Prof. dr hab. Marek Walesiak
Uniwersytet Ekonomiczny we Wrocławiu
Wydział Ekonomii, Zarządzania i Turystyki
Katedra Ekonometrii i Informatyki
58-500 Jelenia Góra, ul. Nowowiejska 3

RECENZJA ROZPRAWY DOKTORSKIEJ
MGR JUSTYNY BRZEZIŃSKIEJ NT.
„Modele logarytmiczno-liniowe i ich zastosowanie w analizie zjawisk ekonomicznych"
Uniwersytet Ekonomiczny w Katowicach, Wydział Zarządzania
(maszynopis powielony, 187 stron)

Temat podjęty w dysertacji pt. „Modele logarytmiczno-liniowe i ich zastosowanie w analizie zjawisk ekonomicznych" jest trudny merytorycznie. Autorka podjęła się problemu budowy modeli logarytmiczno-liniowych dla danych niemetrycznych w tablicach kontyngencji oraz prezentacji przykładowych ich zastosowań w analizie zjawisk ekonomicznych. Na s. 4-5 *Wstępu* w sposób przekonujący doktorantka uzasadniła wybór tematu dysertacji.

Struktura rozprawy doktorskiej obejmuje pięć rozdziałów, wśród których cztery pierwsze składają się na część teoretyczno-metodyczną, a rozdział 5 stanowi część empiryczną. Układ i treść rozprawy doktorskiej odpowiadają jej tematowi.

Na s. 5 *Wstępu* przedstawiono cel główny rozprawy oraz 4 hipotezy badawcze. Oddają one istotę badanej w rozprawie problematyki.

Rozdział pierwszy poświęcono analizie zależności zmiennych niemetrycznych dla dwuwymiarowych i wielowymiarowych tablic kontyngencji. Tytuł podrozdziału 1.1 brzmi „Pojęcie zależności w statystycznej analizie danych" (s. 8-13). Tytuł nie oddaje w pełni jego zawartości. Znajdujemy tutaj również kryteria podziału metod statystycznych. Dopiero na s. 12-13 Autorka przedstawia rozważania dotyczące zależności zachodzących między zmiennymi w analizie danych. W podrozdziale 1.2 zaprezentowano szeroki przegląd miar analizy zależności dla zmiennych nominalnych w przypadku dwuwymiarowych tablic kontyngencji (2x2 i HxJ) oraz wielowymiarowych tablic kontyngencji (2x2xK oraz HxJxK). Analizę zależności zmiennych nominalnych w przypadku wielowymiarowych tablic kontyngencji ograniczono do trzech zmiennych. Podrozdział 1.3 poświęcono charakterystyce miar korelacji dla zmiennych porządkowych. Na s. 40 zapowiedziano charakterystykę miar korelacji obliczanych na

podstawie skonstruowanych tablic kontyngencji. Nie wyjaśniono istoty konstrukcji tablicy kontyngencji dla zmiennych porządkowych. W związku z tym miary Kendalla, a zwłaszcza wzory (1.95)-(1.97) są niejasne. Współczynniki W (wzór 1.99) oraz r_S (wzór 1.104) nie są obliczane na podstawie tablic kontyngencji. Ponadto powoływanie się na opracowanie S. Mynarskiego do wzoru na współczynnik korelacji rang Spearmana jest nieuzasadnione. Jest to wzór powszechnie znany. Podrozdział 1.4 traktuje o problemie zerowych liczebności w tablicach kontyngencji i sposobach ich rozwiązania. Niektóre elementy tego podrozdziału omówiono wcześniej w tym samym rozdziale. W podrozdziale 1.5 (s. 45-46) krótko zasygnalizowano problem interakcji zmiennych i ich wpływu w modelach logarytmiczno-liniowych.

Rozdział drugi zawiera charakterystykę analizy korespondencji. W podrozdziale 2.1 przedstawiono genezę jej powstania, w tym różne nazwy, pod którymi omawia się w literaturze analizę korespondencji. W podrozdziałach 2.2 i 2.3 zaprezentowano analizę korespondencji odpowiednio dla dwóch i wielu zmiennych. W podrozdziale 2.4 (s. 55-60) zaprezentowano cztery sposoby rozkładu macierzy **A** (tj. macierzy różnic standaryzowanych, która pozwala na wyznaczenie optymalnej przestrzeni rzutowania) według wartości osobliwych autorstwa Fishera, Greenacre'a, Andersena i Jobsona. W podrozdziale 2.5 (s. 61-65) poruszono dwa zagadnienia: wybór najlepszej przestrzeni rzutowania (metody: stopień wyjaśniania inercji, kryterium łokcia, kryterium Greenacre'a) oraz zagadnienie interpretacji wyników analizy korespondencji.

