

Wstęp

Zjawiska złożone określane są jako zjawiska niepodlegające bezpośredniemu pomiarowi, które opisywane są najczęściej przez pewną liczbę zmiennych. Jednym z takich zjawisk złożonych jest zrównoważony rozwój.

W ocenie zjawisk złożonych pomocnym narzędziem okazują się metody wielowymiarowej analizy porównawczej (WAP). Otrzymana w wyniku ich zastosowania miara syntetyczna umożliwia porządkowanie liniowe obiektów (np. krajów, województw, powiatów, gmin, przedsiębiorstw itp.) w wielowymiarowych przestrzeniach zmiennych diagnostycznych. Znacznie ułatwia to prowadzenie analiz porównawczych.

O poprawności, wiarygodności i rzetelności przeprowadzonych badań (w tym także zjawisk złożonych) w decydującym stopniu przesądza jakość informacji statystycznej. Działania instytucji gromadzących różnego typu dane (w Polsce przede wszystkim GUS-u) sprowadzają się do zapewnienia użytkownikom informacji o sytuacji ekonomicznej, demograficznej, społecznej oraz na temat środowiska naturalnego, spełniających kryteria jakości stosowane w statystyce publicznej. Jednym z ważniejszych kryteriów jest dokładność danych statystycznych, rozumiana jako bliskość pomiędzy ostatecznie uzyskaną wartością oszacowania parametru a rzeczywistą wartością parametru populacji. Różnica pomiędzy tymi dwiema wartościami jest wartością błędu. W literaturze zwraca się uwagę na dwa główne rodzaje błędów występujących w badaniach statystycznych. Są to błędy losowe (występujące tylko w badaniach reprezentacyjnych) i błędy nielosowe (pojawiające się w każdym badaniu statystycznym). Błędy danych są przede wszystkim związane ze sposobem pozyskiwania informacji statystycznych przez instytucje do tego powołane. Nie ma więc możliwości całkowitego pozbycia się błędów danych statystycznych, a stosowane przez instytucje je gromadzące działania korygujące mogą je jedynie zmniejszyć.

Wielu użytkowników oficjalnie opublikowanych danych statystycznych traktuje je jako dokładne, czyli takie, z którymi nie jest powiązany żaden błąd bądź niepewność co do faktycznej ich wartości. Uwzględnienie w badaniach

empirycznych danych obarczonych błędami może mieć jednak wpływ na końcowe wnioski z prowadzonych analiz.

W świetle przedstawionych przesłanek **celem niniejszej pracy** jest ocena wpływu dokładności danych statystycznych na wyniki wielowymiarowych analiz porównawczych zjawisk złożonych na przykładzie zrównoważonego rozwoju województw Polski. Okresem badawczym są lata 2010-2019.

Cele szczegółowe pracy obejmują:

- przedstawienie wybranych zagadnień z zakresu wielowymiarowej analizy porównawczej,
- zwrócenie uwagi na problem jakości danych statystycznych,
- omówienie najważniejszych zagadnień dotyczących zrównoważonego rozwoju,
- przeprowadzenie analizy porównawczej poziomu zrównoważonego rozwoju województw w Polsce w latach 2010-2019 za pomocą wybranych metod WAP,
- badanie wpływu dokładności danych statystycznych na wyniki porządkowania liniowego i klasyfikacji województw w Polsce pod względem poziomu zrównoważonego rozwoju.

W pracy podjęto próbę odpowiedzi na następujące pytania badawcze:

- Czy koncepcja zrównoważonego rozwoju i pomiar jej realizacji są jednoznacznie określone w literaturze przedmiotu?
- Czy zastosowane w ocenie zrównoważonego rozwoju metody porządkowania liniowego (zarówno bezwzorcowe, jak i wzorcowe) dają podobne wyniki?
- Która z zastosowanych procedur jest optymalna?
- Czy w świetle otrzymanych wyników nastąpiły istotne zmiany w poziomie zrównoważonego rozwoju województw Polski w latach 2010-2019?
- Czy uwzględnienie dokładności danych statystycznych ma wpływ na wyniki porządkowania liniowego województw w Polsce pod względem poziomu zrównoważonego rozwoju?
- Czy uwzględnienie dokładności danych statystycznych ma wpływ na wyniki klasyfikacji województw na grupy podobne pod względem poziomu zrównoważonego rozwoju.

Praca składa się z wstępu, pięciu rozdziałów oraz zakończenia.

