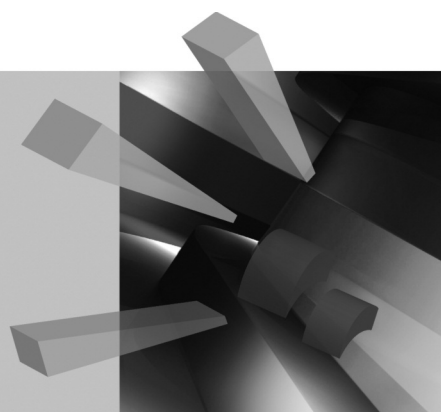


PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

237

Badania marketingowe w zarządzaniu przedsiębiorstwem



pod redakcją
Krystyny Mazurek-Łopacińskiej
Magdaleny Sobocińskiej



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2011

Recenzenci: Henryk Mruk, Andrzej Szromnik, Teresa Żabińska

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl> oraz w The Central and Eastern European Online Library www.ceeol.com, a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się na stronie internetowej Wydawnictwa www.wydawnictwo.ue.wroc.pl

Kopiowanie i powielanie w jakiegokolwiek formie wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2011

ISSN 1899-3192

ISBN 978-83-7695-252-9

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp.....	9
------------	---

Część 1. Badania marketingowe produktu oraz marki

Stanisław Kaczmarczyk: Badania marketingowe produktu w jego cyklu rynkowym	13
Marek Rawski: Możliwości i ograniczenia stosowania metody refleksji strategicznej w procesie planowania nowego produktu	22
Paweł Bryła: Marketingowa konstrukcja jakości regionalnych i ekologicznych produktów żywnościowych – koncepcja badania preferencji konsumentów i dystrybutorów	31
Beata Tarczydło: Narzędzia pomiarowe wizerunku marki. Wybrane przykłady	42
Mariola Grzybowska-Brzezińska, Katarzyna Tadajewska: Badania konsumenckich atrybutów jakości mleka spożywczego	52

Część 2. Badania zachowań konsumentów

Sławomir Smyczek, Artur Turek: Możliwość zastosowania diagnostyki medycznej w badaniach zachowań konsumentów	65
Sylwester Białowąg: Wpływ orientacji na osiągnięcia na zachowania oszczędnościowe gospodarstw domowych	75
Radosław Mącik, Monika Nalewajek: Motywacja racjonalności w korzystaniu z ICT w procesie podejmowania decyzji zakupowych przez konsumenta w świetle wyników badań empirycznych	85
Małgorzata Bombol: Jak zbadać kształtującą się Polską klasę wyższą – pytania i dylematy	99
Maja Jedlińska: Postmodernizm w zachowaniach konsumentów na rynku turystycznym.....	109
Adam Rudzewicz, Magdalena Krawczyk: Opinie konsumentów na temat reklamy internetowej	119

Część 3. Badania relacji, satysfakcji i lojalności oraz wartości dla klienta

Barbara Dobiegała-Korona, Alicja Krzepicka: Badania ukierunkowane na budowę wartości klienta	131
Adam Sagan, Anna Siwy-Hudowska: Wartość dla klienta na rynku konsumpcyjnym – porównanie trzech modeli pomiarowych.....	138

Edyta Rudawska: Więzi relacyjne w badaniach usług finansowych – ujęcie metodologiczne.....	150
Piotr Kwiatek: Metodologia badań relacji w parkach naukowych i technologicznych z wykorzystaniem teorii sieci.....	161
Krzysztof Błoński: Możliwości wykorzystania liczb rozmytych w badaniach satysfakcji klienta	168
Łukasz Skowron: Badania satysfakcji i lojalności klientów w różnych sektorach gospodarczych	178
Ewa Frąckiewicz: Badania satysfakcji klientów jednostek samorządu terytorialnego	187
Renata Nestorowicz: Badanie satysfakcji studentów jako klientów uczelni wyższych.....	197
Hanna Hall: Dydaktyczne kryteria satysfakcji studenta – hierarchia i znaczenie w świetle wyników badań empirycznych	209
Edyta Gołąb-Andrzejak: Badanie satysfakcji gości hotelowych i ich lojalności na przykładzie Grupy Hotelowej Orbis.....	220

