



## **ANALIZA PREFERENCJI KONSUMENTÓW Z WYKORZYSTANIEM PROGRAMU STATISTICA – ANALIZA CONJOINT I SKALOWANIE WIELOWYMIAROWE**

*Adam Sagan, Uniwersytet Ekonomiczny w Krakowie*

### **Preferencje w zachowaniach konsumenta**

Badania preferencji konsumentów stanowią podstawowe pole zainteresowań badaczy marketingowych. Stanowią one podstawową kategorię badawczą w modelowaniu zachowań konsumentów na rynku, uwzględniającą racjonalność procesu podejmowania decyzji. Gintis wymienia szereg założeń związanych z rolą preferencji w procesie racjonalnego zachowania konsumentów na rynku: preferencje dotyczą zarówno efektów procesu decyzyjnego, np. ilości i jakości porównywanych ofert rynkowych (*outcome-related preferences*), jak również dotyczą samego procesu odnoszącego się do dystrybucji i komunikacji ofert (*proces-related preferences*), źródła preferencji mogą odnosić się do własnego systemu wartościowań i zadowolenia z ofert (*self-regarding preferences*), jak i wchodzić w interakcje z preferencjami innych konsumentów i ich systemami wartości (*other-regarding preferences*), preferencje mogą być traktowane jako autonomiczne, wynikające z realizowanych celów konsumenta (*exogenous preferences*) lub są ujmowane jako efekt modelowania poprzez modę, reklamę lub naśladownictwo (*endogenous preferences*)<sup>1</sup>.

W badaniach marketingowych wyróżnia się dwie podstawowe grupy modelowego ujęcia preferencji konsumentów. Do pierwszej należą tzw. modele kompensacyjne, wynikające z założeń mikroekonomicznych racjonalnego zachowania konsumenta: neutralizacji alternatyw wyboru (*cancellation*), przechodniości (*transitivity*), dominacji (*dominance*) i niezmienniczości (*invariance*). Do modeli kompensacyjnych należą model oczekiwanej wartości oraz model idealnej marki. Druga grupa modeli preferencji tzw. modele niekompensacyjne, są częściej rozważane w ramach nurtów ewolucyjnych w ekonomii i ekologicznych w psychologii. Do tych modeli zalicza się model leksykograficzny, model koniunkcyjny, model dysjunkcyjny oraz model determinacji.

---

<sup>1</sup> Gintis H., The individual in economic theory: a research agenda, Dept. of Economics, University of Massachusetts, Amherst, 1998.

## Metody pomiaru preferencji

Preferencje konsumentów są najczęściej traktowane jako nieobserwowalny bezpośrednio konstrukt teoretyczny, wynikający z określonej teorii mikroekonomicznej lub modelu zachowania konsumenta. Ich pomiar dokonywany jest na podstawie określonych deklaracji wyrażonych na odpowiednich skalach pomiarowych (*stated preferences*) lub ujawniane są poprzez obserwacje rzeczywistych rynkowych wyborów konsumentów (*revealed preferences*). Najbardziej popularne ujęcie metod pomiaru preferencji wynikają z teorii danych zaproponowanej przez C. H. Coombsa, który zaproponował klasyfikacje rodzaju uzyskiwanych danych, jak i metod ich gromadzenia i analizy na podstawie dwóch kryteriów: charakteru relacji między danymi oraz liczby porównywanych typów obiektów. Z punktu widzenia charakteru relacji dane mogą mieć charakter relacji podobieństwa (bliskości) lub dominacji (preferencji), a z punktu widzenia typów porównywanych obiektów porównania mogą być dokonywane w obrębie jednego zbioru (jednostka vs obiekt) lub dwóch zbiorów obiektów (np. par punktów A-B vs C-D). Skrzyżowanie tych kryteriów daje w konsekwencji cztery podstawowe rodzaje danych: 1/ pojedynczego bodźca, 2/ preferencyjnego wyboru, 3/ porównania bodźców i 4/ podobieństwa między bodźcami. Pierwszy rodzaj danych jest charakterystyczny dla skal ocen (np. Guttmana czy Likerta), drugi typ danych cechuje wielowymiarowe skalowanie preferencji oparte na teorii rozwijania (*unfolding*), trzeci jest właściwy dla skal porównawczych rangowych, porównań preferencyjnych par obiektów, skali V Thurstone'a, porządkowego sortowania i techniki punktu kotwicznego, a czwarty rodzaj danych obejmuje dane uzyskane na podstawie ocen podobieństw między parami diad obiektów (skala porównań par), sortowania i techniki triad stanowiących podstawę wielowymiarowych skal percepcji<sup>2</sup>. W badaniach preferencji konsumentów typowym rodzajem uzyskiwanych danych są dane oparte na porównywaniu bodźców (skale rangowe, porównań par, punktu kotwicznego) oraz dane preferencyjnego wyboru odnoszącego się do różnic między obiektem a punktem idealnym (jednostką). Są one podstawą wyboru między kompozycyjnymi a dekompozycyjnymi metodami pomiaru preferencji.

Wyróżnia się trzy podstawowe grupy metod pomiaru preferencji. Pierwszą grupę metod stanowią tzw. metody dekompozycyjne, drugą grupę metody kompozycyjne pomiaru preferencji, a trzecią – metody mieszane. Metody dekompozycyjne są głównie związane z danymi uzyskiwanymi na podstawie porównywania obiektów między sobą. Podstawą pomiaru preferencji jest równoczesna ocena, szeregowanie (lub wybór) porównywanego zbioru marek lub kategorii produktów opisanych za pomocą charakteryzującego je zbioru atrybutów o określonych poziomach lub własnościach (realizacjach). Respondenci oceniają, rangują lub dokonują wyborów profili produktów i na tej podstawie szacowane są całkowite ich użyteczności. W kolejnym etapie analizy użyteczności całkowite są dekomponowane (stąd nazwa podejścia) na użyteczności cząstkowe poszczególnych poziomów lub własności atrybutów. Do podstawowych metod analizy struktury preferencji w tym nurcie

<sup>2</sup> C.H. Coombs, R.M. Dawes, A. Tversky, Wprowadzenie do psychologii matematycznej, PWN Warszawa 1977, s. 62.

należy analiza conjoint (zwana także analizą łącznego współwystępowania zmiennych i pomiarem wieloczynnikowym).

Metody kompozycyjne polegają na niezależnych ważonych ocenach lub porównaniach poszczególnych cech produktów dokonywanych na podstawie skal ocen lub rangowych wraz z ogólną oceną preferencji tych produktów. Wpływ oceny danego poziomu cechy na inny nie jest w tym podejściu identyfikowany. Są one silniej związane z danymi dotyczącymi wyborów preferencyjnych. Na podstawie dokonanych ewaluacji poszczególnych cech uzyskiwana jest ogólna struktura preferencji danego zbioru produktów (stąd nazwa podejścia). Podstawową metodą analizy preferencji w tym nurcie jest wielowymiarowe skalowanie preferencji wyrażone w postaci map graficznych zbudowanych w zredukowanej przestrzeni wielowymiarowej.

Metody mieszane stanowią połączenie poprzednich podejść. Do najbardziej znanych metod mieszanych należą hybrydowe i adaptacyjne metody analizy conjoint<sup>3</sup>.

