

Rola statystyki w nauce i życiu społecznym

Mirosław Szreder
Uniwersytet Gdański
Komitet Statystyki i Ekonometrii PAN

Ogólnopolska konferencja naukowa
„Społeczne wyzwania edukacji statystycznej”
Gdańsk – Gdynia, 17-18 marca 2016 r.

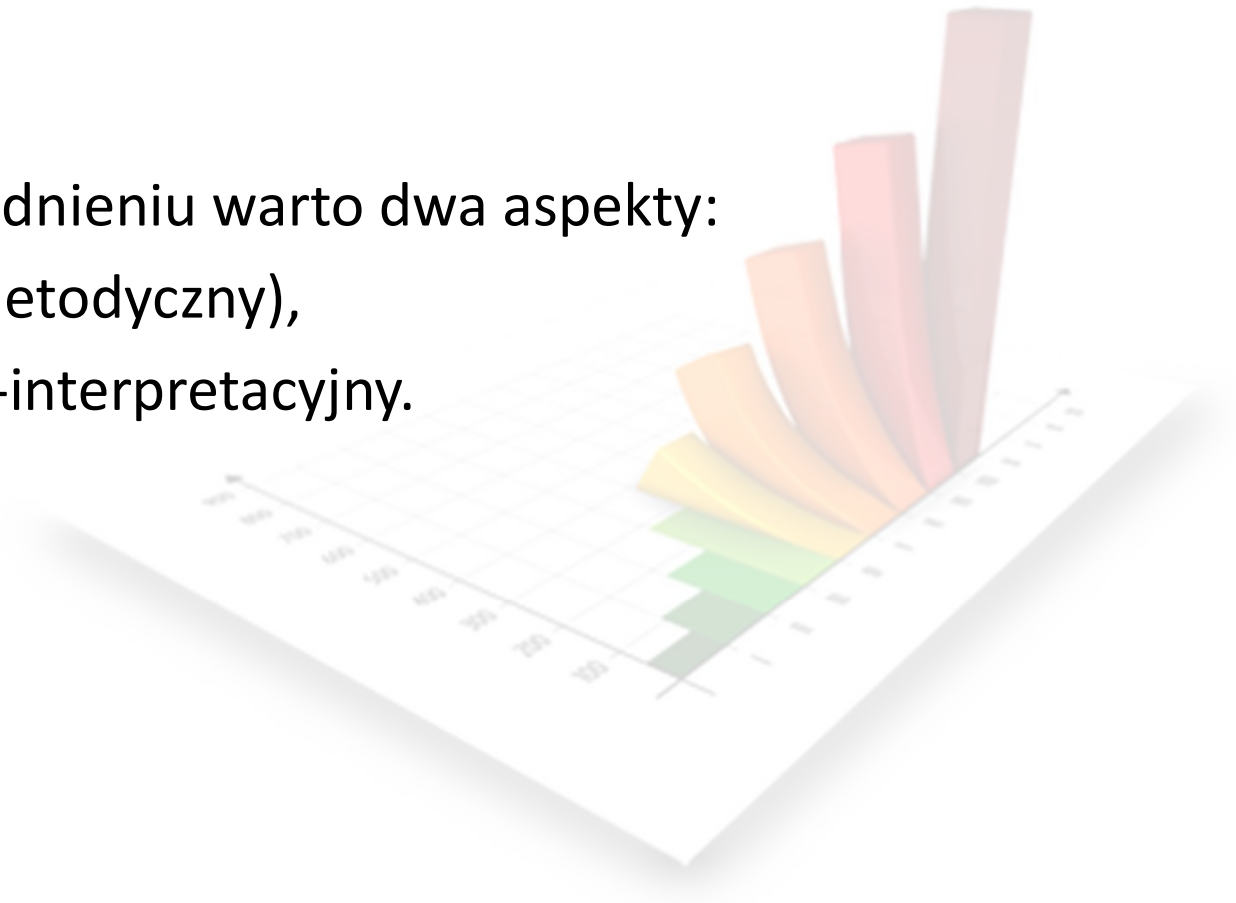


Cel rozważań:

Na ile statystycy oraz odbiorcy ich opracowań i komunikatów, czyli my wszyscy, jesteśmy przygotowani do krytycznego oglądu świata „zanurzonego w danych” (ang. *the world awash with data*)?

Dostrzec w tym zagadnieniu warto dwa aspekty:

- o badawczy (metodyczny),
- o dydaktyczno-interpretacyjny.

