakiej. Wykonanych będzie tyle analiz regresyjnych, ile cech (zmiennych) uwzględniono w skalowaniu wielowymiarowym. Dla przykładu: jeżeli marki pewnych produktów oceniane były pod względem pięciu cech, to dla każdej takiej cechy przeprowadzona będzie analiza regresji.

Standaryzowane współczynniki równań regresyjnych odpowiadające każdej z osi skalowania wielowymiarowego wyznaczają punkt na mapie *MDS* określający współrzędne danej cechy (zmiennej). Innymi słowy, to właśnie one pozwalają ustalić, w jaki sposób ułożone są interesujące nas obiekty ze względu na natężenie danej cechy. Współczynniki te mogą być również interpretowane w kategoriach kwadratów cosinusów kąta nachylenia wektorów danych cech względem osi układu współrzędnych, przy czym początek wektora znajdować będzie się zawsze na przecięciach osi układu współrzędnych. Wektor taki zwrócony jest w kierunku rosnących wartości danej cechy, jednakże dopiero odpowiednia interpretacja pozwala na uszeregowanie obiektów względem takiej cechy. Należy mieć na uwadze fakt, że dla opisu wyników bez znaczenia pozostaje odległość danego obiektu od prostej, na której położony jest wektor, interpretuje się z kolei uszeregowanie rzutów obiektów na takie proste.



Rys. 4. Interpretacja wyników analizy *PROFIT*.

Z zaprezentowanych danych można odczytać, że względem cechy 1 uszeregowanie obiektów wygląda tak, że największym natężeniem wartości zmiennej (cechy) charakteryzuje się obiekt O3, natomiast najmniejszym obiekt O5. Dla przykładu w badaniach marketingowych można by dokonać skalowania wielowymiarowego pewnych marek produktów względem kilku cech, z których jedną stanowiłaby ocena prestiżu danej marki. Rzutując dane marki produktów na wektor prestiżu, otrzymano by informację o tym, która marka cieszy się największym, a która najmniejszym prestiżem wśród respondentów.

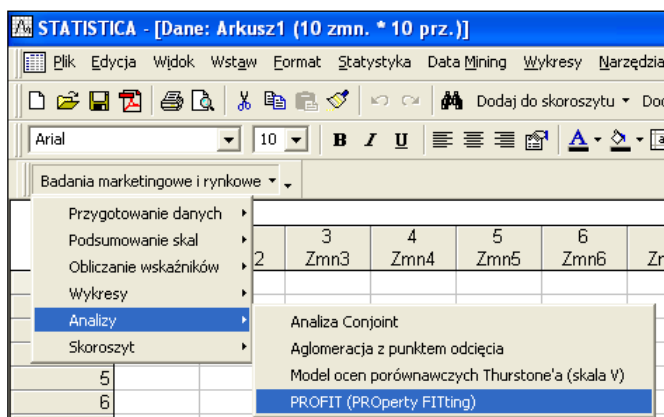
Interpretując wyniki analizy *PROFIT*, należy także rozważyć współczynniki determinacji równań regresyjnych. Pokazują one, w jakim stopniu uszeregowanie obiektów względem wartości danej cechy wyjaśniane jest poprzez położenie tych obiektów na płaszczyźnie.

Przykład analizy *PROFIT* z wykorzystaniem modułu dla analiz marketingowych i rynkowych *STATISTICA 9*

W badaniach marketingowych oraz badaniach rynku procedura skalowania wielowymiarowego, w tym także analiza *PROFIT*, zyskała swoje szczególne uznanie w zakresie budowy map percepcyjnych. Wykorzystuje się je np. wtedy, gdy głównym celem analiz jest ustalenie, w jakim zakresie (względem jakich wymiarów) porównywane między sobą towary, usługi, czy też produkty, uznawane są przez respondentów jako podobne. Podstawą skalowania wielowymiarowego jest w takich przypadkach macierz relacji podobieństwa pomiędzy analizowanymi markami produktów (obiektami). Następnie wyliczane są uśrednione oceny preferencji dla poszczególnych marek produktów, które wykorzystuje się w równaniach regresyjnych. Standaryzowane współczynniki równań regresyjnych wyznaczają

zwrot i kierunek wektora danej cechy produktu i tym samym umożliwiając ustalenie preferencji produktów względem tej cechy.

