



METODY STATYSTYCZNE STOSOWANE DO ANALIZY ZADOWOLENIA I LOJALNOŚCI KLIENTÓW

Janusz Wątroba

StatSoft Polska Sp. z o.o.

Wprowadzenie

Opracowanie zostało poświęcone ogólnej charakterystyce kolejnego etapu badań satysfakcji i lojalności klientów, jakim jest analiza zebranych danych. Na wstępie zwrócimy uwagę na podstawowe czynniki warunkujące uzyskanie takich efektów analizy, które rzeczywiście umożliwiają wspomaganie procesu doskonalenia badanego produktu lub usługi. Następnie scharakteryzujemy ogólnie metody analizy stosowane w badaniach satysfakcji i lojalności. Szczególną uwagę zwrócimy na metody analizy wielowymiarowej, a wśród nich na grupę metod określaną terminem **key driver analysis**, czyli metody analizy kluczowych czynników powodzenia. Są to metody umożliwiające budowanie modeli zależności przeznaczonych do oceny wpływu różnych predyktorów na pojedynczą zmienną zależną (np. analiza regresji, analiza regresji logistycznej czy analiza dyskryminacyjna). Niektóre z przedstawionych metod zostaną zilustrowane krótkimi przykładami analiz w programie *STATISTICA*.

Czynniki warunkujące praktyczną użyteczność wyników analizy

Zaprojektowanie i przeprowadzenie badań wiąże się zazwyczaj z określonymi nakładami finansowymi. Z tego względu bardzo ważną rzeczą jest odpowiednie zadbanie o uzyskanie wymiernych efektów. Praktyczne korzyści z przeprowadzonych badań mogą polegać na lepszym poznaniu poziomu lub struktury czynników determinujących satysfakcję i lojalność konsumentów, jak również na usprawnieniu procesu decyzyjnego, związanego z rozwojem produktu, lub zwiększeniem efektywności jego sprzedaży (Allen i Rao 2000). Nie jest zapewne specjalnie odkrywczym stwierdzenie, że warunkiem koniecznym (choć niekiedy niewystarczającym) uzyskania praktycznie użytecznych wyników analizy jest prawidłowe przeprowadzenie wszystkich etapów badania. W przypadku etapu analizy danych oznacza to w szczególności, że w trakcie doboru lub konstruowania narzędzi mierzących określone aspekty satysfakcji i lojalności uwzględniono takie, które charakteryzują się jak najlepszą trafnością, a podczas przeprowadzania badań zadbano o odpowiednią wiarygodność gromadzonych danych (zagadnienia te były omawiane w poprzednich



referatach). Nawet bowiem najbardziej wyszukane i złożone procedury analityczne nie są w stanie dać użytecznych wyników w przypadku, gdy nie zagwarantujemy odpowiedniej jakości danych wejściowych. Dobrze ilustruje to wyrażenie „*garbage in garbage out*” (w dosłownym tłumaczeniu: jeśli włożysz śmieci, to również śmieci wyjmiesz), które zwraca uwagę na fakt, że uzyskanie sensownych wyników jest możliwe tylko w przypadku poprawnych danych wejściowych (Everitt 1998). W tym miejscu niestety nie sposób oprzeć się refleksji, że niektórzy badacze hołdują raczej zasadzie „*garbage in roses out*” i dość bezkrytycznie wykorzystują metody analizy danych (zwłaszcza te bardziej złożone) do swoiście pojętego uprawomocnienia wymyślanych przez siebie ad hoc prawdziwości czy też teorii (Rao 1994).

