

Reprezentatywność i pobieranie próby - zerwania oraz kontynuacje w zakresie konceptualizacji

SŁAWOMIR PASIKOWSKI

Uniwersytet Łódzki

Artykuł poświęcony jest zagadnieniu pobierania prób w badaniach obserwacyjnych oraz reprezentatywności tych prób w stosunku do populacji pod względem charakterystyk będących przedmiotem badań. Celem artykułu jest zaprezentowanie zmienności w zakresie konceptualizacji tych zagadnień mającej konsekwencje również w obszarze badań pedagogicznych. W artykule odwołano się do klasycznych wykładni opublikowanych w literaturze krajowej i zagranicznej oraz zaprezentowano przykłady i symulację obrazujące konceptualne różnice i podobieństwa. W rezultacie zaproponowano wykorzystywanie istniejących lub tworzenie nowych podejść integracyjnych.

SŁOWA KLUCZOWE: badania obserwacyjne, próbkowanie, reprezentatywność.

Representativeness and Sample Sampling – Breaks and Continuations in the Scope of Conceptualization

The article is dedicated to the issue of sampling in observational studies and the representativeness of these samples in relation to the population in terms of the characteristics under study. The aim of the article is to present the variability in the conceptualization of these issues, which also have consequences in the field of educational research. The article refers to classical interpretations published in both domestic and foreign literature and provides examples and simulations illustrating conceptual differences and similarities. As a result, the utilization of existing or the creation of new integrative approaches is proposed.

KEYWORDS: observational research, representativeness, sampling.

Wprowadzenie

W badaniach obserwacyjnych wybór elementów do próby, czyli próbkowanie, jest warunkiem dostępu do źródła danych, ale nie tej oczywistej kwestii dotyczy ta idea. Celem każdego próbkowania jest umożliwienie dotarcia do danych, lecz w sposób pozwalający na uzyskanie reprezentacji generalnego zbioru. W badaniach pedagogicznych elementami tego zbioru mogą być ludzie, zachowania, ich własności lub relacje, a jakość reprezentacji

powstającej na skutek próbkowania określają cele towarzyszące badaniu i wynikające z nich kryteria (por. Lohr, 2010; Steczkowski, 1995). Dwa podstawowe znajdują zastosowanie podczas klasyfikacji metod wyboru elementów z populacji do próby. Pierwsze kryterium to wyczerpywanie elementów. Jest ono podstawą dychotomicznego podziału na dobór wyczerpujący (całościowy) i dobór wyrwykowy (częściowy). Ten pierwszy obejmuje wszystkie elementy zbioru generalnego, a zatem dostarcza zbiorowości o liczności równej liczności tego zbioru. Ten drugi prowadzi do uzyskania zbioru, którego liczność jest mniejsza niż liczność zbioru generalnego. Taki zbiór nosi nazwę zbioru lokalnego. Drugie kryterium to randomizacja. Pozwala ono wyodrębnić dwie postaci doboru – losowy oraz nielosowy. To kryterium dotyczy przede wszystkim doborów wyrwykowych, gdyż określa możliwość odnoszenia wyników obserwacji podzbioru do zbioru generalnego. Próbkowanie beztrendycyjne – takie, w którym każdy element zbioru generalnego, czyli populacji, ma nieograniczone prawdopodobieństwo pobrania do próby – nazywane jest doбором losowym. Polega on na pojedynczym pobieraniu elementu bezpośrednio z całego zbioru generalnego i zwracaniu tego elementu przed pobraniem kolejnego (losowanie nieograniczone indywidualne niezależne). Uzyskana w ten sposób próba nazywana jest próbą losową prostą, z kolei losowanie – losowaniem prostym (w losowaniu złożonym próbkowanie może być ograniczone i przebiegać w schemacie wielostopniowym lub bezzwrotnym). Dobory nielosowe charakteryzuje poważne ograniczenie tego prawdopodobieństwa powodowane działaniami czynników sprzyjających pobieraniu jednych elementów bardziej niż innych. Jeśli czynnikami tymi są wiedza, preferencje, a także motywacje osób organizujących lub uczestniczących w procesie selekcjonowania elementów do próby, to wówczas mowa o doborze arbitralnym. Połączenie idei doboru losowego i doboru nielosowego jest podstawą wyodrębnienia klasy doborów mieszanych.

U źródeł decyzji o podejmowaniu badań wyrwykowych zamiast wyczerpujących mogą leżeć bardzo duża liczność populacji, niszczący wpływ badania na elementy tej zbiorowości, jak też koszty finansowe i czasowe związane z organizacją badania (Steczkowski, 1995, s. 16). Umiejętny sposób organizacji badania pozwala jednak, by zebrane dane o przedmiotowych cechach w próbie umożliwiły trafną charakterystykę tych cech w populacji, tzn. by zbiorowość próbna była reprezentatywna względem zbiorowości generalnej w zakresie charakterystyki obserwowanych cech.

Reprezentacja i reprezentatywność w statystyce

Metody próbkowania i sposoby wnioskowania o charakterystykach zbiorowości generalnych na podstawie zbiorowości częściowych stanowią przedmiot teorii i badań prowadzonych w ramach metody reprezentacyjnej (MR) – działu matematyki statystycznej. Zgodnie z nią próba jest reprezentacją populacji w zakresie obserwowanej cechy lub cech. Oznacza to, że rozkład cechy w próbie pozostaje w relacji do rozkładu cechy w populacji. Rozkład cechy można przedstawić jako ciąg par wartości, z których pierwsze w parach są uporządkowanymi wartościami cechy, a drugie odpowiadają licznościom lub prawdopodobieństwom występowania tych pierwszych: $(x_1, 4)$, $(x_2, 1)$, $(x_3, 3)$, ..., (x_n, n) . Reprezentatywność stanowi własność tej reprezentacji i informuje o różnicy między populacją a pobraną z niej próbą w zakresie rozkładu danej cechy. Reprezentatywność traktowana jako stan braku różnicy, albo dostatecznie małej różnicy (por. np. Greń, 1987; Hellwig, 1995; Steczkowski, 1995), jest punktem odniesienia, do którego aspiruje badacz w procesie konstruowania reprezentacji zbiorowości generalnej. W tym znaczeniu uchodzi za ideał trudno osiągalny w badaniach

