NR 3 (634)
MARZEC 2014

WIADOMOŚCI STATYSTYCZNE

CZASOPISMO GŁÓWNEGO URZĘDU STATYSTYCZNEGO
I POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO

MIĘDZYNARODOWY ROK STATYSTYKI 2013
KONFERENCJA NAUKOWA
STATYSTYKA — WIEDZA — ROZWÓJ

Jan KORDOS

Od twierdzenia Jakuba Bernoulliego do współczesnych badań reprezentacyjnych[1]

Publikacji *Ars Conjectandi* (*Sztuka Przewidywania*) Jakuba Bernoulliego, wydanej 300 lat temu, poświęcono wiele opracowań i analiz naukowych, podkreślając jej znaczenie dla rozwoju rachunku prawdopodobieństwa i innych dziedzin naukowych[2]. Zapewne w ub. roku z okazji obchodzonego Międzynarodowego Roku Statystyki powstało wiele interesujących opracowań poświęconych tej publikacji, które rzucą nowe światło na rozwój rachunku prawdopodobieństwa i statystyki, a szczególnie na zastosowania tych dyscyplin w różnych dziedzinach.

Od twierdzenia Bernoulliego do współczesnych badań reprezentacyjnych daleka droga, ale wydaje się celowe dokonanie podróży przez trzy stulecia i prześledzenie rozwoju poszczególnych komponentów badań reprezentacyj-

[1] Jest to rozszerzona i uaktualniona wersja referatu wygłoszonego na ogólnopolskiej konferencji naukowej *Statystyka — wiedza — rozwój* zorganizowanej w Łodzi (17 i 18 października 2013 r.), z okazji Międzynarodowego Roku Statystyki.

[2] http: //books.google.pl/books?id=-xgwSAjTh34C&pg=PR9&lpg=PR9&ots=8v1Wj_bd_1&dq==jakob+bernoulli+law+of+large+numbers

nych, wskazując na ważniejsze momenty, które odegrały istotną rolę w tym rozwoju.

Rozpocznę od znaczenia tego twierdzenia dla wyboru próby z populacji, czyli przejścia od teorii do praktyki, prowadzącego do konstrukcji tablic liczb losowych, które umożliwiają spełnienie w praktyce teoretycznego założenia o niezależności zmiennych losowych. Chciałbym przypomnieć, w ogólnym zarysie, wkład różnych matematyków i statystyków, którzy przyczynili się do rozwoju teorii wykorzystywanych w praktycznych zastosowaniach badań reprezentacyjnych w różnych dziedzinach życia gospodarczego i społecznego.

Wykorzystując twierdzenie Bernoulliego spróbuję wykazać, dlaczego próbę z populacji należy wylosować, a nie wybrać w inny sposób, aby uzyskane oceny można było uogólnić na populację z określoną precyzją. Wspomnę przy okazji o tworzeniu tablic liczb losowych przez niektórych badaczy, a także o generatorach liczb losowych wykorzystywanych w symulacjach komputerowych i różnych badaniach naukowych.

Następnie będę rozważał rozwój różnych komponentów badań reprezentacyjnych, zmierzających do zwiększenia precyzji i dokładności uzyskiwanych ocen. Nawiążę do publikacji Bayesa[3], wydanej 250 lat temu, a głównie do *bayesizmu*, jednego z paradygmatów w statystyce, którego podejście wykorzystywane jest w badaniach reprezentacyjnych, a szczególnie przy estymacji złożonych ocen. Podkreślę znaczenie fundamentalnych prac prof. Jerzego Neymana, głównego twórcy drugiego paradygmatu w statystyce, *podejścia częstościowego*, dla rozwoju badań reprezentacyjnych. Wspomnę także o wykorzystywaniu w szerokim zakresie różnego rodzaju modeli przy estymacji parametrów i ich analizie, a także o niebezpieczeństwach występujących w niektórych praktycznych zastosowaniach. Nawiążę tu do słynnego stwierdzenia Boxa, iż: *wszystkie modele są złe, ale niektóre są użyteczne*[4].

Rozważę także znaczenie metod uzyskiwania danych od wybranych jednostek w rozmaitych typach badań, wykorzystywanie dodatkowych źródeł informacji przy wyborze próby i uogólnieniu wyników, a także napotykane trudności oraz postęp przy stosowaniu w szerokim zakresie nowoczesnej techniki obliczeniowej.

W zakończeniu opiszę ogólnie ostatnie zastosowania złożonych metod w badaniach reprezentacyjnych, w których wykorzystane są skomplikowane plany prób, modele, metody symulacyjne i inne techniki badawcze prowadzące do zwiększenia precyzji i dokładności ocen, ale wymagające znacznej uwagi przy interpretacji uzyskanych wyników. W uwagach końcowych wspomnę o pracach, jakie prowadzi ostatnio Eurostat w zakresie jakości, zmierzających stopniowo do wprowadzenia globalnego zarządzania jakością także w badaniach reprezentacyjnych[5].

[3] Bayes T. (1763), s. 370—418.
[4] Box G. E. P. (1976), s. 791—799.
[5] *Handbook on improving...* (2013).

TWIERDZENIE JAKUBA BERNOULLIEGO I JEGO NASTĘPCY

Opublikowanie *Ars Conjectandi* jest niewątpliwie wydarzeniem przełomowym co najmniej z dwóch powodów. Mianowicie, opracowanie zawiera wyraźne określenie pojęcia prawdopodobieństwa oraz opisanie jego własności, a także pierwsze sformułowanie prawa wielkich liczb. Wielu znakomitych uczonych zajmowało się tymi zagadnieniami w poszukiwaniu ilościowych sposobów oceny zjawisk losowych. Nie byli oni jednak w stanie dokonać ostatniego koniecznego kroku — wprowadzenia pojęcia prawdopodobieństwa, choć bliscy tego byli B. Pascal (1623—1662) i C. Huygens (1629—1695). Dokonali oni tego, że powstał fundament pod wprowadzenie tego pojęcia. Było ono analizowane do końca XIX w., w szczególności przez A. Moivre'a (1657—1754) w publikacji *The Doctrine of Chances*, wydanej w Londynie w 1756 r. oraz P. S. Laplace'a (1749—1827), ale dopiero A. N. Kołmogorow w 1933 r. rozwiązał ostatecznie to zagadnienie.

Twórczość Bernoulliego miała dla teorii prawdopodobieństwa znaczenie zasadnicze. Jego odkrycia w tej dziedzinie przedstawione są, jak już wspomniałem, w publikacji *Ars Conjectandi* wydanej pośmiertnie przez Mikołaja Bernoulliego w Bazylei w 1713 r. Książka J. Bernoulliego składa się z czterech części. Pierwszą część stanowi dzieło Huygensa, lecz Bernoulli podał swoje uwagi do prawie wszystkich wypowiedzi Huygensa, przy czym w niektórych przypadkach są one nawet bardziej istotne; w innej części rozwiązane są różnorodne zadania z teorii prawdopodobieństwa. Bernoulli przedstawił przede wszystkim ogólne pojęcia o naturze zdarzeń losowych, a następnie dowiódł twierdzenia, noszącego obecnie jego imię, stanowiącego podstawę wszystkich badań nad prawidłowościami masowych zjawisk losowych.

Twierdzenie Bernoulliego było pierwszym i najprostszym z wielu twierdzeń składających się na „prawo wielkich liczb" — termin ten wprowadził w 1835 r. francuski matematyk S. Poisson. Wraz z twierdzeniem Moivre'a-Laplace'a i jego uogólnieniami, „prawo wielkich liczb" i „centralne twierdzenie graniczne" należą do twierdzeń granicznych teorii prawdopodobieństwa, których pryncypialne znaczenie polega na tym, że na nich opierają się wszystkie zastosowania tej nauki do badania zjawisk przyrodniczych i społecznych.

Twierdzenie Bernoulliego dało matematyczną interpretację dobrze z doświadczenia znanego faktu, że występujące masowo zdarzenia losowe wykazują określone prawidłowości. Jeśli prawdopodobieństwo zajścia zdarzenia losowego A w następstwie niezależnych prób jest stałe i równe p w każdej próbie, to dla każdej dowolnie małej liczby dodatniej ε można twierdzić z prawdopodobieństwem dowolnie zbliżonym do jedności, że różnica $\frac{m}{n} - p$, co do bezwzględnej wartości, okaże się mniejsza niż ε. Twierdzenie to krócej można wyrazić przy pomocy następującego zapisu:

$$\Pr\left\{\left|\frac{m}{n}-p\right|\le\varepsilon\right\}\ge 1-\eta \quad \text{przy dostatecznie wielkich } n$$

gdzie:

m — liczba realizacji zdarzenia losowego A w wyniku n prób niezależnych,
ε i η — dowolnie małe liczby dodatnie.

