Wiadomości Statystyczne. The Polish Statistician, 2025, *70*(2), 84–91 https://doi.org/10.59139/ws.2025.02.4
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

Recenzja książki Mirosława Szredera i Arkadiusza Kozłowskiego *Wnioskowanie na podstawie prób losowych i nielosowych*

Review of Mirosław Szreder and Arkadiusz Kozłowski's book *Inference on the basis of random and non-random samples*

Język/Language: polski/Polish
Wydawnictwo/Publisher: Wydawnictwo Uniwersytetu Gdańskiego
Miejsce i rok wydania / Place and year of publication: Gdańsk 2024
Liczba stron / Number of pages: 323

Formułowanie rzetelnych i użytecznych wniosków na podstawie wyników badań statystycznych stanowi czynnik decydujący o wartości takich analiz. Wysoka jakość zebranych informacji, umiejętność właściwej ich interpretacji i znajomość skutków nieodpowiedniego stosowania niezbędnych narzędzi statystycznych ułatwiają ten proces. Monografia autorstwa znakomitych gdańskich statystyków: prof. dr. hab. Mirosława Szredera i dr. Arkadiusza Kozłowskiego stanowi cenne kompendium wiedzy o różnych aspektach wnioskowania, zwłaszcza korzyściach i problemach związanych z zastosowaniem prób losowych i nielosowych w badaniach statystycznych.

Opracowanie składa się z wprowadzenia, siedmiu rozdziałów, podsumowania i dwóch załączników: opinii Amerykańskiego Towarzystwa Statystycznego dotyczącej istotności i stosowania *p-value* oraz stanowiska Amerykańskiego Stowarzyszenia Badaczy Opinii Publicznej w sprawie zagrożeń płynących z badań opartych na internetowych próbach respondentów chętnych do wzięcia udziału w ankiecie.

We wprowadzeniu autorzy przedstawili cele monografii i przedyskutowali samą ideę wnioskowania, w tym wnioskowania statystycznego. Wyszli od ogólnej definicji wnioskowania („rozumowanie i formułowanie sądów na podstawie określonych przesłanek stanowiących niepełny opis sytuacji oraz okoliczności, do których czynności te się odnoszą") i wskazali specyfikę wnioskowania w statystyce. Właśnie,

w statystyce, a nie *statystycznego*, ponieważ – jak słusznie zauważyli – w klasycznym ujęciu wnioskowanie statystyczne odnosi się do formułowania konkluzji dotyczących kształtowania się badanego zjawiska dla populacji na podstawie próby losowej prostej. Tymczasem w praktyce coraz powszechniej wykorzystuje się w tym celu próby złożone i nielosowe. Co więcej, wydaje się, że ograniczanie znaczenia pojęcia *wnioskowanie statystyczne* tylko do oceny populacji na podstawie próby to nadmierne zawężenie. Statystyka to nie tylko estymacja czy imputacja, lecz także wielowymiarowa analiza danych pozwalająca na wyodrębnianie grup obiektów podobnych, porządkowanie tych obiektów i efektywną redukcję wymiaru danych (analiza skupień, konstrukcja mierników syntetycznych, analiza głównych składowych itp.). Dobrze, że w obliczu współczesnych realiów autorzy podjęli dyskusję nad redefinicją pojęcia *wnioskowanie statystyczne*.

W rozdziale pierwszym *Atrybuty losowości i prawdopodobieństwa* zaprezentowano koncepcyjne fundamenty zdarzeń losowych i prawdopodobieństwa ich zajścia. Na podkreślenie zasługuje wnikliwe omówienie kwestii losowości i reprezentatywności, z uwzględnieniem praktycznego wykorzystania obu mechanizmów oraz ułomności społecznego postrzegania losowości. Jeśli chodzi o ten ostatni aspekt, to autorzy wymienili dwa najczęściej popełniane błędy: wiarę w występowanie zależności między zjawiskami niezależnymi od siebie i przyjmowanie, że pewne zdarzenia są niezależne, mimo że mają tę samą przyczynę. Szkoda, że autorzy podali konkretny przykład jedynie pierwszej sytuacji, ponieważ przydałaby się także egzemplifikacja drugiej.

