

- Bendyk E. [2020], *W Polsce, czyli wszędzie. Rzecz o upadku i przyszłości świata*, Wydawnictwo Polityka.
- ciekawostkihistoryczne.pl/2021/02/09.
- Duda J., Orłowski R. [1999], *Gospodarka polska w dziejowym rozwoju Europy (do 1939 roku)*, Wydawnictwo Uniwersytetu Marii Curie Skłodowskiej, Lublin.
- ekartkazwarszawy.pl.
- Graphic detail Economic history [2023], *The Economist*, January 21st.
- Jakubowicz A. [2020], Spadek i wzrost gospodarczy w okresie transformacji polskiej gospodarki, *Biuletyn PTE*, nr 1.
- Jakubowicz A. [2021], Uwagi na temat wzrostu PKB w okresie transformacji polskiej gospodarki. Czy Polska dogoni Niemcy? *Biuletyn PTE*, nr 4.
- Janicki K. [2021], *Pańszczyzna. Prawdziwa historia polskiego niewolnictwa*, Wydawnictwo Poznańskie, Poznań.
- Karpiński A. [2018], *Prawda i kłamstwa o przemyśle. Polska w obliczu III rewolucji przemysłowej*, Fundacja Oratio Recta.
- Kieżun W. [2012], *Patologia transformacji*, Poltext, Warszawa.
- Kleer J. [2021], *Ekonomiczne i społeczne skutki przesilenia cywilizacyjnych*, Instytut Nauk Ekonomicznych Polskiej Akademii Nauk.
- Kurkiewicz R. [2023], Miliarderów ruch społeczny? *Przeгляд*, 30.01–5.02.
- Maddison A. [2001], *The World Economy: A Millennial Perspective*, OECD Publishing, Paris.
- Mączyńska E. [2014], *Dylematy pomiaru gospodarki globalnej – produkt krajowy brutto*, w: *Księga Jubileuszowa Profesora Włodzimierza Siwińskiego*, WNE, Uniwersytet Warszawski.
- Mączyńska E. [2015], *Inkluzywna gospodarka*, w: *Reforma kulturowa 2020–2030–2040*, Krajowa Izba Gospodarcza.
- Narożniak M. [2021], *Niewolnicy modernizacji. Między pańszczyzną a kapitalizmem*, Wydawnictwo RM.
- Piątkowski M. [2019], *Europejski lider wzrostu. Polska droga od ekonomicznych peryferii do gospodarki sukcesu*, Poltext, Warszawa.
- pl.wikipedia.org/wiki/Holenderskie_imperium_kolonialne.
- pl.wikipedia.org/wiki/Lista_kolonii_brytyjskich.
- pl.wikipedia.org/wiki/Politechnika_Warszawska.
- Pleijt de A.M., Zanden van J.L. [2016], Accounting for the „Little Divergence”: What drove economic growth in pre – industrial Europe, 1300–1800? *European Review of Economic History*, vol. 20, no. 4.
- Slicher van Bath B.H. [1963], *Yield ratios, 1810–1820*, A.A.G. Bijdragen 10, Wageningen.
- Ślżak R. [2021], *Samozagłada polskiej gospodarki 1989–2016*, Wydawnictwo Muzeum Historii Polskiego Ruchu Ludowego.
- turystyka.wp.pl/sokrates-starynkiewicz.
- Unger R. [1992], The tonnage of Europe’s Merchant Fleets 1300–1800, *American Neptune*, vol. 52.
- Wójtowicz G., Wójtowicz A. [2019], *Dlaczego nie jesteśmy bogaci? Dystans gospodarki polskiej do zachodnioeuropejskiej*, wyd. II, CeDeWu, Warszawa.
- Wyczański A. [1973], *Polska w Europie XVI stulecia*, Wiedza Powszechna.
- Zienkowska K. [2003], *Korzenie polskiego zacofania*, w: Zienkowski L. (red.), *Wiedza a wzrost gospodarczy*, Wydawnictwo Naukowe Scholar, Warszawa.

Mirosław Szreder*

Wartość poznawcza mediany i dominanty w analizie płac

Motto:

Variation itself is nature’s only irreducible essence. Variation is the hard reality, not a set of imperfect measures for a central tendency.

