

Iwona Bąk, Agnieszka Sompolska-Rzechuła

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

**WYKORZYSTANIE ANALIZY LOG-LINIOWEJ
DO WYBORU CZYNNIKÓW
DETERMINUJĄCYCH WYJAZDY TURYSTYCZNE
WYBRANEJ KATEGORII
GOSPODARSTW DOMOWYCH**

Streszczenie: Celem artykułu jest próba wyodrębnienia zmiennych, które wpływają na podjęcie decyzji o wyjeździe turystycznym w gospodarstwach domowych emerytów i rencistów. Informacje dotyczące aktywności turystycznej gospodarstw domowych emerytów i rencistów zaczerpnięto z badań ankietowych „Turystyka i wypoczynek w gospodarstwach domowych”, przeprowadzonych przez GUS w 2005 roku. Ponieważ w badaniu wzięto pod uwagę zmienne kategoryzacyjne, zatem do wyboru optymalnego zbioru czynników decydujących o wyjeździe turystycznym wykorzystano analizę log-liniową.

Słowa kluczowe: analiza log-liniowa, emeryci i renciści, wyjazdy turystyczne.

1. Wstęp

Wydłużanie się średniej długości ludzkiego życia oraz systematyczny spadek urodzeń sprawia, że zwiększa się odsetek ludzi starszych. Rosnąca ich liczba może stać się potężnym źródłem koniunktury dla szeroko pojętego przemysłu turystycznego. Ludzi w starszym wieku należy zachęcać do uprawiania turystyki o każdej porze roku i dążyć do wyrobienia trwałego nawyku aktywności. Pozytywny wpływ aktywności turystycznej na zdrowie i samopoczucie starszych osób jest niezaprzeczalny i potwierdzony licznymi wynikami badań. Wiele podmiotów gospodarczych funkcjonujących na rynku usług turystycznych działa intuicyjnie, bez znajomości grup docelowych, do których adresowane są poszczególne produkty rynkowe. Badania prowadzone w sferze konsumpcji turystycznej mogą dostarczyć wielu cennych informacji o motywach i zachowaniach turystów w wieku podeszłym i o ograniczonej sprawności oraz o wielkości wydatków przeznaczonych przez tę grupę na wypoczynek [Bąk 2010].

Celem artykułu jest próba wyodrębnienia zmiennych, które wpływają na podjęcie decyzji o wyjeździe turystycznym w gospodarstwach domowych emerytów i rencistów. Informacje dotyczące aktywności turystycznej gospodarstw domowych emerytów i rencistów zaczerpnięto z badań ankietowych „Turystyka i wypoczynek w gospodarstwach domowych”, przeprowadzonych przez Główny Urząd Statystyczny w 2005 roku. Dane mają charakter reprezentacyjny i pochodzą z badań cyklicznych przeprowadzanych co cztery lata. Ankieta zawierała pytania charakteryzujące gospodarstwa domowe pod względem: liczby osób, liczby dzieci, sytuacji finansowej oraz wyposażenia w sprzęt turystyczny.

2. Metoda

Ze względu na to, że w badaniu wzięto pod uwagę głównie zmienne kategoryzacyjne, do wyboru optymalnego zbioru czynników determinujących wyjazd turystyczny wykorzystano analizę log-liniową. Pozwala ona na dokładniejszy opis zależności między zmiennymi kategoryzacyjnymi w porównaniu z miernikami stosowanymi do oceny współzależności cech jakościowych¹. Dodatkowym atutem analizy log-liniowej jest możliwość uwzględnienia wpływu interakcji pomiędzy zmiennymi [Stanisz 2007, s. 269-276].

Punktem wyjścia w analizie log-liniowej są liczebności oczekiwane dla poszczególnych poziomów zmiennych kategoryzacyjnych [Dobosz 2004, s. 365-376]. Jeżeli człony interakcyjne są nieistotne, a liczebności w klasach zależą tylko od czynników głównych, to oznacza to, że brak jest wzajemnych oddziaływań pomiędzy zmiennymi kategoryzacyjnymi.

