Zeszyty Naukowe Uniwersytet Ekonomiczny w Krakowie
Metody analizy danych
916
ISSN 1898-6447
Zesz. Nauk. UEK, 2013; 916: 47–58

Piotr Tarka
Katedra Badań Rynku i Usług
Uniwersytet Ekonomiczny w Poznaniu

Model czynnikowy ze zmiennymi dychotomicznymi w analizie ukrytej struktury zjawisk rynkowych

Streszczenie

W artykule przedstawiono teoretyczne założenia konstrukcji wraz z praktycznymi implikacjami modelu czynnikowego w kontekście badań marketingowych. W pierwszej części zwrócono uwagę na zależność badań marketingowych od rozwoju teorii pomiaru w ramach analizy zjawisk rynkowych o charakterze ukrytym. Następnie omówiono proces konstrukcji modelu czynnikowego do analizy zmiennych dychotomicznych. Napodstawie przeprowadzonych badań empirycznych zaprezentowano funkcjonalność jednoczynnikowego modelu w sferze pomiaru postaw młodzieży (młodych konsumentów) wobec hedonistyczno-konsumpcyjnego stylu życia.

Słowa kluczowe: model czynnikowy, zmienne ukryte, dane dychotomiczne, skala nominalna.

1. Wprowadzenie

Gdyby projekty badawcze były zdominowane przez antyrealistów (kierujących się zasadą „brzytwy Ockhama")[1], wiele pomysłów w badaniach, w tym także

[1] Zasada ta polega m.in. na tym, że wszystko to, czego nie można zmierzyć, nie istnieje.

wiele interesujących odkryć badawczych w marketingu, nigdy nie doczekałoby się realizacji. Na szczęście w metodologii badań marketingowych (w sposób szczególny w sferze analizy zjawisk rynkowych ukrytych) zaistniały także prace realistów, według których zmienne obserwowalne podlegają statystycznym procesom łączenia i redukcji, co może przyczynić się do odkrywania zjawisk o charakterze ukrytym. Zdaniem realistów to, co jest widoczne i mierzone za pomocą zmiennych obserwowalnych, dotyczy często ukrytej rzeczywistości, którą należy rozpoznawać jedynie drogą pośrednią [Sagan 2000]. Przykładowo w badaniach marketingowych w zakresie analizy konsumentów wyrażających określone postawy i opinie wobec produktów czy usług pomiar bezpośredni byłby mało skuteczny, a otrzymana na jego podstawie informacja – mało wiarygodna. Oznacza to, że prawdziwa informacja zawarta jest w zmiennej ukrytej, która odzwierciedla pośredni wkład szeregu zmiennych obserwowalnych [Sikorska 2012].

W niniejszym artykule podjęto próbę przybliżenia modelu czynnikowego ze zmiennymi dychotomicznymi w celu rozpowszechnienia tego podejścia analitycznego wśród badaczy akademickich i praktyków. W części pierwszej zaprezentowano proces konstrukcji modelu czynnikowego do analizy danych dychotomicznych. W części drugiej zawarto zaś prostą ilustrację modelu, w ramach której ukazano i zinterpretowano wybrane składniki modelu omówionego w części pierwszej. W ramach przykładu posłużono się modelem jednoczynnikowym w sferze pomiaru postaw młodzieży (młodych konsumentów) wobec hedonistyczno-konsumpcyjnego stylu życia.