Część 4. Zastosowanie badań marketingowych w procesach komunikacji i dystrybucji

Małgorzata Rószkiewicz: Metody formalne optymalizujące wybór grup docelowych dla przekazu marketingowego.....	231
Grzegorz Hajduk: Uwarunkowania i zakres całościowej oceny efektów komunikacji marketingowej – wybrane aspekty	241
Paweł Kowalski: Innowacja marketingowa w handlu detalicznym – przegląd obszarów badań	250
Tomasz Olejniczak, Piotr Biela: Wykorzystanie badań marketingowych w zarządzaniu siecią handlową – dylemat menedżera.....	266
Dawid Szostek: Badania opinii pracowników w przedsiębiorstwach zarządzanych marketingowo.....	273
Maja Knecht-Tarczewska: Adaptacja metod pomiaru jakości usług do oceny oferty marketingowej centrów handlowych	286
Lucyna Witek: Możliwości wykorzystania metody <i>Mystery Shopping</i> w ocenie efektywności i skuteczności działań merchandisingowych.....	296
Monika Hajdas: Audyt kreatywny komunikacji marketingowej w oparciu o archetyp marki	305
Katarzyna Bilińska-Reformat: Projekt „Audyt Marketingowy Młodej Firmy” jako narzędzie diagnozujące aktywność marketingową MŚP – ujęcie empiryczne.....	315

Summaries

Part 1. Marketing research of product and brand

Stanisław Kaczmarczyk: Marketing research of a new product in market cycle.....	21
Marek Rawski: Possibilities and limitations of practising the method of strategic reflection in the process of new product planning.....	30
Paweł Bryła: Marketing construction of origin and organic food quality – the concept of a research study concerning preferences of consumers and distributors.....	41
Beata Tarczydło: Measuring tools for brand image. Selected examples.....	51
Mariola Grzybowska-Brzezińska, Katarzyna Tadajewska: Research on consumer quality attributes of milk.....	61

Part 2. Consumers behaviour research

Sławomir Smyczek, Artur Turek: Applicability of medical diagnostics in consumer behaviour research.....	74
Sylwester Białowąs: Impact of achievement drive on the savings behaviour of households.....	84
Radosław Maćik, Monika Nalewajek: Rational motivations of ICT usage in consumer decision processes – empirical investigation.....	98
Małgorzata Bombol: Polish upper class – research, problems and dilemmas	108
Maja Jedlińska: Postmodernism in consumer behaviour on tourism market.	118
Adam Rudzewicz, Magdalena Krawczyk: Consumers' opinions about Internet advertising.....	128

Part 3. Research on the relationships, satisfaction and loyalty and value for the customer

Barbara Dobiegała-Korona, Alicja Krzepicka: Research oriented toward building Customer Lifetime Value.....	137
Adam Sagan, Anna Siwy-Hudowska: Value for the customer on the consumers market – a comparison of three models of measurement.....	149
Edyta Rudawska: Relational bonds in financial services research – methodological perspective.....	159
Piotr Kwiatek: Methodology of research on relationships in science and technology parks in network approach.....	167
Krzysztof Błoński: The possibility of using fuzzy numbers in the study of satisfaction.....	177

Łukasz Skowron: Research on the consumer satisfaction and loyalty in different business sectors	186
Ewa Frąckiewicz: Research on the satisfaction of clients of local government units.....	196
Renata Nestorowicz: Satisfaction survey of students as customers of universities	208
Hanna Hall: Didactic constituents of students' satisfaction – hierarchy and meaning based on results of empirical studies	219
Edyta Gołąb-Andrzejak: Research of hotel guests' satisfaction and their loyalty on the example of the Orbis Hotel Group.....	228

Part 4. The use of marketing research in the processes of communication and distribution

Małgorzata Rószkiewicz: Quantity approach to the optimization of target groups for marketing strategy	240
Grzegorz Hajduk: Conditions and scope of the overall marketing communication effects assessment – selected aspects.....	249
Paweł Kowalski: Marketing innovation in retail sector – review of research areas	265
Tomasz Olejniczak, Piotr Biela: Use of marketing research in trade network management – manager dilemma.....	272
Dawid Szostek: Employee opinion surveys in the marketingly managed enterprises.....	285
Maja Knecht-Tarczewska: Adjustment of the perceived service quality measurement methods in the evaluation of marketing offer of shopping centers.....	295
Lucyna Witek: Possibilities of use of mystery shopping method in the evaluation of merchandising actions efficiency	304
Monika Hajdas: Creative audit of marketing communication based on brand archetype.....	314
Katarzyna Bilińska-Reformat: “Marketing Audit of a Young Enterprise” project as a diagnostic tool of marketing activities of small and medium sized enterprises – empirical approach	325

Małgorzata Rószkiewicz

Szkoła Główna Handlowa w Warszawie

METODY FORMALNE OPTYMALIZUJĄCE WYBÓR GRUP DOCELOWYCH DLA PRZEKAZU MARKETINGOWEGO

Streszczenie: Warunkiem skutecznej orientacji marketingowej stają się m.in. bliskie kontakty z klientem oraz analiza kosztów obsługi klientów i generowanych przez nich wpływów, bazy danych i technologia informatyczna. Analizy te nie mogą obyć się bez modeli warunków działania firmy, ponieważ modele instrumentów jej oddziaływania na rynek pozwalają w warunkach zmiennego środowiska rynkowego przewidywać wyniki planowanych działań. Z tych powodów wiele uwagi poświęca się konstrukcji i weryfikacji empirycznej modeli formalnych umożliwiających optymalizację działań w zakresie przekazu marketingowego adresowanego do klienta, w celu rozwoju jego wartości dla firmy. W referacie dokonano przeglądu modeli formalnych wykorzystywanych do optymalizacji wyboru grup docelowych dla przekazu marketingowego, ilustrując ich własności.