Przejdźmy teraz do praktycznej implementacji procedury *PROFIT* w programie *STATISTICA*. Analiza *PROFIT* dostępna jest w specjalnym module dedykowanym badaniom marketingowym oraz rynkowym.



Rys. 5. Analiza *PROFIT* w programie *STATISTICA*.

Dane wykorzystywane w przykładzie pochodzą z badań opartych na wywiadach kwestionariuszowych prowadzonych wśród odbiorców usług komunalnych w Poznaniu. Celem badań było ustalenie tego, w jaki sposób mieszkańcy miasta oceniają firmy świadczące usługi w zakresie: (1) dostarczania ciepła do mieszkań, (2) dostarczania gazu, (3) dostarczania wody, (4) dostarczania energii elektrycznej, (5) oczyszczania miasta oraz (6) transportu publicznego. Poszczególne firmy oceniane były pod względem: (a) nowoczesności, (b) konkurencyjności, (c) jakości świadczonych usług, (d) dynamiki rozwoju, (e) dbania o klienta oraz (f) wiarygodności. Wykorzystano w tym celu tzw. dyferencjał semantyczny, tzn. proszono respondentów, aby ocenić każdą firmę pod względem tego, czy uważają ją za: (a) nowoczesną czy też przestarzałą, (b) konkurencyjną czy też niekonkurencyjną, itd. dla każdej z takich cech. W ramach każdego wymiaru oceny odpowiedziom respondentów przypisano wagi liczbowe w ten sposób, że odpowiedzi skrajnie pozytywnej przypisano wartość +3, odpowiedzi neutralnej wartość 0, natomiast odpowiedzi skrajnie negatywnej wartość -3. Cała gama wartości obejmowała następujący zbiór wag: -3, -2, -1, 0, +1, +2 +3. Na podstawie jednostkowych odpowiedzi respondentów dokonano uśrednienia wyników dyferencjału semantycznego, uzyskując oceny poszczególnych firm z punktu widzenia wyróżnionych wymiarów oceny.

Pierwszym ułatwieniem, jakie daje moduł analizy *PROFIT* względem modułu analizy skalowania wielowymiarowego, jest sposób przygotowania pliku wejściowego do analizy. Prowadząc skalowanie wielowymiarowe należy przekształcić plik wejściowy do postaci macierzowej, z wyliczonymi odległościami pomiędzy rozpatrywanymi obiektami. Przeprowadzenie skalowania wielowymiarowego możliwe jest więc dopiero po odpowiednim zapisaniu danych w pliku macierzowym.

	Baza danych - przykład 1					
	1	2	3	4	5	6
	Firma 1	Firma 2	Firma 3	Firma 4	Firma 5	Firma 6
Firma 1	0,00000	1,13371	1,69979	1,34696	2,39962	1,20632
Firma 2	1,13371	0,00000	1,40968	1,38470	2,28256	1,42327
Firma 3	1,69979	1,40968	0,00000	1,96209	1,18520	1,43878
Firma 4	1,34696	1,38470	1,96209	0,00000	2,13670	1,54606
Firma 5	2,39962	2,28256	1,18520	2,13670	0,00000	1,92338
Firma 6	1,20632	1,42327	1,43878	1,54606	1,92338	0,00000
Średnie	0,45833	0,04333	-0,00333	0,17333	-0,04167	0,25833
Odch. std	0,45849	0,52110	0,32303	0,59691	0,52674	0,40504
No. Cases	6,00000					
Matrix	3,00000					

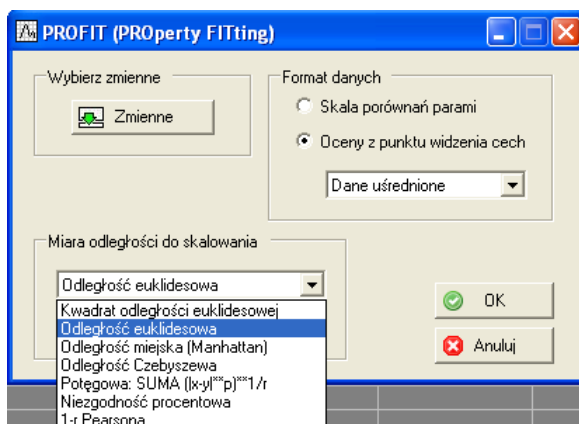
Rys. 6. Dane wejściowe skalowania wielowymiarowego – plik macierzowy.

