

Problemy jakości danych statystycznych w przypadku badania cech rzadkich

Alina Jędrzejczak

– Katedra Metod Statystycznych, Uniwersytet Łódzki

Jan Kubacki

– Ośrodek Statystyki Matematycznej, Urząd Statystyczny w Łodzi



Streszczenie

- ▶ W referacie przedstawiono zagadnienie szacowania parametrów dla cech rzadkich oraz metody wyboru próby dla tego typu zmiennych. Zaprezentowane zostały podstawowe schematy losowania dla cech rzadkich, w tym schemat losowania odwrotnego (*inverse sampling*), schemat losowania wielokrotnego połowu (losowanie typu pojmanie-uwolnienie - *capture-recapture sampling*), losowanie lokacyjne (*location sampling*), schemat linia-przecięcie (*line-intercept sampling*) oraz schemat śledzenia łączy (*link-tracing design*).
- ▶ Przedstawiono też koncepcję operatów wielokrotnych (*multiple frame*) oraz zagadnienie próby o zwiększonym pokryciu (*oversampling*). Trudności wynikające z zastosowania typowych schematów losowania zilustrowano przykładem symulacyjnym opartym na danych z PSR, oraz danych z BAEL i Budżetów Gospodarstw Domowych. Wskazano na istnienie alternatywnych rozwiązań dla takich przypadków, wraz z określeniem korzyści oraz wad takich rozwiązań.

Zjawiska rzadkie – klasyfikacja Purcella i Kisha

Zgodnie z klasyfikacją wprowadzoną przez Purcella i Kisha (1979), można wyróżnić cztery typy domen według ich wielkości w stosunku do całej badanej populacji. Według tej klasyfikacji wyróżnić można:

- ▶ dziedziny główne (*major*) - stanowiące 1/10 populacji,
- ▶ drugorzędne (*minor*) - o liczebności od 1/10 do 1/100 populacji,
- ▶ minidziedziny (*mini*) - o wielkości od 1/100 do 1/10000 populacji
- ▶ jednostki rzadko występujące (*rare types of individuals*) - dla których względna wielkość dziedziny jest mniejsza niż 1/10000.

Zjawiska rzadkie – problemy badawcze

- ▶ W badaniach nakierowanych na cechy powszechne liczba zaobserwowanych jednostek posiadających rzadki atrybut może być zbyt mała aby osiągnąć zadowalającą precyzję szacunku.
- ▶ Standardowe metody, do których zalicza się estymator Horvitz-Thompsona (HT), można odnieść przede wszystkim do populacji stanowiących dziedzinę główną. Natomiast w odniesieniu do dziedzin drugorzędnych, jak i do minidziedzin konieczne jest zastosowanie innych technik w tym metod estymacji pośredniej
- ▶ Metody estymacji pośredniej stosowane w statystyce małych domen często także zawodzą w przypadku populacji rzadkich, gdyż ze względu na ich odrębny charakter „zapożyczanie mocy” z pozostałej części próby nie znajduje uzasadnienia.
- ▶ Istotnym problemem jest także dotarcie do tych populacji, gdyż często brak jest odpowiednich operatów losowania. Często uniemożliwia to zastosowanie klasycznych losowych technik doboru próby.
- ▶ W celu zminimalizowania błędów konieczne jest zastosowaniu specjalnych technik doboru próby oraz estymacji.

Schematy losowania dla rzadkich populacji

Losowanie, którego celem jest dotarcie do populacji rzadkich bądź trudno uchwytnych może być żmudne oraz kosztowne. W literaturze można znaleźć kilka schematów losowania, które można wykorzystać dla tego typu populacji:

- ▶ schemat losowania odwrotnego (*inverse sampling*)
- ▶ schemat losowania wielokrotnego połowu (*capture-recapture sampling*)
- ▶ losowanie lokacyjne (*location sampling*)
- ▶ schemat linia-przecięcie (*line-intercept sampling*)
- ▶ schemat śledzenia łączy (*link-tracing design*) który obejmuje
 - ▶ schemat losowania sieciowego (*network sampling*)
 - ▶ schemat "śnieżnej kuli" (*snowball sampling*)
 - ▶ schemat przypadkowego błędzenia (*random walk sampling*)
 - ▶ schemat losowania adaptacyjnego (*adaptive sampling*)

Losowanie odwrotne

- ▶ Koncepcję losowania odwrotnego, które miało być techniką pozwalającą na losowania dla cech rzadkich w populacji, wprowadził Haldane (1945).
- ▶ W przypadku badania zjawisk rzadkich, próba o ustalonej wielkości może nie pozwolić na wylosowanie dostatecznej liczby jednostek mających wymagany atrybut.
- ▶ Losowanie odwrotne polega na losowaniu jednostka po jednostce z użyciem schematu losowania prostego, dopóki nie wybierze się określonej liczby elementów posiadających daną cechę. Jest więc odwrotnie niż w klasycznych schematach losowania- to liczebność próby a nie liczba wyróżnionych elementów jest zmienną losową.
- ▶ W pracy Haldane rozważano przypadek nieograniczonej populacji. Określono rozmiar próby oraz rozkład prawdopodobieństwa dla takiego przypadku. Przedstawiono też formuły na nieobciążony estymator dla proporcji oraz wariancję tego estymatora.

Losowanie metodą wielokrotnego połowu

- ▶ Metoda wielokrotnego połowu była po raz pierwszy zaproponowana przez Sebera (1973) do szacowania populacji rzadkich gatunków zwierząt.
- ▶ W celu oszacowania całkowitej liczby jednostek w populacji posiadających dany rzadki atrybut pobiera się wstępną próbkę tych jednostek, są one oznaczane i powracają do populacji. Następnie niezależnie wybiera się drugą losową próbkę i zlicza ponownie oznakowane jednostki.
- ▶ Jeśli druga próbka jest reprezentatywna dla całej populacji, to otrzymany udział oznakowanych jednostek powinien być taki sam jak udział tych jednostek w całej populacji. Korzystając z tego założenia, można w prosty sposób określić całkowitą liczbę jednostek w populacji.

Losowanie lokacyjne

- ▶ Technika ta opisana jest w pracy Kaltona (1991), jako metoda losowania jednostek z populacji przemieszczających się (*mobile populations*), jak turyści lub odwiedzający centra handlowe, chorzy na HIV, czy bezdomni.
- ▶ Zwykle stosuje się losowanie dwustopniowe, przy czym jps konstruowane są jako kombinacje miejsc obserwacji i segmentów czasu. Dodatkowo na pierwszym stopniu stosuje się warstwowanie względem np. godzin otwarcia i rodzaju obiektu, a jednostki wybierane są proporcjonalnie do ich wielkości. Jednostki badania (jds) losowane są zazwyczaj systematycznie w ramach wylosowanych segmentów czasu.
- ▶ Otrzymana próba jest próbą probabilistyczną z punktu widzenia charakterystyk badanych jednostek (np. satysfakcji z odwiedzania danego sklepu czy muzeum). Jednak gdy celem badania jest określenie liczebności danej rzadkiej populacji powstają problemy z określeniem prawdopodobieństw inkluzji związane z powtarzalnością obserwacji tych samych obiektów (tzw. *multiplicities*).



