

ZASTOSOWANIE METOD WIELOWYMIAROWYCH W CHARAKTERYSTYCE PREFERENCJI KONSUMENTÓW

Lucyna Błażejczyk-Majka

Zakład Historii Gospodarczej, Uniwersytet w Poznaniu
e-mail: majkal@amu.edu.pl

Paweł Boczar

Katedra Rynku i Marketingu, Uniwersytet Przyrodniczy w Poznaniu
e-mail: pboczar@up.poznan.pl

Streszczenie: W pracy przeprowadzono analizę preferencji polskich konsumentów oleju rzepakowego z wykorzystaniem metod wielowymiarowych. Zwrócono uwagę na zbieżności i różnice zastosowania skalowania wielowymiarowego, analizy skupień i analizy PROFIT. Wykazano, że decyzje zakupowe konsumentów olejów rzepakowych w zależności od fazy życia respondentów różnią się pod względem postrzegania smaku tego oleju, jego ceny, naturalności i dotychczasowego doświadczenia.

Słowa kluczowe: skalowanie wielowymiarowe, analiza PROFIT, konsument, olej rzepakowy, analiza skupień, biplot

WSTĘP

We współczesnym świecie powszechne są badania ankietowe. Stanowią one punkt wyjścia do oceny sytuacji przedsiębiorstwa czy pojedynczego produktu na rynku czy w konsekwencji do podjęcia działań mających na celu wybór dalszej ścieżki rozwoju firmy. Dzięki tego typu badaniom uzyskiwany jest szeroki zestaw informacji, który można analizować pod względem pojedynczych cech [Boczar i Błażejczyk-Majka 2015]. Jednak ze względu na ilość dostępnych danych coraz większym zainteresowaniem cieszy się zastosowanie w tym obszarze analiz wielowymiarowych [Walesiak 1996].

Badanie preferencji konsumentkich z definicji generuje złożoną informację zwrotną [Sagan 2009, Zaborski 2014]. Pytania dotyczące tego obszaru sprowadzają się do tego, że respondenci wybierają spośród wielu czynników te, które dla nich

mają kluczowe znaczenie przy wyborze danego produktu czy firmy. W efekcie tego typu badań uzyskuje się dla każdego respondenta listę cech mających decydujący wpływ na ich decyzje rynkowe, uszeregowaną według, przypisanej im przez respondenta, ważności. Dla przedsiębiorcy istotna jest informacja jak konsumenci dzielą się ze względu na postrzeganie ich produktu i jakie cechy produktu decydują w tym podziale.

Wśród wielu dostępnych metod wielowymiarowych pozwalających na segmentację konsumentów można oczywiście wymienić analizę skupień [patrz np.: Mardia i in. 1979]. Pozwala ona na grupowanie obiektów, w tym przypadku respondentów, którym przypisano, ze względu na udzielane przez nich odpowiedzi, wiele cech. Tego typu problemy można rozwiązać także dzięki wykorzystaniu skalowania wielowymiarowego [patrz np. Zaborski 2001]. W jego wyniku uzyskuje się mapę percepcji, na której rozmieszczone są badane grupy respondentów. Im mniejsza odległość między nimi, tym większe ich podobieństwo ze względu na wszystkie badane cechy. Pozostaje jednak problem wyboru i prezentacji czynników decydujących o danym podziale obiektów na mapie percepcji. Tę możliwość daje analiza PROFIT, dostępna w zestawie PLUS programu *Statistica* [Migut 2012]. Dzięki zastosowaniu tego podejścia, na mapę percepcji uzyskaną w wyniku skalowania wielowymiarowego nakładane są wektory obrazujące znaczenie cech w tym podziale.

Celem pracy jest prezentacja wykorzystania wybranych metod wielowymiarowych w opisie preferencji konsumentów oleju rzepakowego w Polsce. W badaniach zostanie zaprezentowana implementacja skalowania wielowymiarowego, analizy skupień i analizy PROFIT. Do prezentacji możliwości tych metod wybrano przykład związany z oceną czynników wpływających na decyzje zakupowe konsumentów oleju rzepakowego. Wszystkie obliczenia przeprowadzono z wykorzystaniem oprogramowania *Statistica*, a w szczególności zestawu PLUS.

MATERIAŁ BADAWCZY

Dane będące podstawą po prezentacji metod wielowymiarowych powstały w wyniku badań ankietowych¹ pt: „Zwyczaje związane z konsumpcją tłuszczów roślinnych”. Zostały one zrealizowane w dniach od 26 do 30 kwietnia 2012r., a pytania w nich zawarte dotyczyły okresu od maja 2011r. do kwietnia 2012r. Przedmiotem badania było pozyskanie informacji na temat deklaracji spożycia i kupowania poszczególnych rodzajów tłuszczów roślinnych, częstotliwości ich spożycia i kupowania, kryteriów branych pod uwagę przy ich zakupie oraz

¹ Badania ankietowe zostały przeprowadzone w ramach projektu badawczego nr N N112 395940 pt.: „Wpływ zmian kierunków wykorzystania rzepaku na uczestników polskiego rynku roślin oleistych i produktów ich przetwarzania.”

przekonań dotyczących tłuszczów roślinnych. W badaniu wzięła udział ogólnopolska grupa Polaków w wieku 15+.