[Gould, 1985]

Dla ekonomistów jest na ogół jasne, że każda próba zastąpienia informacji o cesze ciągłej, czyli na przykład o wysokości wynagrodzeń w gospodarce lub wysokości wypłacanych emerytur, jedną miarą liczbową, musi

się okazać niesatysfakcjonująca. I nie ma znaczenia czy miarą tą będzie średnia arytmetyczna, mediana, dominanta, czy jakkolwiek inny miernik syntetyczny. Nie są one w stanie wyrazić tego, co w charakterystyce tych cech jest tak samo ważne lub ważniejsze od tendencji centralnej. Nie informują mianowicie w żaden sposób o zróżnicowaniu cechy (ang. *variation*) w opisywanej zbiorowości. Zróżnicowanie zaś, czyli rozproszenie (dyspersja) jest – jak trafnie to ujął w przytoczonym wyżej motcie słynny paleontolog prof. Stephen Jay Gould (1941–2002) – nieredukowalną istotą natury. Jest twardą rzeczywistością, a nie zbiorem niedoskonałych miar tendencji centralnej – dodaje S.J. Gould.

* Prof. dr hab. Mirosław Szreder – dziekan Wydziału Zarządzania, Uniwersytet Gdański.

Mimo tego statystycy wciąż posługują się miarami tendencji centralnej i słusznie nie zamierzają z tego zrezygnować. Ważne jest jednak, aby uświadomić sobie, jaką wiedzę o strukturze danej cechy dostarczają poszczególne miary, a więc nie tylko jakie są ich ograniczenia oraz poprawna interpretacja, ale też jaka powinna im towarzyszyć narracja¹.

O czym nie informuje mediana?

Motto do tego opracowania zaczerpnąłem z eseju S.J. Goulda zatytułowanego *The Median Isn't the Message* (*Mediana nie jest wiadomością*). W eseju tym Autor pisze o medianie w rozkładzie ciągłym – statystycznym rozkładzie czasu życia, jaki pozostał mu po wykryciu w jego organizmie złośliwego nowotworu w 1982 roku. Z literatury medycznej, do której sięgnął rychło po otrzymaniu diagnozy, dowiedział się, że jest to nieuleczalny nowotwór otrzewnej, a mediana czasu dalszego trwania życia chorych wynosi 8 miesięcy. Połowa chorych nie dożywa ośmiu miesięcy, licząc od dnia zdiagnozowania choroby, a połowie udaje się przeżyć okres dłuższy niż 8 miesięcy. Mediana jednak nie daje żadnych wskazań, na jak długie życie mógł jeszcze liczyć Autor. Nie różnicuje ona pacjentów, których dotknęła ta choroba ze względu na wiele istotnych dla rozwoju choroby cech: stadium zaawansowania nowotworu, ogólny stan zdrowotny pacjenta, jego wiek, stan psychiczny, i wiele innych. Wartość mediany wynosząca 8 miesięcy dalszego trwania życia jest prawdopodobnie przez wiele osób dotkniętych tą chorobą – jak słusznie zauważa Autor – interpretowana jako przybliżony czas, jaki pozostał im na kontemplację życia. Cechą mediany jest to, że nie informuje w ogóle o zróżnicowaniu wartości (długości życia) tych jednostek, które znalazły się po jej prawej stronie, czyli chorych, którym dane było przeżyć więcej niż 8 miesięcy. Prof. S.J. Gould poczuł wyraźny niedosyt wiedzy o zróżnicowaniu (dyspersji) w rozkładzie dalszego trwania życia osób cierpiących na tę chorobę. Nieznane mu zróżnicowanie uznał za prawdziwą, twardą rzeczywistość, samą esencję natury. I trudno się z tym nie zgodzić, biorąc pod uwagę, że przeżył jeszcze prawie 20 lat, a dokładnie 230 miesięcy ponad wielkość mediany wynoszącą 8 miesięcy.

Współcześnie, kiedy odczuwamy nadmiar różnych informacji, wielu z nas chciałoby, aby o danym zjawisku opisywanym liczbami informował jeden prosty i zrozumiały wskaźnik. Pragnienie to wzmacniają dodatkowo media, które w przekazach na temat rozkładów różnych zmiennych ekonomicznych preferują nieskomplikowane syntetyczne mierniki, przede

wszystkim tendencji centralnej. Coraz powszechniej zapominamy lub też chcemy uciec od tego co bardziej realne, mianowicie od zróżnicowania i niepewności. A obie te kategorie przypominają o tym samym, że nie da się za pomocą prostych miar statystycznych przekształcić zróżnicowania cech i niepewności wnioskowania w wiarygodny obraz rzeczywistości. Dotyczy to zwłaszcza powracających co jakiś czas dyskusji na temat najodpowiedniejszych charakterystyk rozkładów wynagrodzeń, a także wypłacanych emerytur w Polsce.

