

Analiza skupień w ocenie warunków pracy w krajach Unii Europejskiej

Unia Europejska (UE) uznaje, że zrozumienie warunków, w jakich pracują ludzie jest warunkiem niezbędnym do osiągnięcia poprawy jakości pracy, zwiększenia wydajności i wzrostu zatrudnienia. Jak stanowi art. 151 traktatu o UE¹, państwa członkowskie powinny aktywnie działać na rzecz „promowania zatrudnienia” oraz „poprawy warunków życia i pracy”, tak aby „umożliwić ich wyrównanie z jednoczesnym zachowaniem postępu”.

Do tradycyjnych miar bezpieczeństwa i higieny pracy należą wskaźniki wypadków przy pracy i chorób zawodowych, obliczane najczęściej jako liczba tych zdarzeń przypadających na określoną liczbę osób pracujących albo przepracowanych roboczogodzin. Wskaźniki te są powszechnie wykorzystywane, stanowiąc często jedyne kryterium bezpieczeństwa i higieny pracy (bhp). Ich wspólną cechą jest to, że sygnalizują występowanie problemów poprzez ich skutki, nie identyfikując jednak ich przyczyn². Jednak tradycyjne wskaźniki statystyczne nie są jedynym źródłem danych statystycznych w zakresie bhp. Istotnych informacji na ten temat dostarczają wyniki badań ankietowych prowadzonych przez Europejską Fundację na Rzecz Poprawy Warunków Życia i Pracy (*European Working Conditions Survey* — EWCS)³.

Zgodnie z definicją wyrażoną w normie PN-N-1801 z 2004 r., bhp to: *stan warunków i organizacji pracy oraz zachowań pracowników zapewniający wymagany poziom ochrony zdrowia i życia przed zagrożeniami występującymi w środowisku pracy*⁴. Z kolei Centralny Instytut Ochrony Pracy definiuje bhp następująco: *ogół norm prawnych oraz środków badawczych, organizacyjnych i technicznych mających na celu stworzenie pracownikowi takich warunków pracy, aby mógł on wykonywać pracę w sposób produktywny, bez narażania go na nieuzasadnione ryzyko wypadku lub choroby zawodowej oraz nadmierne obciążenie fizyczne i psychiczne*⁵. Można zatem przyjąć, że pod nazwą bhp kryje się zarówno zbiór przepisów, jak i zasad dotyczących bezpiecznego i higienicznego wykonywania pracy, a także osobna dziedzina wiedzy zajmująca się kształtowaniem właściwych warunków pracy.

¹ Art. 151, Dz. Urz. UE 2010, C83/3, <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:FULL:PL:PDF>.

² Pawłowska (2006), s. 5—7.

³ Europejska Fundacja na Rzecz Poprawy Warunków Życia i Pracy, www.eurofund.europa.eu.

⁴ Portal internetowy Asystent BHP, <http://asystentbhp.pl/art/definycja-bhp/1#.UjdazX9Cmhp>.

⁵ Centralny Instytut Ochrony Pracy, www.ciop.pl.

Począwszy od 1990 r. wspomniana Fundacja przeprowadza co 5 lat badania, które dostarczają jedynych w swoim rodzaju wyników dotyczących jakości warunków pracy. Odbyło się pięć edycji tego badania, co pozwala na zaobserwowanie i przeanalizowanie długofalowych trendów. Cele, jakie przyświecają badaniom EWCS, to:

- ocena i ilościowe ujęcie w sposób zharmonizowany warunków pracy zarówno pracowników, jak i osób pracujących na własny rachunek w całej Europie;
- analiza relacji między różnymi aspektami warunków pracy;
- określenie obaw i grup ryzyka, a także postępu;
- monitorowanie tendencji poprzez zapewnienie jednolitych wskaźników dotyczących tych kwestii;
- przyczynianie się do rozwoju polityki europejskiej.

Zakres tematyczny kwestionariusza znacznie rozszerzono od czasu pierwszej edycji badania, tak aby zapewnić szczegółowy obraz codziennej sytuacji kobiet i mężczyzn w ich miejscu pracy.

W ostatniej edycji badania wyodrębniono następujące zagadnienia:

- I. Kontekst pracy;
- II. Czas pracy;
- III. Natężenie pracy;
- IV. Czynniki fizyczne;
- V. Czynniki kognitywne;
- VI. Czynniki psychospołeczne;
- VII. Zdrowie i dobre samopoczucie;
- VIII. Umiejętność szkolenia i perspektywy rozwoju zawodowego;
- IX. Organizacja pracy;
- X. Relacje społeczne;
- XI. Satysfakcja z pracy;
- XII. Równowaga między życiem zawodowym i osobistym oraz bezpieczeństwo zawodowe;
- XIII. Przemoc, napastowanie, dyskryminacja.

W każdej edycji przeprowadzono bezpośrednie wywiady z losową próbą pracowników pracujących na etatach i osób pracujących na własny rachunek. Próba badawcza wykorzystywana w EWCS jest reprezentatywna dla osób od 15 roku życia⁶ zatrudnionych i mieszkających w kraju, w którym prowadzone jest badanie. W ramach piątej edycji badania w każdym z państw zastosowano wieloetapową warstwową próbę losową⁷. Prace terenowe były prowadzone od stycznia do czerwca 2010 r. W 27 państwach członkowskich UE oraz Norwegii, Chorwacji, Macedonii (b. Jugosłowiańskiej Republice Macedonii), Turcji, Albanii, Czarnogórze i Kosowie przeprowadzono wywiady z blisko 44000 pracowników, przy czym w Polsce wywiadów udzieliło 1500 respondentów.

⁶ W przypadku Hiszpanii, Wielkiej Brytanii oraz Norwegii dolną granicę wieku stanowi 16 lat.

⁷ Eurofund, http://www.eurofound.europa.eu/surveys/ewcs/2010/sampling_pl.htm.

W artykule zaprezentowano zastosowanie wybranych metod analizy skupień do oceny krajów członkowskich UE ze względu na warunki bezpieczeństwa i higieny pracy oceniane w ramach jednego z komponentów EWCS⁸. Dokładnie rzecz ujmując, cechy diagnostyczne pochodzą z badania *Zdrowie i dobre samopoczucie*. Klasyfikację krajów UE przeprowadzono za pomocą metody *k*-średnich — będącej przykładem metod niehierarchicznych oraz metody Warda — będącej przykładem metod hierarchicznych.