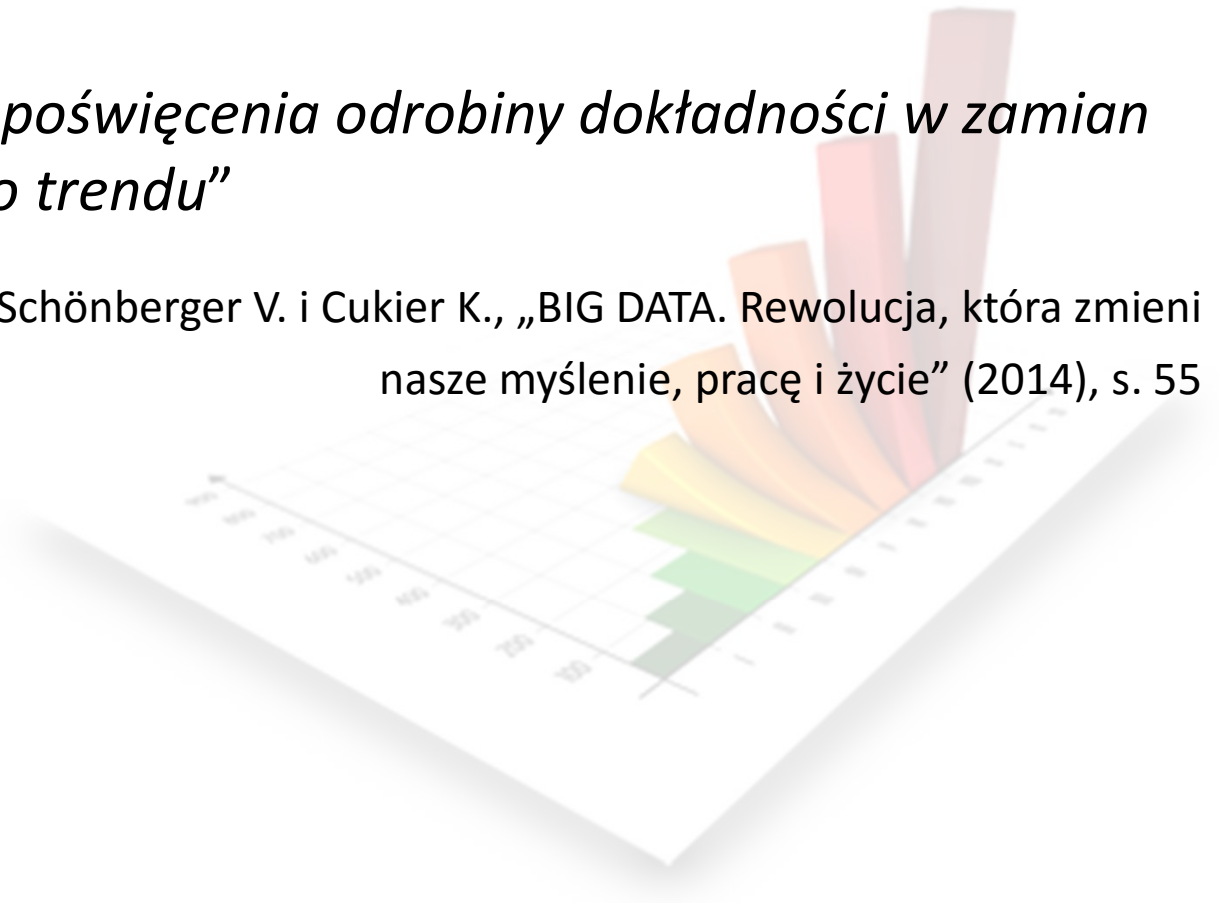


Aspekt badawczy (metodyczny)

Złudzenie że duże liczby obserwacji, bez względu na ich pochodzenie lub mechanizm generowania, są w stanie zrekompensować niedostatki jakości.

„Jesteśmy gotowi do poświęcenia odrobiny dokładności w zamian za poznanie ogólnego trendu”

Mayer-Schönberger V. i Cukier K., „BIG DATA. Rewolucja, która zmienia nasze myślenie, pracę i życie” (2014), s. 55



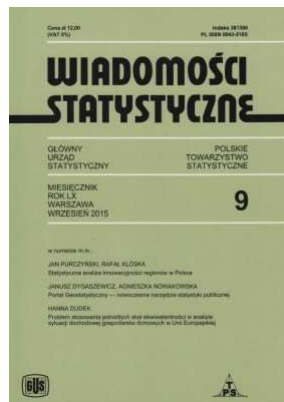
Aspekt badawczy (metodyczny)

Prawdziwym wyzwaniem jest umiejętność wykorzystania stosunkowo łatwo dostępnych danych o populacji w celu poprawy jakości wnioskowania.

Chodzi przede wszystkim o większą kontrolę i redukcję błędów nielosowych w badaniach reprezentacyjnych.



Szreder M., Zmiany w strukturze całkowitego błędu badania próbkowego. *Wiadomości Statystyczne* nr 1, 2015 r., s. 4-12.

