

JAN KORDOS

## WSPÓLZALEŻNOŚĆ POMIĘDZY ROZWOJEM TEORII I PRAKTYKI BADAŃ REPREZENTACYJNYCH W POLSCE

### 1. WSTĘP

Praktyczne problemy napotkane w czasie projektowania, realizacji i analizy badań reprezentacyjnych wpłynęły w znacznej mierze na rozwój teorii tych badań. Z drugiej strony, teoria badań reprezentacyjnych wpływała na praktykę tych badań, co prowadziło często do ich udoskonalenia (O'Muircheartaigh, Wong, 1981; Rao, 2005). Wydaje się więc celowe, aby prześledzić w jakim stopniu występowała współzależność pomiędzy rozwojem teorii i praktyki badań reprezentacyjnych w Polsce, uwzględniając postępy w tej dziedzinie badań na świecie.

Szeroki rozwój badań reprezentacyjnych w ostatnich latach związany jest w znacznym stopniu z rozwojem teorii, na której opierają się badania reprezentacyjne jak również z rozwojem swoistej sztuki związanej z przygotowaniem i prowadzeniem tych badań (Deming, 1950; Hansen et al., 1953; Kruskal i Mosttler, 1979-1980; Mahalanobis, 1944; O'Muircheartaigh et al., 1981; Rao, 2005; Smith, 1976, 1994). Chodzi tu nie tylko o rozwój teorii, na której opiera się metoda reprezentacyjna, ale także o inne istotne komponenty badań reprezentacyjnych, w których znaczną rolę odgrywa nie tylko rachunek prawdopodobieństwa i statystyka matematyczna, ale także prakseologia, socjologia i psychologia (Cochran, 1977; Hansen et al., 1953; Kish, 1965; Kordos, 2000a; O'Muircheartaigh et al., 1981; Särndal et al., 1992; Yates, 1980; Zająca, 1972). Dotyczy to takich komponentów badań jak: a) wybór metody zbierania danych, a w szczególności projektowanie kwestionariuszy, b) szkolenie personelu przeprowadzającego badanie, c) kontrola badania na różnych etapach jego przygotowania, realizacji i opracowania, d) badania na poszczególnych etapach źródeł błędów, ich rodzajów i rozmiarów, a także możliwości ich minimalizacji, e) metod przetwarzania danych, f) metod analizy uzyskanych wyników oraz ich prezentacji i rozpowszechniania.

Dotychczas nie ma jednolitej teorii badań statystycznych, a poszczególne komponenty projektowania badań opierają się na procedurach praktycznych uznawanych za odpowiednie w danych warunkach lub na fragmentarycznych rozwiązaniach teoretycznych, które przystosowuje się do potrzeb praktycznych (Hansen et al., 1953; Kish, 1965, 1987, 1996). Na mocnych podstawach teoretycznych opierają się plany losowania próby, ustalania potrzebnej jej liczebności w konkretnym badaniu, choć zwykle dla wybranych parametrów, wybór metod estymacji i oceny rozmiarów błędów losowych

(Bracha, 1996; Hansen et al., 1953; Smith, 1976, 1994; Zasepa, 1972). Tym zagadnieniom statystycy teoretycy i metodolodzy poświęcali najwięcej uwagi. Inne komponenty projektowania badań reprezentacyjnych opierają się w większym stopniu na sztuce niż nauce.

Z wielu zagadnień metodologicznych i tematów badań reprezentacyjnych wymienimy tu najważniejsze, które dotyczą (Smith, 1976, 1994; Kordos, 2000a): a) wyboru próby, b) metod estymacji, c) oceny błędów losowych i sposobów ich obliczenia, d) metod zbierania danych oraz e) oceny wszelkiego rodzaju błędów nielosowych.

W okresie późniejszym nabrały dużego znaczenia, takie tematy jak: (i) korekta danych i ich imputacja oraz kalibracja, (ii) koordynacja badań, (iii) estymacja dla małych obszarów, (iv) badania powtarzalne, (v) wykorzystanie komputerów. Naszą uwagę skupimy głównie na wykazaniu współzależności pomiędzy rozwojem teorii i praktyki badań reprezentacyjnych w GUS.

Za początek badań reprezentacyjnych w statystyce publicznej w Polsce uważa się rok 1933, w którym J. Neyman opublikował monografię poświęconą zastosowaniu *metody reprezentacyjnej* do przyśpieszenia wyników spisu ludności z 1931 r. (Neyman, 1933). Tak więc, badania reprezentacyjne w sensie losowego wyboru próbki zostały zapoczątkowane w Polsce przez J. Neymana (1933) oraz J. Piekałkiewicza (1934). Współzależność tutaj jest bardzo ścisła, gdyż Neyman opracował zarys teorii i praktyki, a Piekałkiewicz zaprezentował i zaanalizował uzyskane wyniki z badania.

W połowie lat dwudziestych ubiegłego wieku J. Neyman zajmował się już teorią metody reprezentacyjnej, a szczególnie wyborem małych próbek ze skończonych populacji (Neyman, 1925). Sławę światową przyniósł mu dopiero słynny artykuł opublikowany w języku angielskim w 1934 r. (Neyman, 1934), ale główne myśli tego artykułu zostały opublikowane najpierw w języku polskim w 1933 r. (Neyman, 1933). Rozważymy tu więc badania częściowe oparte na próbkach, których jednostki do badania były wybierane według *procedury probabilistycznej*. Zostaną więc pominięte badania monograficzne i inne badania częściowe, do których jednostki badania zostały wybierane w sposób niezgodny z tą procedurą badawczą.

Na początku XX wieku pod pojęciem metody reprezentacyjnej kryła się procedura wyboru próby z całej populacji, która umożliwiała uzyskanie pewnego rodzaju miniatury populacji. Próba mogła być wybrana zarówno w sposób celowy jak i losowy. Dopiero fundamentalna praca J. Neymana z 1934 r. (Neyman, 1934) położyła podstawy pod losowy wybór próby. Przedstawimy więc badania reprezentacyjne, w których jednostki badania zostały wybrane według określonej procedury probabilistycznej, umożliwiającej uogólnienie uzyskanych wyników z próby na całą populację. Metoda reprezentacyjna zdobywała sobie stopniowo uznanie w statystyce publicznej, chociaż ze znacznymi oporami (Kish, 1996). Rozwijała się więc nie tylko teoria badań reprezentacyjnych, ale także, jak już wspomniano, sztuka ich projektowania i realizacji w różnych dziedzinach badań statystycznych (Kordos, 2000a). Szersze zastosowania metody reprezentacyjnej w statystyce oficjalnej w Polsce rozpoczęły się po II wojnie światowej. Znaczną rolę odegrał tu także J. Neyman, który w czasie swych wizyt w Polsce w latach 1950

i 1958 udzielał konsultacji dla GUS w zakresie zastosowania metody reprezentacyjnej w badaniach statystycznych (Fisz, 1950a; Zasepa, 1958). Szczególną rolę odegrały tu przede wszystkim:

- a) Komisja Matematyczna GUS działająca w latach 1950-1993, w skład której wchodziłi specjaliści z ośrodków akademickich i przedstawiciele GUS (Kordos, 1975, 2012a);
- b) Pracownia Metod Matematycznych, a następnie Pracownia ds. Statystyki Matematycznej Zakładu Badań Statystyczno-Ekonomicznych GUS w latach 1966-2004;
- c) Zespół ds. badań reprezentacyjnych w Departamencie Programowania i Organizacji Badań GUS w latach 1995-2004, a od 2005 r. Wydział Operatów Statystycznych w Departamencie Metodologii, Standardów i Rejestrów GUS.

Metodologią badań reprezentacyjnych zajmowały się także w ciągu ostatnich 50-ciu lat pracownicy naukowcy wyższych uczelni oraz jednostek naukowo-badawczych, których prace dotyczyły głównie planów prób, schematów losowania, liczebności prób i ich lokalizacji, metod estymacji oraz oceny błędów losowych i nielosowych. Łącznie literatura przedmiotu jest bardzo bogata i liczy kilkaset pozycji. Z konieczności w tym artykule okolicznościowo ograniczyłem się tylko do wybranych pozycji charakteryzujących prezentowane kierunki badań i praktycznych zastosowań. Z tych względów wiele pozycji teoretycznych z metody reprezentacyjnej nie mogły być tu uwzględnione.

Szczególną uwagę poświęcę:

1. pracom teoretycznym i metodologicznym rozwijanym przez polską statystykę akademicką i w ośrodkach naukowo-badawczych, mających istotny wpływ na praktykę statystyczną,
2. badaniom powtarzalnym w czasie, a głównie *metodzie rotacyjnej*,
3. badaniom jakości danych, a szczególnie metodom prodzadającym do zwiększenia ich dokładności,
4. metodom estymacji dla małych obszarów.

Przed tym jednak przedstawię zarys rozwoju teorii i praktyki badań reprezentacyjnych na świecie, aby na tym tle przedstawić rozwój badań reprezentacyjnych prowadzonych w Polsce.

## 2. ZARYS ROZWÓJU TEORII I PRAKTYKI BADAŃ REPREZENTACYJNYCH NA ŚWIECIE

Należy podkreślić, że *Międzynarodowy Instytut Statystyczny (MIS)* uznał metodę reprezentacyjną w statystyce oficjalnej dopiero w 1925 r., a przecież już w 1891 r. w spisie ludności w Norwegii wykorzystano metodę badania częściowego, która właśnie została nazwaną *metodą reprezentacyjną*. Metoda ta była zaprezentowana na sesji Międzynarodowego Instytutu Statystycznego w Bernie w 1895 r. przez A. Kiaera (1897). Dlatego rok 1895 uważany jest jako początek badań reprezentacyjnych w statystyce oficjalnej. A. L. Bowley – pierwszy profesor statystyki w London School of Economics and Political Science – przez swoje prace o doborze próby (Bowley, 1926) oraz o pomiarze (Bowley, 1915) był pierwszą z głównych postaci rozwoju naukowe-

go doboru próby, tj. opartego na zasadach rachunku prawdopodobieństwa. Gdy dziś powszechnie uznaje się zasady losowego pobierania próby w badaniach społecznych i gospodarczych, warto zauważyć, że propozycja Kiaera na sesji MIS w 1895 r. nie była przychylnie przyjęta (Kish, 1996). 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:

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

W pierwszych dekadach XX wieku badania reprezentacyjne były przeprowadzane przy wykorzystaniu jednej z wyżej podanych metod wyboru, a do połowy lat trzydziestych ubiegłego wieku, na pytanie która z nich jest „lepsza”, nie znano odpowiedzi. Dopiero fundamentalny wkład w rozwój teorii metody reprezentacyjnej wniósł, wspomniany już, *polski statystyk Jerzy Neyman* w swoim słynnym artykule z 1934 r. (Neyman, 1934), który 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* (Neyman, 1935, 1938). Użył on terminu „reprezentatywny” w nowym znaczeniu. Warto dodać, że w literaturze anglojęzycznej termin „*metoda reprezentacyjna*” (representative method) obecnie jest rzadko używany, ale w języku polskim jest dalej dominującym. Wybór losowy w warstwach jest, według Neymana, reprezentatywny, nawet gdy frakcje losowania w poszczególnych warstwach znacznie się różnią. Uświadomienie sobie tego, że równe prawdopodobieństwo wyboru nie jest konieczne dla ważności wnioskowania, było wielkim krokiem naprzód. Neyman zalecał podejście probabilistyczne (tj. losowe) przy wyborze próby, a krytycznie odnosił się do celowego wyboru próby.

Od tego czasu teoria i praktyka badań reprezentacyjnych rozwijały się intensywnie (Kish, 1996; O’Muircheartaigh, et al., 1981; Smith, 1994; Rao, 2005). W latach 1950-tych 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 i Hurwitz, 1943, 1946; Hansen, Hurwitz i Madow, 1953; Hansen, Hurwitz i Bershad, 1961). W tym czasie obserwuje się znaczny rozwój badań reprezentacyjnych, chociaż nie zawsze zgodnie z aktualnym rozwojem teorii. Niejednokrotnie praktyka wyprzedzała teorię (Kruskal i Mosteller, 1979-1980; O’Muircheartaigh, et al., 1981; Smith, 1994). Dotyczy to szczególnie początków lat siedemdziesiątych. Trudno tu ocenić wkład w tym okresie poszczególnych autorów, dlatego ograniczę się do omówienia ważniejszych kierunków prac badawczych.

