



JEDEN OBRAZ UKAZUJE WIĘCEJ NIŻ 10 LICZB, CZYLI JAK BUDOWAĆ MAPY ZADOWOLENIA KLIENTA Z WYKORZYSTANIEM PROGRAMU *STATISTICA*

Adam Sagan

Akademia Ekonomiczna w Krakowie, Katedra Analizy Rynku i Badań Marketingowych

Analiza zadowolenia klienta

Analiza zjawisk zadowolenia konsumentów, postrzeganej jakości i wartości produktów oraz lojalności wobec marki jest jednym z najsilniej rozwijających się kierunków analiz konsumentów. Ich pomiar związany jest z wykorzystywaniem różnorodnych wskaźników zadowolenia, rang i szacunkowych skal ocen.

W badaniach nad zadowoleniem z produktu można wymienić kilka kierunków, wśród których wyróżnić należy: analizę luk w percepcji jakości produktu, modele kompromisów oraz względne skale porównawcze. Pierwszym nurtem w analizie zadowolenia jest wykorzystywanie skal typu Likerta mierzących absolutny poziom satysfakcji klientów lub wielkość luki między percepcją postrzeganej jakości produktu a oczekiwaniami związanymi z pożądanym poziomem jakości. Najbardziej znanym narzędziem pomiaru jest skala SERVQUAL (*Service Quality*) stosowana w pomiarze postrzeganej jakości usług. Twórcy tej skali pracowali model luk jakości usług, który był również podstawą budowy skali SERVQUAL.

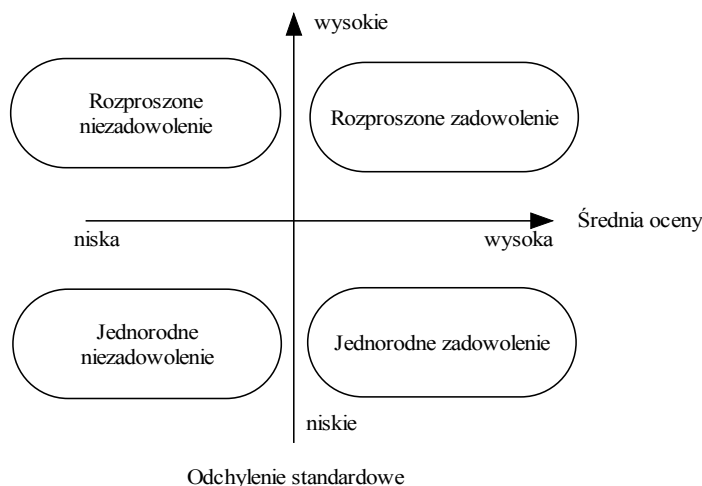
Drugi typ skali służącej do pomiaru zadowolenia klienta to skala zwana SIMALTO (*Simultaneous Multi Attribute Level Trade-Off*), którą można określić jako skalę równoczesnych kompromisów wieloatrybutowych, służy do określenia stopnia kompromisu podejmowanego przez konsumenta przy wyborze różnych kombinacji cech produktu. Skala SIMALTO jest prezentowana w postaci zestawów cech produktu i ich poziomów. Liczba cech w tej metodzie może wynosić nawet ponad 30, a cechy mogą mieć po 8-10 poziomów. Zwykle najlepsze rezultaty przynosi badanie z wykorzystaniem 20-25 atrybutów. Badania z wykorzystaniem skali polegają na ocenie ważności cech produktu. Ocena polega na wyróżnieniu 4 podstawowych charakterystyk poziomów cech: 1/ oczekiwanych przez klienta od „idealnej” firmy, 2/ postrzeganych przez klienta w kontekście rzeczywistego produktu, 3/ poziomu nie do przyjęcia dla klienta, 4/ cech najważniejszych przy wyborze danego produktu (ok. 25% cech).

Do trzeciego nurtu należą skale porównawcze, w których wykorzystywane są różnego typu względne skale satysfakcji, jak np. skale ważności-realizacji (*importance-performance*), względnej realizacji-wpływu (*relative performance-impact*) czy wkładu do wartości i satysfakcji (*contributes to value perception-satisfaction*). Skala ważności-realizacji wpływa w sposób bezpośredni z wieloatrybutowego modelu oczekiwanej wartości preferencji, w którym preferencje wobec produktu są funkcją ważonych ocen atrybutów. Sytuacja pomiaru jest przedstawiona w tabeli 1.

Tabela 1. Model oczekiwanej wartości.

Cecha	Ważność cechy (skala o sumie stałej) = 1,00	Ocena – realizacja cechy w produkcie P1 (skala 1-10)	Ważność x ocena	Ocena – realizacja cechy w produkcie P2 (skala 1-10)	Ważność x ocena
A	0,3	2	0,6	3	0,9
B	0,5	6	3,0	2	1,0
C	0,2	7	1,4	8	1,6
		Suma	5,0	Suma	3,5

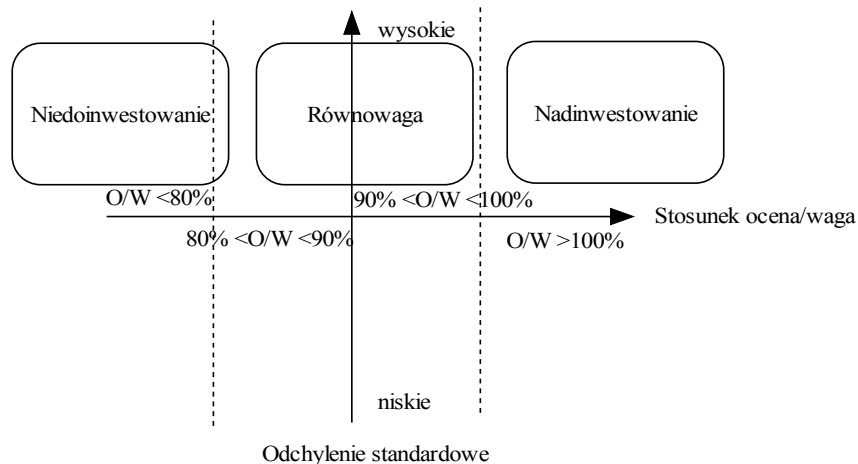
Z tabeli wynika, że oczekiwana wartość satysfakcji z produktu P1 jest wyższa niż z produktu P2, pomimo że P2 ma wyższą ocenę ze względu na cechy A i C, które są jednak tłumione przez niższą ocenę ze względu na ważną cechę B. Zestawienie graficzne cech kształtujących satysfakcję z produktu jest dokonywane na podstawie porównania przeciętnych wartości ich ocen oraz odchyłeń standardowych. Na ich podstawie można uzyskać cztery podstawowe typy zadowolenia klientów.