W opracowaniu zostaną przedstawione zastosowania programu *STATISTICA* w trzech typach analiz związanych z badaniami preferencji konsumentów: 1/ analizy conjoint, 2/ budowy skali V ocen porównawczych Thurstone'a i 3/ analizy dopasowania własności (*property fitting* – PROFIT) z wykorzystaniem połączonych metod wielowymiarowego skalowania percepcji i analizy regresji.

## Analiza conjoint

Analiza conjoint jest zestawem procedur pomiarowo-analitycznych opartych na zasadzie (teorii) pomiaru psychometrycznego, zwanej zasadą równoczesnego addytywnego pomiaru łącznego. Zgodnie z nią pomiar danej cechy wyrażonej poprzez wartości zmiennej (zależnej) jest możliwy z wykorzystaniem przyczynowo związanych zmiennych niezależnych oddziałujących na mierzoną cechę w sposób jednoczesny i addytywny. Podstawowe kroki w analizie conjoint są następujące<sup>4</sup>:

1. Określenie przedmiotu analizy: specyfikacja skali pomiaru preferencji (ocena, ranking lub wybór profilu), wybór badanych produktów, ich cech oraz poziomów. Na tym etapie należy określić także liczbę profili powstałych na podstawie kombinacji cech i ich poziomów. Liczba ta zależy od możliwości percepcji respondentów, metody estymacji modelu i liczebności próby. Dla popularnych w analizie conjoint metod regresyjnych minimalna liczba profili jest równa łącznej liczbie poziomów cech minus liczba atrybutów plus jeden.

<sup>3</sup> Zob. Analiza danych marketingowych. Problemy, metody, przykłady, red. A. Stanimir, AE Wrocław 2006, s. 162.

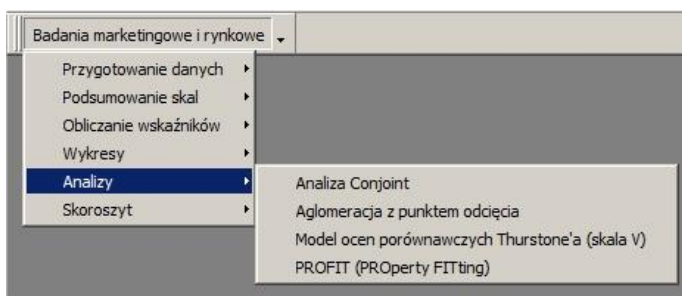
<sup>4</sup> Na temat analizy conjoint zob.: M. Walesiak, A. Bąk, Conjoint analysis w badaniach marketingowych, AE Wrocław 2000, A. Bąk, Analiza conjoint, w.: M. Walesiak, E. Gatnar, Statystyczna analiza danych z wykorzystaniem programu R, PWN Warszawa 2009, s. 283-317.



2. Określenie postaci modelu: zależności między zmiennymi niezależnymi (modele bez interakcji i z interakcjami między poziomami cech) i przyjmowany model preferencji. Najczęściej jest to model kompensacyjny oczekiwanej wartości – liniowy (wektorowy), idealnej marki – kwadratowy lub dyskretny model odrębnych użyteczności częściowych.
3. Sposób gromadzenia danych i generowania profili: wybór między metodą pełnych profili, porównywania profili parami, porównywania atrybutów parami lub wyboru profilu. Profile te mogą być generowane za pomocą planów czynnikowych lub metodą losową.
4. Wybór skali pomiaru preferencji: ma wpływ na metodę estymacji parametrów modelu, uwzględnienie interakcji między poziomami zmiennych niezależnych, sposób reagowania respondenta na przedstawiane profile. Do najczęściej wykorzystywanych skal pomiaru preferencji należą skale: rangowe, szacunkowe skale ocen i skale porównań parami.
5. Metoda estymacji modelu: metody niemetryczne (np. monotoniczna analiza wariancji MONANOVA), metody metryczne (metoda najmniejszych kwadratów) i metody probabilistyczne (regresja logistyczna i probitowa). Wybór metody estymacji zależy przede wszystkim od skali pomiaru zmiennej zależnej. Dla skal metrycznych (skal ocen, stopniowalnych skal porównań parami) stosowana jest metoda najmniejszych kwadratów, dla skal porządkowych (rangowych, porównań par) wykorzystywana jest monotoniczna analiza wariancji, a dla podejść opartych na dyskretnych wyborach – modele logitowe i probitowe.
6. Interpretacja i wykorzystanie wyników: ocena użyteczności całkowitych i częściowych, interpretacja profili użyteczności, ranking ważności atrybutów, analiza kompromisów (*trade-off*). Wyniki analizy mogą być wykorzystywane do symulacji udziałów rynkowych (na podstawie modeli użyteczności maksymalnej, probabilistycznego modelu Bradleya-Terryego-Luce'a lub modelu logitowego), analiz optymalizacyjnych nowego produktu i badań segmentacyjnych konsumentów na podstawie wartości użyteczności częściowych w segmentacji post-hoc.

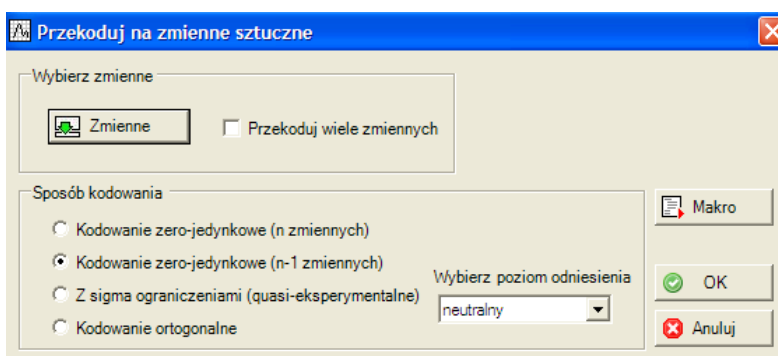
## Analiza conjoint w programie **STATISTICA**

Program *STATISTICA* umożliwia przeprowadzenie analizy conjoint z wykorzystaniem metody najmniejszych kwadratów. Oznacza to, że do analizy przyjmowane są zmienne zależne mierzone na skali co najmniej przedziałowej, zależność między poziomami cech produktu a ujawnionymi preferencjami jest liniowa i identyfikowane są jedynie efekty główne bez interakcji. Program analizy conjoint znajduje się w dodatkowym programie, zwanym *STATISTICA dla badań marketingowych i rynkowych*, który umożliwia wykonanie wielu dodatkowych analiz nie znajdujących się w podstawowym pakiecie programu. Analiza conjoint jest dostępna w grupie *Analizy* (zob. rys. 1).



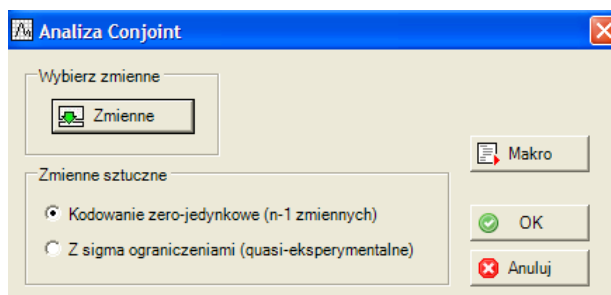
Rys. 1. Program *STATISTICA dla badań marketingowych i rynkowych*.