wyrywkowych. Jeśli jednak reprezentatywność próby uznać za wielkość losową ciągłą, czyli z określonym prawdopodobieństwem przyjmującą wartości z kontinuum, to im ta wartość wyższa, w tym mniejszym stopniu rozkład cechy w próbie różnił będzie się od rozkładu tej cechy w zbiorowości generalnej. Innymi słowy, im wyższa wartość, tym bardziej relacje zachodzące w próbie będą zbliżone do relacji zachodzących w populacji (por. Hellwig, 1995, s. 191). Takie ujęcie pozwala dostrzec, że o uznaniu próby za reprezentatywną decydują kryteria określające granicę, po przekroczeniu której różnica między rozkładem w próbie a rozkładem w populacji staje się nieistotna. W przypadku próbkowania losowego tymi kryteriami są dopuszczana wielkość różnicy oraz założone prawdopodobieństwo, że charakterystyka rozkładu próbkowego pokryje się z charakterystyką rozkładu populacyjnego. Prawdopodobieństwo traktowane jest przy tym jako nieujemna, unormowana i addytywna miara zdarzeń przyjmująca wartości z przedziału $[0,1]$, gdzie 0 oznacza zdarzenie niemożliwe, a 1 – zdarzenie pewne. Dominuje przy tym częstościowa interpretacja prawdopodobieństwa, zgodnie z którą jest ono częstością względną zdarzeń zachodzących w takich samych warunkach (np. Hellwig, 1995; Szreder, 2010).

Różnica nie przekraczająca dopuszczanej wielkości oraz w granicach przyjętego prawdopodobieństwa wskazuje na wysoką reprezentatywność próby. Jednak mimo dysponowania powyższym kryterium sama możliwość oceny faktycznej wielkości tej różnicy zachodzi raczej *ex post*, gdyż zwykle badania wyrywkowe dotyczą populacji, których rozkłady nie są znane.

Wiadomo jednak, kiedy reprezentatywność prób wzrasta. Przede wszystkim rośnie wraz ze wzrostem jakości ramy próbkowania, która jest rejestrem elementów skończonej populacji. Utworzenie optymalnej ramy próbkowania (*sampling frame*), zwanej też operatem losowania, wymaga wiedzy przynajmniej o liczności zbioru generalnego, na którego elementach może realizować się cecha mająca być docelowym przedmiotem obserwacji.

Reprezentatywność wzrasta też na skutek zastosowania procedur gwarantujących bezstronne pobieranie elementów ze zbiorowości generalnej (Hellwig, 1995; Szreder, 2010). Polega to na zapewnieniu równych prawdopodobieństw wyboru, co znosi faworyzowanie jednych elementów lub prób kosztem innych. Dzięki temu próba powstaje w sposób odzwierciedlający wartości i częstości ich realizacji w populacji, a nie oczekiwania lub wyobrażenia osób prowadzących próbkowanie. Elementy charakteryzowane przez wartości cechy rzadko występujące w populacji będą pojawiały się w próbie w mniejszej liczbie, a elementy o wartościach częściej występujących uzyskają odpowiednio liczniejszą reprezentację. Dostatecznie długie procedowanie próbkowania zapewnia, że w próbie znajdą się reprezentacje wszystkich wartości cechy występujących w populacji.

Reprezentatywność rośnie również wraz ze zwiększaniem się rozmiaru próby, czyli jej liczności (Hellwig, 1995; Szreder, 2010). W przypadku skończonych populacji wzrost liczności próby oznacza coraz większe wyczerpywanie elementów ze zbiorowości generalnej, a zatem coraz dokładniejsze jej odwzorowanie w próbie. W przypadku populacji nieskończonych, jak też nieskończonego pobierania ze zwracaniem z populacji skończonych, odwzorowanie w próbie podlega częstości realizacji określonych wartości w populacji. U podstaw obu tych przypadków leży zbieżność stochastyczna, którą wyraża powszechnie znana formuła:

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$$

Oznacza ona, że przy liczbie realizacji doświadczenia (n) dążącej do nieskończoności (∞), jedności (1) równa się graniczna wartość prawdopodobieństwa ($\lim P$), że bezwzględna różnica

między ciągiem niezależnych wielkości losowych X_1, X_2, \dots, X_n (X_n), z których każda odpowiada pojedynczemu elementowi lokalnego zbioru, a wielkością losową (X) dotyczącą elementów generalnego zbioru jest mniejsza od przyjętej dowolnie małej dodatniej liczby (ε). Innymi słowy, elementy zbiorowości (można je oznaczyć kolejnymi numerami 1, 2, ..., n) charakteryzuje wielkość losowa X , która w przypadku każdego z nich ma swój niezależny rozkład prawdopodobieństwa. Ciąg tych wielkości (każda z nich odpowiada osobnemu elementowi zbiorowości), o którym mówi twierdzenie o zbieżności stochastycznej, lub ciąg ich rozkładów, jest wektorem $X' = [X_1, X_2, \dots, X_n]$. Ten wektor to próba losowa. Wektor stanowi zbiór punktów, co oznacza, że każda z tych wielkości losowych, przyjmując dowolną wartość, wyznacza punkt. Ta wartość, gdy zrealizuje się w doświadczeniu, jest wynikiem obserwacji (x_i). Zbiór wyników obserwacji $[x_1, x_2, \dots, x_n]$, z których każdy stanowi realizację niezależnej wielkości losowej ze zbioru tych wielkości $[X_1, X_2, \dots, X_n]$, jest realizacją próby losowej. To tę realizację zwykle ma się na myśli, pisząc lub mówiąc o losowej próbie. Nieprzestrzeganie tego rozróżnienia na próbę losową i jej realizację jest źródłem nieporozumień, także w interpretacji twierdzenia o zbieżności stochastycznej, tym samym utrudnia zrozumienie pojęcia reprezentatywności próby. Próba staje się tym bardziej reprezentatywna, im bardziej maleje różnica między nią a populacją w zakresie rozkładu cechy X , a co następuje wraz ze wzrostem liczności realizacji próby losowej.

