

DYDAKTYKA I NAUKA

MIROSŁAW SZREDER

LOSOWE I NIELOSOWE PRÓBY W BADANIACH STATYSTYCZNYCH

WPROWADZENIE

Jednym z istotnych etapów projektowania badania próbkowego (niewyczerpującego) jest określenie techniki wyboru próby (ang. *sampling technique*). Badacz musi najpierw rozstrzygnąć, czy będzie to któraś z grupy technik probabilistycznych (losowych) czy nieprobabilistycznych (niełosowych). O wyborze tym decydować mogą: zakres posiadanych lub dostępnych informacji o populacji, zamiar wykorzystania wnioskowania statystycznego do uogólnień na badaną populację, budżet przeznaczony na badanie, czas realizacji badania, i inne czynniki. Wskazanie konkretnej techniki albo kombinacji technik próbkowania, wiązać się będzie dalej z tak ważnymi kwestiami, jak: sposób określenia liczebności próby, możliwość włączenia do badania już posiadanych informacji (*a priori*) o populacji, wybór metod uzupełnienia (imputacji) braków w danych zebranych od respondentów.

Celem tego opracowania jest scharakteryzowanie wpływu, jaki na wybór techniki próbkowania mają coraz doskonalsze sposoby gromadzenia i przetwarzania danych o różnych populacjach poddawanych badaniom. Dynamiczny postęp technologiczny, którego doświadczamy od kilku dziesięcioleci, oddziałuje w sposób znaczący na wszystkie etapy procesu projektowania badania próbkowego, w tym na decyzję o sposobie doboru próby badawczej. Dzięki temu postępowi coraz więcej wiemy o wielu zbiorowościach poddawanych badaniu, i naturalna staje się potrzeba wykorzystania tej wiedzy już na etapie wyboru próby badawczej.

1. PRÓBA LOSOWA CZY NIELOSOWA?

Dla statystyka dylemat, czy próba powinna być losowa czy niełosowa, właściwie nie istnieje. Teoria klasycznego wnioskowania statystycznego oparta jest na modelu matematycznym, w którym zakłada się, że do próby dostają się jednostki wygenerowane przez mechanizm losujący, który każdej jednostce populacji daje taką samą szansę znalezienia się w próbie. Statystyka dostarcza znacznie więcej sformalizowanych narzędzi

wnioskowania w sytuacji, gdy do wspomnianego modelu matematycznego może się odwołać, anizeli wtedy, gdy próba zostaje pobrana z populacji w inny sposób. Główną zaletą prób probabilistycznych, w szczególności próby losowej prostej, jest możliwość stosowania w dalszej analizie (wnioskowaniu) zasad rachunku prawdopodobieństwa (stąd nazwa: próba probabilistyczna, ang. *probability* – prawdopodobieństwo). Tam gdzie nie ma próby losowej, mechanizmu losującego, czy wreszcie zdarzeń losowych, tam nie ma zastosowania klasyczny rachunek prawdopodobieństwa. Jaka jest więc rola prawdopodobieństwa we wnioskowaniu statystycznym, i czy oznacza to, że próby inne niż losowe (nieprobabilistyczne) są mniej wartościowe w badaniach próbkowych? Zanim odpowiemy na te pytania i uzasadnimy, że próby nielosowe odgrywają także ważną rolę w praktyce wielu badań, przyjrzyjmy się najpierw roli prawdopodobieństwa we wnioskowaniu statystycznym.

