

Polemika z artykułem Mirosława Błażeja i Emilii Gosińskiej pt. *Dylematy związane z estymacją dominanty wynagrodzeń*

Polemic on the article entitled *Dilemmas relating to mode estimation of wages and salaries* by Mirosław Błażej and Emilia Gosińska

1. Wprowadzenie

Artykuł Mirosława Błażeja i Emilii Gosińskiej pt. *Dylematy związane z estymacją dominanty wynagrodzeń*, opublikowany w nr. 12/2022 „Wiadomości Statystycznych. The Polish Statistician”, wymaga sprostowania i kilku wyjaśnień. Jest on oparty na błędnych założeniach i tezach i przez to wprowadza zamęt do problematyki wyznaczania wartości modalnej oraz jej interpretacji w rozkładach ciągłych, w tym w szczególności w rozkładzie płac.

2. Definicja i rozumienie dominanty

Autorzy omawianego artykułu wykazują daleko idącą niekonsekwencję w samym definiowaniu i rozumieniu dominanty. Przypomnijmy, że dominantą w rozkładzie cechy statystycznej nazywa się wartość, która w tym rozkładzie występuje najczęściej (z największą częstotliwością). Innymi słowy, jest to taka wartość cechy, która powtarza się u największej liczby jednostek badanej zbiorowości. Definicja ta jest jednoznaczna i nie można jej zmieniać w zależności od rodzaju rozpatrywanej cechy statystycznej lub rodzaju rozkładu. Nie jest więc prawdą stwierdzenie, że definicja dominanty zależy od „typu rozpatrywanych zmiennych” (Błażej i Gosińska, 2022, s. 63). Warto było się natomiast w tym miejscu pokusić o refleksję, jaką interpretację oraz jaką wartość poznawczą ma dominanta w rozkładach ciągłych. O ile w rozkładach skokowych sprawa jest prosta – wskazuje się wartość, której odpowiada największa częstotliwość występowania, o tyle w rozkładach ciągłych kwestie te są bardziej złożone (zostaną one omówione w części 5 niniejszego opracowania). Ale, powtórzmy, nie definicja budzi wątpliwości, lecz sposób wyznaczania dominanty, jej interpretacja i walory poznawcze.

3. Wynagrodzenia jako zmienna ciągła

Wypada najpierw ustalić, jakim rodzajem rozkładu cechy jest rozkład wynagrodzeń w Polsce, scharakteryzowany obserwacjami wysokości płac ok. 869 tys. pracowników

(w badaniu Głównego Urzędu Statystycznego z 2018 r.¹, do którego odwołują się Błażej i Gosińska; por. GUS, 2020). Płace pracowników określa się w Polsce z dokładnością do 1 gr i każda wartość z przedziału od ustawowej płacy minimalnej do nieskończoności ma prawo pojawić się jako możliwa wartość tej cechy. Jeżeli cecha (lub zmienna) może przyjmować dowolną wartość rzeczywistą z określonego przedziału (skończonego lub nieskończonego), to oznacza to, że mamy do czynienia z cechą ciągłą. Rozkład wynagrodzeń należy więc uznać za rozkład ciągły. Dopuszczalne byłoby ewentualnie nazwanie go quasi-ciągłym dla podkreślenia faktu, że nie pojawiają się w nim wartości ułamkowe mniejsze od 1 gr (czyli na trzecim i dalszych miejscach po przecinku w wysokości płac, które zwyczajowo wyrażone są w złotych).

Po pierwsze nie można zakładać, że rozkład ciągły lub quasi-ciągły jest rozkładem skokowym (dyskretnym), tak jak nie można przyjąć, że liczba niewymierna jest liczbą wymierną „przy założeniu, że rozkład wynagrodzeń jest dyskretny [...]” (Błażej i Gosińska, 2022, s. 64). Sformułowana przez autorów teza zamieszczona została w podsumowaniu artykułu, co dodatkowo może być dla czytelników mylące. Sugeruje to bowiem, że wykonane i omówione analizy sposobów szacowania dominanty mogą służyć do oceny, czy rozważany rozkład jest ciągły czy mieszany. W rzeczywistości zaś logiczna kolejność działań powinna być odwrotna. Na wstępie badacz określa, z jakim rozkładem ma do czynienia – nie tylko po to, aby dobrać adekwatne metody przybliżeń punktowej oceny wartości dominującej, lecz także aby poprawnie zastosować wiele innych miar opisu statystycznego. Po drugie nie można tworzyć niczym nieuzasadnionego konstruktów mieszanego rozkładu płac, w części skokowego, a w części ciągłego, tylko dlatego że pewna (względnie duża) część jednostek przyjmuje stałą wartość (równą płacy minimalnej): „Analiza otrzymanych wyników daje podstawy do rozważenia tezy, że rozkład wynagrodzeń jest mieszany: ma charakter rozkładu dyskretnego dla wynagrodzeń w wysokości płacy minimalnej i rozkładu ciągłego dla wynagrodzeń powyżej płacy minimalnej” (Błażej i Gosińska, 2022, s. 78).