Warto zwrócić uwagę na czynniki gwarantujące uzyskanie wyników analizy dających praktyczne korzyści. Najważniejsze z nich to:

- ◆ zastosowanie narzędzi pomiarowych, które trafnie mierzą założone przez badacza wielkości,
- ◆ zebrane dane dobrze odzwierciedlają rzeczywisty poziom mierzonych wielkości,
- ◆ poddana badaniom próba klientów jest reprezentatywna dla populacji klientów, na którą mają być uogólnione wyniki badań,
- ◆ cele analizy zostały sformułowane w sposób umożliwiający ich zrealizowanie za pomocą metod analizy danych,
- ◆ osoba lub zespół przeprowadzający analizę wykazuje odpowiednią znajomość możliwości i ograniczeń procedur analizy danych oraz doświadczenie w poprawnej interpretacji uzyskanych wyników,
- ◆ dostęp do odpowiedniego oprogramowania komputerowego (np. programów z rodziny *STATISTICA*).

Powtórzmy jeszcze raz, że są to warunki, których spełnienie jest konieczne, ale nie gwarantuje uzyskania użytecznych praktycznie wyników.

Metody analizy stosowane w badaniach satysfakcji i lojalności

Do analizy danych opisujących różne aspekty (wymiary) satysfakcji i lojalności może być wykorzystany cały szereg różnych technik statystycznych. O stosowaniu określonych metod analizy decyduje kilka czynników. Wydaje się, że najważniejsze z nich to liczba przypadków, liczba zmiennych oraz charakter stawianych pytań lub hipotez badawczych. Biorąc pod uwagę te kryteria, można wyróżnić trzy grupy metod:

- ◆ metody analizy danych jednowymiarowych,
- ◆ metody analizy danych dwuwymiarowych oraz
- ◆ metody analizy danych wielowymiarowych.

Pierwsza grupa metod jest wykorzystywana do wstępnej charakterystyki zebranych danych i sprowadza się albo do określenia częstości występowania poszczególnych kategorii

odpowiedzi (w przypadku skal o małej liczbie kategorii), albo do podania miar charakteryzujących przeciętny poziom i rozrzut badanych wielkości (w przypadku skal o dużej liczbie kategorii lub skal metrycznych). Czasami przy ocenie poziomu satysfakcji bierze się pod uwagę częstości występowania najwyższej (lub dwóch najwyższych) kategorii skal (ang. top-box scores lub top-two boxes scores). Podejście to może budzić pewne wątpliwości ze względu na fakt, iż brane są pod uwagę charakterystyki niepełnego rozkładu badanej pozycji (cechy). Dla oceny przeciętnego poziomu najlepszymi miarami są takie charakterystyki jak: średnia, mediana lub wartość modalna. To, którą z nich należy zastosować, zależy od skali pomiarowej zmiennej oraz charakteru rozkładu jej wartości. Niektórzy autorzy zwracają przy tym uwagę, że czasami wybór odpowiedniej miary jest decyzją o charakterze „politycznym” (Allen i Rao 2000).

Wyniki uzyskiwane w efekcie analiz jednowymiarowych znajdują zastosowanie m.in. w przypadku tzw. badań śledzących, gdy badanie jest przeprowadzane wielokrotnie, w celu śledzenia zmian zachodzących w różnych aspektach percepcji respondentów (Hill i Alexander 2003).

Metody analizy danych dwuwymiarowych są wykorzystywane do oceny współzależności zachodzących pomiędzy dwoma cechami lub pozycjami. Wybór odpowiedniej miary zależy od skali pomiaru cechy. Jeśli obydwie cechy są wyrażone na skali przedziałowej lub ilorazowej to najczęściej stosowaną miarą jest współczynnik korelacji liniowej Pearsona. Z kolei jeśli jedna lub obydwie cechy są mierzone na skali porządkowej, to stosowaną charakterystyką jest współczynnik korelacji rang Spearmana, *Tau* Kendalla lub współczynnik Gamma. W przypadku skal nominalnych stosowane są takie statystyki, jak współczynnik *Fi* lub *V* Cramera.

Przytoczone wyżej miary współzależności dwóch zmiennych są często wykorzystywane do ustalania ważności branych pod uwagę pozycji lub cech w odniesieniu do pewnej „krytycznej” zmiennej zależnej (np. miary całkowitej satysfakcji klienta). Jednak biorąc pod uwagę znaczenie terminu „ważność”, nie jest to podejście poprawne (będzie o tym mowa w dalszej części referatu).

