

**Sugerowany przypis:** Chybalski F., *Analiza struktury wiekowej oraz płciowej członków OFE z wykorzystaniem metod taksonomicznych* [w:] Chybalski F., Staniec I. (red.), *10 lat reformy emerytalnej w Polsce. Efekty, szanse, perspektywy i zagrożenia*. Monografie Politechniki Łódzkiej, Łódź 2009, s. 122-132

## **ANALIZA STRUKTURY WIEKOWEJ ORAZ PŁCIOWEJ CZŁONKÓW OFE Z WYKORZYSTANIEM METOD TAKSONOMICZNYCH**

*Filip Chybalski*  
*Katedra Zarządzania, Politechnika Łódzka*

### **1. Wstęp**

Reforma emerytalna w Polsce, zapoczątkowana w 1999 roku, której głównym elementem było wprowadzenie II filaru emerytalnego i funkcjonujących w jego ramach otwartych funduszy emerytalnych, sprawiła, że obywatele współdecydują o wysokości swojego przyszłego świadczenia emerytalnego między innymi poprzez wybór OFE. Decyzja o wyborze funduszu emerytalnego może wynikać z dwóch przesłanek: podjęcia pracy po raz pierwszy, z czym wiąże się obowiązek przystąpienia do OFE (w drodze własnego wyboru lub w drodze losowania) bądź wskutek zmiany funduszu emerytalnego, do czego każdy ubezpieczony ma prawo. Tym samym na rynku OFE mamy do czynienia ze stałym napływem nowych członków do poszczególnych funduszy oraz z przepływem tych członków pomiędzy funduszami (czyli z tzw. transferami). Przez to struktury wiekowe oraz płciowe

członków funduszy emerytalnych mogą podlegać ciągłym zmianom oraz istotnie różnić się między sobą.

Celem artykułu jest analiza struktury wiekowej oraz płciowej członków otwartych funduszy emerytalnych oraz ocena podobieństwa tych struktur w przekroju całego rynku OFE. Analiza ta pozwoli odpowiedzieć na następujące pytania, istotne przede wszystkim z punktu widzenia towarzystw emerytalnych zarządzających funduszami, zabiegających stale o nowych członków:

- Czy OFE istotnie różnią się między sobą pod względem badanych struktur?
- Czy występują fundusze emerytalne preferowane przez dane grupy wiekowe bądź płć?

## **2. Materiał empiryczny**

Badanie analizy struktur wiekowych oraz płciowych członków OFE przeprowadzono na podstawie danych z 2008 roku (stan na dzień 31.12.2008), pochodzących z Biuletynu Kwartalnego OFE za IV kwartał 2008, publikowanego przez Komisję Nadzoru Finansowego. Dane pierwotne zawierały informacje o ogólnej liczbie członków poszczególnych OFE, liczbie mężczyzn i liczbie kobiet ubezpieczonych w poszczególnych funduszach oraz o liczebności grup rozróżnionych wg płci i wieku. Grupowanie wg wieku następowało wg następujących przedziałów: 17 lat i mniej, 18-20 lat, 21-25 lat, 26-30 lat, 31-35 lat, 36-40 lat, 41-45 lat, 46-50 lat, 51 lat i więcej.

Na podstawie powyższych danych surowych obliczono wskaźniki struktury dla poszczególnych cech, uzyskując w ten sposób wyrażone w % struktury członków OFE ze względu na wiek, płć oraz ze względu na wiek i płć łącznie. W pierwszym przypadku uzyskano 9 grup, w drugim 2 grupy a w trzecim 18 grup (dla każdej płci wyróżniono 9 grup wiekowych, stąd uzyskano 18 następujących grup płć/wiek: mężczyźni w wieku 17 lat i mniej, kobiety w wieku 17 lat i mniej, mężczyźni w wieku 18-20 lat, kobiety w wieku 18-20 lat, itd.). Wszystkie struktury przedstawiono w postaci macierzy (zwanych dalej macierzami struktur: wiekowej, płciowej, płć/wiek), które zawierały wartości badanych cech (zmiennych) dla poszczególnych obiektów, czyli otwartych funduszy emerytalnych.