2. Znaczenie modeli czynnikowych w badaniu marketingowych zjawisk ukrytych

Analiza ukrytych postaw konsumentów, nieobserwowalnych bezpośrednio i mających znaczenie denotacyjne o głębokim „ładunku informacyjnym”, nie byłaby możliwa bez wpływu statystyki i rozwoju teorii psychometrii. Należy w tym miejscu wyraźnie podkreślić fakt, że pomiar zmiennych ukrytych dokonuje się poprzez wyznaczenie zestawu zmiennych obserwowalnych (wskaźników lub indeksów). Oczywiście charakter wskaźników generujących dany czynnik może się wyrażać dwuzakresowo – poprzez wskaźniki refleksyjne bądź formatywne, za pomocą których oceniany jest charakter wzajemnych relacji przyczynowych. W modelu czynnikowym opartym na klasycznej teorii pomiaru wykorzystuje się przede wszystkim wskaźniki refleksyjne, zaś domeną wskaźników formatywnych jest analiza głównych składowych. Najistotniejszą kwestią badań statystycznych są techniczne aspekty rozpoznania relacji przyczynowych. Kwestie sporne dotyczą na ogół tego, jak odnaleźć najlepszy model analityczny odwzorowujący właściwy poziom jakości dopasowania różnego typu danych (pozyskiwanych ze skal mocnych bądź skal słabych) ze względu na wzajemne relacje między zmiennymi obserwowalnymi a czynnikami. W zależności od dokładności rozpoznania tych relacji można mówić o tym, że dany model pomiaru przyjmuje postać niezidentyfikowaną, zidentyfikowaną lub nadidentyfikowaną[2].

Modele czynnikowe (będące elementem szerszego nurtu badań statystycznych, tj. badania współwystępowania) pozwalają nie tylko określić jakość i spójność związków zmiennych, ale także dokonać swoistego rodzaju ilościowej syntezy wniosków w zbyt złożonej strukturze danych. R.L. Gorsuch [1974] podaje np. trzy główne zalety modeli czynnikowych, mianowicie: 1) uproszczoną analizę złożonej macierzy obejmującą wiele wskaźników, 2) możliwość oceny poziomu zmienności związków zachodzących między poszczególnymi wskaźnikami w obrębie badanych czynników, 3) detekcję malejących lub rosnących granicznych obszarów związków, jakie rozdzielają wskaźniki tworzące dany czynnik od pozostałych wskaźników (zmiennych obserwowalnych) i czynników.

3. Konstrukcja modelu czynnikowego na podstawie danych dychotomicznych

Badacze z zakresu marketingu w modelach czynnikowych wykorzystują dane, których pomiar w zakresie cech respondentów dokonuje się na skali porządkowej lub skali interwałowej zgodnie z klasyczną teorią pomiaru. Niewiele osób stosuje w praktyce badawczej modele czynnikowe oparte na danych dychotomicznych, odwołujące się m.in. do teorii reakcji na pozycje skali (*item response theory*).

Teoretyczną podbudowę założeń modeli czynnikowych w ujęciu danych dychotomicznych przedstawili w swoich pracach m.in. D. Bartholomew [1980], I. Moustaki i M. Knott [2000], K.G. Jöreskog i I. Moustaki [2001], I. Moustaki i F. Steele [2005]. W procesie konstrukcji modelu czynnikowego ze zmiennymi dychotomicznymi można wykorzystać funkcję logistyczną i model regresji logistycznej w celu określenia zależności każdej z rozpatrywanych zmiennych obser-

---

[2] Warto w tym miejscu dodać, że związki te (choćby nawet badane statystycznie za pomocą modelu matematycznego) są zawsze trudne do uznania za wystarczające, tzn. w pełni identyfikowalne. Relacje między płaszczyzną zmiennych obserwowalnych a zmiennymi ukrytymi – czynnikami – opierają się bowiem nie tylko na *stricte* matematycznych zależnościach, ale są uwarunkowane istnieniem teorii abstrakcyjnych (opartych wyłącznie na pojęciach teoretycznych z danej dziedziny aplikacji) oraz teoriami empirycznymi poszukującymi określonych faktów (m.in. poprzez pryzmat działań statystycznych). Stąd rozpoznanie ukrytej struktury danego zjawiska dokonuje się na gruncie teoretycznym (z dokładną znajomością danej dziedziny) oraz na poziomie zjawisk obserwowalnych.

wowalnych (pozycji) wobec generowanego czynnika[3] (zob. rys. 1). W funkcji tej dychotomiczna zmienna obserwowalna x_i przyjmuje wartość 1 lub 0 względem danego czynnika *f*, przy czym wartość oczekiwana zmiennej x_i w ujęciu rozpatrywanego przez badacza czynnika będzie określona tak, że: $P\left(x_i = 1 \middle| \mathbf{f}\right) = \pi_i\left(\mathbf{f}\right)$, gdzie $\pi_i(\mathbf{f})$ oznacza prawdopodobieństwo warunkowe rozpatrywanej zmiennej x_i uzyskującej wartość 1, jeśli zmienna ta odpowiada danemu czynnikowi spośród *j*-możliwych czynników $f_1, \ldots, f_j$.