W związku z tym, że w badaniach satysfakcji i lojalności klientów mamy zazwyczaj do czynienia z danymi dotyczącymi większej liczby cech, więc spośród różnych metod analizy danych najczęściej wykorzystywane są metody wielowymiarowe. Jest to grupa metod obejmująca bardzo szeroki zakres narzędzi analitycznych. Dla ich scharakteryzowania wygodnie jest podzielić te metody na trzy szerokie klasy: **modele zależności**, **modele zależności wewnętrznych** oraz **modele hybrydowe**.

Modele zależności

Metody z pierwszej grupy są wykorzystywane do opisu wpływu jednej lub większej liczby predyktorów (zmiennych niezależnych lub objaśniających) na jedną lub większą liczbę zmiennych zależnych (objaśnianych). Zmienną zależną jest zazwyczaj jakaś ogólna ocena satysfakcji, lojalności, stopień zaangażowania („związania”) klienta lub skłonność do ponownego zakupu. Natomiast zbiór zmiennych objaśniających zawiera cały szereg szczegółowych ocen jakości produktów lub usług. W przypadku gdy mamy jedną zmienną



zależną o charakterze ilościowym, wówczas odpowiednią metodą będzie analiza regresji wielorakiej lub metoda oceny względnej ważności Kruskala. W przypadku zmiennej zależnej jakościowej najbardziej znane metody to analiza dyskryminacyjna, analiza regresji logistycznej oraz drzewa klasyfikacyjne. I wreszcie jeśli mamy większą liczbę zmiennych zależnych (ilościowych) powiązanych z wieloma predyktorami (ilościowymi), wówczas stosowaną techniką będzie analiza korelacji kanonicznych.

Modele zależności wewnętrznych

Druga grupa metod jest stosowana w sytuacji, kiedy interesuje nas struktura współzależności występujących w obrębie jednego zbioru zmiennych (przy czym w skład tego zbioru mogą wchodzić zarówno zmienne zależne, jak i zmienne niezależne). Celem tego typu analiz jest redukcja wymiaru danego zbioru danych w celu wykrycia mniejszej liczby ukrytych czynników, wyjaśniających możliwie dużą część zmienności oryginalnych danych. Najczęściej stosowanymi metodami z tej grupy są: analiza czynnikowa i metoda głównych składowych. Zostały one omówione w jednym z poprzednich referatów.

Modele hybrydowe

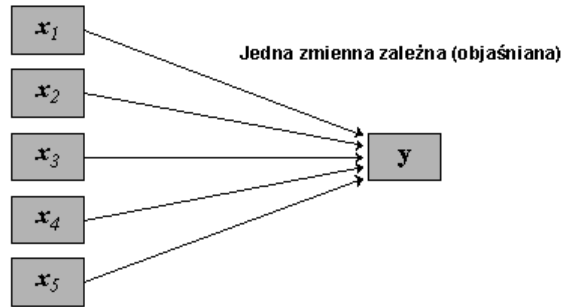
W analizie danych dotyczących satysfakcji klientów są również stosowane techniki łączące w sobie własności zarówno modeli zależności, jak i modeli analizy struktury danych. Najpopularniejsze z nich to modele równań strukturalnych dla zmiennych ukrytych (latent variable structural equation models), regresja głównych składowych (principal components regression) oraz modelowanie ścieżkowe z użyciem cząstkowych najmniejszych kwadratów (latent variable path modelling with partial least squares). Niektóre z tych metod zostaną omówione w dalszych referatach.