4. Błędna percepcja dominanty

Za najpoważniejsze merytoryczne uchybienie w opracowaniu Błażeja i Gosińskiej uważam powielenie błędnego rozumienia dominanty. Dodatkowo na swojej błędnej percepcji wartości dominującej w rozkładzie płac autorzy oparli uzasadnienie celu podjętych przez siebie i opisanych w artykule badań. Owo błędne rozumienie dominanty lub błędna jej percepcja zasadza się na oczekiwaniu, że wartość dominująca

¹ Jest to badanie reprezentacyjne, którym objęte są przedsiębiorstwa o liczbie zatrudnionych powyżej dziewięciu osób.

będzie w rozkładzie reprezentowana przez dużą lub bardzo dużą frakcję jednostek. Tak jednak być nie musi. Z definicji wartości dominującej wynika, że jest to wartość najczęściej występująca w rozkładzie, przy czym nie ma znaczenia, jak często – ważne, aby częściej niż każda inna wartość. Autorzy przyznają dwukrotnie w swoim tekście, że w rozpatrywanym rozkładzie płac najczęściej występującą wielkością wynagrodzenia jest płaca minimalna wynosząca 2100,00 zł, którą w 2018 r. otrzymywało 7,44% jednostek badanej zbiorowości. Nie ma merytorycznych powodów, aby nie uznać tej wysokości płacy za dominującą, czyli określającą dominantę w tym rozkładzie. Natomiast następujący po tym wywód: „Wartość ta reprezentuje jednak małą frakcję przypadków w relacji do liczebności badanej zbiorowości. Z uwagi na to celem artykułu jest eksploracja wybranych metod [...]” (Błażej i Gosińska, 2022, s. 64) wskazuje na niezrozumienie definicji dominanty. A dalsza deklaracja autorów o poszukiwaniu oszacowań „dominanty wynagrodzeń reprezentującej większą frakcję przypadków, która ma uzasadnienie metodologiczne i społeczno-gospodarcze oraz odzwierciedla sytuację na rynku pracy” jest nie do przyjęcia z co najmniej dwóch powodów. Zasadniczy powód to sugerowanie czytelnikowi, że w sytuacji gdy jednoznacznie wskazano dominantę, czyli wartość liczbową najczęściej występującą w rozkładzie, celowe może być poszukiwanie jakiejś lepszej (?) dominanty. Poza tym nie wiadomo, co się kryje za stwierdzeniem, że dominantą – albo większą od 7,44% frakcją jednostek, z którą, zdaniem Błażeja i Gosińskiej, powinna być powiązana dominantą – ma mieć uzasadnienie metodologiczne. O jakie uzasadnienie chodzi? I dalej, jakie uzasadnienie społeczno-gospodarcze autorzy mają na myśli przy wyznaczaniu dominanty, skoro definicja dominanty jest jednoznaczna? Tego rodzaju sugestie narażają ich na zarzut manipulacji miarami statystycznymi. Jeżeli o wyznaczonej poprawnie wartości dominującej pisze się, że „nie jest w pełni satysfakcjonującym, reprezentatywnym wynikiem” (Błażej i Gosińska, 2022, s. 75), to czytelnik może zapytać, czy statystyk ma prawo sięgać po argument „niesatysfakcjonującego” (kogo?) wyniku, aby go unieważnić i poszukiwać innego.