Z analizy tego twierdzenia wynika także, że istnieje możliwość oceny częstości względnej p odnoszącej się do całej zbiorowości generalnej z określonym stopniem dokładności na podstawie wyników uzyskanych tylko od części tej zbiorowości, liczącej n jednostek. Pozostaje tylko do określenia stopień dokładności szacunku — wielkość prawdopodobieństwa. Trzeba określić ufność naszego wnioskowania oraz ustalić minimalną liczebność próbki n. Z takimi właśnie problemami spotykamy się w metodzie reprezentacyjnej, a więc dopuszczamy przy szacunku błąd, którego rozmiary możemy dowolnie regulować zwiększając odpowiednio liczebność próbki n. Zakładamy przy tym, że dane uzyskane od wybranych jednostek są zgodne ze stanem faktycznym.

Następnie matematycy i statystycy dowiedli, że prawo wielkich liczb zachodzi nie tylko dla częstości względnych i średnich, ale także dla innych cech zbiorowości, co ma szczególnie znaczenie w przypadku badań reprezentacyjnych (zagadnienie to rozwiązał matematyk rosyjski P. L. Czebyszew, 1821—1894). Oznacza to, że z prawdopodobieństwem dowolnie bliskim jedności można wnioskować, że różnice co do wartości bezwzględnej między charakterystyką dla całej zbiorowości i charakterystyką dla próbki nie różnią się więcej niż ε, gdy n jest dostatecznie duże, przy określonych warunkach.

Teoria rozkładów granicznych stanowi jeden z większych i ważniejszych działów rachunku prawdopodobieństwa, który rozwinął się szczególnie szybko w ostatnich kilkudziesięciu latach, jakkolwiek niektóre twierdzenia z tego zakresu zostały udowodnione jeszcze w wiekach XVIII i XIX. Jako pionierów prac badawczych w tej dziedzinie należy wymienić A. de Moivre'a, C. F. Gaussa i P. S. Laplace'a, odkrywców rozkładu normalnego odgrywającego w twierdzeniach granicznych podstawową rolę. Dalsze prace wielu wybitnych uczonych, a przede wszystkim prace Levy'ego, Lindeberga, Lapunowa, Fellera i Chinczyna doprowadziły do pełnego rozwiązania tego zagadnienia. Sformułowano i udowodniono kilka ważnych twierdzeń, podając warunki, które muszą być spełnione, aby zmienna losowa miała rozkład normalny lub zbieżny do tego rozkładu. Twierdzenia te od 1920 r. znane są pod wspólną nazwą „centralnego twierdzenia granicznego" (Fischer, 2011).

Prawo wielkich liczb nie daje gwarancji, że przy poszczególnych szacunkach błąd nie wypadnie większy niż przyjęta dowolna liczba. Stwierdza tylko, że przypadek taki staje się tym mniej prawdopodobny, im większa jest liczba obserwacji. Tak np., gdy szacujemy średnie wydatki gospodarstwa domowego na

określony produkt na podstawie próbki wylosowanej ze zbiorowości generalnej gospodarstw domowych to jest rzeczą mało prawdopodobną, że wylosujemy przypadkowo same gospodarstwa o niskich wydatkach i że w skutek tego szacunek średnich wydatków odbiegnie daleko od wartości rzeczywistej. Prawo wielkich liczb twierdzi jednak, że im większa liczba obserwacji, tym mniejsze jest prawdopodobieństwo otrzymania takiego rezultatu.

Do dalszych rozważań podam „słabe prawo wielkich liczb", nieco zmodyfikowane, według S. Zubrzyckiego[6]:

Jeśli $X_1, X_2, \ldots, X_n$ *jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie prawdopodobieństwa z wartością oczekiwaną* μ *i odchyleniem standardowym* σ, *a* $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$ *oznacza średnią pierwszych n zmiennych losowych, to dla każdego* $\varepsilon > 0$ *i* $\eta > 0$ *zachodzi nierówność, gdy n jest dostatecznie duże:*

$$P(|\overline{x}_n - \mu| < \varepsilon) > 1 - \eta$$

W pierwszej kolejności należy wyjaśnić, w jaki sposób próba powinna być wybrana ze zbiorowości generalnej, wykorzystując twierdzenie Bernoulliego. Pierwsza część słabego prawa wielkich liczb głosi (Kordos, 2009):

Jeśli $X_1, X_2, \ldots, X_n$ *jest ciągiem <u>niezależnych</u> zmiennych losowych o jednakowym rozkładzie prawdopodobieństwa z wartością oczekiwaną* μ *i odchyleniem standardowym* σ*".* Podkreśliłem słowo „niezależnych" zmiennych losowych, gdyż w praktyce, aby spełnić ten warunek teoretyczny statystycy zbudowali „tablice liczb losowych". Przy pomocy tablic liczb losowych wybieramy najpierw x_1 i traktujemy jako realizację zmiennej losowej X_1, a następnie x_2 — jako realizację zmiennej losowej X_2 itd. Próba $x_1, x_2, \ldots, x_n$ jest traktowana jako pojedyncza realizacja n-wymiarowej zmiennej losowej $(X_1, X_2, \ldots, X_n)$. Tak więc jednostki badania wybierane są do próby przy pomocy „tablic liczb losowych", aby spełnić warunek niezależności.

Druga część twierdzenia słabego prawa wielkich liczb brzmi: *wtedy dla każdego* $\varepsilon > 0$ *i* $0 < \eta < 1$ *zachodzi* $P\{|\overline{x}_n - \mu| < \varepsilon\} > 1 - \eta$, *gdy n jest dostatecznie duże*. Wydaje mi się, że prof. J. Neyman (1933, 1934) przy tworzeniu przedziałów ufności wykorzystał nierówność $P\{|\overline{x}_n - \mu| < \varepsilon\} > 1 - \eta$ i zastąpił ją $P\{|\overline{x}_n - \mu| < \varepsilon\} = 1 - \alpha$, gdzie ε oznacza precyzję oceny, a 1 — α poziom ufności. Ale tego nie potrafię udowodnić.

Prawo wielkich liczb, jak już podkreślono, stanowi podstawę teoretyczną badań reprezentacyjnych. Można z niego wyciągnąć bardzo ważny i praktyczny

[6] Zubrzycki S. (1966), s. 139.

wniosek, że o parametrach zbiorowości generalnej (populacji) można wnioskować na podstawie badania tylko części tej zbiorowości, wylosowanej z całej populacji. Jeżeli próbka będzie dostatecznie liczna, to precyzja takiego wnioskowania może być dowolnie wysoka (tzn. błędy mogą być dowolnie małe), przy czym prawdopodobieństwo słuszności takiego wnioskowania, zwane poziomem ufności, może być odpowiednio zbliżone do pewności. Jednakże tak sformułowane samo prawo wielkich liczb nie rozwiązuje jeszcze podstawowych problemów badań reprezentacyjnych. Na przykład nie można odpowiedzieć, w jaki sposób należy określić minimalną liczebność próbki, aby zapewnić żądaną precyzję oszacowań przy zadanym poziomie ufności; jak obliczyć prawdopodobieństwo słuszności wypowiadanych sądów (poziomu ufności); w jaki sposób oszacować precyzję badanych parametrów po przeprowadzeniu badania. Do rozwiązania tych problemów wykorzystano, po odpowiednim przystosowaniu, różne twierdzenia graniczne rachunku prawdopodobieństwa, a szczególną rolę odgrywa tu „centralne twierdzenie graniczne". Rozważam dalej także inne problemy metodologiczne badań reprezentacyjnych, a twierdzenie Bernoulliego traktuję jako teoretyczny wstęp do badań reprezentacyjnych.

METODOLOGICZNE PROBLEMY BADAŃ REPREZENTACYJNYCH

Metodologicznie badania reprezentacyjne dotyczą trzech różnych, lecz równie ważnych pytań:

1) w jaki sposób próba powinna być wybrana z danej populacji, aby uzyskane wyniki uogólnić na populację (***metody wyboru próby***),
2) w jaki sposób wyciągnąć wnioski o populacji na podstawie danych otrzymanych z wybranej próby oraz dostępnych informacji oraz określić precyzję takiej oceny (***metody estymacji***) oraz
3) w jaki sposób uzyskać informacje od jednostek wybranych do próby i dostępnych źródeł (***metody uzyskiwania danych***), aby były zgodne ze stanem faktycznym.