2. PRAWDOPODOBIEŃSTWO I PRÓBY LOSOWE

Przede wszystkim warto zauważyć, że najważniejsze elementy wnioskowania statystycznego, w tym określenie wielkości próby, interpretacja wyników estymacji i testowania hipotez, charakterystyka wielkości błędu, wszystkie one w sposób bezpośredni odwołują się do prawdopodobieństwa. Oznacza to, że **u samych podstaw najważniejszych aspektów wnioskowania statystycznego leży założenie o możliwości stosowania rachunku prawdopodobieństwa** (założenie losowości próby badawczej). Prawdopodobieństwo otrzymuje tutaj interpretację częstościową – jedną z czterech najpopularniejszych interpretacji, obok: klasycznej, logicznej i personalistycznej (subiektywnej)¹. Interpretacja częstościowa oznacza, że prawdopodobieństwo zdarzenia definiowane jest jako granica częstości względnej zdarzeń elementarnych sprzyjających temu zdarzeniu, przy liczbie doświadczeń dążącej do nieskończoności. W praktyce rozumieć je należy jako częstość względną odpowiadającą dużej liczbie doświadczeń wykonanych w identycznych warunkach. W tym kontekście interpretuje się własności estymatorów wykorzystywanych we wnioskowaniu. Na przykład, nieobciążenie estymatora oznacza, że przy dużej (rosnącej do nieskończoności) liczbie losowań ustalonej wielkości próby, przeciętna wartość estymatora w tych próbach równa się wartości szacowanego parametru.

W estymacji przedziałowej prawdopodobieństwo występuje *explicite*, jako poziom ufności $(1 - \alpha)$, który interpretować należy jako częstość względną liczby przedziałów ufności pokrywających szacowany parametr populacji w wielokrotnie powtarzanych losowaniach prób, z których każda określa własne granice przedziału ufności. Gdyby tę samą procedurę wielokrotnie powtarzanych prób zastosować do technik nielosowych próbkowania, w których mechanizm generowania obserwacji w próbie nie jest losowy, to nie byłoby żadnych podstaw do zastosowania rachunku prawdopodobieństwa. W konsekwencji, poziom ufności nie posiadałby interpretacji mówiącej cokolwiek o prawdopodobieństwie popełnienia (lub niepopołnienia) błędu w przedziałowej estymacji parametru. Dlatego zastosowanie podanej wyżej interpretacji do estymacji opartej na próbach nieprobabilistycznych (nielosowych) uznać należałoby za nieupoważnione.

¹ Szerzej na ten temat por. Szreder [1994] i [2010a].

Podobnie rzecz się ma z testowaniem hipotez statystycznych². Pojęcie prawdopodobieństwa występuje tu wprost jako poziom istotności (α) – prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu hipotezy, gdy w rzeczywistości jest ona prawdziwa. Prawdopodobieństwo popełnienia tego błędu jesteśmy w stanie ustalić, bo potrafimy obliczyć częstość względną nietypowych prób w długim ciągu losowań. Nietypowych prób, czyli takich, których struktura – różna od struktury populacji – wskazuje na nieprawdziwość sprawdzanej hipotezy, podczas gdy faktycznie hipoteza ta jest prawdziwa. Wiemy z jaką częstotliwością takie nietypowe próby się pojawiają w losowym mechanizmie generowania obserwacji z populacji i na tej podstawie określamy prawdopodobieństwo błędu. Zwróćmy jednak uwagę, że prawdopodobieństwo α odnosi się wyłącznie do niedoskonałości mechanizmu losującego, odpowiedzialnego za tzw. błąd losowania (ang. *sampling error*). Gdyby ktoś zapytał, czy prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu prawdziwej hipotezy uwzględnia też inne okoliczności, które do takiego błędu mogą prowadzić, jak na przykład: brak obserwacji na niektórych jednostkach próby (braki odpowiedzi respondentów), pomyłki respondentów, uchybienia w obliczeniach statystycznych, to oczywiście odpowiedź brzmi – nie. Błąd losowania i przypisane mu prawdopodobieństwo nie uwzględniają niczego poza niedoskonałością samego aktu losowania. A jeżeli tak, to zrozumiałe jest, że błąd ten nie występuje tam, gdzie w ogóle losowania nie ma. W technikach nieprobabilistycznych prawdopodobieństwo to (α) nie miałoby żadnej interpretacji. **Zdarzające się zastosowania teorii weryfikacji hipotez do prób nielosowych powodują, że traci się w tych warunkach możliwość określenia prawdopodobieństwa podjęcia błędnej decyzji, a także możliwość interpretacji przyjętego poziomu istotności.**

