

ANALIZA ZASOBÓW KOMPETENCJI Z WYKORZYSTANIEM GRUPOWANIA ZMIENNYCH NIEMETRYCZNYCH

Alicja Grześkowiak

Wydział Zarządzania, Informatyki i Finansów
Uniwersytet Ekonomiczny we Wrocławiu
e-mail: alicja.grzeskowiak@ue.wroc.pl

Streszczenie: W artykule przedstawiono zastosowanie grupowania zmiennych niemetrycznych do badania zasobów kompetencji w Polsce. Analizę przeprowadzono na podstawie danych z ogólnopolskiego badania pt. Bilans Kapitału Ludzkiego odnoszących się zarówno do kompetencji twardych, jak i miękkich. Zastosowana procedura analityczna pozwoliła na wyróżnienie grup kompetencji podobnie ocenianych i wyznaczenie dla nich syntetycznych zmiennych – reprezentantek, które posłużyły do ewaluacji powiązań z wybranymi charakterystykami społeczno-demograficznymi.

Słowa kluczowe: kompetencje, zmienne niemetryczne, analiza skupień

WSTĘP

Zasoby kompetencji pełnią bardzo istotną funkcję w gospodarce warunkując jakość kapitału ludzkiego, który stanowi jeden z ważniejszych czynników rozwoju społeczno-gospodarczego. Ich znaczenie można także rozpatrywać w kontekście działania organizacji oraz w wymiarze indywidualnym związanym z kształtowaniem kariery poszczególnych jednostek. Ze względu na kluczowe znaczenie kompetencji właściwe wydaje się prowadzenie badań nad ich stanem z wykorzystaniem stosownych procedur analitycznych. Wieloaspektowość i możliwość występowania zależności pomiędzy posiadaniem różnych umiejętności skłania do stosowania metod statystycznej analizy wielowymiarowej do oceny zasobów kompetencji, co stanowi przedmiot rozważań niniejszej pracy. Obszernym źródłem danych dotyczących poziomu kompetencji, wykorzystanych w przedstawionej analizie, są wyniki ostatniej edycji badania Bilans Kapitału Ludzkiego. Zebrane opinie na temat posiadanych kompetencji są wyrażone na skali porządkowej, co wymaga zastosowania specyficznych metod analitycznych, które

są słabiej akcentowane i poznane w przeciwieństwie do bogatego arsenału technik umożliwiających badanie zbiorów zmiennych metrycznych.

Artykuł koncentruje się wokół dwóch zasadniczych celów. Pierwszy z nich ma charakter poznawczy i dotyczy oceny relacji zachodzących pomiędzy różnymi kompetencjami, w szczególności wiąże się z wyodrębnianiem grup umiejętności podobnie ocenianych przez respondentów oraz identyfikacją ich powiązań z charakterystykami społeczno-demograficznymi ankietowanych. Drugi cel jest ukierunkowany metodycznie i odnosi się do wskazania możliwości zastosowania metody grupowania zmiennych niemetrycznych zaproponowanej w pracy [Chavent i in. 2013] do prowadzenia analiz opartych na zmiennych pochodzących ze słabych skal pomiarowych. Ma to duże znaczenie w badaniach społeczno-ekonomicznych, w których często bazuje się na wynikach badań sondażowych zdominowanych przez występowanie zmiennych nominalnych lub porządkowych. Posiadanie tego rodzaju danych nie przekreśla jednakże stosowania podejścia wielowymiarowego, a rozwój procedur analitycznych i ich implementacja w środowisku R owocuje nowymi możliwościami w tym zakresie, na co warto zwrócić uwagę przy doborze metod analizy.

METODYKA BADANIA

Wykonane analizy opierają się na danych wtórnych pochodzących z ogólnopolskiego badania „Bilans Kapitału Ludzkiego” dotyczącego ludności, przeprowadzonego przez Polską Agencję Przedsiębiorczości oraz Uniwersytet Jagielloński w 2014 r., obejmującego swym zasięgiem 17 674 respondentów. Przed wykonaniem obliczeń usunięto nieznaczną liczbę obserwacji ze względu na występowanie braków danych.