WYBÓR CECH DIAGNOSTYCZNYCH

Cechy diagnostyczne wybrano zgodnie z następującymi kryteriami doboru cech do analizy skupień^{9, 10}:

- uniwersalności — cechy diagnostyczne powinny mieć przydatność merytoryczną z punktu widzenia omawianej tematyki oraz mieć znaczenie społeczno-ekonomiczne;
- zmienności — analizowane cechy powinny wykazywać dostateczną zmienność przestrzenną, czyli być nośnikiem informacji, które różnicują badane obiekty. Dlatego też, aby to zweryfikować, dla cech wyjściowych wyznacza się wartość współczynnika zmienności (najczęściej przyjmuje się, że eliminacji podlegają te zmienne, dla których współczynnik zmienności osiąga wartość mniejszą od arbitralnie zadanej małej liczby dodatniej ε , często jako granicę tę przyjmuje się $\varepsilon = 0,1$);
- stopnia skorelowania — zbyt silna zależność korelacyjna pomiędzy dwiema cechami diagnostycznymi powoduje, że są one nośnikiem podobnych informacji, w związku z tym przyjmuje się, iż w przypadku uzyskania zbyt wysokiej wartości współczynnika korelacji pomiędzy analizowanymi cechami należy dokonać doboru ich reprezentanta, kierując się zazwyczaj przesłankami merytorycznymi. Za progowy poziom współczynnika korelacji przyjmuje się $r^* = 0,7$ ¹¹;
- ważności — sprawdzenia ważności cech, a tym samym eliminacji cech nieważnych dokonuje się w następstwie wyliczenia wartości współczynnika asy-

⁸ Celem Europejskiego Badania Warunków Pracy jest uzyskanie porównywalnych danych z zakresu warunków pracy w krajach europejskich. Wyniki uzyskane w ramach 5 edycji badania stanowiły podstawę dokonania grupowania państw członkowskich UE. Zainteresowania Autorki w prezentowanym opracowaniu skupiły się na implementacji wybranych metod analizy skupień do konkretnego tematu badawczego w tym przypadku była to ocena warunków pracy mieszkańców UE. Stąd w artykule Autorka nie dokonywała szerszych porównań wyników EWCS z innymi prowadzonymi badaniami w ramach omawianego zagadnienia badawczego. W kolejnym artykule szerzej natomiast zostaną omówione źródła danych statystycznych, co znajdzie odzwierciedlenie w zestawieniu wyników Europejskiego Badania Warunków Pracy z wynikami prezentowanymi w publikacjach GUS — *Warunki...* (2014); *Wypadki...* (2014).

⁹ Ostasiewicz (1998), s. 118.

¹⁰ Malina, Zieliś (1996), s. 85—89.

¹¹ Nowak (1990), s. 396.

metrii. Jeżeli rozkład zmiennej charakteryzuje się bardzo silną asymetrią lewostronną, wówczas zdecydowana większość badanych obiektów uzyskuje wartości powyżej średniej, zatem analizowana cecha nie różnicuje obiektów, większość z nich osiąga bowiem wysoki stopień nasycenia¹².

Oto przyjęte do analizy cechy diagnostyczne wraz z ich krótką charakterystyką¹³:

X₁ — odsetek osób (bardzo) dobrze poinformowanych o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy.

W zdecydowanej większości krajów członkowskich UE poziom poinformowania pracowników o zagrożeniach dla zdrowia i bezpieczeństwa z powodu wykonywanej pracy wzrósł w roku 2010 w stosunku do roku 2000. Stan wiedzy pracowników najbardziej poprawił się w Hiszpanii i Portugalii (odpowiednio o 7,4 p.proc. i 7,3 p.proc.). Jednak w tym okresie w przypadku 10 państw obserwowano zmniejszenie się odsetka osób dobrze poinformowanych o zagrożeniach w miejscu pracy, najbardziej we Francji (o blisko 7 p.proc.). W 2010 r. najlepszą sytuację pod tym względem odnotowano w Irlandii (97,1%), na Węgrzech (96,3%), Łotwie (95,8%), Słowacji (95,5%), w Wielkiej Brytanii (95,2%) i Polsce (94,0%). We wszystkich tych krajach stwierdzono zwiększenie odsetka pracowników dobrze poinformowanych o zagrożeniach, zarówno w stosunku do roku 2000 jak i 2005. Na końcu listy rankingowej znalazły się Holandia, Belgia, Włochy, Malta i Francja, przy czym w przypadku Malty w porównaniu do 2005 r. odsetek osób poinformowanych o warunkach bhp w miejscu pracy poprawił się dość znacznie (o ponad 10 p.proc.), a w pozostałych krajach (z wyjątkiem Włoch) odnotowano pogorszenie wartości wskaźnika w stosunku do roku 2000;

X₂ — odsetek osób uważających, że ich zdrowie lub bezpieczeństwo nie jest zagrożone z tytułu wykonywanej pracy zawodowej.

Zdecydowana większość ankietowanych uważała, iż ich zdrowie i bezpieczeństwo nie są w żaden sposób narażone na zagrożenia wynikające z wykonywanej pracy. W każdym z krajów UE odsetek osób deklarujących brak zagrożenia zdrowia i bezpieczeństwa z tytułu wykonywanej pracy przewyższał odsetek osób będących na ten temat odmiennego zdania. Mimo to w 9 państwach wskaźnik osób uważających swoją pracę za zagrażającą zdrowiu i życiu był niższy od średniej unijnej (24,2%) — najniższy w Danii (15,5%) oraz Holandii (16,1%) i Irlandii (16,9%). Najbardziej zaniepokojeni o stan swojego zdrowia i bezpieczeństwa byli Łotysze (47,2%), Szwedzi (41,3%), Estończycy (39,2%), Bułgarzy (38,7%) i Grecy (38,5%). W Polsce odsetek osób deklarujących, że stan

¹² Ostasiewicz (1998), s. 118.

¹³ Szczegółowo opracowany kwestionariusz realizowany w ramach piątego badania EWCS dostępny jest pod adresem file:///C:/Users/hp/Downloads/poland_pl.pdf.

zdrowia i bezpieczeństwa są zagrożone przez wykonywaną pracę wyniósł 31,6%. Jednak w naszym kraju odnotowano największą poprawę w stosunku do roku 2005 — odsetek osób deklarujących, że ich praca stanowi zagrożenie dla zdrowia i bezpieczeństwa spadł aż o 16 p.proc;

X₃ — odsetek osób uważających, że wykonywana praca ma przede wszystkim pozytywny wpływ na ich zdrowie.

W 2010 r. większość mieszkańców UE (67,7%) uważała, że praca nie ma wpływu na ich stan zdrowia. Odsetek ten był nawet wyższy w przypadku Brytyjczyków, Irlandczyków, Włochów, Niemców, Holendrów, Belgów i Portugalczyków. Wśród narodów, w których osoby uważające, że praca nie wywiera żadnego wpływu na ich zdrowie nie stanowiły nawet 50%, znaleźli się Finowie, Szwedzi, Grecy, Słowenicy i Łotysze. W Polsce było to 55,6%. Przeciętnie co czwarty mieszkaniec UE uważa, że jego praca ma negatywny wpływ na stan zdrowia, jednak na Łotwie odsetek ten przekroczył 52%. Słowenia i Estonia są natomiast przykładami krajów, w których odsetek osób stwierdzających negatywny wpływ pracy na stan zdrowia przewyższa odsetek tych, którzy są zdania, że wykonywane obowiązki zawodowe nie wpływają na zdrowie. Najrzadziej, bo zaledwie co 10 ankietowany Irlandczyk uznawał pejoratywny wpływ pracy na zdrowie. Również rzadko oceniali negatywnie wpływ pracy na stan zdrowia Brytyjczycy i Holendrzy — odpowiednio 14,4% oraz 17,4% respondentów. Interesujący jest przypadek Finlandii oraz Szwecji, gdzie odsetki osób oceniających pozytywny i negatywny wpływ pracy na zdrowie były bardzo zbliżone, z przewagą ankietowanych, którzy uznali pozytywne oddziaływanie pracy na stan zdrowia;

X₄ — odsetek osób wskazujących, że w okresie ostatnich 12 miesięcy nie zdarzyła im się nieobecność w pracy z powodu problemów zdrowotnych.