Do prawdopodobieństwa odwołujemy się także w ważnym zagadnieniu określenia niezbędnej wielkości próby badawczej. Zwróćmy najpierw uwagę, że pytanie: *Jak duża powinna być w danym badaniu próba?* jest niepełne i trudno jest na nie w ogóle odpowiedzieć, jeżeli nie poda się jakiegoś kryterium precyzji lub dokładności wnioskowania, które mają być spełnione. Dopiero w połączeniu z takim kryterium, pytaniu o liczebność próby można nadać odpowiedni sens logiczny i sformułować je za pomocą właściwych kategorii statystycznych. Na przykład: *Jaka powinna być liczebność próby, aby przeciętne odchylenie uzyskanych w badaniu ocen od prawdziwej wartości szacowanego parametru nie różniło się więcej niż o 2% lub o 5 jednostek miary, w której wyrażona jest badana cecha?* W badaniach próbkowych przyjęło się najczęściej stosować łącznie dwa kryteria:

- średni lub maksymalny błąd, rozumiany jako różnica między oceną z próby a prawdziwą wartością parametru w populacji (np. w badaniach opinii, w których szacowany jest wskaźnik struktury, przyjmuje się zwykle, że maksymalny błąd jest nie większy niż $\pm 3\%$);

- poziom ufności ($1 - \alpha$), czyli prawdopodobieństwo, z jakim przedział o postaci

ocena z próby \pm błąd

zawiera prawdziwą wartość szacowanego parametru.

² Szerzej zagadnienie to zostało omówione w poprzednim numerze „Przeglądu Statystycznego”, M. Szreder [2010b].

Tak sformułowane zadanie określenia minimalnej liczebności próby, dla której spełnione są zadane kryteria, nie stanowi w większości schematów losowania poważnego problemu³.

Zwróćmy wszakże uwagę na istotny element tego wywodu. Założyliśmy *implicite*, że badacz posługuje się próbą probabilistyczną, czyli że przy selekcji elementów do próby stosuje jedną z technik probabilistycznego wyboru. Bez tego założenia nie jest możliwa kontrola ani średniego błędu, o którym wspomnieliśmy, ani współczynnika ufności, który – jak stwierdziliśmy wcześniej – nie ma racji bytu we wnioskowaniu na podstawie prób nieprobabilistycznych (nielosowych). Oba kryteria, które wyżej wypunktowaliśmy, wiążą się bezpośrednio z losowaniem próby, a nie z jakimkolwiek innym sposobem jej wyboru.

3. PRZESŁANKI I KONSEKWENCJE NIELOSOWEGO WYBORU PRÓBY

Jak już wspomnieliśmy, rezygnacja z losowego generowania obserwacji do próby oznacza niemożność stosowania we wnioskowaniu pojęcia prawdopodobieństwa, przynajmniej w jego klasycznej i częstościowej interpretacji. W konsekwencji badacz, decydując się na nielosowy dobór próby, świadomie rezygnuje z rachunku prawdopodobieństwa, a co za tym idzie także z klasycznej Neymanowsko-Pearsonowskiej teorii wnioskowania statystycznego. Mogłoby się więc wydawać, że techniki nielosowego doboru próby są tym gorszym wariantem, wymuszonym okolicznościami, które nie pozwalają na zastosowanie technik probabilistycznych. Do niedawna było to dość powszechne rozumowanie, czego wyrazem był między innymi brak zainteresowania statystyków nielosowymi technikami próbkowania. W rzeczywistości okazuje się, że istnieją sytuacje wymuszające zastosowanie techniki nielosowej, ale **są też takie sytuacje, w których techniki te są szansą, a nie koniecznością**. Koniecznością stają się wówczas, gdy trudne lub niemożliwe jest zapewnienie każdej jednostce populacji równej możliwości dostania się do próby (np. z powodu braku operatu losowania, albo złej jego jakości). Niekiedy inne czynniki mogą decydować o tym wyborze, na przykład krótki czas na realizację badania albo skromne środki finansowe. Wszystko to może sprawić, że – na przykład – dla wewnętrznie zróżnicowanej populacji badacz zdecyduje się na zastosowanie techniki doboru kwotowego próby (nieprobabilistycznego), zamiast bardziej czasochłonnego losowania warstwowego. Szerszego wyjaśnienia wymagają zaś okoliczności, w których wybór techniki nielosowej nie jest niczym wymuszony, lecz jest świadomie stosowany, dzięki walorom danej techniki.