Za bardzo wartościowy element rozdziału należy uznać szczegółowe ukazanie definicyjnych i percepcyjnych podstaw rachunku prawdopodobieństwa (tu drobna uwaga: w zakresie teorii definicja σ-ciała podzbiorów zbioru zdarzeń elementarnych pomija warunek, że to ciało jest zamknięte również ze względu na iloczyny, czyli części wspólne podzbiorów, a nie tylko na ich sumy i dopełnienia). Istotny walor tej części książki stanowi również krytyczne omówienie różnorodnych definicji prawdopodobieństwa – od podejścia Kołmogorowa, przez klasyczną definicję Laplace'a i częstościową interpretację von Misesa, aż po ideę prawdopodobieństwa personalistycznego (subiektywnego). Warto w tym miejscu zauważyć, że nawet w obrębie eksperckich ocen prawdopodobieństwa subiektywnego można się także opierać na pewnych kryteriach obiektywnych, np. w testowaniu hipotez statystycznych, w którym punktem wyjścia jest właśnie subiektywne przekonanie badacza o wartości danego parametru czy występowaniu związku nieparametrycznego. Wynika to z tego, że wartości krytyczne lub *p-value* są generowane z teoretycznego rozkładu statystyki testowej, a więc na podstawie jak najbardziej obiektywnych przesłanek.

Ważne i szczególnie wartościowe jest omówienie najważniejszych – i niestety dość powszechnych – błędów w zakresie percepcji prawdopodobieństwa, takich jak niewłaściwe rozumienie kwestii losowości (zwłaszcza w kontekście posiadanych informacji a priori) i koncentrowanie się na bezwzględnej (zamiast na względnej) liczbie zaobserwowanych przypadków wystąpienia danego zdarzenia. Do tego warto byłoby dodać kłopoty z postrzeganiem prawdopodobieństwa warunkowego, czego najbardziej znanym przykładem jest tzw. problem Serbelloni[1]. Najprościej mówiąc, chodzi o sytuację znaną także z teleturniejów[2], w której mamy troje drzwi, ale tylko za jednymi znajduje się nagroda. Gracz wybiera losowo jedne z drzwi, po czym prowadzący otwiera te z dwojga pozostałych, za którymi nic nie ma, i pyta gracza, czy zmienia decyzję, a zatem czy wybiera drugie z zamkniętych drzwi. Prawidłowa interpretacja prawdopodobieństwa warunkowego prowadzi do konkluzji, że zmiana decyzji zwiększa dwukrotnie szansę na wygraną. Omówieniem tego zagadnienia (choć przedstawionego w innej formie – dylematu skazańców) rozpoczynał zazwyczaj swoje wykłady ze wstępu do rachunku prawdopodobieństwa dla studentów zastosowań matematyki na Uniwersytecie Wrocławskim prof. dr hab. Józef Łukaszewicz (1992). Dość radykalnie na temat rozumienia wskazanego problemu wypowiedział się także prof. dr Lech T. Kubik (1991). W szerszym zakresie zagadnienie to rozpatrywali ponadto m.in. Smoluk (1997a, 1997b) oraz Dniestrzański i Wilkowski (2009).

W rozdziale drugim *Teoretyczne podstawy wnioskowania statystycznego* autorzy odnieśli się przede wszystkim do określenia losowego wyboru próby oraz badania reprezentacyjnego. Warto zauważyć, że podana definicja próby losowej prostej odpowiada klasycznej definicji próby w rozumieniu statystyki matematycznej: skończony ciąg niezależnych zmiennych losowych o jednakowym rozkładzie (zob. np. Bartoszewicz, 1996). Jednak w praktyce statystycznej wyróżnia się co najmniej dwa schematy losowania prób prostych: ze zwracaniem i bez zwracania. W tym drugim przypadku losowanie – rzecz jasna – nie jest niezależne. Skłania to do szerszej dyskusji na ten temat. Warto podkreślić przejrzystość i szczegółowość scharakteryzowania tutaj roli losowości doboru próby w osiąganiu odpowiedniej dokładności i efektywności estymacji.