## **3. Metodologia**

W analizie podobieństwa otwartych funduszy emerytalnych pod względem struktur płciowych ich członków ze względu na małą liczbę wyróżnionych cech

(równą 2), dokonano analizy otrzymanych wskaźników struktury, wyrażających udział kobiet i mężczyzn w ogólnej liczbie członków danego OFE, ze tego typu wskaźnikami dla pozostałych funduszy oraz uśrednionymi dla całego rynku. Natomiast w analizie podobieństwa otwartych funduszy emerytalnych pod względem struktur: wiekowej oraz płeć/wiek, ze względu na dużą liczbę wyróżnionych cech (odpowiednio równą 9 i 18), wykorzystano hierarchiczne metody aglomeracyjne, stanowiące jedną z metod wchodzących w skład szerzej rozumianej analizy skupień. Punktem wyjścia w hierarchicznych metodach aglomeracyjnych jest macierz odległości natomiast ostatecznym rezultatem zazwyczaj drzewo hierarchiczne, grupujące badane obiekty ze względu na wyróżnione cechy. Procedura zastosowania hierarchicznych metod aglomeracyjnych jest następująca [1, s. 30]:

1. Utwórz skupienia, spośród których każde zawiera jeden z badanych obiektów.
2. Na podstawie macierzy odległości między skupieniami znajdź dwa obiekty najbardziej podobne, czyli te, dla których odległość jest najmniejsza.
3. Wybrane w punkcie 2 obiekty połącz w jedno skupienie.
4. Oblicz odległość między nowopowstałym skupieniem a wszystkimi pozostałymi i stwórz nową macierz odległości.
5. Powtarzaj kroki 2-4 aż do momentu, gdy pozostanie jedno skupienie obejmujące wszystkie obiekty.

W ten sposób z liczby skupień równej liczbie porównywanych obiektów na sam koniec otrzymujemy jedno duże skupienie, zawierające wszystkie obiekty.

W hierarchicznych metodach aglomeracyjnych mogą być obliczane różne odległości pomiędzy badanymi obiektami, w tym:

- odległość euklidesowa:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{i,k} - x_{j,k})^2} \quad (1)$$

- odległość miejska,

$$d_{ij} = \sum_{k=1}^n |x_{i,k} - x_{j,k}| \quad (2)$$

- odległość Czebyszewa.

$$d_{ij} = \max |x_{i,k} - x_{j,k}| \quad (3)$$

gdzie:

$d_{ij}$  – odległość pomiędzy  $i$ -tym a  $j$ -tym obiektem,  
 $x_{i,k}$  – wartość  $k$ -tej cechy dla  $i$ -tego obiektu,  
 $x_{j,k}$  – wartość  $k$ -tej cechy dla  $j$ -tego obiektu,  
 $n$  – liczba cech charakteryzujących obiekty.

W procesie grupowania obiektów można zastosować jedną z następujących metod, zwanych metodami wiązania [1, s. 30-31; 2, s. 82-83; 3]:

- metoda pojedynczego wiązania, zwana także metodą najbliższego sąsiedztwa, w której odległość między dwoma skupieniami jest wyznaczana przez odległość pomiędzy dwoma najbliższymi obiektami (sąsiadami), pochodzącymi z tych dwóch skupień,
- metoda pełnego wiązania, zwana również metodą najdalszego sąsiedztwa, w której odległość pomiędzy dwoma skupieniami jest wyznaczana przez odległość pomiędzy dwoma najbardziej oddalonymi obiektami (sąsiadami), pochodzącymi z tych dwóch skupień,
- metoda średnich połączeń, zwana też metodą średniej grupowej, gdzie odległość pomiędzy dwoma skupieniami jest równa średniej odległości pomiędzy wszystkimi parami obiektów należących do tychże skupień,
- metoda średnich połączeń ważonych, która jest podobna do metody średnich połączeń z tą tylko różnicą, że w obliczeniach przyjmuje się wagi proporcjonalne do liczby obiektów zawartych w danym skupieniu,
- metoda środków ciężkości, gdzie odległość pomiędzy dwoma skupieniami jest równa odległości pomiędzy ich środkami ciężkości. Środek ciężkości w przestrzeni wielowymiarowej jest zdefiniowanym przez jej wymiary średnim punktem,
- metoda ważonych środków ciężkości, podobna do poprzedniej, lecz dodatkowo uwzględniająca wielkości porównywanych skupień mierzoną liczbą zawartych w nich obiektów,
- metoda Warda, wykorzystująca w procesie szacowania odległości analizę wariancji.

