



ORIGINAL PAPER / PRACA ORYGINALNA

Agnieszka Ćwirlej-Sozańska^{1 (A,B,C,D,E,F,G,H)}, Bernard Sozański^{2 (A,D,E,F)},
Anna Wilmowska-Pietruszyńska^{1 (A,C,F,H)}

**Application of the log-linear analysis to choose determinants of disability
among the elderly residents of south-eastern Poland**

**Zastosowanie analizy log – liniowej do wyboru czynników determinujących
niepełnosprawność wśród starszych mieszkańców Polski
południowo-wschodniej**

¹ Institute of Physiotherapy, Faculty of Medicine, University of Rzeszow, Poland

² Center for Innovative Research in Medical and Natural Sciences, Faculty of Medicine,
University of Rzeszow, Poland

ABSTRACT

Introduction: Many variables in research in the area of medical and health sciences are qualitative in nature. A common statistical tool used to analyze them is the χ^2 test. However, it does not allow us to assess the relationship between a number of variables and distinguish the factors determining the investigated phenomenon. A more accurate tool is the log-linear analysis, which enable the researcher to evaluate the dependences and interactions between the studied variables.

Purpose: Description of the use of log-linear analysis on the example of the cross-sectional study on disability of the elderly.

Material and methods: The assessment of disability and the choice of the factors that determine it was carried out on the

STRESZCZENIE

Wstęp: Wiele zmiennych w badaniach z obszaru nauk medycznych oraz nauk o zdrowiu ma charakter jakościowy. Powszechnym narzędziem statystycznym stosowanym do ich analizy jest test χ^2 . Nie pozwala on jednak na dokonanie oceny powiązań między większą liczbą zmiennych oraz wyodrębnienie czynników determinujących badane zjawisko. Dokładniejszym narzędziem jest analiza log-liniowa, pozwalająca na ocenę zależności i interakcji między badanymi zmiennymi.

Cel: Przedstawienie zastosowania analizy log-liniowej na przykładzie przekrojowego badania niepełnosprawności osób starszych.

Materiał i metody: Ocenę niepełnosprawności oraz wybór czynników, które ją determinują przeprowadzono na wyni-

Mailing address / Adres do korespondencji: Agnieszka Ćwirlej-Sozańska, Instytut Fizjoterapii, ul. Warszawska 26A, 35-205 Rzeszów, sozanska@ur.edu.pl

Participation of co-authors / Udział współautorów: A – Author of the concept and objectives of paper / autor koncepcji i założeń pracy; B – collection of data / zbieranie materiału; C – implementation of research / realizacja badań; D – elaborate, analysis and interpretation of data / opracowanie, analiza i interpretacja wyników; E – statistical analysis / analiza statystyczna danych; F – preparation of a manuscript / przygotowanie manuskryptu; G – working out the literature / opracowanie piśmiennictwa; H – obtaining funds / pozyskanie funduszy

Received / Artykuł otrzymano: 20.08.2016 | Accepted / Zaakceptowano do publikacji: 8.12.2016

Publication date / Data publikacji: december / grudzień 2016

results of a survey of 800 randomly selected people aged 71-80 years from the area of south-eastern Poland. The research tool was a WHODAS 2.0 questionnaire and a respondent's particulars. The log-linear model was used for the analysis. In order to evaluate the fitting of the model, the Pearson's χ^2 and the χ^2 maximum likelihood statistics, R^2 and A coefficients were used.

Results: Education, adjustments of a house / flat, physical exercises have a significant impact on the prevalence of disability in the study group.

Conclusions: The log-linear analysis allows us to determine the effect not only of individual variables on the formation of an independent variable, but also their interactions and the determination the odds occurrence of a dependent variable according to different qualitative categories of dependent variables. The information obtained in this way is a valuable clue to take practical action to decrease or increase the severity of the studied phenomenon.

Keywords: disability, elderly, log-linear model, Poland

Introduction

Plenty of variables analyzed in the study from the area of medical science and health sciences are qualitative in nature. Basic statistical methods for analyzing qualitative data are based on a tool developed by Karl Pearson, which is a test χ^2 [1]. However, these methods do not allow the researchers to assess the relationship between three or more variables, because their way of dependence analysis of several features is reduced to examine all possible combinations of two quality characteristics on the basis of which value of the χ^2 test of independence is calculated [2]. Information received on the basis of such analyzes is very valuable, but it does not allow us to differentiate crucial factors determining the studied phenomenon among more variables. A more accurate statistical tool is the log-linear analysis, which provides a more precise description of the relation between the studied qualitative variables and it let us assess the impact of the interaction between the variables [3].

This article presents an example of using the log-linear model to determine the most important factors that affect the occurrence of disability among the elderly. Ageing of the body is associated with progressive impairment of bodily functions and reducing the adaptive capacity of the body in relation to environmental changes. It provides to some limitations associated with activities of varying complexity undertaken by the elderly in everyday life. Each person, depending on the environment in which he/she lives and available resources at his/her disposal, may experience different restrictions. Widely understood environmental factors are a key element determining the level of disability [4,5]. The use of more sophisticated methods of statistical data analysis, which allow us to isolate the factors determining the prevalence of disability, provides a precise determination of the most important

of the study 800 randomly selected people aged 71-80 years from the area of south-eastern Poland. The research tool was a WHODAS 2.0 questionnaire and a respondent's particulars. The log-linear model was used for the analysis. In order to evaluate the fitting of the model, the Pearson's χ^2 and the χ^2 maximum likelihood statistics, R^2 and A coefficients were used.

Results: Education, adjustments of a house / flat, physical exercises have a significant impact on the prevalence of disability in the study group.

Conclusions: The log-linear analysis allows us to determine the effect not only of individual variables on the formation of an independent variable, but also their interactions and the determination the odds occurrence of a dependent variable according to different qualitative categories of dependent variables. The information obtained in this way is a valuable clue to take practical action to decrease or increase the severity of the studied phenomenon.