Warto w tym miejscu dodać, że model regresji logistycznej, wyznaczany wraz z właściwym składnikiem systematycznym, w którym zachowuje się pod pewnymi względami liniowość [Bartholomew i Knott 1999], jest następujący:

$$\pi_i\left(x_i, \ldots, x_k\right) = \frac{\exp L}{1 + \exp L}, \tag{1}$$

gdzie:

$$L = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}. \tag{2}$$

Z kolei składnik losowy to:

$$y \left| \left(x_{i1}, \ldots, x_{ik}\right) \right. \sim \text{Bernoulli}\left(\pi_i\left(x_{i1}, \ldots, x_{ik}\right)\right). \tag{3}$$

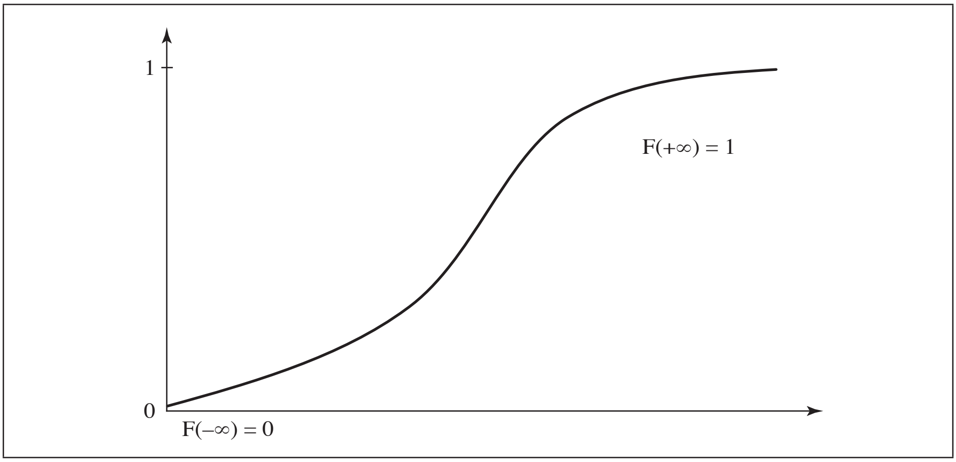
Składnik systematyczny określa, w jakim stopniu π_i zależy od *x*, zaś składnik losowy opisuje, jak *y* zmienia się w obrębie π_i[4]. Sam rozkład Bernoulliego (zwany też rozkładem dwumianowym)[5] wyznacza podejście, zgodnie z którym dana zmienna losowa może uzyskać wartość 1 z prawdopodobieństwem π_i i wartość 0 z prawdopodobieństwem równym $(1 - \pi_i)$. W ten sposób model ujęty w zapisie równań (1) i (2) odzwierciedla prawdopodobieństwo tego, że $y = 1$ jest dane funkcją wartości x_i.

---

[3] Przy funkcji logistycznej zakres prawdopodobieństwa odpowiedzi z przedziału [0, 1] przebiega od granicy dolnej (–∞) do granicy górnej (+∞) i odpowiada wyrazowi funkcji monotonicznej ze względu na dany rozkład odpowiedzi tworzących określony czynnik *f*.

[4] W liniowym modelu regresji składnik losowy jest określany przez rozkład reszt *e*. W modelu tym zmienna *y* to zmienna zależna (zmienna objaśniania), a *x* oznacza zmienną niezależną (zmienną objaśniającą). W modelu regresji logistycznej *y* jest zmienną opisaną na określonych kategoriach, np. w najprostszym ujęciu zmienna ta może przyjmować wartości binarne typu 0 i 1.

[5] Rozkład dwumianowy (w Polsce zwany też rozkładem Bernoulliego, choć w krajach anglojęzycznych termin *Bernoulli distribution* odnosi się do rozkładu zero-jedynkowego) to dyskretny rozkład prawdopodobieństwa, opisujący liczbę sukcesów *k* w ciągu *N* niezależnych prób, z których każda ma stałe prawdopodobieństwo sukcesu równe *p*.