Losowanie metodą linia-przecięcie

- ▶ Losowanie metodą linia-przecięcie (zob. Lucas i Seber, 1977, 618-622) odnosi się do losowania dla rzadkich populacji w celu określenia całkowitej liczebności takiej populacji.
- ▶ Jest to schemat losowania, w którym wyznacza się n linii przecięcia (*transect lines*) przez wybór n pozycji tych linii, wzdłuż pewnej linii bazowej, która poprowadzona jest wzdłuż szerokości badanego obszaru, natomiast linie przecięcia są poprowadzone przez badany obszar prostopadle do linii bazowej, dla każdej z wybranych pozycji.
- ▶ Za każdym razem, gdy obiekt w populacji jest przecięty przez jedną lub więcej wylosowanych linii, wykonuje się zapis dla rozpatrywanej zmiennej.
- ▶ We wnioskowaniu o wielkości populacji można użyć estymatora HT, w którym prawdopodobieństwo inkluzji obiektu wynika z rzutu szerokości obiektu na linię bazową odniesioną do długości tej linii (Thompson, 2002).

Losowanie metodą śledzenia łączy

Losowanie tą metodą polega na stosowaniu rozmaitych technik, które w celu otrzymania próby wykorzystują powiązania (*links* lub *connections*) między jednostkami populacji. Powiązania te dotyczą często więzi rodzinnych lub społecznych, pomocnych w identyfikowaniu przypadków rzadkich chorób lub dotarciu do trudno dostępnych populacji takich jak np. bezdomni.

Schemat śledzenia łączy obejmuje następujące techniki pobierania próby:

1. losowanie sieciowe,
2. losowanie metodą kuli śnieżnej,
3. losowanie metodą przypadkowego błędzenia
4. losowanie adaptacyjne.



Losowanie metodą śledzenia łączy

1. Losowanie sieciowe

- ▶ Schemat losowania sieciowego został opisany po raz pierwszy przez socjologa James'a Colemana (1958) w kontekście badań struktur społecznych opartych często na powiązaniach nieformalnych, a więc trudno uchwytnych dla badacza.
- ▶ Sieć w takim przypadku jest zbiorem obserwowanych jednostek wraz określonym wspólnym wzorcem połączeń, przy czym jednostka może być połączona z więcej niż jedną siecią. Początkowo pobierana jest próba prosta lub warstwowa, a następnie dołączane są wszystkie jednostki powiązane z początkowo wylosowanymi.
- ▶ Nieobciążony estymator dla losowania sieciowego (tzw. *multiplicity estimator*) został zaproponowany w pracy Birnbauma i Sirkena (1965). Autorzy ci podali także alternatywną formułę oparta na estymatorze HT.

2. Losowanie metodą kuli śnieżnej

- ▶ Termin ten odnosi się do dwóch podobnych schematów pobierania próby, opartych na tzw. efekcie kuli śnieżnej.
- ▶ Pierwszy z nich był zaproponowany przez Goodmana (1961) do badania struktur społecznych i polega na wylosowaniu pewnej liczby jednostek, które w następnym etapie wskazują określoną liczbę następnych itd., przy czym liczba etapów jest także określona z góry.
- ▶ Druga wersja metody, stosowana w przypadku badania populacji rzadkich, była przedstawiona w pracy Kaltona i Andersona (1986). Początkowo identyfikowana jest niewielka grupa jednostek, które potem wprowadzają do próby następne itd.. Otrzymana w efekcie próba nieprobabilistyczna może być wykorzystana bezpośrednio do estymacji lub też służy jako operat do pobrania próby losowej.

3. Losowanie metodą „random walk”

- ▶ Schemat błędzenia przypadkowego jest schematem losowania pozwalającym na otrzymanie próby probabilistycznej dla dużych sieci społecznych. (Klov Dahl i inni , 1977).
- ▶ Jednostki do próby losowane są w kilku etapach zwanych falami (*waves*). Jeśli na pewnym etapie wylosowano daną jednostkę, to w następnym etapie losuje się jedną z jednostek wskazanych przez jej łączy. Procedurę powtarza się tak długo aż nie osiągnie się żądanej liczebności próby.
- ▶ W ten sposób otrzymujemy próbę losową o liczebności n . Do oszacowania średniej na podstawie tej próby można wykorzystać asymptotycznie nieobciążony estymator ilorazowy oparty na formule Hansena-Hurwitza.

Losowanie metodą śledzenia łączy

4. Losowanie adaptacyjne

- ▶ Adaptacyjny schemat losowania sieciowego to schemat, w którym decyzja o włączeniu do próby powiązanych z danym węzłem jednostek zależy od wartości zmiennej obserwowanej w węźle.
- ▶ Procedura adaptacyjna została zaproponowana przez Thompsona i Ramseya w 1983 roku. Thompson w 2006 roku przedstawił nowy schemat losowania nazwany adaptacyjnym losowaniem sieciowym (*adaptive web sampling* - AVWS), który można zastosować w przypadku sieci oraz schematów przestrzennych. Wybór jednostek odbywa się sekwencyjnie z zastosowaniem mieszanki rozkładów wykorzystujących zależności sieciowe lub przestrzenne, a także wartości otrzymane z próby.
- ▶ Schemat ten ma pewną przewagę nad opracowanymi wcześniej schematami śledzenia łączy, gdyż m. in. pozwala na sterowanie rozmiarem próby.

Operat wielokrotny (multiple frame)

- ▶ W przypadku, gdy rozważa się badanie cech rzadkich czasami skuteczne może być zastosowanie wielu wykazów jednostek będących przedmiotem badania. Często sytuacja taka zachodzi, gdy badane są osoby mające określone schorzenie względnie badane są specyficzne gospodarstwa rolne. W takim przypadku podczas estymacji uwzględnić trzeba liczebność wszystkich wykorzystywanych operatów oraz występowanie ich części wspólnych. Istnienie pewnych informacji dodatkowych w próbie może też być pomocne przy stworzeniu oddzielnych warstw, których wyodrębnienie może poprawić jakość otrzymanych szacunków.