W UE odsetek osób deklarujących 15-dniową nieobecność w pracy z powodu problemów zdrowotnych zmalał o 20,2%. Można zatem postawić pytanie, czy pracownicy pozostają na zwolnieniu lekarskim krócej czy też ogólnie ich stan zdrowia polepszył się i nie przebywają na zwolnieniu lekarskim? Odpowiedź na to pytanie jest mniej optymistyczna aniżeli statystyka pokazująca absencję długoterminową. W 2010 r. w porównaniu do roku 2000 o blisko 10% spadł odsetek osób, które nie przebywały na zwolnieniu lekarskim, natomiast wskaźnik absencji krótkotrwałej od 1 do 15 dni wzrósł o 27,1%. Najwyższe wartości wskaźnika absencji długoterminowej odnotowano w Polsce (12,7%), Czechach, na Litwie i Słowenii (odpowiednio 12,2%, 12,1% i 12,0%). Poniżej 4% ankietowanych deklarowało ponad 15-dniową nieobecność w pracy z powodu problemów zdrowotnych w Estonii, Rumunii i Irlandii. Warto podkreślić, że długotrwała absencja staje się coraz rzadsza w zdecydowanej większości państw

członkowskich. Jednak w 6 krajach odsetek ten wzrósł, najbardziej znacząco w Polsce (4,5 p.proc.), na Cyprze (3,2 p.proc.), Malcie (2,4 p.proc.) i Litwie (2,3 p.proc.);

X₅ — odsetek osób, którym w okresie ostatnich 12 miesięcy nie zdarzyło się pracować podczas choroby.

Europejczycy dość często wykonują obowiązki służbowe podczas choroby — średnia unijna wynosi dla tej kategorii blisko 40%. Najczęściej pracowali, pomimo choroby, Słowenci (59,2% ankietowanych). W okresie ostatnich 12 miesięcy pracowało pomimo choroby ponad 50% mieszkańców Malty, Danii, Szwecji, Wielkiej Brytanii i Finlandii. Podczas choroby najrzadziej pracowali Bułgarzy, Włosi, Portugalczycy, a także Polacy;

X₆ — odsetek osób uważających, że będą zdolne do wykonywania tej samej pracy, którą wykonują obecnie, po ukończeniu 60 roku życia.

Mieszkańcy UE uważali na ogół, że po osiągnięciu wieku 60 lat będą zdolni do wykonywania tej samej pracy. Zaledwie 16,1% ogółu respondentów badania uznało, że nie będzie chciało pełnić swoich obowiązków zawodowych w wieku 60+. Odsetek ten uległ zwiększeniu w porównaniu z badaniem z roku 2000 i 2005, co czwarty respondent wskazał, że nie będzie w stanie wykonywać swojej dotychczasowej pracy po 60 roku życia. Najmniej optymistycznie na wykonywanie obowiązków zawodowych po 60 roku życia zapatrywali się Portugalczycy (29,7%), Francuzi (27,0%), Słowacy (25%), Luksemburczycy (24,6%) i Łotysze (22,3%). Optymiści to ci, którzy chcieliby utrzymać dotychczasową pracę po osiągnięciu wieku 60 lat. Najrzadziej wyrażali niechęć wobec wykonywania pracy po ukończeniu 60 roku życia — Holendrzy (3,9%), Bułgarzy (6,5%) i Szwedzi (9,6%). W Polsce 15% respondentów uznało, że nie będzie w stanie wykonywać obecnych obowiązków zawodowych w wieku 60+.

ANALIZA SKUPIEŃ — ZAŁOŻENIE TEORETYCZNE

Analiza skupień¹⁴ oznacza segmentację lub klastrowanie danych, nazywana również grupowaniem obiektowym, stanowi jedną z najbardziej znanych metod eksploracji danych. Jest zaawansowanym narzędziem analizy danych służącym do grupowania zmiennych lub obiektów w nieznane grupy. Analiza ta jest metodą klasyfikacji, która polega na dzieleniu (zazwyczaj wielowymiarowym) zbioru danych na grupy (skupienia) w taki sposób, by elementy w tej samej grupie były do siebie podobne, a jednocześnie jak najbardziej odmienne od elementów z pozostałych grup¹⁵. Analiza skupień może być wykorzystywana do wykrywania struktur w danych bez wyprowadzania interpretacji, wykrywa zatem jedynie

¹⁴ Pojęcie zostało wprowadzone w roku 1939 przez Tyrona.

¹⁵ Gatnar, Walesiak (2004), s. 310—317.

struktury bez wyjaśniania, dlaczego one występują. Pośrednim celem takiej analizy jest również weryfikacja jednorodności danych. Jeśli możliwe jest wyróżnienie skupień, wówczas danych nie można uznać za jednorodne.

Podczas przeprowadzania analizy skupień dzielimy zbiór danych na grupy przy użyciu różnorodnych kryteriów, pozwalających na określenie podobieństwa obiektów w danej grupie. Taka kategoryzacja ma na celu maksymalne rozróżnienie elementów pomiędzy grupami, przy jednoczesnym zachowaniu ich spójności — podobieństwa — w ramach tej samej grupy. Można powiedzieć, że dążymy do maksymalnej homogeniczności w grupach, przy zachowaniu jak największej heterogeniczności¹⁶ pomiędzy grupami.

W literaturze przedmiotu istnieje ogromna liczba metod grupowania i zarazem mnogość propozycji odnoszących się do ich klasyfikacji (Sneath, Sokal, 1973). Metody taksonomii numerycznej można podzielić na¹⁷:

- 1) hierarchiczne,
- 2) optymalizacyjno-iteracyjne,
- 3) obszarowe.

Ze względu na zdefiniowany we wstępie cel pracy oraz prezentowane w artykule struktury obiektów, szerzej omówione zostaną jedynie dwa zasadnicze typy algorytmów grupowania danych^{18,19}, tj.:

- metody hierarchiczne, dzielące się ze względu na kryterium rozpoczęcia grupowania na:
 - a) procedury aglomeracyjne (*agglomerative*),
 - b) procedury deglomeracyjne (podziałowe) (*divisive*).