Keywords: disability, elderly, log-linear model, Poland

Wstęp

Bardzo dużo zmiennych analizowanych w badaniach z obszaru nauk medycznych oraz nauk o zdrowiu ma charakter jakościowy. Podstawowe statystyczne metody analizy danych jakościowych są oparte na narzędziu stworzonym przez Karla Pearsona, jakim jest test χ^2 [1]. Metody te nie pozwalają jednak na dokonanie oceny powiązań między trzema lub większą liczbą zmiennych, gdyż oparty na nich sposób analizy zależności kilku cech jest sprawdzany do badania wszelkich możliwych kombinacji dwóch cech jakościowych, na podstawie których jest obliczana wartość testu niezależności χ^2 [2]. Informacje otrzymane na podstawie takich analiz są bardzo wartościowe, ale nie pozwalają wyodrębnić kluczowych czynników determinujących badane zjawisko spośród większej liczby zmiennych. Dokładniejszym narzędziem statystycznym jest analiza log-liniowa, pozwalająca na bardziej precyzyjny opis zależności między badanymi zmiennymi kategoryzacyjnymi oraz ocenę wpływu interakcji pomiędzy zmiennymi [3].

W niniejszym artykule przedstawiono przykład wykorzystania modelu log-liniowego do określenia czynników mających najważniejszy wpływ na występowanie zjawiska niepełnosprawności wśród osób starszych. Starzenie się organizmu związane jest z postępującym upośledzeniem funkcji życiowych oraz zmniejszaniem się możliwości adaptacyjnych organizmu w stosunku do zmian środowiska. Przekłada się to na ograniczenie wykonywania przez osoby starsze czynności o różnym stopniu złożoności w życiu codziennym. Każdy człowiek, zależnie od otoczenia, w którym żyje oraz od zasobów, jakimi dysponuje, doświadcza innych ograniczeń. Szeroko rozumiane czynniki środowiskowe są kluczowym elementem stanowiącym o poziomie niepełnosprawności [4, 5]. Zastosowanie bardziej skomplikowanych metod statystycznej analizy danych pozwalających na wyodręb-

variables and designs targeted preventive interventions arranged in proper order.

The log-linear analysis is often used in economic research, for example in the case of a choice of factors affecting the particular economic situation of the analysed set [6,7]. However, it should be more frequently applied to the research in the area of medical and health sciences, where many of the phenomena are qualitative in nature. Its advantage is the ability to analyze variables in contingency tables of any size and to take account interactions between variables in the performed analysis [8].

The aim of the article is to present the use of log-linear analysis on the example of cross-sectional studies regarding health condition, functioning and disability of people aged 71-80 years.

Material

assessment of disability and the choice of the factors that determine it was carried out on the results of research on 800 people aged 71-80 years. The study was performed by the use of a direct interview with a random sample of residents of south-eastern Poland (the Podkarpacie region). The frame of the survey was represented by the Personal Identification Number (PESEL) database, collected by the Ministry of Internal Affairs and Administration.

The research tool was the WHO Disability Assessment Schedule (WHODAS 2.0) [9] and respondent's particulars which collected various socio-demographic data on the study population. The questionnaire contained questions concerning, among others the state of health, functioning, received benefits, use of health services, adjustment of the residence, physical activity and questions on personal particulars (including the number of people in the household, income level, education, etc.).

By means of WHODAS 2.0 questionnaire, the assessment of disability level was performed on the study population. It is a standardized tool and adapted to Polish conditions. It was developed on the basis of a comprehensive set of items contained in the International Classification of Functioning, Disability and Health (ICF). In order to calculate the overall level of disability as well as in specific areas of WHODAS 2.0, the following scale has been used (according to ICF) [9, 10]:

No disability – 0–4%

Mild disability – 5–24%

Moderate disability – 25–49%

Severe disability – 50–95%

Extreme disability – 96–100%.

For the purposes of the analysis, as for the people with disability, those participants characterized by at least a moderate level of disability are considered.

nienie czynników determinujących występowanie niepełnosprawności pozwala na precyzyjne określenie najważniejszych zmiennych i zaprojektowanie celowych działań prewencyjnych odpowiednio zhierarchizowanych.

Analiza log-liniowa często wykorzystywana jest w badaniach ekonomicznych, np. do wyboru czynników wpływających na określoną sytuację ekonomiczną analizowanego układu [6, 7]. Powinna być jednak zdecydowanie częściej stosowana w badaniach z obszaru nauk medycznych oraz nauk o zdrowiu, w których wiele zjawisk ma charakter jakościowy. Jej zaletą jest możliwość analizy zmiennych w tablicach kontyngencji o dowolnych wymiarach oraz uwzględnienia w prowadzonych analizach interakcji zachodzących między zmiennymi [8].

Celem artykułu jest przedstawienie zastosowania analizy log-liniowej na przykładzie przekrojowego badania stanu zdrowia, funkcjonowania i niepełnosprawności osób w wieku 71-80 lat.

Material

Ocenę niepełnosprawności oraz wybór czynników, które ją determinują przeprowadzono na wynikach badania 800 osób w wieku 71-80 lat. Badanie zostało wykonane z zastosowaniem wywiadu bezpośredniego na losowej próbie mieszkańców Polski południowo-wschodniej (rejon Podkarpacia). Operat badawczy stanowiła baza danych PESEL, gromadzona przez Ministerstwo Spraw Wewnętrznych i Administracji.

Narzędzie badawcze stanowił WHO Disability Assessment Schedule (WHODAS 2.0) [9] oraz kwestionariusz metrykalny, za pomocą którego zebrano różne dane socjodemograficzne dotyczące badanej populacji. Kwestionariusz zawierał pytania dotyczące m.in. stanu zdrowia, stanu funkcjonalnego, pobieranych świadczeń, korzystania z usług ochrony zdrowia, dostosowania miejsca zamieszkania, aktywności fizycznej oraz pytania metryczkowe (m.in. liczbę osób w gospodarstwie domowym, wysokość dochodów, wykształcenie itd.).