W analizie *PROFIT* taka transformacja pliku danych do postaci macierzowej nie jest konieczna, wystarczy wyliczyć statystyki punktowe (np. średnią arytmetyczną, medianę, czy też wskaźnik struktury) charakteryzujące obiekty względem interesujących nas cech. Poniższy rysunek zawiera zestawienie danych wejściowych wykorzystanych w opisywanym przykładzie.

	Baza danych - analiza PROFIT - przykład 1					
	1	2	3	4	5	6
	nowoczesność	konkurencyjność	jakość usług	dynamiczność rozwoju	dbanie o klienta	wiarygodność
Firma 1	0,29	0,22	0,37	-0,14	1,04	0,97
Firma 2	-0,46	-0,13	-0,06	-0,42	0,47	0,86
Firma 3	0,38	-0,08	-0,55	0,12	-0,1	0,21
Firma 4	-0,32	0,54	0,64	-0,66	0,85	-0,01
Firma 5	0,51	0,42	-0,37	0,29	-0,29	-0,81
Firma 6	0,21	-0,46	0,53	0,36	0,72	0,19

Rys. 7. Dane wejściowe analizy *PROFIT*.

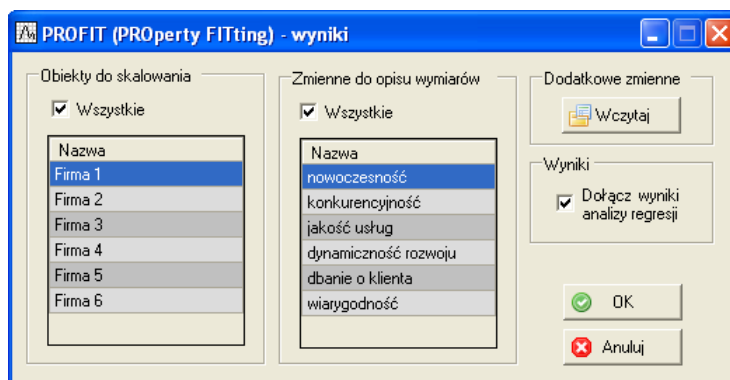
Na tak przygotowanych danych można rozpocząć procedurę analizy *PROFIT*. Po jej uruchomieniu okno programu wygląda następująco.



Rys. 8. Okno analizy *PROFIT* w programie *STATISTICA*.

Konieczne jest ustalenie wejściowych warunków dla wykonania analizy *PROFIT*. W pierwszej kolejności określić należy format danych dla analizy. Ponieważ dane w przykładzie zapisane zostały jako uśrednione oceny poszczególnych firm w ramach każdego z sześciu wymiarów, to format danych ustalamy jako „uśrednione oceny obiektów z punktu widzenia cech”. Do wyboru mamy także siedem miar odległości, w tym: kwadrat odległości euklidesowej, odległość euklidesową, odległość miejską (tzw. Manhattan), odległość Czebyszewa, odległość potęgową, niezgodność procentową, a także odległość opartą na korelacjach liniowych Pearsona. Zauważmy, że w programie *STATISTICA* analiza *PROFIT* oferuje ten sam zestaw miar odległości, który dostępny jest również w procedurze skalowania wielowymiarowego (*PROFIT* jest przecież rozszerzeniem *MDS*), a także w module analizy skupień programu *STATISTICA*. Celem wszystkich tych procedur jest analiza podobieństw, stąd konsekwentnie dostępny jest ten sam zestaw miar odległości.

Po ustaleniu tych podstawowych warunków procedury *PROFIT* możemy przejść do wyboru zmiennych (cech), które będą podstawą skalowania obiektów. W naszym przykładzie do utworzenia mapy percepcji firm świadczących usługi komunalne wybierzemy wszystkie sześć cech, względem których firmy te oceniane były przez mieszkańców. Po zaakceptowaniu listy zmiennych program przechodzi do okna *PROFIT* – wyniki.