Tabela 1. Średnie wartości oraz odchylenia standardowe ocen przypisywanych poszczególnym czynnikom zakupu w zależności od fazy życia respondenta

Faza życia Czynnik zakupu	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Razem
Cena	34,42 (35,92)	27,92 (30,73)	27,43 (28,74)	26,94 (27,47)	31,39 (31,29)	23,92 (29,57)	23,25 (26,15)	36,12 (30,09)	29,92 (29,55)
Walory zdrowotne	5,86 (4,30)	7,96 (9,32)	5,32 (4,90)	7,93 (8,25)	7,91 (8,92)	8,78 (7,55)	9,50 (9,51)	7,21 (7,84)	7,89 (8,19)
Wysoka jakość	15,58 (11,42)	15,32 (10,14)	12,08 (10,95)	13,93 (11,75)	12,75 (12,24)	13,91 (11,68)	14,39 (11,15)	9,45 (8,19)	12,29 (10,66)
Promocja w sklepie	2,22 (3,69)	1,49 (2,30)	4,52 (7,08)	1,83 (3,28)	2,44 (3,71)	1,58 (2,32)	1,99 (2,64)	2,71 (3,67)	2,32 (3,52)
Reklama produktu w mediach	0,24 (0,30)	0,36 (1,11)	1,01 (1,61)	0,23 (0,57)	0,70 (1,76)	0,88 (1,98)	0,41 (1,02)	0,53 (1,35)	0,56 (1,42)
Znana marka	0,70 (0,85)	0,82 (1,35)	1,65 (2,16)	0,83 (1,13)	1,46 (2,89)	1,75 (2,67)	1,07 (1,75)	1,15 (2,54)	1,23 (2,33)
Kraj pochodzenia produktu	0,28 (0,35)	1,07 (1,68)	0,99 (1,45)	0,59 (0,96)	0,88 (1,73)	1,67 (3,78)	0,86 (1,28)	0,78 (1,54)	0,92 (1,97)
Wygląd opakowania	0,58 (0,93)	0,52 (1,39)	1,04 (1,83)	0,35 (0,83)	0,62 (1,18)	0,77 (1,33)	0,52 (1,01)	0,75 (1,81)	0,65 (1,43)
Wielkość opakowania	0,38 (0,48)	0,32 (0,40)	1,71 (2,60)	0,75 (1,55)	0,98 (1,65)	0,77 (1,14)	0,66 (1,04)	0,84 (1,26)	0,83 (1,38)
Praktyczne opakowanie	0,40 (0,45)	0,49 (0,96)	1,15 (1,68)	0,57 (1,15)	0,79 (1,27)	0,79 (1,16)	0,57 (0,93)	0,73 (1,21)	0,71 (1,17)
Uniwersalność produktu	0,86 (0,86)	0,93 (1,01)	1,77 (1,97)	1,14 (2,16)	1,08 (1,53)	1,03 (1,16)	0,79 (0,90)	1,21 (2,38)	1,10 (1,84)
Naturalność produktu	5,64 (3,60)	8,65 (5,84)	8,72 (9,27)	8,99 (9,10)	8,68 (8,55)	8,21 (8,02)	10,01 (9,38)	6,57 (8,00)	8,12 (8,45)
Smak	23,24 (20,72)	9,63 (5,46)	14,33 (13,73)	19,20 (15,19)	13,36 (14,63)	18,73 (17,50)	17,12 (15,21)	16,03 (15,42)	16,37 (15,38)
Przyjemny zapach	1,28 (0,77)	1,33 (0,61)	2,18 (1,61)	1,82 (1,50)	1,76 (1,64)	2,49 (2,34)	2,00 (1,99)	2,43 (2,66)	2,15 (2,17)
Barwa produktu	0,76 (1,21)	1,07 (2,09)	0,87 (0,94)	0,54 (0,62)	1,00 (2,18)	1,10 (1,76)	0,72 (1,25)	0,78 (1,11)	0,84 (1,43)
Dodatek witamin	1,98 (1,36)	2,86 (4,27)	2,97 (3,96)	2,04 (1,81)	2,48 (3,03)	3,27 (4,08)	3,09 (3,90)	2,69 (3,60)	2,73 (3,48)
Dotychczasowe doświadczenie	1,06 (0,69)	9,27 (12,43)	7,54 (9,96)	7,18 (10,82)	4,80 (5,80)	4,43 (6,37)	6,31 (8,01)	5,75 (8,79)	5,77 (8,38)
Rodzaj surowca	4,54 (4,42)	9,95 (13,14)	4,67 (4,77)	5,14 (4,80)	6,88 (9,29)	5,89 (6,93)	6,74 (8,66)	4,27 (5,98)	5,59 (7,33)

() odchylenie standardowe; fazy życia zostały zdefiniowane w rozdziale Materiał badawczy.

Źródło: obliczenia własne na podstawie GfK Polonia

Badanie to polegało na przeprowadzeniu bezpośrednich wywiadów w domach respondentów, w ramach wielotematycznego syndykatowego badania cyklicznego GfK CAPIBUS, Badanie OMNIBUS zrealizowane techniką CAPI - Computer Assisted Personal Interview. Wywiady te odbywały się z osobą prowadzącą gospodarstwo domowe. Liczebność próby wynosiła 910 respondentów, w tym 205 mężczyzn i 705 kobiet. Z bazy wyodrębniono konsumentów oleju rzepakowego, którzy zadeklarowali, że spożywają ten olej raz w tygodniu lub częściej. Osób spełniających to kryterium było 362.

Punktem wyjścia do przeprowadzonych badań były wyniki MaxDiff, w której analizowano odpowiedzi dotyczące oceny czynników zakupu oleju roślinnego. Dzięki tej metodzie oceny poszczególnych czynników zakupu każdego z respondentów były wyrażone w skali ilorazowej, przy czym ocena czynnika na poziomie 0 oznaczała brak jego wpływu na zakup oleju. Wartości ocen dla wszystkich kryteriów sumowały się do 100 [patrz np.: Bartłomowicz 2014].