Słowa kluczowe: optymalizacja grup docelowych, model wyboru dyskretnego, regresja logistyczna, regresja tobitowa, regresja ucięta, drzewa klasyfikacyjne.

1. Wstęp

Współczesne zarządzanie przedsiębiorstwem kładzie nacisk na kreowanie wartości przedsiębiorstwa dla jego akcjonariuszy. Oznacza to konieczność prowadzenia zarządzania zorientowanego na wzrost wartości rynkowej firmy oraz oceny rezultatów jej działalności. Jest to możliwe tylko wówczas, gdy uzyskuje się rosnącą dodatnią relację między zwrotem z zainwestowanego kapitału a średnio ważonym jego kosztem, połączoną z dostatecznie długim okresem utrzymywania trwałej przewagi konkurencyjnej na rynku.

Za główne generatory wartości przedsiębiorstwa uznaje się stopę wzrostu przychodów netto ze sprzedaży [Black, Wright, Bachman 2000, s. 18] oraz okres trwania wzrostu wartości. Pierwszy z czynników dotyczy marketingowych determinant wartości i jest on osiągany w wyniku takich działań operacyjnych w sferze marketingu, jak: zapewnienie rentownego wzrostu z wartością dodaną, wejście na nowe rynki, rozwinięcie i wprowadzanie nowych produktów, globalizacja biznesu, opracowanie programów lojalności klientów oraz oferowanie korzyści cenowych w nowych

punktach sprzedaży. Z kolei drugi z wymienionych generatorów wartości jest ściśle związany z budowaniem przewagi konkurencyjnej.

Ekonomiści wskazują na ścisłą zależność poziomu rentowności i silnej pozycji rynkowej firmy z wysoką jakością prowadzonej działalności. Podstawowym wymogiem jakości jest dostosowywanie wyrobów oraz usług do potrzeb i wymagań klientów. Jak podkreślają jednak Reinartz i Kumar, korzyści rynkowe w zarządzaniu klientem są zależne od rodzaju klientów [Reinartz, Kumar 2003]. Analiza zróżnicowania oczekiwań klientów jest zatem fundamentem dla strategii budowania konkurencyjności przedsiębiorstw (*customer driven strategy*) [Noori, Radford 1995, s. 49–55]. Zadowoleni klienci są niezbędnym zasobem i kapitałem firmy, dającym podstawę do utrzymania wysokiej stopy zwrotu z kapitału oraz wysokich stóp wzrostu. Ich identyfikacja pozwala na pomijanie w lokowaniu zasobów klientów niezyskowych [Venkatesan, Kumar 2004; Thomas, Reinartz, Kumar 2004, s. 117]. Stawia to wymóg zorganizowanego podejścia do zarządzania wartością klienta. Zarządzanie to obejmuje identyfikację klientów firmy, ich różnicowanie z punktu widzenia generowanych wpływów i zysków, interakcje z klientem w celu poznania jego oczekiwań, a także wpływanie na tworzenie wartości dla klienta oraz ciągłe dostosowywanie oferty firmy do zmieniających się potrzeb klienta. Zarządzanie to wymaga oceny kosztów oraz korzyści alternatywnych działań i rozpoznania optymalnej alokacji zasobów skierowanych na działania marketingowe według grup klientów [Berger i in. 2002, s. 48].

W orientacji marketingowej zakłada się, że decyzje menedżerskie mają początek na rynku, dotyczą wyboru rynków docelowych, dopasowywania możliwości firmy do potrzeb nabywców, działalność funkcjonalna zaś obejmuje dobór strategii. Orientacja ta przejmuje z klasycznego marketingu konieczność analizy oczekiwań konsumenta przed wprowadzeniem produktu na rynek. Oznacza to wymóg prowadzenia analizy otoczenia rynkowego, identyfikację celów firmy i jej polityk dziedzinowych, w tym – marketingowej. Oznacza potrzebę korzystania z metod szacowania niepewności, gromadzenia informacji i ich przetwarzania. Tym samym warunkiem skutecznej orientacji marketingowej stają się m.in. bliskie kontakty z klientem oraz ciągła analiza kosztów obsługi klientów i generowanych przez nich wpływów, bazy danych i technologia informatyczna. Analizy te nie mogą obyć się bez modeli warunków działania firmy. Modele instrumentów jej oddziaływania na rynek pozwalają bowiem w warunkach zmiennego środowiska rynkowego przewidywać wyniki planowanych działań.