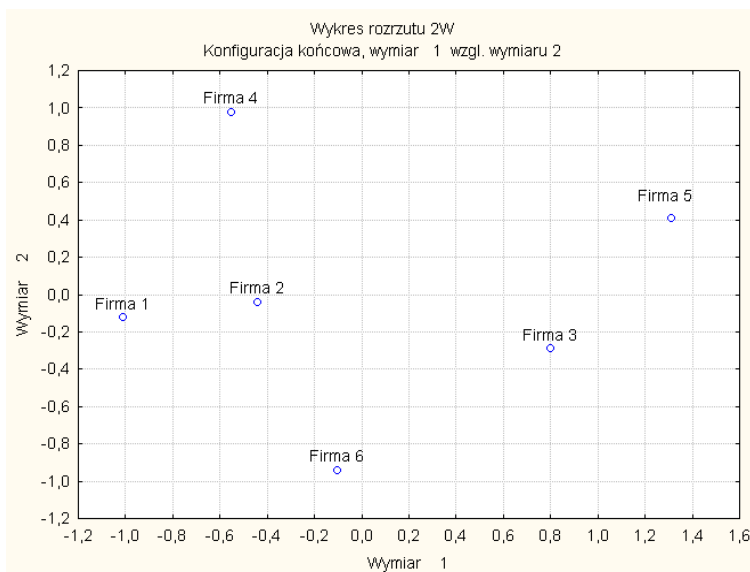


Rys. 9. Okno wyników analizy *PROFIT*.

Na podstawie danych wejściowych algorytm analizy *PROFIT* dokonał już skalowania wielowymiarowego obiektów na płaszczyźnie, chociaż na tym etapie program nie przedstawia jeszcze mapy podobieństw. Co ważne, w klasycznym skalowaniu wielowymiarowym możliwe jest utworzenie mapy o dowolnej liczbie wymiarów, natomiast algorytm procedury *PROFIT* w programie *STATISTICA* został zaprogramowany tak, by zawsze dać rozwiązanie dwuwymiarowe.

Załóżmy zatem, że po przeprowadzeniu klasycznego skalowania wielowymiarowego (analiza ta dostępna jest w module *wielowymiarowe techniki eksploracyjne* programu *STATISTICA*) także otrzymalibyśmy wynik w postaci mapy 2D. Takie rozwiązanie pozwoliłoby wprawdzie ustalić, które firmy są bardziej, a które mniej do siebie podobne, nie sposób jednak byłoby w prosty sposób stwierdzić, jak grupują się te obiekty względem interesujących nas cech. Innymi słowy nie byłibyśmy w stanie intuicyjnie ustalić, wzglę-

dem jakich cech rozpatrywane obiekty są do siebie podobne, a względem jakich cech odmienne.



Rys. 10. Wynik skalowania wielowymiarowego.

Powróćmy zatem do analizy *PROFIT*, która na tym etapie dała „na razie” te same rezultaty co skalowanie wielowymiarowe. W oknie wyników analizy *PROFIT* (rys. 9) dokonujemy teraz ponownie wyboru obiektów oraz ustalamy, które zmienne wykorzystane będą do opisu wymiarów skalowania wielowymiarowego. W naszym przypadku porównywać będziemy wszystkie firmy (obiekty), a do opisu wymiarów wykorzystamy wszystkie zmienne wejściowe.

Zauważmy jednak, że lista zmiennych do opisu wymiarów może być inna niż lista zmiennych na podstawie której dokonano skalowania wielowymiarowego obiektów. Bardzo często jednak do opisu wymiarów wykorzystujemy zarówno dane wejściowe, jak i dodatkowe informacje. Dla przykładu moglibyśmy dla analizowanych przez nas firm zebrać dodatkowe dane dotyczące wydatków na kampanie promocyjne i zdiagnozować, na ile ocena firm jest powiązana z działaniami *Public Relations*.

Jeżeli w oknie wyników analizy *PROFIT* zaznaczymy również opcje dołączenia wyników regresji, to program zamieści w skrócie wyniki tyłu modeli regresyjnych, ile zmiennych wybrano do opisu wyników. Każda z tych zmiennych jest objaśniana poprzez położenie (współrzędne) obiektów na płaszczyźnie (rys. 11).