ROZWÓJ METODYKI BADAŃ REPREZENTACYJNYCH

Badania częściowe, w których jednostki do badania były wybierane przy pomocy jednej z metod wyboru celowego, stosowano zwykle w badaniach społecznych, związanych z badaniami warunków bytu, ubóstwa lub badania opinii. Na przełomie wieków XIX i XX szczególną rolę odegrały badania monograficzne, w których stosowano metodę wywiadu, rejestracji i obserwacji w celu uzyskania informacji od badanych jednostek. W badaniach eksperymentalnych, głównie rolniczych, stosowano zwykle jedną z metod wyboru celowego, a później wykorzystywano także wybór losowy, stosując twierdzenie Bernoulliego.

Po raz pierwszy idea wnioskowania statystycznego została przedstawiona przez francuskiego naukowca P. S. Laplace'a. W 1781 r. opublikował on plan badania częściowego, w którym określił wielkość próby potrzebnej do osiągnięcia pożądanej dokładności w ocenie wyniku. Plan opierał się na zasadzie odwrotnych prawdopodobieństw Laplace'a i wyprowadzeniu centralnego twierdzenia granicznego. Został on opublikowany w 1774 r. i jest jednym z rewolucyjnych dokumentów w historii wnioskowania statystycznego. Model wnioskowania Laplace'a opierał się na próbach Bernoulliego i rozkładzie binominalnym. Zakładał on, że populacja ciągle się zmienia. Został przedstawiony przy założeniu rozkładów *a priori* dla parametrów. Model wnioskowania Laplace'a zdominował myślenie statystyczne przez cały wiek. Dobór próby w badaniach Laplace'a był celowy (Kuusela, 2011). Dalej rozważam tylko losowy wybór próby.

Metody wyboru próby — liczby losowe i ich zastosowania

Tablice liczb losowych zostały zbudowane, aby w praktyce zastosować twierdzenie Bernoulliego i inne twierdzenia graniczne. Rozważmy więc kto i kiedy zbudował tablice liczb losowych oraz ich zastosowanie w różnych dziedzinach badań. Wybór próby z populacji generalnej jest tylko jednym z elementów badań reprezentacyjnych. W praktyce mamy do czynienia z populacjami złożonymi, wymagającymi zastosowania skomplikowanych procedur badawczych, aby uzyskać wiarygodne oceny dla całej populacji. Przedstawmy więc w ogólnym zarysie rozwój planów próby do zbadania złożonych układów populacji, wymagających zastosowania specjalnych metod uzyskiwania danych, a także wykorzystania danych z innych źródeł, w celu zwiększenia jakości wyników. Szczególną uwagę zwrócę na błędy nielosowe, których minimalizacja jest jednym z podstawowych celów badań reprezentacyjnych.

Dlaczego należy wybrać jednostki do badania częściowego przy pomocy tablicy liczb losowych, aby uzyskane wyniki można było uogólnić na populację generalną z określonym błędem losowym, wykorzystując twierdzenie Jakuba Bernoulliego? Kto pierwszy i kiedy zbudował tablice liczb losowych nie potrafiłem ustalić. Losowanie i liczby losowe wykorzystywano już w XVIII w. i później w ruletkach, badaniach eksperymentalnych, a także na mniejszą skalę w badaniach częściowych. Pierwsze tablice liczb losowych wydał w roku 1927 L. H. Tippett pod tytułem *Random Sampling Numbers*. K. Pearson we wstępie do tej publikacji wyjaśnia, w jaki sposób w ostatnich latach dokonywano losowania prób. Zwykle dotyczyło to badań eksperymentalnych i badań naukowych na mniejszą skalę. Omawia też sposoby weryfikacji losowości tych liczb, a także wskazuje na pewne wątpliwości odnośnie losowości tablic Tippetta.

Następnie opublikowano tablice liczb losowych:

- w 1938 r. — *Fisher's and Yates Tables*,
- w 1939 r. — *Kendall and Babington Smith's Random numbers*,

- w 1955 r. — *Rand Corporation Table of Random numbers*,
- w 1966 r. — *C. R. Rao. Mitra and Mathai Table of Random numbers*.

W Polsce w roku 1951 prof. E. Vielrose opracował dla potrzeb GUS tablicę liczb losowych, wykorzystując do tego paski do drukujących maszyn liczących[7]. W późniejszym okresie konstrukcją liczb losowych zajmował się także prof. H. Steinhaus[8].

Liczby losowe wykorzystywane są w reprezentacyjnych badaniach statystycznych (przy losowaniu próby z populacji generalnej lub, szerzej, planowaniu schematu losowania), a zatem we wszelkich badaniach ekonomicznych, społecznych, marketingowych, badaniach opinii, w zagadnieniach statystycznej kontroli jakości produktów itp. Innym sposobem wykorzystania liczb losowych są badania symulacyjne. Są to metody wykorzystujące techniki obliczeniowe, w których przedstawiany jest przebieg realnego zjawiska, opisanego odpowiednimi równaniami, uwzględniając wpływające na nie czynniki losowe. Badania takie są często jedynym sposobem ilościowej analizy skomplikowanego procesu technologicznego lub zjawiska przyrodniczego (Kotulski, 2001; Zieliński, 1972).

W ostatnim czasie liczby losowe zyskały ogromnie na znaczeniu, są one stosowane w kryptografii (np. w wojskowości i dyplomacji) czy też szerzej rozumianej ochronie informacji. W ostatnich latach powstały nowe pola stosowania liczb losowych, dotyczą one współczesnych sposobów przesyłania danych masowych (przy użyciu telefonii komórkowej i sieci komputerowych), w których niezbędna jest ochrona informacji (Kotulski, 2001). Temat zastosowania i tworzenia liczb losowych i pseudolosowych jest przedstawiony szeroko w Internecie[9].

Początki metody reprezentacyjnej

Metodę badania częściowego na szerszą skalę wykorzystano już w 1891 r. w spisie ludności w Norwegii. Została ona nazwana „metodą reprezentacyjną". Metoda ta była zaprezentowana na sesji Międzynarodowego Instytutu Statystycznego (MIS) w Bernie w 1895 r. przez A. Kiaera (1897). Dlatego rok 1895 uważany jest za początek badań reprezentacyjnych w statystyce oficjalnej. Kiaer dowodził, że badanie częściowe może dać rzetelne wyniki pod warunkiem, że obserwacje tworzą reprezentatywny obraz całej badanej zbiorowości. Próba powinna, jego zdaniem, odzwierciedlać populację ze względu na ważne cechy. Uważano wtedy, że reprezentatywność próby może być osiągnięta na dwa sposoby, przez wybór:

a) zrandomizowany (losowy) z całej populacji oraz
b) celowy, który odzwierciedla populację ze względu na wybrane zmienne.

[7] Vielrose E. (1951).
[8] Steinhaus H. (1956), s. 51—65.
[9] Na przykład: http://www.random.org/randomness/.

W pierwszych dekadach XX w. badania reprezentacyjne były przeprowadzane przy wykorzystaniu jednej z tych metod, a do połowy lat 30. ub. wieku na pytanie, która z nich jest „lepsza" nie znano odpowiedzi.

Przełom w metodzie wyboru próby

Fundamentalny wkład w rozwój teorii metody reprezentacyjnej wniósł Jerzy Neyman, światowej sławy statystyk polskiego pochodzenia. W swoim słynnym artykule[10] zmienił teoretyczne podstawy wnioskowania na podstawie badań częściowych, wprowadzając błędy losowe oparte na randomizacji rozkładu. Wprowadził „optymalną lokalizację próby" w losowaniu warstwowym, a także „przedziały ufności"[11]. Użył on terminu „reprezentatywny" w nowym znaczeniu. Jerzy Neyman był współtwórcą, wspólnie z R. A. Fisherem, K. Pearsonem i E. Pearsonem, nowego paradygmatu w statystyce „podejścia częstościowego", który zdominował badania reprezentacyjne w latach 1950—1990 (Kuusela, 2011).

Jednakże w pierwszych dekadach XX w. podejmowane były już badania częściowe na szerszą skalę. Należy wspomnieć o badaniach prowadzonych i opisanych przez A. L. Bowleya (1915, 1926), a także badania bezrobocia prowadzone w latach 1934—1942 w Stanach Zjednoczonych, które w 1943 r. przekształciły się w sławne badania siły roboczej prowadzone przez amerykańskie Biuro Spisów[12].

Należy zaznaczyć, że MIS uznał metodę reprezentacyjną w statystyce oficjalnej w 1925 r., A. L. Bowley, pierwszy profesor statystyki w London School of Economics and Political Science — przez swoje prace o pomiarze[13] oraz o doborze próby[14], był pierwszą z głównych postaci rozwoju naukowego doboru próby, opartego na zasadach rachunku prawdopodobieństwa.