Rys. 1. Graficzna prezentacja funkcji logistycznej

Źródło: opracowanie własne.

Funkcja wykładnicza wykorzystana w równaniu (1) $\exp(L) = M$ umożliwia transformację wykładniczą L do postaci logarytmicznej $\log(M) = L$. L określa łączną sumę wyników przy granicy górnej dodatnich odpowiedzi w zakresie rozpatrywanej pozycji (zmiennej obserwowalnej) w ten sposób, że:

$$\operatorname{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = L = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}\,. \tag{4}$$

W ramach dopasowania modelu potrzebna jest jedynie próbka niezależnych obserwacji i zbiór zmiennych, jak $\{y_i, x_{i1}, \ldots, x_{ik}, i = 1, \ldots, n\}$, gdzie wszystkie zmienne x i y będą wzajemnie niezależne. W dalszej kolejności dopasowanie modelu regresji logistycznej przebiega przykładowo za pomocą metody największej wiarygodności. Następnie wyliczane są odpowiednie szacunki α i β, co do których interpretację przeprowadza się w analogiczny sposób jak w standardowym modelu regresji liniowej, przy czym szacunki tego rodzaju dotyczą już wyrażenia $\operatorname{logit}(\pi_i)$, a nie jedynie wyrażenia π_i.

Z powyższych założeń należy wnioskować, że model regresji logistycznej stanowi bardzo ważny punkt odniesienia w początkowej fazie konstrukcji modelu czynnikowego. Sam model logitowy w kontekście proponowanego modelu czynnikowego **f** definiuje się już jako:

$$\operatorname{logit}\pi_i(\mathbf{f}) = \log_e \frac{\pi_i(\mathbf{f})}{1-\pi_i(\mathbf{f})} = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} f_j\,, \tag{5}$$

gdzie wyrażenie $\pi_i(\mathbf{f})$ to efekt transformacji logitowej, który pozwala sprowadzić model do postaci liniowej w zakresie rozpatrywanych czynników (tj. zmiennych ukrytych). W wyniku przekształcenia modelu (5) otrzymuje się wieloczynnikowy model w następującej formie:

$$\pi_i(\mathbf{f}) = \frac{\exp\left(\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij0} f_j\right)}{1 + \exp\left(\alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij0} f_j\right)}. \tag{6}$$

Jeśli zaś zakładany jest tylko jeden czynnik $j = 1$, wówczas zapis modelu (6) będzie odpowiadać modelowi jednoczynnikowemu w następującej formule [Bock 1972, Moustaki, Jöreskog i Mavridis 2004]:

$$\pi_i(f_1) = \frac{\exp(\alpha_{i0} + \alpha_{i1} f_1)}{1 + \exp(\alpha_{i0} + \alpha_{i1} f_1)}. \tag{7}$$

W literaturze – zgodnie z przesłankami wynikającymi z teorii reakcji na odpowiedzi – model jednoczynnikowy znany jest także pod nazwą modelu dwuparametrycznego. $\pi_i(f_1)$ traktuje się jako prawdopodobieństwo udzielenia przez respondenta odpowiedzi „tak" zgodnie z jego stopniem zdolności i poziomem wiedzy potrzebnej do rozwiązania danego zadania testowego.

4. Interpretacja parametrów modelu

W modelu czynnikowym dla zmiennych dychotomicznych dla każdej pozycji otrzymuje się $q + 1$ parametrów do oszacowania, wartość stałą α_{i0} oraz ładunki czynnikowe $\alpha_{i1}, \ldots, \alpha_{iq}$. Im wyższe są wartości ładunków czynnikowych α_{ij}, tym większy jest efekt j-czynnika w zakresie prawdopodobieństwa uzyskania pozytywnej odpowiedzi na daną pozycję. Jednocześnie im wyższa jest wartość α_{ij} dla danej pozycji, tym większa jest różnica w prawdopodobieństwach uzyskania prawidłowej/pozytywnej odpowiedzi pomiędzy dwiema badanymi jednostkami, znajdującymi się w pewnej odległości od siebie na ukrytych wymiarach. W rezultacie łatwiej można pokazać różnice pomiędzy dwiema jednostkami na podstawie ich odpowiedzi względem ocenianej pozycji. Jeśli wartości poszczególnych pozycji (nakładających się na dany czynnik) będą duże (np. wykraczające poza

3 lub 10), to będzie to oznaczać, że mamy do czynienia z bardzo silnym efektem progowym tych pozycji[6].