Poziom posiadanych kompetencji był oceniany przez ankietowanych w pięciostopniowej skali porządkowej: 1 – niski, 2 – podstawowy, 3 – średni, 4 – wysoki, 5 – bardzo wysoki. Opiniowane były zarówno kompetencje twarde, jak i miękkie [zob. Górniak 2014, s. 192-193, 205-206]. Zestawienie kompetencji wziętych pod uwagę w niniejszej pracy wraz ze skrótami stosowanymi w dalszej części artykułu jest następujące [Bilans Kapitału Ludzkiego 2013]¹:

- kompetencje twarde: (A) wyszukiwanie i analiza informacji oraz wyciąganie wniosków (informacje) – szybkie streszczanie dużej ilości tekstu (tekst); logiczne myślenie, analiza faktów (logika); ciągłe uczenie się nowych rzeczy (nowe); (B) obsługa, montowanie i naprawa urządzeń technicznych (urządzenia); (C) wykonywanie obliczeń (obliczenia) – wykonywanie prostych rachunków (proste); wykonywanie zaawansowanych obliczeń matematycznych

¹ Określenia kompetencji przytoczono w brzmieniu dosłownym z kwestionariusza BKL [https://bkl.parp.gov.pl/pobierz.html/kwestionariusze_IV_2013.7z], w nawiasach podano skrócone określenia stosowane przez Autora w dalszej części opracowania.

- (zaawansowane); (D) obsługa komputera i wykorzystanie Internetu (komputer) – podstawowa znajomość pakietu typu MS Office (Office), znajomość specjalistycznych programów, umiejętność pisania programów czy tworzenia stron internetowych (specjalistyczne),
- kompetencje miękkie: (A) samoorganizacja pracy i przejawianie inicjatywy (samoorganizacja) – samodzielne podejmowanie decyzji (decyzje); przedsiębiorczość i przejawianie inicjatywy (przedsiębiorczość); kreatywność (kreatywność); odporność na stres (stres); terminowa realizacja zaplanowanych działań (terminowość); (B) kontakty z innymi ludźmi, zarówno ze współpracownikami, jak i klientami czy podopiecznymi (kontakty) – współpraca w grupie (grupa); łatwe nawiązywanie kontaktów z współpracownikami czy klientami (relacje); bycie komunikatywnym i jasne przekazywanie myśli (komunikatywność); rozwiązywanie konfliktów pomiędzy ludźmi (konflikty); (C) organizowanie i prowadzenie prac biurowych (biuro); (D) zdolności kierownicze i organizacja pracy innych (kierownicze) – koordynowanie pracy innych pracowników (koordynacja); dyscyplinowanie innych pracowników – przywoływanie ich do porządku (dyscyplinowanie); (E) dyspozycyjność (dyspozycyjność) – gotowość do częstych wyjazdów (wyjazdy); elastyczny czas pracy (elastyczność).

Rozpatrzone zostały cztery kompetencje twarde natury ogólnej oraz siedem szczegółowych. Zestaw kompetencji miękkich jest szerszy – pięć ogólnych oraz trzynaście szczegółowych. Ze względu na ich odmienny charakter analizy przeprowadzono oddzielnie dla kompetencji miękkich i twardych. Oprócz odpowiedzi dotyczących poziomu umiejętności w badaniu uwzględniono również zmienne charakteryzujące respondentów pod względem społeczno-demograficznym, tj.: miejsce zamieszkania (miasto, wieś); wiek skategoryzowany (18-24, 25-34, 35-44, 45-54, 55-59/64); płeć (kobieta, mężczyzna); wykształcenie (gimnazjalne i poniżej, zasadnicze zawodowe, średnie, wyższe); sytuację zawodową według BAEL (pracujący, bezrobotni, nieaktywni).