Za pomocą kwestionariusza WHODAS 2.0 dokonano oceny poziomu niepełnosprawności badanych osób. Jest to narzędzie wystandaryzowane i dostosowane do warunków polskich. Zostało opracowane na podstawie kompleksowego zestawu pozycji zawartych w Międzynarodowej Klasyfikacji Funkcjonowania, Niepełnosprawności i Zdrowia (ICF). W celu obliczenia ogólnego poziomu niepełnosprawności oraz w poszczególnych dziedzinach WHODAS 2.0 zastosowano następującą skalę, zgodnie z ICF [9,10]:

Brak niepełnosprawności – 0–4%

Nieznaczna niepełnosprawność – 5–24%

Umiarkowana niepełnosprawność – 25–49%

Znaczna niepełnosprawność – 50–95%

Skrajnie duża niepełnosprawność – 96–100%

Na potrzeby przeprowadzanej analizy jako osoby niepełnosprawne określono te, które cechuje co najmniej umiarkowany poziom niepełnosprawności.

Method

the starting point in the log-linear analysis are expected quantities for each level of categorical variables [11]. If interactions are irrelevant, and the class quantities depend only on the main factors, it means the absence of interactions between categorical variables [12].

Considering the log-linear model, it is assumed that the natural logarithm of the value of a expected quantity in the table of independence is a linear function of factors. The generalized form of the log-linear model is as follows [7]:

$$\ln(\hat{n}^{P\dots}) = \bar{n} + \sum \lambda_i^P + \dots + \sum \lambda_{ij}^{PQ} + \dots + \sum \lambda_{ijk}^{PQR} + \dots \quad (1)$$

where:

$\hat{n}^{P\dots}$ – vector $[\hat{n}_i^P, \hat{n}_{ij}^{PQ}, \dots]$, where: \hat{n}_i^P is a expected quantity i – of the level of the factor P , \hat{n}_{ij}^{PQ} – a predicted size of interactions i – of the level of the P factor and j – of the level of the Q factor, etc.,

λ_i^P – indicator - of the level of P factor, etc.,

λ_{ij}^{PQ} – interactions indicator of the second order, i – of the level of the P factor, j – of the level of the Q factor.

λ_{ijk}^{PQR} - interactions indicator of the third order i - of the level of the P factor, j – of the level of the Q factor and k – of the level of the R factor, etc.,

and:

\bar{n} – mean of the natural logarithms of the observed quantities determined from the formula:

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

where:

\sum_T – the summation over all the factors,

\sum_i – the summation over all levels of the factors.

It is further assumed that as for the considered model, all the factors sum up to zero over the levels of interaction.

Optimally designed log-linear model allows for the best quantitative prediction taking into account the smallest possible number of interactions. To determine the order of interactions that must be included in the model, Pearson's χ^2 and χ^2 maximum likelihood statistics are used to assess whether the expected quantities are significantly different from the observed ones [3]. χ^2 test becomes important, as the number of expected quantities is getting different from observed quantities. Matching algorithm of the best model, first of all is based on fitting the model without any interactions, and when it does not fit (χ^2 test is statistically significant), next obtaining inclusion of first order interactions, and then checking the fit are performed. The process will be stopped when joining

Metoda

Punktem wyjścia w analizie log-liniowej są liczebności oczekiwane dla poszczególnych poziomów zmiennych kategoryzacyjnych [11]. Jeżeli człony interakcyjne są nieistotne, a liczebności w klasach zależą tylko od czynników głównych, to oznacza to brak wzajemnych oddziaływań pomiędzy zmiennymi kategoryzacyjnymi [12].

W modelu log-liniowym przyjmuje się, że logarytm naturalny wartości oczekiwanej liczebności w tabeli niezależności jest liniową funkcją czynników. Uogólniona postać modelu log-liniowego jest następująca [7]:

gdzie:

$\hat{n}^{P\dots}$ – wektor $[\hat{n}_i^P, \hat{n}_{ij}^{PQ}, \dots]$, gdzie \hat{n}_i^P jest oczekiwaną liczebnością i – tego poziomu czynnika P , \hat{n}_{ij}^{PQ} – przewidywaną liczebnością interakcji i – tego poziomu czynnika P oraz j – tego poziomu czynnika Q , itd.,

λ_i^P – wskaźnik – tego poziomu czynnika P ,

λ_{ij}^{PQ} – wskaźnik interakcji rzędu drugiego i – tego poziomu czynnika P oraz j – tego poziomu czynnika Q ,

λ_{ijk}^{PQR} – wskaźnik interakcji rzędu trzeciego i – tego poziomu czynnika P , j – tego poziomu czynnika Q oraz k – tego poziomu czynnika R , itd.,

zaś:

\bar{n} – średnia z logarytmów naturalnych obserwowanej liczebności wyznaczona ze wzoru:

gdzie:

\sum_T – sumowanie po wszystkich czynnikach,

\sum_i – sumowanie po wszystkich poziomach czynników.

Zakłada się ponadto, że w rozważanym modelu wszystkie czynniki sumują się do zera po poziomach oddziaływania.

Optymalnie skonstruowany model log-liniowy umożliwia najlepszą predykcję liczebności przy uwzględnieniu w nim możliwie najmniejszej liczby interakcji. Do ustalenia rzędu interakcji, jaki należy włączyć do modelu, wykorzystuje się statystyki χ^2 Pearsona i χ^2 największej wiarygodności, które oceniają, czy liczebności oczekiwane są istotnie różne od obserwowanych [3]. Test χ^2 staje się istotny, w miarę jak liczebność oczekiwana zaczyna się różnić od liczebności obserwowanej. Algorytm dopasowania najlepszego modelu polega w pierwszej kolejności na dopasowaniu modelu bez jakichkolwiek interakcji, zaś gdy ten nie będzie pasował (test χ^2 jest statystycznie istotny) włączeniu interakcji pierwszego rzędu, a następnie sprawdzeniu

all the higher-order interactions gives insignificant dispersion (test χ^2 is not statistically significant). In practice, the Pearson's χ^2 and χ^2 maximum likelihood statistics generate similar results [12].