Mankamentem modeli czynnikowych dla zmiennych dychotomicznych są surowe ładunki czynnikowe. Ładunki te nie mają umownie przyjętych wartości w zakresie minimalnej (dolnej) i maksymalnej (górnej) granicy, przez co interpretacja zależności pomiędzy zmiennymi obserwowalnymi a czynnikami jest utrudniona. Wartości takich ładunków nie można więc porównać na wzór współczynników korelacji, które mieszczą się w przedziale od 0 do 1, tak jak ma to miejsce w klasycznej analizie czynnikowej. W związku z powyższym badacz musi je w pierwszej kolejności wystandaryzować[7].

Warto też dodać, że jeśli ładunki czynnikowe odznaczają się zbyt wysokimi wartościami błędów standardowych, to ich wartości są, po pierwsze, słabo określone przez model i najprawdopodobniej źle przyjęte kryteria doboru próby, a po drugie – świadczą o słabej jakości wygenerowanego czynnika.

5. Wskaźniki dobroci dopasowania danych do modelu czynnikowego

Po zakończeniu etapu związanego z interpretacją parametrów modelu należy ocenić dobroć dopasowania danych dychotomicznych do modelu czynnikowego. W tym celu można skorzystać z dwóch testów. Pierwszy test nosi nazwę globalnego testu dobroci dopasowania danych – G^2 i wyraża się wzorem:

$$G^2 = 2\sum_{r=1}^{2^p} O(r)\log_e \frac{O(r)}{E(r)}, \tag{8}$$

gdzie r oznacza zaakceptowany przez badacza w analizie wzorzec odpowiedzi, a $O(r)$ i $E(r)$ – obserwowane i oczekiwane częstości według wzorca odpowiedzi r.

Test ten został zaproponowany przez D.J. Bartholomewa i P. Tzamouraniego [1999]. Umożliwia on porównanie wartości obserwowanych częstości z wartościami oczekiwanymi. Dopasowanie modelu dokonuje się poprzez obliczenie wartości parametrów (odzwierciedlających położenie względem siebie rozkładów rozpatrywanych zmiennych). Im mniejsze są te wartości, tym lepsze jest dopasowanie modelu (tym większa jest dobroć dopasowania). Małe różnice oznaczają, że

[6] Wysokie wartości współczynników dyskryminacji świadczą jednocześnie o tym, że krzywe charakterystyczne pozycji odznaczają się stromym nachyleniem ze względu na znaczne różnice w dwóch binarnych wariantach odpowiedzi.

[7] Zob. punkt dotyczący ilustracji modelu jednoczynnikowego.

związki pomiędzy wszystkimi parami poszczególnych pozycji odpowiedzi zmiennych obserwowalnych są dobrze opisane i przewidziane przez model czynnikowy.

Alternatywnie można także przyjąć test oparty na Pearsonowskiej statystyce chi-kwadrat:

$$\chi^2 = \sum_{r=1}^{2^p} \frac{\left(O(r) - E(r)\right)^2}{E(r)}. \tag{9}$$

W obu wariantach (8 i 9) jeśli model odzwierciedla poziom dopasowania do danych, to wartości obu statystyk G^2 i χ^2 powinny być porównywalne. Statystyki G^2 i χ^2 traktowane są jako stały element badania rozkładu reszt kategorii zmiennych obserwowalnych. Do analizy danych dychotomicznych można wykorzystać podejście, zgodnie z którym reszty danej zmiennej rozpatrywane będą niezależnie od pozostałych reszt – jako statystyka G^2 i χ^2 z 1 stopniem swobody. Wówczas wartości statystyk G^2 i χ^2 większe niż 4 będą oznaczać słabe dopasowanie danych do modelu. Analiza reszt może być wykorzystana w celu identyfikacji oraz ewentualnej eliminacji tych zmiennych obserwowalnych, które nie znajdują w modelu czynnikowym dobrego dopasowania.