Następnie autorzy zajęli się próbami nieprostymi (złożonymi). Omówili przyczyny stosowania tych prób oraz specyfikę losowania warstwowego i zespołowego. W odniesieniu do schematu zespołowego podali, że poszczególne zespoły są loso-

[1] Nazwa pochodzi od hotelu Villa Serbelloni we Włoszech, gdzie w 1966 r. z powodu kontrowersji, jakie wzbudził ten dylemat, omal nie doszło do zerwania odbywającej się tam konferencji biostatystycznej.

[2] Przede wszystkim z teleturnieju *Let's make a deal,* którego prowadzącym był Monty Hall (Maurice Halprin), kanadyjski aktor i piosenkarz; od jego nazwiska zagadnienie nazywa się czasem problemem Monty Halla. W Polsce odpowiednikiem tego programu był teleturniej *Idź na całość*.

wane z prawdopodobieństwami proporcjonalnymi do ich wielkości. Zgodnie z ich opinią tym samym prawdopodobieństwo dostania się danej jednostki do próby nie zależy od wielkości zespołu, do którego ona należy. Ale warto zauważyć, że choć jednostki nie są tutaj bezpośrednio losowane, to w sytuacji gdy losowane są tylko zespoły, a następnie bada się wszystkie jednostki z danego zespołu, prawdopodobieństwo wyboru jednostki do próby jest równe prawdopodobieństwu wylosowania zespołu, do którego ona należy. Wobec tego zależność od rozmiaru zespołu w jakiejś mierze jednak występuje.

Swoiste podsumowanie tej części opracowania stanowi interesująca dyskusja dotycząca stosowania prób nielosowych, w tym warunków i metod efektywnego doboru oraz obciążeń i błędów, jakie się z tym wiążą, a także zalet, które w kontekście jakości badania może w konkretnych sytuacjach mieć nielosowy wybór respondentów.

W rozdziale trzecim *Estymacja punktowa i przedziałowa* na uznanie zasługuje bardzo klarowne i przystępne (ilustrowane odpowiednimi przykładami) omówienie kluczowych metod estymacyjnych: momentów, największej wiarygodności i najmniejszych kwadratów (MNK) oraz szacowania przedziałów ufności – w tym minimalnej liczebności próby losowej, niezbędnej do uzyskania odpowiedniej precyzji oszacowania danej frakcji. W odniesieniu do MNK zabrakło jednak doprecyzowania, względem jakiego argumentu docelowa funkcja kwadratowa jest minimalizowana. Przydałoby się także porównawcze spojrzenie na trzy wskazane metody estymacji punktowej z podkreśleniem, w jakich sytuacjach ich zastosowanie okazuje się szczególnie użyteczne, a w jakich – niepożądane (np. MNK zawodzi w estymacji parametrów modeli ekonometrycznych z heteroskedastycznością lub autokorelacją błędów albo w przypadku występowania zmiennych instrumentalnych).

Tematem rozdziału czwartego *Estymacja z wykorzystaniem informacji spoza próby* jest wykorzystywanie informacji pomocniczych w celu podnoszenia jakości estymacji docelowych wielkości dla populacji. Autorzy położyli nacisk na uwzględnianie dodatkowych danych w modyfikacji wag wynikających ze schematu losowania – w tym ich kalibracji – przy czym zauważyli, że nawet stosowanie estymatorów ilorazowych, regresyjnych czy złożonych opierających się na takich dodatkowych informacjach sprowadza się w istocie do korekty wag.