W celu wykonania analizy conjoint zmienne niezależne stanowiące poziomy cech produktów muszą być wyrażone w postaci tzw. zmiennych sztucznych. Dodatkowy program pozwala na automatyczne przekodowanie zmiennych kategoryalnych w zmienne sztuczne, wykorzystując zero-jedynkowe kodowanie regresyjne (tzw. *reference coding*, *dummy coding*), kodowanie eksperymentalne z sigma ograniczeniami (tzw. *ANOVA coding*, *effect coding*, *deviation coding*) i kodowanie ortogonalne.



Rys. 2. Moduł kodowania zmiennych sztucznych w *STATISTICA dla badań marketingowych i rynkowych*.

W programie analizy conjoint automatycznie można wykonać kodowanie zero-jedynkowe lub eksperymentalne.



Rys. 3. Analiza conjoint w *STATISTICA dla badań marketingowych i rynkowych*.



Po przeprowadzeniu automatycznego rekodowania zmiennych niezależnych zbudowany zostanie plik wejściowy do analizy.

STATISTICA - [Dane: cake (11 zmn. \* 1088 prz.)]

	Profil	Cena 1.5	Cena 2.0	Cena 2.5	Smak neutralny	Smak rodzynkowy	Smak ananasowy	Marka producenta	Marka dystrybutora	Ocena	Respondent
1	1	1	0	0	1	0	0	-1	1	7	1
2	2	1	0	0	0	1	0	1	1	5	1
3	3	1	0	0	0	0	1	1	-1	5	1
4	4	1	0	0	-1	-1	-1	-1	-1	6	1
5	5	0	1	0	1	0	0	1	-1	7	1
6	6	0	1	0	0	1	0	-1	-1	6	1
7	7	0	1	0	0	0	1	-1	1	8	1
8	8	0	1	0	-1	-1	-1	1	1	8	1

Rys. 4. Plik wejściowy do analizy conjoint.

Plik wejściowy składa się ze zmiennej identyfikującej prezentowane w badaniach profile produktów (zmienna Profil), zmiennych sztucznych określających poziomy cech produktu (Cena, Smak, Marka), zmiennej zależnej dotyczącej pomiaru preferencji (Ocena) i zmiennej reprezentującej kod respondenta (Respondent). W przykładzie wykonano kodowanie quasi-eksperymentalne.

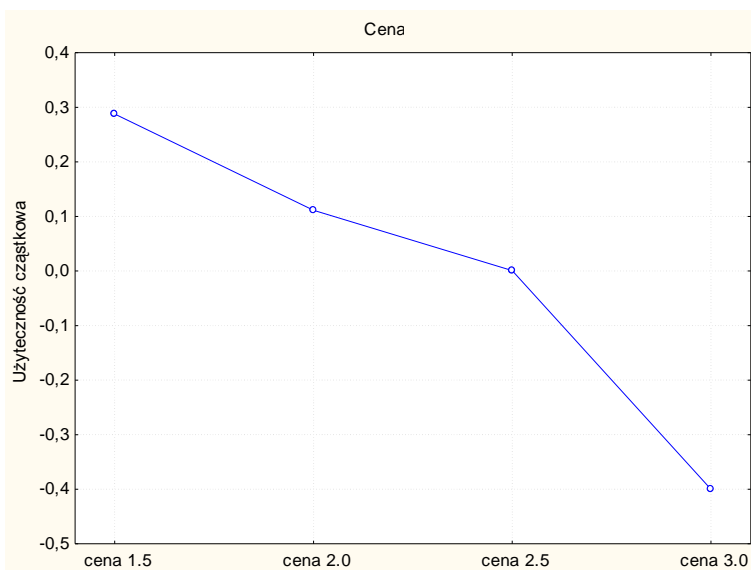
Po wyborze odpowiednich zmiennych do analizy program wykonuje analizę conjoint i przedstawia oszacowane parametry regresji, stanowiące częściowe użyteczności poszczególnych poziomów cech. Macierz użyteczności częściowych dla pierwszych czterech respondentów jest podana poniżej.

	1	2	3	4
	Cena cena 1.5	Cena cena 2.0	Cena cena 2.5	Cena cena 3.0
Respondent: 1	-0.188	1.312	0.813	-1.937
Respondent: 2	-0.000	-0.250	-0.000	0.250
Respondent: 3	-0.062	0.687	-0.313	-0.313
Respondent: 4	0.188	-0.562	0.187	0.188
Respondent: 5	1.500	1.500	-0.500	-2.500
Respondent: 6	1.500	0.250	-0.750	-1.000
Respondent: 7	0.188	0.187	0.188	-0.563

Rys. 5. Macierz użyteczności częściowych.

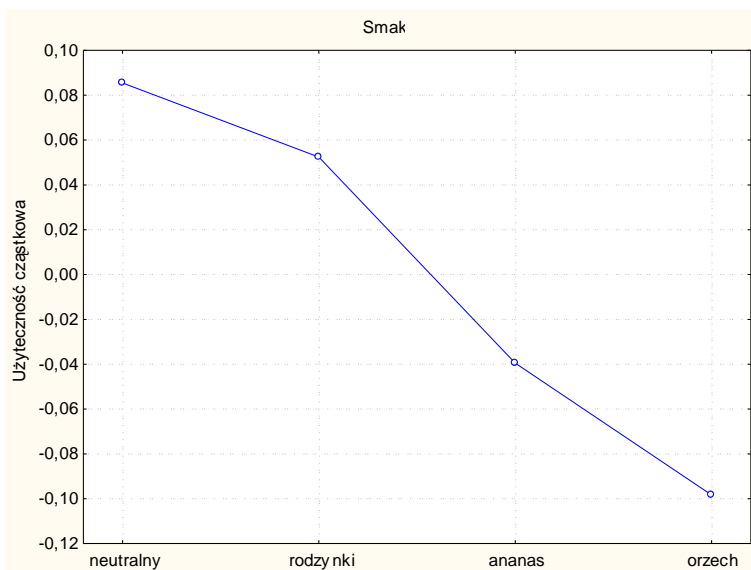
Tabela przedstawia wartości częściowych użyteczności poziomów zmiennych dla poszczególnych badanych. Należy zauważyć, że wartość użyteczności dla referencyjnego poziomu danej zmiennej dla kodowania eksperymentalnego wynosi 1- suma pozostałych użyteczności (w przypadku kodowania regresyjnego zawsze wynosi zero).

Wygodnym sposobem przedstawiania zagregowanych poziomów użyteczności dla poszczególnych atrybutów są wykresy interakcji.



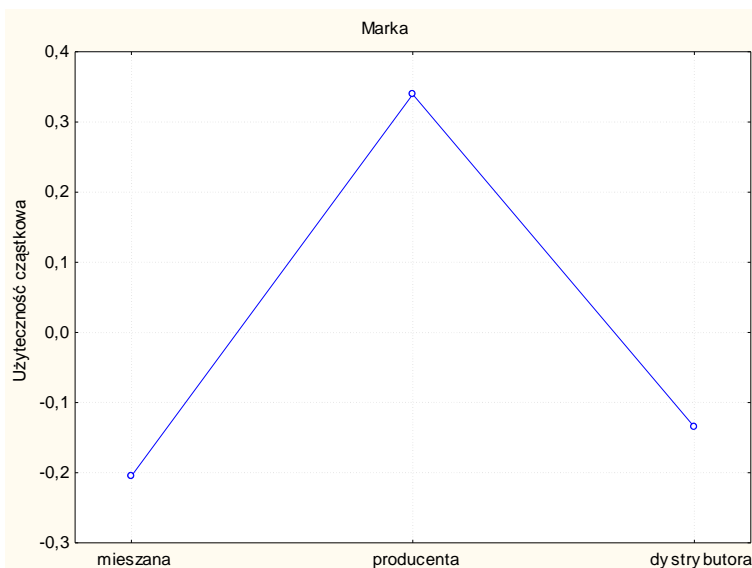
Rys. 6. Wykres użyteczności częściowych dla ceny.