Rozdział trzeci, stanowiący część zasadniczą rozprawy doktorskiej w aspekcie metodycznym, poświęcono modelom logarytmiczno-liniowym. W podrozdziale 3.1 krótko przedstawiono genezę tych modeli budowanych dla danych niemetrycznych w tablicach kontyngencji. W podrozdziale 3.2 zaprezentowano trzy rozkłady probabilistyczne zmiennych niemetrycznych w tablicy kontyngencji, tj. wielomianowy, iloczynowo-wielomianowy i Poissona. Podrozdział 3.3 zawiera charakterystykę modeli logarytmiczno-liniowych budowanych na podstawie tablic kontyngencji. Tytuł punktu 3.3.1 „Rola zmiennych w modelu logarytmiczno-liniowym" jest zgodny z zawartością tylko na s. 70-71 (do końca drugiego akapitu). Dalsza część tego punktu analizuje inny problem, tj. przekształcenia standaryzacyjne w celu usunięcia niejednorodności zbioru danych w tablicy kontyngencji. Autorka nie określiła kiedy w trakcie budowy modelu logarytmiczno-liniowego ten etap występuje. Tylko ostatnie zdanie na s. 72 informuje, że jest to istotne w procesie wizualizacji danych. W punktach 3.3.2 i 3.3.3 zaprezentowano podstawy metodyczne budowy modeli logarytmiczno-liniowych odpowiednio dla tablic dwuwymiarowych (s. 73-78) i trójwymiarowych (s. 79-83). Na s. 81-82 przed-

stawiono modele logarytmiczno-liniowe budowane według zasady hierarchiczności. Tabela 3.4 (s. 82) zawiera 9 modeli logarytmiczno-liniowych budowanych według tej zasady. Według zasady hierarchiczności buduje się wiele modeli, a ostatecznie wybierany jest model o najmniejszej złożoności, który jednocześnie zapewnia wystarczające dopasowanie do danych. W punkcie 3.3.4 przedstawiono charakterystykę modeli hierarchicznych w przypadku trzech zmiennych. W zasadzie w tym punkcie wyznaczono logarytmy ilorazu szans dla poszczególnych modeli. W punkcie 3.3.5 scharakteryzowano metodę największej wiarygodności wykorzystywaną do estymacji parametrów modelu logarytmiczno-liniowego. W podrozdziale 3.4 (s. 94-99) zaprezentowano cztery modele logarytmiczno-liniowe dla zmiennych porządkowych. Podrozdział 3.5 (s. 99-105) obejmuje charakterystykę mierników oceny jakości dopasowania modeli logarytmiczno-liniowych: statystyka chi-kwadrat i miary na niej bazujące, iloraz wiarygodności, kryteria AIC i BIC, indeks niepodobieństwa Giniego, współczynnik determinacji, skorygowany współczynnik determinacji.

Rozdział czwarty przedstawia analizę klas ukrytych. Dwa pierwsze podrozdziały poświęcono wprowadzeniu do zagadnienia analizy klas ukrytych. W podrozdziale 4.3 (s. 110-113) krótko scharakteryzowano analizę klas ukrytych w modelach logarytmiczno-liniowych. Ostatni podrozdział 4.4 (s. 113-115) traktuje o kryteriach wyboru liczby klas w analizie klas ukrytych. Na s. 107 (6-9 w. od dołu) Autorka w zasadzie sama uznała, że ten rozdział nie nawiązuje bezpośrednio do recenzowanej rozprawy doktorskiej, a wskazuje jedynie dalsze możliwości zastosowań modeli logarytmiczno-liniowych. Być może lepiej byłoby ująć ten problem w *Zakończeniu* jako dalsze kierunki badań.

Rozdział piąty przedstawia zastosowania modeli logarytmiczno-liniowych w analizie pięciu zjawisk ekonomicznych (podrozdziały 5.1-5.5). W podrozdziale 5.6 zaprezentowano charakterystykę (nie porównanie, jak jest w tytule) wybranych programów komputerowych pozwalających na analizę danych w tablicach kontyngencji (analiza korespondencji, modele logarytmiczno-liniowe, analiza klas ukrytych).

Rozprawę doktorską wieńczy *Zakończenie*. Autorka odniosła się tutaj do sformułowanych we *Wstępie* hipotez, określiła swój wkład własny do analizowanego problemu badawczego oraz wskazała zalety i wady modeli logarytmiczno-liniowych.

Rozprawa doktorska mgr Justyny Brzezińskiej stanowi indywidualny i oryginalny wkład w problematykę zastosowania modeli logarytmiczno-liniowych w analizie zjawisk ekonomicznych. Do niewątpliwych walorów recenzowanej rozprawy doktorskiej należy zaliczyć:

1. Jasne i precyzyjne sformułowanie problemu badawczego, w tym podstawowego celu pracy.

2. Zwięzłe i syntetyczne ujęcie prezentowanej metodyki badawczej dotyczącej budowy modeli logarytmiczno-liniowych ujętej w rozdziale trzecim rozprawy.