Istnieje ekonomiczne uzasadnienie i zapotrzebowanie na optymalizację grup docelowych kampanii marketingowych oparte na modelowaniu zachowań rynkowych klientów. Z tych powodów wiele uwagi poświęca się konstrukcji i weryfikacji empirycznej modeli formalnych umożliwiających optymalizację działań w zakresie przekazu marketingowego adresowanego do klienta, w celu rozwoju jego wartości dla firmy.

2. Istota modelowania wyborów konsumenckich

Istota każdego algorytmu optymalizacji grup docelowych kampanii marketingowych tkwi w prognozie zachowań klientów, będącej z kolei podstawą przewidywań wyników rynkowych strategii biznesowej. Zachowania klientów polegają bowiem na wyborze jednego z kilku dostępnych wariantów decyzyjnych. Zgodnie z teorią zachowania konsumenta, racjonalność postępowania zakłada ocenę kosztów krańcowych i krańcowych korzyści na podstawie uzyskiwanych użyteczności z podjętej decyzji. Wynik tej kalkulacji jest traktowany zatem jako nieobserwowalna zmienna Y , interpretowana jako miara użyteczności wariantu decyzyjnego [McFadden 1984, s. 1396–1442, Greene 2000, s. 505–513]. Zmienna ta może być przedmiotem modelowania, według modelu regresji liniowej, w którym szacowane są parametry strukturalne. Zmienna Y , która jest wynikiem szacowanych przez klientów korzyści, nie jest bezpośrednio obserwowana, a obserwowalne są jedynie dokonywane przez klientów wybory, więc podejście tradycyjne szacowania parametrów strukturalnych modeli regresji liniowych nie może być stosowane. Zachowania klientów są zatem opisywane przez zmienne dyskretne, czyli np.:

$$y = \begin{cases} 1 & \text{dla } Y \geq y^* \\ 0 & \text{dla } Y < y^* \end{cases}$$

Ogólnie, mogą to być zmienne binarne, wielomianowe, licznikowe lub ograniczone. W pierwszych trzech przypadkach traktowane są jako zmienne jakościowe [Gruszczynski 2002, s. 11–50]. Modele tego typu zmiennych noszą nazwę modeli wyboru dyskretnego [Maddala 2001, s. 317–338].

W kontekście optymalizacji grup docelowych kampanii marketingowych kluczowa jest przynależność jednostki do grupy docelowej, rozumianej jako podzbiór zbioru wszystkich jednostek, dlatego też w wykorzystywanych algorytmach można wyróżnić podejście parametryczne, semiparametryczne i nieparametryczne. Podejście parametryczne polega na estymacji parametrów modelu regresyjnego, w którym zmienna zależna określa przynależność do grupy docelowej. Ogólnie, model grupy docelowej można zapisać jako:

$$P(Y) = f(X_1, X_2, \dots, X_k) + \xi,$$

czyli jako model regresji zmiennych objaśniających X_1, X_2, \dots, X_k . W modelach tego typu zmienna objaśniana Y może być cenzurowana, co oznacza, że obserwowane są tylko jej pewne wartości, takie, które są równe lub większe od pewnej wartości progowej y^* , a jeśli wartości te są mniejsze, to obserwowana jest tylko wartość progowa y^* . Model przybiera wówczas postać modelu tobitowego. Może to być zmienna ukryta, lecz obserwowana za pomocą zmiennej zero-jedynkowej, co oznacza, że jeśli zmienna Y przekroczyła pewną wartość progową, to obserwowana jest wartość 1, w pozostałych zaś przypadkach obserwowana jest wartość 0. Model przybie-

ra wówczas postać modelu dwumianowego. Może to być również zmienna ucięta, co oznacza, że obserwowane są tylko te jednostki, dla których zmienna Y przekroczyła wartość progową y^* . Zmiennymi objaśniającymi mogą być wszystkie dostępne charakterystyki klientów leżące zarówno po stronie klienta (cechy deskryptywne i behawioralne), jak i po stronie przedsiębiorstwa (charakter dotychczas odebranych działań marketingowych przez klienta). Co do funkcji f , to wymaga się, by była funkcją rosnącą. Dysponując oszacowaną funkcją regresji dla zmiennej jakościowej wyboru dyskretnego, można oszacować prawdopodobieństwo wystąpienia zachowania określonego warunkami wyboru.

3. Charakterystyczne modele dla dyskretnej zmiennej zależnej

W zależności od przyjętych założeń o postaci funkcyjnej modelu regresji dyskretnej zmiennej zależnej dostaje się różnej jakości rozwiązania. Najgorszej jakości są tzw. liniowe modele prawdopodobieństwa, z racji heteroskedastyczności składnika losowego w tego typu modelach. W estymacji modelu linowego oszacowania parametrów strukturalnych modelu uzyskuje się metodą najmniejszych kwadratów, czyli poszukuje się takich parametrów modelu, by spełniony był warunek: $\sum (y_i - \mathbf{a}\mathbf{x}_i - b)^2 \rightarrow \min(\mathbf{a}, b)$.