Szczegółowo omówiono specyfikę i użyteczność podejść opartych na modelach regresyjnych, zarówno w ujęciu randomizacyjnym, jak i typowo modelowym, ilustrując to poglądową egzemplifikacją. Znajdziemy tu także – również wzbogaconą stosownymi przykładami – interesującą dyskusję na temat walorów i niedogodności bayesowskiego mechanizmu wnioskowania statystycznego. Warto przy okazji zauważyć, że wzór Bayesa można – wedle definicji rozkładu warunkowego – zapisać

wyłącznie jako funkcję prawdopodobieństw łącznego zajścia zdarzeń podstawowych i pomocniczego, a w przypadku ciągłym – jako wypadkową gęstości łącznych rozkładów tych zdarzeń. Może to ułatwić stosowanie tych metod w praktyce, ponieważ zdarzają się sytuacje, w których cechy łącznego rozkładu są bardziej uchwytne niż cechy rozkładu warunkowego. Pewne niedopowiedzenie można zauważyć w przykładzie dyskretnym, w którym autorzy pominęli sposób obliczenia prawdopodobieństwa warunkowego wylosowania osób z wyższym wykształceniem w 10-elementowej próbie, pod warunkiem że odsetek osób o tej cesze w całej populacji jest dany. Brak też jakiejkolwiek informacji o zastosowanym schemacie losowania. Wskutek tego czytelnik może mieć problem z odtworzeniem sposobu postępowania w tym przypadku.

Rozdział piąty *Weryfikacja hipotez – współczesne wyzwania* rozpoczynają rozważania dotyczące dwóch odmiennych koncepcji klasyfikacji hipotez, a mianowicie wnioskowania indukcyjnego (z którego pochodzi konstrukcja *p-value* Ronalda Fishera) i wnioskowania dedukcyjnego (w którym kluczowy jest lemat Neymana-Pearsona). Autorzy podkreślili, że wedle pierwszej idei test statystyczny jest traktowany jako krok na drodze do głębszego poznania problemu naukowego, a zgodnie z drugą – stanowi przede wszystkim narzędzie podejmowania decyzji. Zasygnalizowali nazbyt mechaniczne (widoczne zwłaszcza w dobie intensywnej informatyzacji statystyki) stosowanie *p-value*, prowadzące do zbytecznego albo niepełnego zakresu formułowania konkluzji o badanych zjawiskach, a także odnieśli się do głównych poglądów wyrażanych w toczących się w ostatnich latach dyskusjach na temat znaczenia i stosowania tego wskaźnika. W opiniach tych zaleca się, aby odejść od dychotomizacji decyzji bazujących na *p-value* (zwłaszcza z progiem 0,05), a formułowanie wniosków oprzeć także na innych miarach (np. przedziałach ufności), które lepiej wychwytują błędy nielosowe. Wydaje się, że łatwiej będzie to zrozumieć, gdy przyjmie się stosowaną w literaturze (zob. np. Luszniewicz i Słaby, 1995; Młodak, 2020) interpretację *p-value* jako poziomu istotności ex post, czyli najmniejszego poziomu istotności, na którym następuje odrzucenie hipotezy zerowej przy danej wartości statystyki testowej. Widać wtedy wyraźniej, że *p-value* jest ciągła i wszelka jej kategoryzacja może się okazać nieefektywna.

Do zakresu istotnych miar służących podejmowaniu decyzji warto byłoby włączyć także moc testu, co postuluje np. Szymczak (2018), który również krytycznie ocenia „kult *p-value*", ponieważ zgodnie z lematem Neymana-Pearsona odpowiednie testy są jednostajnie najmocniejsze. Ewentualna zbyt słaba moc będzie się zatem wiązać z występowaniem niekorzystnych czynników pozalosowych. Nabiera to szczególnego znaczenia w przypadku big data, gdy niezmiernie duża liczebność zbiorów obserwacji bardzo często – poprzez malejące do zera wartości *p-value* – daje

podstawy do odrzucenia hipotezy zerowej wbrew rzeczywistej sytuacji. Wynika to z probabilistycznej zgodności estymatorów, a także – a może przede wszystkim – z odporności zakresu wielkości występujących błędów nielosowych na liczebność próby.