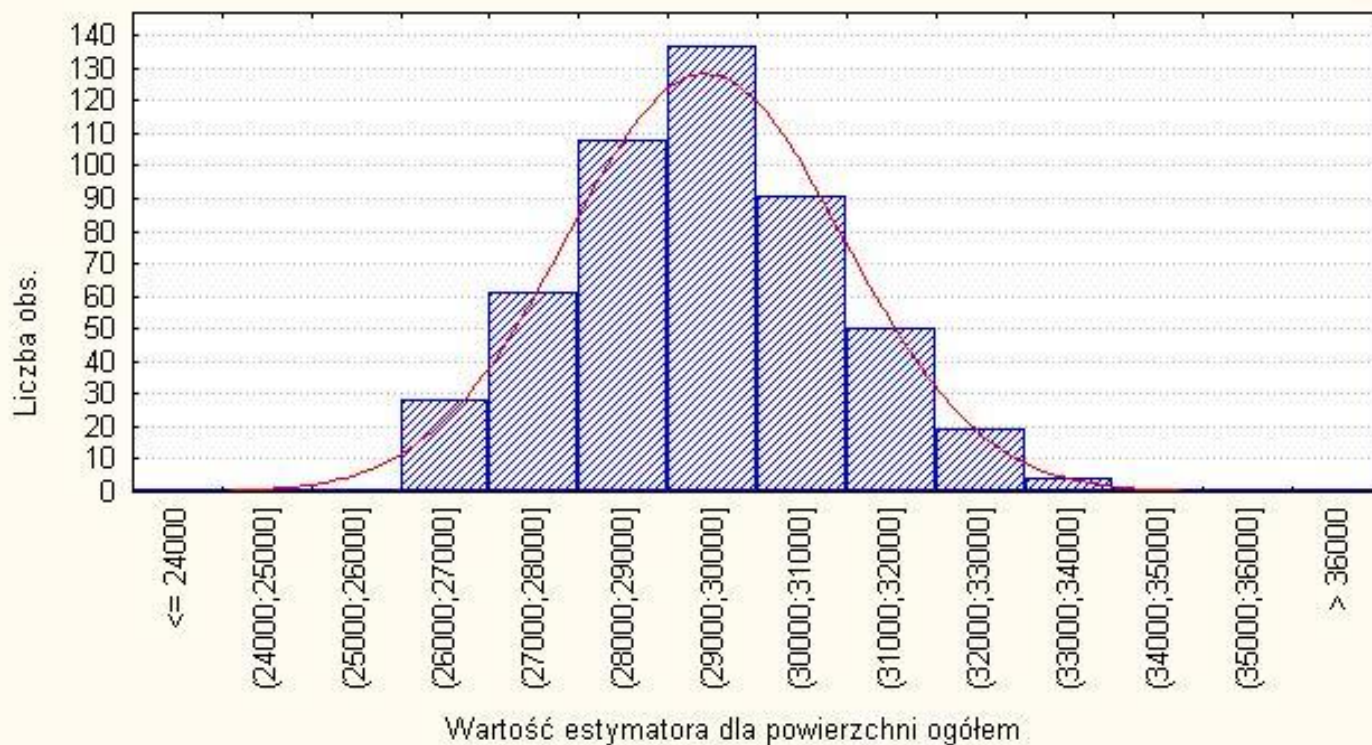


Próba o zwiększonym pokryciu (nadreprezentacja - oversampling)

- ▶ W celu zwiększenia frakcji losowania w pewnych warstwach populacji można wykorzystać zarówno informację dotyczącą lokalizacji geograficznej (np. w przypadku mniejszości etnicznych), jak również inne informacje dotyczące badanych jednostek (np. typ i wielkość gospodarstwa rolnego lub rodzaj działalności wykonywanej przez podmiot ekonomiczny oraz jego wielkość).
- ▶ Przykłady:
 - ▶ Nadreprezentacja obszarów wiejskich w BAEL
 - ▶ Badania rolnicze (w tym Badanie Struktury Gospodarstw Rolnych) - wyodrębnienie warstw dla gospodarstw hodowlanych i sadowniczych, gospodarstw ekologicznych

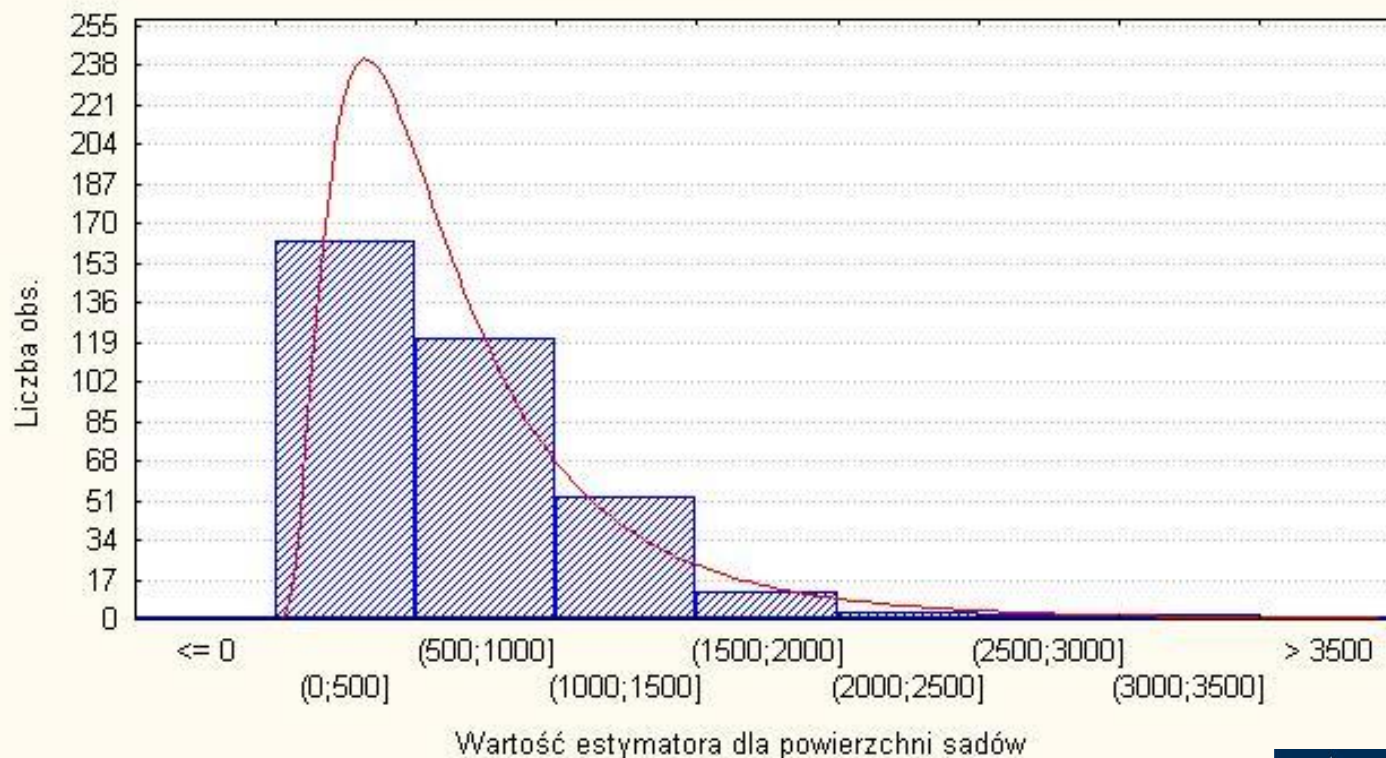
Szacunki dla cechy powszechnej – powierzchnia gosp. ogółem – symulacja - dane PSR 2002