Mediana jako wartość przeciętna w charakterystyce wynagrodzeń

W ostatnich latach w dyskusjach tych pojawiło się sporo uzasadnionych zastrzeżeń do średniej arytmetycznej płac, której wartość zawiązują wysokie i bardzo wysokie wynagrodzenia niewielkiej liczby osób. Mimo że średnia ta była (i jest) przez instytucje statystyki publicznej poprawnie obliczana i interpretowana, to poczucie jej nieadekwatności w opisie rozkładu wynagrodzeń było spowodowane tym, że około 2/3 osób otrzymywało mniejszą płacę niż wynosiła średnia arytmetyczna. Tak jest również obecnie. Przypomnijmy, że w rozkładach o prawostronnej skośności – a takie są rozkłady wynagrodzeń i emerytur w wielu krajach, nie tylko w Polsce – mediana ma niższą wartość niż średnia arytmetyczna. Dla większości osób bardziej wiarygodną miarą przeciętnej płacy jest więc wartość środkowa w rozkładzie (mediana). Połowa zatrudnionych otrzymuje płacę niższą, a połowa wyższą od mediany. Z tego powodu mediana awansowała na wyższe niż do tej pory miejsce w popularności miar przeciętnego wynagrodzenia. Bywa teraz często podawana jako jedyna miara średnich zarobków.

W tym odwróceniu się części ekonomistów od średniej arytmetycznej i przyjęciu mediany za niemal doskonałą miarę średniego poziomu wynagrodzeń tkwi jednak pułapka. Tak jak w dramatycznym pytaniu J.S. Goulda o rozproszenie długości życia wśród osób, które znalazły się po prawej stronie mediany, tak w przypadku wynagrodzeń, mediana nie ujawnia żadnych prawidłowości wśród tych, którzy mają najwyższe zarobki, czyli znajdują się po jej prawej stronie. A dodatkowo jest zupełnie nieczuła na wszelkie wzrosty wynagrodzeń tych właśnie osób. Krótko mówiąc, przywiązując się do mediany jako liczbowej charakterystyki przeciętnego wynagrodzenia, warto zdawać sobie sprawę z tego, że jej wartość nie ulega żadnej zmianie, gdy rosną zarobki 50% tych, którzy zarabiają ponad wielkość mediany. Nawet, gdy zarobki te rosną o dowolnie wysokie (nieograniczone) kwoty! Tej słabości średnia arytmetyczna nie ma. Dlatego za pożądane uzupełnienie mediany, jako miary przeciętnych wynagrodzeń, należy uznać albo średnią

¹ O tym, czy liczby potrzebują narracji pisze m.in. Napiórkowski [2022], a także Szreder [2022].

arytmetyczną, albo iloraz dwóch skrajnych decyli dziewiątego (D9) i pierwszego (D1)² w rozkładzie wynagrodzeń. Iloraz D9/D1 stanowi zarówno uzupełnienie mediany, gdyż reaguje on na zmiany płac wśród najlepiej i najźle zarabiających, a ponadto jest komplementarną w stosunku do współczynnika Giniego miarą nierówności zarobków. Iloraz ten przyciąga uwagę ekonomistów i jest wielkością szeroko komentowaną także przez polityków w wielu krajach. Na przykład w Irlandii zwraca się uwagę na to, że w odniesieniu do godzinowej stawki płacy iloraz D9/D1 jest tam wyższy niż w większości krajów Unii Europejskiej (w 2018 roku wynosił 3,98, podczas gdy np. w Danii tylko 2,37) [Nugent, 2021]. Z kolei z nowszych danych OECD odnoszących się do wynagrodzeń brutto można wyczytać³, że Polska z wartością tego wskaźnika równą 3,45 należała w roku 2020 do kilku krajów o największym zróżnicowaniu wynagrodzeń w Europie. Dziesięć procent najlepiej zarabiających w naszym kraju otrzymywało wynagrodzenie, które było przeciętnie prawie trzy i pół raza wyższe od płac dziesięciu procent zarabiających najmniej. Także w odniesieniu do krajów OECD, dla których wskaźnik ten przeciętnie wynosił 3,33 (w roku 2020), Polska notuje stosunkowo wysokie rozproszenie wynagrodzeń brutto.

