

Próba reprezentatywna – potrzeba i propozycja definicji

Mirosław Szreder^a

Streszczenie. „Próba reprezentatywna” należy do tych nielicznych pojęć wnioskowania statystycznego, które nie doczekały się precyzyjnej i powszechnie uznanej definicji. Potrzeba istnienia takiej definicji jest mniej odczuwalna w naukach ścisłych i przyrodniczych, w których empiryczne badania eksperymentalne, mające głównie cele eksploracyjne, rzadko wymagają uzyskania prób reprezentatywnych. Z kolei cele deskryptywne i konfirmacyjne, właściwe dla nauk społecznych i ekonomicznych, sprawiają, że jednym z najważniejszych wymogów odnoszących się do próby badawczej jest możliwość dokonywania na jej podstawie uogólnień odpowiednich oszacowań dla populacji wraz z oceną ich błędów.

Celem artykułu jest krytyczne omówienie roli próby reprezentatywnej we wnioskowaniu naukowym i statystycznym, a ponadto zaproponowanie definicji „próby reprezentatywnej” i wskazanie jej specyfiki względem prób nielosowych, których reprezentatywność jest ograniczona jedynie do ściśle wskazanych zmiennych. W proponowanej definicji autor kładzie nacisk na zgodność struktury próby ze strukturą populacji oraz na możliwości wnioskowania, w którym ocenom uzyskanym z próby są przypisane odpowiednio oszacowane liczbowe wartości błędów. Postuluje także, aby próbę, która wykazuje zbieżność rozkładów tylko niektórych cech z ich rozkładami populacyjnymi, nazywać nie „próbą reprezentatywną”, lecz „próbą reprezentatywną ze względu na wyszczególnione zmienne”. Uzasadniono także zbędność tworzenia we wnioskowaniu statystycznym kategorii „próby reprezentatywnej”.

Słowa kluczowe: próba reprezentatywna, wnioskowanie statystyczne, wnioskowanie naukowe, cele eksploracyjne, cele deskryptywne, cele konfirmacyjne

JEL: C10, C13, C18

Representative sample – the need for and the proposal of a definition

Abstract. A ‘representative sample’ is one of few terms used in statistical inference which do not have a precise and commonly recognised definition. The need for such a definition is less nagging in sciences and natural sciences, where empirical experimental research has exploratory objectives, and thus rarely requires obtaining representative samples. On the other hand, descriptive and confirmatory objectives of social sciences and economics necessitate samples on the basis of which it would be possible to make generalisations of relevant estimates for a given population and assess their errors.

The aim of this paper is to discuss the role of a representative sample in scientific and statistical inference, and additionally to propose its definition and show its specificity in the case where at the same time it is a non-random sample, whose representativeness is limited to some specified variables only. In the proposed definition, the author focuses on the consistency of the sample structure with the structure of the population and the possibility of making inference

^a Uniwersytet Gdański, Wydział Zarządzania, Polska / University of Gdańsk, Faculty of Management, Poland.
ORCID: <https://orcid.org/0000-0002-7597-0816>. E-mail: miroslaw.szreder@ug.edu.pl.

in which estimates obtained from the sample have numerical error values assigned to them. It is also argued that samples which show convergence of the distributions of only some characteristics with their population distributions should be called 'representative samples in terms of specified variables' instead of just 'representative samples'. Moreover, it is demonstrated that the category of a random sample which is inseparably linked to statistical inference does not need any equivalents.

Keywords: representative sample, statistical inference, scientific inference, exploratory objectives, descriptive objectives, confirmatory objectives

1. Wprowadzenie

W badaniach statystycznych – pełnych i niewyczerpujących (próbkowych) – istnieje wiele kategorii i pojęć od dawna precyzyjnie zdefiniowanych i niebudzących kontrowersji co do ich rozumienia czy stosowania. Dotyczy to zwłaszcza opisu statystycznego, charakterystycznego dla badań pełnych, czyli takich, w których pomiarem (obserwacją) objęte zostają wszystkie jednostki analizowanej populacji. Więcej niejednoznaczności, jeśli chodzi o definiowanie i rozumienie kategorii stosowanych przez statystyków, spotyka się w badaniach niewyczerpujących. Jest to spowodowane zarówno stosunkowo krótkim – bo zaledwie stuletnim – okresem rozwoju i utrwalania się w nauce matematycznego (probabilistycznego) modelu wnioskowania statystycznego, jak i ściślejszymi niż w opisie statystycznym wymogami jednoznaczności rozumienia pojęć w modelowaniu matematycznym. Dokonujący się od trzech dekad dynamiczny rozwój informatycznych technologii przetwarzania danych, któremu towarzyszy również szybki wzrost zainteresowania badaczy niewyczerpującymi badaniami statystycznymi, pogłębia potrzebę jednoznacznego definiowania najważniejszych pojęć występujących w tego rodzaju badaniach. Wydaje się to jednym z podstawowych warunków rzetelności przekazu wyników badań statystycznych oraz poprawnej i jednoznacznej ich interpretacji.

Jedną z najważniejszych kategorii występujących w badaniach niewyczerpujących jest *próba badawcza*, nazywana też *próbą statystyczną*. Od sposobu pobierania próby i konkretnej techniki jej uzyskania zależą nie tylko możliwości uogólniania wyników na populację, z której pochodzi próba, lecz także możliwości określenia precyzji wnioskowania i wielkości błędów otrzymanych ocen czy trafności podjętych decyzji odnoszących się do postawionych wcześniej hipotez. To naturalne, że atrybuty próby, zdefiniowane sposobem jej wyboru, zostają dla wygody ujęte w krótkiej, zwykle dwuczłonowej, nazwie tworzącej nową kategorię, np. *próba losowa*, *próba kwotowa*, *próba reprezentatywna* itp. Wśród tych kategorii za jedną z najważniejszych – a jednocześnie najbardziej niejednoznacznych – należy uznać *próbę reprezentatywną* (zob. np. Analytical Methods Committee, 2016; Rudolph i in., 2023). Pojęcie *próby reprezentatywnej* nie jest ani tożsame z pojęciem *próby losowej*, ani nie jest równoważne z żadną inną popularną kategorią statystyczną.