Pewnym problemem jest określenie punktu przecięcia się osi. Najczęściej jest to średnia dla danego sektora lub wynik najgroźniejszego konkurenta.

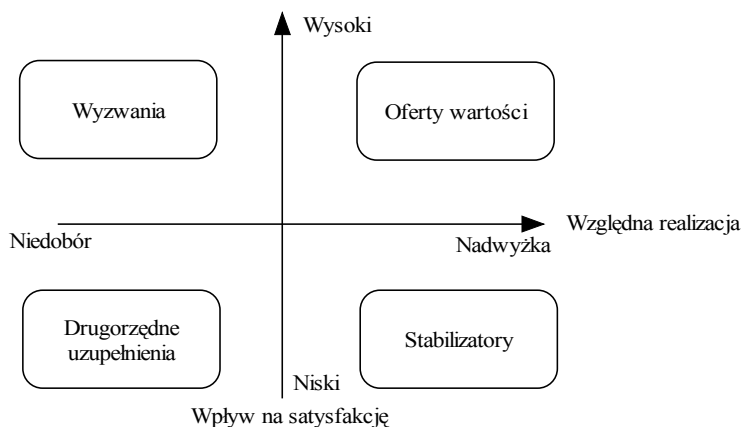
W przypadku gdy wagi i oceny cech produktów są dokonywane w tej samej skali (np. 1-4), można określić stopień akceptacji produktu poprzez stosunek ocena/waga.

Porównanie tego stosunku do odchyłeń standardowych poszczególnych cech pozwala na wyodrębnienie 4 stref relacji O/W.



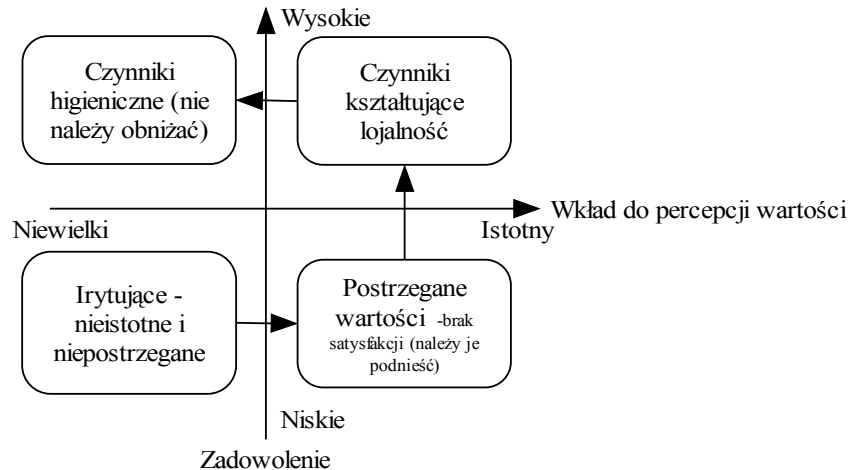
Pierwsza strefa wyznacza obszar, w którym relacja O/W jest większa od 1. Oznacza to, że klienci wyżej oceniają wartość danej cechy w produkcie niż jej ważność, i świadczy o „przeinwestowaniu” w daną cechę. Strefa druga określa sytuację równowagi między oceną a ważnością cechy, w której poziom zadowolenia z ważnych cech produktu jest dobry. Strefa trzecia wyznacza sytuację deprywacji cech. Ich poziom realizacji jest niewystarczający w stosunku do ważności.

Wskaźnik względnej ważności i wpływu zestawia względne oceny realizacji cech (korzyści w produkcie) z ich odczuwanym przez klientów wpływem na satysfakcję z produktu.



W pomiarze realizacji wykorzystywane są względne skale porównawcze, w których produktem odniesienia, do którego cech porównywane są dane cechy produktów, jest lider w sektorze lub tzw. produkt idealny – wyobraźniowy, który tworzą konsumenci w swojej świadomości. Na tej podstawie można wyodrębnić 4 podstawowe obszary cech produktu, które stanowią: podstawowe oferty wartości kształtujące przewagi konkurencyjne, główne wyzwania związane z koniecznością poprawy wizerunku produktu, drugorzędne korzyści niestanowiące o przewagach konkurencyjnych, ale mogące być uzupełnieniem podstawowej oferty oraz stabilizatory ogólnego wizerunku produktu.

Wskaźniki wkładu i zadowolenia określają relacje między subiektywnym wkładem danej cechy do postrzeganej przez konsumenta wartości produktu a stopniem zadowolenia z występowania danej cechy w produkcie.



Czynniki bazowe (higieniczne) wynikają ze spełnienia podstawowych oczekiwań konsumenta. Stanowią o rdzeniu produktu i są uznawane za konieczne w wyposażeniu produktu. Ich wkład do postrzeganej wartości produktu jest niewielki, lecz silnie wpływają na ogólne zadowolenie z produktu. Ich obniżenie prowadzi do rezygnacji z zakupu i odejścia z rynku. Czynniki kształtujące lojalność tworzą wartość produktu i są elementami kształtowania przewagi konkurencyjnej i plasowania produktu. Trzecia grupa cech to cechy, które mają silny wkład w postrzeganą wartość, lecz stopień zadowolenia z nich jest niski. Dotyczą one najczęściej kształtujących się potrzeb i nowych cech produktów je zaspakajających, których postrzegana jakość jest jeszcze niska. Ostatnia grupa odnosi się do nieistotnych lub ukrytych korzyści i cech, które nie są postrzegane jako istotne dla zadowolenia z produktu. Strzałki na rysunku wskazują na „cykl życia satysfakcji”, w którym punktem wyjścia są cechy i korzyści niepostrzegane, będące potencjalnymi możliwościami budowy satysfakcji z produktu. Wartość dodana do tych cech powoduje ich przesunięcie do czynników postrzeganej wartości budujących lojalność klientów. Wzrost walki konkurencyjnej na tym tle prowadzi do dalszego ich przesunięcia w strefę czynników higienicznych będących w wyposażeniu wszystkich konkurujących między sobą produktów.