6. Ilustracja modelu jednoczynnikowego w zakresie badania postaw młodzieży wobec hedonistyczno-konsumpcyjnego stylu życia

W ramach ilustracji praktycznej funkcjonalności modelu jednoczynnikowego do analizy wybrano dane empiryczne pochodzące ze studium przeprowadzonego przez autora [Tarka 2010]. Zbiór danych obejmował 232 respondentów oraz 4 pozycje testowe – ujęte w formie dychotomicznych kategorii odpowiedzi: tak (1), nie (0) – z uwagi na następujące stwierdzenia:

1. Zakupy wpływają korzystnie na moje samopoczucie.
2. Jeśli nie stać mnie na zakupy produktów z powodu braku pieniędzy, zaciągam pożyczki.
3. Jeśli dysponuję już pieniędzmi, to najczęściej odwiedzam galerie handlowe.
4. Wydawanie pieniędzy (podczas zakupów w galerii handlowej) sprawia mi ogromną radość.

Przedmiotem analizy były postawy młodzieży wobec hedonistyczno-konsumpcyjnego stylu życia. Obliczenia przeprowadzono w programie LAMI (*Latent Model Interface*) w module Genlat. Jest to program, który umożliwia generowanie czynników na podstawie danych dychotomicznych. Jego słabą stroną jest ograniczenie w zakresie liczby generowanych czynników. Możliwe jest bowiem wyszczególnienie maksymalnie dwóch czynników[8].

Tabela 1. Zapis częstości występowania poszczególnych wzorców odpowiedzi

Wzorzec odpowiedzi	Częstości obserwowane	Częstości oczekiwane
1 1 1 1	101	102,22
0 0 0 0	63	61,86
0 1 1 1	14	14,12
0 0 1 1	14	8,50
0 0 0 1	8	10,07
1 1 1 0	7	5,83
0 0 1 0	5	8,91
0 1 0 0	4	3,35
0 1 1 0	7	6,26
1 0 1 1	3	3,06
0 1 0 1	1	0,00
1 1 0 1	3	1,40
1 1 0 0	1	0,00
1 0 0 0	1	0,00
–	232	225,59

Źródło: opracowanie na podstawie własnych badań w programie LAMI.

Po zakończeniu obliczeń, na podstawie wyników zamieszczonych w tabeli 1 (przedstawiającej częstości występowania odpowiedzi według określonych wariantów odpowiedzi), można wstępnie określić dwa charakterystyczne wzorce odpowiedzi: 1 1 1 1 („tak") i 0 0 0 0 („nie"). Oba warianty obejmują w sumie największe liczebności obserwacji. Pozostałe wzorce odpowiedzi stanowią jedynie marginalny wkład informacyjny do konstruowanego modelu. Odsetek odpowiedzi z uwzględnieniem obydwu kategorii odpowiedzi (tj. 1 i 0) według wszystkich czterech pozycji testowych wskazuje na przewagę wartości 1 (zob. tabela 2), co

[8] W programie w pliku wejściowym zdefiniowano w pierwszej kolejności zakres danych i poziom obliczeń. Polegało to m.in. na tym, że w wierszu pierwszym (oznaczonym N, NPB) określono liczbę respondentów N = 232 oraz liczbę zmiennych obserwowalnych mierzonych na podstawie danych binarnych (4) – (NPB). Z kolei w wierszu drugim, oznaczonym NFAC, INIT, ITER, PREC, SCOR, wyznaczono liczbę czynników (1) do wygenerowania w toku prowadzonej analizy – (NFAC), wybrano procedurę obliczeń w ramach czynnika (INIT), określono liczbę iteracji (maksimum 2000) – (ITER), określono poziom precyzji wyników (0,0000001) – (PREC) oraz przyjęto ustawione przez program domyślnie opcje wydruku wyników – (SCOR).

oznacza, że rozpatrywane 4 pozycje testowe są zbieżne i dostarczają informacji w ramach jednego wspólnego wymiaru. Ładunki czynnikowe tego modelu zaprezentowano w tabeli 3.