W statystyce nie istnieje jedna, uznana przez większość badaczy definicja *próby reprezentatywnej*. Potrzeba stworzenia takiej definicji istniała od dawna, ale obecnie staje się

szczególnie pilna wobec rosnącej frakcji badań, mniej lub bardziej profesjonalnych, w których zwięzłe stwierdzenie, że wykonano je na próbie reprezentatywnej, bez wskazania metody jej uzyskania, ma stanowić wystarczające uzasadnienie dla ekstrapolacji wniosków na całą populację. Dotyczy to w szczególności niektórych badań rynkowych i społecznych, w tym części badań opinii publicznej, w których użycie kategorii *próba reprezentatywna* ma z jednej strony informować o możliwości uogólnienia wniosków z próby na całą zbiorowość, z której pochodzi próba, a z drugiej – w wielu przypadkach zniechęcić odbiorcę do wnikania w szczegółową procedurę jej wyboru. Trzeba jednak zauważyć, że brak jednoznacznej, ścisłej definicji *próby reprezentatywnej* uniemożliwia w praktyce ocenę rodzajów i skali błędów, którymi mogą być obciążone wyniki badania niewyczerpującego. Gdy termin ten jest różnie definiowany i rozumiany zarówno przez badaczy, jak i odbiorców, wówczas nie jest jasne, jakie konkretnie techniki próbkowania mogą się kryć za tym terminem, a w rezultacie – jakiego rodzaju i jakiej wielkości błędy mogą obciążać wyniki takiego badania.

Dążenie do przypisania próbie badawczej waloru reprezentatywności wiąże się oczywiście z samą istotą wnioskowania statystycznego. Próba dobrze reprezentująca populację, z której pochodzi, daje prawo do formułowania prawidłowości i opisu wzorców, które są charakterystyczne dla całej tej populacji. Z kolei próba, która nie stanowi dobrej reprezentacji populacji, jest pozbawiona prawa do uogólnień. Reprezentatywność ujawnia więc swoje podstawowe zalety we wnioskowaniu statystycznym. Czy również we wnioskowaniu naukowym jako takim? To już jest kwestia warta przedyskutowania.

Celem artykułu jest krytyczne omówienie roli próby reprezentatywnej we wnioskowaniu naukowym i statystycznym, a ponadto zaproponowanie definicji *próby reprezentatywnej* i wskazanie jej specyfiki względem prób nielosowych, których reprezentatywność jest ograniczona jedynie do ściśle wskazanych zmiennych.

2. Próba reprezentatywna we wnioskowaniu naukowym i statystycznym

Relacji między wnioskowaniem naukowym (ang. *scientific inference*) a wnioskowaniem statystycznym (ang. *statistical inference*) nie można przekonująco scharakteryzować bez odwołań do istoty poznania naukowego i celów badań naukowych. Ważną rolę w dyskusji na ten temat odgrywa kategoria próby, a ściślej możliwości dokonywania uogólnień na podstawie próby. „Esencją wiedzy jest uogólnianie”¹ – stwierdza Reichenbach w swoim słynnym dziele na temat filozofii nauki (1951, s. 5). I dodaje: „Uogólnienie ponadto jest samą naturą wyjaśnienia. Przez wyjaśnienie

¹ Oryg.: „The essence of knowledge is generalization”.

zaobserwowanego faktu rozumiemy włączenie tego faktu w ramy ogólnego prawa”² (Reichenbach, 1951, s. 6). Dalej wskazuje, że naukowe objaśnianie wymaga dużej liczby obserwacji i krytycznego myślenia.

Podobnie cele nauki rozumieją współcześni badacze uczestniczący w dyskusji na temat miejsca wnioskowania statystycznego w nauce. Podkreślają oni, że naukowe uogólnienie wiąże się ze znalezieniem i wskazaniem warunków (okoliczności), w odniesieniu do których uzyskane wyniki badawcze lub zaobserwowane empirycznie regularności mają zastosowanie (Rothman i in., 2013). Przy czym w naukowym poznaniu chodzi głównie o dążenie do sformułowania uniwersalnych praw natury, nieograniczonych do szczególnego miejsca i ustalonego czasu obserwacji. W tym kontekście wiele badań statystycznych – zwłaszcza społecznych lub rynkowych – w mniejszym stopniu będzie realizować tego typu cele eksploracyjne, a raczej będzie zorientowanych na cele opisowe. Badania eksploracyjne zmierzają do poznania przyczyn i mechanizmów działań, postawienia nowych pytań oraz doskonalenia metod do dalszych badań (Babbie, 2004; Tong, 2019), a badania opisowe – również dostarczające ważnych informacji – takich celów w większości nie mają. Dotyczy to zwłaszcza badań niewyczerpujących w obszarze zagadnień ekonomicznych i społecznych, które najczęściej odnoszą się do precyzyjnie zdefiniowanej, specyficznej w czasie i przestrzeni populacji. Uogólnienie wykraczające poza populację, z której pobrano próbę, wymagałoby dodatkowej wiedzy lub wsparcia informacyjnego z innych źródeł. Nie wynika z tego oczywiście żaden wniosek, który pozwałaby kwestionować wartość naukową lub poznawczą wiedzy, która jest rezultatem kumulacji osiągnięć ze zrealizowanych poprawnie niewyczerpujących badań statystycznych. Opisowe, a także confirmacyjne znaczenie badań statystycznych (np. w naukach medycznych i przyrodniczych) jest trudne do przecenienia. Niektórzy badacze (np. Tong, 2019) podkreślają właśnie tę ich rolę (confirmacyjną i opisową) w poznaniu naukowym. Stwierdzają, że wnioskowanie statystyczne jest najodpowiedniejsze w analizach confirmacyjnych, w których wcześniej dokładnie określono plan badawczy i metody badawcze, a samo wnioskowanie stanowi ważny końcowy etap iteracyjnego procesu uczenia się (ang. *iterative learning process*), typowego dla poszukiwań naukowych³.