Jeśli rozkład w próbie jest podobny do rozkładu w populacji, to również podobne są parametry tych rozkładów, będące wielkościami liczbowymi, jak wartość oczekiwana, wartość wariancji, skośności i spłaszczenia. Precyzyjnie ujmuje to twierdzenie Chinczyna, zgodnie z którym wspomniany wyżej ciąg niezależnych wielkości losowych o jednakowych rozkładach i skończonych wartościach oczekiwanych ($E(X_i)$) oraz wspólnej wartości przeciętnej μ ma tę własność, że średnia arytmetyczna utworzona z tych wielkości losowych jest stochastycznie zbieżna do tej wspólnej dla nich wartości μ . Własność ta wyrażona jest formułą:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \varepsilon \right) = 1$$

W ocenie reprezentatywności próby można więc odwoływać się do parametrów rozkładów, gdyż – tak jak rozkłady – parametry te podlegają zbieżności tym większej, im liczniejsza jest próba. Parametr rozkładu próby jest estymatorem (Z_n) parametru rozkładu populacji (Q) i charakteryzuje się trzema podstawowymi własnościami. Po pierwsze, estymator powinien być nieobciążony, czyli mieć wartość równą wartości parametru. Po drugie, powinien być zgodny, czyli różnica między wartością estymatora a wartością parametru powinna dążyć do zera. Po trzecie, estymator powinien być efektywny, a taki wśród innych estymatorów jest ten o najmniejszej wariancji. Własności estymatora łatwiej objaśniać, gdy wprowadzi się rozróżnienie na rozkład w próbie i rozkład z próby (por. np. Kot, Jakubowski i Sokołowski, 2011). W pierwszym przypadku mowa o rozkładzie cechy i jego parametrach uzyskanych w pojedynczej próbie. W drugim – o rozkładzie powstałym z więcej niż jednej próby. Może to być np. rozkład średnich arytmetycznych, sum albo wskaźników struktury porachowanych dla każdego z rozkładów uzyskanych w pojedynczych próbach. Również taki rozkład można scharakteryzować parametrycznie, czyli wyznaczyć odpowiadającą mu wartość oczekiwaną oraz wariancję. Jeśli takich rozkładów z próby będzie więcej niż jeden, to możliwe stanie się porównywanie ich parametrów. Estymatorem efektywnym będzie ten o wariancji najmniejszej. W związku z powyższym próba o zmaksymalizowanym stopniu reprezentatywności

charakteryzuje się rozkładem o parametrach, które są nieobciążonymi, zgodnymi i efektywnymi estymatorami parametrów rozkładu danej wielkości w populacji.

Warto zauważyć, że w teorii estymacji z własności estymatorów oraz twierdzenia Chinczy-na wyprowadza się konsekwencję w postaci rozmiaru próby niezbędnej do oszacowania nieznanych parametrów rozkładu zbiorowości generalnej.

Różnica między próbą a populacją w zakresie rozkładu cechy i jego parametrów jest interpretowana jako błąd szacunku, którego pochodzenie może mieć dwa główne źródła: losowe, systematyczne. Najkrócej rzecz ujmując, błąd losowy (*random error, random variability*) jest kwadratem wartości oczekiwanej różnicy estymatora i jego wartości oczekiwanej [$RVar(Z_n) = E(Z_n - E(Z_n))^2$], a błąd systematyczny (*bias*) – kwadratem różnicy wartości oczekiwanej i wartości parametru [$B(Z_n) = (E(Z_n) - Q)^2$]. Suma obu tych błędów zwraca wartość całkowitego błędu szacunku, zwanego też średnim błędem kwadratowym (*mean squared error*) albo średnim odchyleniem kwadratowym (*mean squared deviation*) [$MSE(Z_n) = E(Z_n - Q)^2$]. Błąd losowy nie daje się wyeliminować, gdyż istnieją czynniki, których zachodzenie trudno przewidzieć przed ich realizacją. W zasadzie można go kontrolować jedynie przez zwiększanie liczności próby. Błąd systematyczny w założeniu jest redukowalny do zera, choć z uwagi na liczne jego uwarunkowania nie zawsze do tego dochodzi. Te uwarunkowania mogą być statystyczne, jak np. zbyt mała założona liczność próby, nieodpowiednia metoda próbkowania lub posługiwanie się obciążonym estymatorem. Mogą też być to uwarunkowania pozastatystyczne, jak wadliwość instrumentów i procedur pomiarowych oraz braki danych i ubytki/nadmiary pokrycia, czyli odtwarzania w próbie wartości cechy i ich liczności realizujących się w populacji. Błąd systematyczny spowodowany tym ostatnim uwarunkowaniem nazywany jest błędem kompletności (*non-observation error*) i wyraźnie ogranicza reprezentatywność próby (Steczkowski, 1995). Z kolei, na błąd losowy estymacji parametru, a tym samym na reprezentatywność próby można oddziaływać na dwa sposoby (por. np. Sobczyk, 2007; Szreder, 2010). Można przyjąć odpowiednią wartość współczynnika ufności CLC (*confidence level coefficient*), który informuje o prawdopodobieństwie, z jakim nieznaną wartość parametru jest pokrywana przez przedział ufności zbudowany z uwzględnieniem wartości estymatora. Ten współczynnik wyraża się równaniem $CLC = 1 - \alpha$, gdzie α oznacza prawdopodobieństwo zdarzenia przeciwnego. Można też przyjąć odpowiednią wartość niezbędnej wielkości próby stosownie do przewidywanego błędu szacunku oraz pożądanego poziomu ufności.

Reprezentatywność statystyczna czy typologiczna?

Przedstawienie zagadnienia reprezentatywności statystycznej ułatwia opisywanie i zrozumienie idei reprezentatywności typologicznej, głównie dlatego, że ta druga ma zakorzenienie w pierwszej. Reprezentatywność typologiczna jest identyfikowana w następujący sposób:

- rekonstrukcja możliwie najszerszego zakresu danych pochodzących z różnych kontekstów i sytuacji, odnoszących się do tworzonych kategorii (por. Glaser i Strauss, 2009, s. 53);
- odtwarzanie wszystkich wartości danej cechy lub typów będących efektem kombinatorycznego skrzyżowania cech tak, by w próbie znaleźli się reprezentanci każdej z tych wartości lub każdego z typów bez względu na odpowiadające im liczności w zbiorowości generalnej (Nowak, 2007, s. 300);
- odnoszona raczej do kategorii niż obserwacji własność pozwalająca ocenić, czy badania i ich wyniki obejmują istotne z teoretycznego punktu widzenia aspekty zagadnienia (Flick, 2009, s. 394, 473).