Rozkład empiryczny szacunków estymatora powierzchni gospodarstwa ogółem wyznaczonego w badaniu symulacyjnym na podstawie danych o gospodarstwach poniżej 1 ha użytków rolnych i nie spełniających progów w PSR2010 w gminie Reńska Wieś



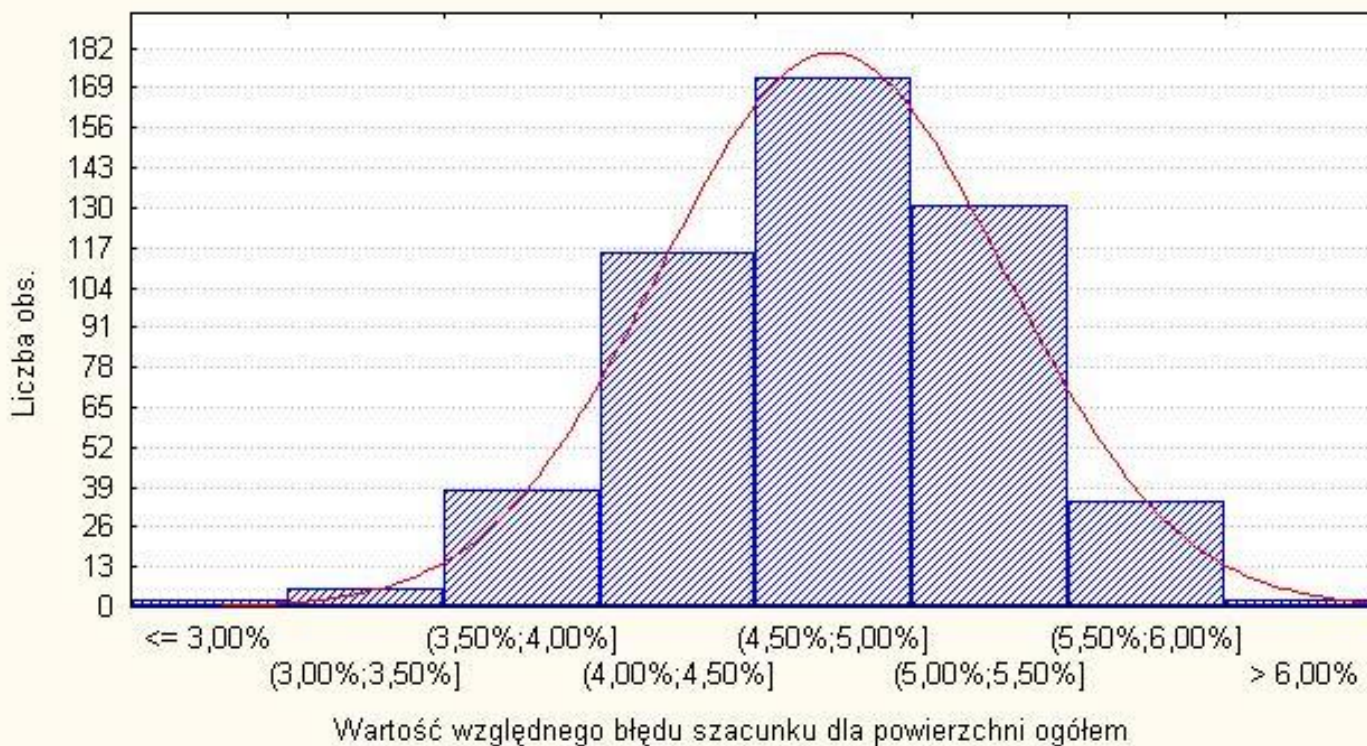
Szacunki dla cechy rzadkiej – powierzchnia sadów – symulacja - dane PSR 2002

Rozkład empiryczny szacunków estymatora powierzchni sadów
wyznaczonego w badaniu symulacyjnym na podstawie
danych o gospodarstwach poniżej 1 ha użytków rolnych i nie
spełniających progów w PSR2010 w gminie Reńska Wieś



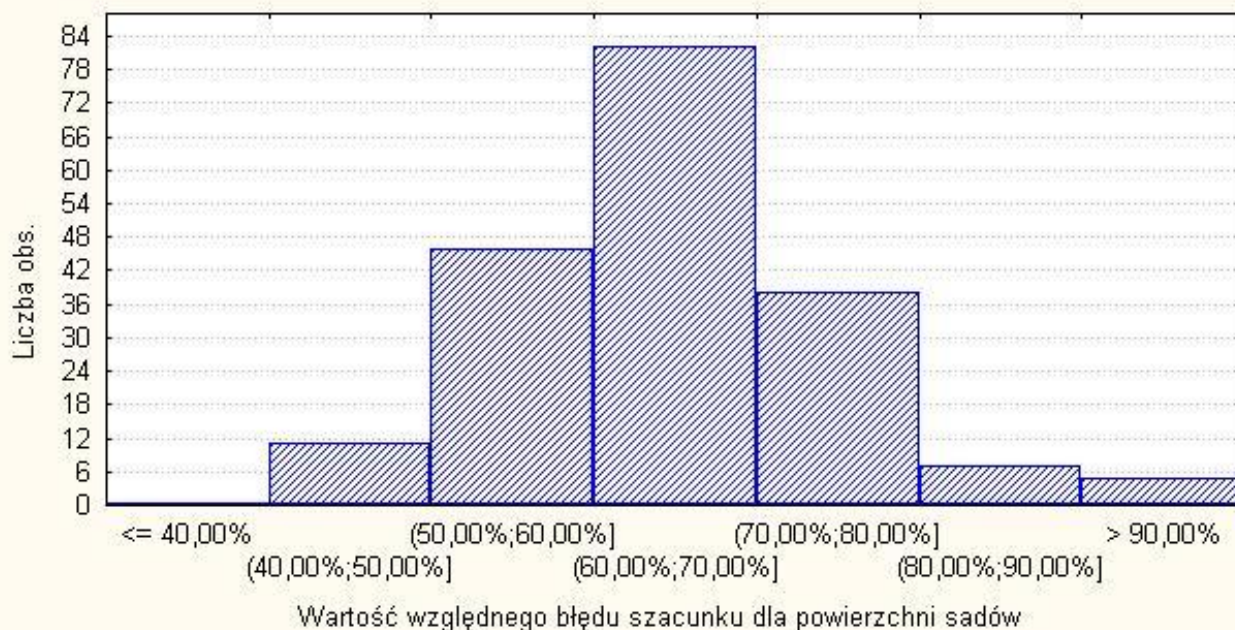
Precyzja dla cechy powszechnej – powierzchnia gosp. ogółem – symulacja - dane PSR 2002