W przytoczonym zestawieniu zmiennych znajdują się wyłącznie zmienne o charakterze niemetrycznym, co musi mieć wpływ na wybór stosownych metod ilościowych zaprojektowanych do badania tego rodzaju danych. Duża liczba rozpatrywanych kompetencji ocenianych na skali porządkowej stanowi swoiste wyzwanie analityczne. Oczywiście możliwe jest badanie każdej kompetencji osobno, np. oceniając średnie wyniki w różnych przekrojach [zob. np. Czarnik, Turek 2015], ale ze względów merytorycznych można podejrzewać, że pomiędzy poziomem różnych kompetencji mogą istnieć powiązania. W literaturze przedmiotu można odnaleźć próby zastosowania różnych technik uwzględniających wieloaspektowe powiązania pomiędzy umiejętnościami. Badanie relacji pomiędzy kompetencjami za pomocą analizy głównych składowych dla zmiennych niemetrycznych (CatPCA) opisano w pracy [Grześkowiak 2014], a zastosowanie klasycznej analizy czynnikowej znaleźć

można w publikacji [Górniak 2014]. W artykule [Dziechciarz-Duda, Dziechciarz 2016] wykorzystano analizę korespondencji, analizę PROFIT oraz grupowanie oparte na mierze odległości GDM. W publikacji [Grześkowiak 2015] do badania związków użyto reguł asocjacyjnych, a wykorzystanie wielorakiej analizy czynnikowej uwzględniającej grupy wieku przedstawiono w pracy [Grześkowiak 2016]. Rezultaty wymienionych prac wskazują na istnienie powiązań pomiędzy deklarowanym poziomem różnych umiejętności. Taka struktura danych skłania do poszukiwań pewnych uogólnień, które w sposób syntetyczny pozwalałyby opisywać złożone zasoby kompetencyjne.

W niniejszej pracy zasugerowano zastosowanie procedury grupowania zmiennych zaproponowanej i omówionej w [Chavent i in. 2013] do identyfikacji powiązań pomiędzy kompetencjami i ich odzwierciedlenia za pomocą nowych reprezentantek. Na korzyści płynące z grupowania zmiennych zwraca się uwagę w [Lasek, Pęczkowski 2010] wskazując na likwidację problemu współliniowości i redundancji informacji, zwiększenie przejrzystości i czytelności związków między zmiennymi oraz umożliwianie budowy modeli o mniejszej złożoności.

W literaturze traktującej o grupowaniu zmiennych spotkać można dwa podejścia: najczęstszym jest zastosowanie takich samych algorytmów, jak przy klasyfikacji obiektów, opisanych np. w pracach [Kaufman, Rousseeuw 2009, Everitt i in. 2011, Kassambara 2017]. Podejście drugie to wykorzystanie procedur opracowanych wyłącznie w tym celu, ale należy zaznaczyć, że metod poświęconych jedynie grupowaniu zmiennych jest stosunkowo niewiele. Jak wskazano w opracowaniu [Chavent i in. 2013] do najważniejszych metod dla zmiennych metrycznych należy zaliczyć: procedurę VARCLUS dostępną w oprogramowaniu SAS, metodę grupowania wokół zmiennych ukrytych [Vigneau, Qannari 2003, Vigneau i in. 2015], diametrical clustering [Dhillon i in. 2003], ujęcie nieparametryczne [Palla i in. 2012] oraz podejście bazujące na analizie kanonicznej [Bühlmann i in. 2013]. Chavent i in. [2013] przedstawiają uniwersalne rozwiązanie problemu grupowania zmiennych, które zostanie wykorzystane w niniejszym artykule. Jest ono szczególnie interesujące dla badaczy stykających się z analizą danych niemetrycznych lub danych o mieszanym charakterze (metrycznych i niemetrycznych). Autorzy koncepcji postulują tworzenie zmiennych syntetycznych c_k reprezentujących wyodrębniane skupienia C_k o postaci [Chavent i in. 2013]:

$$c_k = \arg \max_{u \in \mathbb{R}^n} \left\{ \sum_{x_j \in C_k} r_{u|x_j}^2 + \sum_{y_j \in C_k} \eta_{u|y_j}^2 \right\} \quad (1)$$

gdzie: x_i – zmienne metryczne, y_i – zmienne niemetryczne, $r_{u|x_j}$ – współczynnik korelacji Pearsona, $\eta_{u|y_j}^2$ – stosunek korelacyjny, u – wartości, dla których jest poszukiwane maksimum.