Innymi słowy, reprezentatywność typologiczna polega na odwzorowaniu wartości cechy X i wariacji, jakie one tworzą, bez ścisłego ustalania prawdopodobieństw lub liczności realizacji tych wariacji i wartości. W zależności od celu badania pobieranie próby kierowane reprezentatywnością typologiczną może być nastawione na odzwierciedlenie wszystkich albo wybranych aspektów zjawiska (Glaser i Strauss, 2009, s. 59). Osiąganie stanu reprezentatywności typologicznej wymaga rejestrowania informacji o wartościach lub wariantach cechy. W procesie rejestrowania takich informacji naturalnie występują dane dotyczące liczności obiektów, w przypadku których cecha przyjmuje określoną wartość. Tym samym podczas próbkowania kierowanego reprezentatywnością typologiczną uwidocznia się przynajmniej to, które wartości albo kategorie występują częściej, a które rzadziej. W związku z tym zliczanie obserwacji nie wymaga komplikowania procedury pobierania. Ponadto warto mieć na uwadze, że kategorie nie są swobodnie dobieranymi nazwami. Ich wybór jest wtórny względem zgromadzonych danych, jak również podporządkowany optymalizowaniu relacji zakresu i treści nazwy. Łączenie tych desygnatów w zbiory stanowiące zakres kategorii polega na identyfikowaniu cech, które są przez te desygnaty podzielane i które to cechy składają się na treść nazwy. Co więcej, tworzenie kategorii jest procesem nie tylko jakościowym, ale też ilościowym, bo polega na eksponowaniu kategorii zależnie od liczności ich desygnatów w zgromadzonym materiale empirycznym. Kategorie mogą być rozwijane wraz z pojawianiem się danych wskazujących na obecność nowych własności. Te własności dostarczają podstawy do nowych różnicowań i grupowań zbioru desygnatów, a zatem powoływania nowych kategorii będących reprezentacjami wartości i wariantów badanego zjawiska lub ich wariacji. Jeśli ponadto rejestrowane są liczności lub prawdopodobieństwa występowania tych wartości, możliwe staje się wyznaczenie rozkładu cechy.

Oczywiście różnica w znaczeniu nazw „reprezentatywność typologiczna” i „reprezentatywność statystyczna” jest wyraźna. Pierwsza dotyczy populacji wartości danej cechy, a druga – populacji elementów charakteryzowanych przez wartości tej cechy. W metodzie reprezentacyjnej zabiegi mające na celu odwzorowanie populacji wartości obserwowanej cechy służą minimalizowaniu wspomnianego wyżej błędu pokrycia. Temu samemu celowi służy próbkowanie zgodnie ze strategią doboru teoretycznego. Nie jest więc tak, że metoda reprezentacyjna z założenia prowadzi do uzyskiwania jakościowo innej próby niż dobór kierowany stanem nasycenia teoretycznego kategorii. Poza tym w obu przypadkach zbiorowość generalna może być nieznana lub ukryta. Różnica między tymi dwoma podejściami dotyczy rodzaju, modelu i procedury. Pod tym względem jest zasadnicza. Jednak jakość pozyskiwanej próby może być bardzo podobna w zakresie reprezentatywności względem populacji wartości obserwowanej cechy. Dobór teoretyczny będący strategią weryfikacyjną charakteryzującą specyficzne działania takie, jak stałe porównywanie zbiorów, maksymalizowanie i minimalizowanie różnic i podobieństw, dążenie do replikacji oraz tworzenie struktury kategorii dzielącej zbiór danych na podzbiory. Dlatego porównywanie doboru do próby z doбором teoretycznym może być rozczarowujące, ale porównania własności prób już niekoniecznie. Otóż wbrew argumentacji dotyczącej różnic pod względem tzw. adekwatności prób (Glaser i Strauss, 2009, s. 54–55) próbkowanie kierowane zasadą reprezentatywności statystycznej jest nastawione zarazem na to, co mieści się w pojęciu nasycenia, czyli reprezentacją populacji wartości cechy. Innymi słowy, pokryciem tej populacji.

Warto zauważyć, że oryginalna koncepcja próbkowania teoretycznego dopuszcza łączenie go z doбором losowym (Glaser i Strauss, 2009, s. 54). Uzasadnienie tego łączenia odnaleźć można m.in. w sposobie myślenia o efektywności doboru teoretycznego: „im więcej jest

badań opartych na teoretycznym pobieraniu próbek, tym bardziej efektywne powinno stać się teoretyczne pobieranie próbek i analiza porównawcza w przyszłości” (Glaser i Strauss, 2009, s. 64). Sens tej wypowiedzi bliski jest temu, który wyraża przywołane wyżej twierdzenie o zbieżności stochastycznej.

Wybrane metody nielosowego próbkowania

Wśród nielosowych metod doboru próby popularnych w badaniach społecznych wyróżnić można kilka postaci, z których większość odbierana jest obecnie inaczej niż przewidują to oryginalne koncepcje leżące u ich podstaw, a w związku z tym niekiedy ze szkodą dla celu, któremu te koncepcje miały służyć. Otóż każda z nich podporządkowana została idei podnoszenia reprezentatywności próby pozyskiwanej z trudnodostępnej i zwykle wielowymiarowej populacji, której elementami są poszczególni ludzie albo ich zespoły.