Analiza kluczowych czynników powodzenia (key driver analysis)

Spośród wspomnianych powyżej metod wielowymiarowych stosunkowo największą popularnością cieszą się modele zależności. Wśród badaczy zajmujących się badaniem satysfakcji klientów metody te są znane również pod nazwą analizy kluczowych czynników powodzenia (key driver analysis). W analizie tego typu występuje zazwyczaj jedna zmienna zależna oraz zbiór predyktorów, które są z nią powiązane. Najczęściej zakłada się, że model ujmujący powiązania predyktorów ze zmienną zależną jest modelem liniowym. Dobrym przykładem takiej metody jest model liniowej regresji wielorakiej. Zamieszczony poniżej schemat ilustruje ogólną postać modelu analizy zależności.

Jedną z zasadniczych korzyści, jakie daje stosowanie analizy kluczowych czynników powodzenia jest, możliwość ustalenia **ważności** poszczególnych czynników mających wpływ na zmienna zależną.

Predyktory (zmiennie objaśniające)



Rys. 1. Ogólna postać modeli analizy zależności

Ważność deklarowana i ważność pochodna

W badaniach satysfakcji klientów mamy do czynienia z dwojakim rozumieniem terminu ważność. Tak zwana **ważność deklarowana** (stated importance) oznacza wagę danego czynnika, jaką klient przywiązuje do produktu lub usługi, i która odzwierciedla najważniejsze jego wymagania. Z kolei **ważność pochodna** (derived importance) oznacza wysoki stopień współzmienności (wyrażonej kowariancją lub korelacją) danego czynnika ze zmienną zależną (np. ogólnym poziomem zadowolenia klienta). W tym miejscu trzeba wyraźnie zaznaczyć, że metody analizy kluczowych czynników powodzenia (takie jak np. analiza regresji wielorakiej) pozwalają tylko na modelowanie ważności pochodnej danego zbioru zmiennych objaśniających (predyktorów).

Zastosowanie analizy korelacji i regresji do oceny ważności predyktorów

Techniki analizy korelacji i regresji są obecnie bardzo często wykorzystywane w badaniu różnych aspektów satysfakcji klienta. Poniżej zostanie przedstawiony przykład takiej analizy. Wykorzystamy dane dotyczące firmy produkcyjnej o nazwie HugeCo. Zespół odpowiedzialny za zarządzanie jakością postawił sobie za cel poprawę ogólnej satysfakcji klientów. W tym celu przeprowadzono badania dotyczące pięciu najważniejszych wyznaczników jakości swoich produktów i usług. Badania miały charakter ankiety, w której do oceny wybranych wyznaczników jakości zastosowano skalę 10 punktową. Z bazy klientów firmy wylosowano 100 osób, które następnie poproszono o wypełnienie ankiety. Oryginalna ankieta zawierała dużo więcej pytań, natomiast dla potrzeb prezentowanej analizy wykorzystano następujące pozycje:

- ◆ Ogólne zadowolenie,
- ◆ Niezawodność produktów,
- ◆ Pomoc techniczna,
- ◆ Warunki dostawy,



- ◆ Cena oraz
- ◆ Warunki gwarancji.

Fragment pliku danych przedstawia poniższy rysunek.

Firma produkcyjna <i>HugeCo</i> przeprowadziła badania dotyczące pięciu najważniejszych wyznaczników jakości swoich produktów i usług. Zespół odpowiedzialny za zarządzanie jakością postawił sobie za cel poprawę ogólnej satysfakcji klientów. Przeprowadzono badania ankietowe w których do oceny wybranych wyznaczników jakości zastosowano skale 10 punktowe. Plik zawiera fragment zebranych danych.						
	1 Ogólne zadowolenie	2 Niezawodność produktu	3 Pomoc techniczna	4 Warunki dostawy	5 Cena	6 Warunki gwarancji
1	8	8	5	2	8	8
2	5	5	6	8	5	5
3	6	6	6	9	7	6
4	8	8	5	9	8	8
5	6	5	8	5	8	6
6	10	10	2	10	10	10
7	6	6	4	10	5	1
8	6	6	1	6	8	9

W pierwszym kroku analizy ocenimy powiązanie zmiennej *Ogólne zadowolenie* z pozostałymi zmiennymi, które traktujemy jako zmienne objaśniające (predyktory). W tym celu obliczymy wartości współczynników korelacji liniowej Pearsona.