W przeprowadzonym badaniu obliczono odległości miejskie zgodnie ze wzorem (2), gdzie przyjęto, że:

$d_{ij}$  – odległość miejska między  $i$ -tym a  $j$ -tym OFE pod względem badanej struktury,

$x_{i,k}$  – udział  $k$ -tej wyróżnionej kategorii badanej cechy (wieku lub płci/wieku) dla  $i$ -tego OFE,

$x_{j,k}$  – udział  $k$ -tej wyróżnionej kategorii badanej cechy (wieku lub płci/wieku) dla  $j$ -tego OFE,

$n$  – liczba wyróżnionych kategorii (równa 9 lub 18, w zależności od badanej struktury).

Na tej podstawie skonstruowano macierze odległości między funduszami pod względem dwóch badanych struktur dla danych pochodzących z końca 2008

roku. Wybór właśnie odległości miejskiej był podyktowany faktem, iż może ona występować również w postaci średnich odchyłeń wartości cech [2, s. 39]. W przeprowadzonym badaniu podzielenie odległości miejskiej dla dowolnych dwóch OFE przez liczbę wyróżnionych kategorii, pozwala na interpretację tej miary jako wyrażonej w punktach procentowych średniej różnicy w udziałach poszczególnych kategorii (grup wiekowych lub grup płeć/wiek) w ogólnej liczbie członków tych OFE. Obliczono średnie odległości dla każdej skonstruowanej macierzy (pomijając wartości na głównej przekątnej), czyli średnią odległość między dwoma funduszami, maksymalną i minimalną odległość w macierzy a także maksymalną i minimalną średnią odległość pomiędzy danym OFE a pozostałymi funduszami. Każdą wymienioną wyżej odległość podzielono następnie przez liczbę wyróżnionych kategorii dla badanej cechy i otrzymano odległość w przeliczeniu na liczbę tych kategorii. Następnie skonstruowane macierze odległości miejskich wykorzystano w hierarchicznej metodzie aglomeracji, stosując w procedurze grupowania obiektów metodę średnich połączeń, otrzymując w wyniku drzewa hierarchiczne.

Należy dodać, iż hierarchiczne metody aglomeracji nie są pozbawione wad.

Wśród nich należy wymienić następujące:

- otrzymane raz skupienie nie może być rozłączone, co uniemożliwia skorygowanie popełnionego ewentualnie wcześniej błędu,
- nie można narzucić z góry liczby tworzonych grup bądź liczby obiektów w poszczególnych grupach, co uniemożliwia badaczowi kontrolowanie tworzenia drzewa hierarchicznego o określonej budowie. Przez to konstrukcja tych drzew może utrudniać ocenę zróżnicowania obiektów na podstawie ich analizy wzrokowej,
- ponieważ jest to metoda grupowania a nie porządkowania, w efekcie jej zastosowania można stwierdzić, czy obiekty są podobne, czy też nie, nie można natomiast ich uporządkować ze względu na wyróżnione w analizie cechy,
- są to metody bezwzorcowe, jednak istnieje często możliwość określenia hipotetycznego wzorca (np. jako średniej wartości cechy wśród badanych obiektów) i uwzględnienia tego wzorca w analizie i drzewie hierarchicznym. Dzięki temu można porównać badane obiekty nie tylko pomiędzy sobą, ale także z owym hipotetycznym wzorcem.

Wskazane wady, szczególnie te, dotyczące wyłącznie pomiaru podobieństwa bez możliwości porządkowania obiektów, oraz brak wzorca, nie są bardzo istotne z punktu widzenia przeprowadzonego badania.