Rozwój metodyki badań reprezentacyjnych w drugiej połowie XX w.

Intensywny rozwój metodyki badań reprezentacyjnych rozpoczął się po 1945 r. Ukazało się 5 podręczników na ten temat w języku angielskim, które opracowali: F. Yates (1949), W. E. Deming (1950), W. G. Cochran (1953), M. H. Hansen, W. N. Hurwitch i W. G. Madow (1953) oraz P. V. Sukhatme i B. V. Sukhatme (1954). W szerokim zakresie rozpoczęto również stosowanie metody reprezentacyjnej w rozmaitych dziedzinach. Po ukazaniu się wspomnianych podręczników zorganizowano wiele kursów z metody reprezentacyjnej

[10] Neyman J. (1934), s. 558—606.
[11] Neyman J. (1933); (1938), s. 101—116.
[12] Kish L. (1996), s. 3—16.
[13] Bowley A. L. (1915).
[14] Bowley A. L. (1926), s. 1—62.

w Indiach, Stanach Zjednoczonych, Wielkiej Brytanii i wielu innych krajach. W Polsce współzależność między rozwojem teorii i praktyki badań reprezentacyjnych próbowałem przedstawić na łamach „Wiadomości Statystycznych”[15].

W latach 1940—1960, a okres ten uważany jest za złotą erę w badaniach reprezentacyjnych, szczególną rolę w tej dziedzinie badań odegrali Hansen, Hurwitz, Madow, Waksberg, Bershad, Tepping i in. (Kish, 1996; Kuusela, 2011). Uważa się, że Hansen odegrał główną rolę w rozwoju metodologii badań reprezentacyjnych w XX w. Hansen i Hurwitz[16] zastosowali losowanie warstwowe wielostopniowe z jedną jednostką pierwszego stopnia w warstwie z prawdopodobieństwem proporcjonalnym do wielkości, aby zapewnić zrównoważony schemat losowania. Wprowadzili estymację złożoną do bieżących badań ludności stosując losowanie rotacyjne w celu redukcji obciążenia respondentów (Hansen i in., 1953). Zastosowali więc koncepcję Neymana do praktyki statystycznej na szeroką skalę. Wydaje mi się, że głównie Hansen i jego zespół przyczynili się w znacznym stopniu do rozpowszechnienia metodologii opracowanej przez Neymana. W latach 1950—1990 dominowało stosowanie idei „podejścia częstościowego” w badaniach reprezentacyjnych (Kuusela, 2011). Od 1990 r., po wprowadzeniu na szerszą skalę nowoczesnej techniki obliczeniowej, zaczęła powracać w badaniach reprezentacyjnych dominacja *bayesizmu*, wykorzystująca na szeroką skalę dostępne informacje z różnych źródeł, przy zastosowaniu różnorodnych modeli. W szerszym zakresie zaczęła być wykorzystywana empiryczna i hierarchiczna estymacja bayesowska (Lyberg i in., 1998; Groves i in., 2004; Rao, 2003, 2005).

Instytut Statystyczny w Indiach, kierowany przez Mahalanobisa[17], stał się drugim centrum badań reprezentacyjnych o znaczeniu międzynarodowym. *Indyjskie Narodowe Badanie Reprezentacyjne* (*Indian National Sample Surveys*) zdobyło sławę międzynarodową[18]. Mahalanobis (1944) już w 1937 r. w Indiach wykorzystał losowanie wielostopniowe (wsi, poletek i zdjęć fotograficznych) do badania plonów. Wykorzystał funkcje kosztu badania i wariancje przy planowaniu badania, a więc zastosował w praktyce teorie opracowane przez J. Neymana.

W aspekcie międzynarodowym znaczną rolę odgrywają także sesje naukowe MIS, odbywające się co dwa lata. Na sesjach tych prezentowane są zarówno osiągnięcia teoretyczne, jak i praktyczne w dziedzinie badań reprezentacyjnych uzyskane przez ośrodki badawcze, a także przez urzędy statystyczne różnych krajów. Istotną rolę w rozwoju metodologii badań reprezentacyjnych odgrywa także jedna z sekcji MIS — International Association of Survey Statisticians (IASS), która jest organizatorem lub współorganizatorem wielu konferencji międzynarodowych.

[15] Kordos J. (2012), s. 61—86.
[16] Hansen M. H., Hurwitz W. N. (1943), s. 333—362.
[17] Mahalanbis P. C. (1944), s. 329—451; (1952), s. 1—7.
[18] Sukhatme P. V., Sukhatme B. V. (1954).

Metody pozyskiwania danych

Ważnym komponentem badań reprezentacyjnych są metody uzyskiwania danych od jednostek wybranych do próby. Chodzi o to, aby informacje uzyskane od jednostek badanych były zgodne ze stanem faktycznym, odzwierciedlające badaną rzeczywistość. Okazuje się jednak, że w wielu jednostkach w badaniach reprezentacyjnych występują poważne trudności w uzyskiwaniu danych zgodnych ze stanem faktycznym. Część wylosowanych jednostek, z różnych przyczyn, nie bierze udziału w badaniu, a inne nie udzielają wiarygodnych informacji. Dotyczy to głównie badań społecznych, ale mogą one wystąpić również w badaniach ekonomicznych.

Metody zbierania danych, stosowane w badaniach reprezentacyjnych, można ogólnie podzielić na następujące grupy (Dillman i in., 2009; Groves i in., 2004): a) bezpośrednie wywiady indywidualne, b) wywiady telefoniczne, c) metody oparte na systemie pocztowym, d) metody wykorzystujące Internet, e) metody pomiaru, f) samo spisywanie, g) metody obserwacji, h) metody wykorzystujące różnego rodzaju rejestry.

Problematykę uzyskiwania danych oraz trudności napotykane w badaniach szeroko przedstawiłem w monografiach nt. jakości danych statystycznych i błędów w badaniach społecznych (Kordos, 1987, 1988).

We wczesnym okresie badań społecznych metoda zbierania danych w badaniach w żadnym przypadku nie była zestandaryzowana. W klasycznych studiach nad ubóstwem nie komunikowano się bezpośrednio z jednostkami, o których zbierano dane. Praktykowano gromadzenie zapisów ankietera bez odnoszenia się do respondentów. Obecnie praktyka ta jest kontynuowana w ograniczonym zakresie w przypadku niektórych informacji (Kuusela, 2011).

W projektowaniu kwestionariuszy poczyniono znaczne postępy w porównaniu z klasyczną już dziś pracą Payne'a opublikowaną w 1951 r. na temat sztuki zadawania pytań. Problematyka ta została szeroko przedstawiona w pracach takich autorów, jak Sudman i Brandburn (1982) oraz Kalton i Kasprzyk (1986)[19]. Pytania są znacznie lepiej konstruowane i formułowane, bardziej przejrzyście rozłożone na kwestionariuszu oraz w większym stopniu korzysta się z badań pilotażowych, które przyczyniają się do zmniejszenia błędów nielosowych.

Znaczną korzyścią w kontrolowaniu błędów nielosowych jest wprowadzenie do badań statystycznych wywiadów z zastosowaniem komputera, które zostały rozwinięte przez metodologów, takich jak Bethlehem i Keller z Holandii, z zamiarem kontrolowania całego procesu badawczego (Bethlehem, 2009). Przyczyniło się to do zredukowania kosztów badania, a szczególnie kontroli opracowania (Biemer i in., 1991; Dillman i in., 2009; Groves i in., 2002; Lyberg i in., 1998).

[19] S. 1—16.

We wczesnych latach 60. XX w. ważnym krokiem było wprowadzenie wywiadu telefonicznego. W Stanach Zjednoczonych ok. 1970 r. po raz pierwszy rozwinięto system wywiadów telefonicznych wspomaganych komputerem (CATI), głównie przez agencje badania rynku (Groves i in., 2004).

Ostatnio obserwuje się przesunięcie zainteresowania nad rozwojem metod zbierania danych w kierunku problemów związanych z pomiarem i błędem pomiaru. Znaczny wkład wniósł tu Groves (1989) syntezując psychometrię, statystykę i nauki społeczne w celu uzyskania obszernego przeglądu błędów badania i możliwych ich przyczyn. Kładzie on nacisk na estymacje błędu i połączenie błędu oceny z kosztami badania.

W 1990 r. zorganizowano międzynarodową konferencję poświęconą pomiarom błędów w badaniach, a jej wyniki opublikowano w 1991 r. w specjalnej monografii (Lyberg, Kasprzyk, 1991), w której podano przegląd metod zbierania danych i pomiaru błędów. Nadal prowadzone są prace w kierunku minimalizacji błędów nielosowych, wykorzystując dostępne źródła informacji oraz Internet.