Zmienna	Korelacje (Satysfakcja6_0)					
	Ogólne zadowolenie	Niezawodność produktu	Pomoc techniczna	Warunki dostawy	Cena	Warunki gwarancji
Ogólne zadowolenie	1,00					
Niezawodność produktu	0,71	1,00				
Pomoc techniczna	0,29	0,12	1,00			
Warunki dostawy	0,26	0,32	-0,11	1,00		
Cena	0,35	0,22	0,09	0,10	1,00	
Warunki gwarancji	0,49	0,33	0,17	0,13	0,35	1,00

Na podstawie wyników zamieszczonych w tabeli możemy stwierdzić, że zdecydowanie najmocniejsze powiązanie ze zmienną *Ogólne zadowolenie* wykazuje zmienna *Niezawodność produktu* ($r = 0,71$). Nieco słabszą korelację wykazuje zmienna *Warunki gwarancji* ($r = 0,49$). Najśłabszą, ale istotną statystycznie, korelację wykazuje zmienna *Warunki dostawy* ($r = 0,26$). Dla wszystkich zmiennych współczynnik korelacji przyjmuje wartość dodatnią, co oznacza, że im wyższa ocena danej pozycji tym wyższe ogólne zadowolenie klienta. Oprócz współczynników korelacji zmiennych objaśniających ze zmienną zależną w tabeli są również podane współczynniki korelacji w obrębie zmiennych objaśniających. Na ich podstawie możemy stwierdzić, że w najbardziej podobny sposób były oceniane *Cena* i *Warunki gwarancji*.

W drugim kroku analizy zbudujemy model opisujący łączny wpływ branych pod uwagę zmiennych objaśniających na zmienną *Ogólne zadowolenie*. W tym celu wykorzystamy moduł *Regresja wieloraka* w programie *STATISTICA*. Spośród dostępnych w tym module technik budowania modeli dla potrzeb niniejszego przykładu wybrano metodę krokową postępującą. Polega ona na wprowadzaniu lub usuwaniu z modelu zmiennych



objaśniających, aż do momentu znalezienia najlepszego modelu. Kryterium wejścia danej zmiennej do modelu jest wartość statystyki F , która określa, jak istotny wkład musi wnieść zmienna do modelu, aby można ją było uwzględnić w równaniu regresji. Poniżej zaprezentowano arkusz zawierający najważniejsze wyniki analizy.

		Podsumowanie regresji zmiennej zależnej: Ogólne zadowolenie (Satysfakcja)				
		R= ,78995108 R2= ,62402271 Skoryg. R2= ,60819208				
		F(4,95)=39,419 p<0,0000 Błąd std. estymacji: 1,0226				
N=100	BETA	Błąd st. BETA	B	Błąd st. B	t(95)	poziom p
W. wolny			1,081887	0,508676	2,126867	0,036022
Niezawodność produktu	0,589388	0,067218	0,556871	0,063510	8,768268	0,000000
Warunki gwarancji	0,224607	0,070523	0,159082	0,049949	3,184882	0,001959
Pomoc techniczna	0,166439	0,064011	0,104892	0,040340	2,600178	0,010805
Cena	0,130417	0,067741	0,090236	0,046870	1,925240	0,057190