Budowa map zadowolenia w programie *STATISTICA*

Przedstawione w poprzednim punkcie mapy mogą być budowane na podstawie wskaźników absolutnych (wartości przeciętne – średnie arytmetyczne lub mediany - dla danych cech i ich odchylenia standardowe lub rozstępy międzykwartylowe) lub względnych (w odniesieniu do cech produktu najgroźniejszego konkurenta lub cech produktu odniesienia – najczęściej produktu idealnego). Prezentowane mapy są najczęściej zestawieniami wartości średnich arytmetycznych porównywanych cech i przyjęcia określonych schematów ich klasyfikacji na osiach układu współrzędnych. Pomimo niewątpliwych zalet, jakimi są prostota ich tworzenia i klarowność interpretacji, mają także swoje wady. Należą do nich subiektywizm w ocenie krytycznych wartości skali (np. wyróżnienie poziomu cechy mającej niewielki bądź duży wkład do percepcji wartości), pomijanie współzależności między cechami i nieobserwowalnych bezpośrednio, ukrytych układów cech lub korzyści.

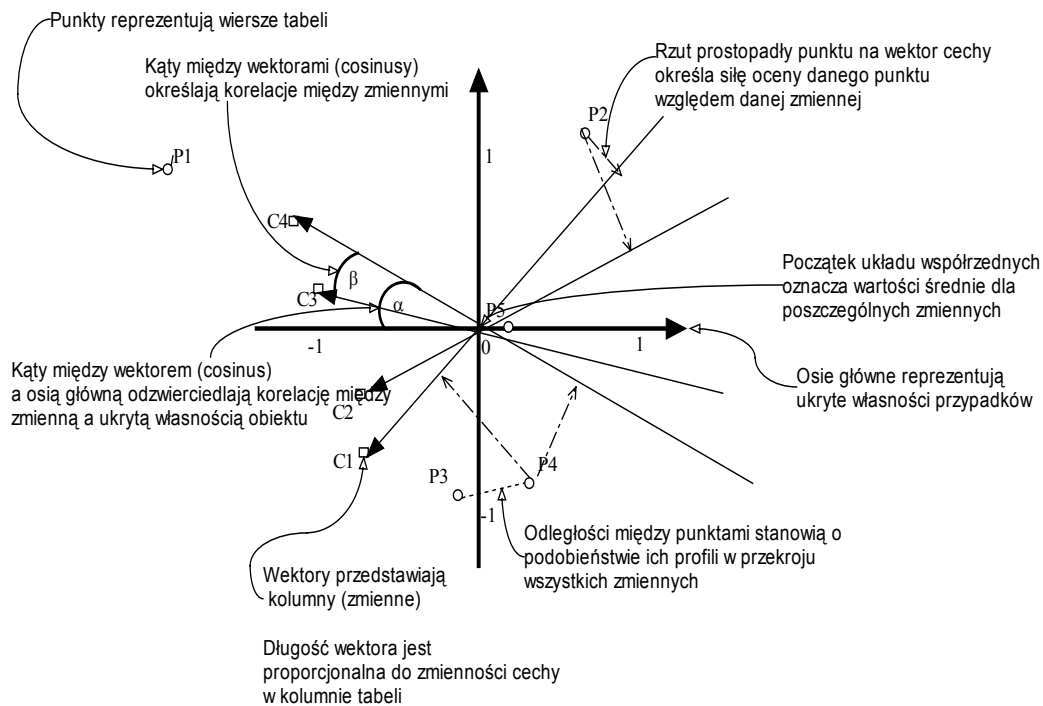
W celu bardziej pogłębionej identyfikacji czynników wpływających na zadowolenie z produktu wykorzystywane są metody analizy wielowymiarowej, których graficzna postać pozwala na pełniejsze zrozumienie występujących zależności. Mapy te pozwalają określić zachodzące zależności jednocześnie między porównywanymi produktami i ich cechami oraz między różnymi grupami respondentów. W języku analizy wielowymiarowej noszą one nazwę biplotów, ponieważ umożliwiają graficzne przedstawienie zarówno elementów wierszy, jak i kolumn tabeli danych na tym samym wykresie. Graficzna prezentacja tabeli (macierzy) danych pozwala na pełniejsze jej zrozumienie i ułatwia interpretację analizowanych zależności. Do przestrzennej reprezentacji badanych zależności służy wiele metod statystycznych znajdujących się w programie *STATISTICA*. Do tego celu najczęściej wykorzystywane są znajdujące się w module „Wielowymiarowe techniki eksploracyjne” metody analizy czynnikowej i głównych składowych, analizy korespondencji oraz skalowania wielowymiarowego. Każda z nich pozwala na graficzną prezentację uzyskanych wyników i na ukazanie możliwie największej ilości informacji z analizowanych danych.

Tabela 2. Tabela danych (PxC).

	1 C1	2 C2	3 C3	4 C4		
P1	5	5	6	6	Zmienna	Korelacje (Biplot) Oznaczone wsp. korel. N=5 (Braki danych usunięte)
P2	1	1	2	2		
P3	5	4	3	2		
P4	4	3	2	1		
P5	4	3	3	3		
ODCH.STD	1,64	1,48	1,64	1,92	C1	1,00 0,94 0,57 0,38
ŚREDNIA	3,8	3,2	3,2	2,8	C2	0,94 1,00 0,80 0,63
					C3	0,57 0,80 1,00 0,96
					C4	0,38 0,63 0,96 1,00

Tabela przedstawia średnie wartości ocen zadowolenia klientów z różnych cech produktów oraz współczynniki korelacji między zmiennymi. Cechy produktów (zmienne w kolumnach) są reprezentowane przez symbole C1...C4, a same produkty (przypadki w wierszach) przez P1...P5. Wartości w tabeli są średnimi wartościami ocen zadowolenia klientów z realizacją cech w produktach w skali 1-6. Celem biplotu jest graficzna

prezentacja tabeli danych (zarówno wierszy, jak i kolumn) w postaci graficznej mapy jako punktów w układzie współrzędnych. Jest ona przedstawiona na rysunku poniżej.