Rozkład empiryczny względnego błędu szacunku estymatora powierzchni gospodarstwa ogółem wyznaczonego w badaniu symulacyjnym na podstawie danych o gospodarstwach poniżej 1 ha użytków rolnych i nie spełniających progów w PSR2010 w gminie Reńska Wieś



Precyzja dla cechy rzadkiej – powierzchnia sadów – symulacja - dane PSR 2002

Rozkład empiryczny względnego błędu szacunku estymatora powierzchni sadów
wyznaczonego w badaniu symulacyjnym na podstawie
danych o gospodarstwach poniżej 1 ha użytków rolnych i nie
spełniających progów z PSR2010 w gminie Reńska Wieś

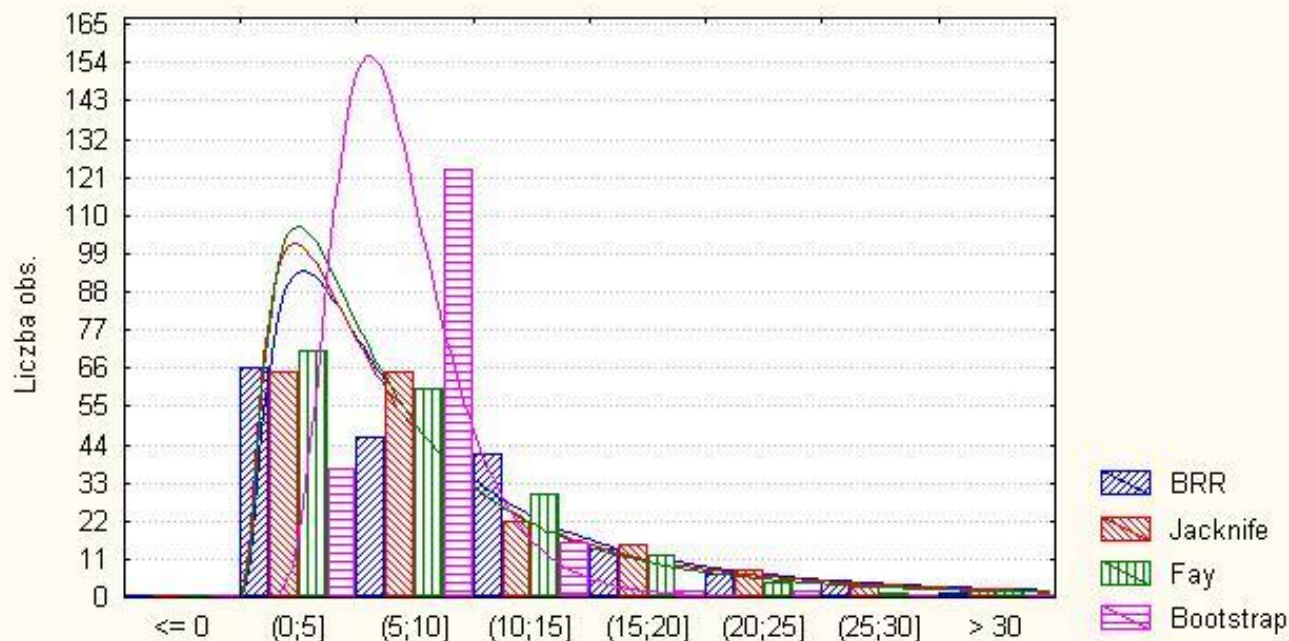


Precyzja dla cechy powszechnej

– dochód rozporządzalny – dane dla powiatów

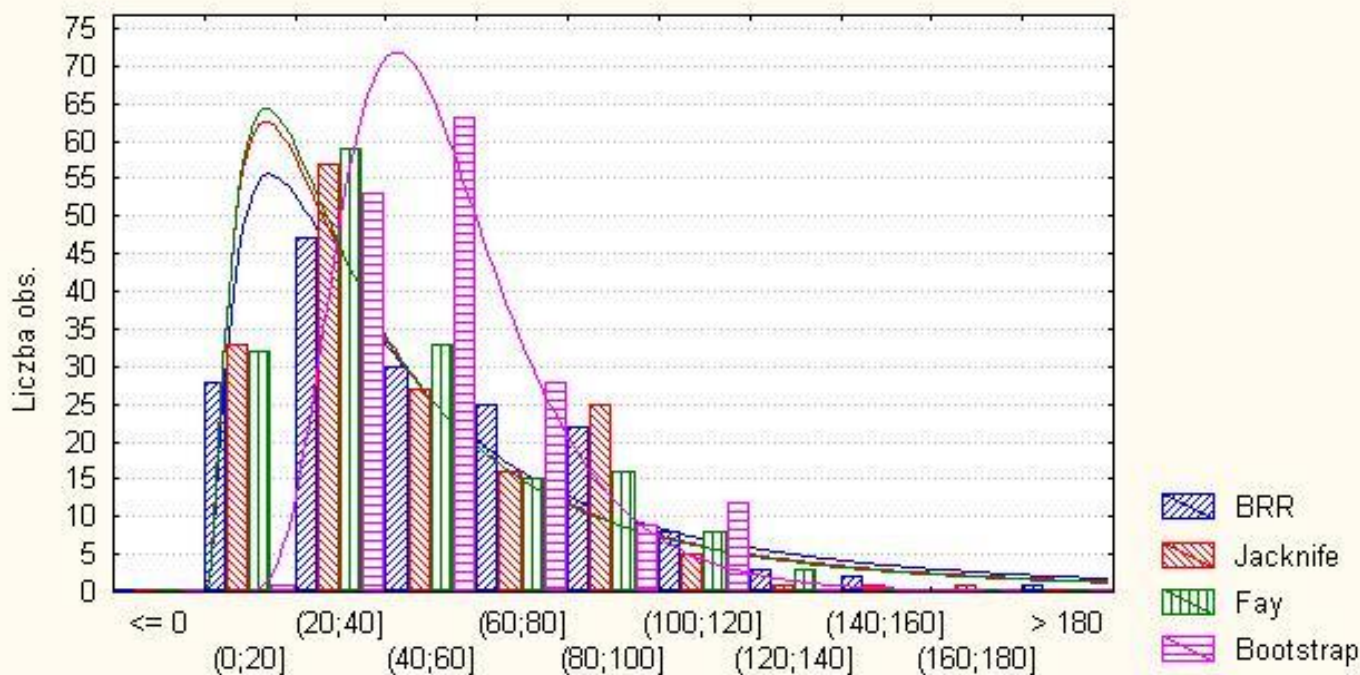
– Badanie Budżetów - 2003

Rozkład współczynnika zmienności szacunków
dla estymatora dochodu rozporządzalnego
otrzymany przy użyciu różnych metod szacowania precyzji
dla Badania Budżetów Gospodarstw Domowych i powiatów w 2003 r