Wprawdzie badania reprezentacyjne na szerszą skalę rozpoczęły się po 1945 r., to jednak w pierwszej połowie XX stulecia warto także odnotować kilka faktów. Należy wspomnieć o badaniach prowadzonych i opisanych przez A.L. Bowleya’a (1913, 1926)

w Anglii, a także badania bezrobocia prowadzone w latach 1934-1942 w USA (Frankel i Stock, 1942), które w 1943 r. przekształciły się w sławne badania siły roboczej prowadzone przez Biuro Spisów USA.

Po 1945 r. w wielu krajach nastąpił intensywny rozwój badań reprezentacyjnych. W tym czasie ukazało się 5 podręczników z metody reprezentacyjnej w języku angielskim: Yates (1949), Deming (1950), Cochran (1953), Hansen et al. (1953) oraz Sukhatme i Sukhatme (1970). Po ukazaniu się wymienionych podręczników zorganizowano wiele kursów z metody reprezentacyjnej w USA, Wielkiej Brytanii, w Indiach i wielu innych krajach.

W USA badania prowadzone przez Biuro Spisów przekształciły się w Bieżące Badania Ludności (*Current Population Surveys*), które pod kierunkiem M.H. Hansena wykształciło wielu specjalistów w dziedzinie badań reprezentacyjnych na poziomie krajowym i międzynarodowym.

Instytut Statystyczny w Indiach, kierowany przez Mahalanobisa (1944, 1952), stał się drugim centrum badań reprezentacyjnych o znaczeniu międzynarodowym. Indyjskie Narodowe Badania Reprezentacyjne (*Indian National Sample Surveys*) zdobyły sławę międzynarodową (Sukhatme i Sukhatme, 1970).

Do rozpowszechnienia się badań reprezentacyjnych, szczególnie w krajach rozwijających, przyczyniło się FAO (Rzym) i ONZ (Nowy Jork), które podjęły prace już w 1950 r. (O'Muircheartaigh et al., 1981, Zarkovich, 1966). Z ramienia FAO pracowałem jako ekspert i konsultant z badań reprezentacyjnych w Etiopii, Chińskiej Republice Ludowej oraz Nepalu, a w latach 200-2005 byłem konsultantem UN Statistics Division z zakresu badań reprezentacyjnych.

Obecnie większość urzędów statystycznych prowadzi badania reprezentacyjne, chociaż ich rozwój jest dość zróżnicowany z punktu widzenia tempa i jakości.

Dość zróżnicowany rozwój badań reprezentacyjnych był w ośrodkach uniwersyteckich. Warto tu wspomnieć o dwóch ośrodkach: Centrum Badań Reprezentacyjnych (*Sample Surveys Center*) na uniwersytecie Iowa (*Iowa State University*), zorganizowane przez Snedecora i Cochran (1977), oraz Centrum Badawcze na Uniwersytecie Michigan (*Survey Research Center*) kierowane przez L. Kisha (Kish, 1965, 1996, 1987).

W aspekcie międzynarodowym znaczną rolę odgrywają także sesje naukowe *Międzynarodowego Instytutu Statystycznego (MIS)*, które odbywają się co dwa lata. Na sesjach tych prezentowane są zarówno osiągnięcia teoretyczne jak i praktyczne używane w tej dziedzinie przez ośrodki badawcze różnych krajów, a także przez urzędy statystyczne. 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 poświęconych różnym aspektom badań reprezentacyjnych. W Polsce organizacja ta była współorganizatorem międzynarodowej konferencji poświęconej metodom estymacji dla małych obszarów w 1992 r. (Kalton et al. 1993).

W ostatnim okresie Eurostat podjął intensywne prace nad integracją i harmonizacją badań reprezentacyjnych prowadzonych w statystyce społecznej i ekonomicznej

(Eurostat, 1996, 1997). Poświęca się wiele uwagi ocenie jakości uzyskanych wyników oraz sposobom ich prezentacji (Eurostat, 2007, 2009).

Zostanie dalej przedstawiony, w ogólnym zarysie, rozwój badań reprezentacyjnych w Polsce, rozpoczynając od prac metodologicznych prowadzonych w ośrodkach akademickich i naukowo-badawczych oraz ich wpływ na praktykę badań reprezentacyjnych w GUS. Wspomnę o ważniejszych badaniach reprezentacyjnych prowadzonych przez GUS, rozważę także niektóre aspekty badań reprezentacyjnych, takie jak *badania powtarzalne*, *jakość uzyskanych wyników* oraz *metody estymacji dla małych obszarów*.

### 3. WPLYW NIEKTÓRYCH STATYSTYKÓW NA ROZWÓJ TEORII I PRAKTYKI BADAŃ REPREZENTACYJNYCH W POLSCE W POCZĄTKOWYM OKRESIE

W Polsce rozwijano zarówno teorię jak i praktykę badań reprezentacyjnych. W początkowym okresie szczególny wpływ na rozwój tych badań miały prace badawcze kilku naukowców z kraju i z zagranicy. Znany jest wpływ Jerzego Neymana, który był kilkakrotnie omawiany przy różnych okazjach (Fisz, 1950a; Kordos, 2009; Zasepa 1958). Znacznie mniej znany jest wpływ profesorów: Marka Fisz i Hugona Steinhausa, którzy oddziaływali w pierwszym okresie na rozwój badań reprezentacyjnych. Istotny wpływ mieli także niektórzy statystycy z innych krajów, z których należy wymienić, moim zdaniem, przede wszystkim Deminga (1950), Daleniusa (1957), Yatesa (1949), Hansena et al. (1953), Cochran (1977) Kisha (1965) i Zarkowicha (1966).

#### 3.1. WPLYW PROFESORA MARKA FISZA

Oprócz Jerzego Neymana, znaczny wpływ na rozwój badań reprezentacyjnych w Polsce w początkowym okresie miał także Marek Fisz, który współpracował z GUS oraz brał udział w konsultacjach z J. Neymanem i opisał ich wyniki (Fisz, 1950a). W Komisji do Spraw Statystyki Matematycznej GUS powołanej zarządzeniem wewnętrznym Prezesa GUS, której przewodniczył prof. S. Szulc, Marek Fisz pełnił funkcję sekretarza Komisji<sup>1</sup>. Następnie brał intensywny udział w szkoleniach pracowników GUS z metody reprezentacyjnej, w których były wykorzystane jego programy szkoleniowe, skrypty, a następnie jego podręcznik pt. „*Rachunek prawdopodobieństwa i statystyka matematyczna*”. Profesor Marek Fisz miał także pewien wpływ na podjęcie w Głównym Urzędzie Statystycznym badań w zakresie jakości danych statystycznych (Fisz, 1950b).

#### 3.2. WPLYW PROFESORA HUGONA STEINHAUSA

Szczególne znaczenie i wpływ na podjęcie przez GUS badań w zakresie jakości danych oraz badań reprezentacyjnych miał profesor Hugo Steinhaus. Muszę tu odwo-

<sup>1</sup> Zarządzenie wewnętrzne Prezesa GUS Nr 87 z dnia 31 grudnia 1949 r. w sprawie powołania Komisji do Spraw Statystyki Matematycznej.

łać się do własnych doświadczeń w tym zakresie. W latach 1953-1955 studiowałem na Uniwersytecie Wrocławskim, a profesor Steinhaus prowadził dla nas wykłady ze *statystycznej kontroli jakości (SKJ)* oraz seminarium z rachunku prawdopodobieństwa. Uczestniczyłem w tym czasie również w seminariach prowadzonych przez prof. H. Steinhausa i prof. J. Perkala w Instytucie Matematycznym PAN we Wrocławiu nt. *zastosowań przyrodniczych i gospodarczych matematyki*. Pod kierunkiem prof. H. Steinhausa napisałem pracę magisterską pt. „*Przedziały ufności i tolerancji dla rozkładu normalnego*”, którą obroniłem w czerwcu 1955 r. Prof. H. Steinhaus miał wpływ na trzy zagadnienia związane z badaniami GUS: a) badanie jakości danych, b) losowanie próbek do badań, oraz c) wycenę statystyczną, które przedstawię w skrócie.

### 3.2.1. BADANIE JAKOŚCI DANYCH STATYSTYCZNYCH

Jak już wspomniałem, na wykładach ze *statystycznej kontroli jakości*, prof. H. Steinhaus niejednokrotnie podkreślał, że metody te mogą być stosowane także w badaniach statystycznych w celu zbadania ich dokładności, gdyż wyniki tych badań mogą być obciążone różnego rodzaju błędami. Starłem się przenieść pewną wiedzę zdobytą na studiach w mojej pracy w GUS, którą rozpocząłem we wrześniu 1955 r., a współpracowałem dodatkowo również z Komisją Matematyczną GUS, na której przedstawiłem swoje uwagi nt. badania jakości w statystyce w oparciu o wykłady prof. H. Steinhausa ze *statystycznej kontroli jakości* i mojej pracy z tego zakresu. Dlatego Komisja Matematyczna GUS skierowała mnie dodatkowo do współpracy z wydziałem kontroli badań statystycznych GUS, a dotyczyło to głównie kontroli jakości sprawozdawczości statystycznej. Współpraca ta trwała kilka lat, a następnie zająłem się także jakością innych badań statystycznych prowadzonych przez GUS, które przedstawię w syntetycznym ujęciu w dalszej części artykułu.

### 3.2.2. LOSOWANIE PRÓBEK – WYKORZYSTANIE TABLIC LICZB LOSOWYCH

W badaniach reprezentacyjnych występuje problem uzyskania próbki możliwie najbardziej losowej. Wykorzystuje się tu różne mechanizmy losowania. Pomocnym narzędziem w tym względzie są *tablice liczb losowych*, które prof. Steinhaus także konstruował. Dyskutowaliśmy niejednokrotnie ten problem na seminarium prowadzonym przez Profesora. Starając się zaradzić rozmaitym mankamentom związanym z wyborem próbki, Profesor projektował zbudowanie tablicy zawierającej wszystkie liczby naturalne od 0000 do 9999, każdą dokładnie jeden raz, a powstałej przez wymieszanie tych liczb za pomocą odpowiedniego i wyraźnie określonego algorytmu. Jeden taki dość zawiły algorytm został użyty do zbudowania tablicy liczb „*przetasowanych*”. Inny algorytm, korzystający ze złotego podziału odcinka, a dokładniej z reszty modulo 1 wielokrotności liczby złotej  $a = (\sqrt{5} - 1)/2 = 0,618\dots$ , został wykorzystany do zbudowania tablic liczb „*złotych*” i „*żelaznych*” (zob. Steinhaus, 1956). Właśnie „*tablice liczb żelaznych*” (TLŻ) zbudowane przez H. Steinhausa zostały wykorzysta-

wane w badaniach reprezentacyjnych prowadzonych przez GUS. Zastosowanie TLŻ w badaniach GUS zaproponował także prof. J. Neyman w czasie swojej konsultacji dla GUS w 1958 r. zamiast losowania systematycznego (zob. Zasępa, 1958). Uważaliśmy, że TLŻ są szczególnie przydatne w losowaniu sekwencyjnym, tj. gdy, z różnych przyczyn, należy dalej kontynuować losowanie próbki. Losowanie sekwencyjne stosowaliśmy, gdy założyliśmy określoną liczebność próby, a na skutek różnych przyczyn nie uzyskania odpowiedzi, trzeba było dalej kontynuować losowanie. Dotyczyło to badań budżetów gospodarstw domowych i innych badań społecznych. Stosowałem także TLŻ w badaniach rolniczych, gdy w latach 1974-1980 pracowałem w Etiopii jako ekspert FAO z zakresu badań reprezentacyjnych.