In order to assess which interactions of the given order should be included in the model, it is necessary to analyze partial and marginal dependence. The partial dependence informs about the significance of the degree of interactions with the assumption that all the other effects of the same degree were included in the model. It can be verified by means of the χ^2 test of partial dependence. The statistical significance of the test leads to the need to take account of the given interaction in the model, since its omission would result that a fit of the model is significantly worse. The marginal dependence informs about the influence of the interaction provided that the model does not include any interactions in the same order. We can verify it by the use of χ^2 test of marginal interdependence [3].

To assess the fit of the model to the data, the Pearson's χ^2 and the χ^2 maximum likelihood statistics are used, as well as R^2 and A coefficients.

The R^2 coefficient is being calculated from the formula:

$$R^2 = \frac{G^2(M_0) - G^2(M)}{G^2(M_0)} \quad (3)$$

where:

$G^2(M_0)$ – the value of the chi-square maximum likelihood statistics for the smallest model M_0 (the model of complete independence of coefficients)

$G^2(M)$ – the value of the chi - square statistics for the analyzed model M

The better the model is fitted to the data, the higher the value of this coefficient is [13].

And while the ratio is calculated as follows:

$$A = G^2(M) - 2 \cdot df \quad (4)$$

where:

df – the number of degrees of freedom of the analyzed model M .

The smaller value of this coefficient indicates a better fit of the model to the data [14].

Results

With regard to the considered model, the dichotomous dependent variable was defined as the occurrence of disability Y . In the set of potential independent variables the following variables were considered:

X_1 – gender (female, male)

X_2 – education (at the most basic vocational education, at least secondary education)

X_3 – place of residence (city, countryside),

jego dopasowania. Proces ulegnie zatrzymaniu w momencie, gdy dołączenie wszystkich interakcji wyższego rzędu daje nieistotne rozszerzenie (test χ^2 jest statystycznie nieistotny). W praktyce statystyki χ^2 Pearsona i χ^2 największej wiarygodności generują zbliżone wyniki [12].

W celu dokonania oceny, które z interakcji danego rzędu należy uwzględnić w modelu należy dokonać analizy zależności cząstkowych oraz brzegowych. Zależność cząstkowa informuje o istotności interakcji danego stopnia przy założeniu, że wszystkie inne efekty tego samego stopnia zostały ujęte w modelu. Weryfikuje się ją testem χ^2 zależności cząstkowej. Statystyczna istotność tego testu powoduje konieczność uwzględnienia danej interakcji w modelu, gdyż jej pominięcie spowoduje, iż dopasowanie modelu będzie istotnie gorsze. Zależność brzegowa informuje o wpływie danej interakcji pod warunkiem, że w modelu nie ma jeszcze zawartych żadnych interakcji tego samego rzędu. Weryfikujemy ją testem χ^2 współzależności brzegowych [3].

Do oceny dopasowania modelu do danych wykorzystuje się statystyki χ^2 Pearsona oraz χ^2 największej wiarygodności, a także współczynniki R^2 oraz A.

Współczynnik R^2 wyznaczamy ze wzoru:

gdzie:

$G^2(M_0)$ – wartość statystyki chi-kwadrat największej wiarygodności dla najmniejszego modelu M_0 (modelu całkowitej niezależności czynników),

$G^2(M)$ – wartość statystyki chi-kwadrat dla analizowanego modelu M ,

Model jest tym lepiej dopasowany do danych, im wyższa jest wartość tego współczynnika [13].

Natomiast współczynnik A obliczamy następująco:

gdzie:

df – liczba stopni swobody analizowanego modelu M .

Mniejsza wartość tego współczynnika świadczy o lepszym dopasowaniu modelu do danych [14].

Wyniki badań

W rozważanym modelu określono dychotomiczną zmienną zależną zdefiniowaną jako występowanie niepełnosprawności Y . W zbiorze potencjalnych zmiennych niezależnych uwzględniono następujące zmienne:

X_1 – płeć (kobieta, mężczyzna),

X_2 – wykształcenie (co najwyżej zasadnicze zawodowe, co najmniej średnie),

X_3 – miejsce zamieszkania (miasto, wieś),

X_4 – dostosowanie mieszkania/domu do potrzeb zwią-

X_4 – adjustment of the flat / house to the needs of daily functioning (fully, not fully),

X_5 – performing daily physical activity at least 150 minutes a week (yes, no),

X_6 – physical exercises strengthening the muscles and improving physical fitness (none, 1-3 times a week, 4 or more times a week),

X_7 – social activity expressed in engagement in at least one organization / group / association (yes, no)

X_8 – possibility to use the help of other people every day (yes, no).

At the outset, the analysis verified which variables $X_1 - X_8$ showed a statistically significant association with the level of disability. For this purpose the χ^2 test of independence was used. Collation of the values of test statistics with corresponding test probabilities are shown in Table 1.

Considering the information presented in Table 1, it indicates that there was no significant relationship ($p > 0,01$) between the level of disability (Y) and sex (X_1), place of residence (X_3), and performing daily physical activity at least 150 minutes a week (X_5). Hence, indicated variables were excluded from further analysis.

Therefore, the following variables: X_2, X_4, X_6, X_7, X_8 were used in the log - linear model. Owing to determine the specifications of the model, the order of interactions were determined. The test results of all interactions are selected in Table 2.

Regarding the information contained in Table 2, it ought to be said that a model which is the subject of the analysis should include, apart from the main components, the interactions of at the most second order.

In order to assess which of the interactions should be included in the model, the partial tests were used (Table 3). Due to a large number of possible interactions in the case of six factors, the analysis was limited up to the second order of the interaction, since all interactions of the third or higher order proved to be statistically insignificant.

zanych z codziennym funkcjonowaniem (w pełni, nie w pełni),

X_5 – wykonywanie codziennej aktywności fizycznej minimum 150 minut tygodniowo (tak, nie),

X_6 – wykonywanie ćwiczeń fizycznych wzmacniających mięśnie i poprawiających sprawność fizyczną (brak, 1–3 razy w tygodniu, 4 i więcej razy w tygodniu),

X_7 – aktywność społeczna wyrażona przynależnością do minimum 1 organizacji/grupy/stowarzyszenia (tak, nie),

X_8 – możliwość skorzystania na co dzień z pomocy innych osób (tak, nie).