Zmienne syntetyczne c_k są otrzymywane jako pierwsze główne składowe otrzymane z procedury PCAMIX, tj. analizy głównych składowych dla mieszanki zmiennych metrycznych i niemetrycznych [Kiers 1991]. Algorytmy grupowania są

tak skonstruowane by zmaksymalizować kryterium homogeniczności, które dla danego skupienia C_k definiowane jest jako [Chavent i in. 2013]:

$$H(C_k) = \sum_{x_j \in C_k} r_{u,x_j}^2 + \sum_{y_j \in C_k} \eta_{u|y_j}^2, \quad (2)$$

natomiast dla całego podziału jako suma miar homogeniczności (2) wyznaczonych dla utworzonych klas. Dokładny opis algorytmów wraz z charakterystyką pakietu `ClustOfVar` programu R pozwalającego na zastosowanie procedur jest przedstawiony w pracy [Chavent i in. 2013]. Wartościowym aspektem, na który zwracają uwagę autorzy idei, jest otrzymanie ilościowych zmiennych syntetycznych reprezentujących poszczególne skupienia. Pakiet `ClustOfVar` daje możliwość wyboru metody grupowania spośród hierarchicznej procedury aglomeracyjnej oraz metody k-średnich. W niniejszej pracy zastosowano pierwsze z wymienionych podejść. Ze stosowaniem hierarchicznych procedur aglomeracyjnych wiąże się problem ustalania ostatecznego podziału na klasy, gdyż metoda w sama w sobie nie daje odpowiedzi na pytanie, ile należy wyodrębnić grup. Twórcy pakietu `ClustOfVar` proponują ocenę stabilności podziału za pomocą metody bazującej na średnim skorygowanym indeksie Randa.

Porównanie dwóch różnych wyników podziału P i P' można oprzeć na spostrzeżeniu, że istnieją cztery możliwości: (a) dwa elementy znajdują się w tej samej grupie zarówno w podziale P , jak i P' , (b) w podziale P są w tej samej grupie, a w podziale P' nie, (c) w podziale P' są w tej samej grupie, a w podziale P nie, (d) znajdują się w różnych grupach w obu podziałach [zob. Wagner, Wagner 2007]. Indeks Randa [Rand 1971] bazuje na porównaniu liczby kombinacji zgodnych do wszystkich wariantów:

$$R = \frac{a+d}{a+b+c+d}, \quad (3)$$

Ze względu na fakt, że wartość oczekiwana indeksu Randa dwóch losowych podziałów nie jest stała, Hubert i Arabie zaproponowali jego modyfikację określaną skorygowanym indeksem Randa [Hubert, Arabie 1985], który można wyrazić formułą [Yeung, Ruzzo 2001]:

$$R_{sk} = \frac{\sum_{i,j} \binom{n_{ij}}{2} - [\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}] / \binom{n}{2}}{0,5[\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}] - [\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}] / \binom{n}{2}} \quad (4)$$

gdzie n_{ij} , $n_{i.}$, $n_{.j}$ – odpowiednio liczebności komórek oraz liczebności brzegowe tabeli kontyngencji odzwierciedlającej przynależność do grup otrzymanych w dwóch podziałach P i P' .

Pakiet `ClustOfVar` oferuje podejście bootstrapowe do obliczania zmodyfikowanego indeksu Randa umożliwiającego ocenę stabilności podziałów. Średni skorygowany indeks Randa według formuły (4) obliczony na podstawie pięćdziesięciu prób typu bootstrapowego stanowił przesłankę do wyboru ostatecznej liczby klas.