Dopasowywany model log-liniowy ma charakter hierarchiczny, co oznacza, że jeżeli określony człony interakcyjny jest włączony do modelu, to wszystkie pozostałe kombinacje czynników występujących w tym członie muszą być uwzględnione w modelu.

W modelu log-liniowym przyjmuje się, że logarytm naturalny wartości oczekiwanej liczebności w komórce w tabeli niezależności jest liniową funkcją czynników. Uogólnioną postać modelu wyraża się równaniem:

$$\ln(\hat{n}^{T\dots}) = \bar{n} + \sum \lambda_i^T + \dots + \sum \lambda_{ij}^{TS} + \dots + \sum \lambda_{ijk}^{TSQ} + \dots,$$

gdzie: \hat{n}^T – wektor $[\hat{n}_i^T \hat{n}_{ij}^{TS} \dots]$, przy czym \hat{n}_i^T jest oczekiwaną liczebnością

i -tego poziomu czynnika T ,

\hat{n}_{ij}^{TS} – przewidywana liczebność dotycząca interakcji i -tego poziomu czynnika T oraz j -tego poziomu czynnika S itd.

¹ Przykład zastosowania analizy log-liniowej można znaleźć m.in. w pracy [Salamaga 2008].

Wartość \bar{n} – średniej z logarytmów naturalnych wszystkich obserwowanych liczebności – wyznacza się na podstawie wzoru:

$$\bar{n} = \frac{1}{n} \sum_T \sum_i \ln(n_i^T), \quad (2)$$

gdzie: \sum_T – sumowanie po wszystkich czynnikach,

\sum_i – sumowanie po wszystkich poziomach czynników,

λ_i^T – wskaźnik i -tego poziomu czynnika T ,

λ_{ij}^{TS} – wskaźnik interakcji rzędu drugiego i -tego poziomu czynnika T oraz j -tego poziomu czynnika S ,

λ_{ijk}^{TSQ} – wskaźnik interakcji rzędu trzeciego pomiędzy i -tym poziomem czynnika T , j -tym poziomem czynnika S oraz k -tym poziomem czynnika Q itd.

Poprawnie zbudowany model log-liniowy umożliwia najlepszą predykcję liczebności przy uwzględnieniu w modelu jak najmniejszej liczby interakcji. Uwzględnienie wszystkich czynników i ich interakcji daje model najlepiej dopasowany, lecz nie zawsze wygodny w praktyce, gdyż wpływ niektórych czynników i interakcji może być niewielki w porównaniu z pozostałymi składnikami modelu. Miarą dopasowania modelu do wyników są statystyki χ^2 Pearsona oraz χ^2 największej wiarygodności.

Po ustaleniu, jakiego rzędu interakcje należy włączyć do modelu, pojawia się problem ustalenia, które interakcje tego rzędu (pomiędzy którymi czynnikami) powinny być wybrane do modelu. W rozstrzygnięciu tego problemu korzysta się z zależności cząstkowych i brzegowych. Zależność cząstkowa informuje o tym, czy odpowiednia interakcja jest istotna, gdy wszystkie inne czynniki tego samego stopnia są już w modelu. Zależność brzegowa informuje, czy odpowiednia interakcja ma wpływ, gdy w modelu nie ma jeszcze żadnych interakcji tego samego rzędu.

3. Wyniki badania

Ze względu na cel artykułu wyróżniono cechę, zależną zdefiniowaną jako wyjazd turystyczny (Y). Jest to zmienna dychotomiczna, której warianty (wystąpienie wyjazdu turystycznego lub nie) zmierzono na skali nominalnej. W zbiorze zmiennych niezależnych uwzględniono:

X_1 – liczbę osób w gospodarstwie domowym (do 2 osób, od 3 do 5 osób, powyżej 5 osób),

X_2 – dzieci w gospodarstwie domowym (tak, nie),

- X_3 – pracujący (tak, nie),
 X_4 – liczbę osób pobierających emeryturę lub rentę (od 1 do 2 osób, powyżej 2 osób),
 X_5 – dochód gospodarstwa domowego w przeliczeniu na jednego członka (do 1 tys. zł, od 1 do 2 tys. zł, powyżej 2 tys. zł),
 X_6 – posiadanie namiotu (tak, nie),
 X_7 – posiadanie działki rekreacyjnej (tak, nie),
 X_8 – posiadanie samochodu osobowego (tak, nie),
 X_9 – posiadanie roweru (tak, nie),
 X_{10} – posiadanie plecaka turystycznego, śpiwora, materaca (tak, nie).