Dobór celowy (*purposive sampling*) jest świadomym i intencjonalnym wyborem konkretnych elementów zbiorowości generalnej na podstawie wiedzy o tej zbiorowości, jej elementach, wstępnego rozpoznania problemu lub konkretnych celów badawczych (por. Lissowski, 1968; Steczkowski, 1995; Szreder, 2010). Motywowany może być potrzebą organizacji próby reprezentatywnej statystycznie (na bazie wiedzy o populacji oraz rozłożeniu i dostępności jej elementów) lub reprezentatywnej typologicznie (na podstawie wiedzy o zjawisku, jego wariantach i reprezentacji w populacji). Nie wystarczy określić docelową zbiorowość generalną, aby uznać próbkowanie za celowe. Definiowanie populacji jest bowiem jedną z pierwszych czynności wykonywanych w procesie doboru próby niezależnie od przyjętej metody próbkowania. Jeśli elementy populacji są pobierane do próby jedynie na podstawie ich przynależenia do kategorii wyznaczonej definicją, to w praktyce dobór taki polega na bieżącej dostępności elementów zbiorowości generalnej, co wiąże go z dużym błędem systematycznym i brakiem kontroli reprezentatywności. Taki dobór nosi nazwę akcydentalnego lub wygodnego (*convenient sampling*). Dobór celowy może służyć podnoszeniu reprezentatywności prób, np. gdy istnieje możliwość podziału populacji na warstwy wyznaczone wartościami cechy różnicującej, która to jest skorelowana z cechą będącą przedmiotem badania albo sama ją stanowi. Do cech stratyfikacyjnych zaliczają się płeć, metrykalność oraz wykształcenie. Zatem zachodzi potrzeba znajomości struktury zbiorowości wyznaczonej cechami stratyfikacyjnymi i liczności podzbiorów wyznaczanych wartościami tych cech. W rezultacie można określić proporcję liczności poszczególnych podzbiorów, czyli stosunek tej liczności do liczności zbioru generalnego. To pomaga w kontrolowaniu reprezentatywności próby. Taka postać doboru celowego nazywana jest doбором celowym proporcjonalnym (*proportionate stratified sampling*) albo doбором kwotowym (*quota sampling*).

Dobór śniegowy (*pure snowball sampling*) przewidziany został dla badań prowadzonych z udziałem ukrytych i trudno dostępnych populacji, w których to badaniach znaczenie odgrywa struktura społeczna (Coleman, 1958/1959). Jego koncepcja odwołuje się do idei łańcucha relacji społecznych stanowiącego w tej metodzie punkt wyjścia pozyskiwania kolejnych respondentów na podstawie rekomendacji udzielanych przez poprzednich respondentów (Erickson, 1979). Procedowanie tego doboru jest powszechnie kojarzone ze swobodnym próbkowaniem początkowym, a nawet scedowaniem na respondentów wykonania zadania pozyskiwania danych. W istocie oryginalny model doboru metodą kuli śnieżnej przewiduje wstępną identyfikację środowiska pozostającego w zasięgu badacza oraz losowy dobór próby

pierwotnej z tego środowiska (Frank i Snijders, 1994; Goodman, 1961). Następne elementy pobierane są na bazie rekomendacji pochodzących z próby pierwotnej i odbywa się to w tzw. falach pierwszego oraz kolejnych rzędów. Niedostatki doboru śniegowego doprowadziły do powstania doboru celowanego oraz doboru kluczowych informatorów. Pod koniec XX wieku opracowana została zaawansowana postać próbkowania oparta na modelu czystego doboru śniegowego, ale wyposażona w instrumenty poprawy dopasowania próby do zbiorowości generalnej, która nosi nazwę *Respondent-Driven Sampling* (Heckathorn, 1997).

Dobór celowany (*targeted sampling*), którego koncepcja stanowi odpowiedź na niedostatki doboru metodą kuli śnieżnej, przewiduje kilkufazowe postępowanie ukierunkowane na podniesienie reprezentatywności próby pochodzącej z ukrytej lub trudno dostępnej populacji (Watters i Biernacki, 1989). Pierwsze fazy postępowania polegają na podjęciu działań umożliwiających zlokalizowanie i zidentyfikowanie docelowej zbiorowości generalnej. W ten sposób określone zostają kryteria ukierunkowujące pobieranie elementów ze zbiorowości generalnej. Działania te polegają na pracy w terenie obejmującej tworzenie geograficznej i etnograficznej mapy populacji. W rezultacie powstaje wstępna rama próbkowania, której jakość uzależniona jest od nakładu pracy w terenie oraz która ulega doskonaleniu na skutek sekwencyjnie prowadzonych próbkowań. Po zmapowaniu populacji dokonywane są próbkowania z określonych miejsc w określonym czasie oraz w ustalonej kolejności, dzięki czemu w zbiorze pobranych elementów reprezentację mają szansę uzyskać różne grupy, warstwy i kategorie społeczne.

Dobór kluczowych informatorów (*key informant sampling*) przewidziany został dla ukrytych i trudno dostępnych populacji. Jest metodą polegającą na identyfikacji osób mających wiedzę o populacji docelowej, które potrafią scharakteryzować wzorce zachowań występujące w tej populacji oraz – ewentualnie – jej reprezentantów wykazujących takie wzorce (Deaux i Callaghan, 1985; Heckathorn, 1997). Materiał empiryczny pozyskiwany jest jednak przede wszystkim od kluczowych informatorów, co wiąże się z ograniczeniami jakości i ilości tego materiału. Ta arbitralna postać doboru nazywana jest też metodą ekspertów (*expert sampling*), choć jako taka bywa odróżniana od doboru kluczowych informatorów (Patton, 2018) m.in. z uwagi na elementy procedury, jak np. tworzenie panelu ekspertów z dobranych informatorów, którzy współpracują z badaczem nad tworzeniem reprezentacji badanego zjawiska.

Dobór teoretyczny (*theoretical sampling*) jest powszechnie uznawany za dobór do próby kierowany ideą nasycenia teoretycznego generowanych kategorii. Jednak nie służy on gromadzeniu surowego materiału empirycznego, choć to stanowi warunek jego uruchomienia, lecz budowaniu teorii przez dopasowywanie wyłaniającego się obrazu zjawiska do danych empirycznych w procesie ich jednoczesnego gromadzenia i analizowania (Charmaz, 2009; Creswell, 2007; Glaser i Strauss, 2009; Konecki, 2000). Jest więc strategią weryfikacyjną, a nie metodą próbkowania (por. Morse, Barrett, Mayan, Olson i Spiers, 2002), która pozwala rekonstruować daną wielkość. Jaśniejszy staje się charakter tego doboru, gdy zauważyć, że obejmuje on także początkowy dobór próby (por. Glaser i Strauss, 2009, s. 41–42). Ten początkowy dobór dostarcza materiału, z którego wyprowadzane są wstępne kategorie teoretyczne. Pobieranie kolejnych elementów zbiorowości ma już charakter teoretyczny, gdyż kierowane jest wskazaniem teorii wyłaniającej się z dotychczas zgromadzonych danych. W kolejnych pobraniach do już istniejących kryteriów dokładane są następne, które filtrują proces dostawiania się elementów zbiorowości do powstającej próby (por. Glaser i Strauss, 2009; Konecki, 2000; Ritchie, Lewis i Elam, 2003). Decyzja o zakończeniu pobierania podejmowana jest w chwili uznania, że nowe dane replikują materiał pozyskany