Koncentracja na błędach nielosowych

Błędy nielosowe w znacznym stopniu popełniane są przy uzyskiwaniu danych od wylosowanych jednostek do próbki, na skutek różnych przyczyn. Jak już wspomniano, po drugiej wojnie światowej teoria i praktyka badań reprezentacyjnych rozwijały się intensywnie (Kish, 1996; O'Muircheartaigh, Wong, 1981[20]; Smith, 1994[21]; Rao, 2005). W latach 50. ub. wieku prawie zakończono rozwój teorii randomizacji, a zainteresowanie statystyków przesunęło się stopniowo z błędów losowych w kierunku błędów nielosowych, do czego w szczególnym stopniu przyczynił się M. H. Hansen i jego współpracownicy (Hansen, Hurwitz, 1943, 1946[22]; Hansen i in., 1953; Hansen i in., 1961[23]; Hansen, Waksberg, 1970[24]; Kuusela, 2011). W tym czasie obserwuje się znaczny rozwój badań reprezentacyjnych, chociaż nie zawsze zgodnie z aktualnym rozwojem teorii. Niejednokrotnie praktyka wyprzedzała teorie, które opracowali: O'Muircheartaigh, Wong (1981); Smith (1994). Dotyczy to szczególnie początków lat 80. ub. wieku.

W historycznym rozwoju badań reprezentacyjnych występują trzy wyraźne kierunki związane z badaniem błędów nielosowych:

a) badania związane ze statystyką oficjalną,
b) badania akademickie (zwykle związane ze studiami społecznymi),
c) badania rynku (związane z handlem i biznesem).

[20] S. 437—446.
[21] S. 3—34.
[22] S. 517—529.
[23] S. 359—374
[24] S. 317—332.

Każdy z tych kierunków wniósł do badań własny potencjał intelektualny oraz własną perspektywę dyscypliny, a także kryteria oceny sukcesu lub niepowodzenia (Groves i in., 2002, 2004; O'Muircheartaigh, Wong, 1981).

Zasadnicze wpływy na te kierunki rozwoju wywarły ruch reform społecznych i wyłaniająca się dyscyplina socjologii. Niektórzy pionierzy badań reprezentacyjnych łączyli zarówno statystykę oficjalną, jak i badania społeczne. W szczególności Bowley (mający istotny wpływ na rozwój doboru próby do badań) w 1915 r. przedstawił pracę o pomiarze społecznym, która pomogła w zdefiniowaniu parametrów jakości danych i błędu dla informacji faktologicznych i behawiorystycznych. Tak więc dziedzina studiów nad badaniem ustalona w latach 40. i 50. XX w. objęła trzy sektory: rządowy, akademicki oraz biznesu. Miały one trzy bazy dyscyplinarne — statystykę, socjologię oraz psychologię eksperymentalną — a w tych dziedzinach rozwinięto różne struktury i terminologię.

Jasne jest, że jakikolwiek model procesu badania, a przeto jakiś ogólny model błędu badania będzie musiał włączyć te elementy. Jedna z możliwych struktur zarysowana jest w książce Kisha o statystycznych planach prac badawczych (Kish, 1987).

Kish sugeruje, że występują trzy zagadnienia, w zakresie których badacz powinien zastosować plan pracy badawczej. Proponuje, aby podobną typologię można było wykorzystać do klasyfikacji rozmiarów, które mogłyby objąć większość źródeł błędów. Każdy z tych rozmiarów jest wielowymiarowy. Oto one: reprezentacja, randomizacja oraz realizm. „Reprezentacja" obejmuje zagadnienia takie, jak: użycie probabilistycznego wyboru próby, warstwowanie, unikanie obciążenia wyboru oraz rozważania o braku odpowiedzi. „Randomizacja" (a jej odwrotnością w tym kontekście jest kontrola) obejmuje zagadnienia eksperymentowania i kontroli badanych zmiennych. Warto podkreślić, że randomizacja lub przynajmniej wybór losowy jest także użyta w celu osiągnięcia reprezentywności w probabilistycznym wyborze próby. „Realizm" powstaje jako problem w tym kontekście na dwa sposoby. „Realizm odnośnie zmiennych" pokazuje, w jakim stopniu mierzone lub obserwowane zmienne odnoszą się do konstrukcji, które mają opisywać. „Realizm w otoczeniu" dotyczy stopnia, w jakim układ zbieranych danych lub eksperymentu jest podobny do kontekstu rzeczywistości, którą badacz jest zainteresowany. „Replikacje" pozostają podstawowym narzędziem statystyków w przypadku błędów. Tradycyjny podział między „wariancję", czyli zmienność przez replikacje oraz „obciążenie", czyli systematyczne odchylenie od pewnej poprawnej lub prawdziwej wartości, informuje o statystycznym podejściu do błędów.

Replikacja (lub interpenetracja Mahalanobisa) jest metodą produkującą mierzalność w sensie możliwości oceny zmienności z wnętrza procesu. W losowaniu dokonywane to jest przez wybór niezależny pewnej liczby jednostek losowania oraz wykorzystanie różnic między nimi jako przewodnika właściwej stabilności lub niestabilności ogólnych ocen. W przypadku prostej wariancji odpowiedzi statystycy po prostu powtarzają obserwację podzbioru respondentów

i porównują wyniki. Zwykle nazywa się to programem ponownych wywiadów, co daje replikacje w sensie powtórzenia.

We wczesnych latach 60. ub. wieku widzieliśmy kolejny postęp w rozważaniach statystyków na temat błędu badania. Hansen i in. (1961) przedstawili pracę, która znana jest jako model błędu Biura Spisów Stanów Zjednoczonych (*U.S. Census Bureau model of survey error*). Wariancja została zdefiniowana w stosunku do tych podstawowych warunków badania. Obserwacja widziana jest jako złożona z dwóch części, jej wartości prawdziwej i odchylenia od tej wartości — odchylenie odpowiedzi.

Chociaż Hansen i jego współpracownicy dobrze zdawali sobie sprawę, że pojęcie „prawdziwej wartości" jest problematyczne, to proponowali je jako użyteczną podstawę definicji, a następnie estymacji błędu. Ich model jest w zasadzie modelem wariancji/kowariancji, bo pozwala na znaczne uogólnienia[25] i był niezwykle popularny wśród statystyków badań reprezentacyjnych. Umożliwia on włączenie efektów ankieterów i kontrolerów oraz możliwość skorelowanych błędów wewnątrz gospodarstw domowych. Tak więc w podejściu tym mogą być włączone: błędy opracowania, brak odpowiedzi, niedoszacowanie zbiorowości i błędy pomiaru. Dla ogólności, jakiekolwiek obciążenia (bez względu na ich źródło) mogą również być dodane, dając średni błąd kwadratowy jako całkowity błąd oceny. Ta koncentracja ocen pojedynczego parametru opisowego powstała częściowo ze statystyki oficjalnej (która często dotyczyła oceny pojedynczej ważnej cechy populacji, takiej jak wskaźnik bezrobocia), a częściowo z ogólnej tradycji statystycznej przedziałowej estymacji parametrów rozkładu.

Tematyce błędów nielosowych w badaniach reprezentacyjnych poświęcono wiele uwagi i dostępna jest obszerna literatura, ale w artykule wymieniam tylko wybrane pozycje (Biemer i in., 1991; Fellegi, 1974; Groves, 1989; Groves i in., 2002; Hansen i in., 1961; Lessler, Kalsbeek, 1992; Lyberg i in., 1998; Zarkovich, 1966).

AKTUALNY ROZWÓJ METODYKI WSPÓŁCZESNYCH BADAŃ REPREZENTACYJNYCH

Metodyka badań reprezentacyjnych ulega usprawnieniom zarówno w zakresie wyboru próby, metod estymacji, jak i sposobów pozyskiwania danych z różnych źródeł. Wynika to z ogólnego rozwoju teorii, jak i sztuki badań reprezentacyjnych. W znacznym stopniu przyczynia się do tego rozwój nowoczesnej techniki obliczeniowej i informatyki, dostarczający odpowiednich algorytmów przetwarzania danych, a także rozwój różnych źródeł danych, które wykorzystywane są w coraz szerszym zakresie w badaniach reprezentacyjnych. Podam tu tylko niektóre kierunki badań, które, moim zdaniem, odgrywają istotną rolę w rozwoju badań reprezentacyjnych.

[25] Fellegi I. (1974), 496—501.