Analogicznie jak w punkcie 2, przyjrzyjmy się najpierw kwestii prawdopodobieństwa. Nielosowy sposób generowania zdarzeń – powtórzmy – nie upoważnia do stosowania klasycznej bądź częstościowej interpretacji prawdopodobieństwa. W praktyce jednak coraz rzadziej mamy do czynienia z „modelową” sytuacją, w której każde zdarzenie jest jednakowo możliwe. Mimo to jednak, pragniemy nadal posługiwać się pojęciem prawdopodobieństwa. Śledząc rozwój teorii prawdopodobieństwa nietrudno zauważyć, że zarówno filozofowie, jak i matematycy poszukiwali takiego sposobu

³ Por. Barnett [1991], Bracha [1998], Steczkowski [1995], Zasepa [1972].

pomiaru prawdopodobieństwa zdarzeń, który odzwierciedlałby bieżący stan wiedzy o określonych zdarzeniach⁴. Powszechnie było i jest dążenie do wykorzystania w ocenie prawdopodobieństwa całej wiedzy o zdarzeniu, nie tylko o jego prostym modelu (takim jak w interpretacji klasycznej) i o obserwowanej częstości względnej realizacji tego zdarzenia w niezmiennych okolicznościach (jak w interpretacji częstościowej). W ten sposób, w wyniku badań m.in. T. Bayesa, L.J. Savage'a, B. de Finetti'ego⁵ sformułowana została interpretacja personalistyczna prawdopodobieństwa, zwana inaczej subiektywną. Przez **subiektywne prawdopodobieństwo** (ang. *subjective probability*) tego, że jakiś sąd na temat zdarzenia *A* jest prawdziwy, rozumie się stopień pewności (ang. *degree of belief*) lub przekonania danej osoby o prawdziwości tego sądu. Zgodnie z tą interpretacją prawdopodobieństwo na temat zdarzenia *A* jest przypisane do danej osoby i może być różne dla różnych osób (ekspertów) w zależności od ich stanu wiedzy, doświadczenia, a nawet intuicji. W przypadku zdarzeń jednostkowych lub rzadko powtarzalnych jest to najczęściej stosowana interpretacja prawdopodobieństwa⁶. Przede wszystkim jednak jej popularność wiąże się z rosnącymi współcześnie zasobami informacji o różnych zdarzeniach, różnych populacjach, będących przedmiotem zainteresowania statystyków. Nieznane wcześniej możliwości gromadzenia, przetwarzania i przesyłania ogromnych zbiorów danych każą zwrócić uwagę na fakt, że coraz rzadziej badacz znajduje się w sytuacji zupełnego braku wiedzy o badanej populacji. Najczęściej wiedzę taką, mimo że częściową i niedoskonałą posiada lub może osiągnąć. I problemem nie jest to, czy ją wykorzystać, lecz jak ją wykorzystać. Jest to kluczowa sprawa dla zrozumienia zarówno coraz śmielszego odwoływania się statystyków do personalistycznej interpretacji prawdopodobieństwa, jak i coraz większej popularności prób nielosowych w badaniach niewyczerpujących. Ostatecznym bowiem **celem wnioskowania statystycznego nie jest osiągnięcie doskonałości w próbkowaniu, lecz jak najlepsze poznanie badanej populacji**. Gdyby mechanizm losowania próby był doskonały, to badacz nie miałby powodów, by w niego ingerować. A ingeruje coraz silniej, gdyż coraz bogatszą posiada wiedzę o populacji, użyteczną wiedzę, którą pragnie włączyć do próbkowania w celu poprawy jakości wnioskowania. Wiedzę tę ma prawo ująć także w formie probabilistycznej, stosując personalistyczną interpretację prawdopodobieństwa.