Wydaje się, że dobrym punktem wyjścia do określenia pożądanych właściwości próby badawczej w badaniach naukowych i statystycznych mogą być dwa wyżej

² Oryg.: „Generalization, furthermore, is the very nature of explanation. What we mean by explaining an observed fact is incorporating that fact into a general law”.

³ Podobną opinię wyraża George Box, który stwierdza m.in.: „Nie ma innego uzasadnienia dla istnienia statystyki poza byciem katalizatorem dla dociekań naukowych, w których tylko ostatnie stadium, gdy wykonana została już cała praca twórcza, jest związane z ustalonym ostatecznym modelem i z rygorystycznym przetestowaniem wniosków” (Spiegelhalt i in., 1995, s. 77; oryg.: „Statistics has no reason for existence except as a catalyst for scientific enquiry in which only the last stage, when all the creative work has already been done, is concerned with a final fixed model and a rigorous test of conclusions”).

wymienione, odrębne dla tych badań, cele: eksploracyjny i konfirmacyjny. Przesądzają one o tym, że innych właściwości oczekuje się od prób w badaniach eksploracyjnych, a innych – w badaniach konfirmacyjnych. Dotyczy to także reprezentatywności próby – atrybutu, którego rola w badaniach eksploracyjnych, zwłaszcza w naukach ścisłych i przyrodniczych, jest inna niż w badaniach konfirmacyjnych czy opisowych.

Marcia McNutt, w latach 2013–2016 redaktorka naczelna prestiżowego czasopisma „Science”, obecnie prezeska amerykańskiej National Academy of Sciences, stwierdziła w wywiadzie sprzed kilku lat, że: „Badanie eksploracyjne raczej odkrywa i analizuje nowe problemy niż koncentruje się na testowaniu istniejących hipotez. Ale badacze niekiedy czują, że powinni swoje badania eksploracyjne zaprezentować w formie testowania hipotez, co jest całkowicie nierzetelne”⁴ (Shell, 2016). Dla lepszego rozróżnienia celów badań eksploracyjnych i badań konfirmacyjnych niektórzy badacze sugerują nazywanie badań eksploracyjnych *badaniami generującymi nowe hipotezy* (ang. *hypothesis-generating research*), a badań konfirmacyjnych – *badaniami skoncentrowanymi na testowaniu hipotez* (ang. *hypothesis-testing research*). W szczególności statystyczna weryfikacja hipotez – mająca charakter dedukcyjny (z wyjątkiem podejścia bayesowskiego; zob. Szreder, 2019) – nie może być uważana za badanie eksploracyjne. Nie oznacza to jednak, że wnioskowanie statystyczne nie tworzy nowej wiedzy, np. poprzez estymację parametrów populacji, opis zależności między cechami lub falsyfikację hipotez. W całym przedsięwzięciu badawczym, w szczególności w naukach przyrodniczych i medycznych, kognitywna rola wnioskowania statystycznego nie jest jednak decydująca (Tong, 2019). Najważniejsze są cele i badania wyjaśniające pierwotne przyczyny i mechanizmy oddziaływania na siebie zmiennych, a więc rezultaty dociekań wywodzących się z nauk merytorycznie właściwych dla danego zagadnienia badawczego. Wnioskowanie statystyczne jest włączone w ten proces badawczy w jednej z ostatnich faz. Nie powinno więc dziwić, że inaczej na rolę wnioskowania statystycznego – i w dalszej kolejności na wymogi odnoszące się do właściwości próby – patrzą naukowcy, których domeną są badania eksploracyjne, a inaczej badacze specjalizujący się w zastosowaniach wnioskowania statystycznego.

O ile reprezentatywność próby jest uznawana za warunek *sine qua non* wiarygodności niewyczerpujących badań społecznych i ekonomicznych, o tyle w naukach medycznych i przyrodniczych kategoria ta ma inny status. Reprezentatywność nie zawsze jest w nich wskazana lub pożądana. Interesującą dyskusję na ten temat zamieszczono w 2013 r. w czasopiśmie „International Journal of Epidemiology”. Rozpoczyna ją artykuł Rothmana i in. (2013) o sugestywnym i nieco zaskakującym tytule *Why representativeness should be avoided* (Dlaczego powinno się unikać reprezentatywności). Autorzy przekonują, że próba reprezentatywna stanowi dobrą podstawę do

⁴ Oryg.: „An exploratory study explores new questions rather than tests an existing hypothesis. But scientists have felt that they had to disguise an exploratory study as hypothesis testing and that is totally dishonest”.

wnioskowania statystycznego o ściśle zdefiniowanej populacji, ale nie upoważnia do dalej idących uogólnień, charakterystycznych dla poszukiwań stricte naukowych. Eksploracyjne cele nauki obejmują niezmiennie prawa natury wykraczające poza daną zbiorowość, którą statystyk musi dokładnie określić przed pobraniem próby. „Błędem jest myślenie, że wnioskowanie statystyczne jest tym samym co wnioskowanie naukowe”⁵ (Rothman i in., 2013, s. 1013). Jeżeli cel badania nie ogranicza się do opisu stanu i cech pewnej populacji (co ma miejsce w przypadku wielu badań społecznych), lecz obejmuje wykorzystanie metod eksperymentalnych, to próba reprezentatywna często nie jest w ogóle brana pod uwagę. Immunolodzy wykonujący eksperymenty na zwierzętach (łac. *in vivo*) dobierają zwykle próby silnie niereprezentatywne, homogeniczne, składające się z osobników mających podobne geny, żyjących w tych samych warunkach i korzystających z identycznej diety. Cechy immunologiczne tych zwierząt mogą nie być takie same jak u ludzi, ale poszukiwanie dróg uogólnień wyników tych badań na zbiorowości ludzkiej wiąże się z odkrywaniem zależności między kontrolowanymi czynnikami i środowiskiem życia a odpornością organizmów. Decydującą rolę mają tu do odegrania nauki chemiczne i biologiczne. O takich próbach mówi się, że nie są reprezentatywne w oszacowaniach (ocenach) odnoszących się do danej populacji, lecz na podstawie szerszej wiedzy o mechanizmach funkcjonowania organizmów zwierząt i ludzi stwierdza się, że mogą być reprezentatywne, jeśli chodzi o interpretację. Dlatego w naukach eksperymentalnych stosowane jest niekiedy rozróżnienie między *uogólnianiem ocen* (ang. *generalisability of estimate*) i *uogólnianiem interpretacji wniosków* (ang. *generalisability of interpretation*; zob. Rudolph i in., 2023).