Ogólnie procedury aglomeracyjne polegają na iteracyjnym łączeniu obiektów w coraz to większe lub coraz to mniejsze skupienia. Pozwalają one na budowanie hierarchii skupień w zależności od odległości między nimi, prowadząc do stworzenia dendrogramu, czyli hierarchii drzewkowej elementów analizowanego zbioru. Początkowo zakłada się, że każdy z obiektów stanowi osobne skupienie, następnie etapowo łączy się je w podzbiory, podgrupy najbardziej do siebie podobne, aż do otrzymania finalnie jednego skupienia zawierającego wszystkie obserwacje. W ten sposób powstaje uporządkowane zestawienie podziałów na segmenty.

W procedurach deglomeracyjnych algorytm postępowania jest odwrotny. Początkowy zbiór obiektów traktowany jest jako jedno skupienie, w kolejnych krokach dzieli się jedno ze skupień na dwie części, a liczba skupień ule-

¹⁶ Marek (1989), s. 19—22.

¹⁷ Frątczak (2009), s. 120—122.

¹⁸ Harańczyk (2005), s. 77—79.

¹⁹ W opracowaniu nie zastosowano dwóch innych typów algorytmów analizy skupień, tj.: metody rozmytej analizy skupień — jest to grupa metod, w których każdy obiekt może zostać przydzielony do więcej niż jednej kategorii z podaniem prawdopodobieństwa przynależności; *biclustering* (*2-way cluster analysis*) — metoda, w której analiza skupień odbywa się poprzez jednoczesne oddziaływanie na obiekty i ich cechy.

ga zwiększeniu o jeden. W konsekwencji po $n-1$ krokach uzyskuje się n skupień jednoelementowych²⁰. Wyniki uzyskane za pomocą metod podziałowych mają analogiczną strukturę do uzyskanych za pomocą procedur aglomeracyjnych²¹.

Grupowanie za pomocą metod hierarchicznych nie wymaga wcześniejszego ustalenia liczby skupień. Wybór liczby skupień może zostać dokonany na samym końcu analizy. Dopiero dysponując gotowym dendrogramem trzeba zdecydować, na jakiej wysokości zostanie on przecięty;

- metody optymalizacyjno-iteracyjne (niehierarchiczne) — polegają na przenoszeniu obiektów z jednego skupienia do innego w poszukiwaniu najlepszego zestawu skupień według zadanego kryterium i wymagają, w odróżnieniu od metod hierarchicznych, wcześniejszego podania liczby skupień, na które ma być podzielony zbiór.

Wybór liczby skupień ma duży wpływ na jakość uzyskanej segmentacji. Podanie dużej liczby skupień powoduje, że wyznaczone skupienia są wewnętrznie jednorodne, jednak utrudniona jest interpretacja uzyskanych wyników i stosowanie ich w praktyce. Z kolei mała liczba skupień prowadzi do znacznie mniejszej wewnętrznej homogeniczności skupienia. Istotną wadą metod jest brak wewnętrznego porządkowania w obrębie skupienia, co jest możliwe do wykazania w metodach hierarchicznych. Ponadto jakkolwiek zmiana liczby skupień powoduje zupełnie nowy podział, który nie zawiera w sobie wcześniej uzyskanych rezultatów. Do najbardziej znanych metod optymalizacyjno-iteracyjnych należą metody: k -średnich, k -centroidów, Forgy-Jancey'a, Wisharta, Thorndike'a (Grabiński, 1992).

Grupowanie metodą k -średnich

Metoda k -średnich wykorzystywana jest do analizy dużych ilości danych i jest najbardziej klasycznym algorytmem analizy skupień. Jej istota polega na zredukowaniu dużej ilości nagromadzonych informacji do kilku podstawowych kategorii, dzięki czemu łatwiejsze staje się zorientowanie w danym zjawisku i wyciągnięcie wniosków uogólniających. Jej zastosowanie może prowadzić do zmniejszenia nakładów czasu i kosztów badań przez ograniczenie rozważań do najbardziej typowych faktów, zjawisk czy obiektów, przy stosunkowo niewielkich stratach informacji²².

Działanie metody k -średnich można opisać w następujący sposób — w pierwszym etapie należy ustalić wartość parametru k , która decyduje o liczbie grup wyodrębnionych ze zbioru danych; następnie losowo wybierana jest liczba k reprezentantów w taki sposób, aby byli oni możliwie jak najbardziej od siebie oddaleni. Wybrane elementy stanowią załączki grup (prototypy). W następnej

²⁰ Frątczak (2009), s. 121.

²¹ Malina (2004), s. 59 i 60.

²² Józwiak, Podgórski (1997), s. 330—340.

kolejności każdy element zbioru przypisywany jest do najbliższej mu grupy. Na tym etapie wyznaczone są grupy początkowe. Dla każdej z grup obliczany jest jej środek na podstawie średniej arytmetycznej współrzędnych obiektów do niej należących. W kolejnym kroku następuje wyznaczenie nowych środków grup i powtórne przydzielanie do najbliższej (ze względu na odległość od poszczególnych centroidów) grupy każdego z obiektów. Tak długo nowe środki grup będą wyznaczane i sprawdzana będzie poprawność przynależności elementów do grup, jak długo występować będzie przenoszenie obiektów pomiędzy skupieniami. Jeśli w kolejnych dwóch przebiegach algorytmu nie nastąpi żadna zmiana w dokonanym podziale (mówi się wówczas, że została osiągnięta stabilizacja) przetwarzanie danych zostaje zakończone²³. Wadą tej metody jest konieczność odgórnego określenia liczby skupień występujących w danym zbiorze, dlatego też zaleca się powtórzenie metody dla różnych wartości k i wybranie tej, dla której zbiór danych podzielony jest najlepiej²⁴.

Problem, z jakim spotyka się badacz w metodzie k -średnich, to ustalenie wstępnego podziału na liczbę skupień. Po sprawdzeniu rozdziału obiektów na różną liczbę skupień, ostatecznego podziału można dokonać w sposób losowy lub na podstawie oceny ekspertów, która wynika z intuicji lub znajomości przedmiotu badań. Można także wykorzystać inne metody taksonomiczne.

W dużym stopniu o jakości uzyskanych wyników decyduje liczba skupień, wstępne ustalenie środków skupień oraz określenie, w jaki sposób będzie obliczona odległość między obiektami.

W przypadku zmiennych ilościowych najczęściej w metodzie k -średnich stosuje się odległość euklidesową, czyli odległość geometryczną w przestrzeni wielowymiarowej, obliczaną jako pierwiastek z sumy kwadratów różnicy pomiędzy wartościami i -tej cechy dla dwóch badanych obiektów x i y .

Odległości euklidesowe, podobnie jak kwadraty odległości euklidesowych, wyliczane są na podstawie surowych danych, a nie danych standaryzowanych. Należy zatem pamiętać, że na odległości duży wpływ mają różnice jednostek między wymiarami, według których są one wyliczane. Przeprowadzenie procedury normalizacji zmiennych zapewnia eliminację ograniczeń formalnych i trudności interpretacyjnych²⁵. W praktyce dostępne są następujące miary:

- standaryzacja,
- unitaryzacja,
- rangowanie,
- przekształcenia ilorazowe²⁶.