Ocenianymi kryteriami były: cena, walory zdrowotne, wysoka jakość, promocja w sklepie, reklama produktu w mediach, znana marka, kraj pochodzenia produktu, wygląd opakowania, wielkość opakowania, praktyczne opakowanie/wygoda użycia, wielość zastosowań/universalność produktu, naturalność produktu, smak, przyjemny zapach, barwa produktu, dodatek witamin, dotychczasowe doświadczenie, rodzaj surowca, z którego wytworzono produkt.

W poniższym artykule wykorzystano tylko te wyniki, które dotyczyły oleju rzepakowego. Co więcej respondentów odpowiadających na wspomniane powyżej pytania o ich decyzje zakupowe podzielono ze względu na fazę życia, w której się znajdowali w momencie przeprowadzania badania. Wśród wymienionych grup znaleźli się: (1) studenci, wolni, mieszkający z rodzicami, (2) pracujący, wolni, mieszkający z rodzicami, (3) młodszy, bez dzieci, samodzielne gospodarstwo, (4) rodzina z małymi dziećmi do 6 lat, (5) rodzina z dziećmi 7-14 lata, (6) rodzina z dziećmi 15-25 lat, (7) starsza rodzina, pracująca, bez dzieci, (8) starsza rodzina, niepracująca, bez dzieci.

Średnie wartości ocen, przypisywanych każdej z wymienionych grup respondentów, w zależności od udzielanych przez nich odpowiedzi dotyczących czynników determinujących zakupy oleju rzepakowego, przedstawiono w tabeli 1. Im wyższa wartość liczbową oceny przypisanej danemu czynnikowi, tym respondent przypisywał mu większe znaczenie w swoich decyzjach zakupowych.

WYBRANE METODY WIELOWYMIAROWE

Skalowanie wielowymiarowe

W celu pogłębienia charakterystyki konsumentów olejów roślinnych wykorzystano skalowanie wielowymiarowe. Metoda ta znajduje szerokie zastosowanie w wielu obszarach, zwłaszcza w badaniach marketingowych [Zaborski 2001, 2014]. Stosuje się je także w psychologii, socjologii, ekonomii,

medycynie i rolnictwie [Green i in. 1989, Walesiak 1996]. Wśród protoplastów tej metody wymienia się przede wszystkim Torgersona [1952] oraz Sheparda [1962].

Idea skalowania wielowymiarowego sprowadza się do graficznej prezentacji, zwykle w przestrzeni dwu- lub trójwymiarowej, zbioru badanych obiektów wielocechowych. Te ostatnie są traktowane jako punkty w przestrzeni. Mogą to być grupy konsumentów, produkty, marki, przedsiębiorstwa, rynki testowe itp. Obiekty powinny być rozmieszczone w taki sposób, aby w jak największym stopniu odwzorowane zostało ich podobieństwo w przestrzeni wielowymiarowej [Gordon 1981, Malina i Wanat 2000].

Punktem wyjścia w procedurze skalowania wielowymiarowego jest wyznaczenie macierzy odległości obserwowanych pomiędzy badanymi n obiektami [Walesiak 1996]. Niech d_{ik} oznacza odległość pomiędzy i -tym i k -tym obiektem. Spośród wielu znanych procedur wyznaczania odległości metrycznych w pracy skorzystano z klasycznej odległości euklidesowej [patrz np. Mardia i in. 1979, 381].

$$d_{ik} = \sqrt{\sum_{r=1}^m (x_{ir} - x_{kr})^2}, \quad r = 1, 2, \dots, m; \quad (1)$$

gdzie x_{ir} oznacza wartość r -tej cechy i -tego obiektu. Należy pamiętać, że jeżeli w analizie brane są pod uwagę zmienne wyrażone w różnych mianach i o zróżnicowanym poziomie wartości, to procedura skalowania wielowymiarowego powinna być poprzedzona standaryzacją cech [Walesiak 2004].

W kolejnym kroku procedury wyznaczana jest r -wymiarowa macierz produktów skalarnych. Jak wspomniano wcześniej, jest to zwykle macierz dwu- lub trójwymiarowa. Jeśli zmienne mierzone są na skali przedziałowej, to każdy element tej macierzy – odległość odtworzona - będzie liniowo zależny od odległości obserwowanej:

$$\hat{d}_{ik} = a_0 + a_1 d_{ik}, \quad (2)$$

przy czym $a_0 = 0$ dla zmiennych mierzonych w skali ilorazowej.

Wartości współczynników funkcji liniowej przedstawionej w równaniu (2) wyznaczane są w taki sposób, aby minimalizowały wartość funkcji dopasowania STRESS (Standardized Residual Sum of Squares). Szczegółowy opis tej estymacji zawierają monografie: Mardia i in. [1979, 395-423], Zaborskiego [2001, 56-62] czy Walesiaka i Gatnara [2009, 354-369].

Współczynnik *STRESS* [Cox i Cox 1991] służy także do oceny jakości przeprowadzonej procedury skalowania wielowymiarowego. Jego wyznaczenie sprowadza się do obliczenia sumy kwadratów odchyłeń obserwowanych odległości (lub pewnej transformacji monotonicznej tych odległości) od odległości odtworzonych:

$$STRESS = \sqrt{\frac{\sum_{i=1}^n (d_{ik} - \hat{d}_{ik})^2}{\sum_{i=1}^n d_{ik}^2}}. \quad (3)$$

Zatem im mniejsza wartość współczynnika *STRESS*, tym lepsze dopasowanie macierzy odległości odtworzonych do macierzy odległości obserwowanych. Zaborski [2009, 355] wskazuje, że o dopasowaniu bardzo dobrym można mówić, jeżeli współczynnik ten przyjmuje wartości od 0 do 0,02.