W badaniach medycznych na nieadekwatność postulatu reprezentatywności próby zwraca się uwagę również w nieco innym kontekście. Otóż prawidłowości i współzależności cech zaobserwowane w próbie reprezentatywnej odnoszą się do całej populacji, którą próba reprezentuje, ale w poszczególnych grupach osób (np. grupach wieku) mogą się istotnie różnić. Samo stwierdzenie, że określona terapia jest skuteczna dla zróżnicowanych jednostek zbiorowości, znajdujących swoje odzwierciedlenie w próbie reprezentatywnej, nie oznacza, że jest ona wskazana dla każdej podgrupy tej zbiorowości. Ogólny efekt danej terapii jest bowiem uśrednionym efektem, ważonym rozkładami liczebności poszczególnych podgrup jednostek zbiorowości. Ważniejsze w naukach medycznych jest odkrycie oddziaływania rozważanej terapii w wyszczególnionych podgrupach, np. u dzieci, osób dorosłych i starszych.

⁵ Oryg.: „The mistake is to think that statistical inference is the same as scientific inference”. Inni dyskutanci prezentują podobny, chociaż nieco bardziej wyważony punkt widzenia na kwestię reprezentatywności próby. Sprowadza się on do stwierdzenia stanowiącego tytuł jednego z tekstów w tej dyskusji: „Reprezentatywność zwykle nie jest konieczna i często powinno się jej unikać” (Richiardi i in., 2013). Wydaje się jednak, że najważniejsza jest konkluzja Ebrahima i Smitha (2013), że nie należy ani unikać reprezentatywności, ani jej bezkrytycznie adoptować, lecz każdorazowo oceniać jej wartość w zaprojektowanym przedsięwzięciu badawczym.

Jednak o tym, czy w określonym badaniu naukowym pożądana jest próba reprezentatywna, decyduje nie dziedzina badań, lecz cel badawczy. W podanym wyżej przykładzie z zakresu medycyny próba reprezentatywna będzie preferowana wtedy, gdy cel badań będzie miał charakter opisowy (deskryptywny), a nie eksploracyjny. Typową sytuacją tego rodzaju jest charakterystyka natężenia występowania chorób lub czynników ryzyka zachorowań na określonym terenie lub w zbiorowości określonej pewnymi cechami. Warto przy tym zauważyć, że wyniki takiego badania opisowego nie tylko wzbogacają wiedzę o rozpowszechnianiu się danej choroby, ale zidentyfikowane w nim prawidłowości statystyczne mogą stanowić podstawę formułowania nowych, oryginalnych hipotez badawczych. Nie są więc te dwa cele – opisowy i eksploracyjny – tak bardzo oddalone od siebie, jak mogłyby sugerować wskazane wcześniej klasyfikacje.

Abstrahując w dalszej części opracowania od nauk eksperymentalnych, warto zauważyć, że w naukach społecznych i ekonomicznych reprezentatywność próby bywa dość często utożsamiana z losowym wyborem próby. W rzeczywistości zaś losowanie jest jednym, ale nie jedynym sposobem uzyskania próby upoważniającej do uogólnień na całą populację. Wiele zależy od tego, jaką wstępną wiedzę o populacji dysponuje badacz, zanim podejmie decyzję o sposobie próbkowania. Wartość tej apriorycznej wiedzy o poddanej badaniu populacji ma decydujące znaczenie w rozstrzygnięciu nie tylko o tym, jaką techniką losowania należy się posłużyć, lecz czy w ogóle wybór losowy jest w świetle tej wiedzy najlepszym rozwiązaniem. I mimo że coraz więcej wiemy o populacjach poddawanych badaniom statystycznym, to losowy wybór próby był i pozostaje najpowszechniejszą metodą próbkowania.

3. Losowość a reprezentatywność próby

Z losowością mamy do czynienia w dwojakim znaczeniu. Z jednej strony otaczają nas naturalne zjawiska i procesy, na których konsekwencje nie mamy żadnego wpływu, takie jak burze, trzęsienia ziemi czy wybuchy wulkanów, których czas i miejsce wystąpienia oraz skutki uznajemy za losowe. Tak rozumiana losowość jest nam narzucona przez naturę, a dokładniej – wynika z niedostatecznej wiedzy na temat przyczynowo-skutkowych mechanizmów powstawania określonych zjawisk i zdarzeń. Z drugiej zaś strony losowość – a właściwie wybór losowy – włączamy do naszych działań intencjonalnie dla osiągnięcia określonego celu. Innymi słowy, by rozwiązać pewien problem, który sam w sobie nie musi mieć żadnych atrybutów stochastycznych, odwołujemy się niekiedy do wyboru losowego – losowania. Celem jest zwykle dokonanie takiego wyboru spośród określonego zestawu opcji, który będzie wolny od wszelkiego racjonalizmu i subiektywizmu człowieka, np. losowanie sędziów do rozpatrywania spraw wpływających do sądów. Chodzi więc o taki mechanizm wyboru, który nie jest w żaden

sposób zdeterminowany przez czynniki dające się przypisać czyjemuś racjonalnemu działaniu, ludzkim sympatiom czy emocjom. Dowlen (2008) określa taki wybór mianem *aracjonalny* dla podkreślenia tego, że nie jest on ani racjonalny (oparty na przesłankach rozumowych), ani irracjonalny (podjęty z pobudek emocjonalnych).