Autorka zdecydowała się poddać wszystkie zmienne diagnostyczne procedurze standaryzacji. W wyniku przeprowadzonego procesu zmienne diagnostyczne X zostały przekształcone w zmienną Z . Doprowadziło to w konsekwencji do wy-

²³ Jajuga (1993), s. 55—60.

²⁴ Migut (2009), s. 77 i 78.

²⁵ Borys (1978), s. 132.

²⁶ Bąk (1999), s. 52.

równania zróżnicowania oraz poziomu wartości cech diagnostycznych. Po dokonaniu standaryzacji wariancje cech są równe 1, a średnie arytmetyczne — 0. W związku z tym każda ze zmiennych w jednakowym stopniu wpływa na finalne wyniki prowadzonej analizy²⁷.

Metoda Warda

Metoda Warda jest jedną z aglomeracyjnych metod grupowania i stosowana jest w badaniach empirycznych zarówno w odniesieniu do klasyfikacji obiektów, jak i cech. W metodzie tej odległość między grupami jest definiowana jako moduł różnicy między sumami kwadratów odległości punktów od środków grup, do których punkty te należą²⁸.

Punktem wyjścia jest tu macierz euklidesowych odległości między klasyfikowanymi obiektami. Odległość między dwiema klasami mierzy się jako wartość, o jaką zwiększy się suma kwadratów odległości od nowych środków ciężkości po połączeniu. Metoda Warda polega na łączeniu takich skupień, które zapewniają minimum sumy kwadratów odległości od środka ciężkości nowego skupienia, które tworzą. Dąży ona do uzyskania raczej małych skupień i jest uznawana za bardzo efektywną. W wyniku analizy otrzymujemy dendrogram będący graficzną interpretacją uzyskanych efektów. W zależności od przyjętych założeń badania, w tym zwłaszcza akceptowanej odległości taksonomicznej między obiektami ze względu na zaproponowany zestaw cech, możemy wyróżniać większe lub mniejsze skupienia, a co za tym idzie, mniejszą lub większą ich liczbę²⁹.

Wśród wielu metod graficznego zobrazowania analizy skupień do najpopularniejszych należą drzewa binarne — dendrogramy, które są wynikiem grupowania hierarchicznego. Dendrogram to diagram w kształcie drzewa ukazujący związki pomiędzy grupami na podstawie przyjętego kryterium określającego podobieństwo. „Listki” takiego drzewa przedstawiają elementy zbioru poddawanego analizie, natomiast „gałęzie” są graficznym przedstawieniem ich grup. Dendrogram nie jest więc osobną metodą, a jedynie sposobem zilustrowania analizy.

ANALIZA SKUPIEŃ — WYNIKI BADAŃ

Analizę klasyfikacji wykonano przy zastosowaniu opisanych od strony teoretycznej metody *k*-średnich jako metody niehierarchicznej oraz metody Warda — jako przykład metody hierarchicznej.

²⁷ Pluta (1986), s. 117.

²⁸ Malina (2004), s. 62 i 63.

²⁹ Marek (1989), s. 110 i 119.

TABL. 1. MIERNIKI STATYSTYCZNE WEDŁUG WYBRANYCH CECH DIAGNOSTYCZNYCH

Cechy diagnostyczne	Minimum	Maksimum	Średnia	Odchy- lenie standar- dowe	Współ- czynnik zmien- ności	Współ- czynnik asymetrii
	w %					
Jestem (bardzo) dobrze poinformowany o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy	80,90 (Francja)	97,10 (Irlandia)	90,93	4,045	4,448	-0,491
Uważam, że moje zdrowie lub bezpieczeństwo nie jest zagrożone przez wykonywaną przeze mnie pracę	52,80 (Łotwa)	84,40 (Dania)	71,79	8,740	12,173	-0,301
Moja praca ma głównie pozytywny wpływ na moje zdrowie	2,70 (Portugalia)	28,70 (Finlandia)	9,46	6,468	68,374	2,020
W okresie ostatnich 12 miesięcy nie zdarzyła mi się nieobecność w pracy z powodu problemów zdrowotnych	34,20 (Finlandia)	84,20 (Rumunia)	58,89	12,903	21,910	-0,386
W okresie ostatnich 12 miesięcy nie zdarzyło mi się pracować podczas choroby	40,80 (Słowenia)	77,30 (Bułgaria)	58,90	10,341	17,558	0,174
Uważam, że będę zdolny do wykonywania tej samej pracy, którą wykonuję obecnie, kiedy skończę 60 lat	25,60 (Słowenia)	75,40 (Holandia)	55,66	10,613	19,068	-0,528

Źródło: opracowanie własne na podstawie EWCS przy wykorzystaniu pakietu statystycznego SPSS 21.0.

W celu zidentyfikowania grup podobnych pod względem oceny warunków pracy wykorzystano metodę Warda. Na wykresie przedstawiono dendrogram uzyskany na podstawie wyników EWCS z badania *Zdrowie i dobre samopoczucie*. W literaturze przedmiotu najczęściej podaje się trzy metody pozwalające na wyznaczenie krytycznej wartości odległości, przy której powinno nastąpić przerwanie łączenia w klasy^{30,31}:

³⁰ Panek (2009), s. 120—123.

³¹ Stanisław (2007), s. 141 i 142.

- wskazanie maksimum miernika $d_i - d_{i-1}$,
- obliczenie miernika zaproponowanego przez T. Grabińskiego:

$$q_i = \max \left\{ \frac{d_i}{d_{i-1}} \right\}$$

- zastosowanie reguły R. Mojeny

$$d_{i+1} > \bar{d} + k \cdot S(d)$$

gdzie:

$i = 2, 3, \dots, n-1$,

d_i = długość i -tego wiązania (i -tej gałęzi drzewa),

$\bar{d}$, $S(d)$ — średnia arytmetyczna i odchylenie standardowe długości wiązań,

k — stała, której wartość według reguły R. Mojeny powinna zawierać się w przedziale $[2,5; 3,5]$.

W innych pracach sugerowana jest optymalna wartość parametru równa 1,25^{32, 33}.

Autorka zdecydowała się na zastosowanie I reguły R. Mojeny (*Upper tail*), przyjmując wartość stałej k na poziomie 1,25, przy $\bar{d} = 30,98$, $S(d) = 37,57$. W konsekwencji uzyskano wynik $d_{i+1} > 77,94$, co odpowiada przecięciu dendrogramu powyżej wiązania o numerze 23. W wyniku metody Warda wspartej drzewem klasyfikacyjnym, wyznaczony moment optymalnego cięcia dendrogramu wskazał, że zbiorowość należałoby podzielić na cztery grupy (wykres). W grupie 1, najliczniejszej spośród wyróżnionych na drzewie klasyfikacyjnym, znalazły się przede wszystkim nowe kraje członkowskie, spośród państw dawnej UE-15 pojawiły się również Grecja, Hiszpania, Portugalia i Włochy. Grupa 2 zdominowana była przez państwa, które tworzyły dawną Wspólnotę Europejską. Spośród nowych członków UE w tej grupie są Czechy. Grupa 3 jest najmniejsza, złożona jedynie z Finlandii i Szwecji. Ostatnia grupa 4 ma w swojej strukturze trzech przedstawicieli z dawnych państw UE-15 oraz Maltę i Słowenię, których akces do UE odbył się w 2004 r.