### 3.2.3. WYCENA STATYSTYCZNA – DOKŁADNOŚĆ METOD ZBIERANIA DANYCH STATYSTYCZNYCH

Prof. H. Steinhaus w czasie wykładów na temat statystycznej kontroli jakości podkreślał, że metody SKJ mogą być stosowane także do określenia jakości badań masowych, takich jak spisy ludności, sprawozdawczość statystyczna i badania statystyczne, w których metody zbierania danych statystycznych, z różnych powodów, nie zapewniają uzyskanie dokładnych danych. W takim przypadku proponował zastosowanie tzw. *wyceny statystycznej*, która, przeciętnie rzecz biorąc, da dokładniejsze wyniki niż metoda wywiadu indywidualnego. Jako przykład wymieniał uzyskanie informacji o dochodach gospodarstw rolnych lub pracujących na rachunek własny. Respondenci nie są w stanie podać dokładnej informacji, gdyż faktycznie jej nie znają lub z innych przyczyn podają informacje nieprawdziwe. Podają więc informacje obciążone, znacznie odbiegające od stanu faktycznego. Proponował zbudowanie pewnego rodzaju modelu uwzględniającego poszczególne komponenty związane z oszacowaniem dochodów danej jednostki lub zbiorowości. Następnie komponenty te byłyby szacowane na podstawie dostępnych dokumentów lub nawet przez rzeczoznawców. Wiedzieliśmy, że dochody podawane przez respondentów są znacznie zaniżane. Podejście Profesora próbowałem zastosować przy badaniu rozkładów dochodów ludności według wysokości zatrudnionej w gospodarce pozarolniczej (Kordos, 1959, 1963). Budowałem modele wykorzystujące dane z badań budżetów gospodarstw domowych, z badania rozkładów płac według wysokości, które w tym okresie były badaniami pełnymi, z danych z ZUS oraz ze spisów ludności. Podejmowaliśmy w Pracowni Regionalnej Zakładu Badań Statystyczno-Ekonomicznych GUS we Wrocławiu próby zastosowania metody wyceny statystycznej do oceny dochodów w gospodarstwach rolnych, ale prace w tym zakresie nie zostały zakończone. Dopiero po 2000 r. dowiedziałem się, że prof. H. Steinhaus zastosował wycenę statystyczną jako metodę odbioru towarów produkcji masowej, stosując zasadę „minimax” (zob. Steinhaus, 2000). Wydaje mi się, że podejście to można by zastosować w rozwijanym ostatnio globalnym zarządzaniu jakością (TQM), które wkracza również stopniowo do statystyki (Kordos, 2001b).



### 3.3. WPŁYW INNYCH STATYSTYKÓW

Oprócz wymienionych wyżej polskich statystyków na rozwój badań reprezentacyjnych w Polsce w pierwszym okresie poważny wpływ wywarło kilku statystyków, z których przede wszystkim należy wymienić Deminga (1950), Yatesa (1949), Hansena et al., (1953), Daleniusa (1957), Kisha (1965), Cochрана (1977) i Zarkovicha (1966). Osiągnięcia tych statystyków były intensywnie studiowane i dyskutowane na zebraniach i seminariach organizowanych przez Komisję Matematyczną GUS oraz Pracownię Metod Matematycznych ZBS-E GUS. Szczególną uwagę poświęciliśmy pracy Deminga (1950, 1987), a głównie jego podejściu do projektowania badań reprezentacyjnych i odpowiedzialności statystyka za projektowanie próbki w całości kształcie badania reprezentacyjnego (Kordos, 2009). Zauważyliśmy, że podejście to było zbliżone do podejścia prof. H. Steinhausa, który również kładł nacisk na jakość danych, sposoby wyboru próbki oraz metody zbierania informacji. Intensywnie studiowaliśmy także książkę Daleniusa (1957) dotyczącą teorii i praktyki badań reprezentacyjnych w Szwecji. Chodziło tu głównie o optymalną lokalizację próbki ze względu na kilka cech. Interesujące wyniki w tej dziedzinie uzyskał J. Greń (1964, 1966, 1969, 1970). Należy tu podkreślić, że w sprawie zastosowania optymalnej lokalizacji próbki w praktyce napotkaliśmy w tym okresie na pewną barierę ze strony opracowujących wyniki badań. Twierdzili, że występują trudności techniczne zastosowania proponowanych metod, tłumacząc to „*zmiennym przecinkiem*”. Nie mogliśmy więc naszych osiągnięć teoretycznych w tym zakresie zastosować w praktyce. Problem ten został rozwiązany dopiero po wprowadzeniu w szerokim zakresie elektronicznego przetwarzania danych. Ze względów dydaktycznych cieszyły się popularnością podręczniki Cochрана (1977) i Kisha (1965), które były wykorzystywane zarówno w naszych pracach badawczych jak i szkoleniowych. Prof. S. Zarkovich (1966) przyczynił się w znacznym stopniu do pogłębienia badań w zakresie jakości danych (Kordos, 1973, 1987, 1988). W latach 1974-1980 pracowałem w Etiopii jako ekspert FAO, a prof. Zarkovich był moim przełożonym. Kładł szczególny nacisk na metody badania jakości danych statystycznych, a następnie uzyskane doświadczenie w tym zakresie przenieśliśmy w mojej pracy w GUS.

### 4. DZIAŁALNOŚĆ KOMISJI MATEMATYCZNEJ GUS

Jak już podkreśliłem, szczególną rolę w rozwoju badań reprezentacyjnych w statystyce publicznej miała Komisja Matematyczna GUS<sup>2</sup>, która działała w latach 1950-1993, jako organ doradczy i opiniodawczy Prezesa GUS w zakresie wdrażania i stosowania do praktyki statystycznej metod matematyczno-statystycznych, a w szczególności

<sup>2</sup> Pierwsza Komisja została powołana zarządzeniem wewnętrznym Prezesa GUS Nr 87 z dnia 31 grudnia 1949 r. pod nazwą „*Komisja do Spraw Statystyki Matematycznej*” w składzie: prof. Stefan Szulc – przewodniczący, mgr Marek Fisz – sekretarz, dr Egon Vielrose, mgr Ryszard Zasępa – członkowie. Faktycznie działalność Komisji rozpoczęła się w 1950 r. i stopniowo rozszerzała zakres działania i skład osobowy.

metody reprezentacyjnej. Komisji Matematycznej GUS, w różnych okresach, przewodniczyli znani w Polsce statystycy: prof. Stefan Szulc, prof. Zbigniew Pawłowski, prof. Władysław Welfe i prof. Ryszard Zasepa. Stopniowo Komisja rozszerzała zakres swej działalności i od 1969 r. była Komisją Matematyczną międzyresortową, służącą nie tylko badaniom GUS, ale także innym resortom<sup>3</sup>.

Wyniki działalności Komisji Matematycznej GUS były często publikowane w wydawnictwach GUS, a głównie w serii „*Biblioteka Wiadomości Statystycznych*”. Publikacje te dotyczyły: a) zastosowania metod matematycznych w statystyce (GUS, 1969); b) metodologii badań reprezentacyjnych (GUS, 1970b, 1971a, 1971b, 1972, 1978, 1987b, 1989); c) wybranych problemów prognoz statystycznych (GUS, 1970a, 1979) d) zastosowania metod ekonometrycznych (GUS, 1976, 1979); e) perspektyw rozwoju statystyki GUS (1973, 1987a). Artykuły indywidualne były zwykle publikowane w „*Wiadomościach Statystycznych*” lub „*Przeglądzie Statystycznym*”. W latach 1966-1969, jako sekretarz naukowy Komisji, publikowałem na łamach „*Wiadomości Statystycznych*” informacje pt. „*Z prac Komisji Matematycznej GUS*”<sup>4</sup>. W „*Przeglądzie Statystycznym*” omówiłem działalność Komisji za okres 25-lecia (Kordos, 1975). Znaczny udział w pracach Komisji Matematycznej GUS w latach 1972-1993 oraz w rozwoju badań reprezentacyjnych w Polsce, miał mgr Bronisław Lednicki, który w latach 1990-1993 pełnił funkcje sekretarza naukowego Komisji. Był autorem wielu ważnych badań reprezentacyjnych prowadzonych przez GUS (Lednicki, 1973, 1974, 1979, 1982, 1987, 1989). Najdłuższą funkcję przewodniczącego Komisji pełnił prof. Ryszard Zasepa, który wniósł największy wkład w rozwój badań reprezentacyjnych w Polsce, a także organizował intensywne szkolenia pracowników GUS z zakresu badań reprezentacyjnych.

W Komisji Matematycznej GUS pełniłem różne funkcje od członka Komisji, sekretarza naukowego, zastępcy przewodniczącego do przewodniczącego Komisji w latach 1990-1993<sup>5</sup>. Szerzej działalność Komisji Matematycznej GUS w latach 1950-1993 przedstawiałem w opublikowanym ostatnio artykule na łamach „*Wiadomości Statystycznych*” (Kordos, 2012 a).

## 5. ZARYS ROZWÓJU BADAŃ REPREZENTACYJNYCH W POLSCE

Pomimo podjęcia dość wczesnie prac w zakresie zastosowania metody reprezentacyjnej w Polsce, bo już w pierwszej połowie lat trzydziestych ubiegłego wieku, to jednak początkowo wykorzystanie tej metody badań w naszej praktyce statystycznej

<sup>3</sup> Zarządzenie, nr 38 Prezesa Głównego Urzędu Statystycznego z dnia 10 czerwca 1969 r. (znak: I-2-0200-38) w sprawie powołania Komisji Matematycznej przy Głównym Urzędzie Statystycznym.

<sup>4</sup> Patrz: J. Kordos, *Z prac Komisji Matematycznej GUS*, *Wiadomości Statystyczne*, z. 9, 1966, s. 20-22; z. 10, s. 22-24.; z. 12, s. 14-16; 1967, z. 3, s. 25-27; z. 6, s. 18-19; z. 9, s. 12-13; 1969, z. 5, s. 21-22.

<sup>5</sup> Zarządzenie Nr 31 Prezesa Głównego Urzędu Statystycznego a dnia 4 lipca 1990 r. w sprawie powołania Komisji Matematycznej „*Dziennik Urzędowy Głównego Urzędu Statystycznego*”, Nr 12 (172), s. 333-334.

było dość ograniczone. Przed II wojną światową zastosowanie metody reprezentacyjnej ograniczyło się do przyspieszenia opracowania wyników spisu ludności z 1931 r. (Neyman, 1933; Piekałkiewicz, 1934). Zastosowania metody reprezentacyjnej w badaniach GUS po ostatniej wojnie były konsultowane, jak już wspomniałem, z prof. J. Neymanem w czasie jego wizyt w Polsce w latach 1950 i 1958. Brałem udział w konsultacjach w 1958 r. w ramach Komisji Matematycznej GUS, które trwały przez 6 tygodni. Było to dla mnie wielkim przeżyciem naukowym, a także jako najmłodszy z zespołu Komisji prowadziłem protokoły i dokładne notatki oraz miałem bezpośrednie kontakty z Profesorem w przygotowaniu i ocenie spotkań. Z prof. R. Zasępą podziwialiśmy nie tylko wielką wiedzę prof. Neymana i jego oryginalne podejście do rozwiązywania problemów statystycznych, ale także ogromną pracowitość. Niejednokrotnie kończyliśmy pracę w późnych godzinnych wieczornych, gdy była taka potrzeba. Swoje refleksje z tych spotkań przedstawiłem ostatnio w artykule w języku angielskim (Kordos, 2011).