W wyniku zastosowania tej metody uzyskuje się dwu- lub trzywymiarową mapę percepcji skalowania wielowymiarowego, która przedstawia uzyskaną konfigurację analizowanych obiektów względem określonego zbioru zmiennych. Zatem im bliżej siebie znajdują się badane obiekty, tym okazały się one bardziej do siebie podobne. Niemniej jednak, jak wskazuje między innymi Jabkowski [2010], nie sposób jednoznacznie odnieść do układu współrzędnych mapy skalowania wielowymiarowego zmiennych, które były podstawą generowania odległości pomiędzy badanymi obiektami. Co więcej uzyskaną w wyniku skalowania wielowymiarowego strukturę obiektów trudno jest podzielić na grupy w sposób obiektywny i jednoznaczny.

Analiza skupień

Analiza skupień [Cormack 1971, Marek 1989, 23-54 lub Timm 2002, 515–541] pozwala na łączenie wielowymiarowych obiektów w grupy (skupienia), które spełniają warunek wewnętrznej jednorodności i zewnętrznej niejednorodności. Ważną grupę metod tej analizy tworzą hierarchiczne metody aglomeracyjne. W metodach tych układ skupień, zapisany w postaci dendrogramu, wskazuje na kolejność ich tworzenia. Pozwala także określić, jakie mniejsze skupienia wchodziły w skład skupienia większego, nadrzędnego.

Wstępnie zakłada się, że każdy obiekt tworzy osobne skupienie s_1, s_2, \dots, s_n . Punktem wyjścia do przeprowadzenia analizy skupień jest, podobnie jak w skalowaniu wielowymiarowym, macierz odległości [patrz np.: Mardia i in. 1974, 360-393, Härdle i Simar 2012, 331-349]. Spośród wielu możliwości, podobnie jak w przypadku skalowania wielowymiarowego wybrano przedstawioną we wzorze (1), metrykę euklidesową, która tym razem odnosi się do poszczególnych skupień. Na podstawie macierzy odległości łączone są ze sobą dwa najbliższe obiekty. W ten sposób liczba skupień zostaje zredukowana o jeden. Przed kolejnym połączeniem skupień należy, korzystając z wybranego algorytmu, odpowiednio skorygować macierz odległości. W literaturze tematu dostępnych jest wiele procedur [patrz np. Marek 1989]. Na potrzeby artykułu wybrano metodę najdalszego sąsiedztwa:

$$d_{s_i s_j} = \max(d_{s_i s_k}; d_{s_j s_k}), \quad (4)$$

gdzie $d_{s_i s_k}$, oznacza odległość pomiędzy skupieniem s_i oraz s_k . Dzięki tej procedurze wymiar macierzy odległości również ulegnie zmniejszeniu.

W wyniku zastosowania analizy skupień powstaje dendrogram, który po odpowiednim przecięciu wskazuje na grupy obiektów, uzyskane w wyniku grupowania [Błażejczyk-Majka i Kala 2005]. Zwykle stosuje się podejście,

w którym uznaje się układ skupień, powstałych po przecięciu dendrogramu na poziomie połowy odległości maksymalnej. Jeżeli w procedurze analizy skupień zastosowano tę samą miarę odległości, co w skalowaniu wielowymiarowym, to rozmieszczenie badanych obiektów powinno prowadzić do zbliżonych wniosków. Niemniej jednak podział dendrogramu prowadzi do obiektywnie jednoznacznych rozstrzygnięć, co do składu poszczególnych skupień.

Analiza PROFIT

Analiza PROFIT (PROperty FITting) łączy procedurę skalowania wielowymiarowego i analizę regresji wielorakiej [Jabkowski 2010]. Dzięki temu połączeniu, można nie tylko ocenić podobieństwo badanych obiektów pod względem ich cech, ale także wyodrębnić te cechy, które miały kluczowy wpływ na uzyskaną formację. Analiza PROFIT pozwala także na graficzną prezentację wyników grupowania obiektów i ich relacji do badanych cech w postaci biplotów.

W pierwszym etapie analizę PROFIT stanowi opisane wyżej skalowanie wielowymiarowe. W oprogramowaniu *Statistica*, oprócz skal metrycznych, można zastosować w module PLUS skale niemetryczne. Dzięki temu możliwe jest uwzględnienie w analizie nie tylko cech ilościowych badanych obiektów, ale i jakościowych [patrz np. Zaborski 2001, Jabkowski 2010].

Jak wspomniano wcześniej w efekcie zastosowania skalowania wielowymiarowego, w przypadku dwuwymiarowej mapy percepcji, każdy obiekt zostaje opisany dwoma współrzędnymi. W drugim etapie analizy PROFIT współrzędne te traktowane są jako zmienne niezależne, natomiast wartości poszczególnych cech obiektów jako zmienne zależne. Analiza regresji liniowej pozwala na oszacowanie w jaki sposób na płaszczyźnie rozłożone są badane obiekty ze względu na natężenie każdej z opisujących go cech. Liczba przeprowadzonych w drugim etapie takich estymacji jest równa liczbie cech badanych obiektów. Efektem tego jest możliwość selekcji zmiennych i prezentacji wyników na wykresie nazywanym biplotem.