Wykres 6 przedstawia strukturę preferencji dla poszczególnych poziomów ceny. Wynika z niego, że użyteczność ceny dosyć dobrze opisuje model wektorowy – im wyższy poziom ceny ciastek, tym niższa wartość funkcji użyteczności.



Rys. 7. Wykres użyteczności częściowych dla smaku.

Dla poszczególnych kategorii smaku ciasta zdecydowanie najsilniej preferowanym jest smak neutralny (bez dodatków smakowych), w następnej kolejności smak rodzynekowy, ananasowy i orzechowy.



Rys. 8. Wykres użyteczności cząstkowych dla marki.

Ostatnią analizowaną cechą jest charakter marki. Najbardziej preferowane są marki producenta, w następnej kolejności marki dystrybutora i mieszane.

Na podstawie informacji o strukturze cząstkowych użyteczności można obliczyć użyteczności całkowite dla każdej kombinacji cech produktu i każdego respondenta.

Użyteczność dla wszystkich profili i respondentów						
	1	2	3	4	5	6
Cena	Smak	Marka	Resp. 1	Resp. 2	Resp. 3	
1	cena	neutral	mieszana	4.625	6.25	4.125
2	cena	neutral	producenta	5.375	6.25	5.875
3	cena	neutral	dystrybutora	4.875	6.25	5.875
4	cena	rodzynek	mieszana	4.125	7	4.625
5	cena	rodzynek	producenta	4.875	7	6.375
6	cena	rodzynek	dystrybutora	4.375	7	6.375
7	cena	ananas	mieszana	6.375	6.5	5.625
8	cena	ananas	producenta	7.125	6.5	7.375
9	cena	ananas	dystrybutora	6.625	6.5	7.375

Rys. 9. Użyteczności całkowite.

Wygodnym sposobem interpretacji atrybutów jest prezentacja względnej ich ważności dla każdego respondenta.

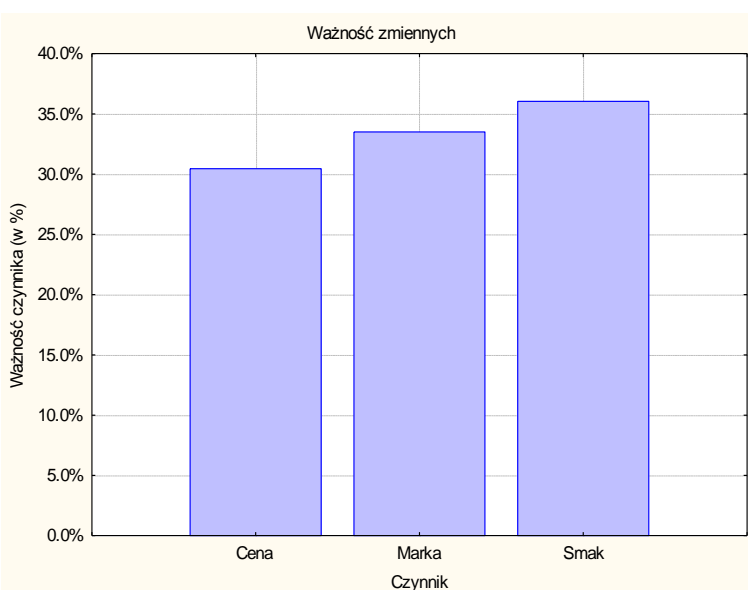


	Relatywne ważności czynników w %		
	1 Respondent: 1	2 Respondent: 2	3 Respondent: 3
Cena	52.0%	40.0%	23.5%
Smak	36.0%	60.0%	35.3%
Marka	12.0%	0.0%	41.2%

Rys. 10. Tabela względnej ważności czynników.

Z tabeli wynika, że dla respondenta nr 1 najważniejszym atrybutem jest cena, następnie smak i marka, a np. dla respondenta nr 3 najważniejszą cechą okazała się marka, w dalszej kolejności smak i cena.

Sumaryczne zestawienie względnej ważności atrybutów w przekroju wszystkich badanych jest uwidocznione na rys. 11.



Rys. 11. Wykres względnej ważności czynników.

Z rysunku wynika, że najważniejszym czynnikiem przy wyborze badanych ciastek jest smak, a kolejne miejsca zajmują marka i cena.

Uszeregowanie wszystkich kombinacji atrybutów wraz z całkowitymi użytecznościami dla danego profilu jest przedstawione w tabeli.

	Ranking profili			
	1 Cena	2 Smak	3 Marka	4 MW
1	cena 1.5	neutralny	producent	7.105
2	cena 1.5	rodzynki	producent	7.072
3	cena 1.5	ananas	producent	6.98
4	cena 2.0	neutralny	producent	6.928
5	cena 1.5	orzech	producent	6.921
6	cena 2.0	rodzynki	producent	6.895
7	cena 2.5	neutralny	producent	6.818
8	cena 2.0	ananas	producent	6.803
9	cena 2.5	rodzynki	producent	6.785
10	cena 2.0	orzech	producent	6.744
11	cena 2.5	ananas	producent	6.693
12	cena 1.5	neutralny	dystribut	6.645
13	cena 2.5	orzech	producent	6.634
14	cena 1.5	rodzynki	dystribut	6.612
15	cena 1.5	neutralny	mieszana	6.546

Rys. 12. Struktura użyteczności profili marek.

Najwyższymi preferencjami charakteryzują się ciastka o marce producenta, cenie na poziomie 1.5 i smakiem neutralnym, następnie inne produkty o tej samej cenie i marce o innych smakach.

## Metoda ocen porównawczych Thurstone'a

Metoda ocen porównawczych Thurstone'a umożliwia zbudowanie jednowymiarowej metrycznej skali preferencji na podstawie danych o preferencjach uzyskanych z wykorzystaniem skali porównań parami. Metoda ta nosi też nazwę modelu V ocen porównawczych Thurstone'a, który wraz z modelem III Thurstone'a i modelem Takane-Thurstone'a jest najczęściej stosowanym podejściem w pomiarze preferencji traktowanych jako ciągła metryczna zmienna ukryta. Taki sposób ujęcia preferencji wymaga spełnienia pewnych założeń<sup>5</sup>:

- ♦ preferencje ujawnione na podstawie wyborów z par porównywanych marek produktów mają charakter ciągły (ciągły proces dyskryminacyjny),
- ♦ w ramach danej pary obiektów wybór danego obiektu jest dokonany na zasadzie maksymalizacji ciągłych preferencji dotyczących porównywanych marek,
- ♦ rozkład ukrytych i ciągłych preferencji w populacji jest normalny,
- ♦ preferencje są niezależne od siebie i mają wspólne źródło wariancji,
- ♦ prawdopodobieństwo nieprzechodnich preferencji jest różne od zera,
- ♦ rozkład błędów pomiaru jest normalny i są one nieskorelowane.

<sup>5</sup> Zob. C.H. Coombs, R.M. Dawes, A. Tversky, Wprowadzenie do psychologii matematycznej, PWN Warszawa 1977, s. 71-83.