#### 5.1. BADANIA REPREZENTACYJNE DO 1989 ROKU<sup>6</sup>

W poprzednim systemie społeczno-gospodarczym, badania statystyczne były głównie oparte o pełną sprawozdawczość statystyczną, a w ograniczonym stopniu wykorzystywano badania częściowe. Metodę reprezentacyjną stosowano głównie w badaniach społecznych (badania budżetów gospodarstw domowych i badania warunków życia), a także w badaniach rolniczych. Badania te zostały opisane w kilku publikacjach Głównego Urzędu Statystycznego (GUS, 1978, 1987a, 1987b, 1989). Polscy statystycy rozwijali metodologię badań reprezentacyjnych w ośrodkach akademickich oraz jednostkach naukowo-badawczych GUS. Powstało wiele interesujących opracowań naukowych, które w znacznym stopniu zostały wykorzystane w praktyce. Prace te wymagają specjalnego opracowania monograficznego, a tu wymienię tylko niektóre z nich.

Metoda reprezentacyjna była w Polsce twórczo rozwijana, wykładana na uczelniach o profilu ekonomicznym, a w 1962 r. został opublikowany w Polsce pierwszy podręcznik metody reprezentacyjnej (Zasępa, 1962), a później dalsze książki poświęcone tej metodzie badawczej (Pawłowski, 1972; Zasępa, 1972; Steczkowski, 1988; Bracha, 1996; Szreder, 2004; Wywiół, 2010). W 1987 r. została opublikowana moja monografia nt. *dokładności danych w badaniach społecznych* (Kordos, 1987), a w 1988 r. nt. *jakości danych statystycznych* (Kordos, 1988). Z inicjatywy Komisji Matematycznej GUS, pod kierunkiem jej przewodniczącego, prof. R. Zasępy, prowadzono dla pracowników GUS intensywne szkolenia z metody reprezentacyjnej, co przyczyniło się do lepszego zrozumienia tej metody badania statystycznego w praktyce. Powstanie w 1966 r. Zakładu Badań Statystyczno-Ekonomicznych GUS (ZBS-E GUS) umożliwiło podjęcie prac naukowo-badawczych, wdrożenie a także doskonalenie tej metody badawczej w praktyce. Próbowaliśmy także przekazywać nasze doświadczenia w tym zakresie innym krajom, organizując w 1970 r. konferencję międzynarodową (GUS, 1971a i 1971b).

<sup>6</sup> Podano także publikacje wydane po 1989 roku.

Z publikacji powstałych w ośrodkach akademickich oraz naukowo-badawczych, które miały istotny wpływ na praktykę statystyczną, należałoby tu wymienić prace R. Zasepy (1958, 1962, 1968, 1972, 1993ab), Cz. Brachy (1972, 1994, 1996, 1998, 2003), Cz. Brachy et al. (2004a, 2004b), J. Grenia (1964, 1966, 1969, 1970), B. Lednickiego (1973, 1974, 1979, 1982, 1987, 1989), B. Lednickiego et al., (1994, 2003.), J. Kordosa (1967, 1968, 1971, 1974, 1982, 1985, 1991), L. Kursy, B. Lednickiego (2006), J. Pawłowskiej (1969), A. Szarkowskiego, J. Witkowskiego (1994) oraz J. Wywiśla (1988, 1995, 1996, 1998).

Polscy specjaliści z badań reprezentacyjnych byli współorganizatorami i brali aktywny udział w trzech międzynarodowych konferencjach, które odbyły się w Polsce na początku okresu przejściowego do gospodarki rynkowej:

1) w 1991 r. nt. *pomiaru ubóstwa w okresie przemian gospodarczych w krajach Europy Wschodniej (poverty measurement for economies in transition in Eastern European Countries)*, która odbyła się w Jachrance k. Warszawy w dn. 7-9 października 1991 r. (GUS, 1992);

2) w 1992 r. nt. *problemów statystyki małych obszarów i planowania badań (small area statistics and survey design)*, która odbyła się w Warszawie w dn. 30 września – 3 października 1992 r. (Kalton et al., 1993);

3) w 1994 r. *międzynarodowa konferencja statystyczna dla upamiętnienia setnej rocznicy urodzin Jerzego Neymana (International Conference in Memory of the Hundredth Anniversary of the Birth of Jerzy Neyman)*, która odbyła się w dn. 25-26 listopada 1994 r. w Jachrance k. Warszawy.

W czasie tych międzynarodowych konferencji oraz spotkań z ekspertami Eurostatu i innych organizacji międzynarodowych okazało się, że polscy statystycy specjalizujący się w metodyce badań reprezentacyjnych byli dość dobrze przygotowani do przekształcenia polskiej statystyki do wymagań gospodarki rynkowej.

## 5.2. BADANIA REPREZENTACYJNE PO 1989 ROKU

Od 1989 r. rozpoczęto w GUS dostosowywanie statystyki publicznej do międzynarodowych standardów, a także podjęto starania zharmonizowania jej ze statystyką Unii Europejskiej (Kordos, 1998). Zostały zaadaptowane podstawowe klasyfikacje i nomenklatury, program badań został w znacznym stopniu zmieniony, zastosowano nowe metody i metodologie badań dostosowane do wymagań gospodarki rynkowej i standardów europejskich GUS (1995).

Przerwano większość sprawozdawczości statystycznej, a rozpoczęto nowe badania reprezentacyjne. Dostosowano *badania budżetów gospodarstwach domowych* do wymagań Eurostatu (Eurostat, 1997; GUS, 2011a) oraz podjęto nowe badania, takie jak: *badania aktywności ekonomicznej ludności* (Popiński, 2006; Szarkowski, Witkowski, 1994), *badania przedsiębiorstw, koniunktury w przemyśle, budownictwie i handlu* (Barczyk et al., 1994), *badania w rolnictwie* (Kordos, Kursy, 1997; Kursy, Lednicki, 2006), *badania zdrowia* (Bracha, 1998; GUS, 2006, 2011b), *badania warunków życia* (GUS,

2002, 2005a) i badania wykorzystania czasu (GUS, 2005b). Od 2005 r. wprowadzono Europejskie Badanie Dochodów i Warunków Życia (EU-SILC) przygotowane według koncepcji Eurostatu (GUS, 2008; Verma & Betti, 2006). Studiowano i wdrożono do praktyki metody estymacji precyzji dla złożonych schematów losowania (Kordos, Szarkowski, 1973; Popiński, 2006; Szarkowski, Witkowski, 1994; Kordos, Zięba-Pietrzak, 2010). Należy tu także wymienić prace z zakresu lokalizacji próby i statystyki małych obszarów, które wskazały kierunki doskonalenia praktyki statystycznej (np. Kozak, 2004, 2006; Kubacki, 2004, 2006; Niemiro et al., 2001, 2002, 2012).

### 5.3. REPREZENTACYJNE BADANIA POWTARZALNE W CZASIE – METODA ROTACYJNA

Większość badań statystyki publicznej są badaniami powtarzalnymi w różnych okresach. Próbuje się tu stosować badania panelowe, w których wybrana jednostka badania bierze udział w badaniu w kilku kolejnych rundach, gdy głównym celem badania jest ocena zmian w kolejnych okresach. Jednakże w większości badań chodzi zarówno o badania zmian w czasie jak również o oceny w poszczególnych okresach. Kompromisem jest tu *metoda rotacyjna*, której na świecie, a także w Polsce, poświęcono wiele uwagi. Prace teoretyczne rozpoczęto w tej dziedzinie już na początku lat pięćdziesiątych ubiegłego wieku (Patterson, 1950). Z późniejszych prac warto tu wymienić prace Ecklera (1955), Rao i Grahama (1964), Kaltona i Kasprzaka (1987), Kaltona i Citro (1993).

W Polsce problematyce badań powtarzalnych poświęcono wiele uwagi, a szczególnie metodzie rotacyjnej, którą próbowano zastosować w kilku badaniach, a głównie w badaniach budżetów gospodarstw domowych (GUS, 1972, 1986, 1989; Kordos, 1967, 1971, 1982; Lednicki, 1982), w badaniu aktywności ekonomicznej ludności (Szarkowski i Witkowski, 1994; Popiński, 2006) oraz warunków życia GUS (2008). Należy także wspomnieć o pracach teoretycznych w zakresie metody rotacyjnej podjętych w ostatnich latach przez kilku statystyków (np. Ciepiera et al., 2012; Kowalczyk, 2003; Kowalski, 2006, 2009; Kowalski, Wesołowski, 2011; Wesołowski, 2010). Szerszego przeglądu prac teoretycznych w zakresie badań reprezentacyjnych w czasie, a szczególnie metody rotacyjnej, prowadzonych w Polsce w ostatnich latach, a także ich praktycznych zastosowań, przedstawiłem w opracowaniu monograficznym (Kordos, 2012b).

### 5.4. PROBLEMY JAKOŚCI DANYCH W BADANIACH REPREZENTACYJNYCH

Prace nad jakością danych statystycznych podjęto w GUS dość wcześnie, bo już w końcu lat 1950-tych. Znaczny wpływ, jak już wspomniałem, miał tu prof. Hugo Steinhaus, którego propozycje w tym zakresie próbowano zastosować w statystyce publicznej. Pierwsze wyniki prac z zakresu analizy błędów losowych i nielosowych w Polsce przedstawiłem na sesji naukowej Międzynarodowego Instytutu Statystycznego w Wiedniu w 1973 r. (Kordos, 1973). Problematykę jakości danych statystycznych

przedstawiłem później w dwóch monografiach (Kordos, 1987, 1988) i w kilku artykułach (Kordos, 1995, 2001a, 2007).

W ostatnich latach na świecie wiele uwagi poświęcano jakości danych statystycznych, a w ostatnim okresie – *jakości statystyki*. Wiąże się to ściśle ze stale rosnącym zapotrzebowaniem na rzetelne informacje statystyczne, które wykorzystywane są w szerokim zakresie w różnych dziedzinach życia gospodarczego i społecznego, a także do porównań międzynarodowych.

Szersze zainteresowanie jakością statystyki, a nie tylko danymi statystycznymi wynika także z rewolucji jakościowej w społeczeństwie. W przedsiębiorstwach przemysłowych, a także w innych jednostkach organizacyjnych wprowadzane jest stopniowo *globalne zarządzanie jakością*<sup>7</sup> (*Total Quality Management – TQM*), które obejmuje wszystkie procesy i jednostki organizacyjne oddziałujące na jakość produkcji lub usług. Organizacje statystyczne nie stanowią tu wyjątku. Przecież produkty statystyczne – dane statystyczne, ich interpretacje i analizy, wykorzystywane wykresy i diagramy, publikacje statystyczne oraz udzielane konsultacje w zakresie statystyki, podlegają podobnym zasadom i regułom procesów masowych jak produkty przemysłowe czy usługi.

Problematyką jakości danych zajmowało się kilku naszych statystyków (Kordos, 1973, 1987, 1988; 2001a, 2004, 2007; Szutkowska, 2012; Wierchosławski, 1995; Zasepa, 1962, 1968, 1972, 1993ab).

W ostatnich latach wiele uwagi poświęca się problematyce jakości w statystyce. Ograniczę się tu tylko do wspomnienia ostatnich prac prowadzonych pod patronatem Eurostatu. Prace te dotyczą udoskonalenia jakości Europejskiego Systemu Statystycznego. Ważne jest, że pojęcie jakości w statystycznych organizacjach zmieniło się w czasie ostatniej dekady. Tak, więc dokładność nie jest już jedyną miarą jakości. W takim podejściu jakość może być zdefiniowana w wielu wymiarach, gdzie dokładność jest jednym z nich Eurostat (2007; 2009). GUS próbuje stopniowo uwzględnić te zalecenia (Szutkowska, 2012). W publikacjach GUS, w których prezentowane są wyniki z badań reprezentacyjnych, podawane są zwykłe błędy standardowe lub względne błędy standardowe dla wybranych parametrów. Z błędów nielosowych podawane są niekiedy frakcje realizacji badania, tj. iloraz liczby zbadanych jednostek do liczby jednostek wylosowanych, lub frakcje braku odpowiedzi. Błędy odpowiedzi są rzadko badane, a próbę oszacowania błędów tego rodzaju podejmuje się tylko w kontrolnych badaniach spisów ludności. Jednakże oceny jakości ostatnich spisów ludności, w których zastosowano pospisowe badania kontrolne, tj. w latach 1978, 1988 i 2002, nie zostały opublikowane. Należy mieć nadzieję, że analiza jakości spisów z tych badań zostanie wkrótce opublikowana, łącznie z oceną jakości Narodowego Spisu Ludności 2011. W sprawie oceny jakości spisów ludności została ostatnio opublikowana specjalna monografia przez UN Statistics Division (2010).