Rozważając rozkład płac z 2018 r., w którego lewym krańcu ulokowało się najwięcej jednostek, ponieważ otrzymywały one ustalone urzędowo wynagrodzenie, autorzy powinni byli postąpić inaczej, nie prowokując czytelników do podejrzeń o manipulację. Pierwsza z możliwości to rozpatrzenie rozkładu wynagrodzeń z wyłączeniem płacy minimalnej, czyli oznajmienie czytelnikom, że autorzy świadomie i celowo pomijają (ucinają) część rozkładu odpowiadającą płacy minimalnej i poszukują wartości modalnej w tak przekształconym rozkładzie. Drugie możliwe podejście to poszukiwanie innej, różnej od dominanty, miary statystycznej, która spełniałaby oczekiwania autorów. Trzeba jednak przyznać, że oczekiwania te wyrażone są niejednoznacznie, o czym świadczą zacytowane wyżej fragmenty, jak również stwierdzenie

o poszukiwaniu oszacowania dominanty „która opisuje rozkład wynagrodzeń w oczekiwany sposób, tzn. tak, aby reprezentowała odpowiednio wysoką frakcję przypadków” (Błażej i Gosińska, 2022, s. 78). Nieprecyzyjne są w tym ostatnim zdaniu oba wyrażenia: „oczekiwany sposób” oraz „odpowiednio wysoka frakcja”.

5. Dominanta w rozkładach ciągłych

Rzadko zwraca się uwagę na to, że dominanta – jako liczbowa charakterystyka rozkładu cechy statystycznej – nie ma interpretacji ekonomicznej w rozkładach ciągłych². W tych rozkładach, w przeciwieństwie do skokowych (dyskretnych), można jedynie interpretować dominujący przedział, ale nie pojedynczą wartość liczbową. Przypisanie jakiejś liczbie interpretacji, że pojawia się ona w rozkładzie ciągłym najczęściej, stałoby w sprzeczności z właściwością modelu probabilistycznego opisującego rozkład ciągły, a najczęściej także z obserwacjami statystycznymi.

Modelem probabilistycznym dla rozkładu ciągłej cechy statystycznej jest rozkład ciągły zmiennej losowej, opisywany funkcją gęstości lub dystrybuantą. Dla każdego ciągłego rozkładu zmiennej losowej prawdą jest, że prawdopodobieństwo przyjęcia przez tę zmienną jakiegokolwiek pojedynczej wartości liczbowej jest równe 0. Gdyby zaś zamiast pojedynczej liczby utworzyć wokół niej przedział liczbowy o dowolnie małej rozpiętości, to oczywiście prawdopodobieństwo przyjęcia przez zmienną losową wartości z tego przedziału ma prawo być większe od 0. Innymi słowy, przedziałowi, a nie pojedynczej wartości z tego przedziału, można przypisać niezerowe prawdopodobieństwo realizacji. Nie można więc żadnej wartości liczbowej z rozkładu ciągłego interpretować jako najczęściej występującej, ponieważ prawdopodobieństwo jej wystąpienia jest równe 0.

Interpretowanie punktowej wartości dominanty w rozkładzie ciągłym będzie też często przeczyło rzeczywistym obserwacjom. Dana wartość dominująca nie tylko nie będzie występowała najczęściej, ale może w ogóle nie wystąpić w rozkładzie. Weźmy na przykład jedną z wartości dominaty oszacowanych przez autorów omawianego artykułu, którzy raz podają wielkość 2379,67 zł, a w innym wariantcie – 3013,79 zł. Zinterpretowanie którejś z tych wartości jako wynagrodzenia najczęściej występującego w badanej zbiorowości w październiku 2018 r. jest niepoprawne. Wynagrodzenie dokładnie tej wysokości mogło nie pojawić się ani razu albo pojawić się w odosobnionych, rzadkich przypadkach.

² Ścisłej rzecz ujmując, nie da się o niej powiedzieć, że występuje w rozkładzie najczęściej. Celem poszukiwania punktowej oceny dominanty w rozkładzie ciągłym może być natomiast dążenie do oszacowania takiej wartości D , wokół której liczebności w jej obustronnym sąsiedztwie będą przejawiały tendencję do zmniejszania się wraz z oddalaniem się wartości zmiennej od D .