dotychczas i kategorie, które zostały utworzone na jego podstawie. Strategia weryfikacyjna „dobór teoretyczny”, o ile prowadzona jest zgodnie z przypisanym jej rygiorem, służy więc podnoszeniu reprezentatywności próby w zakresie cech poddawanych obserwacji i rozwijającej się teorii zjawiska konstytuowanego przez te cechy. Przypomina procedurę, na której opiera się dobór celowany, choć z tą różnicą, że polega na bieżącym obrazie i nie wiąże się z potrzebą tworzenia ramy próbkowania. Warto zauważyć, że opublikowanie założeń doboru teoretycznego przez Barneya Glasera i Anselma Straussa poprzedzone zostało zaprezentowaniem przez Abrahama Walda w latach 40. XX wieku metody wnioskowania statystycznego, którą autor nazwał analizą sekwencyjną (Wald, 1947–2013). W metodzie tej wyniki analizy danych pozyskiwanych w bieżącym próbkowaniu decydują o pobieraniu kolejnego elementu do próby. Sekwencyjne próbkowanie (*sequential sampling*) opiera się na losowaniu indywidualnym. Prowadzone jest do czasu osiągnięcia np. założonej z góry wartości błędu standardowego, założonej częstości względnej albo uzyskania wartości statystyki spoza granic określonych przedziałem wyznaczonym na podstawie wartości I i II błędu wnioskowania statystycznego oraz wartości tej statystyki przyjęte hipotezami zerową i alternatywną. Może to nasuwać skojarzenia z nasycaniem kategorii (por. Pasikowski, 2019). Analiza sekwencyjna pozwala na zmniejszenie liczności próby potrzebnej do wnioskowania o rozkładach cech zbiorowości generalnych oraz zwiększenia kontroli nad przebiegiem postępowania badawczego, co w istocie należy rozumieć jako dążenie do maksymalizowania trafności wewnętrznej i zewnętrznej badania. Jeśli badanie jest trafne, to znaczy, że uprawnione jest formułowanie sądów o rozkładzie cechy w populacji na podstawie obserwacji rozkładu cechy w próbie lub z próby. Wypada zauważyć, że pojęcie trafności jest obecne także w teorii metod jakościowych i występuje pod nazwą ufności (*trustworthiness*). Charakteryzuje takie aspekty wyniku obserwacji jak realna wartość, generalizowalność, porównywalność i rzetelność (Creswell, 2007; Guba, 1981).

Podejście integracyjne

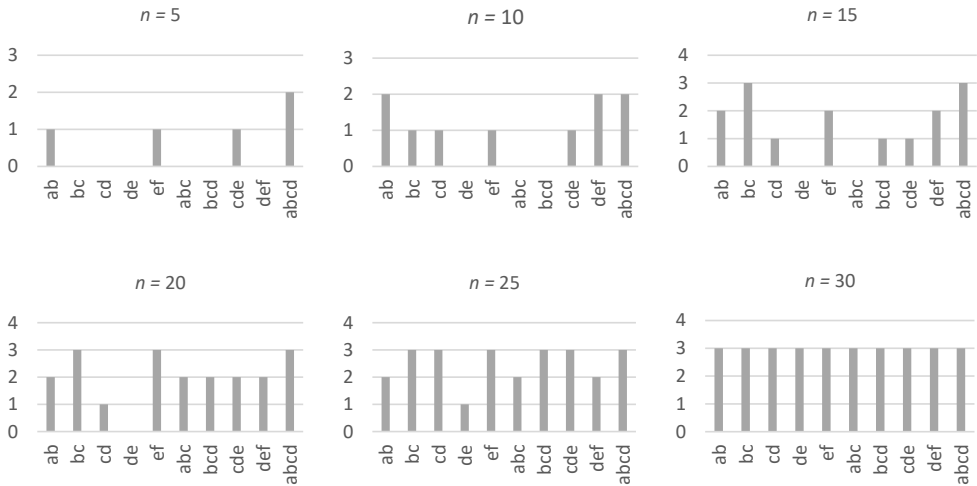
Przegląd wybranych metod nielosowego próbkowania wskazuje, że idea podnoszenia reprezentatywności próby przez łączenie metod doboru nielosowego z metodą doboru losowego nie jest nowa i leży u podstaw przynajmniej kilku modeli doboru próby, które obecnie identyfikowane są z badaniami wolnymi od statystyki. Odcinanie korzeni tych modeli może działać zubożająco na rezultaty ich zastosowania. Wykażę to na przykładzie pobierania danych kierowanego teorią, jak też na podstawie porównania pokrycia populacji wartości cechy tą metodą oraz metodą odwołująca się do idei zbieżności stochastycznej i beztrendyjnej selekcji elementów zbioru.

Im większa wiedza o populacji oraz jej elementach, tym większa pewność, że próba pobrana z tej populacji będzie charakteryzowała się wysoką reprezentatywnością. Jeśli jednak próbkowanie kierowane jest wyłącznie informacjami uzyskiwanymi w toku jego procedowania, to nie wystarczy poleganie na tym, co przynoszą bieżące dane, bo te są silnie zdeterminowane warunkami i własnościami tworzenia się i utrzymywania sieci społecznych oraz zlokalizowaniem rekrutera próby w przestrzeni geograficznej i społecznej. Te okoliczności mogą nie być brane pod uwagę albo nie być dostępne rozpoznaniu w bieżącym postępowaniu badawczym, a mimo to wyznaczać granice teoretycznego nasycenia kategorii i uniemożliwiać dostęp do kolejnych wariacji zjawiska (por. Pasikowski, 2015). Reprezentatywność próby i zgromadzonego materiału ograniczona będzie zatem do bieżącej