Jak widać, do modelu nie weszła tylko zmienna Warunki dostawy. Na podstawie wartości współczynnika determinacji R^2 możemy stwierdzić, że uwzględnienie pozostałych zmiennych objaśniających pozwala wyjaśnić nieco powyżej 62,5% oryginalnej zmienności zmiennej zależnej. Gdybyśmy chcieli prognozować poziom satysfakcji klientów na podstawie oszacowanego modelu regresji, to przeciętny błąd prognozy wyniósłby około 1 (lub 13,8 % w stosunku do średniej wartości zmiennej objaśnianej). Otrzymany model możemy zapisać w następującej postaci:

$$\hat{Y} = 1,082 + 0,557 \cdot x_1 + 0,159 \cdot x_5 + 0,105 \cdot x_2 + 0,090 \cdot x_4$$

gdzie \hat{Y} oznacza prognozowaną (na podstawie modelu) ocenę satysfakcji klientów, x_1 to Niezawodność produktu, x_2 - Pomoc techniczna, x_4 - Cena, a x_5 - Warunki gwarancji. Na podstawie podanych wyżej współczynników regresji cząstkowej możemy ocenić wpływ, jaki do modelu wnoszą poszczególne zmienne objaśniające. W przypadku gdyby zmienne były wyrażone za pomocą różnych jednostek, wówczas porównanie ich wpływu należałoby oprzeć na wartości standaryzowanych współczynników regresji (beta).

Względna ważność predyktorów

Jednym z ograniczeń metody regresji wielorakiej jest to, że przy ustalaniu ważności predyktorów w ich wpływie na zmienną zależną interpretuje się wagi regresji, nazywane wagami beta. Są to miary pozwalające jedynie na uporządkowanie predyktorów pod względem ich współzmienności ze zmienną zależną. Nie dają natomiast możliwości ustalenia względnej ważności cech. Metodą umożliwiającą to zadanie jest algorytm względnej ważności Kruskala lub analiza dominacji. Metody te zostały omówione w książce Allena i Rao (2000).

Przykład zastosowania analizy dyskryminacyjnej

W badaniach dotyczących satysfakcji i lojalności klientów zdarzają się również zagadnienia, w których zmienna zależna (objaśniana) nie jest zmienną ilościową, lecz jakościową (np. wysoki poziom zadowolenia i niski poziom zadowolenia). W takich sytuacjach nie jest możliwe zastosowanie klasycznych technik regresyjnych, lecz stosuje się inne stosowne



metody. Najczęściej wykorzystywane są: regresja logistyczna, analiza dyskryminacyjna oraz drzewa klasyfikacyjne. Scharakteryzujemy krótko analizę dyskryminacyjną.

Główna idea leżąca u podstaw analizy dyskryminacyjnej to rozstrzygnięcie, czy grupy obiektów różnią się ze względu na średnią pewnej zmiennej (lub zmiennych), a następnie wykorzystanie tej zmiennej do przewidywania przynależności do grupy. Przykładowo możemy być zainteresowani ustaleniem cech różniących grupy intensywnych i sporadycznych użytkowników danego produktu, nabywców produktów określonych marek, klientów lojalnych i nielojalnych wobec marki itp. Zazwyczaj predyktory, czyli zmienne umożliwiające modelowanie klasyfikacji obiektów do klas, są zmiennymi ilościowymi, zostały jednak również opracowane techniki umożliwiające uogólnienie analizy dyskryminacyjnej na wypadek, gdy wśród predyktorów występują także zmienne jakościowe.

Dla zilustrowania przebiegu i korzyści związanych z zastosowaniem analizy dyskryminacyjnej wykorzystamy dane dotyczące zadowolenia klientów pewnych linii lotniczych. Losowo wybranej grupie 120 osób korzystających z usług tej firmy zadano pytanie: Czy na podstawie doświadczeń z obecnego lotu zdecydowałyby się Pani (Pan) na ponowne skorzystanie z naszych usług? Możliwe były tylko dwie odpowiedzi: Tak lub Nie. Oprócz tego pasażerom zadano pytanie, czy wcześniej korzystali już z usług tej firmy oraz poproszono ich o określenie zadowolenia z następujących aspektów: jakości posiłków, czystości, bezpieczeństwa, uprzejmości obsługi oraz wygody w czasie trwania lotu. Odpowiedzi były udzielane przy zastosowaniu 10-stopniowej skali. Poniżej na rysunku zamieszczono fragment zbioru danych.