W modelach liniowych prawdopodobieństwa teoretyczne wartości zmiennej zależnej wychodzą poza przedział $(0, 1)$. W zasadzie model linowy jest w literaturze tematu rozważany jedynie teoretycznie ze względu na brak spójności rozwiązania z opisem rzeczywistości, której to rozwiązanie dotyczy. Wad tych nie mają modele wykorzystujące dystrybucję rozkładu normalnego oraz funkcję logistyczną. Przy tym, jak podkreśla Greene [2000, s. 505–513] różnice w przebiegu tych funkcji dotyczą jedynie wartości krańcowych. W modelach z nieliniową funkcją f parametry modelu (podejście parametryczne) szacuje się metodą największej wiarygodności. Warunek optymalizacyjny dotyczy maksymalizacji wiarygodności próby, w której zaobserwowano zmienną dychotomiczną, przybierającą jedynie wartości 0 (porażka) lub 1 (sukces) oraz wartości zmiennych niezależnych. Co do zmiennych niezależnych, to nie przyjmuje się żadnych ograniczeń. Mogą to być zmienne dowolnego typu.

W estymacji modelu tobitowego oszacowania parametrów strukturalnych modelu uzyskuje się również metodą największej wiarygodności, poszukując takich parametrów modelu, by spełniony był warunek:

$$\max_{\mathbf{a}, b, \sigma} \sum_{y > y^*} \ln \left(\frac{f(y - \mathbf{a}^T \mathbf{x} - b) / \sigma}{\sigma} \right) + \sum_{y = y^*} \ln \left(\Phi \left(\frac{y^* - \mathbf{a}^T \mathbf{x} - b}{\sigma} \right) \right),$$

gdzie: f oraz Φ – odpowiednio gęstość i dystrybuanta standaryzowanego rozkładu normalnego.

Również w estymacji regresji uciętej oszacowanie parametrów strukturalnych modelu uzyskuje się metodą największej wiarygodności, poszukując z kolei takich parametrów modelu, by spełniony był warunek:

$$\max_{\mathbf{a}, b, \sigma} \sum_i \ln \left(f \left(\frac{y - \mathbf{a}^T \mathbf{x} - b}{\sigma} \right) / \sigma \right) - \ln \left(\frac{y^* - \mathbf{a}^T \mathbf{x} - b}{\sigma} \right).$$

W modelu regresji logistycznej funkcja wiarygodności przybiera postać:

$$L = P(y_1, y_2, \dots, y_n) = P(y_1) \cdot P(y_2) \cdot \dots \cdot P(y_n) = \prod_{i=1}^n \left(\frac{1}{1 + e^{-(\mathbf{a}\mathbf{x} + b)}} \right)^{y_i} \left(1 - \frac{1}{1 + e^{-(\mathbf{a}\mathbf{x} + b)}} \right)^{(1-y_i)}.$$

Poszukuje się takich wartości parametrów będących współczynnikami kombinacji liniowej wykładnika równania logistycznego, dla których funkcja wiarygodności osiąga maksimum.

W modelach nieliniowych, gdy ocena współczynnika spełnia warunek $\alpha_j > 0$, uznaje się, że czynnik opisywany przez zmienną niezależną X_j działa stymulująco na prawdopodobieństwo (możliwość) wystąpienia badanego zdarzenia, przy kontrolowanym wpływie pozostałych zmiennych uwzględnionych w równaniu, jeśli zaś $\alpha_j < 0$, to uznaje się, że czynnik opisywany przez zmienną niezależną X_j działa ograniczająco na prawdopodobieństwo (możliwość) wystąpienia badanego zdarzenia, przy kontrolowanym wpływie pozostałych zmiennych uwzględnionych w równaniu. Gdy $\alpha_j = 0$, to uznaje się, że czynnik opisywany przez zmienną niezależną X_j nie ma wpływu na prawdopodobieństwo (możliwość) wystąpienia badanego zdarzenia, przy kontrolowanym wpływie pozostałych zmiennych uwzględnionych w równaniu.

Poprawność rozwiązania oceniana jest za pomocą ilorazu wiarygodności między funkcją wiarygodności dla modelu zawierającego jedynie stałą ($L(0)$) oraz dla oszacowanego modelu z k zmiennymi niezależnymi ($L(k)$), czyli:

$$-2 \ln \left(\frac{L(0)}{L(k)} \right) = -2 [\ln L(0) - \ln L(k)].$$

Dla dużych prób zarówno każdy ze składników tej różnicy, jak i sama różnica wartości funkcji wiarygodności mają rozkład zbliżony do rozkładu chi-kwadrat o liczbie stopni swobody $\nu = k - 1$, co umożliwi ocenę istotności uzyskanego rozwiązania. Ocenie podlega również istotność poszczególnych współczynników regresji logistycznej. Podstawą ich oceny jest wartość statystyki Walda,

$$\chi^2 = \left(\frac{\hat{\alpha}_j}{S(\hat{\alpha}_j)} \right)^2,$$

gdzie: $\hat{\alpha}_j$ – oszacowana na podstawie danych empirycznych wartość współczynnika,

$S(\hat{\alpha}_j)$ – standardowy błąd szacunku tego współczynnika.