Tabela 2. Rozkłady odpowiedzi według poszczególnych pozycji

Wyszczególnienie	Pozycja testowa 1	Pozycja testowa 2	Pozycja testowa 3	Pozycja testowa 4
Kategoria 1	0,55	0,60	0,65	0,62
Kategoria 0	0,45	0,40	0,35	0,38

Źródło: opracowanie na podstawie własnych badań w programie LAMI.

Tabela 3 przedstawia wartości Alpha (surowe ładunki czynnikowe) dla każdej z czterech rozpatrywanych pozycji testowych. Dla przypomnienia – im wyższe są wartości tych ładunków, tym większy jest efekt danego czynnika w zakresie prawdopodobieństwa jego połączenia z daną pozycją testową. W analizowanym przykładzie wartości tych ładunków są bardzo wysokie i dodatnie (zob. kolumnę Alpha(1, I)), przez co jednoznacznie wskazują na jeden wspólny ukryty czynnik.

Tabela 3. Ładunki czynnikowe modelu w zakresie poszczególnych pozycji

Pozycje	Alpha (0, I)	Błąd standardowy	Alpha (1, I)	Błąd standardowy	$P(X = 1\|Z = 0)$	Standaryzowana Alpha
1	–0,04	0,34	6,79	0,84	0,49	0,99
2	1,49	0,29	6,22	0,74	0,82	0,99
3	2,29	0,66	5,81	0,42	0,91	0,98
4	1,18	0,36	5,52	0,54	0,77	0,97

Miara dobroci dopasowania danych do modelu – $G = 76{,}85\%$.

Źródło: opracowanie na podstawie własnych badań w programie LAMI.

W modelach czynnikowych dla zmiennych dychotomicznych surowe ładunki czynnikowe nie podlegają interpretacji, co oznacza, że nie można ich ze sobą w prosty sposób porównać. Ich wartości wykraczają poza umownie przyjęty przedział współczynników, np. korelacji od 0 do 1. Stąd potrzebna jest standaryzacja, w wyniku której otrzymuje się możliwość interpretacji ładunków czynnikowych w zakresie rozpatrywanych pozycji w relacji do badanego czynnika. Ostatnia kolumna w tabeli 3 zawiera wystandaryzowane wartości ładunków czynnikowych, które mieszczą się w przedziale od 0,97 do 0,99 i świadczą o wysokim poziomie ładunków czynnikowych opisujących poziom korelacji między daną pozycją (zmienną obserwowalną) a zmienną ukrytą (czynnikiem).

W praktyce badawczej zdarza się i tak, że duże wartości ładunków czynnikowych dla niektórych pozycji mogą być obarczone dużymi błędami standardowymi, co wynika z rozproszenia wartości odpowiedzi, jakie respondenci przydzielają poszczególnym pozycjom testowym. Wartości te w konsekwencji wpływają negatywnie na sam konstruowany czynnik[9]. Z zaprezentowanego przykładu wynika, że taką pozycją jest m.in. pozycja nr 1 „Zakupy wpływają korzystnie na moje samopoczucie”, której błąd standardowy wyniósł 0,84 przy wartości Alpha(1, I) równej 6,79. Ponadto w przypadku pozycji nr 1 wartość oszacowanego prawdopodobieństwa odpowiedzi ($P(X = 1|Z = 0)$ wyniosła jedynie 0,49. Oszacowane prawdopodobieństwa określają w tym wypadku, na ile jednostki badane w próbie są zgodne w zakresie pozytywnych odpowiedzi względem czterech rozpatrywanych pozycji.

Ostatecznie, kierując się miarą dobroci dopasowania danych do modelu, tj. rozpatrując wskaźnik G (który osiągnął poziom 76,85%), można stwierdzić, że jakość tego modelu w układzie czterech pozycji testowych kształtuje się na dobrym poziomie. Przeanalizowane w ten sposób pozycje należy traktować jako odpowiednie składowe czynnika charakteryzującego „hedonistyczno-konsumpcyjny styl życia” młodzieży.