Podejście wspomagane modelem

Estymację opartą na modelach stosowano od połowy lat 60. XX w. Estymacja ta nie wykorzystywała planu próby (łączne prawdopodobieństwo wyboru elementów), lecz formalny model, który odnosił się do różnych zmiennych w analizie i ich związków między sobą, niezależnie od konfiguracji próby. Metody statystyczne wykorzystywane w tych analizach są zwykle oparte na uogólnionym modelu liniowym. Taki model prawie zawsze zakłada niezależne i identyczne rozkłady obserwacji, a także normalność rozkładu błędów. Nacisk położony jest często na specyfikację modelu i testowanie (Biemer i in., 1991; Groves i in., 2004; Hansen i in., 1983[26]; Särndal i in., 1992; Särndal, Lundström, 2005).

Należy zauważyć, że w badaniach reprezentacyjnych niekiedy uzyskuje się oceny bardzo precyzyjne, a zarazem niedokładne. Dotyczy to niektórych ocen w badaniach społecznych uzyskanych metodą wywiadu, np. z zakresu dochodów, w których obciążenie nieraz przewyższa 10%, a precyzja kształtuje się poniżej 2%.

W takich przypadkach nie ma sensu budowania przedziałów ufności, gdyż wprowadza się użytkownika w błąd, twierdząc, że: *podany przedział ufności pokrywa prawdziwą wartość szacowanego parametru przy danym poziomie ufności*. Warto o tym pamiętać przy planowaniu badania reprezentacyjnego, a następnie przy analizie uzyskanych wyników. Z tych powodów prof. J. Neyman podczas konsultacji dla Komisji Matematycznej GUS w 1958 r. nie zalecał stosowania metody reprezentacyjnej w badaniach budżetów rodzinnych[27].

A przecież w wielu badaniach reprezentacyjnych prowadzonych przez urzędy statystyczne na wielką skalę pewna frakcja badanych jednostek nie podejmuje badań, występują błędy odpowiedzi lub pomiaru oraz innego rodzaju błędy nielosowe. W takich przypadkach nie można twierdzić, iż podany przedział ufności pokrywa prawdziwą wartość parametru na określonym poziomie ufności. Jesteśmy w stanie tylko określić wielkość precyzji. Mój pogląd w tej sprawie wyrażałem na kilku konferencjach, a także zaznaczyłem go wyraźnie w publikacji w języku angielskim[28]. Przed szerokim wykorzystaniem modeli statystycznych ostrzegał także Box (1976), kładąc nacisk na przyjęte założenia. Wiadomo, że modele statystyczne wykorzystywane są efektywnie w różnych analizach i zastosowaniach praktycznych. Chodzi tu głównie o to, aby przy ich budowie oraz interpretacji uzyskanych wyników uwzględnić realność założeń, na których się one opierają.

Metody estymacji dla małych obszarów

Od ponad trzydziestu lat statystycy zajmują się problematyką estymacji dla małych obszarów, gdy wyniki z badań reprezentacyjnych w tych układach oka-

[26] S. 776—793.
[27] Zasępa R. (1958), s. 7—12.
[28] Kordos J. (2011), s. 112—122.

zują się nierzetelne. W ostatnich latach odbyło się kilka międzynarodowych konferencji i seminariów naukowych, na których prezentowano dorobek w tej dziedzinie badań. Bibliografia dotycząca estymacji wyników dla małych obszarów jest dość obszerna. Na szczególną uwagę zasługuje międzynarodowe sympozjum zorganizowane w 1985 r. w Ottawie (Platek i in., 1987) oraz międzynarodowa konferencja naukowa w 1992 r. w Warszawie (Kalton i in., 1993). Ta ostatnia miała znaczący wpływ na szersze zainteresowanie się tą tematyką badawczą w Polsce[29].

Metodologia estymacji dla małych obszarów (SAE) zaczyna odgrywać ważną rolę w nowoczesnym systemie informacji, który ma na celu zmniejszenie kosztów badań, obniżenie obciążenia respondentów oraz dostarczenie wiarygodnych oszacowań na niższych poziomach agregacji. Powstała Europejska Grupa Robocza ds. Estymacji dla Małych Obszarów (*European Working Group on Small Area Estimation*), która skupia specjalistów zarówno z kręgów akademickich, jak i statystyki oficjalnej oraz organizuje międzynarodowe konferencje. Dotychczas odbyły się takie konferencje: w 2005 r. w Finlandii (Jyväskylä), w 2007 r. we Włoszech (Piza), w 2009 r. w Hiszpanii (Elche), w 2011 r. w Niemczech (Trier), a w Polsce odbędzie się międzynarodowa konferencja od 3 do 5 września 2014 r. w Poznaniu. Teoria estymacji dla małych obszarów przedstawiona jest obszernie również w książce J. N. K Rao (2003).

Złożone metody estymacji — imputacja, kalibracja, metody bootstrapowe

Występowanie braków odpowiedzi w wielu badaniach statystycznych jest istotnym problemem związanym z ich realizacją, a sposoby radzenia sobie z tym zjawiskiem i ograniczania jego negatywnych konsekwencji dla wyników badań stanowią ważne wyzwanie metodologiczne. W praktyce stosuje się więc różne techniki obliczeniowe. Prowadzone są prace badawcze w celu minimalizacji skutków braku odpowiedzi. Należy tu wymienić przede wszystkim różne metody imputacji i kalibracji, a przy trudnościach w ocenie wariancji w złożonych metodach konfiguracjach prób — metody bootstrapowe.

1. Imputacja

W stosunku do pozycyjnych braków odpowiedzi (takich, gdy jednostka statystyczna bierze udział w badaniu, ale z jakichś powodów nie udziela odpowiedzi na niektóre pytania) możliwe jest przyjęcie przez badacza różnej strategii. Przyjęcie restrykcyjnych wymogów co do kompletności zbieranych wywiadów (ograniczenie możliwości nieudzielenia odpowiedzi przez respondenta na poszczególne pytania) pozwala uzyskać spójne i kompletne zbiory danych wynikowych, jednak może powodować wzrost częstości występowania jednostkowych braków odpowiedzi, czyli takich, gdy jednostka statystyczna w ogóle nie

[29] Domański C., Pruska K. ((2000), s. 1—26.

bierze udziału w badaniu. Dopuszczenie odmów odpowiedzi na poszczególne pytania pozwala ograniczyć występowanie odmów udzielenia wywiadu w ogóle, pogarsza jednak kompletność i spójność wewnętrzną oraz może zmniejszać użyteczność uzyskanego zbioru danych wynikowych. W takim przypadku w zbiorze danych wynikowych pojawiają się pozycyjne braki danych, wymagające przyjęcia jakiejś procedury postępowania z nimi.

Rozwiązaniem przywracającym w dużym stopniu zbiorowi danych niekompletnych użyteczność i funkcjonalność zbliżoną, jak w przypadku danych kompletnych (choć z zastrzeżeniami co do wnioskowania, o których należy pamiętać) może być „imputacja". Statystycy dalej zajmują się intensywnie tymi problemami i prowadzą prace badawcze mające na celu ich rozwiązanie. Rozwijana jest nowa metodologia imputacji (*The New Imputation Methodology*)[30] (Chambers, 2003; Rubin, 1987).

2. Kalibracja

Kalibracja jako metoda szacowania parametrów w populacji generalnej jest techniką statystyczną wykorzystywaną w badaniach reprezentacyjnych (Deville, Särndal, 1992). Punkt wyjścia do jej zastosowania stanowi odpowiedni sposób ustalania wag. Kalibracja stała się ważnym narzędziem metodologicznym. Polega ona na skorygowaniu wag wynikających ze schematu losowania próby z wykorzystaniem zmiennych pomocniczych tak, aby spełnione były odpowiednie równania kalibracyjne. W tradycyjnym podejściu, na etapie poszukiwania wag kalibracyjnych, zakłada się, że powinny być one bliskie — w sensie przyjętej funkcji odległości — wyjściowym wagom wynikającym ze schematu losowania próby (Särndal, Lundström, 2005).

Wiele urzędów statystycznych stosuje podejście kalibracyjne wykorzystując specjalne algorytmy w celu redukcji obciążenia i poprawy efektywności, w sensie mniejszej wariancji stosowanych estymatorów. Z doświadczeń państw stosujących kalibrację w badaniach statystycznych wynika, że jest to metoda, która może stanowić remedium na ten rodzaj błędów nielosowych.