Ponadto badając występowanie ewentualnych zależności pomiędzy wybranymi zmiennymi, analizowanymi w badaniu, wykorzystano współczynnik korelacji liniowej Pearsona oraz test jego istotności.

#### **4. Wyniki i wnioski z badania**

W tabeli 1 przedstawiono strukturę płciową członków poszczególnych OFE oraz całego rynku.

**Tabela 1.** Struktura płciowa członków poszczególnych OFE oraz całego rynku (udziały w %)

OFE	Mężczyźni	Kobiety
AEGON	57,88	42,12
AIG	55,83	44,17
Allianz	55,14	44,86
AXA	51,40	48,60
Bankowy	54,83	45,17
CU	51,32	48,68
Generali	55,22	44,78
ING	50,06	49,94
Nordea	50,95	49,05
Pekao	56,99	43,01
Pocztylion	51,37	48,63
Polsat	59,52	40,48
PZU	51,31	48,69
WARTA	58,53	41,47
<b>Razem</b>	<b>52,65</b>	<b>47,35</b>

*Źródło: obliczenia własne na podstawie danych KNF.*

W przekroju całego rynku, 52,65% członków otwartych funduszy emerytalnych stanowią mężczyźni a 47,35% kobiety. Największe odchylenie od tych wskaźników rynkowych miało miejsce na koniec 2008 roku w następujących OFE: Polsat (6,87 punktu procentowego), Warta (5,88 punktu procentowego) oraz Aegon (5,23 punktu procentowego) i w każdym z wymienionych przypadków oznaczało większy udział mężczyzn wśród członków tych OFE aniżeli w przekroju całego rynku. Można więc uznać, iż w przypadku tych OFE występowała największa preferencja mężczyzn wśród członków funduszu. Natomiast największa preferencja kobiet wśród członków funduszu miała miejsce w funduszu ING (2,59 punktu procentowego). W przypadku tego OFE kobiety stanowiły niemal połowę jego członków (49,05%). W pozostałych OFE, w przypadku których wskaźnik struktury udziału kobiet był

większy od wskaźnika rynkowego, różnica wynosiła poniżej 2 punktów procentowych.

W tab. 2 przedstawiono wartości wybranych odległości, obliczone na podstawie skonstruowanych macierzy odległości badanych struktur członków OFE: wiekowej oraz płciowo-wiekowej.

**Tabela 2.** Wybrane odległości dla struktury wiekowej oraz struktury płć/wiek członków OFE (w punktach procentowych)

Rodzaj struktury	Wybrane charakterystyki liczbowe	Odległość	Odległość w przeliczeniu na kategorię
Wiekowa	Średnia odległość między dwoma OFE	23,78	2,64
	Maksymalna odległość między dwoma OFE	59,79	6,64
	Minimalna odległość między dwoma OFE	4,54	0,50
	Maksymalna średnia odległość między danym OFE a pozostałymi OFE	37,45	4,16
	Minimalna średnia odległości między danym OFE a pozostałymi OFE	18,43	2,05
Wiek/płeć	Średnia odległość między dwoma OFE	32,64	1,81
	Maksymalna odległość między dwoma OFE	76,20	4,23
	Minimalna odległość między dwoma OFE	8,39	0,47
	Maksymalna średnia odległość między danym OFE a pozostałymi OFE	49,56	2,75
	Minimalna średnia odległość między danym OFE a pozostałymi OFE	25,21	1,40

*Źródło: obliczenia własne na podstawie danych KNF.*

Analiza danych zawartych w tabeli 2 pozwala sformułować następujące wnioski dla struktury wiekowej członków OFE:

- średnia odległość między dwoma OFE, równa 23,78 punktu procentowego, oznacza, po podzieleniu przez liczbę wyróżnionych kategorii (równą 9), że dwa OFE różniły się pomiędzy sobą udziałem liczebności poszczególnych grup wiekowych w ogólnej liczbie członków każdego z tych OFE przeciętnie o 2,64 punktu procentowego. Jest to niska wartość oznaczająca, iż przeciętnie ujmując struktury wiekowe członków OFE były do siebie zbliżone;
- maksymalna odległość między dwoma OFE wyniosła 59,79 punktu procentowego, co w przeliczeniu na wyróżnione kategorie było równe