Ingerencja badacza w klasyczny schemat próby losowej prostej jest widoczna w powszechnie znanych schematach próbkowania, takich jak losowanie systematyczne, czy losowanie warstwowe. W tym ostatnim szczególnie dobrze widać przekonanie o niedoskonałości modelu próby losowej prostej w sytuacji, w której badacz zna (*a priori*) zróżnicowanie wewnętrzne populacji ze względu na cechy istotne dla celu badania.

⁴ G.W. Leibniz w 1955 r. pisał: *Ale to, co prawdopodobne sięga dalej; trzeba je wydobyć z natury rzeczy, a mniemanie osób o wielkim autorytecie jest jedną z rzeczy, które mogą się przyczynić do uprawdopodobnienia jakiegoś mniemania* (G.W. Leibniz [1955], s. 307). Również J. Bernoulli w swojej wcześniejszej pracy z 1713 r. pt. *Ars Conjectandi* („Sztuka przewidywania”) pisał o prawdopodobieństwie jako o stopniu zaufania do realizacji danego zdarzenia na podstawie posiadanej wiedzy o ogólnych okolicznościach tego zdarzenia.

⁵ Szerzej na ten temat, wraz z odpowiednimi odwołaniami bibliograficznymi por. M. Szreder [1994].

⁶ Przykłady zastosowań subiektywnej interpretacji prawdopodobieństwa w różnych zagadnieniach ekonomicznych podają m.in.: H. Kowalczyk [2010], A. Hołda i J. Pocięcha [2009], oraz M. Szreder [2004].

Ingerowanie zaś w mechanizm losowania, to nic innego jak stopniowe jego zastępowanie wyborem opartym na informacjach apriorycznych o badanej zbiorowości. Dlatego wraz ze zwiększaniem się zasobów informacji o badanych populacjach widoczna staje się tendencja do coraz częstszego korzystania w praktyce ze schematów losowania nieprostego⁷, a także z technik nielosowego doboru próby, takich jak wybór kwotowy. Gdy o wielokrotnie badanej populacji badacz posiada stosunkowo bogatą wiedzę, to kwotowy (nieprobabilistyczny) wybór próby może się okazać lepszy od schematów probabilistycznych.

Obawy przed stosowaniem technik nielosowego doboru próby dotyczą przede wszystkim kwestii możliwości oszacowania błędów pojawiających się przy uogólnianiu prawidłowości zaobserwowanych w próbie na całą populację. Techniki próbkowania nieprobabilistycznego nie zawierają tego składnika błędu, który w dotychczasowym rozwoju statystyki został najlepiej poznany i opisany – błąd losowania (nazywanego też błędem losowym). A ten właśnie błąd jest nie tylko podstawowym i zwykle jedynym błędem, który w sposób wymierny zostaje przypisany wynikowi wnioskowania w badaniach próbkowych⁸, ale jest ponadto podstawowym kryterium służącym określeniu liczebności próby (por. wyżej p. 2). W praktyce więc ośrodki badawcze stosujące nielosowe techniki próbkowania, jeżeli tylko doświadczenia ich są wystarczająco bogate, odwołują się do zgromadzonych doświadczeń i do wiedzy ekspertów, aby w sposób skwantyfikowany określić błąd, jaki może wynikać z posłużenia się w opisie populacji informacjami z próby (błąd próbkowania). Tak postępują znane polskie i zagraniczne ośrodki badawcze w odniesieniu do często stosowanego próbkowania kwotowego. Podkreślić jednak należy, że w każdym innym przypadku, gdy informacje o możliwym błędzie próbkowania są niewystarczające lub mało wiarygodne, charakterystykę populacji ograniczyć należy do opisu statystycznego, a nie wnioskowania statystycznego. Zaobserwowane prawidłowości w próbie odnieść można nadal do zbiorowości, której próba ta jest reprezentantem, jednakowoż bez przypisywania im wielkości błędu lub prawdopodobieństw ich prawdziwości. Nieuzasadnione jest w tych okolicznościach posługiwanie się metodami wnioskowania statystycznego.