Stosując metodę k -średnich podjęto próbę podziału zbiorowości na 3, 4 lub 5 klas. Z punktu widzenia statystycznego najwłaściwszą będzie ta liczba klas, dla której w zestawieniu statystyki F wszystkie z nich będą statystycznie istotne (23)³⁴. W przypadku wszystkich podziałów na grupy uzyskano satysfakcjonujące wyniki z punktu widzenia wartości statystyki F oraz jej poziomów istotności (tabl. 2). O grupowaniu zbiorowości na 3 klasy zdecydowały jednak przesłanki merytoryczne.

³² Miligan, Cooper (1985), s. 159—179.

³³ Stanisław (2007), s. 141 i 142.

³⁴ Malarska, Nowak-Sapota (1999), s. 22—30.

TABL. 2. ANALIZA WARIANCJI DLA METODY k -średnich

Cechy diagnostyczne	Analiza wariancji — podział zbioru na grupy					
	3 grupy		4 grupy		5 grup	
	F	istotność p	F	istotność p	F	istotność p
Jestem (bardzo) dobrze poinformowany o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy	10,056	0,001	6,095	0,003	6,901	0,001

TABL. 2. ANALIZA WARIANCJI DLA METODY k -ŚREDNICH (dok.)

Cechy diagnostyczne	Analiza wariancji — podział zbioru na grupy					
	3 grupy		4 grupy		5 grup	
	F	istotność p	F	istotność p	F	istotność p
Uważam, że moje zdrowie lub bezpieczeństwo nie jest zagrożone przez wykonywaną przeze mnie pracę	6,081	0,007	14,458	0,000	12,122	0,000
Moja praca ma głównie pozytywny wpływ na moje zdrowie	4,765	0,018	21,365	0,000	19,324	0,000
W okresie ostatnich 12 miesięcy nie zdarzyła mi się nieobecność w pracy z powodów problemów zdrowotnych	19,097	0,000	14,337	0,000	11,571	0,000
W okresie ostatnich 12 miesięcy nie zdarzyło mi się pracować podczas choroby	5,581	0,010	6,694	0,002	3,714	0,019
Nie uważam, abym był zdolny do wykonywania tej samej pracy, którą wykonuję obecnie, kiedy skończę 60 lat	14,902	0,000	8,227	0,001	12,672	0,000

Ź r ó d ł o: jak przy tabl. 1.

W grupie 1 znalazło się (zestawiono) 13 państw, w tym 8 krajów włączonych do struktur UE w 2004 r. oraz Bułgaria i Rumunia (dołączyły do Wspólnoty w 2007 r.). Z krajów dawnej UE-15 w grupie tej znalazły się Grecja, Hiszpania i Portugalia, zatem kraje, które najbardziej odczuły skutki kryzysu finansowego. Grupa 2 to 7 państw, przy czym nie znalazł się w niej żaden kraj, którego akcesja odbyła się po roku 2003. Grupa 3 jest tak samo liczna jak grupa 2 — składa się zarówno ze „starych” jak i „nowych” krajów członkowskich.

ZESTAWIENIE PRZYNALEŻNOŚCI DO GRUPY WEDŁUG WPŁYWU PRACY NA ZDROWIE I BEZPIECZEŃSTWO UZYSKANEJ METODĄ k -ŚREDNICH

Wyszczególnienie	Kraje
Grupa 1 — umiarkowane warunki	Bułgaria, Cypr, Czechy, Estonia, Grecja, Hiszpania, Litwa, Łotwa, Polska, Portugalia, Rumunia, Słowacja, Węgry
Grupa 2 — sprzyjające warunki	Dania, Finlandia, Irlandia, Holandia, Niemcy, Szwecja, Wielka Brytania
Grupa 3 — niekorzystne warunki	Austria, Belgia, Francja, Luksemburg, Malta, Słowenia, Włochy

Ź r ó d ł o: jak przy tabl. 1.

Analiza międzygrupowa ze względu na wybrane cechy diagnostyczne

W grupie 1 znalazły się państwa, dla których poziom poinformowania o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy jest o 3 p.proc. wyższy aniżeli średnia unijna (tabl. 3). Grupa ta jest jednocześnie najbardziej homogeniczna w ww. zakresie, co potwierdza najniższa wartość współczynnika zmienności. Współczynnik zmienności klasyczny lub pozycyjny stanowi ważne źródło informacji na temat stopnia zróżnicowania grup³⁵. Generalnie mieszkańcy tych krajów wypadają lepiej pod względem odsetka osób deklarujących, że nie wystąpiły u nich problemy zdrowotne, które doprowadziłyby do ich nieobecności w pracy, jak i odsetka osób wskazujących, że nie zdarzyło im się pracować podczas choroby. Niemniej mieszkańcy tych krajów znacznie rzadziej, aniżeli wynika to ze średniej unijnej, deklarowali pozytywne oddziaływanie pracy na ich stan zdrowia — wartość wskaźnika w grupie 1 wynosi 6,8%, wobec 9,46% dla UE-27.

TABL. 3. CHARAKTERYSTYKA GRUPY 1 UZYSKANEJ METODĄ *k*-średnich W %

Mienniki statystyczne	Jestem (bardzo) dobrze poinformowany o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy	Uważam, że moje zdrowie lub bezpieczeństwo nie jest zagrożone przez wykonywaną przeze mnie pracę	Moja praca ma głównie pozytywny wpływ na moje zdrowie	W okresie ostatnich 12 miesięcy nie zdarzyła mi się nieobecność w pracy z powodu problemów zdrowotnych	W okresie ostatnich 12 miesięcy nie zdarzyło mi się pracować podczas choroby	Uważam, że będę zdolny do wykonywania tej samej pracy, którą wykonuję obecnie, kiedy skończę 60 lat
Średnia	93,70	66,90	6,80	69,70	65,00	50,90
Odchylenie standardowe	2,69	7,06	2,03	7,41	7,72	5,93
Współczynnik zmienności	2,87	10,55	30,05	10,63	11,89	11,65

Źródło: jak przy tabl. 1.