6,64 punktu procentowego i stanowiło 251% przeciętnej odległości z całej macierzy w przeliczeniu na kategorię . Odległość ta dotyczyła funduszy CU oraz Polsat, czyli odpowiednio największego i najmniejszego funduszu na rynku i oznaczała, iż różnią się one między sobą zdecydowanie bardziej aniżeli wynosi przeciętna różnica pomiędzy dwoma OFE w przekroju całego rynku;

- minimalna odległość pomiędzy dwoma wyniosła 4,54 punktu procentowego, co w przeliczeniu na kategorię wiekową wyniosło 0,50 punktu procentowego i oznaczało bardzo istotne podobieństwo tych dwóch OFE: Warta i Nordea, pod względem struktur wiekowych ich członków;
- maksymalna średnia odległość pomiędzy danym OFE a pozostałymi funduszami wyniosła 37,45 punktu procentowego i dotyczyła funduszu Polsat. W przeliczeniu na kategorię było to 4,16 punktu procentowego, co oznacza, że różnica w udziale kategorii wiekowej członków w tym funduszu i w pozostałych funduszach wyniosła przeciętnie 4,16 punktu procentowego i była zdecydowanie wyższa, aniżeli w pozostałych OFE;
- minimalna średnia odległość pomiędzy danym OFE a pozostałymi funduszami wyniosła 18,43 punktu procentowego i dotyczyła funduszu Generali. W przeliczeniu na kategorię wiekową było to 2,05 punktu procentowego, czyli o połowę mniej aniżeli w przypadku funduszu najbardziej oddalonego od pozostałych.

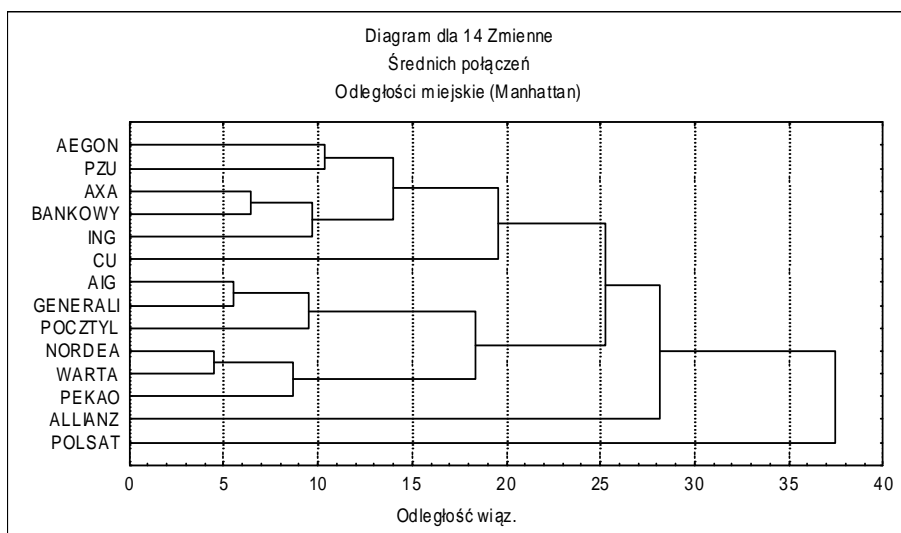
W przypadku struktury płeć/wiek można natomiast sformułować następujące wnioski:

- średnia odległość w macierzy wyniosła 32,64 punktu procentowego, co w przeliczeniu na kategorię płeć/wiek (było ich 18) wyniosło 1,81 punktu procentowego. Oznacza to, iż uwzględniając kombinację dwóch cech: płci i wieku członków funduszy, zróżnicowanie OFE pod tym względem było jeszcze mniejsze, gdyż przeciętna różnica udziału kategorii płeć/wiek w dwóch OFE wyniosła 1,81 punktu procentowego, a w przypadku struktury wiekowej było to 2,64 punktu procentowego;
- maksymalna odległość pomiędzy dwoma OFE wyniosła 76,20 punktu procentowego (w przeliczeniu na kategorię – 4,23 punktu procentowego) i dotyczyła, jak w przypadku struktury wiekowej, OFE CU i OFE Polsat. Odległość ta w przeliczeniu na kategorię była mniejsza aniżeli w przypadku struktury wiekowej;
- minimalna odległość pomiędzy dwoma OFE wyniosła 8,39 punktu procentowego (0,47 punktu procentowego w przeliczeniu na kategorię) i dotyczyła OFE AIG i OFE Generali;