<sup>7</sup> Warto zauważyć, że prof. W.E. Deming uważany jest za ojca TQM, który miał znaczny wpływ na rozwój badań reprezentacyjnych w Polsce w początkowym okresie. Omówiłem ten fakt w moim artykule Kordos (2009).

### 5.5. ESTYMACJA DLA MAŁYCH OBSZARÓW A BADANIA REPREZENTACYJNE

Problem ocen wartości parametrów dla małych obszarów występował w statystyce już od początku prowadzenia badań częściowych. Dla badań częściowych, prowadzonych metodą reprezentacyjną, w których jednostki próby losowano według określonego schematu, powstaje problem rzetelności ocen, szczególnie, gdy są one oparte na niewielkiej liczbie badanych jednostek. Niekiedy precyzja takich ocen jest bardzo niska i ich wartość informacyjna znikoma. Dotyczy to zwykle małych obszarów, takich jak podregiony, powiaty, miasta, a nawet niektórych województw, gdy próbka, na podstawie, której szacowane są wyniki, jest zbyt mała.

Wprawdzie metodami ocen dla małych obszarów zajmowano się już od dość dawna, to jednak szersze zainteresowania tymi problemami wśród statystyków obserwuje się w ciągu ostatnich 40 lat. Zaczęły ukazywać się prace metodyczne i naukowe w tym zakresie, a także rozpoczęto organizowanie międzynarodowych konferencji naukowych i seminariów, na których rozważano różne problemy związane z estymacją parametrów dla małych obszarów. Warto tu wymienić międzynarodowe sympozjum zorganizowane w 1985 r. w Ottawie (Platek, et. al., 1987), międzynarodowe konferencje naukowe, które odbyły się w 1992 r. w Warszawie (Kalton et al., 1993) i w Rydze (1999). Od 2005 roku organizowane są co dwa lata międzynarodowe konferencje poświęcone metodom estymacji dla małych obszarów, a dotychczas odbyły się konferencje: w Finlandii (2005), we Włoszech (Piza, 2007), w Hiszpanii (Elche, 2009) oraz w Niemczech (Trier, 2011). Wiele artykułów prezentowanych na tych konferencjach opublikowano w naszym międzynarodowym czasopiśmie w języku angielskim *Statistics in Transition*. W latach 2002-2005 polscy statystycy, pod kierunkiem prof. J. Paradysza i prof. E. Gołaty, uczestniczyli w międzynarodowym konsorcjum Eurarea (2005), sponsorowanym przez Eurostat. Wymienię tu tylko niektóre publikacje z tego zakresu wydanych w Polsce: Dehnel, (2010); Domański, Pruska (2001); Gołata (2004a, 2004b, 2012); Kalton, et al. (1993); Kordos (1991, 1999, 2004); Kordos, Paradysz (2000); Paradysz (1998); Pekasiewicz, Pruska (2002); Wesołowski (2004). Jednakże, do wdrożenia tej metody badawczej do praktyki statystycznej na szerszą skalę niezbędne są dalsze prace badawcze (Gołata, 2012).

### 5.6. MIĘDZYNARODOWE KONFERENCJE W KATOWICACH POŚWIĘCONE METODZIE REPREZENTACYJNEJ

Wiele interesujących opracowań teoretycznych i metodologicznych z metody reprezentacyjnej prezentowali polscy statystycy na różnych konferencjach międzynarodowych. Szczególne znaczenie dla rozwoju badań reprezentacyjnych w Polsce miały międzynarodowe konferencje organizowane przez Katedrę Statystyki Uniwersytetu Ekonomicznego w Katowicach, pod kierunkiem prof. Janusza Wywiśla i jego współpracowników pt: *Metoda reprezentacyjna w badaniach ekonomiczno-społecznych (Survey Sampling in Economic and Social Research)*.

Wykładowcami na tych konferencjach byli często światowej sławy eksperci, jak np. prof. Malay Ghosh z USA, prof. Jean-Claude Deville z Francji i prof. Yves Tillé ze Szwajcarii. Dotychczas zorganizowano 7 konferencji międzynarodowych, a wyniki konferencji z lat 2008 i 2009 zostały wydane w publikacjach zwartych (patrz Wywiał, Żądło, 2009; Wywiał, Gamrot, 2010). Ostatnia konferencja odbyła się w dn. 18-20 września 2011 r. Należy mieć nadzieję, że konferencje te będą w przyszłości kontynuowane.

#### 5.7. REPREZENTACYJNY SPIS LUDNOŚCI 2011

Największym badaniem reprezentacyjnym w dziejach GUS był przeprowadzony w ramach Narodowego Spisu Powszechnego 2011 *reprezentacyjny spis ludności* (GUS, 2012). Spis reprezentacyjny został przeprowadzony na próbie losowej w ok. 20% mieszkań w skali kraju. Jednostką losowania było mieszkanie, a dokładniej jego adres. Operat losowania, który został utworzony w oparciu o rejestry i systemy informacyjne, został odpowiednio uporządkowany i podzielony na poszczególne warstwy.

Podstawowym celem spisu reprezentacyjnego realizowanego w ramach NSP 2011 było uzyskanie informacji o sytuacji społeczno-demograficznej na poziomie powiatu. Konieczne było więc dokonanie podziału założonej 20%-owej próby mieszkań dla Polski pomiędzy powiaty. Dokonano tego przy wykorzystaniu metody alokacji pierwiastkowej. Metoda ta stanowi kompromis pomiędzy alokacją proporcjonalną, a alokacją zapewniającą jednakową precyzję dla subpopulacji. Przy założeniu, że zastosowano by proporcjonalne losowanie próby, w każdym powiecie próba stanowiłaby 20% populacji. Ponieważ precyzja wyników tj. wielkość błędu losowego, zależy od liczebności próby, uzyskano by w efekcie niedostateczną precyzję dla wielu małych powiatów. Z kolei, w metodzie alternatywnej uzyskujemy, w przybliżeniu, jednakową precyzję wyników dla wszystkich powiatów, ale za cenę istotnego „spłaszczenia” liczebności próby. W efekcie liczebność próby, a tym samym pracochłonność przy realizacji spisu byłaby mało zróżnicowana pomiędzy dużymi i małymi powiatami. Z tych powodów jako metodę rozdziału próby przyjęto alokację pierwiastkową, w której np. liczba mieszkań losowanych w p-tym powiecie jest proporcjonalna do pierwiastka kwadratowego z populacyjnej liczby mieszkań. W celu wylosowania, w każdym z powiatów, próby o ustalonej wcześniej liczebności zastosowany został schemat losowania jednostopniowego warstwowego. Jednostki losowania – mieszkania, zostały przed losowaniem pogrupowane w warstwy w celu zwiększenia efektywności losowania. Zastosowano zróżnicowane podejście do warstwowania w zależności od typu powiatu i gminy. Schemat losowania próby oraz ogólny zarys metody estymacji wyników zostały przedstawione w publikacji GUS (2012).

*Spis kontrolny do NSP 2011* przeprowadzono od 1 do 11 lipca 2011 r. Jego celem było sprawdzenie kompletności przeprowadzonego spisu, poprawności danych uzyskanych w spisie oraz ich zgodności ze stanem faktycznym. Spis kontrolny przeprowadzono w formie spisu reprezentacyjnego. Spośród 2744 tys. mieszkań, które wcześniej zostały wylosowane do spisu reprezentacyjnego, wylosowano 80 tys. mieszkań. Spisem



kontrolnym objęto mieszkania, w których respondenci dokonali samospisu przez Internet, mieszkania spisane bezpośrednio przez rachmistrzów spisowych lub telefonicznie przez ankieterów jak i takie, w których spis nie został przeprowadzony (z różnych powodów). Ze względu, jak twierdzą autorzy, na zapisy ustawowe spis kontrolny został przeprowadzony przez ankieterów poprzez telefon (metodą CATI). Spowodowało to objęcie spisem kontrolnym wyłącznie tych mieszkań, w których przynajmniej dla jednej osoby udało się ustalić numer telefonu, przy czym nie miało znaczenia, czy był to telefon stacjonarny, czy komórkowy. Formularz do spisu kontrolnego zawierał 14 pytań. Zwykle w badaniu kontrolnym stosuje się dokładniejszą metodę zbierania danych niż w spisie podstawowym. Wydaje się jednak wątpliwe, aby metoda wywiadu telefonicznego umożliwiła sprawdzenie dokładności udzielanych odpowiedzi w badaniu podstawowym (UN Statistics Division, 2010). Jednakże analizę krytyczną uzyskanych wyników należy przeprowadzić i wyciągnąć stąd odpowiednie wnioski na przyszłość.

## 6. UWAGI KOŃCOWE

Przedstawiłem tu w bardzo ogólnym zarysie, współzależność pomiędzy rozwojem teorii i praktyki badań reprezentacyjnych w Polsce. Zdaje sobie sprawę z subiektywizmu mojej oceny, gdyż w badaniach tych uczestniczyłem jako praktyk, zarówno w kraju jak i za granicą, ponad 50 lat. Uważam, że w okresie tym polscy statystycy, zarówno akademicy jak i oficjalni, wnieśli istotny wkład w światowy rozwój badań reprezentacyjnych, pomimo znacznych trudności występujących w poszczególnych okresach. Brali aktywny udział w międzynarodowym życiu statystycznym, prezentując swój dorobek na międzynarodowych konferencjach, seminariach lub biorąc udział w konsorcjach, a także w różnych projektach międzynarodowych.

Wydaje się, że największym wyzwaniem jest stale rosnące zapotrzebowanie na różnego rodzaju rzetelne dane statystyczne w różnorodnych układach, porównywalne w czasie i przestrzeni. Powstaje poważny problem w jaki sposób można zaspokoić te potrzeby biorąc pod uwagę rozwój teorii badań statystycznych, dotychczasową praktykę statystyczną i istniejące różnego rodzaju rejestry administracyjne i inną dostępną dokumentację. Chodzi tu nie tylko o porównywalność między poszczególnymi krajami i regionami w układach europejskich zalecanych przez Eurostat, OECD i ONZ, ale także w układach dla poszczególnych krajów. Danych tych przecież nie można uzyskać z badań pełnych oraz istniejących rejestrów administracyjnych, więc niezbędne staje się szersze wykorzystanie badań reprezentacyjnych.

Badania te, jak wykazano powyżej, odegrały dotychczas ogromną rolę, jednakże wystąpiło wiele zaniedbań i uchybień, które powinny być stopniowo usuwane. Chodzi tu przede wszystkim o jakość tych badań, przystosowanie ich do analiz przekrojowych i longitudinalnych, a także o zwiększenie ich precyzji dla małych obszarów. Te zagadnienia nie zostały dotychczas dostatecznie rozwiązane. We wszystkich krajach obserwuje się ograniczenia środków finansowych na badania statystyczne i wydaje się, że tendencja ta utrzyma się dalej.

Poważnym dylematem jest w jaki sposób zaspokoić wymagania użytkownika. Przecież perspektywa *użytkownika* jest zupełnie inna niż statystyka. Użytkownik interesuje się udostępnionymi wynikami z badania, jeśli one spełniają jego osobiste pojęcia o jakości i użyteczności. Będzie on stawiał następujące pytania: (a) czy wyniki są porównywalne lub zgodne z informacjami z innych podobnych źródeł danych? (b) czy informacja jest odpowiednia i aktualna? (c) czy jest porównywalna z poprzednimi obserwacjami? (d) czy mogę zaufać procedurze pomiarowej, korektom ze względu na brak odpowiedzi i procedurze estymacji wyników? (e) czy dostępne dane są przygotowane w sposób łatwo zrozumiały, solidny i użyteczny dla moich celów?