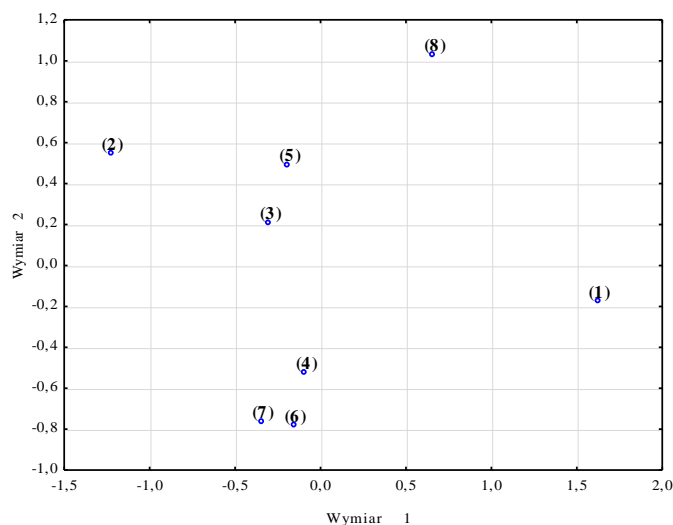
Biplot [Gabriel 1971] jest formą podwójnego wykresu, który w sposób graficzny przedstawia wzajemne relacje pomiędzy obiektami i opisującymi ich cechami. Poszczególne obiekty są przedstawiane jako punkty. Odległości między punktami na płaszczyźnie odwzorowują odległości między obiektami w przestrzeni wielowymiarowej. W formie wektorów są przedstawione badane cechy obiektów, wychodzące ze środka ciężkości zbioru tych punktów. Zwrot wektorów wskazuje na wzrost wartości cechy, którą wektor oznacza. Natomiast ich długości są proporcjonalne do wariancji cech, których dotyczą. Z kolei korelacje między cechami wyrażają cosinusy kątów między dwoma wektorami. Zatem im większy kąt pomiędzy wektorami tym mniejsza korelacja pomiędzy cechami [Malina i Wanat 2000]. Należy także pamiętać, że w interpretacji wyników nie ma znaczenia odległość badanych obiektów od wektora cechy, ale uszeregowanie ich rzutów na takim wektorze [Borgatti 1997].

Na wykresach przedstawiano tylko te cechy których jednowymiarowe wyniki regresji przeprowadzonych dla poszczególnych cech odznaczały się dopasowaniem na poziomie co najmniej $R^2 > 0,75$. Skalowanie wielowymiarowe wraz z szczegółowymi analizami przeprowadzono z wykorzystaniem analizy PROFIT dostępną w pakiecie PLUS w programie *Statistica*.

WYNIKI BADAŃ

Ze względu na fakt, że wszystkie uwzględnione w badaniu cechy były tego samego rodzaju, to analizy tej nie poprzedzono standaryzacją zmiennych. W wyniku zastosowania skalowania wielowymiarowego (por. wzór 1), dzięki zastosowaniu odległości euklidesowej (por. wzór 2) osiemnaście cech opisujących osiem badanych grup respondentów sprowadzono do dwóch wymiarów. W wyniku tej operacji każda z badanych jednostek otrzymała dwie współrzędne, dzięki czemu możliwe okazało się przedstawienie jej w postaci dwuwymiarowej mapy percepcji (rysunek 1). Wartość współczynnika STRESS dla skalowania wielowymiarowego uwzględniającego wszystkie cechy wynosiła 0,01. Wyniki procedury skalowania wielowymiarowego można zatem uznać za wiarygodne.

Rysunek 1. Wynik skalowania wielowymiarowego z uwzględnieniem wszystkich cech



Źródło: obliczenia własne na podstawie GfK Polonia

Uzyskany w wyniku skalowania wielowymiarowego obraz pozwala jedynie w sposób subiektywny pogrupować badanych respondentów ze względu na wskazywane przez nich czynniki decydujące o zakupie przez nich oleju rzepakowego. Niemniej jednak w prezentowanym przykładzie takich czynników jest osiemnaście. Uwzględnienie ich wszystkich na wykresie zaciemniłoby

otrzymany wynik. Jeszcze bardziej złożona w takim przypadku wydaje się analiza ilościowa wszystkich otrzymanych wyników. Z tego powodu wydaje się konieczne wyodrębnienie kluczowych czynników zakupowych, które różnicowały badane jednostki. W tym celu przeanalizowano wyniki regresji, w której zmienną objaśnianą była kolejna cecha determinująca wybór konsumenta, a zmiennymi objaśniającymi wartości dwóch wymiarów dla każdej jednostki, otrzymanych w wyniku skalowania wielowymiarowego: WYM. 1 oraz WYM. 2. Wyniki tych analiz zestawiono w tabeli 2. Na podstawie wartości współczynników determinacji w tych analizach można powiedzieć, że cechami wpływającymi na zróżnicowanie jednostek były smak, cena, naturalność produktu i dotychczasowe doświadczenie.

Tabela 2. Wyniki analizy regresji pomiędzy czynnikami zakupu a uzyskanymi w wyniku analizy regresji wymiarami badanych jednostek

Czynniki zakupu	b ₀	b _(WYM. 1)	b _(WYM. 2)	R ²
Cena	28,924*	3,669*	5,110*	0,967
Walory zdrowotne	7,558*	-0,777	-0,878	0,391
Wysoka jakość	13,427*	-0,283	-1,725	0,352
Promocja w sklepie	2,347*	0,145	0,501	0,135
Reklama produktu w mediach	0,545*	-0,089	0,038	0,071
Znana marka	1,179*	-0,118	-0,028	0,062
Kraj pochodzenia produktu	0,890*	-0,288	-0,099	0,385
Wygląd opakowania	0,644*	0,011	0,084	0,074
Wielkość opakowania	0,802*	-0,065	0,091	0,036
Praktyczne opakowanie/wygoda użycia	0,685*	-0,066	0,067	0,090
Uniwersalność produktu	1,100*	-0,039	0,146	0,116
Naturalność produktu	8,185*	-1,387*	-0,774	0,813
Smak	16,455*	4,018*	-3,512*	0,976
Przyjemny zapach	1,911*	-0,026	-0,051	0,080
Barwa produktu	0,855*	-0,095	0,068	0,223
Dodatek witamin	2,674*	-0,326	-0,077	0,338
Dotychczasowe doświadczenie	5,793*	-2,557*	0,802	0,792
Rodzaj surowca	6,010*	-1,709*	0,240	0,583