Losowo wybranej grupie 120 osób, korzystających z usług pewnej linii lotniczej zadano pytanie: Czy na podstawie doświadczeń z obecnego lotu zdecydowałyby się Pani (Pan) na ponowne skorzystanie z usług? Możliwe były tylko dwie odpowiedzi: Tak lub Nie. Oprócz tego pasażerom zadano pytanie czy wcześniej korzystali już z usług tej firmy oraz poproszono ich o określenie zadowolenia z jakości posiłków, czystości, bezpieczeństwa, uprzejmości obsługi oraz wygody w czasie trwania lotu. Odpowiedzi były udzielane przy zastosowaniu 10-stopniowej skali. Celem analizy było zbudowanie modelu ujmującego czynniki wpływające na decyzję o ponownym skorzystaniu z usług.							
	1 Ponowny lot	2 Jakość posiłków	3 Czystość	4 Bezpieczeństwo	5 Uprzejmość obsługi	6 Wygoda	7 Czy wcześniej już latał
1	Nie	3	3	8	6	3	Tak
2	Nie	2	7	5	6	6	Nie
3	Tak	9	5	3	7	10	Nie
4	Tak	8	4	5	6	9	Tak
5	Nie	1	6	8	5	6	Nie
6	Nie	2	6	7	6	7	Nie
7	Tak	10	5	7	5	9	Tak

Do wykonania odpowiednich analiz wykorzystano moduł **Ogólna analiza dyskryminacyjna**, zawarty w programie *STATISTICA*. Poniżej w tabeli zamieszczono wyniki analizy, pozwalające na ocenę, które z predyktorów mają wpływ na przynależność ankietowanych osób do grup zmiennej zależnej. Okazuje się, że tylko w przypadku zmiennej **Bezpieczeństwo** odpowiedni poziom istotności przekracza 0,05, co oznacza, że jej udział w dyskryminacji jest nieistotny.



Efekt	Wielowymiarowe testy istotności (Dyskryminacja.sta) Parametryzacja z sigma-ograniczeniami Dekompozycja efektywnych hipotez					
	Test	Wartość	F	Efekt df	Błąd df	p
Wyraz wolny	Wilksa	0,587116	79,4664	1	113	0,000000
Jakość posiłków	Wilksa	0,521705	103,5974	1	113	0,000000
Czystość	Wilksa	0,951543	5,7545	1	113	0,018083
Bezpieczeństwo	Wilksa	0,988112	1,3595	1	113	0,246074
Uprzejmość obsługi	Wilksa	0,964479	4,1618	1	113	0,043678
Wygoda	Wilksa	0,945785	6,4775	1	113	0,012273
Czy wcześniej już latał	Wilksa	0,964460	4,1640	1	113	0,043622

W związku tym zbudujemy nowy model, uwzględniający tylko te zmienne, których udział w dyskryminacji jest wystarczający. Poniżej podano współczynniki funkcji dyskryminacyjnych dla takiego modelu. Funkcje te umożliwiają przeprowadzenie klasyfikacji.

Efekt	Funkcje klasyfikacyjne dla Ponowny lot (Dyskryminacja.sta) Parametryzacja z sigma-ograniczeniami			
	Poziom Efekt	Kolumna	Nie p=,3417	Tak p=,6583
Wyraz wolny		1	-15,0503	-26,1781
Jakość posiłków		2	1,1569	2,7612
Czystość		3	1,5553	1,9591
Uprzejmość obsługi		4	1,2219	0,8876
Wygoda		5	1,3848	1,7838
Czy wcześniej już latał	Nie	6	0,0867	-0,6542