Użytkownicy chcą *zaufać* do jakości danych statystycznych im dostarczanych. Zaufanie może być rozszerzone przez dostarczenie użytkownikom obiektywnych mierników dotyczących ważnych aspektów planu i procesu badania. Wykorzystuje się wiele mierników. Niektóre urzędy statystyczne podają w pewnych badaniach określone mierniki jakości danych. Praktyka w tym zakresie, tj. informowania użytkownika o jakości danych, jest różnorodna. Powstaje więc pytanie, czy ujednoczyć informacje o badaniu dla potrzeb użytkownika? Wiadomo, że za wiele numerycznych szczegółów może być dla użytkownika raczej kłopotliwe niż objaśniające. Jaką praktykę przyjąć w tym zakresie? Na te pytania trzeba odpowiedzieć, ale nie ma gotowych recept. Należy podkreślić, że metodologia badań statystycznych, w tym badań reprezentacyjnych, jest podstawową częścią ogólnej efektywności urzędów statystycznych (Eurostat, 2007, 2009; Fellegi, 1999). Reputacja urzędu statystycznego zależy w poważnym stopniu od solidności metodologii statystycznej. Słysz się niekiedy opinie, że w czasach ograniczeń budżetowych nie powinno się rozwijać metodologii badań, gdyż wiele z nowych badań nie będzie prowadzonych. Niejednokrotnie można się było przekonać, że ulepszenie metody badania może być ważnym czynnikiem ogólnej efektywności urzędu. Może to dotyczyć lepszego projektowania badania oraz udoskonalenia narzędzi pomiaru i przetwarzania. Podczas gdy metodologię można zniszczyć w przeciągu kilku miesięcy, to odbudowanie właściwej zdolności metodologicznej urzędu zabierze lata a nawet dekady (Fellegi, 1999). Główny Urząd Statystyczny powinien odpowiedzieć na te pytania. Podjęte prace w zakresie globalnego zarządzania jakością statystyki, zgodnie z koncepcją Eurostatu, powinny być transparentne dla wszystkich statystyków i użytkowników danych. Wiadomo, że rok 2013 będzie obchodzony jako *Międzynarodowy Rok Statystyki*, okazja ta powinna być wykorzystana do wzmocnienia naszej statystyki publicznej, zawodu statystyka, a szczególnie zwiększenia publicznej świadomości znaczenia i wpływu statystyki na wszystkie aspekty życia społecznego. Jakość badań reprezentacyjnych powinna być udoskonalana w całokształcie systemu statystyki publicznej.

*Szkoła Główna Handlowa w Warszawie*  
*Wyższa Szkoła Menedżerska w Warszawie*

#### LITERATURA

- [1] Barczyk A., Gaca I., Zagoździńska I., (1994), *Badanie koniunktury, Zeszyty metodyczne i klasyfikacje GUS*, Warszawa.

- [2] Bowley A.L., (1913), Working Class Households in Reading, *Journal of the Royal Statistical Society*, 672-691.
- [3] Bowley A.L., (1915), *The Nature and Purpose of the Measurement of Social Phenomena*, P.S. King and Son, Ltd, London.
- [4] Bowley A.L., (1926), Measurement of the Precision Attained in Sampling, *Bulletin of the International Statistical Institute*, 22 (1), 1-62.
- [5] Bracha Cz., (1972), Reprezentacyjne badanie struktury zapasów w uspołecznionym handlu detalicznym, *Wiadomości Statystyczne*, 1, 21-25.
- [6] Bracha Cz., (1994), Metodologiczne aspekty badania małych obszarów, *Z prac Zakładu Badań Statystyczno-Ekonomicznych GUS*, 43.
- [7] Bracha Cz., (1996), *Teoretyczne podstawy metody reprezentacyjnej*, PWN, Warszawa.
- [8] Bracha Cz., (1998), Schemat losowania próby i metoda estymacji w badaniu ankietowym „stan zdrowia ludności”, *Wiadomości Statystyczne*, 3, 9-15.
- [9] Bracha Cz., (2003), *Estymacja danych z Badania Aktywności Ekonomicznej Ludności na poziomie powiatów dla lat 1995-2002*, GUS, Warszawa.
- [10] Bracha Cz., Jakubowski J., Szarkowski A., (2004a), Analiza porównawcza estymatorów regresyjnych w reprezentacyjnych badaniach statystycznych, *Seria Studia i Prace – Z prac Zakładu Badań Statystyczno-Ekonomicznych GUS i PAN*, 292, Warszawa.
- [11] Bracha Cz., Lednicki B., Wieczorkowski R., (2004b), Wykorzystanie złożonych metod estymacji do dezagregacji danych z Badania Aktywności Ekonomicznej Ludności w roku 2003, *Studia i Prace – Z prac Zakładu Badań Statystyczno-Ekonomicznych GUS i PAN*, 299, Warszawa.
- [12] Ciepela P., Gniado K., Wesołowski J., Wojty M., (2012), Dynamic K-Composite estimator for an arbitrary rotation scheme, *Statistics in Transition-new series*, 13 (1), 7-20.
- [13] Cochran W.G., (1977), *Sampling Techniques*, 3rd ed., Wiley, New York.
- [14] Dalenius T., (1957), *Sampling in Sweden. Contributions to the Methods and Theories of Sample Survey Practice*. Uppsala.
- [15] Dehnel G., (2010), *Rozwój mikroprzedsiębiorczości w Polsce w świetle estymacji dla małych domen*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań.
- [16] Deming W.E., (1950), *Some Theory of Sampling*, Wiley, New York.
- [17] Deming W.E., (1987), On the Statistician's Contribution to Quality, *Bulletin of the International Statistical Institute, Proceedings of the 46th Session*, 2, 355-369.
- [18] Domański Cz., Pruska, K., (2001), *Metody statystyki małych obszarów*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- [19] Eckler A.R., (1955), Rotation Sampling, *Annals of Mathematical Statistics*, 26, 664-685.
- [20] Eurostat (1996), *The Future of European Social Statistics: Use of Administrative Registers and Dissemination Strategies*, Luxembourg.
- [21] Eurostat (1997), *Family Budget Surveys in the EC: Methodology and Recommendations for Harmonisation*, Population and Social Conditions 3, Methods E, Luxembourg.
- [22] Eurostat (2007), *Handbook on Data Quality Assessment: Methods and Tools*, Luxembourg.
- [23] Eurostat (2009), *Handbook on Quality Report*, Luxembourg.
- [24] Fellegi I., (1999), Statistical Services – Preparing for the Future, *Survey Methodology*, 25 (2), 113-128.
- [25] Fisz M., (1950a), Konsultacje prof. Neymana i wnioski z nich wypływające, *Studia i Prace Statystyczne*, 3-4, 14-27.
- [26] Fisz M., (1950b), Kontrola jakości produkcji masowej na cechę ciągłą, *Studia i Prace Statystyczne*, 2, 123-160.
- [27] Frankel L.R., Stock J.S., (1942), On the sample surveys of employment, *Journal of the Statistical American Association*, 37, 77-80.
- [28] Gołata E., (2004a), *Pośrednia estymacja bezrobocia na lokalnym rynku pracy*, Wydawnictwo Poznańskiej Akademii Ekonomicznej, Prace habilitacyjne, 11, Poznań.

- [29] Gołata E., (2004b), Problems of Estimate Unemployment for Small Domains in Poland, *Statistics in Transition*, 6 (5), 755-776.
- [30] Gołata E., (2012), Data integration and small domain estimation in Poland – experiences and problems, *Statistics in Transition-new series*, 13 (1), 107-142.
- [31] Greń J., (1964), O pewnych metodach wyznaczania lokalizacji próby w losowaniu warstwowym wieloparametrycznym, *Przegląd Statystyczny*, 3, 361-369.
- [32] Greń J., (1966), O pewnym zastosowaniu programowania nieliniowego do metody reprezentacyjnej, *Przegląd Statystyczny*, 3, 361-369.
- [33] Greń J., (1969), Efektywność powtarzalnych badań metodą reprezentacyjną, *Biblioteka Wiadomości Statystycznych*, 7, 97-104.
- [34] Greń J., (1970), Wielowymiarowy estymator regresyjny średniej dla skończonej populacji, *Przegląd Statystyczny*, 1, 73-78.
- [35] GUS (1969), Zastosowanie metod matematycznych w statystyce, *Biblioteka Wiadomości Statystycznych*, 7, Warszawa.
- [36] GUS (1970a), Wybrane problemy prognoz statystycznych, *Biblioteka Wiadomości Statystycznych*, 11, Warszawa.
- [37] GUS (1970b), Statystyczna ocena wyników badań budżetów rodzinnych, *Studia i Prace Statystyczne*, Warszawa.
- [38] GUS (1971a), Badania statystyczne metodą reprezentacyjną w krajach socjalistycznych, *Biblioteka Wiadomości Statystycznych*, 14, Warszawa.
- [39] GUS (1971b), Wybrane problemy metodologiczne badań reprezentacyjnych, *Biblioteka Wiadomości Statystycznych*, 15, Warszawa.
- [40] GUS (1972), Eksperymentalne badania budżetów rodzinnych metodą rotacyjną, *Biblioteka Wiadomości Statystycznych*, 18, Warszawa.
- [41] GUS (1973), Stan i perspektywy rozwoju statystyki w Polsce, *Biblioteka Wiadomości Statystycznych*, 17, Warszawa.
- [42] GUS (1976), Statystyka i ekonometria w Polsce Ludowej, *Biblioteka Wiadomości Statystycznych*, 25, Warszawa.
- [43] GUS (1978), Metodologia badań reprezentacyjnych w GUS – Prace Komisji Matematycznej, *Biblioteka Wiadomości Statystycznych*, 29, Warszawa.
- [44] GUS (1979), Statystyka i ekonometria w Polsce Ludowej (wydanie drugie – zmienione), *Biblioteka Wiadomości Statystycznych*, 31, Warszawa.
- [45] GUS (1986), Metodyka i organizacja badań budżetów gospodarstw domowych, *Zeszyty Metodologiczne*, 62, Warszawa.
- [46] GUS (1987a), Problemy integracji statystycznych badań gospodarstw domowych, *Biblioteka Wiadomości Statystycznych*, 34, Warszawa.
- [47] GUS (1987b), Zastosowanie metody reprezentacyjnej w badaniach statystycznych GUS (1981-1986), *Z prac Zakładu Badań Statystyczno-Ekonomicznych GUS*, 166, Warszawa.
- [48] GUS (1989), Problemy badań statystycznych metodą reprezentacyjną, *Biblioteka Wiadomości Statystycznych*, 36, Warszawa.
- [49] GUS (1992), *Poverty Measurement for Economies in Transition in Eastern European Countries*, Polish Statistical Association, Warsaw.
- [50] GUS (1995), Rozwój metodologii badań statystycznych w Polsce, *Biblioteka Wiadomości Statystycznych*, 44, Warszawa.
- [51] GUS (1998), Metodologia i organizacja mikrospisów, *Statystyka w Praktyce*, Warszawa.
- [52] GUS (2002), Warunki życia ludności w 2001 r., *Studia i Analizy Statystyczne*, Warszawa.
- [53] GUS (2005a), Warunki życia ludności w 2004 r., *Studia i Analizy Statystyczne*, Warszawa.
- [54] GUS (2005b), Budżet czasu ludności, *Studia i Analizy Statystyczne*, Warszawa.
- [55] GUS (2006), Stan zdrowia ludności Polski w 2004 r., *Informacje i Opracowania Statystyczne*, Warszawa.