Współcześnie można zaobserwować rosnącą popularność włączania mechanizmu losowego do różnych dziedzin życia. Losuje się nie tylko sędziów do prowadzenia poszczególnych spraw w wymiarze sprawiedliwości, recenzentów w przewodach na stopnie i tytuły naukowe, terminy rozpatrywania wiz dla obcokrajowców, lecz także przedstawicieli społeczeństwa do sformowania paneli dyskusyjnych i doradczych dla sprawujących władzę w danym kraju lub regionie. W ostatnich latach szczególnie popularne jest korzystanie z opinii losowo wybranych przedstawicieli społeczeństwa do pracy zespołowej i wypracowania stanowiska w różnych ważnych sprawach gospodarczych, społecznych i związanych z ochroną środowiska naturalnego. Korzysta z tej możliwości Parlament Europejski w Strasburgu (Pub Affairs Bruxelles, 2021) oraz niektóre kraje: Francja⁶, Wielka Brytania i Niemcy, zainteresowane opiniami paneli obywatelskich (ang. *citizens' panels*). Podmioty te uznają losowanie za skuteczny mechanizm pozwalający formować panele reprezentatywne (stanowiące adekwatną reprezentację) dla zbiorowości danego kraju lub regionu. Warto jednak postawić ogólniejsze pytanie: czy losowanie jest najlepszym – a może jedynym – sposobem na uzyskanie próby reprezentatywnej z określonej zbiorowości?

Relacje między pojęciami *reprezentatywny* i *losowy* są ważne, ponieważ dotyczą samych fundamentów wnioskowania statystycznego. Warto zdać sobie sprawę z tego, że formalne wnioskowanie statystyczne jest nieodłącznie związane z aktem losowania z dwóch zasadniczych powodów:

1. W wyniku losowania uzyskuje się przeciętnie – w hipotetycznym długim ciągu powtarzalnych losowań – próbę, w której rozkłady cech statystycznych są zbliżone do rozkładów tych cech w populacji, czyli próbę dobrze odzwierciedlającą strukturę populacji, bez względu na to, które z tych cech są przedmiotem zainteresowania badacza.
2. Probabilistyczny (matematyczny) model wnioskowania statystycznego, odwołujący się do częstościowej interpretacji prawdopodobieństwa, wymaga losowego generowania zdarzeń (obserwacji) w próbie, co z kolei umożliwia badaczowi wyrażenie w kategoriach probabilistycznych wielkości błędów oszacowań lub ryzyka podjęcia błędnej decyzji w odniesieniu do testowanych hipotez.

⁶ We Francji jednym z pierwszych ogólnokrajowych paneli obywatelskich, stworzonych z inicjatywy prezydenta Emmanuela Macrona po protestach tzw. żółtych kamizelek w 2019 r., była 150-osobowa grupa losowo wybranych osób, której zadaniem było wypracowanie stanowiska w sprawie reform związanych ze zmianami klimatycznymi, a także z postulatami zwiększenia demokracji bezpośredniej we Francji (Trian, 2021).

Wśród podstawowych przesłanek losowego wyboru próby podkreśla się na ogół pierwszy z wyżej wymienionych powodów – gwarancję reprezentatywności próby (przeciętnie). Rzadziej wspomina się o tym, że równie ważna jest możliwość przypisania wynikom wnioskowania probabilistycznych miar dokładności uzyskanych ocen. W statystyce – tak jak w każdej innej dyscyplinie naukowej – znaczenie mają nie tylko wyniki przeprowadzonych badań, lecz także towarzyszące im błędy. Dlatego rozpatrując inne (nieprobabilistyczne) sposoby uzyskania próby będącej dobrą reprezentacją populacji, należy mieć na uwadze także to, czy umożliwiają one oszacowanie błędów wyników, które miałyby odpowiednie uzasadnienie naukowe.

W relacjach między kategoriami *losowości* i *reprezentatywności* istotne znaczenie ma to, że nie każda konkretna próba wylosowana ze zdefiniowanej wcześniej populacji będzie wykazywała taki sam stopień podobieństwa rozkładów cech co rozkłady analogicznych cech w populacji. Innymi słowy, niektóre próby będą miały strukturę bardziej, a niektóre mniej przypominającą strukturę populacji. Jedne będą bardziej, a inne mniej reprezentatywne. Przeciętnie natomiast mechanizm indywidualnego, niezależnego losowania obserwacji z populacji gwarantuje uzyskanie próby o rozkładach cech takich jak rozkłady tych cech w populacji⁷. To, że w statystyce posługujemy się różnymi schematami wyboru próby – a nie wyłącznie losowaniem prostym indywidualnym – wiąże się z dążeniem do uzyskania nie tylko przeciętnie, lecz także w konkretnym losowaniu, takiej próby, której charakterystyki będą jak najbardziej zbliżone do analogicznych charakterystyk w populacji. I mimo że matematyczny model wnioskowania oparty jest na definicji prostej próby losowej, to w praktyce, dla zwiększenia szans na otrzymanie w konkretnym losowaniu próby o znacznym stopniu reprezentatywności, korzysta się z kilku innych schematów losowania. Ich wspólną cechą jest włączenie do każdego z tych schematów informacji wstępnych (a priori), które badacz posiadał przed przystąpieniem do wyboru próby. Im bogatsza jest wstępna wiedza o populacji – np. pochodząca z wcześniejszych badań na temat tych samych lub innych jej cech albo wynikająca z informacji zawartych w rejestrach administracyjnych lub innych zbiorach danych – tym większe są szanse na wysoki stopień reprezentatywności próby, gdy wiedza ta zostanie wykorzystana w losowaniu. Dobrą tego ilustracją może być znany i z powodzeniem stosowany w wielu badaniach schemat losowania warstwowego, w którym na podstawie apriorycznych informacji o populacji jej jednostki zostają pogrupowane w homogeniczne warstwy (powarstwowane),