3. Jej nowatorski charakter na rynku polskim w zakresie modelowania logarytmiczno-liniowego na podstawie tablic kontyngencji budowanych dla danych niemetrycznych.

4. Podjęcie próby uporządkowania terminologii polskiej z zakresu modelowania logarytmiczno-liniowego na podstawie tablic kontyngencji.

5. Opracowanie autorskich skryptów w programie **R** oraz umiejętne wykorzystanie istniejących skryptów służących budowie modeli logarytmiczno-liniowych w analizie zjawisk ekonomicznych. Skrypty nie zostały udostępnione przez Autorkę.

6. Podjęcie w rozdziale piątym udanej próby (mimo zgłoszonych dalej uwag krytycznych do tego rozdziału) zastosowania modeli logarytmiczno-liniowych w analizie takich zjawisk ekonomicznych, jak bezrobocie, aktywność zawodowa ludności, wybrane aspekty zachowań społecznych, analiza wypadków w górnictwie.

Ta niewątpliwie wartościowa praca doktorska skłania do kilku uwag, choć niektóre z nich mają z pewnością zabarwienie dyskusyjne:

1. W pierwszym akapicie na s. 8 (9-14 w. od dołu) Autorka powołała się na pracę Walesiaka (2011), s. 13. Sformułowanie „W badaniach empirycznych wykorzystujących metody statystycznej analizy wielowymiarowej nie wychodzi się poza trzeci wymiar" nie dotyczy liczby analizowanych zmiennych. Chodzi tutaj o pojęcie kostki danych, w której kolejne wymiary to obiekty, zmienne oraz czas.

2. Wzór na współczynnik V Cramera (1.50) na s. 29 jest niepoprawny. Prawidłową postać tego wzoru podano m.in. w pracach:

a. Steczkowski J., Zeliaś A. (1997), *Metody statystyczne w badaniach cech jakościowych*, Wydawnictwo AE, Kraków, s. 175,

b. Goodman L.A., Kruskal W.H. (1954), *Measures of association for cross classifications*, "Journal of the American Statistical Association", Vol. 49, December, s. 740.

3. W punkcie 1.2.2 (s. 23-31) dobrze byłoby zaprezentować wymagania, jakie powinna spełniać dobrze zdefiniowana miara zależności. Ponadto brakuje w tym miejscu prezentacji miary zależności Z. Hellwiga:

a. Hellwig Z. (1969), *On the measurement of stochastical dependence*, „Zastosowania Matematyki", T. 10, s. 233-247,

b. Hellwig Z. (1975), *Elementy rachunku prawdopodobieństwa i statystyki matematycznej*, PWN, Warszawa, s. 143.

4. Przedstawione w podrozdziale 5.2 wykresy dla analizowanego przykładu aktywności zawodowej ludności opisanej 4 zmiennymi są mało przydatne. Wykresy czteropolowy i asocjacji (rys. 5.3 na s. 132) dotyczą tylko dwóch wybranych zmiennych. Wykres sitkowy (rys. 5.4 na s. 134) wprawdzie uwzględnia trzy zmienne występujące w modelu logarytmiczno-liniowym, ale jak sama Autorka stwierdza na s. 133, wykres jest nieczytelny.

Do recenzowanej pracy zgłaszam następujące uwagi szczegółowe mające głównie charakter uchybień natury redakcyjnej, stylistycznej i językowej:

– błędy językowe i redakcyjne: na s. 10 (19 w. od góry) jest „Kruskalka", a powinno być „Kruskala"; na s. 16 (1 w. od góry) jest „liczebności empiryczne", a powinno być „liczebności oczekiwane (teoretyczne)"; na s. 21 (11 w. od góry) jest „Everit", a powinno być „Everitt"; m.in. na s. 28 oraz w spisie literatury jest „Blauris (1994)", a powinno być „Blasius (1994)"; na s. 39 (7 w. od góry) zamiast „(179)" powinno być „(1.79)"; na s. 39 (9 w. od góry" zamiast „... prawdopodobieństw" powinno być „... prawdopodobieństw dla warstw"; na s. 44 w pierwszym akapicie nie wszystkie nazwiska Autorów zostały odmienione zgodnie z zasadami języka polskiego; na s. 71 (13 w. od dołu) zamiast „dużej liczebność" powinno być „dużej liczebności"; na s. 85 (3 w. od dołu) zamiast „okreslony" powinno być „określony"; na s. 89 (7 w. od góry) zamiast „minimin" powinno być „minimum"; na s. 94 (1 w od góry) zamiast „Finberg" powinno być „Fienberg"; na s. 102 (18 w. od góry) zamiast „hipotezę zerowa" powinno być „hipotezę zerową"; na s. 102 (25 w. od góry) zamiast „złożoności)" powinno być „złożoności"; na s. 103 (20 w. od dołu) zamiast „Sakkamoto" powinno być „Sakamoto"; sformułowania „okres czasu" i „moment czasu" (s. 118, 2-3 w. od dołu) są niepoprawne w języku polskim; na s. 164 (spis literatury) zamiast „Fiebnerg" powinno być „Fienberg"; na rys. 5.5 (s. 136) zamiast kategorii „0-24" powinno być „18-24" (zob. s. 135);