Kłopoty z dominantą wynagrodzeń

Najbardziej kontrowersyjną miarą tendencji centralnej w przypadku wynagrodzeń jest wartość modalna, inaczej zwana dominantą. Z jednej strony budzi ona duże zainteresowanie, zwłaszcza w środowisku dziennikarskim, gdyż uwagę odbiorców łatwo przyciągnąć informacją o tym, że najczęściej otrzymywanym w Polsce wynagrodzeniem jest X złotych. Z drugiej jednak strony, posługiwanie się dominantą wyrażoną punktowo jest w rozkładach cech ciągłych (wynagrodzeń, emerytur, czasu trwania życia) nie tylko dyskusyjne, ale wprowadzające w błąd odbiorców.

Dominantą w rozkładzie cechy statystycznej nazywa się wartość, która w rozkładzie tym występuje najczęściej (z największą częstotliwością). Innymi słowy, jest to taka wartość cechy, która powtarza się u największej liczby jednostek badanej zbiorowości. W rozkładach cech jakościowych, a także w przypadku cech ilościowych skokowych dominantą, jeżeli istnieje, ma jednoznaczną interpretację i na ogół łatwo się ją wyznacza (wskazuje). W rozkładach ciągłych

natomiast poprawnie udaje się jedynie zinterpretować dominujący przedział, ale nie pojedynczą wartość liczbową. Przypisanie bowiem 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 empirycznie dokonanymi obserwacjami statystycznymi.

Modelem probabilistycznym do opisu ciągłej cechy statystycznej jest rozkład ciągły zmiennej losowej, charakteryzowany funkcją gęstości lub dystrybuantą. Dla każdego ciągłego rozkładu zmiennej losowej prawdą jest, że prawdopodobieństwo przyjęcia przez tę zmienną dowolnej pojedynczej wartości liczbowej jest równe zero. 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 zera. Innymi słowy, przedziałowi, a nie pojedynczej wartości z tego przedziału, można przypisać niezerowe prawdopodobieństwo realizacji. Nie można więc żadnej wartości liczbowej z rozkładu ciągłego przyjąć, a dalej interpretować jako najczęściej występującą, bo prawdopodobieństwo jej wystąpienia jest równe zero.

Interpretowanie punktowej wartości dominanty w przypadku cechy ciągłej będzie też często przeczyło rzeczywistym obserwacjom. Dana wartość dominująca nie tylko nie będzie występowała najczęściej, ale może w ogóle ani razu nie wystąpić w danej zbiorowości. Tak jest nie tylko w przypadku wynagrodzeń, ale także emerytur. Gdy więc ZUS [2022, s. 9 i 22] podaje, że w marcu 2022 roku najwyższy odsetek emerytów otrzymywał emerytury z przedziału 2000–2200 zł i odsetek ten wynosił 8,5%, to oznacza, że jest to poprawnie wskazany i zinterpretowany przedział dominanty. Natomiast, gdy w tym samym komunikacie dalej stwierdza się, że emeryturą pobieraną przez największą liczbę świadczeniobiorców, tj. przez 253,4 tys. osób (4,2%), była emerytura w wysokości 2051,77 zł, to nie jest to ścisła informacja. Po pierwsze jest mało prawdopodobne, aby aż 4,2% osób otrzymywało emeryturę równą dokładnie 2051,77 zł i niemal taki sam odsetek (4,3%) wszelkie inne kwoty z przedziału 2000–2200 zł. Po drugie, nie mamy pewności, czy podając tę kwotę, ZUS nie posługuje się przybliżoną wielkością, a w rzeczywistości liczba osób pobierających emeryturę równą dokładnie 2051,77 zł jest znacznie mniejsza lub osób takich w ogóle nie ma. Gdyby była to przybliżona wielkość emerytury, czyli w domyśle „około 2051,77 zł”, to słowo „około” zastępuje w tym stwierdzeniu pewien przedział liczbowy, tyle że z nieujawnionymi granicami. Rodzi się więc pytanie, po co podawać przybliżoną wielkość punktową dominanty, skoro w rzeczywistości reprezentuje ona przedział liczbowy o nieznanym odbiorcy rozpiętości?

² W literaturze anglojęzycznej, w tym także w komunikatach Eurostatu i OECD częściej decyl dziewiąty jest nazywany 90. percentylem (P90) i analogicznie decyl pierwszy 10. percentylem (P10).

³ OECD, https://stats.oecd.org/Index.aspx?DataSetCode=DEC_I (data pobrania).