Otrzymane grupy zmiennych skupiające kompetencje o zbliżonych ocenach są reprezentowane przez ilościowe zmienne syntetyczne postaci (1), które stanowią

podstawę dalszych analiz wiążących ocenę umiejętności z cechami demograficzno-społecznymi. W celu oceny siły powiązań posłużono się współczynnikiem eta, adekwatnym do pomiaru stopnia związku pomiędzy zmienną metryczną a niemetryczną, który można wyrazić jako [zob. Malarska 2005, s.89]:

$$\eta_z = \frac{s(\bar{z}_j)}{s(z)}, \quad s^2(\bar{z}_j) = \frac{1}{n} \sum_{j=1}^k (\bar{z}_j - \bar{z})^2 n_{.j}, \quad (5)$$

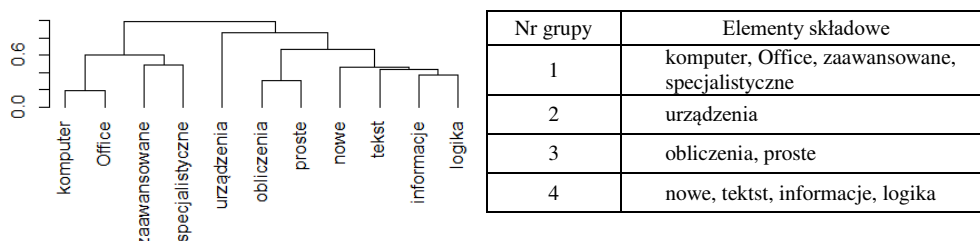
gdzie: $s(z)$ – odchylenie standardowe rozpatrywanej zmiennej metrycznej, k – liczba kategorii zmiennej niemetrycznej, \bar{z}_j – średnia zmiennej metrycznej dla danej kategorii zmiennej niemetrycznej. Współczynnik eta jest miarą unormowaną w przedziale $\langle 0,1 \rangle$, co pozwala na ocenę i porównywanie siły zależności.

WYNIKI BADAŃ EMPIRYCZNYCH

W niniejszym rozdziale przedstawiono wyniki analizy zasobów kompetencji w podziale na kompetencje twarde i miękkie. W odniesieniu do każdego z obszarów przeprowadzono grupowanie zmiennych niemetrycznych i wyodrębniono ich skupienia, których liczbę określono na podstawie kształtowania się skorygowanego indeksu Randa. Dla otrzymanych w ten sposób klas obliczono zmienne syntetyczne je reprezentujące, które następnie wykorzystano do oceny powiązań ze zmiennymi demograficznymi.

Na rysunku 1 zilustrowano wyniki procedury grupowania zmiennych niemetrycznych odzwierciedlających kompetencje twarde.

Rysunek 1. Wyniki hierarchicznej procedury aglomeracyjnej dla kompetencji twardych



Źródło: opracowanie własne

Otrzymano cztery klasy, w tym jedną jednoelementową – jest to skupienie nr 2 zawierające wyłącznie umiejętności w zakresie obsługi urządzeń technicznych, co oznacza, że poziom tej kompetencji charakteryzuje się dużą odrębnością i nie wiąże się z deklarowaniem innych umiejętności. Skupienie nr 3 łączy kompetencje matematyczne jako kategorię ogólną z wykonywaniem prostych rachunków. Warto zauważyć, że wykonywanie zaawansowanych obliczeń nie zostało przyłączone do tej klasy, lecz stanowi element skupienia nr 1 obejmującego ponadto wszystkie rozpatrywane kompetencje informatyczne, zarówno na poziomie podstawowym,

jak i specjalistycznym. Ostatnia z wyodrębnionych grup zawiera umiejętności związane z wyszukiwaniem i analizą informacji oraz uczeniem się nowych rzeczy.