⁷ W statystyce podejście, w którym zakłada się, że źródłem losowości ocen parametrów populacji jest przyjęty plan losowania, a wartości cech w populacji nie są losowe (czyli są ustalone), nazywa się *podejściem randomizacyjnym*. Obok niego istnieje drugie podejście, nazywane *modelowym*, którym się w tym opracowaniu nie zajmujemy. Jego najważniejszym założeniem jest to, że wektor wartości badanej cechy traktowany jest jako wektor realizacji zmiennych losowych, a źródłem losowości jest model nadpopulacji (zob. np. Bracha, 1996; Steczkowski, 1995).

a losowanie odbywa się odrębnie z każdej warstwy. Schemat ten pozwala na uzyskanie próby, której struktura ze względu na cechy warstwujące jest identyczna ze strukturą populacji. Zwłaszcza w przypadku populacji silnie niejednorodnych jest to jedna z dobrze znanych możliwości zwiększenia efektywności wnioskowania.

Wiedza o populacji, którą badacz posiadał przed wyborem jednostek do próby, jest tym źródłem informacji, które odgrywa najważniejszą rolę w dążeniu do zapewnienia możliwie największego podobieństwa struktury próby do struktury populacji. Jeżeli dla badacza wystarczające jest ograniczenie tego podobieństwa do ściśle określonych cech (np. w populacji przedsiębiorstw do rodzaju działalności według PKD, wielkości zatrudnienia lub formy organizacyjnoprawnej), to może się on zadowolić próbą, której struktura będzie zgodna ze strukturą populacji tylko ze względu na te cechy. Spełnienie postulatu identycznej lub podobnej struktury próby i populacji w odniesieniu do kilku konkretnych cech statystycznych jest zadaniem prostszym niż uzyskanie próby, która odzwierciedlałaby strukturę populacji ze względu na wszystkie cechy. Przykładem takiego podobieństwa ograniczonego do wybranych cech próby i populacji jest technika wyboru kwotowego – jedna z nieprobabilistycznych technik próbkowania. Za jej podstawową zaletę uważa się zgodność rozkładów cech (charakterystyk kontrolnych) wskazanych przez badacza na wstępie z rozkładami tych cech w populacji. Znając – zwykle na podstawie danych wtórnych – strukturę populacji ze względu na wyróżnione cechy, dokonuje się nielosowego wyboru jednostek do próby w takich proporcjach, aby udziały jednostek o określonych wariantach tych cech były takie same w próbie i w populacji. O takich próbach często mówi się, że są reprezentatywne ze względu na określony, ograniczony zestaw cech. Wynika z tego, że kategoria reprezentatywności pojawia się także w odniesieniu do nielosowych sposobów wyboru próby.

4. Propozycja zdefiniowania próby reprezentatywnej

W literaturze anglojęzycznej definicje *próby reprezentatywnej* przybierają raczej formę definicji realnej niż nominalnej (zob. np. Babbie, 2004). Definicje te nie objaśniają charakteru takiej próby, sposobu jej wyboru lub struktury, lecz koncentrują się na jej atrybucie praktycznej użyteczności. Na przykład Lavrakas (2008) stwierdza, że próba reprezentatywna daje możliwość wiarygodnej ekstrapolacji wyników na populację, którą reprezentuje⁸. Podobnie próbę reprezentatywną charakteryzują Rudolph i in. (2023), pisząc, że jest to próba, z której uzyskane oszacowania mogą być uogólniane

⁸ Oryg.: „A representative sample is one that ensures external validity in relationship to the population of interest the sample is meant to represent”.

na populację, z której pochodzi⁹. Definicje te słusznie podkreślają najważniejszy walor próby reprezentatywnej – możliwość uogólniania na jej podstawie wyników na reprezentowaną populację – ale nie są one, jak się wydaje, dostatecznie precyzyjne i jednoznaczne.

Przytoczone definicje w niewystarczający sposób charakteryzują samą próbę, ograniczając się jedynie do jej zastosowań w procesie wnioskowania statystycznego. Z tego powodu, odwołując się wyłącznie do tych definicji, trudno byłoby przekonać badacza wykorzystującego na przykład próbę kwotową, że jego próba nie spełnia warunków definicji próby reprezentatywnej. Korzystający z prób kwotowych twierdzą dość często, że posługują się próbami reprezentatywnymi. Tak jednak nie jest, ponieważ samo kontrolowanie zgodności struktury próby ze strukturą populacji jedynie w odniesieniu do wyselekcjonowanych zmiennych (charakterystyk) jest niewystarczające, aby próbę nazwać *reprezentatywną*. Można natomiast powiedzieć, że jest reprezentatywna dla danej populacji ze względu na kilka konkretnych zmiennych, uznanych za ważne dla celu badań, a które w wyborze kwotowym pełnią funkcję charakterystyk kontrolnych. Podnoszony nieraz przez zwolenników próbkowania kwotowego argument, że dla badacza kluczowa jest zgodność właśnie ze względu na zmienne najważniejsze dla celu badań, co ma czynić tę próbę reprezentatywną, jest trudny do obrony. A to dlatego, że ten rodzaj rozumowania – czyli przekonanie o tym, że badacz potrafił bezbłędnie wskazać wszystkie istotne dla celu badania cechy jednostek populacji – był w przeszłości częstą przyczyną poważnych błędów wnioskowania na podstawie prób kwotowych (Szreder, 2010, 2012). Okazywało się mianowicie, że niewystarczająco poznane i niedoceniane inne cechy, których rozkłady w populacji nie są w żaden sposób weryfikowane z rozkładami z próby (ponieważ nie należą do charakterystyk kontrolnych), powodowały niereprezentatywność próby, mimo że była ona reprezentatywna w odniesieniu do kilku wskazanych na wstępie zmiennych. W zdefiniowaniu próby reprezentatywnej trzeba więc to rozróżnienie brać pod uwagę.