Na wstępie zweryfikowano, które ze zmiennych $X_1 - X_8$ wykazują statystycznie istotny związek z poziomem niepełnosprawności. W tym celu wykorzystano test niezależności χ^2 . Zestawienie wartości statystyk testowych wraz z odpowiadającymi im prawdopodobieństwami testowymi przedstawiono w tabeli 1.

Z zaprezentowanych w tabeli 1 informacji wynika, że nie stwierdzono istotnego związku ($p > 0,01$) pomiędzy poziomem niepełnosprawności (Y) a płcią (X_1), miejscem zamieszkania (X_3) oraz wykonywaniem codziennej aktywności fizycznej minimum 150 minut tygodniowo (X_5). Stąd też wskazane zmienne zostały wykluczone z dalszych analiz.

Zatem do modelu log – liniowego wprowadzono następujące zmienne: X_2, X_4, X_6, X_7, X_8 . W celu określenia specyfikacji modelu określono rząd interakcji. Wyniki testów wszystkich interakcji zestawiono w tabeli 2.

Z informacji zawartych w tabeli 2 wynika, że w modelu, który będzie przedmiotem analizy powinny znaleźć się, oprócz składowych głównych, interakcje co najwyżej drugiego rzędu.

W celu oceny, które z interakcji powinny być włączone do modelu wykorzystano testy cząstkowe (tabela 3). Ze względu na dużą liczbę możliwych interakcji w przypadku sześciu czynników, ograniczono się do zaprezentowania interakcji maksymalnie drugiego rzędu, gdyż wszystkie interakcje trzeciego bądź wyższych rzędów okazały się statystycznie nieistotne.

Table 1. Analysis of the relationship between the dependent variable and potential independent variables $X_1 - X_8$

Tabela 1. Analiza zależności pomiędzy zmienną objaśnianą a potencjalnymi zmiennymi objaśniającymi $X_1 - X_8$

Variable Zmienna	Value of χ^2 statistics Wartość statystyki χ^2	Value of probability test p wartość prawdopodobieństwa testowego p
X_1	2,816	0,093
X_2	24,815	< 0,001
X_3	4,450	0,035
X_4	151,382	< 0,001
X_5	3,114	0,078
X_6	94,500	< 0,001
X_7	8,053	0,005
X_8	19,057	< 0,001

Note: potential independent variables significantly associated with the dependent variable are written in bold.

Uwaga: pogrubioną czcionką zaznaczono potencjalne zmienne objaśniające istotnie powiązane ze zmienną objaśnianą.

Table 2. Test results of interactions between the variables $Y, X_2, X_4, X_6, X_7, X_8$
 Tabela 2. Wyniki testów interakcji pomiędzy zmiennymi $Y, X_2, X_4, X_6, X_7, X_8$

The degree of interaction Stopień interakcji	Degrees of freedom Stopnie swobody	The value of χ^2 Maximum Likelihood Wartość χ^2 największej wiarygodności	Probability of p value Wartość prawdopodobieństwa p	Value of Pearson's χ^2 Wartość χ^2 Pearsona	Probability of p value Wartość prawdopodobieństwa p
1	7	1601,957	< 0,001	4437,606	< 0,001
2	20	412,239	< 0,001	703,931	< 0,001
3	30	28,294	0,555	33,415	0,305
4	25	18,315	0,829	16,331	0,905
5	11	5,464	0,907	6,022	0,872
6	2	1,746	0,418	1,791	0,408

Note: potential independent variables significantly associated with the dependent variable are written in bold.

Uwaga: pogrubioną czcionką zaznaczono potencjalne zmienne objaśniające istotnie powiązane ze zmienną objaśnianą.

Table 3. Selected results of partial and marginal tests between the variables $Y, X_2, X_4, X_6, X_7, X_8$

Tabela 3. Wybrane wyniki testów cząstkowych i brzegowych pomiędzy zmiennymi $Y, X_2, X_4, X_6, X_7, X_8$

Factors Czynniki	Degrees of freedom Stopnie swobody	Partial association χ^2 Związek cząstkowy χ^2	p-value in the partial association Wartość p w związku cząstkowym	Marginal association χ^2 Związek brzegowy χ^2	p-value in the marginal association Wartość p w związku brzegowym
Y	1	41,131	< 0,001	41,131	< 0,001
X_2	1	60,966	< 0,001	60,966	< 0,001
X_4	1	204,678	< 0,001	204,678	< 0,001
X_6	2	763,464	< 0,001	763,464	< 0,001
X_7	1	381,090	< 0,001	381,090	< 0,001
X_8	1	150,629	< 0,001	150,629	< 0,001
YX_2	1	7,951	0,005	23,931	< 0,001
YX_4	1	80,574	< 0,001	141,207	< 0,001
YX_6	2	31,663	< 0,001	86,599	< 0,001
YX_7	1	0,112	0,738	8,573	0,003
YX_8	1	2,127	0,145	16,367	< 0,001
X_2X_4	1	1,675	0,196	16,601	< 0,001
X_2X_6	2	15,953	< 0,001	33,091	< 0,001
X_2X_7	1	2,445	0,118	8,267	0,004
X_2X_8	1	1,916	0,166	-0,000	1,000
X_4X_6	2	22,238	< 0,001	86,467	< 0,001
X_4X_7	1	11,297	0,001	27,787	< 0,001
X_4X_8	1	23,224	< 0,001	41,227	< 0,001
X_6X_7	2	16,447	< 0,001	33,836	< 0,001
X_6X_8	2	5,901	0,052	17,651	< 0,001
X_7X_8	1	0,287	0,592	0,686	0,408

Note: interactions, for which partial and marginal dependence is important are written in bold..

Uwaga: pogrubioną czcionką zaznaczono interakcje, dla których zależność cząstkowa i brzegowa jest istotna.