Ogólne zasady interpretacji biplotów są przedstawione na rysunku.

1. Przypadki (elementy w wierszach tabeli) są przedstawiane za pomocą punktów (P), a zmienne (elementy w kolumnach) reprezentowane są przez strzałki (wektory) kończące się w reprezentowanym punkcie (C) (lub przez niego przechodzące).
2. Długość wektora jest odzwierciedleniem odchylenia standardowego danej zmiennej w kolumnie i określa znaczenie tej zmiennej z punktu widzenia jej tzw. mocy dyskryminacyjnej, tj. zdolności zmiennej do różnicowania elementów w wierszach (np. odpowiedzi respondentów lub różnicowania ocen zadowolenia z produktów). Z rysunku wynika, że największe znaczenie w tym względzie ma cecha C4 (najdłuższy wektor), a najmniejsze – cecha C2 (najkrótszy wektor). Potwierdzają to wartości odchyłeń standardowych tych zmiennych podane w tabeli.
3. Kąty między wektorami wskazują na skorelowanie zmiennych. Wszystkie strzałki są zorientowane w tym samym kierunku, świadczy to o dodatnim skorelowaniu wszystkich zmiennych. Wskazuje na to również zamieszczona macierz korelacji. Najsilniej skorelowane są pary zmiennych C3, C4 oraz C1, C2, w przypadku których kąty są najmniejsze (odpowiadające im cosinusy kątów zaś największe). Ważniejsza korelacja występuje między zmiennymi C1, C4 (kąąt między wektorami jest największy). Gdyby



wektory zmiennych były skierowane pod kątem 90° , oznaczałoby to brak korelacji między tymi zmiennymi.

Istnieje ścisła zależność między współczynnikami korelacji między zmiennymi a kątami między wektorami reprezentującymi te zmienne. Dla przykładu macierzy korelacji podanej poniżej przekątnej odpowiadają kąty podane powyżej przekątnej (współczynniki korelacji stanowią cosinusy tych kątów)

1/0°	30°	80°
0,87	1/0°	50°
0,17	0,64	1/0°

4. Odległości euklidesowe (w linii prostej) między punktami reprezentującymi wiersze tabeli (przypadki) wskazują na podobieństwo reakcji w przekroju wszystkich zmiennych. Dla przykładu najbardziej podobnymi profilami reakcji w przekroju wszystkich zmiennych charakteryzują się produkty P3 (5,4,3,2) i P4 (4,3,2,1).
5. Odległość od początku układu współrzędnych wskazuje na podobieństwo danego wiersza do wartości średnich poszczególnych zmiennych. Początek układu współrzędnych jest tak określony, że stanowi on średnią wartość dla każdej zmiennej. Z wykresu wynika, że wiersz P5 (4,3,3,3) jest najbardziej zbliżony do wartości średnich dla poszczególnych zmiennych (3.8, 3.2, 3.2, 2.8).
6. Rzuty prostopadłe punktów reprezentujących wiersze na wektory zmiennych wskazują na siłę wartości ocen elementów w wierszach w przekroju każdej zmiennej. Z punktu widzenia zmiennej C1 najwyższe oceny mają punkty P1 i P3, a najniższą – punkt P2. Ze względu na cechę C4 najwyższa wartość należy do punktu P1, a najniższa – do punktu P4. Rzuty prostopadłe elementów wierszy pozwalają więc na uporządkowanie porównywanych obiektów z punktu widzenia ich siły ocen w przekroju poszczególnych zmiennych. Początek układu współrzędnych dzieli wartości rzutów prostopadłych poszczególnych przypadków na wartości większe od średniej dla wszystkich przypadków (w kierunku grota wektora) i wartości mniejsze od średniej (w kierunku przeciwnym).
7. Kąty nachylenia wektorów zmiennych do osi głównych związane są z identyfikacją ukrytych własności porównywanych obiektów. Na rysunku widać, że wszystkie kąty nachylenia dla pierwszej – poziomej osi są niewielkie w porównaniu z kątami nachylenia do osi pionowej oraz wektory cech są ukierunkowane względem tylko jednego bieguna osi głównej. Świadczy to o tym, że wszystkie korelacje z pierwszą osią są wyższe niż z drugą, oraz że wszystkie korelacje z tą osią są jednakowe (tu: dodatnie, bowiem zorientowanie osi głównej jest umowne). Można wysunąć więc wniosek, że wszystkie cechy, ze względu na które oceniane były produkty, można sprowadzić do jednej, ukrytej własności (jednego wymiaru). Najwyższe korelacje z tą osią mają zmienne C2 i C3, stąd ich nazwy mogą być pomocne w nazwaniu tego ukrytego wymiaru. Reprezentacja wierszy i kolumn tabeli danych w układzie dwuwymiarowym jest



najwygodniejszym sposobem prezentacji biplotu. Można jednak dokonać podobnej prezentacji, wykorzystując większą liczbę osi głównych (wymiarów).¹

W *STATISTICA* procedurami służącymi do jednoczesnej reprezentacji tabeli danych w postaci biplotów jest analiza głównych składowych oraz analiza korespondencji. Są one dość zbliżonymi metodami analizy wielowymiarowej opartej na tzw. dekompozycji wartości osobliwej (*singular value decomposition*), w której macierz danych wejściowych do analizy (standaryzowanych jak w macierzy korelacji, centrowanych jak w macierzy kowariancji lub reszt standaryzowanych jak w tabeli kontyngencji) jest dekomponowana na tzw. wektory osobliwe i wartości osobliwe. Podstawowe różnice między metodami związane są z postacią danych wejściowych oraz rodzajem dekomponowanej macierzy danych. Pokrewność tych metod podkreśla również fakt, że analiza korespondencji jest często nazywana uogólnioną analizą głównych składowych dla danych niemetrycznych. Podstawowe podobieństwa i różnice między tymi metodami są przedstawione w tabeli 3.

Tabela 3. Porównanie analizy głównych składowych i analizy korespondencji.