W rozdziale szóstym *Próby nieproste jako podstawa wnioskowania* autorzy podali formuły oszacowań średniej arytmetycznej i frakcji oraz stosownych wariancji z próby uzyskanej poprzez losowanie warstwowe ze zwracaniem i bez zwracania, a także wskazali na użyteczność szacowania minimalnej liczebności próby na podstawie przedziałów ufności wobec arbitralnie zakładanej dopuszczalnej precyzji oszacowań. Podobne rozważania przedstawili w kontekście losowania zespołowego. Można się jedynie zastanowić nad stwierdzeniem, że w przypadku estymacji frakcji próba wstępna (niezbędna do oszacowania odpowiednich frakcji w zespołach) musi być znacznie liczniejsza niż w estymacji średniej arytmetycznej. Jest to oczywiście prawda, gdy mamy na myśli szacowanie średniej dla całej populacji. Jeżeli jednak interesują nas także średnie w zespołach lub innego rodzaju grupach (a te są zazwyczaj predefiniowane i mają znaczenie praktyczne), to w obu przypadkach liczebność prób powinna być raczej znaczna.

W rozdziale siódmym *Nielosowe próby jako podstawa wnioskowania o populacji* na szczególne uznanie zasługuje dokładne, oparte na fundamentach formalnych, omówienie specyfiki szacowania obciążenia i wariancji w przypadku stosowania prób nieprobabilistycznych. Zauważyć należy zwłaszcza wskazanie komponentów błędu wnioskowania, a mianowicie: korelacji między faktem udziału w badaniu a wartością rozpatrywanej cechy, relacji udziału jednostek niezbadanych do udziału jednostek zbadanych w populacji i wariancji analizowanej zmiennej w populacji. W przeciwieństwie do badań opierających się na losowaniu probabilistycznym szacunki te (zwłaszcza dotyczące pierwszego ze wskazanych wyżej komponentów) wymagają informacji spoza próby, a co więcej – przy zachowaniu oczekiwań co do efektywności optymalna liczebność próby nielosowej może być nawet kilkanaście razy większa od liczebności próby losowej. Pewną niedogodnością jest jedynie brak szczegółów dotyczących wyprowadzenia niektórych wzorów i przejścia od wielkości parametrycznych (np. numerów jednostek) do odpowiednich zmiennych losowych bez niezbędnych objaśnień dotyczących postaci właściwych momentów.

Następnie autorzy szczegółowo omówili rolę podejść quasi-randomizacyjnego i opartego na modelu w redukcji obciążenia płynącego z odpowiednich prób nielosowych. Do tych drugich (bazujących na koncepcji nadpopulacyjnej) autorzy zaliczyli także imputację masową, w przypadku której zakładają całkowity brak informacji dla badanych zmiennych. Założenie to wydaje się jednak nadmiernie restrykcyjne, ponieważ o imputacji masowej można mówić także w sytuacji uzupełniania braków

danych, gdy braki te występują dla rekordów w liczbie znacznie przewyższającej liczbę rekordów, dla których odpowiednie dane są dostępne (zob. np. Lańduch, 2023).

Finalnie porównano efektywność badań opartych na próbach losowych i nielosowych z wykorzystaniem przykładów empirycznych zaczerpniętych z literatury. Opisane eksperymenty wykazały, że próby losowe dają na ogół lepsze jakościowo rezultaty, nawet w przypadku występowania obciążeń odpowiedzi. Jakość i użyteczność badań z prób nieprobabilistycznych można jednak ulepszyć, wykorzystując odpowiednie zmienne, mechanizm kalibracji i dane pomocnicze. Nasuwają się jedynie wątpliwości, czy konieczność badania w sondażach przedwyborczych dwóch cech – tzn. preferencji wyborczych i chęci uczestnictwa w głosowaniu – stanowi aż tak duży problem. W badaniach ankietowych statystyki publicznej występuje często większa liczba (i nierzadko bardziej złożonych) zmiennych, a jednak nie stoi to na przeszkodzie w uzyskiwaniu – w obrębie aktualnych ram metodologicznych – wysokiej jakości rezultatów. Problem z rozbieżnością pomiędzy prognozowanymi a rzeczywistymi wynikami wyborów[3] może wynikać raczej z doboru próby (np. sondażownie stosują często dużą próbę, tzw. matkę, z której losują próby do badania, jednak próba ta powinna być aktualizowana z wykorzystaniem alternatywnych źródeł danych) i obciążenia odpowiedzi (ukrywania lub zmiany decyzji respondentów dotyczącej tego, na kogo oddać głos).