Powiązania zmiennych syntetycznych reprezentujących poszczególne skupienia z charakterystykami demograficzno-społecznymi odzwierciedlają współczynniki eta zawarte w tabeli 1.

Tabela 1. Wartości współczynników eta pomiędzy zmiennymi syntetycznymi dla kompetencji twardych i charakterystykami demograficzno-społecznymi

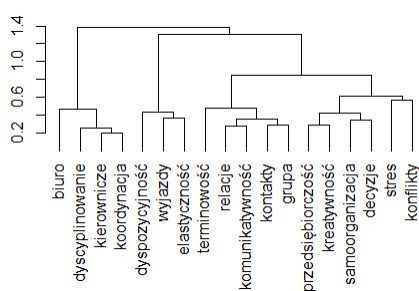
Nr grupy	Płeć	Wiek	Miejsce zamieszkania	Wykształcenie	Sytuacja zawodowa
1	0,053	0,423	0,166	0,497	0,170
2	0,294	0,048	0,009	0,091	0,152
3	0,063	0,163	0,102	0,407	0,183
4	0,107	0,241	0,147	0,501	0,211

Źródło: opracowanie własne

Z wyjątkiem skupienia nr 2 (obsługa urzędów) najwyższe wartości miary otrzymano dla poziomu wykształcenia. Można więc stwierdzić, że ten czynnik w największym stopniu jest związany z oceną kompetencji twardych, w szczególności wysoko z umiejętnościami dotyczącymi analizy informacji oraz kompetencjami informatycznymi w powiązaniu z wykonywaniem zaawansowanych obliczeń. Należy również podkreślić wysoki stopień powiązania zmiennej syntetycznej dla skupienia nr 1 z wiekiem ankietowanych, co wiąże się z lepszą znajomością nowoczesnych technologii przez ludzi młodych. Z kolei umiejętność obsługi urzędów pozostaje w zależności z płcią respondentów, co wynika z lepszej samooceny mężczyzn w tym zakresie.

Analogiczna procedura przeprowadzona dla zestawu kompetencji miękkich wskazała na istnienie sześciu grup (rysunek 2).

Rysunek 2. Wyniki hierarchicznej procedury aglomeracyjnej dla kompetencji miękkich



Nr grupy	Elementy składowe
1	biuro, dyscyplinowanie, kierownicze, koordynacja
2	dyspozycyjność, wyjazdy, elastyczność
3	terminowość, relacje, komunikatywność, kontakty, grupa
4	przedsiębiorczość, kreatywność, samoorganizacja, decyzje
5	stres
6	konflikty

Źródło: opracowanie własne

Istnieją dwa rodzaje umiejętności, które stanowią odrębne klasy, mimo iż w badaniu Bilans Kapitału Ludzkiego były rozpatrywane jako kompetencje

szczegółowe stanowiące składowe kompetencji o charakterze ogólnym. Należą do nich odporność na stres (skupienie nr 5) oraz radzenie sobie w sytuacjach konfliktowych (skupienie nr 6). Wyniki aglomeracji wskazują na podobieństwo w ocenie zdolności kierowniczych oraz organizowania i prowadzenia prac biurowych (skupienie nr 1). W skupieniu nr 2 znalazły się wszystkie zmienne reprezentujące kwestie związane z dyspozycyjnością. Grupa nr 3 składa się z umiejętności interpersonalnych oraz przestrzegania terminów wykonywania zadań. Skupienie 4 obejmuje kompetencje związane z operatywnością i inwencją twórczą.

W tabeli 2 ujęto wartości współczynników eta pomiędzy zmiennymi syntetycznymi reprezentującymi skupienia a zmiennymi demograficzno-społecznymi.