Precyzja dla cechy rzadkiej – zasiłki dla bezrobotnych – dane dla powiatów – Badanie Budżetów 2003

Rozkład współczynnika zmienności szacunków
dla estymatora dochodu z zasiłków dla bezrobotnych
otrzymany przy użyciu różnych metod szacowania precyzji
dla Badania Budżetów Gospodarstw Domowych i powiatów w 2003 r

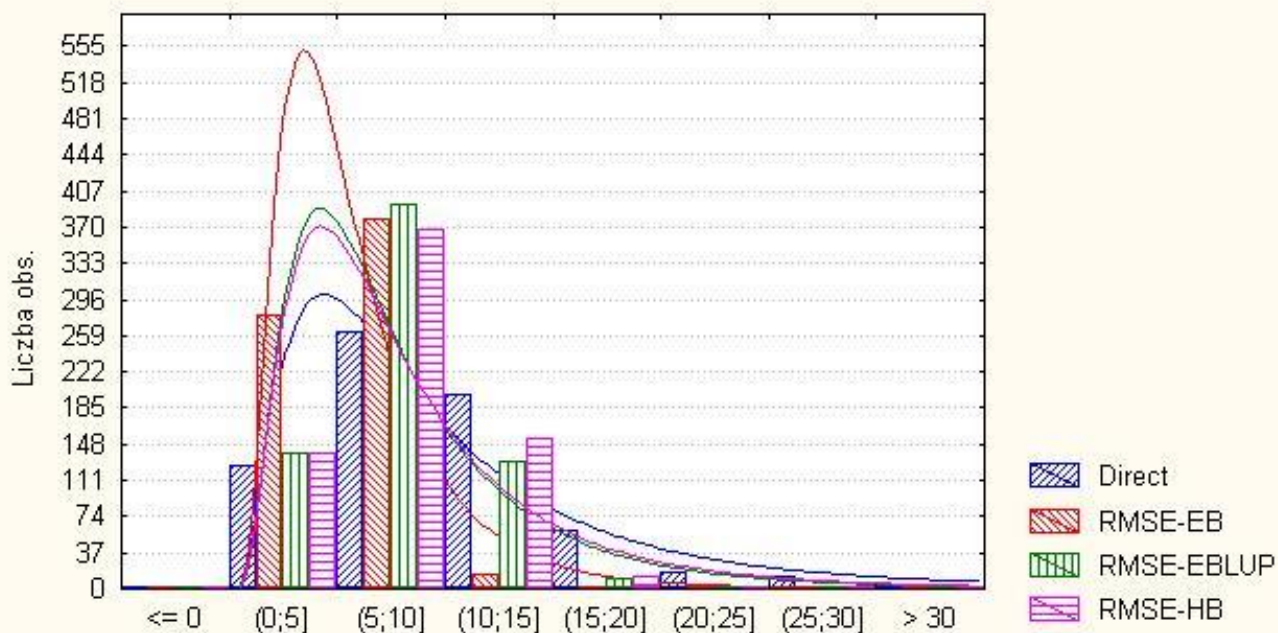


Precyzja dla cechy powszechnej

– dochód rozporządzalny – dane dla powiatów

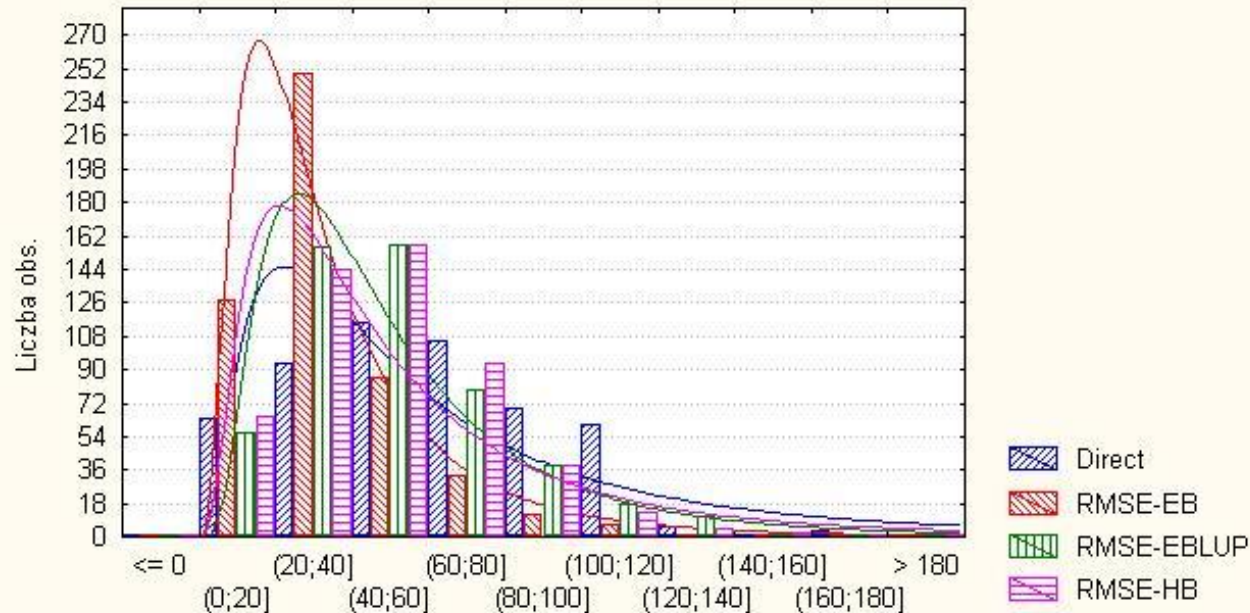
– Badanie Budżetów 2003 i 2004–z uwzględnieniem modeli

Względny błąd szacunku dla estymatora bezpośredniego oraz dla estymatora EB, EBLUP (wariant REML) oraz HB dla dochodu rozporządzalnego w powiatach na podstawie Badania Budżetów oraz danych z systemu POLTAX dla lat 2003 i 2004



Precyzja dla cechy rzadkiej – dochód z zasiłku dla bezrobotnych – dane dla powiatów – Badanie Budżetów 2003 i 2004 – z uwzględnieniem modeli

Względny błąd szacunku dla estymatora bezpośredniego oraz dla estymatorów EB, EBLUP (wariant REML) oraz HB dla dochodu z zasiłków dla bezrobotn. w powiatach na podstawie Badania Budżetów oraz danych z systemu POLTAX dla lat 2003 i 2004