7. Podsumowanie

Cechy poszczególnych jednostek badawczych mogą być obserwowalne i nieobserwowalne. Pierwsze wyróżniają się poprzez fakt, że można je zmierzyć, drugie natomiast dotyczą ukrytych postaw, opinii, emocji i podświadomych wyborów jednostek, przez co ich bezpośredni pomiar jest niemożliwy. Analiza zmiennych ukrytych (a tym samym rozpoznanie zjawisk rynkowych o charakterze ukrytym) jest trudnym, ale też bardzo pożądanym przedsięwzięciem. Sama analiza oparta na modelach czynnikowych dla zmiennych dychotomicznych jest bardzo istotna dla dalszego rozwoju badań i analiz marketingowych. Jest ona szczególnie przydatna wobec rosnącej liczby danych jakościowych (pochodzących z takich obszarów analiz, jak pomiar reakcji konsumentów na reklamę, deklaracji konsumenckich, wzorców postaw i zachowań). Dane tego typu dotyczą coraz częściej pomiaru cech jednostek z wykorzystaniem skal słabych, tj. skal porządkowych i skal nominalnych. Zaprezentowany w pracy model czynnikowy dla zmiennych dychotomicznych można zatem uznać za niezwykle przydatny w opisie nieobserwowalnych zjawisk rynkowych na gruncie tradycyjnego pomiaru, tj. skali nominalnej.

[9] Czynnik taki jest niespójny i tym samym słabo zdefiniowany przez model.

Literatura

Bartholomew D.J. [1980], *Factor Analysis of Categorical Data*, „Journal of the Royal Statistical Society", vol. 42, nr 3.

Bartholomew D.J., Knott M. [1999], *Latent Variable Models and Factor Analysis*, wyd. 2, Arnold, London.

Bartholomew D.J., Tzamourani P. [1999], *The Goodness-of-fit of Latent Trait Models in Attitude Measurement*, „Sociological Methods and Research", vol. 27, nr 4.

Bock R.D. [1972], *Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories*, „Psychometrika", vol. 37, nr 1.

Gorsuch R.L. [1974], *Factor Analysis*, W.B. Saunders Co., Toronto.

Jöreskog K.G., Moustaki I. [2001], *Factor Analysis of Ordinal Variables: A Comparison of Three Approaches*, „Multivariate Behavioral Research", vol. 36, nr 3.

Moustaki I., Jöreskog K.G., Mavridis D. [2004], *Factor Model for Ordinal Variables with Covariate Effects on the Manifest and Latent Variables: A Comparison LISREL and IRT Approaches*, „Structural Equation Modeling", vol. 11, nr 4.

Moustaki I., Knott M. [2000], *Generalized Latent Trait Models*, „Psychometrika", vol. 65, nr 3.

Moustaki I., Steele F. [2005], *Latent Variable Models for Mixtures of Categorical and Duration Responses, with an Application to Fertility Preferences and Family Planning in Bangladesh*, „Statistical Modelling", vol. 5, nr 4.

Sagan A. [2000], *Wybrane problemy identyfikacji i pomiaru struktur ukrytych*, Zeszyty Naukowe Akademii Ekonomicznej w Krakowie, nr 543, Kraków.

Sikorska I. [2012], *Analiza zmiennych ukrytych* [w:] *Zaawansowane metody analiz statystycznych*, red. E. Frątczak, Szkoła Główna Handlowa w Warszawie, Warszawa.

Tarka P. [2010], *Latent Variable Models – Issues on Measurement and Finding Exact Constructs*, „Przegląd Statystyczny", nr 4.

Factor Analysis for Binary Data in the Analysis of the Latent Structure of Market Phenomena

The article describes a number of theoretical and practical issues pertaining to the factor analysis model and its application in the field of marketing research. The first part discusses the dependence of marketing research on the theory of measurement e.g., the course of research development in the sphere of latent events analysis. The next part looks at factorial modeling in research on consumer attitudes and opinions, followed by a discussion of the possibilities for binary data analysis using the latent trait model. Finally, using empirical data, the author presents a one-factor model to measure youth (young consumers) attitudes towards a hedonistic-consumer lifestyle.

Keywords: factor analysis, latent variables, binary data, nominal scale.