W pierwszym kroku zbadano, które zmienne wykazują istotny związek ze zmienną zależną. Do tego celu wykorzystano test niezależności χ^2 . W tabeli 1 podano wartości statystyki testującej χ^2 oraz prawdopodobieństwa testowe p .

Tabela 1. Wartości statystyki χ^2 oraz prawdopodobieństwa testowe p

Zmienna	Wartość statystyki χ^2	Wartość prawdopodobieństw testowych p	Zmienna	Wartość statystyki χ^2	Wartość prawdopodobieństw testowych p
X_1	27,99	0,00000	X_6	71,55	0,00000
X_2	14,13	0,00017	X_7	10,81	0,00101
X_3	44,48	0,00000	X_8	69,34	0,00000
X_4	0,18	0,67174	X_9	5,80	0,01598
X_5	47,89	0,00000	X_{10}	130,02	0,00000

Źródło: obliczenia własne.

Na podstawie przeprowadzonych obliczeń można stwierdzić, że zmienna określająca wystąpienie wyjazdu turystycznego nie jest powiązana tylko ze zmienną dotyczącą liczby osób pobierających emeryturę lub rentę, dlatego zmienna ta została wykluczona z badania.

W celu znalezienia czynników wpływających istotnie na podjęcie decyzji o wyjeździe turystycznym w gospodarstwach domowych emerytów i rencistów dokonano podziału analizowanych zmiennych na dwie grupy. W pierwszej znalazły się zmienne charakteryzujące sytuację społeczno-ekonomiczną gospodarstwa domowego:

- X_1 – liczba osób w gospodarstwie domowym,
 X_2 – dzieci w gospodarstwie domowym,
 X_3 – pracujący,
 X_5 – dochód gospodarstwa domowego w przeliczeniu na jednego członka.

Natomiast pozostałe zmienne dotyczyły wyposażenia gospodarstwa w sprzęt turystyczny i weszły w skład grupy drugiej.

W celu dokonania specyfikacji modelu log-liniowego należy określić rząd interakcji zmiennych. W tym celu wykonano testowanie wszystkich interakcji. Wyniki

zamieszczono w tab. 2, z której wynika, że w modelu log-liniowym, oprócz czynników głównych, należy uwzględnić interakcje maksymalnie trzeciego rzędu. Wskazują na to wartości prawdopodobieństw testowych p , odpowiadające wartościom statystyk testowych χ^2 największej wiarygodności i χ^2 Pearsona. Aby ocenić, które interakcje czynników powinny być włączone do modelu log-liniowego, wykorzystano testy cząstkowe i brzegowe (tab. 3).

Tabela 2. Wyniki testów dla interakcji między zmiennymi

Stopień interakcji	Stopnie swobody	Wartość χ^2 NW	Prawdopodobieństwo p	Wartość χ^2 Pearsona	Prawdopodobieństwo p
1	7	3141,833	0,0000	5704,503	0,0000
2	19	1676,762	0,0000	2352,750	0,0000
3	25	58,776	0,0002	76,709	0,0000
4	16	5,869	0,9895	6,479	0,9821
5	4	0,241	0,9933	0,242	0,9932

Źródło: obliczenia własne.