Baza danych - analiza PROFIT - przykład 1								
	1 nowoczesność	2 konkurencyjność	3 jakość usług	4 dynamiczność rozwoju	5 dbanie o klienta	6 wiarygodność	7 WYM. 1	8 WYM. 2
Firma 1	0,29	0,22	0,37	-0,14	1,04	0,97	-1,01050	-0,121828
Firma 2	-0,46	-0,13	-0,06	-0,42	0,47	0,86	-0,44366	-0,037936
Firma 3	0,38	-0,08	-0,55	0,12	-0,1	0,21	0,79771	-0,287516
Firma 4	-0,32	0,54	0,64	-0,66	0,85	-0,01	-0,55312	0,979472
Firma 5	0,51	0,42	-0,37	0,29	-0,29	-0,81	1,31197	0,407734
Firma 6	0,21	-0,46	0,53	0,36	0,72	0,19	-0,10240	-0,939925

Rys. 11. Akusz do wyliczania regresji.

Na podstawie takich danych przeprowadzane są analizy regresyjne. Przykładowy wynik analizy regresji dla wymiaru oceny *konkurencyjności* rozpatrywanych firm przedstawiony został na poniższym rysunku.

Podsumowanie regresji zmiennej zależnej: konkurencyjność (Arkusz183)						
R= ,92897699 R^2= ,86299825 Skoryg. R^2= ,77166374						
F(2,3)=9,4488 p<,05071 Błąd std. estymacji: ,17981						
N=6	BETA	Bł. std. BETA	B	Bł. std. B	t(3)	poziom p
W. wolny			0,085000	0,073405	1,157953	0,330698
WYM. 1	0,042770	0,213699	0,018241	0,091142	0,200142	0,854169
WYM. 2	0,927992	0,213699	0,537785	0,123842	4,342519	0,022541

Rys. 12. Wynik analiz regresyjnych.

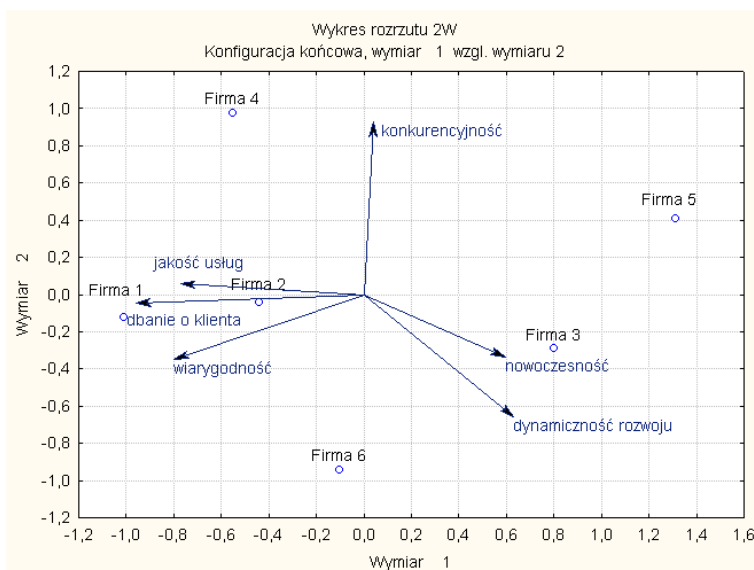
Po przeprowadzeniu analiz regresyjnych dla wszystkich zmiennych algorytm analizy *PROFIT* naniesie współczynniki kierunkowe określające zwrot i kierunek wektora, odpowiadającego każdej z cech wybranych do opisu wymiarów.

	1 nowoczesność	2 konkurencyjność	3 jakość usług	4 dynamiczność rozwoju	5 dbanie o klienta	6 wiarygodność
Wymiar 1	0,594409927	0,0427701069	-0,771437882	0,63148127	-0,958990711	-0,797056435
Wymiar 2	-0,334341627	0,927991899	0,0581444096	-0,657889285	-0,0450272172	-0,348072289

Rys. 13. Współrzędne osi.