Dla oszacowanego modelu wyznaczany jest również współczynnik dopasowania Coxa-Snella:

$$R_{C-S}^2 = 1 - \left(\frac{L(\text{stała})}{L(\text{zmienna}, \text{stała})} \right)^2$$

oraz współczynnik Nagelkerke'a będący jego normalizacją do przedziału $\langle 0; 1 \rangle$:

$$R_N^2 = \frac{R_{C-S}^2}{1 - \{L(\text{stała})\}^2}.$$

Test Hosmera-Lemeshowa ocenia zbieżność oszacowanego modelu do danych empirycznych. Podstawą testu jest ocena zbieżności obserwowanej liczby przypadków (sukces i porażka oddzielnie) z teoretyczną liczbą przypadków, wynikającą z oszacowanego modelu, w kolejnych grupach decylowych wartości zmiennej zależnej, jaką jest prawdopodobieństwo wystąpienia zdarzenia. Statystyka ta ma rozkład chi-kwadrat o $v = (g - 2)$ stopniach swobody, gdzie g określa liczbę grup decylowych.

Przedstawione modele nieliniowe, których parametry szacowane są metodą największej wiarygodności, wymagają spełnienia restrykcyjnych założeń o składniku losowym. W modelu regresji logistycznej wymagane jest, by składnik losowy był homoskedastyczny, a w modelach tobitowym i regresji uciętej dodatkowo wymaga się normalności jego rozkładu. Wymagana jest też niezależność zmiennych objaśniających.

W podejściu semiparametrycznym dokonuje się poluzowania założeń dotyczących składnika losowego modelu regresji zmiennej objaśnianej względem zmiennych objaśniających, co upraszcza procedury estymacyjne i poprawia własności uzyskiwanych estymatorów szacowanych parametrów równania regresji. Utrzymywane jest założenie o postaci funkcyjnej modelu, lecz nie zakłada się konkretnej postaci rozkładu składnika losowego. Założenia dotyczyć mogą jedynie pewnych cech tego rozkładu, np. wartości niektórych parametrów pozycyjnych lub kształtu rozkładu [Manski 1975; Owczarczuk 2009, s. 31–35].

W podejściu nieparametrycznym przedmiotem oceny są jedynie wartości zmiennej zależnej, a budowany model zależności jest wolny od parametrów i wiążących założeń. Do grupy tego typu metod należą procedury data miningowe, a wśród nich sieci neuronowe, programowanie genetyczne, a także metody rekurencyjnego podziału, zwane procedurami drzew klasyfikacyjnymi. Drzewa klasyfikacyjne są procedurą, która umożliwia podział zbioru na podzbiory o różnej intensywności występowania o ustalonej wartości zmiennej objaśnianej. Jeśli zmienna ta jest nominalna lub porządkowa, to procedura określana jest mianem drzewa klasyfikacyjnego lub dyskryminacyjnego, jeśli zaś jest to zmienna przedziałowa lub ilorazowa, to procedura określana jest mianem drzewa regresyjnego. Podejście z wykorzystaniem podziału rekurencyjnego ma charakter nieparametryczny, w tym sensie, że nie

jest wymagana znajomość rozkładów analizowanych zmiennych. Metoda drzewa klasyfikacyjnego ma również inne zalety. Należy do nich przede wszystkim brak wymagań co do znajomości klasy funkcji opisujących zależność między zmiennymi objaśniającymi i zmienną objaśnianą oraz specyfikacji zmiennych objaśniających (ich dobór dokonywany jest wraz z rozwojem drzewa). Ponadto nie ma żadnych ograniczeń co do typu tych zmiennych (mogą to być zarówno zmienne nominalne, porządkowe, jak i przedziałowe oraz ilorazowe), model jest odporny na transformacje monotoniczne predyktorów oraz na problemy wynikające z jakości materiału statystycznego, tj. obserwacje nietypowe oraz braki danych.

Istota metody sprowadza się do podziału na podgrupy zbioru jednostek, charakteryzowanych przez zbiór zmiennych objaśniających, tak by wyodrębnione grupy charakteryzowały się jak najwyższą jednorodnością ze względu na zmienną objaśnianą. Procedura ma charakter rekurencyjny w tym sensie, że w kolejnych iteracjach wybierana jest tylko jedna zmienna ze zbioru zmiennych objaśniających, ze względu na którą dokonuje się podziału badanej zbiorowości na podgrupy. E. Gatnar [2001, s. 24] procedurę drzewa klasyfikacyjnego ujął w schemat następujących pięciu kroków:

1. Ocena stopnia heterogeniczności zbioru pod względem zmiennej objaśnianej. Znaczna jednorodność podważa zasadność dokonywania klasyfikacji.