– przy odwoływaniu się do pozycji książkowych bezwzględnie należy podawać numer lub numery stron. Autorka w całej rozprawie nie stosuje się do tego zalecenia;

– w wielu miejscach w rozprawie do wzorów nie ma odwołań literaturowych (np. w podrozdziale 2.3);

– cytowane na s. 14 (13-15 w. od góry) pozycje literaturowe mają charakter bardzo wybiórczy;

– na s. 18 (4-5 w. od góry) zastosowano niepoprawne merytorycznie sformułowanie „orientacja tablicy kontyngencji";

– nie bardzo wiadomo co oznacza na s. 45 we wzorze (1.105) T_e – całkowita liczebność tablicy;

– zamiennie w pracy stosuje się terminy „zmienna" i „cecha" (np. s. 52);

– w tablicy Burta (2.3 na s. 52) w ostatnich dwóch wierszach niektóre zera są w złych miejscach;

– do wzoru (2.14) na s. 53 nie objaśniono symbolu $\mathbf{D}_r$;

– na s. 59-60 przykład został pobieżnie zaprezentowany. Ponadto czy wnioski z przeprowadzonej analizy czterech metod rozkładu macierzy $\mathbf{A}$ według wartości osobliwych wysunięte na podstawie jednego przykładu można uogólnić?

– na s. 73 (6 w. od góry) podano, że w punkcie 3.3.2 zaprezentowano podstawy metodyczne budowy modeli logarytmiczno-liniowych dla trójwymiarowych tablic kontyngencji. W rzeczywistości jest to ujęte w punkcie 3.3.3;

– na s. 82 (10 w od dołu) zamiast „w tabeli 3.3" powinno być „odpowiednio w tabeli 3.2 i 3.3";

– na s. 110 (6-13 w. od góry) wskazano własności formalne, które pełnią szczególną rolę w procesie segmentacji w badaniach ekonomicznych. Nie bardzo widać tutaj związek z segmentacją rynku;

– z uwagi na to, że w podrozdziale 5.1 nie zamieszczono odpowiednich tablic kontyngencji dotyczących bezrobocia ze względu na płeć, wykształcenie i rejon Polski w latach 2004-2013 wnioski zawarte na s. 117-118 są niejasne i niemożliwe do potwierdzenia;

– na s. 120 (20 w. od dołu) zamiast [EP][R] powinno być [WP][R];

– rys. 5.1 (s. 119) i 5.2 (s. 125) są mało czytelne z uwagi na wprowadzenie pełnych nazw kategorii. Lepiej byłoby zastosować tutaj prostą symbolikę z legendą pod rysunkami. Ponadto na rys. 5.2 powinno być „wykształcenie" a nie „edukacja";

– tytuł podrozdziału 5.3 (s. 134-138) jest zbyt ogólny w stosunku do uwzględnionych w badaniu zmiennych (s. 135).

Recenzowana rozprawa doktorska została obudowana bardzo wieloma pozycjami literaturowymi (551 pozycji, w tym zdecydowana większość w języku angielskim). Na szczególne wyróżnienie zasługuje to, że wśród pozycji literaturowych znajdujemy 16 pozycji Autorki rozprawy, w tym aż 9 w języku angielskim.

Podniesione uwagi nie umniejszają walorów naukowych i poznawczych recenzowanej rozprawy doktorskiej. Biorąc pod uwagę jej wysoki poziom (w tym: samodzielne rozwiązanie przez Doktorantkę problemu naukowego, dobrą wiedzę w dyscyplinie naukowej – ekonomia, wyniki własnych badań zawartych w rozdziale 5, ujęcia syntetyzujące i porządkujące wiedzę) stwierdzam, że dysertacja mgr Justyny Brzezińskiej pt. „Modele logarytmiczno-liniowe i ich zastosowanie w analizie zjawisk ekonomicznych" odpowiada wymogom stawianym rozpra-

wom doktorskim określonym w art. 13 ustęp 1 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. Nr 65, poz. 595 z późn. zm.) i wnoszę o dopuszczenie do publicznej jej obrony w dziedzinie nauk ekonomicznych w dyscyplinie ekonomia.

Jelenia Góra, 11 czerwca 2014 r.

Marek Malewski