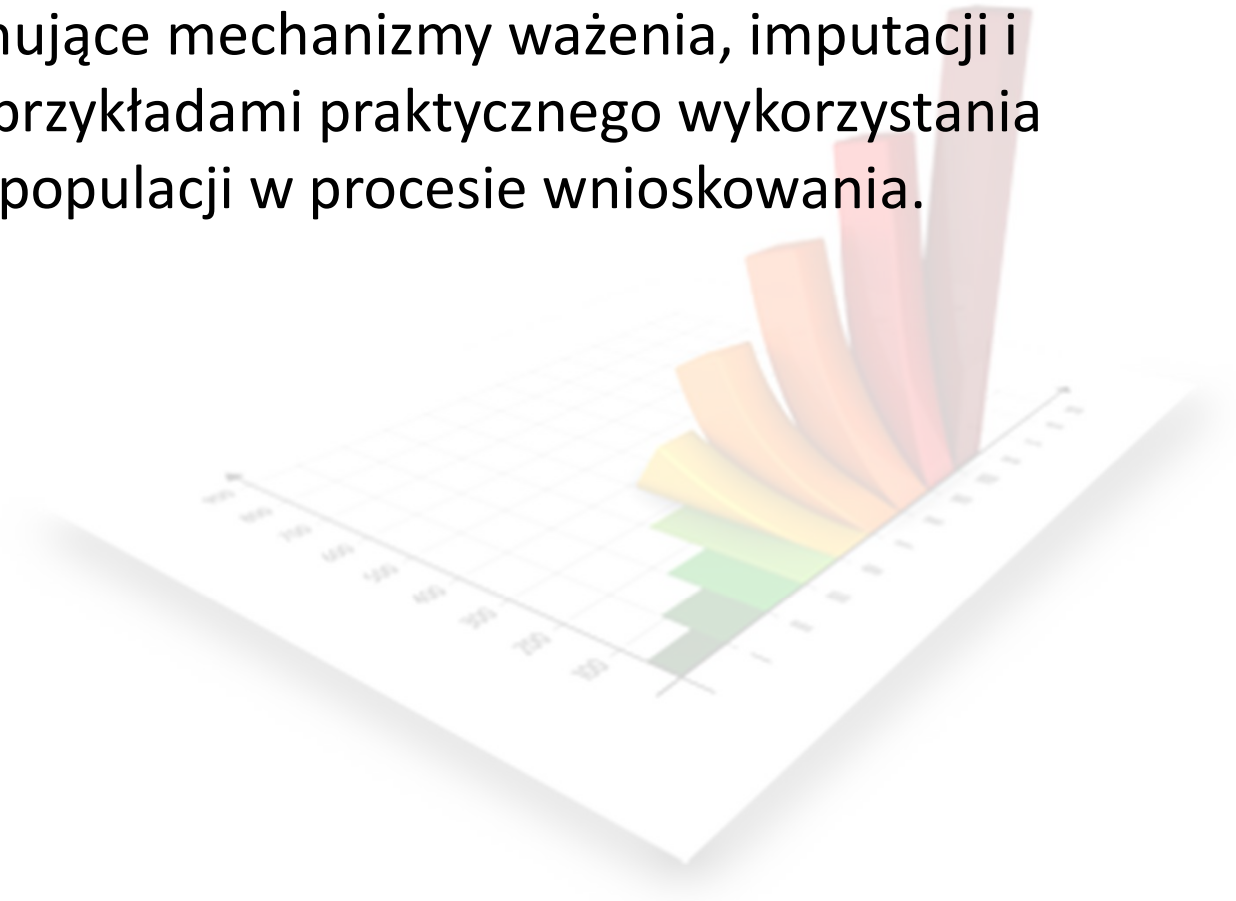


Stefanowicz B., Cierpiął-Wolan M., Błędy przetwarzania danych. *Wiadomości Statystyczne* nr 9, 2015 r., s. 23-29.

Aspekt badawczy (metodyczny)

Duże zbiory danych nie stanowią poważnego zagrożenia dla badań reprezentacyjnych.

Coraz lepiej funkcjonujące mechanizmy ważenia, imputacji i kalibracji danych są przykładami praktycznego wykorzystania informacji a priori o populacji w procesie wnioskowania.