Tabela 2. Wartości współczynników eta pomiędzy zmiennymi syntetycznymi dla kompetencji miękkich i charakterystykami demograficzno-społecznymi

Nr grupy	Płeć	Wiek	Miejsce zamieszkania	Wykształcenie	Sytuacja zawodowa
1	0,071	0,157	0,132	0,444	0,233
2	0,136	0,165	0,072	0,248	0,315
3	0,077	0,156	0,072	0,352	0,237
4	0,052	0,208	0,089	0,420	0,274
5	0,018	0,090	0,015	0,137	0,130
6	0,045	0,074	0,066	0,237	0,114

Źródło: opracowanie własne

Najwyższe wartości współczynników wystąpiły w przypadku poziomu wykształcenia, choć zależności te są słabsze niż dla kompetencji twardych. W najwyższym stopniu z wykształceniem korelują umiejętności kierownicze i organizowania prac biurowych (skupienie 1) oraz kompetencje ze sfery przedsiębiorczości i samoorganizacji pracy (skupienie 4). Kompetencje tworzące odrębne skupienia (nr 5 i nr 6) nie są mocno związane z żadną z rozpatrywanych cech demograficzno-społecznych. Warto zauważyć, że ocena umiejętności w bardzo niewielkim stopniu jest powiązana z płcią i miejscem zamieszkania respondentów. Z kolei kategoria wieku wykazuje najsilniejszą relację ze skupieniem nr 4, którego domena to przedsiębiorczość.

PODSUMOWANIE

Zasadniczym wkładem artykułu jest zastosowanie rzadko używanego podejścia w analizie danych ze słabych skal pomiarowych – procedury grupowania zmiennych niemetrycznych w celu wyodrębnienia klas cech podobnych i wyznaczenia odpowiadających im zmiennych syntetycznych. Otrzymane w ten sposób reprezentantki skupień posłużyły do identyfikacji powiązań z charakterystykami

społeczno-demograficznymi respondentów, które okazały się być najsilniejsze w przypadku wykształcenia i wieku.

Należy zaznaczyć, że aplikacje metod klasyfikacji obejmują najczęściej wyodrębnianie skupień obiektów, a nie zmiennych. Natomiast przedstawione w artykule odmienne podejście, tj. grupowanie zmiennych pozwoliło na redukcję wymiarów i przedstawienie struktury ocen kompetencji w uogólniony sposób. Procedura ujawniła istnienie kompetencji, które pozostają w izolacji od pozostałych, oraz umiejętności, które pozostają ze sobą w bliskiej relacji.

Efekty analiz zaprezentowanych w pracy wskazują, że warto sięgać po specyficzne metody analityczne adekwatne dla zmiennych o charakterze niemetrycznym. Ze względu na ograniczone ramy niniejszej pracy analizy zostały przeprowadzone wyłącznie za pomocą hierarchicznej procedury aglomeracyjnej. Wykorzystana metoda daje natomiast podstawy do pogłębienia rozważań o inne metody analiz wielowymiarowych, np. grupowanie metodą k-średnich, co wskazuje na potencjalne dalsze kierunki badań. Kolejnym obszarem zainteresowania może być wykorzystanie omówionych metod w celu oceny zmian zachodzących w zasobach kompetencji w wymiarze czasowym.