Można podać kontrargument, że opatrzenie punktowej wartości dominanty w rozkładzie ciągłym słowem „około” pozwoliłoby zasugerować odniesienie się do pewnego przedziału dominanty, a nie jej konkretnej wartości liczbowej. Tyle że tego typu uzupełnienie interpretacyjne domaga się doprecyzowania, o jakim przedziale dominanty mowa, w szczególności – jaka ma być jego rozpiętość. Wydaje się, że w tych okolicznościach znacznie lepiej byłoby pozostać przy interpretacji wyznaczonego wcześniej w rozkładzie dominanty przedziału niż szacować jej wartość punktową, pozbawioną interpretacji ekonomicznej. Przy założeniu jednakowej rozpiętości przedziałów wartości cechy wskazanie jako dominującego (przedziału dominanty) tego, któremu odpowiada największa liczebność, nie rodzi żadnych kontrowersji interpretacyjnych podobnych do wyżej opisanych³. Z tym jednak zastrzeżeniem, że niedopuszczalne jest wtedy sugerowanie – tak jak to zrobili Błażej i Gosińska – że zbyt mała frakcja jednostek reprezentujących taki przedział czyni go „nie w pełni satysfakcjonującym”.

W rozkładach o niewielkim stopniu koncentracji należy się spodziewać, że przedział dominanty będzie reprezentowany przez mały procent jednostek zbiorowości. W Polsce takimi rozkładami są zarówno rozkład wynagrodzeń, jak i rozkład emerytur. Jeśli chodzi o ten drugi przypadek, to z raportu *Struktura wysokości świadczeń wypłacanych przez ZUS po waloryzacji w marcu 2022 roku* (ZUS, 2022) wynika, że największy odsetek emerytur (16,6%) mieści się w przedziale od 2000,01 zł do 2400,00 zł. W rozkładach wynagrodzeń i emerytur przedziały dominanty reprezentowane są przez kilku- lub kilkunastoprocentowe frakcje analizowanych zbiorowości. W niczym nie odbiera im to prawa do miana dominujących przedziałów wartości cechy. Przy czym warta namysłu jest kwestia walorów poznawczych takiego przedziału i stopnia zgodności jego interpretacji z powszechną percepcją terminu *dominujący*. Wiele osób, w tym także przedstawiciele mediów, utożsamia terminy *najczęściej występująca wartość* lub *wartość dominująca w zbiorowości* z dużymi frakcjami jednostek reprezentujących te wartości. Niekiedy wręcz zakładają, że mowa jest o większości jednostek w populacji, czyli o frakcji przewyższającej 50%. W konfrontacji z małymi frakcjami, np. kilkunastoprocentowymi, tracą spójność w postrzeganiu rozkładu danej cechy. Z tego powodu należy uznać za ważne, aby informację o granicach przedziału dominującego uzupełniać każdorazowo o wskazanie, jaki odsetek zbiorowości on zawiera. Wówczas odbiorca sam będzie mógł ocenić poznawczy walor przedziału dominanty w rozkładach ciągłych. Dotyczy to także informacji GUS o charakterystykach rozkładu wynagrodzeń.

³ Nie oznacza to wszakże, że taki przedział jest określony w sposób jednoznaczny, niezależny od grupowania obserwacji, w tym zwłaszcza rozpiętości przedziałów. W tym sensie może to powodować niejednoznaczności o charakterze raczej poznawczym niż interpretacyjnym.

Bibliografia

- Błażej, M., Gosińska, E. (2022). Dylematy związane z estymacją dominanty wynagrodzeń. *Wiadomości Statystyczne. The Polish Statistician*, 67(12), 62–79. <https://doi.org/10.5604/01.3001.0016.1615>.
- Główny Urząd Statystyczny. (2020). *Struktura wynagrodzeń według zawodów w październiku 2018 r.* <https://stat.gov.pl/obszary-tematyczne/rynek-pracy/pracujacy-zatrudnieni-wynagrodzenia-koszty-pracy/struktura-wynagrodzen-wedlug-zawodow-w-pazdzierniku-2018-roku,4,9.html>.
- Zakład Ubezpieczeń Społecznych. (2022). *Struktura wysokości świadczeń wypłacanych przez ZUS po waloryzacji w marcu 2022 roku.* <https://www.zus.pl/documents/10182/39637/Struktura+wysoko%C5%9Bci+%C5%9Bwiadcze%C5%84+wyp%C5%82acanych+przez+ZUS+po+waloryzacji+w+marcu+2022+r..pdf/75b859ca-4130-0372-9d23-27c1882aced6?t=1653904662501>.

Mirosław Szreder

Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk,
Faculty of Management, Poland
ORCID: <https://orcid.org/0000-0002-7597-0816>