Grupa 2 wypada na tle pozostałych dwóch grup najkorzystniej ze względu na zadane cechy diagnostyczne (tabl. 4). W przypadku 4 spośród 6 zmiennych wartości średnich wewnątrzgrupowych są wyższe od wartości średniej ogólnej. Jest to grupa zdecydowanie najbardziej jednorodna ze względu na deklarowaną przez mieszkańców zdolność do wykonywania tej samej pracy, którą wykonują obecnie osoby po ukończeniu 60 roku życia. Współczynnik zmienności w przypadku tej zmiennej wynosił 5,6%, wobec 19% dla UE ogółem. Najbardziej znacząca różnica w tej grupie w stosunku do wartości średnich dla UE była widoczna w przypadku odsetka osób deklarujących pozytywny wpływ pracy na stan zdrowia (średnia wewnątrzgrupowa przewyższa średnią UE o ponad 60%) oraz odsetka tych, któ-

³⁵ Szerzej na temat homogeniczności grup uzyskanych w analizie skupień w publikacji A. Młodaka (2006), s. 80.

rzy uważają, że będą zdolni do wykonywania tej samej pracy po ukończeniu 60 roku życia (średnia wewnątrzgrupowa przewyższa średnią UE o ponad 23%).

TABL. 4. CHARAKTERYSTYKA GRUPY 2 UZYSKANEJ METODĄ k -średnich W %

Mienniki statystyczne	Jestem (bardzo) dobrze poinformowany o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy	Uważam, że moje zdrowie lub bezpieczeństwo nie jest zagrożone przez wykonywaną przeze mnie pracę	Moja praca ma głównie pozytywny wpływ na moje zdrowie	W okresie ostatnich 12 miesięcy nie zdarzyła mi się nieobecność w pracy z powodu problemów zdrowotnych	W okresie ostatnich 12 miesięcy nie zdarzyło mi się pracować podczas choroby	Uważam, że będę zdolny do wykonywania tej samej pracy, którą wykonuję obecnie, kiedy skończę 60 lat
Średnia	91,30	78,20	15,20	45,60	52,30	68,70
Odchylenie standardowe	3,69	9,29	9,71	9,91	6,93	3,84
Współczynnik zmienności	4,04	11,87	63,97	21,74	13,26	5,60

Źródło: jak przy tabl. 1.

W grupie 3 jedynie w przypadku oceny braku zagrożenia dla zdrowia lub bezpieczeństwa ze strony wykonywanej pracy wartość średniej wewnątrzgrupowej jest wyższa niż średnia ogólna dla UE (tabl. 5). W tej kategorii obserwowano jednocześnie największą homogeniczność grupy 3, co potwierdza wartość współczynnika zmienności będącą o 32 p.proc. niższą w stosunku do wartości ogółem. Ponadto właśnie w grupie 3 wewnątrzgrupowa wartość współczynnika zmienności jest najniższa spośród rozpatrywanych pozostałych wartości oceniających dyspersję wewnątrzgrupową w zakresie deklaracji, że wykonywana praca nie stanowi zagrożenia dla zdrowia lub bezpieczeństwa. We wszystkich pozostałych zmiennych wartości średnich wewnątrzgrupowych są niższe od średniej ogółem. Zatem należy przyjąć, że mieszkańcy krajów, które znalazły się w tej grupie najgorzej oceniają warunki swojej pracy ze względu na wybrane cechy diagnostyczne.

TABL. 5. CHARAKTERYSTYKA GRUPY 3 UZYSKANEJ METODĄ k -średnich W %

Mienniki statystyczne	Jestem (bardzo) dobrze poinformowany o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy	Uważam, że moje zdrowie lub bezpieczeństwo nie jest zagrożone przez wykonywaną przeze mnie pracę	Moja praca ma głównie pozytywny wpływ na moje zdrowie	W okresie ostatnich 12 miesięcy nie zdarzyła mi się nieobecność w pracy z powodu problemów zdrowotnych	W okresie ostatnich 12 miesięcy nie zdarzyło mi się pracować podczas choroby	Uważam, że będę zdolny do wykonywania tej samej pracy, którą wykonuję obecnie, kiedy skończę 60 lat
Średnia	86,60	74,60	7,70	54,00	54,60	49,70
Odchylenie standardowe	3,30	6,16	4,15	9,41	12,51	11,57
Współczynnik zmienności	3,81	8,26	53,65	17,42	22,91	23,28

Źródło: jak przy tabl. 1.

Wnioski

Tabl. 6 stanowi podsumowanie analizy porównawczej uzyskanych wyników, w kolumnie 1 znalazły się kolejno wyszczególnione kraje członkowskie UE, w kolumnie 2 wyniki klasyfikacji uzyskane metodą Warda, a w kolumnie 3 wyniki uzyskane metodą k -średnich (cyfry w kolumnach odpowiadają numerom grup, do których zostały przyporządkowane poszczególne kraje). Wybrany zestaw cech diagnostycznych oceniających warunki pracy mieszkańców UE zarówno metodą Warda (jako przykład metody hierarchicznej), jak i metodą k -średnich (jako przykład metody niehierarchicznej) dostarczył bardzo podobnych rezultatów. Jedynie w przypadku dwóch krajów spośród uwzględnionych w analizie nastąpiły wyraźne różnice w przyporządkowaniu do odpowiedniego klastra. Podział wykonany za pomocą metody Warda wskazał, że Czechy znalazły się w grupie 2, podczas gdy w metodzie k -średnich trafiły do grupy 1, w której znajduje się większość nowych państw członkowskich. W przypadku Włoch metoda k -średnich przyporządkowała to państwo do grupy 3 zatem wypadającej najsłabiej ze względu na zadane kryteria diagnostyczne, a metoda Warda do grupy 1.

Należy zwrócić uwagę, że w wyniku analizy przeprowadzonej metodą Warda wyodrębniono 4 skupienia. W przypadku podziału wykonanego metodą k -średnich wartości testu F oraz odpowiadające im poziomy istotności *ex post* pozwoliły na podział zbiorowości na 3 grupy. Warto jednak zwrócić uwagę, że w klasyfikacji uzyskanej przy zastosowaniu metody Warda, jedna z grup jest tylko dwuelementowa i oba kraje (Finlandia i Szwecja) w przypadku metody k -średnich znalazły się w tej samej grupie, czyli w 2. Autorka przyjęła, że podział zbioru będzie optymalny, gdy wyróżnione zostaną 3 grupy, nie tylko ze względu na wartości parametrów uzyskane w analizie wariancji, ale przede wszystkim mając na względzie przesłanki merytoryczne i potrzebę przeprowadzenia dalszej analizy międzygrupowej. Dodatkowo w metodzie k -średnich każdy podział powyżej 3 grup prowadził do wyodrębnienia grup jedno- bądź dwuelementowych, co stanowiło znaczne utrudnienie w prowadzeniu analiz porównawczych.

TABL. 6. ANALIZA PORÓWNAWCZA WYNIKÓW UZYSKANYCH METODĄ HIERARCHICZNĄ I NIEHIERARCHICZNĄ (numery grup, do których należą kraje)

K r a j e	Metoda Warda	Metoda k -średnich
Austria	2	3
Belgia	4	3
Bułgaria	1	1
Cypr	1	1
Czechy	2	1
Dania	2	2

TABL. 6. ANALIZA PORÓWNAWCZA WYNIKÓW UZYSKANYCH METODĄ HIERARCHICZNĄ I NIEHIERARCHICZNĄ (numery grup, do których należą kraje) (dok.)