zbiorowości próbnej, bez możliwości uzasadnionego przenoszenia wniosków z obserwacji na inne obiekty lub konteksty. To ograniczenie można znieść na przynajmniej dwa sposoby. Pierwszy polega na replikacji próbkowania prowadzonego w sposób kierowany teorią, co skutkuje powiększeniem zbiorowości próbnej. Jednak prowadzenie badań replikacyjnych w pedagogice i innych naukach społecznych nie jest popularne. Możliwe, że przyczynia się do tego konieczność postępowania zgodnie z celem i warunkami określonymi w pierwszym badaniu. Drugi sposób polega na dopasowaniu metody próbkowania do rodzaju populacji oraz przewidywanego zakresu wnioskowania. Jeśli populacja obiektów lub wartości jest trudno dostępna albo niemożliwa do rozpoznania, to można zastosować którąś z postaci przywołanych wyżej metod próbkowania i przede wszystkim rozważyć czas, lokalizację i liczbę punktów startowych tak, by minimalizować ryzyko ograniczeń w dostępie do istotnych wariantów cechy i typów zjawiska. Optymalizowanie doboru sposobu próbkowania polega również na łączeniu metod. Już zdobycie informacji o zlokalizowaniu populacji, a na pewno jej fasadowych charakterystykach, umożliwi przygotowanie ramy próbkowania stanowiącej podstawę randomizowanego próbkowania zespołowego. Zmniejsza to tendencyjność rekrutowania początkowej próby. Na podstawie wyników analizy materiału pozyskanego z tej próby kolejne fale próbkowania mogą być już kierowane teorią albo polegać na wykorzystywaniu teorii do tworzenia ramy próbkowania umożliwiającej przeprowadzenie doboru w sposób wielostopniowy lub warstwowy. Można też procedować dobór celowy proporcjonalny w wariancie powtarzanych próbkowań. Na skutek powiększania liczności próby wzrasta reprezentatywność, choć z uwagi na niespełnienie warunku beztendencji i związane z tym duże ryzyko powstawania błędów systematycznych związek liczności próby celowej z jej reprezentatywnością nie jest liniowy. Poza tym powtarzane próbkowanie celowe proporcjonalne dostarcza danych umożliwiających porównywanie rozkładów obserwowanych cech, charakteryzowanie parametrów tych rozkładów, obserwowanie rozkładu z próby, a nie tylko w próbie, i na tej podstawie szacowanie reprezentatywności próby powstałej z prób. Szczególne znaczenie ma to w przypadku małych zbiorowości próbnych, które pobierane metodą doboru celowego mogą charakteryzować się wyższą reprezentatywnością niż małe zbiorowości próbne złożone z elementów pobieranych losowo (Hoeven, Janssen, Roes i Koffijberg, 2015).

Jeśli chodzi o pokrycie populacji, to jego efektywność zależy od znajomości zbiorowości generalnej, dostępności jej elementów oraz zastosowanych metod próbkowania. W przypadku nieznanych lub ukrytych populacji próbkowanie kierowane wyłącznie teorią i regulowane odwoływaniem się do nasycenia teoretycznego kategorii obciążone jest dużym błędem systematycznym wynikającym m.in. z ekologicznego usytuowania rekrutera. W takiej sytuacji zasięg działań może ograniczać dotarcie do nowych danych i w związku z tym zdecydować o zaprzestaniu dalszego próbkowania z powodu uznania, że kategorie teoretyczne uległy nasyceniu.

Tę kwestię można zobrazować za pomocą prostej symulacji. Przewiduje ona pobieranie kolejno pięcioelementowych prób z nieznannej populacji, na której opisana została cecha X . Każdy element pobierany jest bez zwracania i z przyjętym założeniem, że podczas próbkowania rekruter nie zna liczności populacji ani zakresu wartości cechy X . Rysunek 1 przedstawia odwzorowanie wartości tej cechy postępujące ze wzrostem liczności próby. Oś Y widocznych na rysunku histogramów opisuje liczbę elementów (n), a oś X – wartości cechy.



Rys. 1: Rozkład cechy w zależności od wielkości próby

Źródło: opracowanie własne.

W projekcie symulacji uwzględniono zbiorowość generalną o liczności $N = 30$, którą charakteryzuje cecha X przyjmująca 10 wartości będących konfiguracjami czterech podstawowych własności (a, b, c, d). Nieznane rekruterowi prawdopodobieństwo wyboru pojedynczego elementu spośród trzydziestoelementowej populacji wynosiło w pierwszym kroku $p(x_i) = 0,033$ i stanowiło iloraz $1/30$, a w każdym kolejnym rosło proporcjonalnie do liczby elementów pobranych w sposób opisany wzorem $p(x_i) = 1/(N - n)$, gdzie N oznacza licznosc populacji, a n liczbę pobranych elementów. Wyczerpanie zbiorowości generalnej w próbkowaniu ($n = 30$) wskazuje, że każda z wartości cechy realizuje się w trzech obserwacjach na trzydzieści. Innymi słowy, że prawdopodobieństwo *a posteriori* każdej z wartości cechy wynosi $p(x_i) = 0,1$.

Zaprezentowana symulacja pokazuje odtwarzanie w próbie rozkładu cechy w populacji. Jednak zwykle rozkład nie jest znany. Skoro tak, to nie są znane nie tylko licznosci, ale i wartości, jakie cecha przyjmuje. Dopiero próbkowanie oraz analiza danych przybliżają obraz tego rozkładu. Przy zaistnieniu okolicznosci skłaniających do uznania, że zachodzi nasycenie kategorii, próbkowanie może zakończyć się zbyt wcześnie, by uzyskać dostateczne odwzorowanie wartości cechy, co daje fałszywy obraz zarówno populacji wartości, jak i struktury populacji elementów zbioru generalnego. Symulacja ukazuje również kłopotliwość podziału reprezentatywności na typologiczną i statystyczną. Próba statystycznie reprezentatywna jest zarazem reprezentatywna typologicznie, ale nie odwrotnie. Wartości cechy, ich wariacje oraz licznosci odtwarzają się w próbie jednocześnie. W związku z tym należałoby dobierać sposoby, które zapewnią podobną efektywność odtwarzania każdej z nich zamiast stosować procedury służące sztuczemu oddzielaniu jednych od drugich. Uzasadnienie takiego dobierania procedur znaleźć można również w tym, że wiedza o częstości występowania każdej z wartości cechy rozszerza i pogłębia wgląd w badane zjawisko. Umożliwia również optymalizowanie struktury kategorii wyłaniających się z danych, która – jak zauważono wyżej – uzależniona jest od obserwowanej częstości danych.