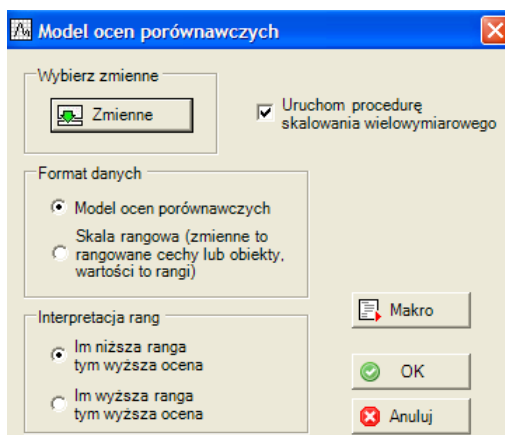
Skala porównań parami obok prostej skali rangowej jest najczęściej wykorzystywaną skalą służącą do gromadzenia danych o preferencjach. W odróżnieniu od absolutnych i monadycznych skal ocen mają one charakter skal względnych i porównawczych, określających miejsce danej cechy lub marki ze względu na inną analizowaną cechę lub markę. Skale rangowe ze względu na swoją konstrukcję uwzględniają jedynie preferencje przechodnie i nie pozwalają na uwzględnienie nieprzechodnich preferencji (jeżeli A jest preferowane nad B i B nad C, to zawsze A jest preferowane nad C, co jest uwzględniane w uporządkowaniu rangowym  $A - 1, B - 2, C - 3$ ). Skala porównań parami umożliwia również identyfikację preferencji nieprzechodnich, bowiem każda para marek jest porównywana niezależnie, co pozwala na ocenę sytuacji, w której A jest preferowane nad B, B jest preferowane nad C, a C jest preferowane nad A. W skali porównań parami dla  $n$  porównywanych marek wszystkich porównań jest  $n(n-1)/2$ , a liczba par nieprzechodnich wynosi  $n!$ . Podstawowe etapy w metodzie ocen porównawczych są następujące<sup>6</sup>:

1. Wybór zbioru porównywanych marek.
2. Określenie sposobu prezentacji: skala rangowa, porządkowanie marek, skala porównań parami. Ostatnia skala jest najbardziej uniwersalnym sposobem identyfikacji preferencji, ponieważ wyniki skali rangowej lub porządkowania zawsze można przedstawić w postaci porównania parami, natomiast odwrotna procedura jest możliwa jedynie w przypadku par przechodnich.
3. Zestawienie ocen porównań: proporcje odpowiedzi wskazujących na preferowane marki w danej kolumnie nad marką w wierszu. Macierz quasi-symetryczna ze wskaźnikami powyżej i poniżej przekątnej dopełniającymi się do jedności.
4. Stworzenie tablicy rozkładu zmiennej standaryzowanej  $Z$  odczytanych z tablicy rozkładu normalnego z wartościami dodatnimi dla marek dominujących i ujemnymi dla ich dominowanych odpowiedników poniżej przekątnej.
5. Obliczenie średniej wartości  $Z$  dla każdej kolumny.
6. Centrowanie wartości  $Z$  - przyjęcie za układ odniesienia marki (kolumny) o najniższej średniej.
7. Przedstawienie jednowymiarowej skali na osi graficznej.

## Budowa skali ocen porównawczych w programie *STATISTICA*

W programie *STATISTICA* skala ocen porównawczych Thurstone'a może być zbudowana na podstawie prostej skali rangowej lub skali porównań parami. Znajduje się ona w programie *STATISTICA dla badań marketingowych i rynkowych* w grupie *Analizy* (zob. rys. 1).

<sup>6</sup> J. Bazarnik, T. Grabiński, E. Kąciak, S. Mynarski, A. Sagan, *Badania marketingowe. Metody i oprogramowanie komputerowe*, Fogra Kraków 1991, s. 82-83.



Rys. 13. Model ocen porównawczych.

W przykładzie zastosowano skalę rangową napojów chłodzących, w której respondenci szeregowali 7 typów napojów chłodzących na skali rangowej, gdzie ocena 1 oznaczała najmniej preferowany napój, a 7 – najbardziej preferowany.

	1 Herbata mrozona	2 Kompot	3 Woda	4 Soki	5 Napoje gazowane	6 Typu Cola	7 Kefir
1	7	3	6	5	1	2	4
2	3	2	4	7	1	6	5
3	2	3	4	5	1	7	6
4	7	4	5	1	2	3	6
5	6	4	3	1	2	5	7
6	7	4	3	1	2	6	5
7	3	5	4	7	2	1	6
8	3	7	4	5	2	1	6
9	5	2	3	4	6	7	1
10	7	6	4	5	1	2	3

Rys. 14. Dane wejściowe – ranking napojów.

Dane w postaci skali rangowej są następnie przetwarzane na wyniki porównań parami poszczególnych obiektów.

	1 Herbata mrozona-Kompot	2 Herbata mrozona-Woda	3 Herbata mrozona-Soki
1	1	1	1
2	1	0	0
3	0	0	0
4	1	1	1
5	1	1	1
6	1	1	1
7	0	0	0
8	0	0	0
9	1	1	1
10	1	1	1

Rys. 15. Tabela porównań parami.

Rysunek przedstawia wyniki porównań. Dla przykładu: respondent nr 1 w parze “herbata – kompot” preferuje herbatę mrożoną jako odpowiedni napój chłodzący (herbata ma rangę 7, a kompot otrzymał rangę 3). Dla badanego nr 3 lepszym napojem chłodzącym jest kompot (ranga 3) w porównaniu do herbaty (ranga 2)<sup>7</sup>.

Na podstawie ocen porównawczych tworzona jest tabela proporcji, w jakiej dany napój (w kolumnie) jest preferowany nad inny (w wierszu).

	1 Herbata mrozona	2 Kompot	3 Woda	4 Soki	5 Napoje gazowane	6 Typu Cola	7 Kefir
Herbata mrozona	0	0.3	0.4	0.4	0.1	0.3	0.5
Kompot	0.7	0	0.5	0.5	0.1	0.5	0.7
Woda	0.6	0.5	0	0.6	0.1	0.5	0.7
Soki	0.6	0.5	0.4	0	0.4	0.5	0.5
Napoje gazowane	0.9	0.9	0.9	0.6	0	0.8	0.9
Typu Cola	0.7	0.5	0.5	0.5	0.2	0	0.6
Kefir	0.5	0.3	0.3	0.5	0.1	0.4	0

Rys. 16. Tabela proporcji.