The results of partial and marginal tests indicated the need to include six main factors in the model, as well as second order interactions between variables X_2, X_4, X_6 and the dependent variable. Furthermore, in order to avoid a significant reduction in the degree of fitting associated with removing the interactions between the independent variables, the effect representing all interactions between independent variables was included in the

Rezultaty testów cząstkowych i brzegowych wskazały na potrzebę uwzględnienia w modelu sześciu czynników głównych, a także interakcji rzędu drugiego pomiędzy zmiennymi X_2, X_4, X_6 i zmienną zależną. Ponadto, w celu uniknięcia istotnego zmniejszenia stopnia dopasowania związanego z usunięciem interakcji pomiędzy zmiennymi niezależnymi, do modelu włączono efekt reprezentujący wszystkie interakcje pomiędzy zmiennymi niezależnymi

model [3]. That is why, a log-linear model was assumed for further analysis:

$$\ln(\hat{n}_{(ijklmn)}) = \bar{n} + \lambda_i^Y + \lambda_j^{X_2} + \lambda_k^{X_4} + \lambda_l^{X_6} + \lambda_m^{X_7} + \lambda_n^{X_8} + \lambda_{ij}^{YX_2} + \lambda_{ik}^{YX_4} + \lambda_{il}^{YX_6} + \lambda_{ijklmn}^{X_2X_4X_6X_7X_8} \quad (5)$$

The values of χ^2 maximum likelihood statistics (amounting 37.304, $p = 0.716$) and Pearson's χ^2 statistics (amounting 37.193, $p = 0.720$) were not statistically significant, which allowed us to conclude that the suggested model is well designed to the empirical data. This is also confirmed by the diagram 1, which shows the size of the observed quantities compared to the fitting quantities.

In addition, regarding the model (5) based on the formulas (3) and (4), the measures R^2 and A were determined. They amounted, respectively, $R^2 = 0.92$ and $A = -48.696$, which confirmed its very good quality of the fit.

The received model let us state that the second order interactions may influence (apart from the above mentioned factors) classification of an individual into a group of people with at least moderate disability for the following pairs of variables:

[3]. Wobec tego do dalszych analiz przyjęto model log-liniowy postaci:

Wartości statystyki χ^2 największej wiarygodności wynoszące 37,304 ($p = 0,716$) oraz statystyki χ^2 Pearsona równa 37,193 ($p = 0,720$) okazały się statystycznie nieistotne, co pozwoliło wnioskować, iż zaproponowany model jest dobrze dostosowany do danych empirycznych. Potwierdza to także wykres 1, na którym przedstawiono liczebności obserwowane względem liczebności dopasowanych.

Ponadto, dla modelu (5) na podstawie wzorów (3) i (4) wyznaczono miary R^2 oraz A. Wynosiły one odpowiednio: $R^2 = 0,92$ oraz $A = -48,696$, które potwierdziły jego bardzo dobrą jakość dopasowania.

Do czynników, które warunkują zaklasyfikowanie danej osoby do grupy osób z co najmniej umiarkowaną niepełnosprawnością według WHODAS mogą mieć wpływ, oprócz wskazanych wyżej czynników występujących w modelu, interakcje rzędu drugiego dla następujących par zmiennych:

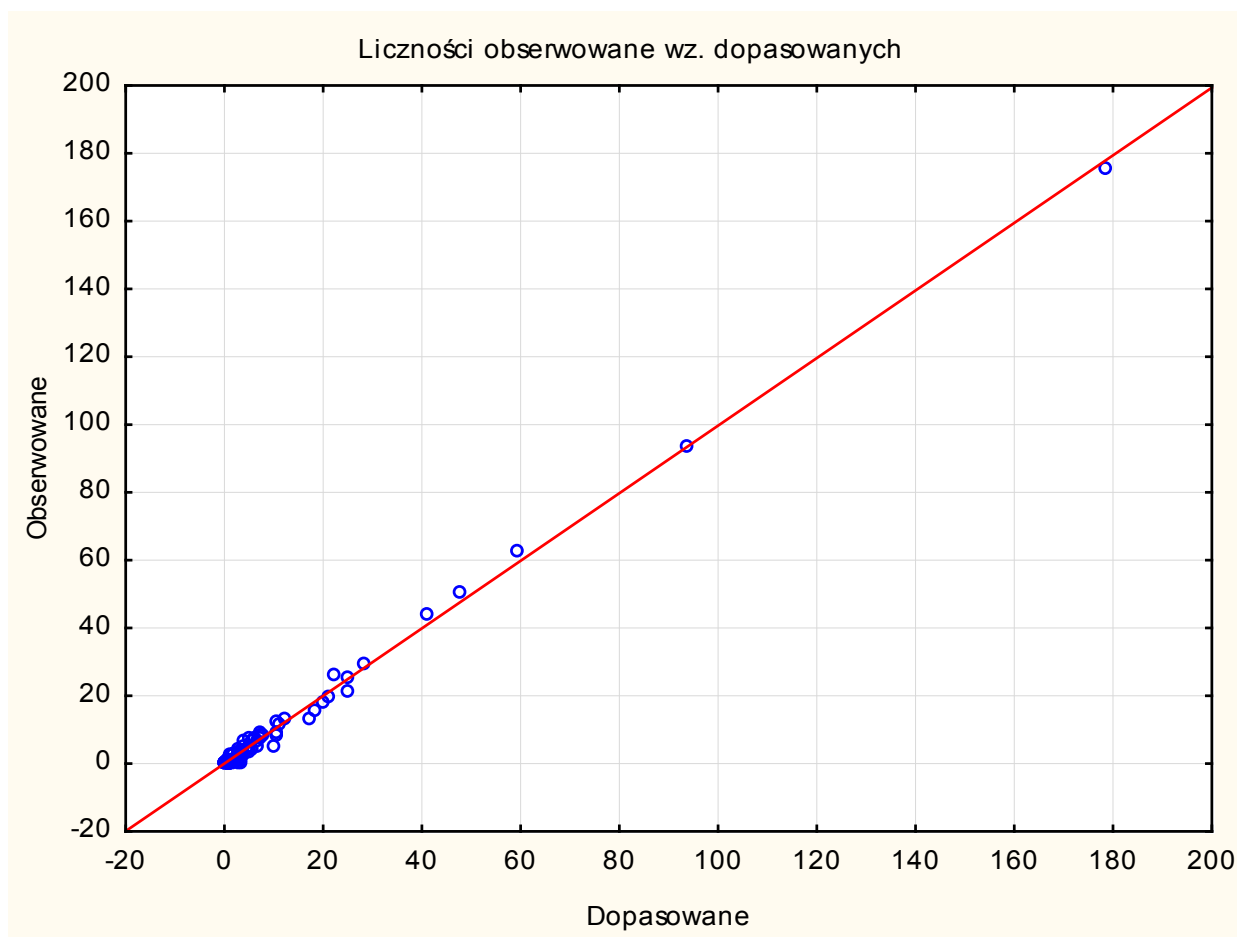


Diagram 1. Spread of observed quantities compared to the number of fitting quantities (5)

Wykres 1. Rozrzut liczebności obserwowanych względem liczebności dopasowanych w modelu (5)

1. education and doing physical exercises to strengthen the muscles and improve physical performance,
2. adjustment of the flat/house to the needs of daily functioning and doing physical exercises to strengthen the muscles and improve physical performance,
3. adjustment of the flat/house to the needs of daily functioning and social activity,
4. adjustment of the flat/house to the needs of daily functioning and the ability to use the help of other people every day,
5. physical exercises strengthening the muscles and improving physical and social activity.

Thus, the dependent variable which most often interacts significantly with other dependent variables, is doing physical exercises to strengthen the muscles and improving physical performance. In contrast, education interacts least often with other dependent variables.

Results

The assumptions of the log-linear analysis were developed over 50 years ago. It is based on the estimation method of maximum likelihood which analyzes the independence and the interaction between the qualitative variables [15].

In this paper, the log - linear analysis allowed us to identify the factors that determine the occurrence of at least a moderate level of disability in people aged 71-80 years in south-eastern Poland. These include: education, adjustments of a house/flat to meet the needs of daily functioning and doing physical exercises to strengthen the muscles and improve physical performance. Moreover, the interactions of dependent variables that may affect the dependent variable have been indicated. Then, based on the estimated and verified model, it is possible to obtain a detailed analysis of received dependencies as well as conclusions, useful in practice can be drawn, which will be the subject of a separate article.

Implementation of the log-linear analysis allowed us to make a more detailed description of the interactions between the level of disability and the independent variables, than it would take place in the situation of using standard measures evaluating the relationship of qualitative features. The log-linear analysis is more powerful than regular study of dependence according to the probability value [16–19].

The choice of data analysis techniques should be adapted to the specific needs of the research problem. The log-linear analysis is useful especially when it is necessary to examine the relationship between three or more variables (the immeasurable variables) and when the analyzed sample is quite large. It results from the fact that due to the number of factors and their categories, the sample will be divided into very many subgroups. In order to estimate correctly the distributions in population on the basis of frequency distributions observed in the sample,

1. wykształcenie oraz wykonywanie ćwiczeń fizycznych wzmacniających mięśnie i poprawiających sprawność fizyczną,
2. dostosowanie mieszkania/domu do potrzeb związanych z codziennym funkcjonowaniem oraz wykonywanie ćwiczeń fizycznych wzmacniających mięśnie i poprawiających sprawność fizyczną,
3. dostosowanie mieszkania/domu do potrzeb związanych z codziennym funkcjonowaniem oraz aktywność społeczna,
4. dostosowanie mieszkania/domu do potrzeb związanych z codziennym funkcjonowaniem oraz możliwość skorzystania na co dzień z pomocy innych osób,
5. wykonywanie ćwiczeń fizycznych wzmacniających mięśnie i poprawiających sprawność fizyczną oraz aktywność społeczna.

Zatem zmienną objaśniającą wchodzącą najczęściej w istotne interakcje z innymi zmiennymi objaśniającymi jest wykonywanie ćwiczeń fizycznych wzmacniających mięśnie i poprawiających sprawność fizyczną. Natomiast najrzadziej w interakcje z innymi zmiennymi objaśniającymi wchodziło wykształcenie.

Omówienie wyników

Założenia analizy log-liniowej zostały opracowane już ponad 50 lat temu. Opiera się ona na estymacyjnej metodzie największej wiarygodności, analizuje niezależność i interakcje pomiędzy zmiennymi jakościowymi [15].

W niniejszej pracy analiza log-liniowa pozwoliła na wskazanie czynników, które determinują występowanie co najmniej umiarkowanego poziomu niepełnosprawności u osób w wieku 71–80 lat w Polsce południowo-wschodniej. Należą do nich: wykształcenie, dostosowanie mieszkania/domu do potrzeb związanych z codziennym funkcjonowaniem oraz wykonywanie ćwiczeń fizycznych wzmacniających mięśnie i poprawiających sprawność fizyczną. Wskazano także interakcje zmiennych objaśniających, które mogą mieć wpływ na zmienną zależną. Na podstawie oszacowanego i zweryfikowanego modelu możliwa będzie szczegółowa analiza otrzymanych zależności oraz wyciągnięcie wniosków przydatnych w praktyce, co będzie przedmiotem odrębnego artykułu.

Wykorzystanie analizy log-liniowej pozwoliło na dokonanie dokładniejszego opisu zależności pomiędzy poziomem niepełnosprawności a zmiennymi niezależnymi, niż miałyby to miejsce w sytuacji zastosowania standardowych mierników oceny współzależności cech jakościowych. Analiza log-liniowa posiada większą moc niż zwykle badanie zależności oparte na prawdopodobieństwie testowym [16–19].

Wybór techniki analizy danych powinien być dostosowany do specyficznego sformułowania problemu badawczego. Analiza log-liniowa jest przydatna

these subgroups must be sufficiently numerous. Speaking of carrying out the log-linear analysis, particularly valuable is to consider more than just one model, which allows for a closer examination of the problem and avoids hasty conclusions [20].