Pogrubioną czcionką zaznaczono w tab. 3 interakcje, dla których zależności cząstkowe i brzegowe są istotne. Rezultaty testów cząstkowych i brzegowych wskazują na potrzebę uwzględnienia w modelu log-liniowym pięciu czynników głównych oraz tych interakcji rzędu drugiego i trzeciego, które występują pomiędzy zmiennymi określonymi jako niezależne a zmienną zależną. Aby uniknąć braku dopasowania związanego z usunięciem interakcji pomiędzy zmiennymi niezależnymi, do modelu należy włączyć efekt reprezentujący wszystkie interakcje pomiędzy zmiennymi niezależnymi [Stanisz 2007, s. 287]. Ostatecznie uwzględniono zależności pomiędzy wyjazdem turystycznym a:

- liczbą osób w gospodarstwie domowym,
- dziećmi w gospodarstwie domowym,
- pracującymi,
- dochodem gospodarstwa domowego w przeliczeniu na jednego członka,
- dziećmi i pracującymi,
- pracującymi i dochodem gospodarstwa domowego w przeliczeniu na jednego członka.

Model log-liniowy przyjmuje wówczas następującą postać:

$$\ln(\hat{n}_{ijklm}) = \bar{n} + \lambda_i^Y + \lambda_j^{X_1} + \lambda_k^{X_2} + \lambda_l^{X_3} + \lambda_m^{X_5} + \lambda_{ij}^{YX_1} + \lambda_{ik}^{YX_2} + \lambda_{il}^{YX_3} + \lambda_{im}^{YX_5} + \lambda_{ijm}^{YX_3X_5} + \lambda_{iklm}^{X_1X_2X_3X_5}.$$

Oszacowany model jest dobrze dopasowany do danych empirycznych, świadczą o tym wartości statystyki χ^2 największej wiarygodności, która wynosi 15,155

($p = 0,938$), oraz statystyki χ^2 Pearsona, która równa się 14,962 ($p = 0,942$). Wartości obu statystyk nie są istotne, zatem model został dobrze dopasowany do danych empirycznych. Potwierdza to także rys. 1 przedstawiający liczebności obserwowane względem liczebności dopasowanych.

Tabela 3. Wyniki testów związków cząstkowych i brzegowych między czynnikami

Czynniki	Stopnie swobody	Związek cząstkowy χ^2	Wartość p w związku cząstkowym	Związek brzegowy χ^2	Wartość p w związku brzegowym
Y	2	851,379	0,000	851,379	0,000
X_1	1	892,054	0,000	892,054	0,000
X_2	1	3,948	0,047	3,948	0,047
X_3	2	1384,454	0,000	1384,454	0,000
X_5	1	9,999	0,002	9,999	0,002
YX_1	2	429,715	0,000	708,320	0,000
YX_2	2	416,598	0,000	647,080	0,000
YX_3	4	95,120	0,000	166,013	0,000
YX_5	2	16,159	0,000	27,675	0,000
X_1X_2	1	1,827	0,176	217,927	0,000
X_1X_3	2	18,191	0,000	97,324	0,000
X_1X_5	1	7,310	0,007	13,831	0,000
X_2X_3	2	7,469	0,024	28,571	0,000
X_2X_5	1	18,111	0,000	43,819	0,000
X_3X_5	2	70,226	0,000	46,077	0,000
YX_1X_2	2	0,380	0,827	1,022	0,600
YX_1X_3	4	2,946	0,567	3,982	0,408
YX_1X_5	2	2,127	0,345	1,687	0,430
YX_2X_3	4	28,047	0,000	27,825	0,000
YX_2X_5	2	0,858	0,651	0,347	0,841
YX_3X_5	4	21,910	0,000	15,101	0,004
$X_1X_2X_3$	2	0,479	0,787	5,005	0,082
$X_1X_2X_5$	1	0,356	0,551	2,291	0,130
$X_1X_3X_5$	2	0,064	0,968	0,388	0,823
$X_2X_3X_5$	2	6,734	0,034	0,142	0,931
$YX_1X_2X_3$	4	0,896	0,925	0,573	0,966
YX_1XX_5	2	0,451	0,798	0,810	0,667
$YX_1X_3X_5$	4	2,282	0,684	2,857	0,582
$YX_2X_3X_5$	4	1,528	0,822	1,636	0,802
$X_1X_2X_3X_5$	2	0,068	0,966	1,004	0,605