Kryterium	Analiza głównych składowych	Analiza korespondencji
Rodzaj gromadzonych danych	Oceny na skalach, mierniki, dane metryczne	Rangi, oceny, odpowiedzi na pytania zamknięte, wybory cech, dane niemetryczne
Tabela danych wejściowych	Macierz kowariancji lub korelacji	Tabela kontyngencji (reszty standaryzowane), dowolna tabela danych nieujemnych
Interpretacja wymiaru	Zakres wyjaśnionej wariancji w zbiorze danych	Zakres wyjaśnionej bezwładności (wartości statystyki χ^2/N)
Rola wierszy i kolumn (zmiennych i przypadków) w analizie	Aktywna i pasywna	Aktywna i pasywna
Korelacja między zmienną a osią główną	Ładunek czynnikowy, kwadrat cosinusa	Kwadrat cosinusa

W przypadku analizy głównych składowych interpretacja uzyskanej mapy zależy od charakteru macierzy danych wejściowych oraz roli zmiennych i przypadków w analizie.

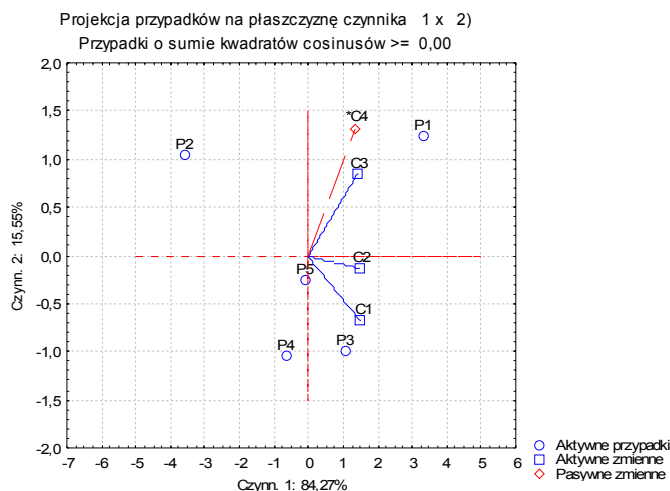
Jeżeli macierzą wejściową jest macierz kowariancji wówczas interpretacja mapy jest tożsama z interpretacją dokonaną w poprzednim przykładzie. Interpretując uzyskany wykres, należy zwrócić uwagę na dwie dodatkowe informacje podane na wykresie.

¹ Nazwa biplot pochodzi od możliwości jednoczesnego przedstawienia dwóch źródeł zmienności danych (w wierszach i kolumnach) w wielowymiarowym układzie współrzędnych, a nie od liczby osi głównych, których może być więcej niż 2. Podejście to najsilniej związane jest z nazwiskiem K. G. Gabriela, stąd mówi się często o tzw. biplotach Gabriela.



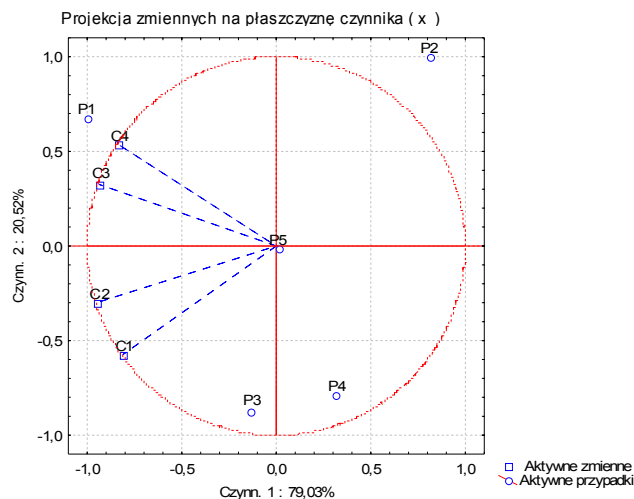
Po pierwsze na mapie występują tzw. zmienne aktywne i pasywne oraz przypadki aktywne (mogą również występować przypadki pasywne). Podział na zmienne (lub przypadki) aktywne i pasywne (dodatkowe) wynika z ich roli w budowaniu osi głównych. W analizie głównych składowych osie główne (tzw. główne składowe) są obliczane jako nowe zmienne, które stanowią liniową kombinację zmiennych biorących udział w analizie maksymalizując ilość wyjaśnianej całkowitej ich wariancji. Zmienne, które biorą udział w tworzeniu tych głównych składowych, noszą nazwę zmiennych aktywnych. One „budują” owe osie czynnikowe. Zmienne pasywne natomiast są jedynie „lokowane” w przestrzeni osi głównych zbudowanych przez zmienne aktywne i nie biorą udziału w ich tworzeniu (obliczane są jedynie współrzędne tych zmiennych w układzie osi głównych). Na rysunku zmienną pasywną jest zmienna C4, która nie była brana pod uwagę przy tworzeniu osi czynnikowych, a jedynie obliczone zostały współrzędne dla głównych składowych zbudowanych za pomocą zmiennych C1, C2, C3.

Po drugie wartości podane w procentach oznaczają procent wyjaśnianej ogólnej wariancji zmiennych (C1-C3). Podstawą obliczenia są tzw. wartości własne będące wynikiem dekompozycji wartości osobliwej i określają ilość wyjaśnianej wariancji. Z rysunku wynika, że pierwsza oś tłumaczy ponad 84% całkowitej wariancji zmiennych, a druga oś ponad 15%. W sumie dwie osie główne wyjaśniają blisko 100% całkowitej wariancji 3 zmiennych. Informacja o % wyjaśnianej wariancji jest użyteczna do identyfikacji ukrytych wymiarów leżących u podstaw analizowanych zmiennych. W naszym przykładzie można wysunąć wniosek, że wszystkie zmienne biorące udział w analizie mogą być sprowadzone praktycznie do jednego wymiaru, bowiem pierwsza oś główna ma zdecydowanie największe znaczenie w tłumaczeniu ogólnej zmienności danych w analizie.



Jeżeli macierzą wejściową w budowie mapy jest macierz korelacji, to wówczas interpretacja mapy ulega pewnej modyfikacji związanej ze standaryzowaną postacią danych

wejściowych.² Na rysunku pojawia się tzw. koło jednostkowe wskazujące na korelacyjny charakter danych.