* istotność na poziomie $\alpha = 0,05$

Źródło: obliczenia własne podstawie GfK Polonia

Dla tych cech przeprowadzono analizę PROFIT i jej wyniki przedstawiono tym razem na biplocie (rysunek 2). Jak wspomniano wcześniej analiza skupień prowadzi do bardziej jednoznacznych wyników grupowania. Dendrogramy uzyskane w wyniku zastosowania tej metody przedstawiono na rysunku 3. Pierwszy z nich jest rezultatem grupowania badanych respondentów z uwzględnieniem

wszystkich czynników decyzyjnych. Podstawą drugiego były tylko cztery cechy wyłonione w wyniku oceny współczynników regresji, przeprowadzonej w ramach analizy PROFIT (tabela 2). Porównanie obu dendrogramów wskazuje na dobrą selekcję cech, które różnicowały respondentów pod względem czynników wpływających na podejmowane przez nich decyzje zakupowe dotyczące olejów rzepakowych. Należy jednak pamiętać, że są to czynniki, którym przeciętnie przypisywano w badaniu ankietowym najwyższe wartości (por. ostatnią kolumnę tabeli 1). Z tego względu odległości pomiędzy obiektami generowane na ich podstawie były z definicji wyższe niż te, które wyznaczono w oparciu o cechy o niższych wartościach [Walesiak 2004, Błażejczyk-Majka, Kala 2005]. Niemniej jednak ta sytuacja w prezentowanym badaniu koresponduje z oczekiwaniami jakie się im stawia.

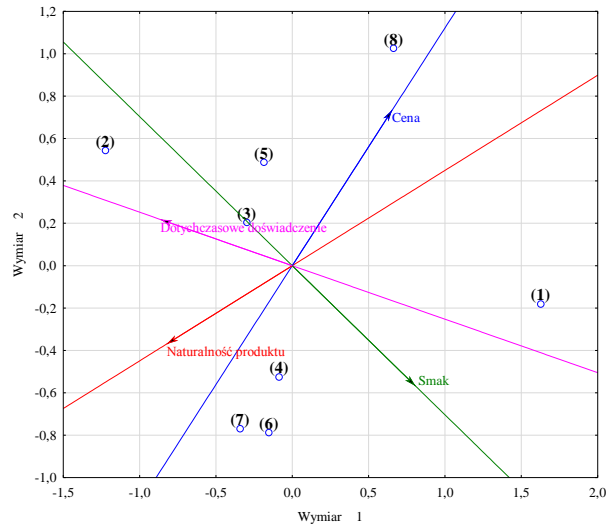
Na podstawie wyników analizy PROFIT wykazano, że tylko cztery czynniki mają decydujący wpływ na podejmowane przez konsumentów decyzje zakupowe dotyczące oleju rzepakowego (por. tabelę 2). Interpretacja biplotu przedstawianego na rysunku 2 pozwala zauważyć, że postrzeganie ceny było w analizowanych grupach respondentów niemal przeciwne, w porównaniu z oceną naturalności oleju rzepakowego. Z kolei respondenci, którzy przypisywali wyższe znaczenie w swoich decyzjach dotychczasowemu doświadczeniu w zakupach oleju rzepakowego, niżej oceniali wpływ smaku tego oleju na swoje decyzje. Dla przykładu, traktując wektor ceny jako oś interpretacji, można powiedzieć, że najmniejsze znaczenie przypisywali temu czynnikowi respondenci z grupy (7) starsza rodzina, pracująca, bez dzieci i (6) rodzina z dziećmi 15-25 lat oraz (4) rodzina z małymi dziećmi do 6 lat. Nieco wyżej czynnik ten wycenili respondenci należący do grupy (2) pracujący, wolni, mieszkający z rodzicami, (3) młodzi, bez dzieci, samodzielne gospodarstwo, (5) rodzina z dziećmi 7-14 lata oraz (1) studenci, wolni, mieszkający z rodzicami. Natomiast najwyższy wpływ czynnika ceny na decyzje zakupowe oleju rzepakowego wyceniły starsze rodziny, niepracujące, bez dzieci, które tworzyły grupę (8).

Połączenie wyników skalowania wielowymiarowego (rysunek 2) i analizy skupień (rysunek 3) pozwala na podział respondentów ze względu na postrzeganie czynników wpływających na ich decyzje zakupowe. Uśrednione wartości wyodrębnionych cech z podziałem na skupienia zestawiono w tabeli 3.

Należy podkreślić, że w każdej z wyodrębnionych grup czynnik cenowy okazał się najwyższej oceniany, drugie miejsce zajmował smak. W dalszej kolejności pojawiały się naturalność produktu i dotychczasowe doświadczenie. Niemniej jednak wielkości różnic pomiędzy ocenami przypisywanymi wymienionym czynnikom w wyodrębnionych skupieniach rozkładały się różnie. Pierwsze skupienie tworzą respondenci którzy (2) pracują, są wolni i mieszkają z rodzicami, (3) są młodzi, nie mają dzieci, ale prowadzą samodzielne gospodarstwo oraz (5) tworzą rodzinę z dziećmi w wieku 7 do 14 lat. Respondenci ci, w porównaniu z innymi skupieniami, niżej oceniali wpływ smaku. Za to relatywnie wyżej