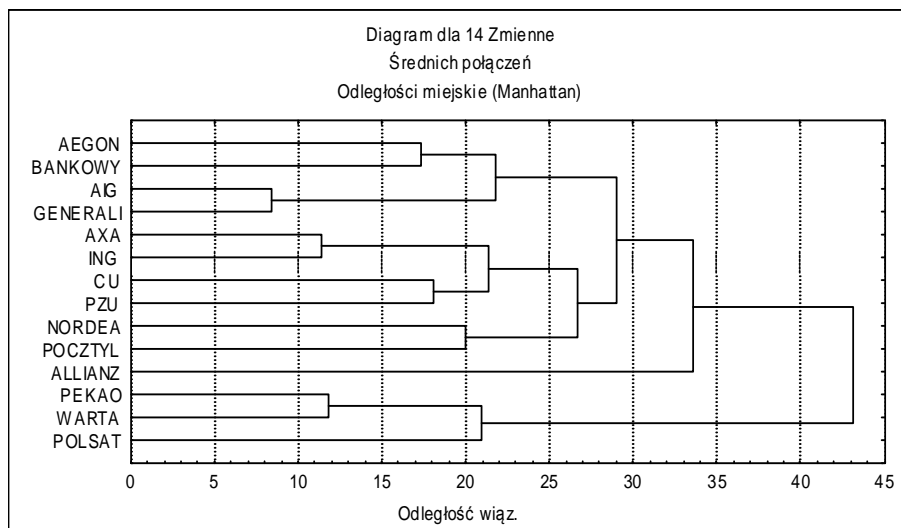


- maksymalna średnia odległość pomiędzy danym OFE a pozostałymi funduszami wyniosła 49,56 punktu procentowego (2,75 punktu procentowego w przeliczeniu na kategorię) i dotyczyła ponownie OFE Polsat;
- minimalna średnia odległość pomiędzy danym OFE a pozostałymi wyniosła 25,21 punktu procentowego (1,40 punktu procentowego w przeliczeniu na kategorię) i dotyczyła OFE Generali, podobnie jak w przypadku struktury wiekowej.

Aby dokonać dokładnej analizy podobieństwa otwartych funduszy emerytalnych pod względem badanych struktur, oprócz macierzy odległości dodatkowo posłużono się drzewami hierarchicznymi, skonstruowanymi za pomocą opisanej w poprzednim punkcie hierarchicznej metody aglomeracji (zob. rys. 1 i 2).



**Rys. 1.** Drzewo hierarchiczne dla struktur wiekowych członków OFE  
*Źródło: opracowanie własne.*



**Rys. 2.** Drzewo hierarchiczne dla struktur płęć/wiek członków OFE  
*Źródło: opracowanie własne.*

W przypadku struktury wiekowej członków OFE podobne są do siebie następujące OFE (podane parami): Nordea i Warta, AIG i Generali, AXA i Bankowy. Natomiast najbardziej oddalonymi pod tym względem funduszami od pozostałych są: Polsat, CU i Allianz. W przypadku struktury płęć/wiek najbardziej podobne są do siebie następujące fundusze: AIG i Generali; AXA i ING oraz PEKAO i WARTA. Natomiast funduszami najbardziej różniącymi się od pozostałych są: Polsat, CU i Allianz.