Inną interpretację mają też wektory zmiennych, których długość nie wskazuje już na zmienność danych (jest ona już ustalona na poziomie 1), lecz na tzw. jakość reprezentacji zmiennych przez punkty na wykresie. Wiąże się to z tym, że w analizie wielowymiarowej występuje zasada kompromisu – możemy złożony układ zależności między wieloma zmiennymi pokazać w uproszczonym układzie o silnie zredukowanej liczbie wymiarów, lecz to uproszczenie będzie nas kosztować większą niedokładność odwzorowania poszczególnych zmiennych pierwotnych przez punkty na wykresie. Na rysunku widać jednak, że wszystkie wektory zmiennych mają długość w przybliżeniu równą 1 (leżą na kole jednostkowym). Stąd wniosek, że wszystkie zmienne są poprawnie reprezentowane przez punkty w 2-wymiarowym układzie osi głównych (jakość reprezentacji kategorii przez punkt wynosi 1).

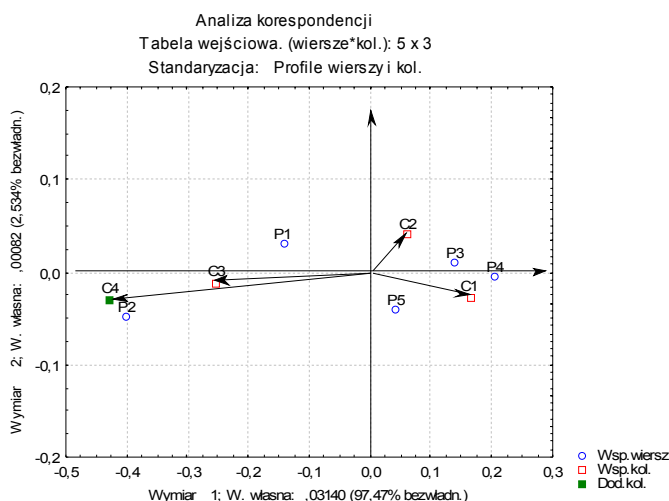
W analizie korespondencji interpretacja mapy uzależniona jest od wybranego sposobu skalowania elementów wierszy i kolumn. Najczęściej wyróżnia się 2 podstawowe rodzaje skalowania (normalizacji): wierszowo-kolumnowe oraz kanoniczne.³

Skalowanie wierszowo-kolumnowe jest podstawowym typem skalowania w analizie korespondencji. Polega ono na nałożeniu na siebie dwóch rodzajów map: mapy, w której punkty reprezentujące wiersze są reprezentowane w przestrzeni (na osiach) zdefiniowanej

² Macierz korelacji jest macierzą standaryzowanych kowariancji, w której kowariancje (korelacje) między zmiennymi zawierają się w przedziale $<-1;+1>$, a wariancje wszystkich zmiennych są stałe i wynoszą 1.

³ Analiza korespondencji jest jedną z metod dekompozycji prostokątnych macierzy danych. Najczęściej jest ona stosowana dla graficznej reprezentacji złożonych tabel kontyngencji zawierających częstości, jednakże może być – jak w poniższym przykładzie – stosowana dla dowolnych tabel zawierających nieujemne dane. Podobnymi metodami i stosowanymi zamiennie w stosunku do klasycznej analizy korespondencji są metody wzajemnego uśredniania, skalowania dualnego i analizy homogeniczności.

przez punkty odzwierciedlające kolumny, oraz mapy, w której punkty reprezentujące kolumny są zlokalizowane w przestrzeni stworzonej przez punkty reprezentujące wiersze.⁴ W tym rodzaju skalowania punkty reprezentujące wiersze i kolumny tabeli danych w odpowiednich przestrzeniach mają wyliczone tzw. współrzędne główne, a odległości między nimi są wyrażone w metryce Chi-Kwadrat, co umożliwia bezpośrednie porównywanie punktów należących do danego rodzaju danych między sobą. Możemy więc sensownie porównywać odległości tylko między punktami z jednego zbioru, a punkty drugiego zbioru służą do zdefiniowania osi głównych. Wybór należy już do badacza i wynika z przesłanek merytorycznych.⁵



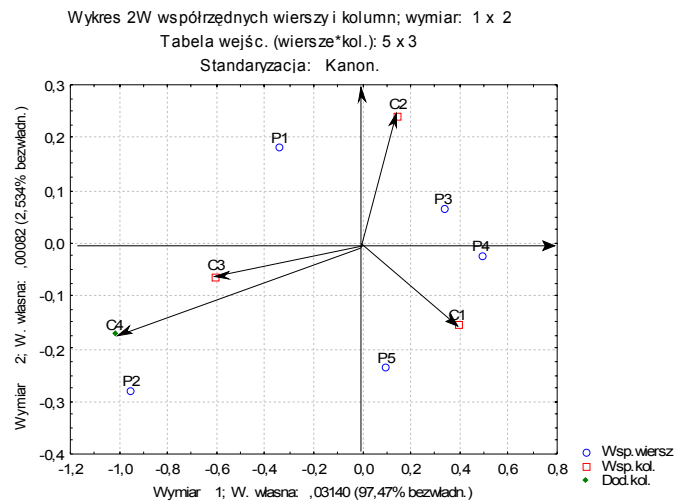
Dla przykładu w powyższej mapie przedstawione zostały współrzędne główne wierszy i kolumn tabeli danych. Jeżeli przyjmiemy, że w celu identyfikacji osi głównych posłużą nam zmienne (C1-C3), to interpretacja wzajemnych relacji między punktami wierszowymi jest analogiczna jak przedstawiona w poprzednim przykładzie.

Skalowanie kanoniczne umożliwia identyfikację relacji między kategoriami wierszy i kolumn. Pozwala na określenie, w jakim stopniu dany element w wierszu (lub kolumnie) wpływa na odchylenie elementów brzegowych kolumny (lub wiersza) od wartości przeciętnych. Elementy wierszy i kolumn powiązane znacznie między sobą będą leżały blisko siebie, a niepowiązane – daleko.⁶

⁴ Ten rodzaj map jest związany z tzw. szkołą francuską analizy korespondencji reprezentowaną przez Benzecri.