2. Rozważenie wszystkich możliwych podziałów zbioru ze względu na kolejne proponowane zmienne objaśniające.

3. Ocena jakości każdego z proponowanych podziałów według wybranego kryterium oceny homogeniczności tworzonych grup.

4. Wybór jednej ze zmiennych objaśniających według ustalonego kryterium jakości i dokonanie podziału zbioru na podgrupy.

5. Powrót do p. 1 w ramach każdej z utworzonych w p. 4 podgrup.

Kolejne pętle 1–5 powtarzane są tak długo, aż dezagregacja jednostek doprowadzi do wydzielenia grup o całkowitej homogeniczności ze względu na zmienną objaśnianą lub wyczerpie się liczba zadanych iteracji (pętli) bądź też wyodrębniane w kolejnych podziałach grupy osiągnęły liczebności niewystarczających do kontynuowania podziałów. Kryteria te mogą prowadzić do bardzo rozbudowanych schematów podziału, czyli ich wynikiem są drzewa o znacznych rozmiarach. Dlatego też innym kryterium zatrzymania procedury (kryterium stopu) jest osiągnięcie jedynie satysfakcjonującej homogeniczności segmentów, co oznacza, że wyodrębnione podgrupy nie będą całkowicie jednorodne. Konsekwencją takiego rozwiązania jest brak całkowitej pewności w przewidywaniu przynależności jednostki do grup na podstawie zestawu zmiennych objaśniających. Niedokładność takiego typu określa się mianem błędu klasyfikacji, a mierzy go wskaźnik ryzyka błędnej klasyfikacji. Jego poziom w opisywany jest przez frakcję jednostek, które w wyodrębnionych grupach nie mają oczekiwanej wartości zmiennej objaśnianej. Jest on szacowany wraz z błędem standardowym $\sqrt{\frac{w(1-w)}{n}}$.

W zależności od tego, na którym poziomie dezagregacji zostanie zatrzymana procedura, w profilowaniu wyróżnionych grup uczestniczą wszystkie wyróżnione zmienne objaśniające lub tylko ich część.

Wśród miar oceny stopnia homogeniczności wyodrębnianych grup pod względem wartości lub kategorii zmiennej zależnej najpopularniejsze to:

- entropia rozkładu, wyznaczana według wzoru:

$$H(S_k) = - \sum_{i=1}^L p(i|k) \log_2 p(i|k),$$

gdzie: $p(i|k)$ – empiryczne prawdopodobieństwo (częstość) występowania i -tej kategorii zmiennej zależnej w k -tym segmencie, gdzie $i = 1, \dots, L$;

- wskaźnik zróżnicowania Giniego:

$$H(S_k) = 1 - \sum_{i=1}^L p^2(i|k);$$

- miara dobroci podziału algorytmu CART, w którym rozważane są tylko podziały binarne [Larose 2006, s. 112]:

$$\Phi(s|k) = 2p_{1k} \cdot p_{2k} \sum_i |p_{j1k} - p_{j2k}|,$$

gdzie: p_{1k} – prawdopodobieństwo empiryczne (częstość) dla podgrupy o numerze 1 w podziale binarnym k -tej grupy,

p_{2k} – prawdopodobieństwo empiryczne (częstość) dla podgrupy o numerze 2 w podziale binarnym k -tej grupy,

p_{j1k} – prawdopodobieństwo empiryczne (częstość) dla podgrupy o numerze 1 w podziale k -tej grupy według j -tej kategorii rozważanego deskryptora,

p_{j2k} – prawdopodobieństwo empiryczne (częstość) dla podgrupy o numerze 2 w podziale k -tej grupy według j -tej kategorii rozważanego deskryptora;

- miara zależności między kryterium segmentacji i predykatorem algorytmu CHAID:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}},$$

gdzie: n_{ij} – liczebność empiryczna odnosząca się do i -tej kategorii kryterium segmentacji i j -tej kategorii deskryptora,

$$\hat{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

– liczebność teoretyczna odnosząca się do i -tej kategorii kryterium segmentacji i j -tej kategorii deskryptora, gdy:

$$n_{i.} = \sum_j n_{ij}; n_{.j} = \sum_i n_{ij}.$$

W algorytmie CHAID ocena statystyki χ^2 dokonywana jest na podstawie jej istotności. Własności rozkładu χ^2 , zależne od stopni swobody, a dokładniej asymetryczność tego rozkładu, sprawiają, że oceny te są nieporównywalne. Dlatego też, by sprowadzić je do porównywalności zaleca się stosowanie poprawki Bonferroniego o postaci [Gatnar 2001, s. 57]:

$$\alpha^* = \frac{\alpha}{B(T)},$$

$$B(T) = 1 + \sum_{k=1}^T N(T-k, T-k),$$

$$N(T-k+1, T-k) = \sum_{i=0}^{T-k-1} (-1)^i \frac{(T-k-i)^{T-k-i}}{i!(T-k-i)!},$$

gdzie: α – krytyczna istotność związana z obliczoną wartością statystyki χ^2 ,
 $B(T)$ – mnożnik wyznaczony dla wszystkich możliwych podziałów T ,
 $N(T-k+1, T-k)$ to liczba grup w kroku k oraz $k+1$.