- [56] GUS (2008), *Incomes and Living Conditions of the Population in Poland (report from the EU-SILC survey of 2006)*, Warsaw.
- [57] GUS (2011a), Metodologia badania budżetów gospodarstw domowych, *Zeszyty Metodologiczne*, Warszawa.
- [58] GUS (2011b), Stan zdrowia ludności Polski w 2009 r., *Informacje i Opracowania Statystyczne*, Warszawa.
- [59] GUS (2012), *Raport wyników, Narodowy Spis Powszechny Ludności i Mieszkań 2011*, Warszawa.
- [60] Hansen M.H., Hurwitz W.N., (1943), On the theory of sampling from a finite population, *Annals of Mathematical Statistics*, 14, 333-362.
- [61] Hansen M.H., Hurwitz W.N., (1946), The problem of non-response in sample surveys, *Journal of the American Statistical Association*, 41, 517-529.
- [62] Hansen M.H., Hurwitz W.N., Bershad M.A., (1961), Measurement errors in censuses and surveys, *Bulletin of the International Statistical Institute*, 38, 359-374.
- [63] Hansen M.H., Hurwitz W.N., Madow W.G., (1953), *Sample Survey Methods and Theory*, Vol. I i II, Wiley, New York.
- [64] Kalton G., Citro C.F., (1993), Panel surveys: adding the fourth dimension, *Survey Methodology*, 19, 205-215.
- [65] Kalton G., Kasprzyk D., (1986), The treatment of missing survey data, *Survey Methodology*, 12, 1-16.
- [66] 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.
- [67] Kiaer A., (1897), The representative method of statistical surveys (angielskie tłumaczenie z norweskiego oryginału z 1976 r., *Central Bureau of Statistics of Norway*, Oslo.
- [68] Kish L., (1965), *Survey Sampling*, New York.
- [69] Kish L., (1987), *Statistical Design for Research*, John Wiley and Sons, New York.
- [70] Kish L., (1996), Stulecie zmagania o badania reprezentacyjne, *Wiadomości Statystyczne*, 8, 3-16.
- [71] Kordos J., (1959), Szacunek rozkładu ludności według grup zamożności, *Wiadomości Statystyczne*, 3, 4-8.
- [72] Kordos J., (1963), Rozkład ludności pozarolniczej według wysokości dochodów na osobę w 1960 r., *Biuletyn Komitetu Przestrzennego Zagospodarowania Kraju PAN*, 8.
- [73] Kordos J., (1967), Metoda rotacyjna w badaniach reprezentacyjnych, *Przegląd Statystyczny*, 4, 373-394.
- [74] Kordos J., (1968), Zastosowanie metody reprezentacyjnej w statystyce rolniczej, *Przegląd Statystyczny*, 3, 251-264.
- [75] Kordos J., (1971), Eksperymentalne badania budżetów rodzinnych metodą rotacyjną, *Przegląd Statystyczny*, 3-4, 227-246.
- [76] Kordos J., (1973), On Analysis of Sampling and Non-sampling Errors in Official Statistics in Poland. *Proceeding of the 40th Session of the International Statistical Institute*, XLV, Vienna, 609-616.
- [77] Kordos J., (1974), Reprezentacyjne badania warunków bytu ludności, *Przegląd Statystyczny*, 2, 195-209.
- [78] Kordos J., (1975), 25 lat działalności Komisji Matematycznej GUS, *Przegląd Statystyczny*, 1, 171-173.
- [79] Kordos J., (1982), Metoda rotacyjna w badaniach budżetów rodzinnych w Polsce. *Wiadomości Statystyczne*, 9, 1-6.
- [80] Kordos J., (1985), Towards an Integrated System of Household Surveys in Poland, *Bulletin of the International Statistical Institute, (invited paper), vol. 51, Amsterdam, Book 2, 1.3.1-18.*
- [81] Kordos J., (1987), Dokładność danych w badaniach społecznych, *Biblioteka Wiadomości Statystycznych, GUS*, 35, Warszawa.
- [82] Kordos J., (1988), *Jakość danych statystycznych*, PWE, Warszawa.

- [83] Kordos J., (1991), Statystyka małych obszarów a badania reprezentacyjne, *Wiadomości Statystyczne*, 4, 1-5.
- [84] Kordos J., (1993), Refleksje nad rozwojem metod badania budżetów gospodarstw domowych, *Wiadomości Statystyczne*, 7, 15-20.
- [85] Kordos J., (1995), Metodologiczne problemy kontroli badań statystycznych, W: *Rozwój metodologii badań statystycznych w Polsce*, Biblioteka Wiadomości Statystycznych, 44, 45-56, Warszawa.
- [86] Kordos J., (1997), 40 lat badań budżetów gospodarstw domowych w Polsce, *Wiadomości Statystyczne*, 7, 27-42.
- [87] Kordos J., (1998), Social Statistics in Poland and its Harmonisation with the European Union Standards, *Statistics in Transition*, 3 (4), 617-639.
- [88] Kordos J., (1999), Problemy estymacji danych dla małych obszarów, *Wiadomości Statystyczne*, 1, 85-101.
- [89] Kordos J., (2000a), Teoria i sztuka badań reprezentacyjnych, *Wiadomości Statystyczne*, 1, 1-13.
- [90] Kordos J., (2000b), Nowy projekt zastosowania estymacji dla małych obszarów, *Wiadomości Statystyczne*, 8, 1-10.
- [91] Kordos J., (2001a), Some Data Quality Issues in Statistical Publications in Poland, *Statistics in Transition*, 5 (3), 475-489.
- [92] Kordos J., (2001b), Globalne zarządzanie jakością wkracza do statystyki, *Kwartalnik Statystyczny*, nr 2, 9-12.
- [93] Kordos J., (2004), Niektóre aspekty jakości w statystyce małych obszarów, W: A. Zeliaś (red.), *Tradycje i obecne zadania statystyki w Polsce*, 95-124, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- [94] Kordos J., (2007), Some Aspects of Post-Enumeration Surveys in Poland, *Statistics in Transition-new series*, 8 (3), 563-576.
- [95] Kordos J., (2009), Impact of J. Neyman and W.E. Deming on Sample Survey Practice in Poland, The third Conference of the European Survey Research Association, Warsaw, 2009, <http://www.europeansurveyresearch.org/sites/default/files/abstracts.pdf>, Chapter II.
- [96] Kordos J., (2011), Professor Jerzy Neyman – some reflections, *Lithuanian Journal of Statistics*, 2011, No. 1, s. 112-122. [www.statisticsjournal.lt](http://www.statisticsjournal.lt).
- [97] Kordos J., (2012a), Działalność Komisji Matematycznej GUS w latach 1950-1993, *Wiadomości Statystyczne*, 9, 11-25.
- [98] Kordos J., (2012b), Application of Rotation Methods in Sample Surveys in Poland, *Statistics in Transition-new series*, 13 (2), 47-64.
- [99] Kordos J., Szarkowski A., (1974), Metoda oceny precyzji podstawowych estymatorów stosowanych w badaniach budżetów rodzinnych, *Wiadomości Statystyczne*, 5, 18-20.
- [100] Kordos J., Kurska L., (1997), Metoda reprezentacyjna w rolniczych badaniach statystycznych w Polsce, *Wiadomości Statystyczne*, 9, 12-34.
- [101] Kordos J., Paradysz J., (2000), Some Experiments in Small Area Estimation in Poland, *Statistics in Transition*, 4 (4), 963-977.
- [102] Kordos J., Zięba-Pietrzak A., (2010), Development of Standard Error Estimation Methods in Complex Household Sample Surveys in Poland, *Statistics in Transition-new series*, 11(2), 231-252.
- [103] Kowalczyk B., (2003), Estimation of the population total on the current occasion under second stage unit rotation pattern, *Statistics in Transition*, 6 (4), 503-513.
- [104] Kowalski J., (2006), Rotation in sampling patterns, in review for *Journal of Statistical Planning and Inference*, 2006.
- [105] Kowalski J., (2009), Optimal Estimation in Rotation Pattern, *J. Statist. Plann. Infer.*, 139 (4), 2429-2436.
- [106] Kozak M., (2004), Multivariate Sample Allocation Problem in Two Schemes of Two-Stage Sampling, *Statistics in Transition*, 6 (7), 1047-1054.

- [107] Kozak M., (2006), Multivariate Sample Allocation: Application of Random Search Method, *Statistics in Transition*, 7 (4), 889-900.
- [108] Kruskal W.H., Mosteller F., (1979-1980), Representative Sampling I, II, III i IV. *International Statistical Review*, 1895-1939.
- [109] Krzysztofik W., (1939), Gospodarstwa karłowate w świetle ankiety losowej, Warszawa.
- [110] Kubacki J., (2004), Application of the Hierarchical Bayes Estimation to the Polish Labour Force Survey, *Statistics in Transition*, 6 (5), 785-796.
- [111] Kubacki J., (2006), Remarks on Using the Polish LFS Data for Unemployment Estimation by County, *Statistics in Transition*, 7 (4), 901-916.
- [112] Kurza L., Lednicki, B., (2006), The agricultural sample surveys in Poland in transition period, *Statistics in Transition*, 7(5), 981-1008.
- [113] Lednicki B., (1973), Zastosowanie metody reprezentacyjnej w badaniu składników wynagrodzeń w przemyśle, *Wiadomości Statystyczne*, 12, 25-26.
- [114] Lednicki B., (1974), Badanie efektywności losowania zespołowego na podstawie Narodowego Spisu Powszechnego 1970 roku, *Z prac Zakładu Badań Statystyczno-Ekonomicznych GUS*, 78.
- [115] Lednicki B., (1979), Zastosowanie metody reprezentacyjnej w kwartalnych spisach pogłowia zwierząt gospodarskich, W: Metodologia badań reprezentacyjnych w GUS. Prace Komisji Matematycznej, *Biblioteka Wiadomości Statystycznych*, 29, 35-38.
- [116] Lednicki B., (1982), Schemat losowania i metoda estymacji w rotacyjnym badaniu budżetów gospodarstw domowych, *Wiadomości Statystyczne*, 9, 7-15.
- [117] Lednicki B., (1987), Schemat losowania „próby-matki” do Zintegrowanego Systemu Badań Gospodarstw Domowych w Polsce. W: Problemy integracji badań gospodarstw domowych. *Biblioteka Wiadomości Statystycznych*, 34, 269-279.
- [118] Lednicki B., (1989), Badanie efektywności różnych schematów losowania przy wykorzystaniu informacji o cechach dodatkowych, *Wiadomości Statystyczne*, 1, 18-22.
- [119] Lednicki B., Wesołowski J., (1994), Lokalizacja próby pomiędzy subpopulacje, *Wiadomości Statystyczne*, 9, 2-4.
- [120] Lednicki B., Wieczorkowski R., (2003), Optimal Stratification and Sample Allocation Between Sub-population and Strata, *Statistics in Transition*, 6 (2), 287-305.
- [121] Mahalanobis P.C., (1944), On large-scale sample surveys, *Philosophical Transactions of the Royal Society of London*, 231, 329-451.
- [122] Mahalanobis P.C., (1952), Some aspects of the design of sample surveys. *Sankhya*, 12, 1-7.
- [123] Neyman J., (1925), Contributions to the Theory of Small Samples Drawn from a Finite Population, *Biometrika*, 17.
- [124] Neyman J., (1933), Zarys teorii i praktyki badania struktury ludności metoda reprezentacyjna, *Instytut Spraw Społecznych*, Warszawa.
- [125] 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*, 97, 558-625.
- [126] Neyman J., (1935), On the Problem of Confidence Intervals, *The Annals of Mathematical Statistics*, 6, 111-116.
- [127] Neyman J., (1938), Contribution to the Theory of Sampling Human Population, *Journal of American Statistical Association*, 33, 101-116.
- [128] Niemirow W., Wesołowski J., (2001), Fixed Precision Optimal Allocation in Two-Stage Sampling, *Applicationes Mathematicae*, 28 (1), 73-82.
- [129] Niemirow W., Popiński W., Wesołowski J., Wieczorkowski R., (2002), Optymalna alokacja próby w warunkach niekompletnej realizacji badania, *Wiadomości Statystyczne*, 9, 1-9.
- [130] Niemirow W., Wesołowski J., (2012), Linear estimation and prediction under model-design approach with small area effects, *Statistics: A Journal of Theoretical and Applied Statistics*, 46 (4), 523-547.

