

Coraz więcej analiz danych społeczno-ekonomicznych przeprowadzanych jest na podstawie badań ankietowych gospodarstw domowych, czy też tej samej grupy respondentów indywidualnych powtarzanych kilkakrotnie w następujących po sobie okresach czasu. Przykładem takich badań są chociażby badania panelowe tj.: Diagnoza Społeczna¹, Eurobarometr², World Value Survey³, czy też Panel Study of Income Dynamics⁴. Zadawane respondentom pytania, w tego typu badaniach mierzone są najczęściej na skalach jakościowych (nominalnej, czy porządkowej). Jednym z możliwych, nadal bardzo popularnych sposobów opisu struktury danych jakościowych jest ich graficzna prezentacja. Takie rozwiązanie stosuje się głównie w przypadku obserwacji dwuwymiarowych, które można przedstawić na płaszczyźnie i dokonać wstępnej analizy wzrokowej całego zbioru. Dynamiczny rozwój technologii informacyjnych, przyczynił się do coraz większej popularności wielowymiarowych metod statystycznej analizy danych. Jednak założeniem wielu z nich jest jednorodność badanego zbioru obserwacji. Problem niejednorodności analizowanego zbioru nabiera szczególnego znaczenia, gdy przeprowadzane badania powtarzane są kilkakrotnie, tj., w kilku następujących po sobie okresach czasu. Wtedy bardzo często, po prostu zakłada się (bez wcześniejszej weryfikacji), że analizowany zbiór danych jest zbiorem jednorodnym, a analizy statystyczne przeprowadzane są dla każdego z okresów z osobna.

W świetle powyższego, a także w oparciu o wyniki przeprowadzonej przez autorkę krytycznej analizy literatury przedmiotu **głównym celem** projektu badawczego jest zbadanie możliwości wykorzystania modeli dyskretnych zmiennych ukrytych w badaniu jednorodności panelowych danych społeczno-ekonomicznych mierzonych na słabych skalach pomiaru.

Wybór **podjęcia metodologicznego** został zdeterminowany specyfiką założeń badawczych. Identyfikacja jednorodnych struktur w społeczno-ekonomicznych zbiorach danych nastąpi z wykorzystaniem różnych wariantów modeli dyskretnych zmiennych ukrytych.

Modele zmiennych ukrytych zaliczyć należy do metod wielowymiarowej analizy danych jakościowych. Modele te można wpisać również w nurt podejścia modelowego w taksonomii, wykorzystującego ideę mieszanek rozkładów. Modele mieszanek rozkładów dla zmiennych jakościowych (mierzonych na skalach słabych), zwane są zatem modelami (*latent class models*) lub analizą klas ukrytych (*latent class analysis*). Główną ideę tego podejścia jest podział niejednorodnej zbiorowości statystycznej na klasy, które nie są bezpośrednio mierzalne (tzw. zmiana ukryta modelu). Do grupy modeli dyskretnych zmiennych ukrytych zaliczyć należy również coraz bardziej popularne ukryte modele Markowa. Modele te wykorzystywane są do analiz panelowych zbiorów danych o niejednorodnej strukturze, w których liczba klas (stanów) jest nieznaną. Celem badań jest już wtedy nie tylko identyfikacja jednorodnych struktur, ale również oszacowanie prawdopodobieństw przejścia pomiędzy nimi. Niejednorodność zbioru przedstawiana jest wtedy w ujęciu dynamicznym.

Proces testowania hipotez będzie realizowany m.in. na podstawie analiz porównawczych różnych wariantów modeli dyskretnych zmiennych ukrytych (np. wykorzystujących teorię reakcji na pozycję, uwzględniających zmienne towarzyszące, ograniczenia parametrów, czy też wielopoziomową specyfikę zbioru danych). Jakość dopasowania każdego z szacowanych modeli zostanie poddana ocenie m.in. na podstawie kryteriów informacyjnych AIC [Akaike 1974]⁵, BIC [Schwarz 1978]⁶, czy też indeksu *S* [Bartolucci i in. 2013, s. 68]⁷.

W badaniach wykorzystane zostaną rzeczywiste zbiory społeczno-ekonomiczne. Całość obliczeń w projekcie wykonana zostanie z wykorzystaniem bibliotek oraz autorskich procedur programu **R**.

¹ <http://www.diagnoza.com/>

² <http://www.gesis.org/eurobarometer-data-service/home/>

³ <http://www.worldvaluessurvey.org/WVSContents.jsp>

⁴ <http://psidonline.isr.umich.edu/>

⁵ Akaike H., 1974, *A new look at statistical model identification*, IEEE Transactions on Automatic Control, 19, 716-723.

⁶ Schwarz G., 1978, *Estimating the dimension of a model*, Annals of Statistics, 6,461-464.

⁷ Bartolucci F., Farcomeni A., Pennoni F., 2013, *Latent Markov Models for Longitudinal Data*, Chapman and Hall/CRC press, Boca Raton.