Dodatkowo, posługując się współczynnikiem korelacji liniowej Pearsona zbadano, czy występuje związek pomiędzy średnią odległością danego funduszu od pozostałych pod względem struktury wiekowej oraz struktury płęć/wiek a wielkością funduszu mierzoną liczbą jego członków. Obliczone współczynniki korelacji były bliskie zeru, co oznacza, że nie występuje związek pomiędzy wielkością funduszu a jego mierzonym odległością miejską odróżnieniem się od średniej rynkowej struktury wiekowej oraz płciowo-wiekowej. Ponadto sprawdzono, czy dane grupy wiekowe charakteryzują się preferencją funduszy większych lub mniejszych. W tym celu zbadano korelacje pomiędzy udziałem poszczególnych grup wiekowych w ogólnej liczbie członków poszczególnych OFE a ogólną liczbą członków tych OFE, będącą miarą ich wielkości. Wyniki tej analizy zawarto w tab. 3.

**Tabela 3.** Korelacja pomiędzy udziałem danej grupy wiekowej w ogólnej liczbie członków danego OFE a wielkością OFE

Grupa wiekowa	Korelacja
17 lat i mniej	-0,13
18-20	-0,43
21-25	<b>-0,51</b>
26-30	-0,24
31-35	<b>0,67</b>
36-40	<b>0,61</b>
41-45	<b>0,55</b>
46-50	0,42
51 lat i więcej	0,12

*Źródło: obliczenia własne na podstawie danych KNF.*

Grubą czcionką oznaczono współczynniki korelacji istotne statystycznie na poziomie istotności  $\alpha=0,05$ . Wnioski, jakie można sformułować na podstawie wyników analizy korelacji zawartych w tabeli 3, są następujące:

- w przypadku grup wiekowych 31-35 lat, 36-40 la oraz 41-45 lat występuje dodatnia korelacja pomiędzy udziałem tych grup w ogólnej liczbie członków OFE a wielkością tych OFE, co może oznaczać, iż w tych grupach wiekowych preferowane są duże fundusze;
- w grupie wiekowej 21-25 lat występuje ujemna korelacja pomiędzy udziałem tej grupy w ogólnej liczbie członków OFE a wielkością OFE, co może oznaczać, iż osoby w wieku 21-25 lat preferują mniejsze fundusze. Ta preferencja może wynikać z faktu, iż w losowaniu funduszu emerytalnego z powodu nie dokonania jego wyboru, udział biorą osoby wchodzące na rynek pracy, czyli w znaczącej części osoby właśnie z tej grupy wiekowej. Jednocześnie z losowania wykluczone są największe fundusze, więc osoby te nie mają możliwości stać się ich członkami w drodze losowania. Stąd może występować preferencja małych OFE w tej grupie wiekowej.

## 5. Podsumowanie

Przeprowadzone badanie pokazało, iż analiza odległości oraz hierarchiczne metody aglomeracji mogą mieć istotne zastosowanie do badania struktur

wiekowych i płciowych członków OFE, w tym szczególnie do ich porównywania pod względem podobieństwa. Wykorzystać ponadto można współczynnik korelacji, który umożliwia badanie ewentualnych zależności pomiędzy odległościami czy też strukturami a wielkościami OFE, co umożliwia udzielenie odpowiedzi na pytanie, czy odróżnianie się OFE pod względem badanych struktur jest powiązane z jego wielkością oraz czy wybrane grupy wiekowe preferują fundusze większe bądź mniejsze.

Analiza struktur członków OFE może mieć istotne znaczenie dla osób zajmujących się polityką akwizycyjną i reklamową w powszechnych towarzystwach emerytalnych, zarządzających funduszami. Pozwala ona bowiem wskazać grupy wiekowe, które preferują określone fundusze a także odpowiedzieć na pytanie, czy dane OFE są chętniej wybierane przez kobiety czy też przez mężczyzn. Poza tym analiza podobieństwa struktur umożliwia porównywanie funduszy pomiędzy sobą i identyfikowanie mniej lub bardziej podobnych obiektów.

## **Literatura**

- [1] **Gatnar E.:** *Symboliczne metody klasyfikacji danych*, Wydawnictwo Naukowe PWN, Warszawa 1998
- [2] **Nowak E.:** *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1990
- [3] **Sneath P.H.A., Sokal R.R.:** *Numerical Taxonomy*, Freeman, San Francisco 1973