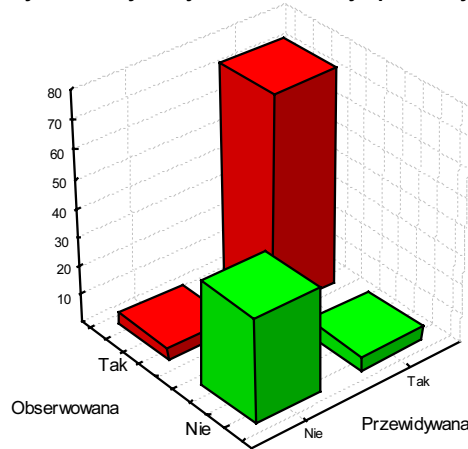
Najlepszym praktycznym sposobem oceny otrzymanego modelu jest ocena wyników klasyfikacji, jakie daje oszacowany model. Wyniki takiej klasyfikacji przedstawia poniższa tabela.

Klasa	Macierz klasyfikacji (Dyskryminacja.sta) Wiersze: obserwowana klasyfik. Kolumny: przewidywana klasyfikacja		
	Procent Poprawne	Nie p=,3417	Tak p=,6583
Nie	87,80	36	5
Tak	94,94	4	75
Ogół	92,50	40	80

Na podstawie wyników zamieszczonych w tabeli możemy stwierdzić, że model pozwala poprawnie sklasyfikować 36 ankietowanych, którzy deklarowali niewybranie ponownie usług tej firmy (stanowi to 87,8% wszystkich tych, którzy deklarowali niewybranie ponownie usług tej firmy), oraz błędnie przypisuje 5 osób do grupy tych, którzy wybraliby ponownie usługi tej firmy (stanowi to pozostałe 12,2% tych, którzy deklarowali niewybranie ponownie usług tej firmy). Jednocześnie model pozwala poprawnie sklasyfikować 75 osób, które deklarowały ponowne wybranie usług tej firmy (stanowi to 94,9% wszystkich tych, którzy deklarowali niewybranie ponownie usług tej firmy) oraz błędnie przypisuje 4 osoby do grupy tych, którzy nie wybraliby ponownie usług tej firmy (stanowi to pozostałe 5,1% tych, którzy deklarowali ponowne wybranie usług tej firmy).

Wyniki tego typu analizy mogą zostać wykorzystane do skonstruowania specjalnej oferty do określonych grup osób. Wyniki klasyfikacji można przedstawić graficznie za pomocą zamieszczonego poniżej wykresu.

Wykres klasyfikacji obserwowanej i przewidywanej



Podsumowanie

W ostatnich latach obserwuje się coraz większe zainteresowanie wykorzystaniem metod ilościowych do wspomaganie procesu poprawiania różnych aspektów zadowolenia i lojalności klientów. W artykule zwrócono uwagę na warunki, których spełnienie zwiększa prawdopodobieństwo uzyskania realnych korzyści z takiego podejścia. Spośród różnych metod zwrócono uwagę na techniki analizy pozwalające ustalić hierarchię ważności czynników mających wpływ na ogólne miary zadowolenia i lojalności klientów (key driver analysis). Dla przykładu zilustrowano użycie analizy regresji wielorakiej (w przypadku gdy zmienna zależna ma charakter ilościowy) oraz analizy dyskryminacyjnej (gdy zmienna zależna jest zmienną jakościową).

Literatura

1. Allen D. R., Rao T. R., 2000, *Analysis of Customer Satisfaction Data. A comprehensive guide to multivariate statistical analysis in customer satisfaction, loyalty, and service quality research*, ASQ Quality Press.
2. Everitt B. S., 1998, *The Cambridge Dictionary of Statistics*, Cambridge University Press.
3. Hill N., Alexander J., 2003, *Pomiar satysfakcji i lojalności klientów*, Oficyna Ekonomiczna.
4. Rao C. R., 1994, *Statystyka i prawda*, PWN
5. Rószkiewicz M., 2002, *Narzędzia statystyczne w analizach marketingowych*, Wydawnictwo C.H. Beck.
6. StatSoft, Inc., 2003, *STATISTICA* (data analysis software system), version 6. www.statsoft.com.