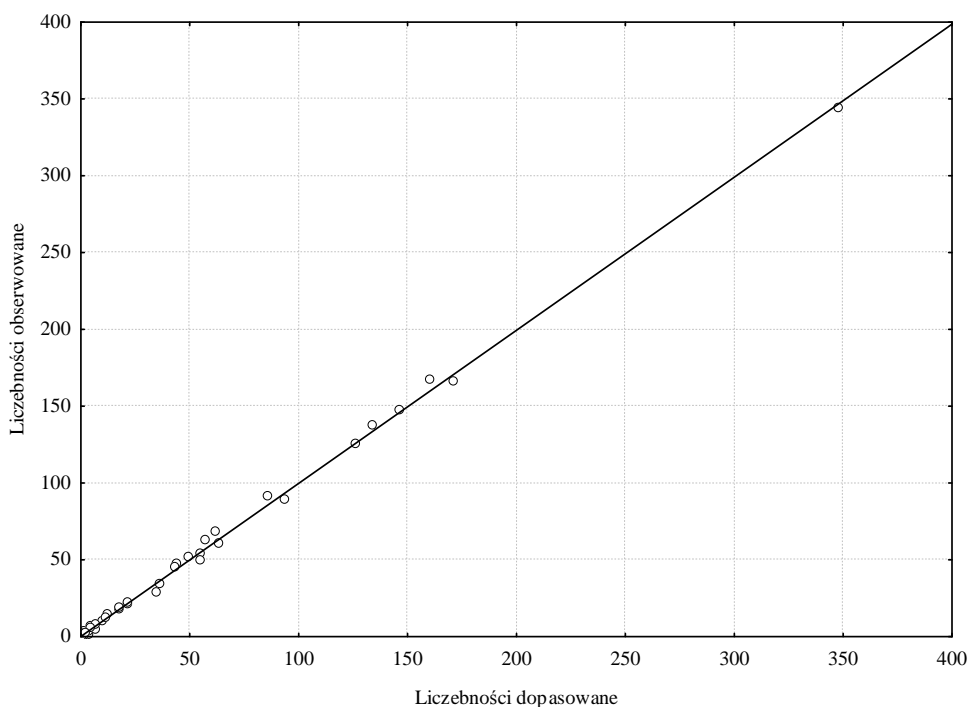
Źródło: obliczenia własne.

Na podjęcie decyzji dotyczącej wyjazdu turystycznego w gospodarstwach domowych emerytów i rencistów mogą mieć również wpływ interakcje rzędu drugiego dla następujących par zmiennych (tab. 3):

- liczba osób w gospodarstwie domowym i pracujący,

- liczba osób w gospodarstwie domowym i dochód gospodarstwa domowego w przeliczeniu na jednego członka,
- dzieci w gospodarstwie domowym i pracujący,
- dzieci w gospodarstwie domowym i dochód gospodarstwa domowego w przeliczeniu na jednego członka,
- pracujący i dochód gospodarstwa domowego w przeliczeniu na jednego członka.

Wynika stąd, że czynnikiem, który najczęściej wchodzi w istotne interakcje z innymi zmiennymi determinującymi podjęcie decyzji o wyjeździe turystycznym jest dochód gospodarstwa domowego w przeliczeniu na jednego członka.



Rys. 1. Wykres rozrzutu liczebności obserwowanych względem liczebności dopasowanych

Źródło: opracowanie własne.

W celu stwierdzenia, które czynniki charakteryzujące wyposażenie gospodarstw domowych w sprzęt turystyczny wpływają na podjęcie decyzji o wyjeździe turystycznym, ponownie przeprowadzono analizę log-liniową. Dokonano testowania wszystkich interakcji i okazało się, że w modelu, oprócz czynników głównych, należy uwzględnić interakcje maksymalnie drugiego rzędu. Aby sprawdzić, które z dwuwymiarowych zależności są istotne, skorzystano z zależności cząstkowych i brzegowych. Ostatecznie otrzymano następujący model:

$$\ln(\hat{n}_{(ijklmn)}) = \bar{n} + \lambda_i^Y + \lambda_j^{X_6} + \lambda_k^{X_7} + \lambda_l^{X_8} + \lambda_m^{X_9} + \lambda_n^{X_{10}} + \\ + \lambda_{ij}^{YX_6} + \lambda_{ik}^{YX_7} + \lambda_{il}^{YX_8} + \lambda_{im}^{YX_9} + \lambda_{in}^{YX_{10}} + \lambda_{jklmn}^{X_6X_7X_8X_9X_{10}}.$$