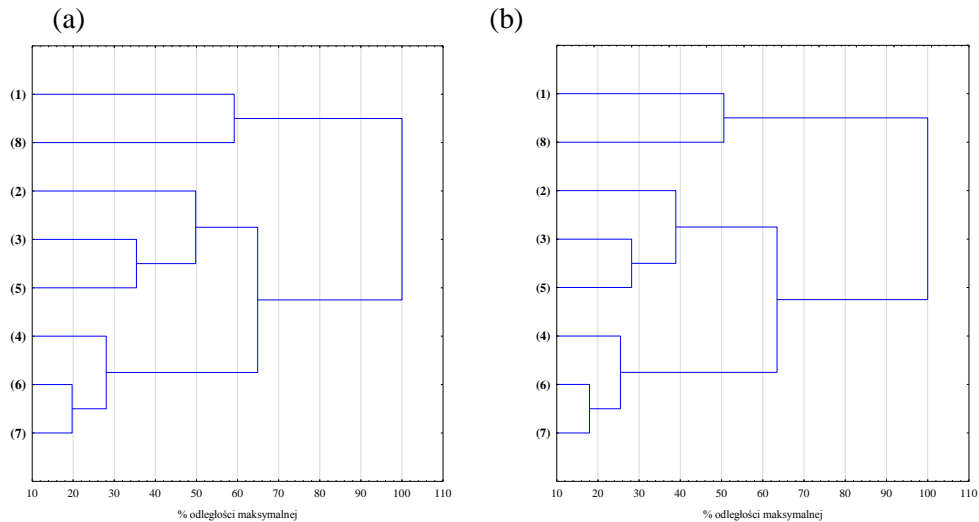
oceniają wpływ dotychczasowego doświadczenia na podejmowane przez nich decyzje zakupowe.

Rysunek 2. Biplot uwzględniający wynik skalowania wielowymiarowego w oparciu o cztery cechy determinujące zakupy oleju rzepakowego: cenę, smak, doświadczenie i naturalność produktu



Źródło: obliczenia własne na podstawie GfK Polonia

Rysunek 3. Dendrogramy przedstawiające wyniki analizy skupień badanych grup respondentów uwzględnieniem (a) 18 cech oraz (b) 4 cech



Źródło: obliczenia własne na podstawie GfK Polonia

Tabela 3. Uśrednione wartości i odchylenia standardowe ocen czynników zakupowych dotyczących oleju rzepakowego w skupieniach respondentów wyodrębnionych w wyniku skalowania wielowymiarowego i analizy skupień

Faza życia	Smak	Cena	Naturalność produktu	Dotychczasowe doświadczenie
młodszy, bez dzieci, samodzielne gospodarstwo (3)	14,33 (13,73)	27,43 (28,74)	8,72 (9,27)	7,54 (9,96)
rodzina z dziećmi 7-14 lat (5)	13,36 (14,63)	31,39 (31,29)	8,68 (8,55)	4,80 (5,80)
pracują, wolni, mieszkają z rodzicami (2)	9,63 (5,46)	27,92 (30,73)	8,65 (5,84)	9,27 (12,43)
rodzina z dziećmi 15-25 lat (6)	18,73 (17,50)	23,92 (29,57)	8,21 (8,02)	4,43 (6,37)
rodzina z małymi dziećmi do 6 lat (4)	19,20 (15,19)	26,94 (27,47)	8,99 (9,10)	7,18 (10,82)
starsza rodzina, pracująca, bez dzieci (7)	17,12 (15,21)	23,25 (26,15)	10,01 (9,38)	6,31 (8,01)
starsza rodzina, niepracująca, bez dzieci (8)	16,03 (15,42)	36,12 (30,09)	6,57 (8,00)	5,75 (8,79)
studenci, wolni, mieszkają z rodzicami (1)	23,24 (20,72)	34,42 (35,92)	5,64 (3,60)	1,06 (0,69)

() odchylenie standardowe.

Źródło: obliczenia własne na podstawie GfK Polonia

Drugie skupienie, w skład którego wchodzi (6) rodziny z dziećmi 15-25 lat, (4) rodziny z małymi dziećmi do 6 lat oraz (7) starsza rodzina, pracująca, bez dzieci, charakteryzuje się z kolei w swoich decyzjach zakupowych relatywnie wysokim wpływem smaku oleju rzepakowego i jego naturalności. Z kolei najwyższe oceny dla ceny przypisali respondenci należący do grupy (8) starsza rodzina, niepracująca, bez dzieci. Studenci, wolni, mieszkający z rodzicami (1) w porównaniu z innymi grupami relatywnie wysoko ocenili wpływ smaku i ceny na swoje decyzje i ekstremalnie nisko ocenę wpływu doświadczenia i naturalność tego produktu.

PODSUMOWANIE I WNIOSKI

W pracy przedstawiono zastosowanie wybranych metod wielowymiarowych do oceny preferencji zakupowych polskich konsumentów olejów rzepakowych. W badaniu wzięto pod uwagę osiemnaście takich czynników. Grupę respondentów analizowano pod względem fazy życia, w której znajdowali się w momencie

przeprowadzania ankiety. W badaniu wykorzystano skalowanie wielowymiarowe, analizę skupień i analizę PROFIT. Punktem wyjścia w wymienionych metodach jest wyznaczenie odległości skalarnej pomiędzy badanymi obiektami. Niemniej jednak każda z tych metod pozwala na prezentację wyników pod nieco innym kątem.

Skalowanie wielowymiarowe pozwoliło na przedstawienie grup respondentów na mapie percepcji w przestrzeni dwuwymiarowej. W wyniku analizy skupień uzyskano precyzyjny i obiektywny podział respondentów. Natomiast w wyniku przeprowadzenia analizy PROFIT wyodrębniono cechy obiektów decydujące o zróżnicowaniu respondentów i prezentację wyników na biplotach. Ten ostatni rodzaj wykresu pozwala jednocześnie analizować wyniki grupowania jednostek oraz ich relacje do wyodrębnionych w wyniku skalowania wielowymiarowego kluczowych cech, decydujących o wynikach grupowania.