Kalibracja jest metodą bardzo intuicyjną w zastosowaniu i może być wykorzystana zarówno w badaniach częściowych prowadzonych metodą reprezentacyjną, jak i w badaniach pełnych. Może ona przy tym nabrać szczególnego znaczenia, zwłaszcza w tych badaniach, które opierać się będą na maksymalnym wykorzystaniu danych pochodzących z różnych źródeł, w tym z rejestrów administracyjnych. Jej wykorzystanie będzie również ważne z punktu widzenia badań pełnych, w których także występuje problem braków danych. Metoda kalibracji, polegająca na korygowaniu sztucznie utworzonych wag, może stanowić cenną technikę statystyczną, wykorzystywaną w tego typu badaniach na etapie tworze-

[30] Zob.: Euredit: http://www.cs.york.ac.uk/euredit/.

nia tablic wynikowych. W konsekwencji przyczyni się to do poprawy jakości publikowanych wyników, z punktu widzenia ich dokładności, a także zapewni ich większą akceptację przez odbiorców informacji statystycznych. Należy jednak pamiętać o założeniach, na których opiera się ta metoda badawcza (Särndal, 2007).

3. Metody bootstrapowe

Od opublikowania przez Efrona (1979) metody bootstrapowe znajdują coraz szersze zastosowania w statystyce. Bootstrap stanowi wysokiej klasy technikę statystyczną opartą na tzw. ponownym próbkowaniu, którego celem jest ocena szukanej wielkości na podstawie próby, często niewielkiej i w warunkach braku dodatkowych założeń. Ponowne próbkowanie albo replikacja bootstrapowa w różnych wariantach ma na celu oszacowanie rozkładu błędów estymacji, jest przydatna szczególnie w warunkach nieznajomości rozkładu w populacji. W istocie bootstrap jest pojęciem bardzo pojemnym, rozróżnia się wiele jego typów, w tym osobne podejście do analizy szeregów czasowych (Efron, Tibshirani, 1993). W ostatnich latach stosuje się coraz powszechniej tę metodę. Można jej używać także do wyznaczania przedziałów ufności określonych parametrów[31].

Ocena całkowitego błędu badania

Osiągnięcie wysokiej jakości lub zredukowanie całkowitego błędu jest ważnym celem jakości. Modele całkowitych błędów badania stały się dla statystyków próbą pomiaru najważniejszych komponentów całkowitego błędu. Dla niektórych powolny postęp w modelowaniu całkowitego błędu badania jest zniechęcający. Smith (1976) zauważa, że: *(...) potrzebny jest systematyczny plan redukowania błędu całkowitego, jeśli mamy mieć na uwadze w badaniach ogólne zarządzanie jakością*. Ten sam autor kilkanaście lat później (1994) zauważył: *Porażka zarówno statystyków teoretyków jak i praktyków w zintegrowaniu błędów losowych i nielosowych w wyrażeniu całkowitego błędu badania nawet po 50 latach intensywnych prac badawczych, musi być uważana jako jedna z bardziej dokuczliwych porażek tej ważnej branży statystyki*.

Lessler i Kalsbeek (1992) oraz Linacre i Trewin (1993) podkreślają, że należy tak zaplanować badanie, aby zredukować „całkowite błędy badania", a nie tylko błędy losowe. Oczywiście, w celu zaplanowania badania, aby zminimalizować całkowity błąd, należy wiedzieć, jakie są główne komponenty błędów. Wynika stąd, że planowanie badania wymaga podejścia interdyscyplinarnego. Niezbędna jest wiedza i doświadczenie z zakresu statystycznej teorii badań złożonych, pla-

[31] Shao J. (2003), s. 191—198.

nowania eksperymentów, statystycznego procesu kontroli, złożonych modeli, psychologii eksperymentalnej, zarządzania i etnografii.

Potrzebę uwzględnienia całkowitego błędu badania podniósł dość wcześnie W. E. Deming (1944) rozważając błędy losowe spowodowane: wywiadem, zbieraniem danych, planowaniem kwestionariusza, brakiem odpowiedzi i pomiarem. Napisał on jeden z pierwszych podręczników z metody reprezentacyjnej (Deming, 1950), który wyraźnie podkreśla odpowiedzialność statystyka za dokładność całego badania. Należy zaznaczyć, że Deming uważany jest za ojca globalnego zarządzania jakością, a wiele pomysłów z jego prac (Deming, 1986, 1987[32]) zostało uwzględnionych przy współczesnych badaniach jakości w statystyce (*Handbook on Data…*, 2007; *Handbook on Quality…*, 2009). Globalne zarządzanie jakością wkracza również do badań reprezentacyjnych (*Handbook on improwving…*, 2013).

Uwagi końcowe

Z przeprowadzonego ogólnego przeglądu rozwoju podstawowych komponentów badań reprezentacyjnych wyraźnie otrzymujemy obraz, w jaki sposób rozwijały się te elementy w poszczególnych okresach. W dużym stopniu miał na to wpływ rozwój teorii statystycznej, zespołów naukowych, postęp techniki obliczeniowej, praktyki urzędów statystycznych, a także osobowości poszczególnych badaczy. Wraz z rozwojem ekonomicznym i społecznym, doskonaleniem techniki obliczeniowej oraz dostępności danych z różnych źródeł, zmieniało się także zapotrzebowanie na dane statystyczne, które stale wzrasta.

W wyniku globalizacji, wzrostu konkurencji, rozwoju metod zarządzania, a w szczególności globalnego zarządzania jakością, rozwija się myślenie statystyczne. Stanowi ono pewnego rodzaju pomost pomiędzy użytkownikami danych a wykorzystaniem metod statystycznych. W coraz szerszym zakresie wykorzystuje się ogromne zbiory danych przy pomocy Internetu i nowoczesnej techniki obliczeniowej, jak również zaawansowanej techniki analiz statystycznych, w której z pewnością będą stosowane metody próbkowania, modele statystyczne, symulacja, imputacja, kalibracja i metody bootstrapowe. W szerszym zakresie będą wykorzystywane metody badania jakości uzyskanych wyników. Jak już wspomniano Eurostat podjął obecnie intensywne badania w dziedzinie jakości statystyki (Eurostat, 2007, 2009), a także w zakresie jakości badań reprezentacyjnych, w których próbuje zastosować globalne zarządzanie jakością w szerokim zakresie (Eurostat, 2013).

prof. dr hab. Jan Kordos — *Główny Urząd Statystyczny*, *Wyższa Szkoła Menedżerska w Warszawie*

[32] S. 355—369.

LITERATURA

Bayes T. (1763), *An essay towards solving a problem in the doctrine of chances*, „Philosophical Transactions of the Royal Society", No. 53

Bethlehem J. G. (2009), *Applied Survey Methods, A Statistical Perspective*, John Wiley, New York

Biemer P. R., Groves L., Lyberg N., Mathiowetz and Sudman S. (Eds.) (1991), *Measurement Errors in Surveys*, John Wiley, New York

Bowley A. L. (1915), *The Nature and Purpose of the Measurement of Social Phenomena*, London, P. S. King and Son, Ltd

Bowley A. L. (1926), *Measurement of the Precision Attained in Sampling*, „Bulletin of the International Statistical Institute", Vol. 22, No. 1

Box G. E. P. (1976), *Science and Statistics*, „Journal of the American Statistical Association", Vol. 71

Chambers R. (2003), *Evaluation Criteria for Statistical Editing and Imputation*, [w:] *Methods and Experimental Results from the EUREDIT Project* (ed. J. R. H. Charlton), http://www.cs.york.uk/ /euredit

Cochran W. G. (1953), *Sampling Techniques*, 3rd ed., John Wiley, New York

Dalenius T. (1957), *Sampling in Sweden. Contributions to the Methods and Theories of Sample Survey Practice*, Uppsala

Deming W. E. (1944), *On errors in surveys*, „American Sociological Review", Vol. 9

Deming W. E. (1950), *Some Theory of Sampling*, John Wiley, New York

Deming W. E. (1986), *Out of the crisis*, MA: Massachusetts Institute of Technology

Deming W. E. (1987), *On the Statistician's Contribution to Quality*, „Bulletin of the International Statistical Institute", Proceedings of the 46th Session, No. 2

Deville J-C., Särndal C-E. (1992), *Calibration Estimators in Survey Sampling*, „Journal of the American Statistical Association", Vol. 87

Dillman D. A., Smyth J. D & Christian L. M. (2009), *Internet, Mail and Mixed-mode Surveys, The Tailored Design Method*, John Wiley, New York

Domański C., Pruska K. (2000), *Nieklasyczne metody statystyczne*, PWE, Warszawa

Efron B. (1979), *Bootstrap methods: another look at the jackknife*, „The Annals of Statistics", Vol. 7, No. 1

Efron B., Tibshirani R. J. (1993), *An introduction to the Bootstrap*, Chapman & Hall, London

Fellegi I. (1974), *An Improved Method of Estimating the Correlated Response Variance*, „Journal of the American Statistical Association", No. 69

Fischer H. (2011), *A History of the Central Limit Theorem*, Sources of Studies in the History of Mathematics and Physical Sciences, Springer Science + Business Media, LLC 2011