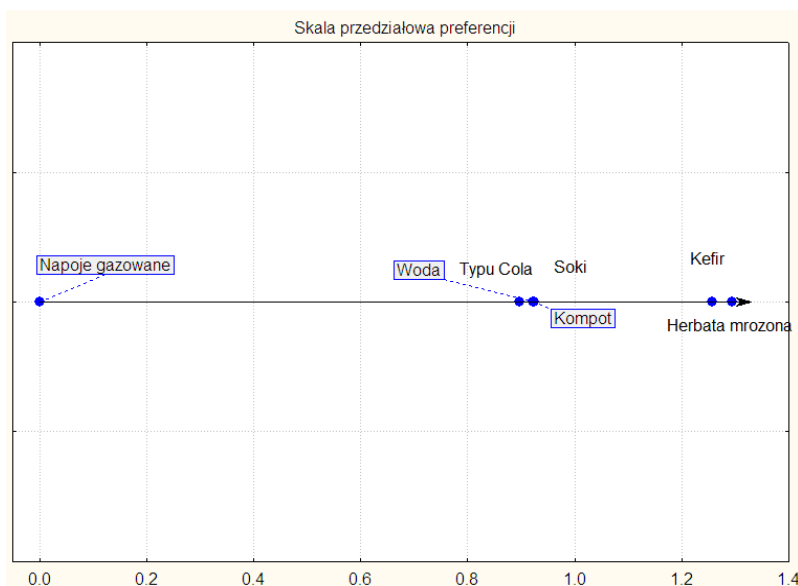
Na rysunku przedstawione są wyniki porównań, w których np. 50% badanych preferuje kompot nad soki, a 90% respondentów uznaje, że lepszym napojem chłodzącym jest woda niegazowana niż napoje gazowane. Na podstawie wyników sumarycznych obliczane są wartości Z rozkładu normalnego, które są następnie uśredniane i centrowane.

	1 Herbata mrozona	2 Kompot	3 Woda	4 Soki	5 Napoje gazowane	6 Typu Cola	7 Kefir
Herbata mrozona	0	-0.5244	-0.25335	-0.25335	-1.28155157	-0.5244005	0
Kompot	0.524400513	0	0	0	-1.28155157	0	0.524401
Woda	0.253347103	0	0	0.253347	-1.28155157	0	0.524401
Soki	0.253347103	0	-0.25335	0	-0.253347103	0	0
Napoje gazowane	1.28155157	1.281552	1.281552	0.253347	0	0.84162123	1.281552
Typu Cola	0.524400513	0	0	0	-0.841621234	0	0.253347
Kefir	0	-0.5244	-0.5244	0	-1.28155157	-0.2533471	0
Średnie Z	0.4052924	0.03325	0.03578	0.036192	-0.888739228	0.0091248	0.3691
Średnie Z*	1.29403163	0.921989	0.924519	0.924932	0	0.89786403	1.257839

Rys. 17. Wartości Z rozkładu normalnego dla proporcji.

W końcowym etapie analizy, na podstawie uśrednionych i centrowanych wartości Z, tworzona jest graficzna przedziałowa jednowymiarowa skala ocen porównawczych.

<sup>7</sup> Skala rangowa umożliwia identyfikację jedynie relacji przechodnich w strukturze preferencji. Skala porównań parami pozwala na uzyskanie informacji zarówno o preferencjach przechodnich jak i nieprzechodnich. Kodowanie zero-jedynkowe skali porównań parami jest wygodnym sposobem kodowania danych rangowych. Ten rodzaj danych wejściowych jest podstawą wielu analiz wielowymiarowych takich jak analiza czynnikowa i modelowanie strukturalne danych rangowych. Zob. A. Sagan, *Modelowanie strukturalne w testowaniu nowego produktu – analiza ocen porównawczych Thurstone’a*, w: red. S. Kaczmarczyk, M. Schultz, *Zastosowania badań marketingowych w procesie tworzenia nowych produktów*, Dom Organizatora Toruń 2008.



Rys. 18. Przedziałowa jednowymiarowa skala ocen porównawczych.

Na podstawie uzyskanej skali można nie tylko szeregować preferencje respondentów, ale również określać „odległości” między szeregowanymi napojami. Z rysunku wynika, że w strukturze napojów chłodzących można wyodrębnić 3 grupy preferowanych napojów. Do pierwszej należy kefir i herbata mrożona (najbardziej preferowane), do drugiej o podobnych preferencjach zaliczyć można: soki, kompot, napoje typu cola i wodę, a zdecydowanie najsłabiej preferowanym typem napojów są napoje gazowane. Relacje powyższe można przedstawić w macierzy odległości między napojami.

	1	2	3	4	5	6	7
	Herbata mrożona	Kompot	Woda	Soki	Napoje gazowane	Typu Cola	Kefir
Herbata mrożona	0	0.372042	0.369513	0.3691	1.29403163	0.3961676	0.036192
Kompot	0.372042322	0	0.002529	0.002942	0.921989306	0.02412527	0.33585
Woda	0.36951285	0.002529	0	0.000413	0.924518778	0.02665475	0.33332
Soki	0.369099956	0.002942	0.000413	0	0.924931672	0.02706764	0.332908
Napoje gazowane	1.29403163	0.921989	0.924519	0.924932	0	0.89786403	1.257839
Typu Cola	0.396167597	0.024125	0.026655	0.027068	0.897864031	0	0.359975
Kefir	0.0361924433	0.33585	0.33332	0.332908	1.25783918	0.35997515	0

Rys. 19. Macierz odległości.

Obliczona macierz odległości pozwala na reprezentację układów preferencji za pomocą skalowania wielowymiarowego.

## Analiza PROFIT

Analiza PROFIT (*PRO*perty *FIT*ting) jest rodzajem „zewnętrznej” mapy preferencji, w której na podstawie mapy percepcji uzyskanej za pomocą klasycznego niemetrycznego

skalowania wielowymiarowego (MDS) i informacji o pozycji marek na stworzonej mapie, wprowadzone są „z zewnątrz” dane o preferencjach analizowanych marek z punktu widzenia charakteryzujących je cech. Metoda została opracowana przez J. D. Carroll i J. J. Changa. Łączy ona wyniki skalowania wielowymiarowego percepcji i analizy regresji wielorakiej<sup>8</sup>.

W pierwszym etapie analizy PROFIT wykorzystywane jest skalowanie wielowymiarowe służące do budowy map percepcyjnych. Podstawą budowy takiej mapy jest macierz reprezentująca relacje podobieństwa /bliskości między analizowanymi markami produktów. Najczęściej spotyka się dwa typy macierzy danych: 1/ odległości euklidesowych, Manhattan lub innych, zbudowane na podstawie ocen porównywanych marek z punktu widzenia przyjętych atrybutów, dających syntetyczną ocenę niepodobieństw (odmienności) między markami; 2/ porównań parami każdej kombinacji pary marek z każdą bez wyróżniania ich cech, dającą ogólny obraz podobieństw między markami wyrażony w macierzy rang (para najbardziej podobna ma najniższą rangę, a para najmniej podobna – najwyższą). W pierwszym podejściu stosowane są procedury metrycznego skalowania wielowymiarowego (dane wejściowe są metryczne i dane wyjściowe z analizy są metryczne), a w drugim stosuje się niemetryczne skalowanie wielowymiarowe (dane wejściowe są niemetryczne, a dane wyjściowe są metryczne). W metodzie PROFIT stosuje się najczęściej niemetryczne skalowanie wielowymiarowe na podstawie skal porównań parami lub porównania w triadach (bez uwzględnienia cech), chociaż stosować można także skalowanie metryczne. Kolejnym etapem analizy jest określenie liczby wymiarów, która ze względów praktycznych wynosi od 2 do 3. Algorytm skalowania wielowymiarowego na podstawie początkowej konfiguracji marek w przestrzeni 2- lub 3-wymiarowej, wyznaczonej na podstawie odległości lub szeregu rangowego w macierzy wejściowej, poszukuje takich współrzędnych marek w tej przestrzeni, które w sposób optymalny odtwarzają odległości zamieszczone w macierzy danych wejściowych. Jakość dopasowania mierzy się za pomocą współczynnika STRESS (standaryzowanej sumy kwadratów reszt między odległościami wejściowymi a odtworzonymi przez algorytm skalowania wielowymiarowego). Efektem finalnym analizy są współrzędne marek w 2- lub 3-wymiarowym układzie współrzędnych.