Próby nieprobabilistyczne, którym przypisuje się walor reprezentatywności w odniesieniu do kilku określonych przez badacza cech populacji, stanowią w praktyce podstawę uogólnień, o których mowa w przytoczonych wyżej definicjach. Za niedostatek tych definicji trzeba więc uznać dopuszczanie powiązania próby reprezentatywnej z bliżej niesprecyzowanymi sposobami uogólnień na populację. Co prawda w niektórych definicjach akcentuje się „upoważnione, wiarygodne uogólnienia” (ang. *external validity*), ale wydaje się to wciąż za mało jak na precyzję pożądaną w każdej definicji. Przede wszystkim wnioskowanie na podstawie prób nielosowych nie daje możliwości wiarygodnego oszacowania błędów ocen. Nie istnieje bowiem matematyczny

⁹ Oryg.: „We define a study sample to be representative of a well defined target population if the results estimated in that sample are generalisable to the target population”.

model dla próbkowania nieprobabilistycznego, który dawałby szansę na wiarygodne liczbowe określenie błędów wnioskowania. Dlatego ogólna, uniwersalna definicja próby reprezentatywnej powinna uwzględniać także i ten aspekt. Innymi słowy, proponuję, aby definicja próby reprezentatywnej odnosiła się szczegółowo do dwóch aspektów:

- zgodności struktury próby ze strukturą populacji;
- atrybutów wnioskowania statystycznego na podstawie próby.

Oczywiście w takiej definicji nie powinno być mowy o wyróżnionych cechach lub zmiennych, których zgodności rozkładów oczekuje się w próbie i w populacji. Powinna to być definicja na tyle uniwersalna, aby poprawnie wskazywać na właściwości próby niezależnie od tego, które rozkłady zmiennych i w jakim stopniu interesują badacza lub odbiorcę wyników wnioskowania.

Proponuję następującą definicję próby reprezentatywnej:

Próba reprezentatywna to taka, której struktura jest identyczna ze strukturą populacji lub do niej zbliżona oraz która może stanowić podstawę uogólnień wyników na populację i wiarygodnej oceny ich błędów.

Tak definiowana próba reprezentatywna odpowiada z jednej strony dotychczasowemu rozumieniu tej kategorii, a z drugiej nie pozwala na to, by krótkim terminem *próby reprezentatywne* nazywać te próby, których struktury są identyczne ze strukturą populacji w odniesieniu jedynie do niektórych cech. Te ostatnie powinny być nazywane *próbami reprezentatywnymi ze względu na wyszczególnione zmienne*, a nie *reprezentatywnymi* w ogóle (zgodnie z zaproponowaną wyżej definicją). W niczym nie deprecjonuje to takich prób, w tym m.in. kwotowych, a jedynie porządkuje terminy używane w statystyce. Próby reprezentatywne ze względu na konkretne zmienne (cechy) mogą być próbami losowymi i nielosowymi. Popularny w empirycznych zastosowaniach wnioskowania statystycznego mechanizm ważenia i kalibracji wag stanowi dobry przykład dążenia do uzyskania właśnie tego rodzaju „ograniczonej” reprezentatywności prób¹⁰.

Warto przywołać jeszcze jeden termin zbliżony brzmieniem do *próby reprezentatywnej*, a mianowicie wyrażenie *próba reprezentacyjna*. Nie jest on co prawda często stosowany, a jego istnienie wiąże się przede wszystkim z badaniami reprezentacyjnymi, oraz z działem statystyki zwanym *metodą reprezentacyjną*. O ile termin *badanie reprezentacyjne* jest jasno określony w literaturze, o tyle *próba reprezentacyjna* jedynie zastępuje i powiela termin *próba losowa*. Jest więc terminem zbędnym, który może prowadzić do nieporozumień we wnioskowaniu statystycznym. Z samej definicji

¹⁰ Szeroko na temat sposobów ważenia i kalibracji oraz zastosowań tych mechanizmów w statystyce piszą Kozłowski i Szreder (2020), a także Szymkowiak (2019).

badania reprezentacyjnego wynika bowiem, że jest to rodzaj badań niewyczerpujących opartych na próbach losowych. Tak rozumieją to Bracha (1998, s. 18): „Częściowe badanie statystyczne, do którego próba została wybrana zgodnie z pewnym planem losowania, nazywamy badaniem reprezentacyjnym”, Steczkowski (1995, s. 11): „Metoda reprezentacyjna polega na losowym doborze próby (reprezentacji) ze skończonej zbiorowości generalnej, opisie tej próby za pomocą charakterystyk statystycznych, a następnie na uogólnieniu otrzymanych wyników na zbiorowość generalną, z której próba ta pochodzi” czy Rószkiewicz i in. (2013, s. 143): „Wnioskowaniem na podstawie prób losowych, które cechuje reprezentatywność statystyczna, zajmuje się nauka zwana metodą reprezentacyjną, zaś badanie oparte na próbie losowej określane jest mianem badania reprezentacyjnego”. Wynika więc z tego jednoznacznie, że próbą wykorzystywaną w badaniach reprezentacyjnych jest próba losowa. Nie ma potrzeby nazywania jej *próbą reprezentacyjną* ani wprowadzania takiej kategorii do wnioskowania statystycznego.

5. Podsumowanie

Próba reprezentatywna jest jedną z podstawowych kategorii występujących w zastosowaniach wnioskowania statystycznego. W teorii wnioskowania – a ściślej w jego modelu matematycznym – kategoria ta nie jest obecna, ponieważ zastępuje ją kluczowe w probabilistyce pojęcie *próby losowej*. Dlatego potrzeba precyzyjnego zdefiniowania *próby reprezentatywnej* odnosi się nie do teorii, lecz do empirycznych zastosowań wnioskowania statystycznego. Potrzeba ta jest ważniejsza w zastosowaniach ekonomicznych i społecznych, w których cele wnioskowania mają głównie charakter deskryptywny lub confirmacyjny. Mniejsze jest natomiast jej znaczenie w badaniach eksperymentalnych – w naukach ścisłych i przyrodniczych – dla których najważniejsze są cele eksploracyjne. W tych ostatnich rzadko postuluje się planowanie i realizację eksperymentów na próbach reprezentatywnych dla całej populacji.