W modelu uwzględniono zależności pomiędzy wyjazdem turystycznym i posiadaniem:

- namiotu (X_6),
- działki rekreacyjnej (X_7),
- samochodu osobowego (X_8),
- roweru (X_9),
- plecaka turystycznego, śpiwora, materaca (X_{10}).

Oszacowany model jest dobrze dopasowany do danych empirycznych, świadczą o tym wartość statystyki χ^2 największej wiarygodności, która wynosi 15,697 ($p = 0,958$), oraz wartość statystyki χ^2 Pearsona, która równa się 14,963 ($p = 0,970$).

4. Wnioski

Zastosowanie modelu log-liniowego umożliwiło wykrycie istotnych związków pomiędzy zmienną zależną a analizowanymi w artykule zmiennymi niezależnymi. Z przeprowadzonych badań wynika, że wszystkie zmienne dotyczące zarówno sytuacji społeczno-ekonomicznej, jak i wyposażenia w sprzęt turystyczny gospodarstw domowych emerytów i rencistów wpływają istotnie na wystąpienie wyjazdu turystycznego. Statystycznie istotne okazały się jedynie interakcje rzędu drugiego, wyjątek stanowią interakcje rzędu trzeciego, dotyczące powiązań zmiennej zależnej z dziećmi i pracującymi oraz z pracującymi i dochodem gospodarstwa domowego w przeliczeniu na jednego członka. Na podjęcie decyzji dotyczącej wyjazdu turystycznego w gospodarstwach domowych emerytów i rencistów mogą mieć również wpływ interakcje pomiędzy zmiennymi niezależnymi, opisującymi zarówno sytuację społeczno-ekonomiczną, jak i wyposażenie gospodarstwa w sprzęt turystyczny.

Wykorzystanie modelu log-liniowego pozwoliło na bardziej dokładny opis zależności między zmiennymi kategoryzacyjnymi w porównaniu z miernikami stosowanymi do oceny współzależności cech jakościowych.

Literatura

- Bąk I., *Badanie wyjazdów turystycznych emerytów i rencistów w Polsce – analiza statystyczna*, [w:] *Potencjał turystyczny – zagadnienia ekonomiczne*, Zeszyty Naukowe nr 591, Ekonomiczne Problemy Usług nr 53, Wyd. Naukowe Uniwersytetu Szczecińskiego, Szczecin 2010, s. 347-359.
- Dobosz M., *Wspomagana komputerowo statystyczna analiza wyników badań*, Akademicka Oficyna Wydawnicza ELIT, Warszawa 2004, s. 365-377.

Salamaga M., *Wykorzystanie analizy log-liniowej do wyboru czynników opisujących sytuację ekonomiczną gospodarstw domowych*, „Przegląd Statystyczny” 2008, t. 55, nr 4, s. 40-51.

Stanisz A., *Przystępny kurs statystyki z zastosowaniem Statistica PL na przykładach z medycyny*, t. 3, *Analizy wielowymiarowe*, StatSoft, Kraków 2007, s. 269-307.

THE USE OF LOG-LINEAR ANALYSIS FOR THE SELECTION OF FACTORS DETERMINING THE JOURNEYS OF SELECTED CATEGORY OF HOUSEHOLDS

Summary: The article attempts to select variables which influence the decision to go on holiday in pensioners' households. The information on tourist activities of pensioners' households has been taken from the survey entitled *Tourism and Recreation in Households*, conducted by the Central Statistical Office (GUS) in 2005. A log-linear analysis was used for the purpose of selecting the optimal set of factors determining the decision to go on holiday, as categorical variables were taken into account in the study.