Conclusions

The log-linear analysis is a valuable tool for the analysis of qualitative data and should be used more often in research in the field of medical and health sciences. It allows us to determine the effect not only of individual variables on the formation of an independent variable, but also their interactions. Moreover, it gives the opportunity to state the odds occurrence of a dependent variable according to the different qualitative categories of dependent variables. The information obtained in this way might be a valuable clue to take practical action to decrease or increase the severity of the studied phenomenon.

szczególnie wtedy, gdy zachodzi konieczność zbadania powiązań pomiędzy trzema i więcej zmiennymi, przy czym są to zmienne niemierzalne oraz gdy analizowana próba jest dostatecznie duża. Wynika to z faktu, iż ze względu na liczbę czynników i ich kategorii próba zostanie podzielona na bardzo wiele podgrup. Aby na podstawie rozkładów frekwencji zaobserwowanych w próbie rzetelnie szacować rozkłady w populacji, podgrupy te muszą być odpowiednio liczne. Szczególnie cenne w przeprowadzaniu analizy log-liniowej jest rozważanie większej liczby modeli niż tylko jednego, co pozwala na dokładniejsze zbadanie problemu oraz uniknięcie wysuwania pochopnych wniosków [20].

Wnioski

Analiza log-liniowa jest wartościowym narzędziem analizy danych jakościowych i powinna być stosowana zdecydowanie częściej w badaniach w obszarze nauk medycznych oraz nauk o zdrowiu. Pozwala ona bowiem na ustalenie wpływu nie tylko pojedynczych zmiennych na kształtowanie się zmiennej niezależnej, lecz także ich interakcji. Ponadto pozwala na określenie szansy wystąpienia zmiennej objaśnianej w zależności od poszczególnych kategorii jakościowych zmiennych objaśniających. Uzyskiwane w ten sposób informacje mogą być bardzo cenną wskazówką do podjęcia konkretnych praktycznych działań w celu zmniejszenia lub zwiększenia stopnia nasilenia badanego zjawiska.

Bibliography / Bibliografia

1. Kwasiborski PJ, Sobol M. Test niezależności chi-kwadrat i jego zastosowanie w interpretacji wyników badań klinicznych. *Kardiochir Torakochi* 2011;4:550–554.
2. Stanisław A. Przystępny kurs statystyki z zastosowaniem Statistica PL na przykładach z medycyny, tom 1, Statystyki podstawowe. Kraków: StatSoft; 2006.
3. Stanisław A. Przystępny kurs statystyki z zastosowaniem Statistica PL na przykładach z medycyny, tom 3, Analizy wielowymiarowe. Kraków: StatSoft; 2007.
4. Officer A, Groce NE. Key concepts in disability. *Lancet* 2009;374(9704):1795–1796.
5. World Health Organization. *World Report on Disability*. Geneva, Switzerland: World Health Organization; 2011.
6. Salamaga M. Wykorzystanie analizy log-liniowej do wyboru czynników opisujących sytuację ekonomiczną gospodarstw domowych. *Przegląd Statystyczny* 2008;55(4):40–51.
7. Bąk I. Wybór czynników determinujących wyjazdy turystyczne rencistów. W: Borkowski B, redaktor. *Metody ilościowe w badaniach ekonomicznych*. Warszawa: Szkoła Główna Gospodarstwa Wiejskiego Wydział Zastosowań Informatyki i Matematyki Katedra Ekonometrii i Statystyki; 2011:1–10.
8. Brzezińska J. Wykorzystywanie modeli logarytmiczno-liniowych do analizy bezrobocia w Polsce w latach 2004–2012. *Pr Nauk Uniw Ekon Wroc* 2014;15(327):95–103.
9. Üstün TB, Kostanjsek N, Chatterji S, Rehm J. *Measuring Health and Disability Manual for WHO Disability Assessment Schedule WHODAS 2.0*. Geneva, Switzerland: World Health Organization; 2010.
10. World Health Organization. *International Classification of Functioning, Disability and Health*. Geneva, Switzerland: World Health Organization; 2011.
11. Dobosz M. *Wspomagana komputerowo statystyczna analiza wyników badań*. Warszawa: Akademicka Oficyna Wydawnicza ELIT; 2004.
12. Bąk I. Wykorzystanie analizy log-liniowej do wyboru czynników opisujących aktywność turystyczną domowych gospodarstw emerytów w Polsce. *Folia Pomer. Univ. Technol. Stetin. Oeconomica* 2012;297(68):5–16.
13. Goodman LA. *Analyzing Qualitative/Categorical Data*. In: Magidson J, edotors. *Log-Linear Models and Latent-Structure Analysis*. Cambridge; 1978.
14. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Czaki F, edotors. *Proceedings of the 2nd International Symposium on Information*. Budapest; 1973.
15. Kaczmarek E, Małkiewicz E. Przydatność modelowania log-liniowego do wieloczynnikowej analizy epidemiologicznej – przykład badania populacyjnego wystę-

- powania rozszczepów podniebienia pierwotnego i/lub wtórnego w latach 1998–1999 w Polsce. *Dent Med Probl* 2005;42(3):419-424.
16. Peng W, Zhang H, Li Y, Li C, Yan W. Log-linear model used in the hybrid design of case-parents triad/control-mother dyad. *Zhonghua Liu Xing Bing Xue Za Zhi* 2016;37(6):886-890.
 17. Zhu M, Nie S, Yan W. Application of log-linear model in the case-parent triad study. *Zhonghua Liu Xing Bing Xue Za Zhi* 2012;33(4):435-438.
 18. Helmy A, Timofeev I, Palmer C, Gore A, Menon D, Hutchinson P. Hierarchical log linear analysis of admission blood parameters and clinical outcome following traumatic brain injury. *Acta Neurochirurgica* 2010;152(6):953-957.
 19. Nyman H, Pensar J, Koski T, Corander J. Context-specific independence in graphical log-linear models. *Computational Statistics* 2016;31(4):1493-1512.
 20. Póltorak M. Modele log-liniowe i ich zastosowania w psychologii. *Przegląd Psychologiczny* 2007;50(1):25-44.