Za lepsze rozwiązanie należy uznać poprzestanie na wskazanym na wstępie przedziale 2000–2200 zł, który wśród przedziałów o rozpiętości 200 zł występuje w rozkładzie emerytur najczęściej. W tym ostatnim stwierdzeniu nie ma żadnych niejasności podobnych do tych, jakie rodzą się przy próbie interpretacji wartości punktowej dominanty.

Interpretacja dominanty i jej rozumienie mogą być uważane za kłopotliwe w przypadku cech ciągłych nawet wówczas, gdy zgodzimy się na sformułowaną wyżej propozycję, aby dla tej grupy cech wskazywać nie dominującą wartość punktową, lecz przedział dominanty. W interpretacji pozostają bowiem wciąż obecne wyrażenia: najwyższy odsetek, najwięcej emerytów, największa liczba zatrudnionych itp. U wielu odbiorców wyrażenia te, informujące o najwyższym odsetku lub największej liczebności, będą utożsamiane z dużymi liczbami, często obejmującymi większość (ponad 50%) jednostek całej zbiorowości, do której się odnoszą. Największy lub najczęstszy oznacza przecież duży – byliby gotowi dowodzić. Tymczasem, jak widać to na przykładzie wysokości wypłacanych emerytur, w przedziale dominanty mieści się zaledwie 8,5% wszystkich emerytów. Najwyższy odsetek oznacza tu odsetek równy 8,5%. Podobnie jest w rozkładzie wynagrodzeń. Jeśli wskażemy przedział o niewielkiej rozpiętości, któremu w rozkładzie wynagrodzeń odpowiada najwyższy odsetek zatrudnionych, to odsetek ten wyniesie kilka lub kilkanaście procent. Gdy zaś ten dominujący przedział będzie miał większą rozpiętość, np. około 1500 zł, to odsetek mieszczących się w nim zatrudnionych będzie wyższy i może sięgać 30% lub więcej. Według danych GUS [GUS, 2022] najwyższy był w Polsce w 2020 roku odsetek zatrudnionych, którzy otrzymywali wynagrodzenie brutto w przedziale 50–75% przeciętnego miesięcznego

wynagrodzenia, czyli od 2874,13 zł do 4311,18 zł. Odsetek ten wynosił 29,5%. Ale i ta wielkość (29,5%) może być w odczuciu niektórych osób zbyt mała, by przypisywać jej termin: najwyższy odsetek. Z tych właśnie powodów, wskazując przedział płacy dominującej, za ważne należy uznać nie tylko podanie jego interpretacji statystycznej, ale także informacji dodatkowych: o odsetku jednostek, jaki przedział ten reprezentuje, oraz o rozpiętościach przedziałów w szeregu służącym do wyznaczenia przedziału dominującego. W przypadku każdego miernika statystycznego warto zadbać nie tylko o jego poprawną interpretację, ale także o wzbogacenie go informacjami, które ułatwią właściwe jego zrozumienie i właściwą narrację wśród odbiorców.

Bibliografia

- Gould S.J. [1985], The median isn't the message, *Discover*, no. 6, s. 40–42.
- GUS [2022], *Struktura wynagrodzeń według zawodów w październiku 2020 r.*, Główny Urząd Statystyczny.
- Napiórkowski M. [2022], *Naprawić przyszłość. Dlaczego potrzebujemy lepszych opowieści, żeby uratować świat*, Wydawnictwo Literackie.
- Nugent C. [2021], *Wages in Ireland are more unequally distributed than in any other high-income EU country*, Nevin Economic Research Institute, 24 June, <https://www.nerinstitute.net/blog/wages-ireland-are-more-unequally-distributed-any-other-high-income-eu-country>.
- OECD, https://stats.oecd.org/Index.aspx?DataSetCode=DEC_I.
- Szreder M. [2022], Czy liczby potrzebują narracji? *Biuletyn PTE*, nr 4(99), s. 44–46.
- ZUS [2022], *Struktura wysokości świadczeń wypłacanych przez ZUS po waloryzacji w marcu 2022 roku*, Zakład Ubezpieczeń Społecznych.



Unikatowy numizmat na 500. rocznicę ogłoszenia traktatu „De aestimatione monetarum”.

Źródło: <https://www.mennica.com.pl/o-nas/aktualnosci/szczegoly/unikatowy-numizmat-na-500-rocznice-wyglaszenia-traktatu-de-aestimatione-monetae>.