W drugim etapie analizy PROFIT wykorzystywane są informacje o preferencjach analizowanych marek z punktu widzenia przyjętych w badaniach cech. Respondenci oceniają preferencje marek ze względu na przyjęte cechy i na tej podstawie buduje się uśrednione oceny preferencji dla poszczególnych marek i cech w przekroju wszystkich badanych lub poszczególnych podgrup. Po stworzeniu tabeli średnich ocen buduje się modele regresyjne (regresji wielorakiej), w których zmiennymi zależnymi są oceny marek ze względu na poszczególne cechy (w analizie występuje tyle równań regresji, ile jest badanych cech produktu), a zmiennymi objaśniającymi preferencje marek są ich współrzędne na mapie percepcyjnej (dla modelu 2-wymiarowego są więc dwie zmienne niezależne, a dla modelu 3-wymiarowego występują trzy takie zmienne). Standaryzowane współczynniki regresji (beta) dla poszczególnych wymiarów wyznaczają punkt na mapie percepcji określający współrzędne danego atrybutu i tym samym preferencje marek ze względu na daną cechę.

<sup>8</sup> S. P. Borgatti, PROFIT, <http://www.analytictech.com/borgatti/profit.htm> (12.08.2009).

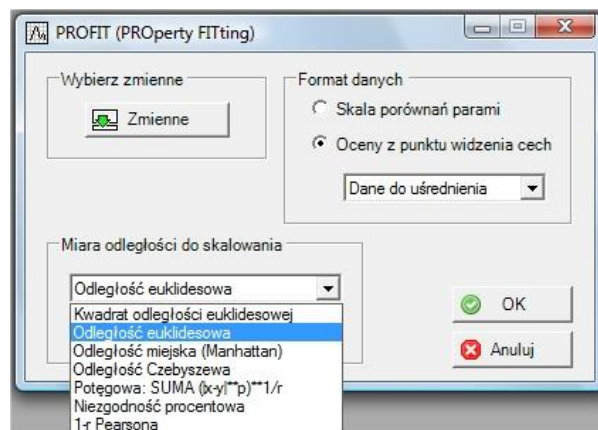
Współczynniki te są również interpretowane w kategoriach kwadratów cosinusów kąta nachylenia wektorów danych cech względem osi układu współrzędnych. Pozwalają na ocenę wkładu danej osi głównej w wyjaśnienie zmienności danej cechy.

Projekcja punktów reprezentujących poszczególne marki na wektory cech pozwala na określenie położenia marek ze względu na intensywność występowania tych cech w danych markach i tym samym na ustalenie szeregu preferencyjnego.

Informacja o stopniu, w jakim położenie marek na mapie percepcji wyjaśnia oceny preferencyjne z punktu widzenia określonych cech, jest związana ze współczynnikiem determinacji  $R^2$ . Im wyższy współczynnik determinacji, w tym większym stopniu zestaw zmiennych niezależnych (współrzędnych na mapie percepcji) wyjaśnia oceny marek ze względu na daną cechę (zmienną zależną).

## Analiza PROFIT w programie *STATISTICA*

Analiza PROFIT znajduje się również w programie *STATISTICA dla badań marketingowych i rynkowych*. W pierwszym kroku analizy należy określić charakter danych wejściowych do skalowania wielowymiarowego (skala porównań parami czy oceny marek z punktu widzenia cech). Jeżeli wybierzemy pierwsze rozwiązanie, program na podstawie wyników porządkowania par zbuduje macierz rang określających subiektywne ogólne podobieństwa między parami marek (bez uwzględnienia cech). W przypadku wyboru skali ocen program wybierze uśrednione dane lub dokona uśrednienia ocen respondentów i obliczy wybraną macierz odległości (np. euklidesowych).



Rys. 20. Okno analizy PROFIT.

Na rysunku przedstawione są uśrednione oceny (mediany) ośrodków narciarskich z punktu widzenia wyróżnionych charakterystyk.



Dane: SKI (4 zm. * 11 prz.)				
	1	2	3	4
	Śnieg	Trasy	Infrastruktura	Koszt
Szczyrk	4.000	4.000	3.500	3.000
Szklarska Poręba	4.000	3.500	3.500	3.000
Karpacz	2.000	3.000	4.000	4.000
Krynica	4.000	4.000	2.000	3.000
Bukowina Tatrzańska	3.000	3.000	1.000	4.000
Piwniczna	3.000	3.000	1.500	4.000
Korbielów	2.000	4.000	2.000	3.000
Szczawnica	4.000	2.000	3.000	3.500
Zakopane	3.000	3.000	4.000	1.000
Białka Tatrzańska	1.000	1.000	2.000	3.000
Zieleniec	1.000	1.000	2.000	3.000

Rys. 21. Dane wejściowe do analizy PROFIT.

Na podstawie danych wejściowych dokonywane jest obliczenie miary podobieństwa między analizowanymi obiektami.

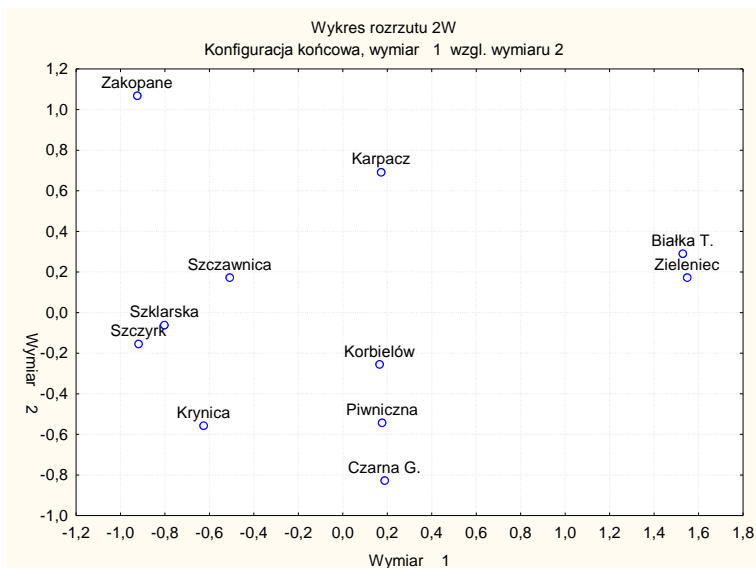
Przykładowa macierz odległości euklidesowych obliczonych na podstawie uśrednionych ocen wybranych ośrodków narciarskich w Polsce jest przedstawiona na rysunku.