Monografię wieńczy podsumowanie, w którym autorzy przedstawili najważniejsze wnioski płynące z omówionych zagadnień. Podkreślili, że wobec rozwoju nowych, mniej kosztownych narzędzi uzyskiwania danych (np. internetu) wzrasta popularność prób nielosowych, jednak otrzymanie na ich podstawie wiarygodnych rezultatów wymaga podjęcia wielu dodatkowych działań korygujących.

Dzięki wszechstronnej analizie wszystkich aspektów i konsekwencji stosowania prób losowych i nielosowych omawiana praca jest wartościowym poradnikiem nie tylko dla osób zajmujących się projektowaniem i prowadzeniem badań statystycznych, lecz także tych, które korzystają z ich wyników. Wysoce krytyczna i pogłębiona dyskusja oraz odniesienia do opinii wyrażanych na forum międzynarodowym stanowią o unikatowej wartości tej pozycji.

Bibliografia

Bartoszewicz, J. (1996). *Wykłady ze statystyki matematycznej*. Wydawnictwo Naukowe PWN.

Dniestrzański, P., Wilkowski, A. (2009). O paradoksie Halla i rzucaniu monetą. *Didactics of Mathematics*, (5–6), 43–52.

[3] Był on widoczny np. podczas wyborów prezydenckich w Stanach Zjednoczonych w 2024 r., kiedy sondaże przeprowadzone w niektórych stanach przez uznane ośrodki badania opinii publicznej dały wyniki zupełnie nieodpowiadające faktycznym.

Kubik, L. T. (1991, 20 listopada). Drzwi do wygranej. *Spotkania*, 7.

Lańduch, P. (2023). *Wykorzystanie technik imputacyjnych w szacowaniu informacji wynikowych oraz w analizie struktury danych w statystyce przedsiębiorstw* [niepublikowana rozprawa doktorska]. Szkoła Główna Handlowa w Warszawie.

Luszniewicz, A., Słaby, T. (1995). *Statystyka stosowana*. Polskie Wydawnictwo Ekonomiczne.

Łukaszewicz, J. (1992). Problem Serbelloni, czyli o warunkowym prawdopodobieństwie ułaskawienia skazańca. *Wiadomości Matematyczne*, *29*, 179–186.

Młodak, A. (2020). *Statystyka w pracach badawczych. Roztropność. Narzędzia. Etyka*. Państwowa Wyższa Szkoła Zawodowa im. Prezydenta Stanisława Wojciechowskiego w Kaliszu, Kaliskie Towarzystwo Przyjaciół Nauk.

Smoluk, A. (1997a). Prognozy intuicyjne. *Ekonomia Matematyczna*, *1*, 71–82.

Smoluk, A. (1997b). Problem Serbelloni i sprawy pokrewne. W: *Dynamiczne modele ekonometryczne* (s. 17–27).

Szymczak, W. (2018). *Praktyka wnioskowania statystycznego*. Wydawnictwo Uniwersytetu Łódzkiego.

Andrzej Młodak

Uniwersytet Kaliski im. Prezydenta Stanisława Wojciechowskiego, Międzywydziałowy Zakład Matematyki i Statystyki; Urząd Statystyczny w Poznaniu, Ośrodek Statystyki Małych Obszarów, Polska / University of Kalisz, Inter-faculty Department of Mathematics and Statistics; Statistical Office in Poznań, Centre of Small Area Estimation, Poland. ORCID: https://orcid.org/0000-0002-6853-9163