⁵ Pewien rodzaj skalowania zaproponowanego przez Carrolla, Greena i Shaffer zwanym również skalowaniem CGS próbował rozwiązać problem porównywania punktów należących do różnych zbiorów.

⁶ Współrzędne wierszy znajdują się w środkach ciężkości współrzędnych kolumn, a współrzędne kolumn znajdują się w środkach ciężkości współrzędnych wierszy.



Na rysunku widzimy, że np. zmienna C1 jest silnie powiązana z przypadkiem P5 (wartość zmiennej C1 silnie wpływa na odchylenie wartości przypadku P5 od wartości przeciętnej dla przypadków).

Na zakończenie należy zwrócić uwagę na relację między omówionymi metodami konstruowania biplotów a również wykorzystywaną metodą skalowania wielowymiarowego. W odróżnieniu od analizy głównych składowych i analizy korespondencji, metody skalowania wielowymiarowego (wielowymiarowego rozwijania – *multidimensional unfolding*) dokonują bezpośredniego odwzorowania graficznego danych wejściowych, które są traktowane jako bezpośrednie miary odmienności lub podobieństwa poszczególnych wierszy i kolumn tabeli danych.

Wykorzystanie analizy głównych składowych do budowy mapy zadowolenia klienta

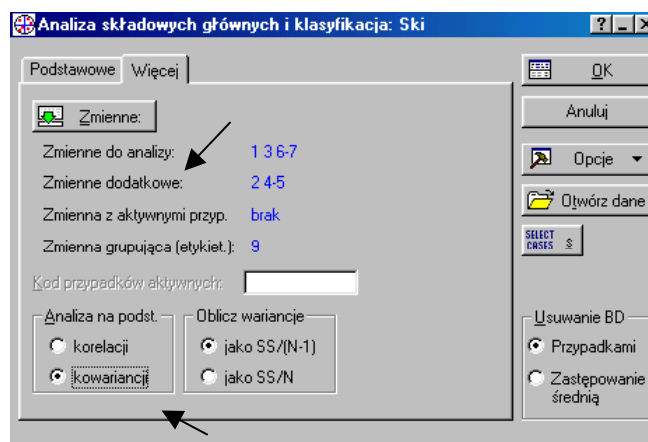
Na zakończenie zostanie zaprezentowany prosty przykład empiryczny ilustrujący proces budowy mapy zadowolenia klienta z ośrodków narciarskich w Polsce na podstawie metody analizy głównych składowych znajdującej się w module „Analiza głównych składowych i klasyfikacja” pakietu *STATISTICA*. Tabela danych przedstawia mediany ocen zadowolenia wśród klientów wybranych ośrodków narciarskich.

Poniższa tabela prezentuje 11 ośrodków narciarskich i 7 zmiennych kształtujących zadowolenie klientów oraz dwie dodatkowe zmienne demograficzne.



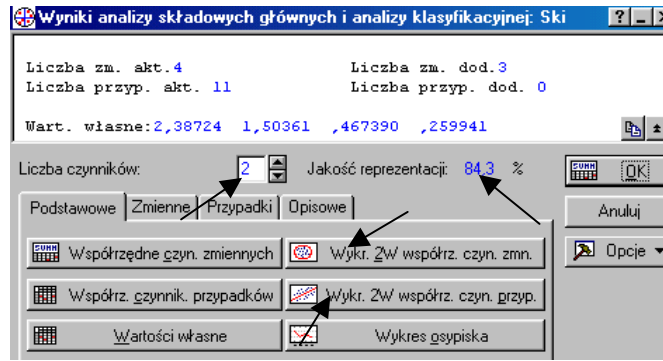
	1	2	3	4	5	6	7	8	9
	Warunki śniegowe	Atrakcyjność geograf.	Przygotowanie tras	Infrastruktura noclegowa	Koszty pobytu	Koszty zjazdów	Promocja regionu	Dochody	Region zamieszkania
Szczyrk	4	3	4	3	3	1	3	wysokie	MAZ
Szklarska Poręba	4	2	3	3	3	3	2	wysokie	MAL
Karpacz	2	4	3	4	4	2	3	średnie	PDK
Krynica	4	4	4	2	3	3	3	wysokie	SLA
Czarna Góra	3	1	3	1	4	4	1	niskie	SLA
Piwniczna	3	3	3	2	4	3	2	niskie	SLA
Korbielów	2	3	4	2	3	3	3	średnie	MAL
Szczawnica	4	2	2	3	4	4	3	średnie	MAL
Zakopane	3	4	3	4	1	1	4	to wysokie	MAZ
Białka Tatrzańska	1	2	1	2	3	2	3	średnie	MAL
Zieleniec	1	1	1	2	3	4	1	średnie	SLA

W pierwszym etapie budowy mapy należy określić charakter wybranych zmiennych w panelu początkowym analizy.

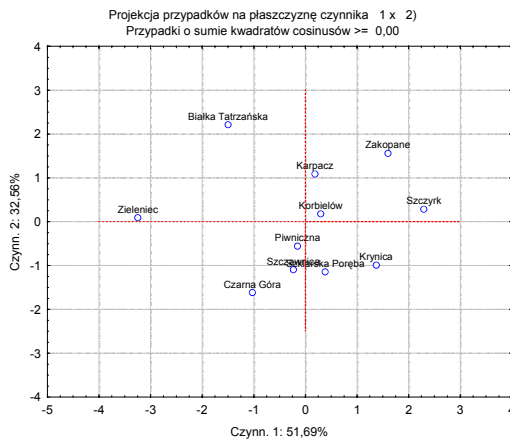


Zmienne aktywne w analizie są zmienne opisujące bezpośrednio ośrodki narciarskie, zmiennymi pasywnymi (dodatkowymi) zmienne, które były podstawą oceny zadowolenia, lecz stanowią ogólny opis miejscowości, w których ośrodek jest zlokalizowany. Będą one znajdować na biplotcie, nie wezmą jednakże udziału w budowaniu osi głównych. Dodatkowo wybrano zmienną miejsce zamieszkania (województwo) jako zmienną etykietującą przypadki (wartości modalne). Biplot jest zbudowany na podstawie macierzy kowariancji w celu uzyskania informacji o zróżnicowaniu poszczególnych zmiennych.⁷

⁷ Wszystkie zmienne w analizie są jednomianowe. Jeżeli analiza byłaby prowadzona na podstawie różnomianowych zmiennych, wówczas powinna być przeprowadzana w oparciu o dane standaryzowane (macierz korelacji).