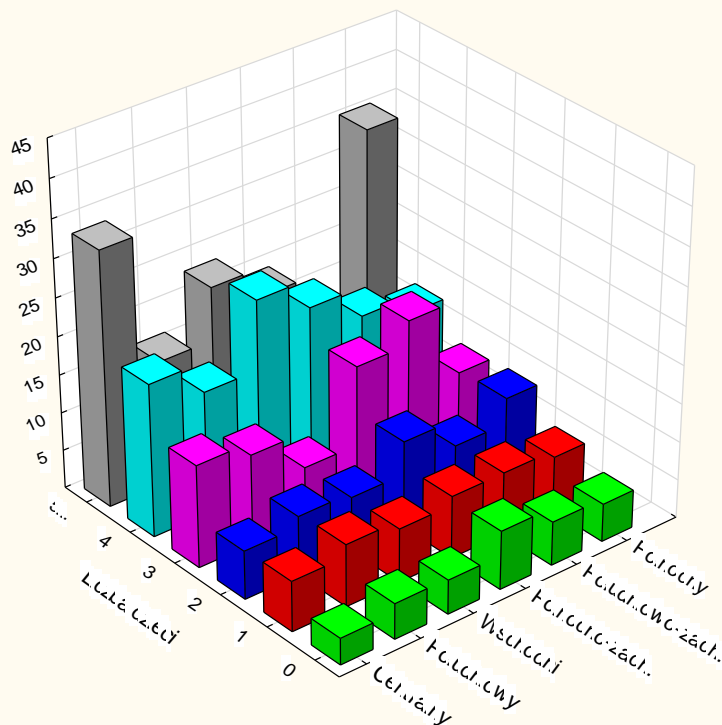
Komentarz do wyników estymacji cech rzadkich i powszechnych

- ▶ Ilustracje przedstawione na poprzednich wykresach pokazują wyniki badania symulacyjnego wykonanego w ramach prac przygotowawczych do PSR 2010. W eksperymencie użyto danych ze spisu rolnego PSR 2002 wraz z kryteriami doboru próby przyjętymi w PSR 2010 dla badania małych gospodarstw (Kubacki, Piasecki, 2009) oraz rozkłady precyzji z użyciem technik bezpośrednich otrzymane dla dochodu rozporządzalnego i dochodu z zasiłku dla bezrobotnych w powiatach na podstawie Badania Budżetów Gospodarstw Domowych (Kubacki, Jędrzejczak, Piasecki, 2011)
- ▶ Wykresy pokazują znaczne różnice w poziomie precyzji między szacunkami bezpośrednimi dla cech powszechnych i cech rzadkich oraz ilustrują brak skuteczności technik estymacji pośredniej (w tym opartych na modelach) dla cech rzadkich.
- ▶ W przypadku cechy rzadkiej wskazano też - wykorzystując wyniki badania symulacyjnego - na odmienny charakter rozkładu precyzji wynikającej ze schematu losowania niż dla cech powszechnych, która ma często inny charakter, niż dla cechy powszechnej. Jest on nierzadko asymetryczny, a niekiedy wręcz nieregularny, co jest szczególnie widoczne dla małych gmin i cech rzadkich.
- ▶ Dodatkowo w wielu przypadkach estymacja precyzji była niemożliwa w wyniku zerowania się estymatora parametru i estymatora jego wariancji.



Precyzja dla populacji rzadkiej i powszechnej. At-Risk-of-Poverty Rate, Badanie Budżetów 2009

Względne błędy szacunku dla estymatora ARPR według liczby dzieci w rodzinie i makroregionów na podstawie Badania Budżetów 2009



Wartości globalne (w tysiącach) wielkości bezrobocia i ich względny błąd standardowy (w procentach) na podstawie BAEL dla Łodzi w IV kwartale 1999

Typ estymatora		Ogółem	Kobiety	Mężczy- żni	Do 24 lat	25-34 lat	35-44 lat	45-54 lat	55 lat i więcej
Wielkość populacji		431,40	235,50	195,90	81,50	62,30	80,00	85,60	122,10
Rejestrowane bezrobocie		29,80	15,70	14,10	6,60	5,00	8,70	8,40	1,20
Udział w próbie		0,64	0,37	0,27	0,13	0,09	0,12	0,13	0,17
Estymator Horwitza – Thompsona	Ocena	40,5	23,1	17,3	13,8	8,7	5,1	9,4	3,6
	CV	13,1	5,7	23,8	10,2	23,0	14,5	41,1	36,9
Estymator POS – liczebnościowy	Ocena	37,6	20,4	17,2	12,1	8,3	4,6	8,5	3,5
	CV	14,8	9,6	25,2	15,4	19,8	22,2	43,6	61,0
Estymator POS – ilorazowy	Ocena	29,3	16,3	13,0	8,9	5,5	6,8	5,8	1,4
	CV	17,2	21,5	9,8	19,6	26,1	51,0	17,0	60,0
Estymator POS – regresyjny	Ocena	31,8	18,1	13,5	10,1	6,4	6,4	6,0	3,0
	CV	2,4	2,9	5,1	10,7	8,4	15,5	9,1	-
Estymator syntetyczny – liczebnościowy	Ocena	38,0	20,7	17,3	10,6	9,6	6,4	7,4	3,6
	CV	8,8	6,3	13,1	14,2	5,9	8,9	27,0	26,4
Estymator syntetyczny – ilorazowy	Ocena	28,8	14,4	14,5	6,0	5,0	7,7	7,1	2,4
	CV	1,5	4,6	2,8	3,2	9,8	7,8	9,7	22,7
Estymator syntetyczny – regresyjny	Ocena	35,5	18,5	17,0	9,6	8,4	7,5	6,6	2,9
	CV	0,7	1,1	2,4	1,7	3,5	1,9	4,9	7,9
Uogólniony estymator regresyjny	Ocena	37,0	19,7	17,3	10,6	9,4	6,2	7,3	3,4
	CV	2,9	2,8	3,8	2,6	2,8	7,0	5,9	5,4

Komentarz do wyników estymacji wielkości bezrobocia

- ▶ Rezultaty zaprezentowane w przedstawionej tabeli wskazują, że w większości przypadków wielkość względnego błędu szacunku jest mniejsza dla estymatorów pośrednich (syntetycznych) wykorzystujących cechy dodatkowe, w szczególności dla syntetycznych estymatorów regresyjnych.
- ▶ Poprawa precyzji dla zwykłego i ilorazowego estymatora wykorzystującego warstwowanie po losowaniu (estymatora POS) jest stosunkowo niewielka, wartości błędu względnego są niekiedy duże w porównaniu z estymatorem HT.
- ▶ Dla przedstawionej tabeli charakterystyczne są także dość duże wartości błędów dla cechy rzadszej (liczba bezrobotnych w grupie wiekowej powyżej 55 lat). Praktycznie tylko dla jednego estymatora można uznać szacunki za wystarczająco precyzyjne. Problem ten pogłębia się dla mniejszych powiatów (tabela przedstawia wyniki dla Powiatowego Urzędu Pracy w Łodzi)