Ostatecznym wynikiem analizy *PROFIT* jest mapa percepcji z wektorami opisującymi poszczególne wymiary płaszczyzny.

Wynik analizy *PROFIT* z naniesionymi współrzędnymi poszczególnych cech pozwala teraz na bardzo intuicyjną interpretację podobieństw i odmienności pomiędzy analizowanymi firmami. Dla przykładu z wykresu tego można odczytać, że Firma 4 jest oceniana jako najbardziej konkurencyjna i pod tym względem podobna jest do Firmy 5. To, co odróżnia obie firmy, to np. ocena ich nowoczesności i dynamiczności rozwoju; Firma 5 oceniana jest w obu wymiarach najwyżej, natomiast Firma 4 – najniżej. Podobnie można zauważyć, że Firma 1, ma najwyższe oceny w zakresie wiarygodności działań, dbania o klienta oraz jakości świadczenia usług, z kolei najgorsze oceny w tym względzie otrzymała Firma 3 oraz Firma 4. Analogiczne wnioskowanie można przeprowadzić dla dowolnego obiektu oraz dowolnego wymiaru osi skalowania wielowymiarowego.



Rys. 14. Mapa percepcji firm – wynik analizy *PROFIT*.

Podsumowanie i wnioski

Głównym celem artykułu była z jednej strony prezentacja podstaw teoretycznych i metodologicznych analizy *PROFIT*, a z drugiej praktyczna implementacja tego nowego modułu pakietu *STATISTICA* dedykowanego analizom marketingowym i rynkowym. Podsumowując rozważania w artykule, można wyraźnie zaznaczyć, że:

- ◆ Nowy moduł pakietu *STATISTICA dla analiz marketingowych i rynkowych* zawiera wiele przydatnych rozwiązań wykorzystywanych w ramach analizy preferencji konsumentów, percepcji marek produktów oraz percepcji rynku.
- ◆ Jednym z takich dodatkowych modułów jest procedura *PROFIT*, łącząca skalowanie wielowymiarowe i analizę regresji. W porównaniu do skalowania wielowymiarowego analiza *PROFIT* ułatwia wyjaśnienie osi skalowania wielowymiarowego, poprzez naniesienie cech obiektów na przestrzeń mapy MDS.
- ◆ Ważną zaletą omawianego modułu jest możliwość wyjaśnienia osi skalowania wielowymiarowego nie tylko w oparciu o zbiór cech, które były podstawą skalowania obiektów, ale także przy wykorzystaniu dodatkowych zmiennych opisujących analizowane marki, produkty, firmy, itd. Np. w ten sposób zyskujemy możliwość analizy wpływu kampanii marketingowych na percepcję marek produktów przez poszczególne grupy klientów.
- ◆ Procedura *PROFIT* jest świetnym narzędziem do budowy map percepcyjnych, program daje przejrzyste rozwiązania graficzne, których interpretacja jest całkowicie intuicyjna.



Literatura

1. Bogarti, S.P., *Multidimensional Scaling*, 1997a, <http://www.analytictech.com/borgatti/mds.htm>.
2. Bogarti, S.P., *Profit*, 1997b, <http://www.analytictech.com/borgatti/profit.htm>.
3. Holland, S. M., *Non-metric Multidimensional Scaling*, 2008, www.uga.edu/strata/software/pdf/mdsTutorial.pdf.
4. Internetowy podręcznik *STATISTICA*: <http://www.statsoft.pl/textbook/stathome.html>, hasło: *Skalowanie wielowymiarowe*.
5. Sagan, A., *Analiza preferencji konsumentów z wykorzystaniem programu STATISTICA – analiza Conjoint i skalowanie wielowymiarowe*, w: *Zastosowania nowoczesnej analizy danych marketingowych w badaniach rynku*, StatSoft Polska Sp. z o.o., 2009, ss. 3-22.
6. Sagan, A., *Jeden obraz ukazuje więcej niż 10 liczb, czyli jak budować mapy zadowolenia klienta z wykorzystaniem programu STATISTICA*, StatSoft Polska Sp. z o.o., 2004, ss. 35-50.
7. Takane, Y., Young, F.W., *Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method With Optimal Acing Features*, *Psychometrika*, Vol. 42(1), 1977, ss. 7-67.