Wyniki przeprowadzonych analiz wskazują, że decydujące przy wyborze oleju okazały się cena, smak, naturalność tego produktu czy dotychczasowe doświadczenie respondentów. Zatem w kampaniach reklamowych skierowanych do rodzin wielopokoleniowych z aktywnymi zawodowo dziadkami i małymi wnukami powinno się kłaść nacisk na smak oleju rzepakowego i jego naturalność. Jeśli bohaterami takiej kampanii byłyby młode pracujące osoby, tworzące już rodziny, lub osoby jeszcze wolne, to należy podkreślać takie walory polskiego oleju rzepakowego jak jego cena i tradycja występowania na polskich stołach. Do starszych ludzi, już niepracujących przemawia przede wszystkim argument ceny, a do młodych osób na starcie życia zawodowego czynnikami decydującymi o wyborze przez nich oleju rzepakowego będzie nie tylko jego cena ale i smak.

Należy podkreślić, że wśród decydujących czynników o wyborze oleju rzepakowego przez respondentów nie znalazł się żaden, który można byłoby wiązać z poczuciem etnocentryzmu czy świadomością respondentów w obszarze jego wysokiej jakości. Wydaje się zatem, że w tym kierunku powinny być prowadzone dalsze działania mające na celu podniesienie świadomości konsumentów na temat jakości krajowego oleju rzepakowego.

BIBLIOGRAFIA

- Bartłomowicz T. (2014) Implementacja metody Maximum Difference Scaling w pakiecie MaxDiff programu R. *Ekonometria*, 4 (46), 189-198.
- Błażejczyk-Majka L., Kala R. (2005) Metody analizy skupień do charakterystyki użytków rolniczych wybranych państw unijnych i Polski. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 7, 5-10.
- Boczar P., Błażejczyk-Majka L. (2015) Characteristics of vegetable oil consumers in Poland in a view of sustainable consumption principles. *Acta Scientiarum Polonorum. Seria: Oeconomia*, 14 (3), 15-26.
- Borgatti S. P. (1997) PROFIT. <http://www.analytictech.com/borgatti/profit.htm>

- Cormack R. M. (1971) A review of classification. *Journal of the Royal Statistical Society, Series A*, 134, 321-367.
- Cox T. F., Cox M. A. A. (1991) *Multidimensional Scaling*. Chapman and Hall, London.
- Gabriel K. R. (1971) The Biplot Graphic Display of Matrices with Applications to Principal Components Analysis. *Biometrika*, 58, 453-467.
- Gordon A. D. (1981) *Classification*. Chapman and Hall, London.
- Green P. E., Carmone F. J., Smith S. M. (1989) *Multidimensional Scaling. Concept and Applications*, Allyn and Bacon, Boston, London, Sydney, Toronto.
- Härdle W. K., Simar L. (2012) *Applied Multivariate Statistical Analysis*. Springer Verlag, Berlin Heidelberg.
- Jabkowski P. (2010) O korzyściach wynikających z zastosowania analizy PROFIT. [w:] *Praktyczna analiza danych w marketingu i badaniach rynku*. StatSoft Polska, Kraków, 89-102.
- Malina A., Wanat St. (2000) Metody skalowania wielowymiarowego w badaniach rozwoju społeczno-gospodarczego Polski. *Prace naukowe Akademii Ekonomicznej we Wrocławiu*, 874, Taksonomia 7: Klasyfikacja i analiza danych. Teoria i zastosowania, 71-81.
- Mardia K. V., Kent J. T., Bibby J. M. (1979) *Multivariate analysis*. Academic Press Inc., London.
- Marek T. (1989) *Analiza skupień w badaniach empirycznych. Metody SAHN*. PWN, Warszawa.
- Migut G. (2012) Nowe możliwości analizy danych – Statistica zestaw Plus. [w:] *Analiza danych w programie Statistica – przegląd*. StatSoft Polska Sp. z o.o., 39-58.
- Sagan A. (2009) 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., 3-22.
- Shepard R. N. (1962) Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. Part I. *Psychometrika*, 27 (2), 125-140.
- Timm N. H. (2002) *Applied Multivariate Analysis*. Springer-Verlag, New York.
- Torgerson W. S. (1952) Multidimensional Scalling: Theory and Method. *Psychometrika*, 17, 401-419.
- Walesiak M. (1996) *Metody analizy danych marketingowych*. PWN, Warszawa.
- Walesiak M. (2004) Problemy decyzyjne w procesie klasyfikacji zbioru obiektów. *Ekonometria*, 13, *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 1010, 52-71.
- Zaborski A. (2001) *Skalowanie wielowymiarowe w badaniach marketingowych*. Wydawnictwo Akademii Ekonomicznej im. Oskara Lange, Wrocław.
- Zaborski A. (2009) *Skalowanie wielowymiarowe*. [w:] *Statystyczna analiza danych z wykorzystaniem programu R*. M. Walesiak (red.), E. Gatnar. PWN, Warszawa.
- Zaborski A. (2014) Analizy preferencji słuchaczy Uniwersytetu Trzeciego Wieku z wykorzystaniem wybranych metod niesymetrycznego skalowania wielowymiarowego. *Studia Ekonomiczne*, 195, 216-224.

APPLICATION OF THE MULTIVARIATE METHODS TO THE CHARACTERISTICS OF CONSUMERS

Abstract: In the paper the use of multivariate methods in the analysis of consumer preferences was presented. The convergences and differences of the multidimensional scaling to cluster analysis and PRPFIT method were shown. As an example of using those methods were the results of a survey about the consumption of rapeseed oil in Poland. In a result of this research it has been shown that purchasing decisions consumers were varied, especially in terms of consumers' perception of the taste of rapeseed oil, its naturalness and price and previous experience in its consumption.

Keywords: multidimensional scaling, PROperty FITting method, cluster analysis, consumer, rapeseed oil