Jednoznaczne rozumienie pojęcia *próby reprezentatywnej* jest ważne przede wszystkim dlatego, że pozwala na pełną i poprawną interpretację zarówno wyników badania próbkowego, jak i błędów, którymi te wyniki mogą być obciążone. W artykule zaproponowano definicję *próby reprezentatywnej*, która bierze pod uwagę zgodność struktury próby ze strukturą populacji oraz możliwości takiego wnioskowania na podstawie próby, w którym ocenom uzyskanym z próby towarzyszą naukowo uzasadnione oszacowania wielkości błędów. Autor postuluje, aby próby, które wykazują zbieżność rozkładów tylko niektórych cech z ich rozkładami populacyjnymi, nazywać *próbami reprezentatywnymi ze względu na wyszczególnione zmienne*.

Bibliografia

- Analytical Methods Committee. (2016). Representative sampling? Views from a regulator and a measurement scientist. *Analytical Methods*, 8(24), 4783–4784. <https://doi.org/10.1039/c6ay90077a>.
- Babbie, E. (2004). *Badania społeczne w praktyce* (tłum. W. Betkiewicz, M. Bucholc, P. Gadomski, J. Haman, A. Jasiewicz-Betkiewicz, A. Kloskowska-Dudzińska, M. Kowalski, M. Mozga). Wydawnictwo Naukowe PWN.
- Bracha, C. (1996). *Teoretyczne podstawy metody reprezentacyjnej*. Wydawnictwo Naukowe PWN.
- Bracha, C. (1998). *Metoda reprezentacyjna w badaniu opinii publicznej i marketingu*. Efekt.
- Dowlen, O. (2008). *The political potential of sortition. A study of the random selection of citizens for public office*. Imprint Academic.
- Ebrahim, S., Smith, G. D. (2013). Commentary: Should we always deliberately be non-representative?. *International Journal of Epidemiology*, 42(4), 1022–1026. <https://doi.org/10.1093/ije/dyt105>.
- Kozłowski, A., Szreder, M. (2020). *Informacje spoza próby w badaniach statystycznych*. Wydawnictwo Uniwersytetu Gdańskiego.
- Lavrakas, P. J. (red.). (2008). Representative sample. W: *Encyclopedia of Survey Research Methods* (t. 2). SAGE Publications. <https://doi.org/10.4135/9781412963947>.
- Pub Affairs Bruxelles. (2021). *Conference on the Future of Europe: European Citizens' Panels set to start*. <https://www.pubaffairsbruxelles.eu/eu-institution-news/conference-on-the-future-of-europe-european-citizens-panels-set-to-start/>.
- Reichenbach, H. (1951). *The Rise of Scientific Philosophy*. University of California Press. <https://doi.org/10.2307/jj.8501244>.
- Richiardi, L., Pizzi, C., Pearce, N. (2013). Commentary: Representativeness is usually not necessary and often should be avoided. *International Journal of Epidemiology*, 42(4), 1018–1022. <https://doi.org/10.1093/ije/dyt103>.
- Rothman, K. J., Gallacher, J. E. J., Hatch, E. E. (2013). Why representativeness should be avoided. *International Journal of Epidemiology*, 42(4), 1012–1014. <https://doi.org/10.1093/ije/dys223>.
- Rószkiewicz, M., Perek-Białas, J., Węziak-Białowolska, D., Zięba-Pietrzak, A. (2013). *Projektowanie badań społeczno-ekonomicznych. Rekomendacje i praktyka badawcza*. Wydawnictwo Naukowe PWN.
- Rudolph, J. E., Zhong, Y., Duggal, P., Mehta, S. H., Lau, B. (2023). Defining representativeness of study samples in medical and population health research. *BMJ Medicine*, 2(1), 1–7. <https://doi.org/10.1136/bmjmed-2022-000399>.
- Shell, E. R. (2016). Hurdling Obstacles: Meet Marcia McNutt, scientist, administrator, editor, and now National Academy of Sciences president. *Science*, 353(6295), 116–119. <https://doi.org/10.1126/science.353.6295.116>.
- Spiegelhalt, D. J., Grieve, A. P., Lindley, D. V., Copas, J. B., Cairns, A. J. G., Chatfield, C., Box, G., Cox, D., Tukey, J. W., Raftery, A. E., Aitkin, M., Barnard, G. A., Donnelly, P., Ehrenberg, A. S. C., Gelfand, A. E., MaHlick, B. K., Gelman, A., Meng, X.-L., Glasbey, C. A., ..., Zellner, A. (1995). Discussion of the paper by Draper. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57(1), 71–97. <https://doi.org/10.1111/j.2517-6161.1995.tb02016.x>.

- Steczkowski, J. (1995). *Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych*. Wydawnictwo Naukowe PWN.
- Szreder, M. (2010). *Metody i techniki sondażowych badań opinii* (wyd. II zmienione). Polskie Wydawnictwo Ekonomiczne.
- Szreder, M. (2012). Wybór próby badawczej – aspekt etyczny i naukowy. *Marketing i Rynek*, (11), 24–27.
- Szreder, M. (2019). Istotność statystyczna w czasach big data. *Wiadomości Statystyczne. The Polish Statistician*, 64(11), 42–57. <https://doi.org/10.5604/01.3001.0013.7583>.
- Szymkowiak, M. (2019). *Podejście kalibracyjne w badaniach społeczno-ekonomicznych*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. <https://doi.org/10.18559/978-83-66199-89-7>.
- Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science. *The American Statistician*, 73(S1), 246–261. <https://www.tandfonline.com/doi/full/10.1080/00031305.2018.1518264#abstract>.
- Triandis, H. C. (2021, 1 marca). *Macron's 'direct democracy' to be tested as citizens' panel on climate wraps up*. <https://www.france24.com/en/france/20210301-macron-s-direct-democracy-to-be-tested-as-citizens-panel-on-climate-wraps-up>.