K r a j e	Metoda Warda	Metoda <i>k</i> -średnich
Estonia	1	1
Finlandia	3	2
Francja	4	3
Grecja	1	1
Hiszpania	1	1
Irlandia	2	2
Litwa	1	1
Luksemburg	4	3
Łotwa	1	1
Malta	4	3
Holandia	2	2
Niemcy	2	2
Polska	1	1
Portugalia	1	1
Rumunia	1	1
Słowacja	1	1
Słowenia	4	3
Szwecja	3	2
Węgry	1	1
Wielka Brytania	2	2
Włochy	1	3

Ź r ó d ł o: jak przy tabl. 1.

Uzyskany za pomocą metody *k*-średnich podział wykazał, że najlepiej wypada ze względu na zadane kryteria diagnostyczne grupa 2, zdominowana przez państwa dawnej UE-15. Grupa 1, gdzie znalazły się przede wszystkim kraje, których akces nastąpił w ostatnim dziesięcioleciu jest najliczniejsza i zarazem najmniej zróżnicowana ze względu na stopień poinformowania o zagrożeniach zdrowia i bezpieczeństwa wynikających z wykonywanej pracy. Grupa ta jednocześnie została określona jako umiarkowana ze względu na syntetyczną ocenę warunków pracy i zatrudnienia. Przedstawiciele grupy 3 byli najmniej zadowoleni z warunków pracy opisanych za pomocą cech diagnostycznych z zakresu *Zdrowie i dobre samopoczucie*.

dr Elżbieta Roszko-Wójtowicz — Uniwersytet Łódzki

LITERATURA

- Bąk A. (1999), *Wykorzystanie metod wielowymiarowej analizy porównawczej w analizie finansowej*, „Prace Naukowe Akademii Ekonomicznej we Wrocławiu”, nr 811
- Borys T. (1978), *Metody normowania cech w statystycznych badaniach porównawczych*, „Przegląd Statystyczny”, nr 2, [w:] Kunasz M., *Przykład zastosowania metod WAP do analizy procesów gospodarowania zasobami ludzkimi w przedsiębiorstwie*, http://mikroekonomia.net/system/publication_files/904/original/11.pdf?1315223800
- Frątczak E. (red.) (2009), *Wielowymiarowa analiza statystyczna. Teoria — przykłady zastosowań z systemem SAS*, SGH
- Gatnar E., Walesiak M. (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej, Wrocław
- Grabiński T. (1992), *Metody taksonometrii*, Akademia Ekonomiczna w Krakowie
- Harańczyk G. (2005), *Analiza skupień na przykładzie segmentacji nowotworów*, StatSoft Polska, Kraków, <http://www.statsoft.pl/czytelnia.html>
- Jajuga K. (1993), *Statystyczna analiza wielowymiarowa*, Wydawnictwo Naukowe PWN, Warszawa
- Józwiak J., Podgórski J. (1997), *Statystyka od podstaw*, PWE, Warszawa
- Małarska A., Nowak-Sapota W. (1999), *Warunki mieszkaniowe ludności miast w Polsce w latach 1978—1995 (empiryczna weryfikacja metody k-średnich)*, „Wiadomości statystyczne”, Nr 11
- Malina A. (2004), *Wielowymiarowa analiza przestrzennego różnicowania struktury gospodarki Polski według województw*, Wydawnictwo AE w Krakowie
- Malina A., Zieliński A. (1996), *Taksonomiczna analiza przestrzennego różnicowania jakości życia ludności w Polsce w 1994 r.* [w:] *Ekonometryczne modelowanie danych finansowo-księgowych*, Nowak E., Urbaniak M. (red.), UMCS Lublin, [za:] Kunasz M., *Przykład zastosowania metod WAP do analizy procesów gospodarowania zasobami ludzkimi w przedsiębiorstwie*, http://mikroekonomia.net/system/publication_files/904/original/11.pdf?1315223800
- Marek T. (1989), *Analiza skupień w badaniach empirycznych*, PWN, Warszawa
- Migut G. (2009), *Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku*, StatSoft Polska, http://www.statsoft.pl/czytelnia/artykuly/Zastosowanie_teknik.pdf
- Miligan G. W., Cooper M. C. (1985), *An examination of procedures for determining the number of groups in a data set*, „Psychometrika”, nr 50
- Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, Centrum Doradztwa i Informacji DIFIN, Warszawa
- Nowak E. (1990), *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWN, Warszawa, [za:] Wojnar J., *Zastosowanie metod taksonomicznych do klasyfikacji państw Unii Europejskiej*, Zeszyty Naukowe, nr 8, Świętokrzyskie Centrum Edukacji na Odległość, http://www.sceno.edu.pl/cms_tmp/2438_Jolanta%20Wojnar%20-%20III%20SCENO.pdf
- Ostasiewicz W. (red.) (1998), *Statystyczne metody analizy danych*, Akademia Ekonomiczna we Wrocławiu
- Panek T. (2009), *Statystyczne metody wielowymiarowej analizy porównawczej*, SGH
- Pawłowska Z. (2006), *Jak oceniać funkcjonowanie przedsiębiorstwa w obszarze bezpieczeństwa i higieny pracy*, „Bezpieczeństwo Pracy — nauka i praktyka”, nr 2, <http://www.ciop.pl/18490>
- Pluta W. (1986), *Wielowymiarowa analiza porównawcza w modelowaniu ekonometrycznym*, PWN, Warszawa
- Sneath P. H. A., Sokal R. R. (1973), *Numerical Taxonomy*, W. H. Freeman and Co., San Francisco

Stanisz A. (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, „Analizy wielowymiarowe”, t. 3, StatSoft Polska, Kraków
Warunki pracy w 2013 r. (2014), GUS
Wypadki przy pracy i problemy zdrowotne związane z pracą (2014), GUS

SUMMARY

The aim of the article is to describe the use of selected methods of cluster analysis — hierarchical and non-hierarchical — to assess the member states of the European Union due to health and safety conditions of work. The analysis is based on data of the European Working Conditions Survey. Based on the answers given by the respondents for the health and well-being were constructed six diagnostic features. Assess working conditions were made using methods Ward and k-average. The use of the two methods yielded very similar results. Only in the case of two of the 27 countries included in the analysis, there were clear differences in the allocation to the respective groups.

РЕЗЮМЕ

Целью статьи является представление использования избранных методов анализа кластеров – иерархических и неиерархических — для оценки стран членов Европейского союза в отношении к условиям безопасности и гигиены труда. В анализе были использованы данные Европейского обследования условий труда.

На основе ответов респондентов касающихся здоровья и хорошего самочувствия было разработано 6 диагностических признаков. Оценка условий труда была сделана методом Уорда и k-средних. В использовании обоих методов были получены сходные результаты. Только в случае двух из 27 стран, включенных в анализ, выступил существенные различия в распределении оценок условий труда по соответствующим группам.