4. PODSUMOWANIE

Zwiększające się zasoby informacji na temat różnych zbiorowości poddawanych badaniom statystycznym sprawiają, że w badaniach próbkowych źródłem informacji nie jest już wyłącznie próba statystyczna, jak często bywało w przeszłości. Posiadana informacja *a priori* o badanej populacji staje się równie ważnym źródłem informacji dla statystyka. W tym upatrywać należy rosnącej popularności nielosowych technik

⁷ O wnioskowaniu dla prób nieprostych w wielu swoich pracach pisze prof. Cz. Domański wraz z zespołem, por. np. Cz. Domański i K. Pruska [2000].

⁸ Pozostałe składniki błędu całkowitego, jak wiadomo, o wiele trudniej poddają się pomiarowi. Ilustracją tego faktu może być twierdzenie niektórych znanych statystyków, iż błąd losowy jest nadmiernie badany (*sampling error is „over-researched”*). Sformułowanie to pojawiło się m.in. w artykule znanych statystyków Richarda Platka i Carla-Erika Särndala pt. *Can a statistician deliver?* opublikowanym w jęz. polskim wraz z dyskusją przez „Wiadomości Statystyczne” nr 4, 2001 r.

próbki, a także innych niż klasyczna i częstościowa interpretacji prawdopodobieństwa. Rosnące możliwości gromadzenia i przetwarzania informacji statystycznych powodować będą – jak się wydaje – upowszechnianie się modelu wnioskowania opartego na łącznej wiedzy o badanej populacji: *a priori* i z próby. Wiedza wstępna (*a priori*) wykorzystywana będzie na wszystkich etapach badania próbkowego.

LITERATURA

- Barnett V., [1991], *Sample Survey. Principles and Methods*, E. Arnold.
- Bracha Cz., [1998], *Metoda reprezentacyjna w badaniu opinii publicznej i marketingu*, Wyd. Efekt.
- Domański Cz., Pruska K., [2000], *Nieklasyczne metody statystyczne*, PWE, Warszawa.
- Hołda A., Pocięcha J., [2009], *Probabilistyczne metody badania sprawozdań finansowych*, Wyd. Uniwersytetu Ekonomicznego w Krakowie, Kraków 2009.
- Kowalczyk H., [2010], *O eksperckich ocenach niepewności w ankietach makroekonomicznych*, „Bank i Kredyt”, nr 41, s. 101-122.
- Leibniz W.G., [1955], *Nowe rozważania dotyczące rozumu ludzkiego*, t. 2, PWN, Warszawa.
- Steczkowski J., [1995], *Metoda reprezentacyjna w badaniach ekonomiczno-społecznych*, PWN, Warszawa-Kraków.
- Szreder M., [1994], *Informacje a priori w klasycznej i bayesowskiej estymacji modeli regresji*, Wydawnictwo Uniwersytetu Gdańskiego.
- Szreder M., [2010a], *Metody i techniki sondażowych badań opinii*, PWE, Warszawa.
- Szreder M., [2010b], *O weryfikacji i falsyfikacji hipotez*, „Przegląd Statystyczny” nr 2-3, t. 57, s. 82-88.
- Szreder M., [2004], *Od klasycznej do częstościowej i personalistycznej interpretacji prawdopodobieństwa*, „Wiadomości Statystyczne” nr 8, s. 1-10.
- Zasępa R., [1972], *Metoda reprezentacyjna*, PWE, Warszawa.