Zakończenie

Optymalizowanie doboru procedur próbkowania rzadko wymaga wytwarzania nowych rozwiązań. Z jednej strony, przywołane wyżej i opracowane już kilka dziesięcioleci wcześniej modele próbkowania wychodzą naprzeciw potrzebom pobierania próby z nieznanymi zbiorowości generalnych w sposób, który minimalizował będzie jej niedostatki w zakresie reprezentatywności. Z drugiej strony, możliwość pozyskiwania informacji o populacji z innych źródeł niż próba dostarcza wiedzy *a priori*, która może wzbogacać wiedzę *a posteriori*, czyli tę pochodzącą z obserwacji próby. W związku z tym kurczone trzymanie się metody próbkowania opartej na mechanizmie losowania prostego więcej ma wspólnego z nieuzasadnionym puryzmem metodologicznym niż troską o poznanie zbiorowości generalnej (por. Szreder, 2010). Celem próbkowania losowego jest uzyskiwanie próby umożliwiającej to poznanie w sposób zapewniający kontrolę nad ryzykiem błędów oceny i wnioskowania dotyczących rozkładu cechy w populacji, zwłaszcza w sytuacjach, gdy ten rozkład nie jest znany. Jednak jeśli próbkowanie poprzedza wiedza o innych cechach populacji powiązanych z przedmiotem badania, to możliwe staje się wybranie rozwiązań korzystniejszych z punktu widzenia poznania docelowej charakterystyki populacji. Wykorzystując taką wiedzę próbkowanie w złożonym schemacie losowania, np. warstwowe lub systematyczne, a nawet próbkowanie kwotowe, może dostarczać prób stanowiących lepsze reprezentacje niż te pochodzące wyłącznie z próbkowań opartych na mechanizmie losowania prostego. Przy obecnych możliwościach pozyskiwania i rozpowszechniania informacji mieszane metody pobierania prób nierzadko mogą dawać lepsze rezultaty niż próbkowanie losowe proste oraz to kierujące się wyłącznie ideą nasycenia teoretycznego.

Bibliografia

- Coleman, J.S. (1958/1959). Relational Analysis: The Study of Social Organizations with Survey Methods. *Human Organization*, 17(4), 28–36.
- Creswell, J.W. (2007). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. London: SAGE Publications.
- Deaux, E., Callaghan, J.W. (1985). Key Informant Versus Self-Report Estimates of Health Behaviour. *Evaluation Review*, 9(3), 365–368.
- Erickson, B.H. (1979). Some Problems of Inference from Chain Data. *Sociological Methodology*, 10, 276–303. [DOI:https://doi.org/10.2307/270774]
- Flick, U. (2009). *An Introduction to Qualitative Research*. London: SAGE Publications.
- Frank, O., Snijders, T. (1994). Estimating the Size of Hidden Populations Using Snowball Sampling. *Journal of Official Statistics*, 10(1), 53–67.
- Glaser, B.G., Strauss, A.L. (2009). *Odkrywanie teorii ugruntowanej. Strategie badania jakościowego*. Kraków: Zakład Wydawniczy NOMOS.
- Goodman, L.A. (1961). Snowball Sampling. *The Annals of Mathematical Statistics*, 32, 148–170. [DOI:10.1214/aoms/1177705148]
- Greń, J. (1987). *Statystyka matematyczna. Podręcznik programowany*. Warszawa: Wydawnictwo Naukowe PWN.
- Guba, E.G. (1981). Criteria for Assessing the Trustworthiness of Naturalistic Inquiries. *Educational Communication and Technology*, 29(2), 75–91.
- Heckathorn, D.D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, 44(2), 174–199.
- Hellwig, Z. (1995). *Elementy rachunku prawdopodobieństwa i statystyki matematycznej*. Warszawa: Wydawnictwo Naukowe PWN.

- Hoeven, L.R., Janssen, M.P., Roes, K.C., Koffijberg, H. (2015). Aiming for a Representative Sample: Simulating Random Versus Purposive Strategies for Hospital Selection. *BMC Medical Research Methodology*, 15(90), 1–9. [DOI:10.1186/s12874-015-0089-8]
- Kot, S.M., Jakubowski, J., Sokołowski, A. (2011). *Statystyka*. Warszawa: Wydawnictwo Difin.
- Lissowski, G. (1968). Z zagadnień doboru próby. W: K. Szaniawski (red.), *Metody statystyczne w socjologii. Wybrane zagadnienia* (s. 11–69). Warszawa: Wydawnictwo Naukowe PWN.
- Lohr, S.L. (2010). *Sampling: Design and Analysis*. Boston: Brooks/Cole, Cengage Learning.
- Morse, J.M., Barrett, M., Mayan, M., Olson, K., Spiers, J. (2002). Verification Strategies for Establishing Reliability and Validity in Qualitative Research. *International Journal of Qualitative Methods*, 1(2), 13–22. [DOI:10.1177/16094069020010]
- Pasikowski, S. (2015). Czy wielkość jest niezbędna? O rozmiarze próby w badaniach jakościowych. *Przegląd Badań Edukacyjnych/Educational Studies Review*, 21(2), 195–211. [DOI:10.12775/PBE.2015.055]
- Pasikowski, S. (2019). Funkcja nasycenia teoretycznego i jej nienasycenie – implikacje metodologiczne dla badań w pedagogice. *Przegląd Pedagogiczny*, 2, 153–161. [DOI:10.34767/PP.2019.02.12]
- Patton, M.Q. (2018). Expert Sampling. W: B.B. Frey (red.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (s. 648). Los Angeles: SAGE Publications.
- Ritchie, J., Lewis, J., Elam, G. (2003). Designing and Selecting Samples. W: J. Ritchie, J. Lewis (red.), *Qualitative Research Practice. A Guide for Social Science Students and Researchers* (s. 77–108). New York: SAGE Publications.
- Sobczyk, M. (2007). *Statystyka*. Warszawa: Wydawnictwo Naukowe PWN.
- Steczowski, J. (1995). *Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych*. Warszawa–Kraków: Wydawnictwo Naukowe PWN.
- Szreder, M. (2010). Losowe i nielosowe próby w badaniach statystycznych. *Przegląd Statystyczny*, 57(4), 168–174.
- Wald, A. (1947–2013). *Sequential Analysis*. New York: Dover Publications.
- Watters, J.K., Biernacki, P. (1989). Targeted Sampling: Options for the Study of Hidden Populations. *Social Problems*, 36(4), 416–430.