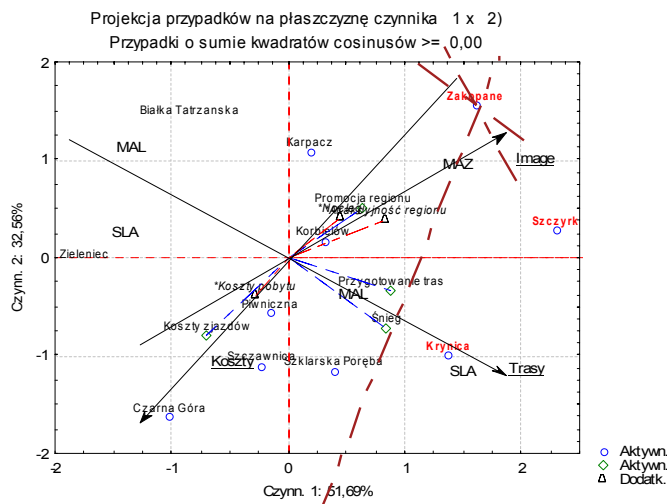
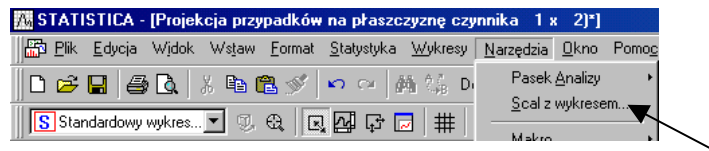


Po wyborze zmiennych i uruchomieniu programu zostały obliczone osie główne oraz pozycje punktów reprezentujących wiersze i kolumny tabeli danych. Nie wchodząc w techniczną interpretację metody analizy głównych składowych, interesować nas będzie jej interpretacja graficzna. Wyodrębnione zostały dwie główne składowe stanowiące osie główne układu, które wyjaśniają ponad 84% całkowitej wariancji 4 zmiennych aktywnych wziętych do analizy. W celu uzyskania reprezentacji zmiennych i przypadków należy uruchomić wykres 2W współrzędnych czynnikowych zmiennych oraz wykres 2W współrzędnych czynnikowych przypadków.



W celu uzyskania jednego wspólnego wykresu należy nałożyć jeden wykres na drugi, wykorzystując w menu opcję skalania.⁸

⁸ Można w tym celu skorzystać również z gotowego programu BiPlot.svb przygotowanego w STATISTICA Basic i zamieszczonego na stronie internetowej StatSoft.



Uzyskana mapa przedstawia na wspólnym układzie współrzędnych położenie zmiennych (wektorów) aktywnych (romby) i pasywnych (trójkąty) oraz punktów reprezentujących ośrodki narciarskie (kółka). Linie przerywane przedstawiają wektory zmiennych, a strzałki trzy podstawowe cechy ośrodków narciarskich wynikające ze silnego skorelowania poszczególnych zmiennych, tj. image ośrodka (promocja, atrakcyjność, infrastruktura), trasy (przygotowanie tras, warunki śniegowe) oraz koszty (koszty zjazdów i koszty pobytu). Przerywane linie oznaczają osie główne.

Z rysunku wynika, że wektory wizerunku ośrodków i jakości tras są silniej nachylone do poziomej osi głównej, a wektor kosztów – do pionowej osi głównej. Stąd 2 podstawowe ukryte właściwości ośrodków kształtujące zadowolenie można nazwać „jakość oferty” i „koszty”. Pierwszy wymiar jest ważniejszy dla oceny zadowolenia, bowiem wyjaśnia ponad połowę (51%) całkowitej wariancji zbioru danych, a wymiar kosztu odzwierciedla ponad 32% wariancji.

Najdłuższe wektory charakteryzują zmienne warunki śniegowe i koszty zjazdów, a najkrótsze – infrastruktura noclegowa i koszty pobytu. Klienci ośrodków najwyżej oceniają z punktu widzenia wizerunkowego Zakopane, Krynicy i Karpacz (oceny powyżej średniej), a najniżej Czarną Górę i Zieleniec (oceny poniżej średniej). Z punktu widzenia kosztów najlepiej wypadają Zieleniec i Czarna Góra, a najbardziej niekorzystnie – Zakopane. Z punktu widzenia podobieństwa profili ocen najbliższe są Szczawnica i Szklarska Poręba. Dodatkowo przyjęto Zakopane jako ośrodek stanowiący układ odniesienia dla pozostałych, co zostało zaznaczone linią przerywaną wynikającą z rzutów prostopadłych. Wynika z niej,



że poza wymiarem kosztowym najsilniejszym konkurentem Zakopanego jest Szczyrk i Krynica, z którymi Zakopane przegrywa konkurencję w zakresie tras, oraz Szklarska Poręba, i Szczawnica, gdzie klienci bardziej są zadowoleni z warunków śniegowych. Dodatkowo umieszczone są również punkty identyfikujące dominujące województwo, z którego przybywają turyści do poszczególnych ośrodków.

Literatura

1. Carroll J. D., P. E. Green, C. M. Schaffer, Interpoint Distance Comparisons in Correspondence Analysis, *Journal of Marketing Research* 1986/August, ss. 271-280.
2. Gabriel, K. R., Goodness of Fit of Biplot and Correspondence Analysis, *Biometrika* 2002/89, ss. 423-436.
3. Gabriel, K. R., The Biplot – Graphical Display of Matrices with Application to Principal Component Analysis, *Biometrika* 1971/58, ss. 453-467.
4. Hoffman, D. L., G. R. Franke, Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research, *Journal of Marketing Research* 1986/August, ss. 213-227.
5. Kroonenberg, P. M., Introduction to Biplots for GxE Tables, Research Report, Centre for Statistics The University of Queensland 1995/51.
6. Sokołowski, A., Identyfikacja osi w skalowaniu wielowymiarowym, *Taksonomia* 1995/2, ss. 97-100.