BIBLIOGRAFIA

- Bilans Kapitału Ludzkiego w Polsce. Badania ludności (2013) Polska Agencja Rozwoju Przedsiębiorczości, https://bkl.parp.gov.pl/pobierz.html/kwestionariusze_IV_2013.7z.
- Bühlmann P., Rütimann P., van de Geer S., Zhang C. H. (2013) Correlated Variables in Regression: Clustering and Sparse Estimation. *Journal of Statistical Planning and Inference*, 143(11), 1835-1858.
- Chavent M., Kuentz V., Liqueur B., Saracco L. (2013) ClustOfVar: An R Package for the Clustering of Variables. *Journal of Statistical Software*, 50(13), 1-16.
- Czarnik Sz., Turek K. (2015) Polski rynek pracy – aktywność zawodowa i struktura wykształcenia. PARP, Warszawa.
- Dhillon I. S., Marcotte E. M., Roshan U. (2003) Diametrical Clustering for Identifying anti-Correlated Gene Clusters. *Bioinformatics*, 19(13), 1612-1619.
- Dziechciarz-Duda M., Dziechciarz J. (2016) The Identification of Training Needs for Human Capital Quality Improvement in Poland – a Statistical Approach. *Statistics in Transition New Series*, 17(4), 723-736.
- Everitt B. S., Landau S., Leese M., Stahl D. (2011) *Cluster Analysis*. Wiley, Chichester.
- Górniak J. (red.) (2014) *Kompetencje Polaków a potrzeby polskiej gospodarki. Raport podsumowujący IV edycję badań BKL z 2013 roku*. Polska Agencja Rozwoju Przedsiębiorczości, Warszawa.
- Grześkowiak A. (2014) Deklarowane umiejętności zawodowe Polaków – analiza na podstawie danych sondażowych o charakterze niemetrycznym. *Acta Universitatis Nicolai Copernici. Nauki Humanistyczno-Społeczne. Ekonomia*, 45 (2), 2014, 161-171.
- Grześkowiak A. (2015) Competences of Adult Poles - Evaluation of the Age as a Differentiating Factor. [w:] Rotschedl J., Cermakova K. (red.) *Proceedings of the 15th*

- International Academic Conference. International Institute of Social and Economic Sciences, Prague, 381-391.
- Grzeškowiak A. (2016) Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 427, 122-130.
- Hubert L., Arabie P. (1985) Comparing Partitions. *Journal of Classification*, 2(1), 193-218.
- Kassambara A. (2017) *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA.
- Kaufman L., Rousseeuw P. J. (2009) *Finding Groups in Data: an Introduction to Cluster Analysis*. John Wiley & Sons, Hoboken.
- Kiers H. A. (1991) Simple Structure in Component Analysis Techniques for Mixtures of Qualitative and Quantitative Variables. *Psychometrika*, 56(2), 197-212.
- Lasek M., Pęczkowski M. (2010) Grupowanie zmiennych w procesach eksploracji danych (Data Mining). *Ekonomia i Zarządzanie*, 2(1), 83-94.
- Malarska A. (2005) *Statystyczna analiza danych wspomagana programem SPSS*. SPSS Polska, Kraków.
- Palla K., Ghahramani Z., Knowles D. A. (2012) A Nonparametric Variable Clustering Model. [w:] *Advances in Neural Information Processing Systems*, 2987-2995.
- Rand W. M. (1971) Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336), 846-850.
- Vigneau E., Chen M., Qannari E. M. (2015) ClustVarLV: An R Package for the Clustering of Variables Around Latent Variables. *The R Journal*, 7 (2), 134-148.
- Vigneau E., Qannari E. M. (2003) Clustering of Variables Around Latent Components. *Communications in Statistics-Simulation and Computation*, 32(4), 1131-1150.
- Wagner S., Wagner D. (2007) *Comparing Clusterings: an Overview*. Karlsruhe: Universität Karlsruhe, Fakultät für Informatik. http://www.cs.ucsb.edu/~veronika/MAE/wagner07_comparingclusterings.pdf.
- Yeung K. Y., Ruzzo W. L. (2001) Details of the Adjusted Rand Index and Clustering Algorithms, Supplement to the Paper An Empirical Study on Principal Component Analysis for Clustering Gene Expression Data. *Bioinformatics*, 17(9), 763-774.

ANALYSIS OF COMPETENCES RESOURCES USING THE CLUSTERING METHOD FOR NONMETRIC VARIABLES

Abstract: The article presents the application of a nonmetric variables clustering method to a study of competences resources in Poland. The analysis is based on results from a nationwide survey on Human Capital in Poland, including data both on hard and soft skills. The analytical procedure allowed for grouping competences assessed similarly into clusters and assigning to them synthetic variables – representatives, which were used to evaluate the relationships with chosen socio-demographic characteristics.

Keywords: competences, nonmetric variables, cluster analysis