Prognoza zmiennej objaśnianej w tej procedurze odbywa się na podstawie frakcji zdarzeń $Y = 1$ w każdej z wygenerowanych podgrup. Dla prognozy tej wyznacza się wskaźnik ryzyka błędnej klasyfikacji.

4. Uwagi końcowe

Modele dla prawdopodobieństwa określonych zachowań klientów nie tworzą uniwersalnego zestawu narzędzi do optymalizacji grup docelowych przekazu marketingowego. Każdy z nich przez swe własności zarówno wyznacza korzyści analityczne, jak i określa ograniczenia zastosowanego podejścia. Ich dobór powinien uwzględniać specyfikę problemu, którego dotyczą. Przedstawione wyżej dwa główne nurty optymalizacji grup docelowych przekazu marketingowego jednocześnie reprezentują dwa odmienne podejścia z poznania naukowego. Podejście parametryczne jest głęboko zakorzenione w metodzie indukcyjnej, gdzie falsyfikacja pozwala określić ramy poznania naukowego. Z kolei podejście nieparametryczne koncentruje się na eksploracji danych i prowadzi do budowania wiedzy z danych. Trzeba być w pełni świadomym ograniczeń i zagrożeń poznawania prawdy metodą odkrywania reguł, akcentując konieczność weryfikacji budowanych w ten sposób modeli.

Literatura

Berger P., Bolton R., Bosman D., Briggs E., Kumar V., Parasuraman A., Terry C., *Marketing actions and the value of customer base. A framework for customer asset management*, „Journal of Service Research” 2002, vol. 5, no. 1.

- Black A., Wright P., Bachman J.E., *W poszukiwaniu wartości dla akcjonariuszy. Kształtowanie wyników działalności spółek*, ABC, Warszawa 2000.
- Gatnar E., *Nieparametryczne metody dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa 2001.
- Greene W.H., *Econometric Analysis*, Prentice Hall, Upper Saddle River, NJ, 2000.
- Gruszczynski M., *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, SGH, Warszawa 2002.
- Larose D.T., *Odkrywanie wiedzy z danych*, Wydawnictwo Naukowe PWN, Warszawa 2006.
- Maddala G., *Introduction to Econometrics*, John Wiley & Sons, Chichester 2001.
- Manski C.F., *Maximum score estimation of the stochastic utility model of choice*, „Journal of Econometrics” 1975, vol. 3, s. 205–228.
- McFadden D.L., *Econometric analysis of qualitative response models*, [w:] Z. Griliches, M.D. Intriligator (red.), *Handbook of Econometrics*, Elsevier Science Publishers, Amsterdam 1984.
- Noori H., Radford R., *Production and Operation Management. Total Quality and Responsiveness*, McGraw-Hill, New York 1995.
- Owczarczuk M., *Estymatory typu Maximum score dla wybranych modeli mikroekonometrycznych*, praca doktorska, Kolegium Analiz Ekonomicznych, SGH, Warszawa 2009.
- Reinartz W., Kumar V., *The impact of customer relationship characteristics on profitable lifetime duration*, „Journal of Marketing” 2003, vol. 67, s. 77–99.
- Thomas J., Reinartz W., Kumar V., *Getting the most of all your customers*, „Harvard Business Review” 2004, July-August.
- Venkatesan R., Kumar V., *A customer lifetime value framework for customer selection and resource allocation strategy*, „Journal of Marketing” 2004, vol. 68, s. 106–125.

QUANTITY APPROACH TO THE OPTIMIZATION OF TARGET GROUPS FOR MARKETING STRATEGY

Summary: Close relations with the customer and the analysis of customer service costs and revenues generated by them, as well as databases and information technology become the condition for an effective marketing orientation. The analysis cannot do without models referring to terms of companies' activities. The models of the instruments of their influence on the market allow under conditions of the changing market environment to predict the results of the planned activities. For these reasons, much attention is paid to the design and formal verification of empirical models to optimize operations in the field of marketing communication addressed to a client in order to develop its value for the company. The paper reviews formal models used to optimize the selection of target groups for marketing communication.

Keywords: optimization of target groups, models of discrete choice, logistic regression, tobit models, truncated regression.