Groves R. M. (1989), *Survey Errors and Survey Costs*, John Wiley, New York

Groves R. M., Dillman D., Eltinge J., Little R. (2002), *Survey Nonresponse*, John Wiley, New York

Groves R. M., Fowler F., Couper Jr. M., Lepkowski J., Singer E., Tourangeau R. (2004), *Survey Methodology*, John Wiley, New York

Handbook on Data Quality Assessment: Methods and Tools (2007), Eurostat

Handbook on Quality Reports (2009*)*, Eurostat

Handbook on improving quality by analysis of process variables (2013), Eurostat

Hansen M. H., Hurwitz W. N. (1943), *On the theory of sampling from a finite population*, „Annals of Mathematical Statistics", No. 14

Hansen M. H., Hurwitz W. N. (1946), *The problem of non-response in sample surveys*, „Journal of the American Statistical Association", Vol. 17

Hansen M. H., Hurwitz W. N., Madow W. G. (1953), *Sample Survey Methods and Theory*, Vol. 1 i 2, New York

Hansen M. H., Hurwitz W. N., Bershad M. A. (1961), *Measurement errors in censuses and surveys*, „Bulletin of the International Statistical Institute", Vol. 38

Hansen M. H., Waksberg J. (1970), *Research on non-sampling errors in censuses and surveys*, „Review of International Statistical Institute", Vol. 38

Hansen M. H., Madow W. G., Tepping B. J. (1983), *An evaluation of model dependent and probability-sampling inferences in sample surveys*, „Journal of the American Statistical Association", Vol. 78

Kalton G., Kasprzyk D. (1986), *The treatment of missing survey data*, „Survey Methodology", No. 12

Kalton G., Kordos J., Platek R. (1993), *Small Area Statistics and Survey Designs*, Vol. I: *Invited Papers*; Vol. II: *Contributed Papers and Panel Discussion*, Central Statistical Office, Warsaw

Kiaer A. (1897), *The representative method of statistical surveys* (angielskie tłumaczenie z norweskiego oryginału z 1976 r.), Central Bureau of Statistics of Norway, Oslo

Kish L. (1987), *Statistical Design for Research*, John Wiley, New York

Kish L. (1996), *Stuletnie zmagania o badania reprezentacyjne*, „Wiadomości Statystyczne", nr 8

Kordos J. (1987), *Dokładność danych w badaniach społecznych*, „Biblioteka Wiadomości Statystycznych", t. 35, GUS

Kordos J. (1988), *Jakość danych statystycznych*, PWE, Warszawa

Kordos J. (2009), *Z doświadczeń nauczania statystyki matematycznej i wnioskowania statystycznego na uczelniach ekonomicznych*, „Folia Oeconomica", nr 227, „Metody ilościowe w globalnej gospodarce", Wydawnictwo Uniwersytetu Łódzkiego

Kordos J. (2011), *Professor Jerzy Neyman — some Reflections*, „Lithuanian Journal of Statistics", No. 1, www.statisticsjournal.lt

Kordos J. (2012), *Współzależność pomiędzy rozwojem teorii i praktyki badań reprezentacyjnych w Polsce*, „Przegląd Statystyczny", Numer specjalny, nr 1

Kotulski Z. (2001), *Generatory liczb losowych: algorytmy, testowanie, zastosowania*, „Matematyka stosowana", nr 2, http://www.ippt.gov.pl/~zkotulsk/MS.pdf

Kuusela V. (2011), *Paradigms in Statistical Inference for Finite Populations Up to the 1950*, „Research Reports", No. 257, Statistics Finland, http://helda.helsinki.fi/bitstream/handle/10138/27416/paradigm.pdf?sequence=3

Lessler J. T., Kalsbeek W. (1992), *Nonsampling errors in surveys*, John Wiley, New York

Linacre S. J., Trewin D. J. (1993), *Total survey design — Application to a collection of the construction industry*, „Journal of official Statistics", Vol. 9

Lyberg L. E., Kasprzyk D. (1991), *Data Collection Methods and Measurement Error: An Overview*, „Chapter", No. 13, [w:] Biemer P. P., Groves R. M., Lyberg L. E., Mathiowetz N. A., Sudman S. (eds.), *Measurement Errors in Surveys*, John Wiley, New York

Lyberg L., Biemer P., Japec L. (1998), *Quality improvement in surveys — process perspective*, Joint Statistical Meetings, American Statistical Association, Dallas

Mahalanobis P. C. (1944), *On large-scale sample surveys*, „Philosophical Transactions of the Royal Society of London", vol. 231

Mahalanobis P. C. (1952), *Some aspects of the design of sample surveys*, „Sankhya", Vol. 12

Neyman J. (1933), *Zarys teorii i praktyki badania struktury ludności metodą reprezentacyjną*, Instytut Gospodarstwa Społecznego, Warszawa

Neyman J. (1934), *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*, „Journal of the Royal Statistical Society", nr 97

Neyman J. (1938), *Contribution to the Theory of Sampling Human Population*, „Journal of American Statistical Association", nr 33

O'Muircheartaigh C., Wong S. T. (1981), *The impact of sampling theory on survey sampling practice: a review*, „Bulletin of International Statistical Institute", Invited Paper 49 (I)

Platek R., Rao J. N. K, Särndal C-E., Singh M. P. (1987), *Small Area Statistics — An International Symposium*, John Wiley, New York

Rao J. N. K. (2003), *Small area Estimation*, John Wiley, New York

Rao J. N. K. (2005), *Interplay between sample survey theory and practice: An appraisal*, „Survey Methodology", Vol. 31

Rubin D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, John Wiley, New York

Särndal C-E., Swensson B., Wretman J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Varlag

Särndal C-E. (2007), *The Calibration Approach in Survey Theory and Practice*, „Survey Methodology", Vol. 33, No. 2

Särndal C-E., Lundström S. (2005), *Estimation in Surveys with Nonresponse*, John Wiley, New York

Shao J. (2003), *Impact of the Bootstrap on Sample Surveys*, „Statistical Science", Vol. 18, No. 2

Smith T. M. F. (1976), *The foundations of Survey Sampling: a review*, „Journal of Royal Statistical Society", No. 139

Smith T. M. F. (1994), *Sample Surveys 1975—1990; an age of reconciliation?*, „International Statistical Review", No. 62

Steinhaus H. (1956), *Liczby złote i żelazne*, „Zastosowania Matematyki", z. 3

Sudman S., Brandburn N. M. (1982), *Asking questions*, San Francisco, Jossey-Bass

Sukhatme P. V., Sukhatme B. V. (1954), *Sampling Theory of Surveys with Applications*, Asia Publishing House, London

Vielrose E. (1951), *Tablice liczb losowych*, GUS

Yates F. (1949), *Sampling Methods for Censuses and Surveys*, London

Zarkovich S. S. (1966), *Quality of Statistical Data*, FAO, Rome

Zasępa R. (1958), *Problematyka badań reprezentacyjnych GUS w świetle konsultacji z prof. J. Neymanem*, „Wiadomości Statystyczne", nr 6

Zieliński R. (1972), *Generatory liczb losowych*, WNT, Warszawa

Zubrzycki S. (1966), *Wykłady z rachunku prawdopodobieństwa i statystyki matematycznej*, PWN

SUMMARY

The author discusses the importance of Jacob Bernoulli's theorem, published in 1713, for the development of sample surveys, next mentioning the many prominent mathematicians, statisticians and other scientists who developed the idea of the methods used in implementation of the modern sample surveys in the last three hundred years. He starts with showing on the basis of this theorem, why a sample from a population should be selected, using random numbers, to generalize the estimates for the whole population with a certain precision.

Next, different methods of sample selection, estimation methods, modes of data collection, using data from different sources, and methods of data analysis in the last three hundred years are generally discussed. At the end some conclusions related to data quality undertaken by Eurostat are given.

РЕЗЮМЕ

Автор обсуждает значение теоремы Якуба Бернольлего (публикация в 1713 г.) для развития выборочных обследований. Одновременно знакомит нас с многими выдающимися математиками и статистиками, которые разработали его идею используемую в выборочном методе. В дальнейшем используя эту теорему показывает, почему мы должны делать выборку, чтобы полученные результаты можно было обобщать на всю совокупность с определенной точностью. Рассматривает также основные стороны выборочных обследований, подчеркивая влияние значения методов сбора данных на качество полученных оценок.

Кроме того автор рассматривает использование в выборочных обследованиях сложных методов, а также разных моделей, которые должны учитываться в интерпретации результатов. В конце статьи были указаны выводы, касающиеся работы предпринятой Евростатом в области качества данных.