Potencjalne dalsze możliwości zastosowania omawianych metod w praktyce

- ▶ Zastosowanie metody losowania typu wielokrotnego połowu (por.: L. Fattorini, G. Ghellini, 2009), oraz operatów wielokrotnych (por.: S. Lohr, 2007), w badaniu bezdomności.
- ▶ Zastosowanie metody losowania sieciowego w badaniu wiktylizacji (zob. R. Czaja, J. Blair, 1988)
- ▶ Zastosowanie metody losowania sieciowego w badaniach zdrowia i niepełnosprawności (zob. M.G. Sirken, 1998)
- ▶ Dalsze doskonalenie operatów do badań społecznych (por.: Budowa operatu statystycznego do badań społecznych - praca metodologiczna 3.154) oraz rolniczych (por.: Analiza możliwości zastępowania i łączenia danych administracyjnych oraz statystycznych z zakresu rolnictwa. Praca metodologiczna 3.101, a także FAO, 1996, Multiple Frame Agricultural Surveys)

Podsumowanie

- ▶ Przedstawione przykłady wskazują na znacznie gorszą precyzję estymatorów parametrów dla cech rzadkich w porównaniu z powszechnymi.
- ▶ Częstym problemem jest ponadto niemożność estymacji precyzji szacunku w wyniku zerowania się estymatora parametru i estymatora jego precyzji, co jest związane z egzotycznością cechy.
- ▶ Z przedstawionego w referacie przeglądu metod losowania dla cech rzadkich wynika iż możliwe jest zastosowanie metod badania cech rzadkich (często wykorzystujących metody biometryczne) także w przypadku badań społecznych.
- ▶ Zastosowanie metod takich jak operaty wielokrotne oraz nadreprezentacja potwierdza skuteczność stosowania obu tych metod w praktyce.
- ▶ Wydaje się, że celowe jest dalsze doskonalenie metodologii statystycznej pozwalającej na pełniejszą ocenę zjawisk rzadkich.
- ▶ Wykorzystanie takich metod musi uwzględniać zarówno korzyści jak i trudności (w tym dotyczące kosztów badania) związane z zastosowaniem takich metod.

Literatura

- ▶ Birnbaum, Z.W., Sirken, M.G. (1965) Design of Sampling Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates. Vital and Health Statistics. Washington, DC: U.S. Government Printing Office.
- ▶ Coleman, J.S., (1958) Relational Analysis: the Study of Social Organizations with Survey Methods. Human Organization. 17, 28-36. <http://sfaa.metapress.com/content/q5604m676260q8n7/>
- ▶ Czaja R., Blair J., (1988), Using Network Sampling for Rare Populations: An Application to Local Crime Victimization Surveys, Proceedings of the Survey Research Methods Section, American Statistical Association, 38-43, http://www.amstat.org/sections/SRMS/Proceedings/papers/1988_007.pdf
- ▶ FAO (1996), Multiple Frame Agricultural Surveys, Volume I
- ▶ Fattorini L., Ghellini G., (2009), Estimating the size of street-dwelling populations by means of mark-resighting counts: theoretical considerations and empirical results, Proceedings of the XXVI IUSSP International Population Conference, <http://iussp2009.princeton.edu/abstracts/90984>
- ▶ Goodman, L.A. (1961) Snowball Sampling. Annuals of Mathematical Statistics. 32, 148-170. <http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aoms/1177705148>
- ▶ Haldane, J.B.S. (1945) On a Method of Estimating Frequencies. Biometrika. 33 (3) 222-225. <http://www.jstor.org/stable/2332299>
- ▶ Jędrzejczak, A., Kubacki, J., (2013), Estimation of Income Inequality and Poverty in Poland by Region And Family Type, Statistics in Transition-new series (w przygotowaniu).
- ▶ Kalton G., Anderson D.W., (1986), Sampling Rare Populations, Journal of the Royal Statistical Society A 149, 65-82
- ▶ Kalton, G. (1991). Sampling flows of mobile human populations. Survey Methodology, 17, 183—194.
- ▶ Klovdahl, A.S. et al. (1977) Social Networks in an Urban area: First Canberra Study. Australian and New Zealand Journal of Sociology. 3 (2), 169-172.

Literatura

- ▶ Kubacki,J., Piasecki,T., (2009) Opracowanie modelu/estymacji danych w PSR 2010 z wykorzystaniem danych z systemów informacyjnych i badań statystycznych - Ośrodek Realizacji Badań, Urząd Statystyczny w Łodzi, raport powstały w ramach prac podgrupy d/s Statystyczno Matematycznych w ramach prac przygotowawczych do PSR2010
- ▶ Kubacki,J., Jędrzejczak,A., Piasecki,T., (2011) Wykorzystanie metod statystyki małych obszarów do opracowania wyników badań statystycznych, Raport z pracy metodologicznej 3.065, Ośrodek Statystyki Matematycznej, Urząd Statystyczny w Łodzi, 2011
- ▶ Lohr, S. (1999), Sampling: Design and Analysis, Duxbury Press
- ▶ Lohr, S. (2007), Recent Developments in Multiple Frame Surveys, Proceedings of the Survey Research Methods Section, American Statistical Association, 3257-3264, <http://www.amstat.org/sections/SRMS/Proceedings/y2007/Files/JSM2007-000580.pdf>
- ▶ Lucas, H.A., Seber, G.A.F. (1977) Estimating Coverage and Particle Density Using the Line Intercept Method. Biometrika. 64 618-622, <http://www.jstor.org/stable/2345342>
- ▶ Purcell N.J., Kish L., (1979) Estimation for Small Domains, Biometrics, 35, s. 365-384, <http://www.jstor.org/stable/2530340>
- ▶ Seber, G.A.F. (1973) The Estimation of Animal Abundance. London, Griffin.
- ▶ Sirken, M.G., (1998) A Short History of Network Sampling, Proceedings of the Survey Research Methods Section, American Statistical Association, 1-6, http://www.amstat.org/sections/SRMS/Proceedings/papers/1998_001.pdf
- ▶ Thompson, S.K., (2002) Sampling. 2nd ed., New York, Wiley
- ▶ Thompson, S.K., (2006) Adaptive web sampling. Biometrics. 62, 1-24. <http://www.jstor.org/stable/4124545>
- ▶ Thompson, S. K., i Ramsey, F. L. (1983). Adaptive sampling of animal populations. Technical Report 82 . Corvallis, OR: Department. of Statistics, Oregon State University.)