Rozdział pierwszy zawiera ogólną charakterystykę wielowymiarowej analizy porównawczej, w tym jej istotę oraz etapy postępowania. Zwrócono uwagę na kryteria doboru zmiennych diagnostycznych, charakter i normalizację zmiennych. Przedstawiono także ogólne założenia wybranych metod porządkowania liniowego obiektów, zarówno bezwzorcowych, jak i wzorcowych. W końcowej części rozdziału zaprezentowano miary zgodności rankingów.

Problematyka jakości danych statystycznych została przedstawiona w rozdziale drugim. Omówiono tu kryteria jakości danych w statystyce publicznej oraz źródła błędów danych statystycznych. Zaprezentowano teoretyczne podstawy analizy błędów i niepewności danych pomiarowych. Przedstawiono także autorskie podejście do wyznaczania prawdopodobieństwa wystąpienia sytuacji problematycznych w tym zakresie. W ostatniej części rozdziału omówiono istotę metody Monte Carlo, jako podejścia symulacyjnego zastosowanego do wyznaczania niepewności zmiennych i mierników syntetycznych według autorskiego algorytmu autorki.

W rozdziale trzecim zawarto podstawowe zagadnienia teoretyczne dotyczące zrównoważonego rozwoju. Wyjaśniono jego istotę i genezę. Zwrócono uwagę na problemy pomiaru zrównoważonego rozwoju a w końcowej części rozdziału zamieszczono przegląd badań nad zrównoważonym rozwojem województw Polski w latach 2010-2019.

Wielowymiarową analizę porównawczą poziomu zrównoważonego rozwoju województw w Polsce w latach 2010-2019 zawiera rozdział czwarty. Wyszczególniono w nim zmienne określające poziom zrównoważonego rozwoju województw oraz dokonano ich porządkowania za pomocą wybranych metod WAP, zarówno bezwzorcowych, jak i wzorcowych. Metody bezwzorcowe oparte były na średniej arytmetycznej znormalizowanych wartości zmiennych, z wykorzystaniem ośmiu sposobów normalizacji. Natomiast jako wzorcowe przyjęto – klasyczną metodę wzorca rozwoju Z. Hellwiga (MWR) oraz Uogólnioną Miarę Odległości M. Walesiaka (GDM) z ośmioma wariantami normalizacji. W końcowej części rozdziału zamieszczono wyniki analizy porównawczej województw pod względem poziomu zrównoważonego rozwoju.

Rozdział piąty zawiera wyniki badań empirycznych dotyczących wpływu dokładności danych statystycznych na porządkowanie liniowe oraz klasyfikację województw w Polsce pod względem poziomu zrównoważonego rozwoju. W pierwszej części rozdziału omówiono wpływ sposobu pozyskiwania danych

statystycznych na ich dokładność. W następnych częściach rozdziału dokonano oceny wpływu błędów pomiarowych i zaokrąglania liczb na niepewność zmiennych i mierników syntetycznych. Analiza objęła ocenę prawdopodobieństwa wystąpienia sytuacji problematycznych spowodowanych błędami w zależności od zastosowanej metody WAP.

W zakończeniu zamieszczono syntetyczne wnioski zarówno z teoretycznej, jak i empirycznej części pracy.

Praca ma charakter monograficzny i jest kompleksową oceną poziomu zrównoważonego rozwoju województw Polski w latach 2010-2019 z wykorzystaniem podejścia dynamicznego. Wątkiem pracy podnosi uwzględnienie wpływu dokładności danych na wyniki porządkowania liniowego oraz klasyfikacji obiektów pod względem badanego zjawiska złożonego.

Zaprezentowane w pracy wyniki badań empirycznych mają także duży walor praktyczny. Mogą być przydatne do podejmowania określonych decyzji na szczeblu krajowym i regionalnym, związanych np. z prowadzeniem i monitorowaniem polityki regionalnej oraz dostosowywaniem się polskich regionów do postanowień Agendy na rzecz Zrównoważonego Rozwoju 2030 czy innych programów rozwojowych.

Dane statystyczne będące podstawą badań empirycznych zaczerpnięto z bazy danych Banku Danych Lokalnych GUS oraz Roczników Statystycznych GUS. Przygotowanie danych oraz część analiz zrealizowano w programach MS Excel oraz STATISTICA v. 12. Obliczenia prezentowane w monografii wykonane zostały także z wykorzystaniem własnej aplikacji napisanej na platformie programistycznej R¹. Natomiast rysunki przygotowano w MS Excel, a końcowej ich korekty dokonywano wykorzystując również program Corel Draw.

¹ Program R jest niekomercyjnym pakietem do analizy danych, stosowanym m.in. w statystyce i ekonometrii [Biecek, 2014; Walesiak, Dudek, 2014].