Nr przypadku	Kwadratowa odl. euklidesowa (SKI)										
	Szczyrk	Szklarska Poręba	Karpacz	Krynica	Czarna Góra	Piwniczna	Korbielów	Szczawnica	Zakopane	Białka Tatrzańska	Zieleniec
Szczyrk	0.0	0.3	6.3	2.3	9.3	7.0	6.3	4.5	6.3	20.3	20.3
Szklarska Poręba	0.3	0.0	5.5	2.5	8.5	6.3	6.5	2.8	5.5	17.5	17.5
Karpacz	6.3	5.5	0.0	10.0	10.0	7.3	6.0	6.3	10.0	10.0	10.0
Krynica	2.3	2.5	10.0	0.0	4.0	3.3	4.0	5.3	10.0	18.0	18.0
Czarna Góra	9.3	8.5	10.0	4.0	0.0	0.3	4.0	6.3	18.0	10.0	10.0
Piwniczna	7.0	6.3	7.3	3.3	0.3	0.0	3.3	4.5	15.3	9.3	9.3
Korbielów	6.3	6.5	6.0	4.0	4.0	3.3	0.0	9.3	10.0	10.0	10.0
Szczawnica	4.5	2.8	6.3	5.3	6.3	4.5	9.3	0.0	9.3	11.3	11.3
Zakopane	6.3	5.5	10.0	10.0	18.0	15.3	10.0	9.3	0.0	16.0	16.0
Białka Tatrzańska	20.3	17.5	10.0	18.0	10.0	9.3	10.0	11.3	16.0	0.0	0.0
Zieleniec	20.3	17.5	10.0	18.0	10.0	9.3	10.0	11.3	16.0	0.0	0.0

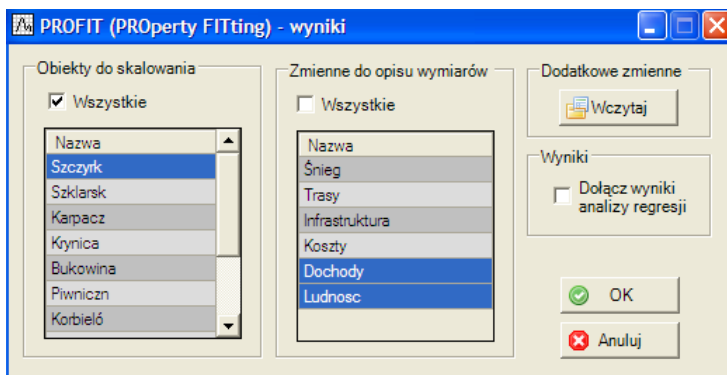
Rys. 22. Macierz odległości.

Macierz odległości euklidesowych kwadratowych jest wprowadzona następnie do programu skalowania wielowymiarowego, który określa położenie marek w układzie współrzędnych o ustalonej liczbie wymiarów.

Wyniki analizy w postaci współrzędnych marek są podstawą obliczenia modeli regresji wielorakiej. W tym celu użytkownik jest proszony o wskazanie listy zmiennych opisujących marki produktu. Lista ta może pochodzić tego samego zbioru cech, stanowić nowy zbiór cech opisujących obiekty lub ich kombinację.



Rys. 23. Wyniki skalowania wielowymiarowego.



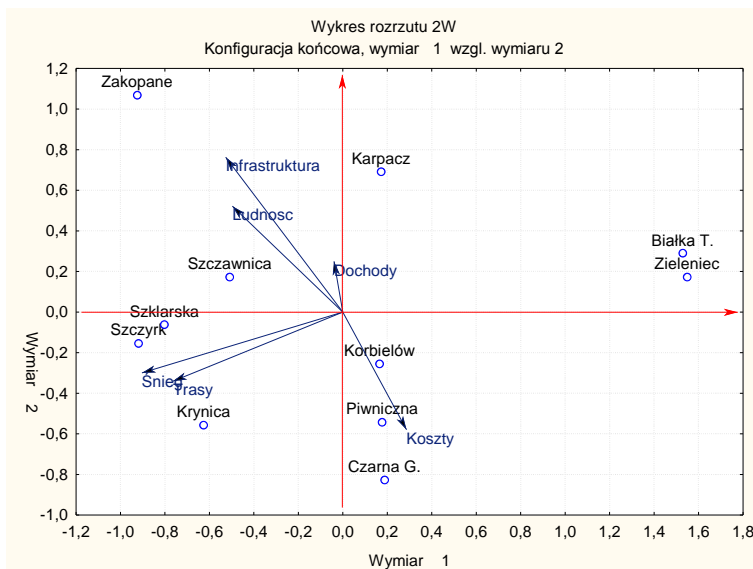
Rys. 24. Okno PROFIT – wyniki.

Program wykona tyle analiz regresji wielorakiej, ile wskazano zmiennych do analizy. Przykładowo podane są wyniki analizy regresji dla zmiennej zależnej „śnieg”. Zmienna ta jest objaśniana z wykorzystaniem zmiennych objaśniających będących współrzędnymi konfiguracji punktów w 2 wymiarach.

Podsumowanie regresji zmiennej zależnej: Śnieg (Arkusz73)						
R= .95170714 R^2= .90574648 Skoryg. R2= .88218310						
F(2,8)=38.439 p<.00008 Błąd std. estymacji: .40082						
N=11	BETA	Bł. std. BETA	B	Bł. std. B	t(8)	poziom p
W. wolny			2.81818	0.120853	23.31913	0.000000
WYM. 1	-0.903266	0.108543	-1.19014	0.143017	-8.32170	0.000033
WYM. 2	-0.299762	0.108543	-0.62416	0.226008	-2.76167	0.024611

Rys. 25. Wyniki regresji.

Po wykonaniu analizy regresji współrzędne współczynników kierunkowych są nałożone na zbudowaną wcześniej mapę percepcji.



Rys. 26. Mapa percepcji z osiami opisującymi wymiary.

Z wykresu wynika, że najbardziej preferowanym ośrodkiem z punktu widzenia infrastruktury jest Zakopane. Krynica i Szczyrk mają względnie wysokie oceny ze względu na warunki śniegowe i przygotowanie tras, a Piwniczna, Korbielów i Czarna Góra są dogodnym miejscem odnośnie kosztów pobytu.

## Podsumowanie i wnioski

Przedstawione w artykule wybrane kompozycyjne i dekompozycyjne metody analizy są często spotykanymi metodami analizy preferencji konsumentów. Istniejące w programie *STATISTICA* narzędzia programowania analiz danych pozwalają na łatwe wykorzystanie szerokiej gamy metod, które nie stanowią podstawowego wyposażenia programu, lecz mogą być implementowane z wykorzystaniem standardowych narzędzi analitycznych.

## Literatura

1. *Analiza danych marketingowych. Problemy, metody, przykłady*, red. A. Stanimir, AE Wrocław 2006.
2. J. Bazarnik, T. Grabiński, E. Kąciak, S. Mynarski, A. Sagan, *Badania marketingowe. Metody i oprogramowanie komputerowe*, Fogra Kraków 1991, s. 82-83.



3. A. Bąk, *Analiza conjoint*, w:., M. Walesiak, E. Gatnar, *Statystyczna analiza danych z wykorzystaniem programu R*, PWN Warszawa 2009.
4. S. P. Borgatti, *PROFIT*, <http://www.analytictech.com/borgatti/profit.html>.
5. C.H. Coombs, R.M. Dawes, A. Tversky, *Wprowadzenie do psychologii matematycznej*, PWN Warszawa 1977, s. 71-83.
6. H. Gintis, *The individual in economic theory: a research agenda*, Dept. of Economics, University of Massachusetts, Amherst, 1998.
7. M. Walesiak, A. Bąk, *Conjoint analysis w badaniach marketingowych*, AE Wrocław 2000.
8. A. Zaborski, *Skalowanie wielowymiarowe w badaniach marketingowych*, AE Wrocław 2001.
9. *Zastosowania badań marketingowych w procesie tworzenia nowych produktów*, red. S. Kaczmarczyk, M. Schultz, Dom Organizatora Toruń 2008.