- [131] 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 (1), 437-446.
- [132] Paradysz J., (1998), Small Area Statistics in Poland – First Experiences and Application Possibilities, *Statistics in Transition*, 3 (5), 1003-1015.
- [133] Patterson H.D., (1950), Sampling on successive occasions with partial replacements of units, *Journal of the Royal Statistical Society*, ser. B, 12, 241-255.
- [134] Pawłowska J., (1969), Efektywność metod estymacji i schematów losowania w badaniach pogłowia zwierząt gospodarskich, *Z prac Zakładu Badań Statystyczno-Ekonomicznych*, GUS, 15, Warszawa.
- [135] Pawłowski Z., (1972), *Wstęp do statystycznej metody reprezentacyjnej*, PWN, Warszawa
- [136] Pekasiewicz D., Pruska K., (2002), Analysis of Distribution of Some Estimators in Small Area Statistics, *Folia Oeconomica*, 156, 91-111.
- [137] Piekalkiewicz J., (1934), Sprawozdanie z badań składu ludności robotniczej w Polsce metodą reprezentacyjną, *Instytut Spraw Społecznych*, Warszawa.
- [138] Platek R., Rao J.N.K., Särndal C.E., Singh, M.P., (1987), *Small Area Statistics – An International Symposium*, John Wiley & Sons, New York.
- [139] Popiński W., (2006), Development of the Polish Labour Force Survey, *Statistics in Transition*, 7 (5), 1009-1030.
- [140] Rao J.N.K., (2005), Interplay between sample survey theory and practice: An appraisal, *Survey Methodology*, 31(2), 117-138.
- [141] Rao, J.N.K., Graham, J.E. (1964), Rotation designs for sampling on repeated occasions. *Ann. Math. Statist.* 35, 492-509.
- [142] Riga (1999), *Small Area Estimation – Conference Proceedings*, Riga, Latvia.
- [143] Särndal C.-E., Swensson B., Wretman, J., (1992), *Model Assisted Survey Sampling*, Springer, New York.
- [144] Smith T.M.F., (1976), The foundations of Survey Sampling: a review, *Journal of Royal Statistical Society*, A. 139, 183-204.
- [145] Smith T.M.F., (1994), Sample Surveys 1975-1990; an age of reconciliation? *International Statistical Review*, 62, 3-34.
- [146] Steczkowski J., (1988), *Zastosowanie metody reprezentacyjnej w badaniach społeczno-ekonomicznych*, PWN, Warszawa.
- [147] Steinhaus H., (1956), Liczby złote i żelazne, *Zastosowania Matematyki*, 3, 51-65.
- [148] Steinhaus H., (2000), *Między duchem a materią pośredniczy matematyka*, PWN, Warszawa.
- [149] Sukhatme P.V., Sukhatme, B.V. (1970), *Sampling Theory of Surveys with Applications*. Asia Publishing House, London.
- [150] Szablowski P.J., Wesołowski J., Wiczorkowski R., (1996), Indeks zgodności jako miara jakości danych, *Wiadomości Statystyczne*, 4, 43-49.
- [151] Szarkowski A., Witkowski J., (1994), The Polish Labour Force Survey, *Statistics in Transition*, 1 (4), 467-483.
- [152] Szreder M., (2004), *Metody i techniki sondażowych badań opinii*, PWE, Warszawa.
- [153] Szulc S. (1967), *Metody statystyczne*, PWE, Warszawa.
- [154] Szutkowska J., (2012), Zarządzanie jakością w statystyce publicznej: modele, standardy, metody i narzędzia, *Wiadomości Statystyczne*, 11, s. 38-51.
- [155] UN Statistics Division (2010), *Post Enumeration Surveys*, Operational guidelines, Technical Report, New York, April 2010.
- [156] Verma J. i Betti G. (2006), EU Statistics on Income and Living Conditions (EU-SILC): Choosing the Survey Structure and Sample Design, *Statistics in Transition*, 7 (5), 935-970.
- [157] Wesołowski J. (2004), Problemy estymacji dla małych obszarów, *Wiadomości Statystyczne*, 3, 9-14.
- [158] Wesołowski J. (2010), Recursive optimal estimation in Szarkowski rotation scheme, *Statistics in Transition-new series*, 11(2), 267-285.



- [159] Wierchosławski S., (1995), Węzłowe problemy metodologii badań statystycznych zjawisk społeczno-ekonomicznych w Polsce, W: *Rozwój metodologii badań statystycznych w Polsce*, Biblioteka Wiadomości Statystycznych, 44, 18-44, Warszawa.
- [160] Wywił J., (1995), *Wielowymiarowe aspekty metody reprezentacyjnej*. Ossolineum, Wrocław–Warszawa–Kraków.
- [161] Wywił J., (1996), On two-phase sampling for stratification, *Statistics in Transition*, 2 (6), 971-977.
- [162] Wywił J., (1998), Estimation of population average on the basis of strata formed by means of discrimination functions. *Statistics in Transition*, 3 (5), 903-912.
- [163] Wywił J., (1999), Generalisation of Singh and Srivastava's schemes providing unbiased regression estimators. *Statistics in Transition*, 4 (2), 259-281.
- [164] Wywił J., (2000), On precision of Horvitz-Thompson strategies, *Statistics in Transition*, 4 (5), 779-798.
- [165] Wywił J., (2001), Estimation of population mean on the basis of non-simple sample when non-response error is present. *Statistics in Transition*, 5(3), 443-450.
- [166] Wywił J., (2007), Simulation analysis of accuracy estimation of population mean on the basis of strategy dependent on sampling design proportionate to the order statistic of an auxiliary variable. *Statistics in Transition-new series*, 8(1), 125-137.
- [167] Wywił J., (2010), *Wprowadzenie do metody reprezentacyjnej*, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice.
- [168] Wywił J., Gamrot W., (red.), (2010), *Survey Sampling Methods in Economic and Social Research*, Katowice University of Economics, Katowice.
- [169] Wywił J., Żądło T., (red.), (2009), *Survey Sampling Methods in Economic and Social Research*, Katowice University of Economics, Katowice.
- [170] Yates F., (1980), *Sampling Methods for Censuses and Surveys*, (4th ed.), London.
- [171] Zarkovich S.S., (1966). *Quality of Statistical Data*. FAO, Rome.
- [172] Zasępa R., (1958), Problematyka badań reprezentacyjnych GUS w świetle konsultacji z prof. J. Neymanem, *Wiadomości Statystyczne*, 6, 7-12.
- [173] Zasępa R., (1962), *Badania statystyczne metodą reprezentacyjną*, PWN, Warszawa.
- [174] Zasępa R., (1968), Zastosowanie metody reprezentacyjnej przy spisach rolnych, *Studia i Prace Statystyczne*, GUS, 10, Warszawa.
- [175] Zasępa R., (1972), *Metoda reprezentacyjna*, PWE, Warszawa.
- [176] Zasępa R., (1993a). Precyzja wyników badań budżetów rodzinnych, „*Wiadomości Statystyczne*”, 3, 30-32.
- [177] Zasępa R., (1993b), Use of Sampling Methods in Population Censuses in Poland, *Statistics in Transition*, 1(1), 69-78.

## WSPÓLZALEŻNOŚĆ POMIĘDZY ROZWOJEM TEORII I PRAKTYKI BADAŃ REPREZENTACYJNYCH W POLSCE

### Streszczenie

Autor bada współzależności między rozwojem teorii i praktyki badań reprezentacyjnych w Polsce w ciągu ponad 60 lat. Rozpoczyna od klasycznej pracy Neymana (Neyman, 1934), która dała teoretyczne podstawy probabilistycznego podejścia wyboru próbki umożliwiające wnioskowanie z badań reprezentacyjnych. Główne idee tej pracy były najpierw opublikowane po polsku w 1933 r. (Neyman, 1933) i miały istotny wpływ na a praktykę badań reprezentacyjnych Polsce przed i po II Wojnie Światowej. Badania reprezentacyjne prowadzone w latach 1950-tych i 1960-tych były konsultowane z J. Neymanem w czasie jego wiz w Polsce w latach 1950 i 1958 (Fisz, 1950; Zasępa, 1958).

Praktyczne problemy występujące przy planowaniu i analizie badań reprezentacyjnych Polsce były częściowo rozwiązywane przez Komisję Matematyczną GUS powołaną w końcu 1949 r. jako organ doradczy i opiniodawczy Prezesa GUS. Komisja ta składała się ze specjalistów zarówno z GUS jak i ośrodków naukowo-badawczych w kraju (Kordos, 2012a). Komisja działała do 1993 r. i wpływała w istotny sposób na praktykę badań reprezentacyjnych Polsce.

Specjalną uwagę poświęcono wpływowi teorii badań próbkowych na świecie i w Polsce na praktykę badań reprezentacyjnych w Polsce, a w szczególności na plany i schematy losowania i metody estymacji w badaniach prowadzonych *w czasie*, a głównie *metodzie rotacyjnej* (Greń, 1969; Kordos, 1967, 2012b; Kowalczyk, 2004; Kowalski, 2006; Lednicki, 1982; Popiński, 2006; Wesołowski, 2010); *lokalizacji próby* (Bracha et al., 2004b, Greń, 1964, 1966; Lednicki, 1979, 1989; Lednicki i Wesołowski, 1994); *metodom estymacji* (Bracha, 1996, 1998; Greń, 1970; Kordos, 1982; Lednicki, 1979, 1987, Wesołowski, 2004, Zasepa, 1962, 1972), *jakości danych* (Kordos, 1973, 1988; Zasepa, 1993) oraz *metodom estymacji dla małych obszarów* (Bracha, 1994, 2003; Bracha et al., 2004b; Dehnel, 2010; Domański, Pruska, 2001; Golata, 2004ab, 2012; Kalton et al., 1993; Kordos, 1991, 2000b; Kordos, Paradysz, 2000; Niemirow, Wesołowski, 2012; Paradysz, 1998; Wesołowski, 2004). W zakończeniu prowadzone są rozważania na temat przyszłych zapotrzebowań na informacje z badań reprezentacyjnych.

**Słowa kluczowe:** badanie reprezentacyjne, metoda estymacji, złożone badanie reprezentacyjne, jakość danych, statystyka małych obszarów, metoda rotacyjna, metoda alokacji próby

## THE INTERPLAY BETWEEN SAMPLE SURVEY THEORY AND PRACTICE IN POLAND

### Abstract

The author examines interplay between sample survey theory and practice in Poland over the past 60 years or so. He begins with the Neyman's (1934) classic landmark paper which laid theoretical foundations to the probability sampling (or design-based) approach to inference from survey samples. Main ideas of that paper were first published in Polish in 1933 (Neyman, 1933) and had a significant impact on sampling practice in Poland before and after the World War II. Sample surveys conducted in 1950s and 1960s were consulted with J. Neyman during his visits in Poland in 1950 and 1958 (Fisz, 1950a; Zasepa, 1958).

Some practical problems encountered in the design and analysis of sample surveys were partly solved by the Mathematical Commission of the CSO which was established in 1949, as an advisory and opinion-making body to the CSO President in the field of sample surveys. The Commission concentrated specialists in the sampling methods both from the CSO and research centres in the country (Kordos, 2012a). The Commission had a significant impact on sampling practice in Poland and was active till 1993.

Special attention is devoted to the impact of sampling theory on sampling practice in Poland, and particularly on sample designs and estimation methods in: *sampling in time and rotation methods* (Kordos, 1967, 2012b; Kowalczyk, 2004; Kowalski, 2006; Wesołowski, 2010); *sample allocation and estimation methods* (Bracha, 1994, 1996, 2003; Greń, 1964, 1966, 1969, 1970; Kordos, 1969, 1973, 1982; Wesołowski, 2004, Zasepa, 1962, 1972, 1993); *data quality* (Kordos, 1973, 1988); *estimation methods for small areas* (Dehnel, 2010; Domański, Pruska, 2001; Golata, 2004ab, 2012; Kalton et al., 1993; Kordos, 1991, 2000b, 2004; Kordos, Paradysz, 2000; Niemirow, Wesołowski, 2012; Paradysz, 1998; Wesołowski, 2004). Concluding remarks are given at the end.

**Key words:** sample survey, estimation methods, complex sample surveys, data quality, estimation for small area, rotation sampling, sample allocation methods