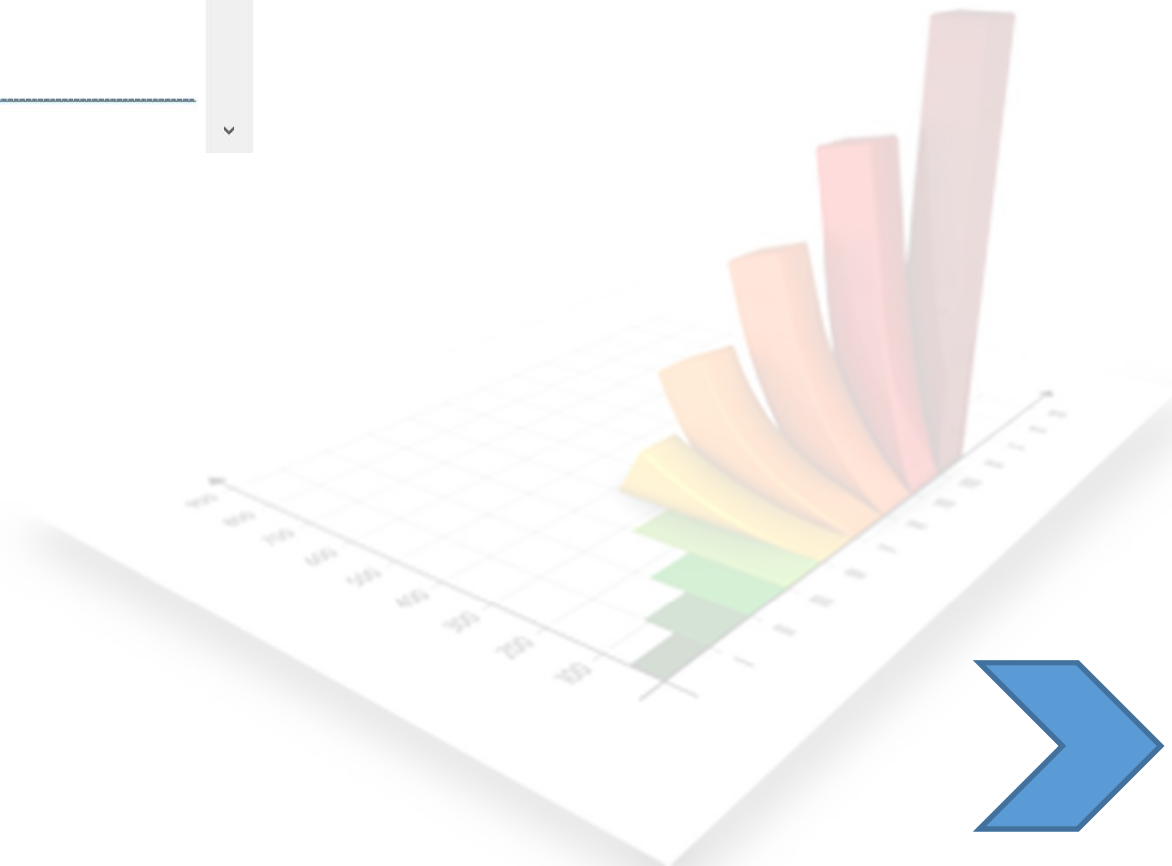
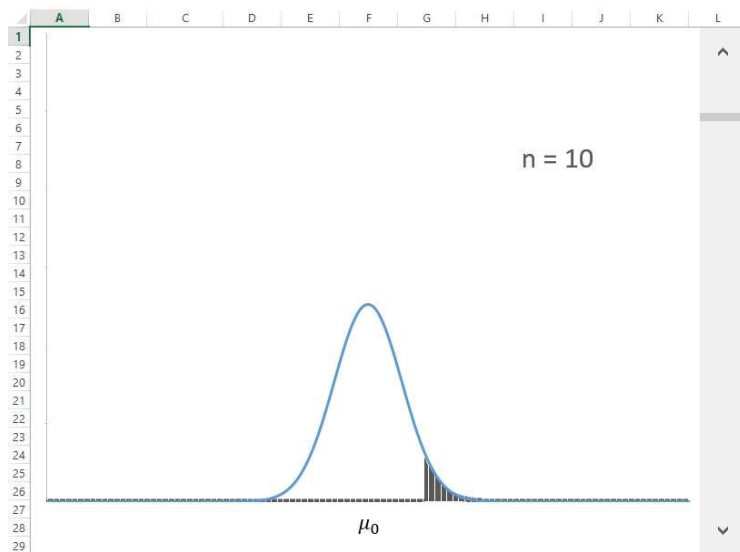
Liczebność próby a jakość wnioskowania

Znany jest pozytywny wpływ rosnącej liczebności próby na moc testu, czyli na jego zdolność do prawidłowego rozstrzygnięcia o prawdziwości lub nieprawdziwości testowanej hipotezy.

Błędy I i II rodzaju mają charakter losowy.

Ze wzrostem liczebności próby maleje dyspersja rozkładu statystyki testowej, co w warunkach braku zakłóceń spowodowanych czynnikami nielosowymi oznacza lepszą moc testu.

Przy bardzo małej dyspersji rozkładu statystyki z próby (a w konsekwencji małym obszarze nieodrżucenia hipotezy), każde nawet niewielkie uchybienie pomiarowe, albo inny błąd systematyczny, będą prowadzić do odrżucenia hipotezy zerowej.



Wniosek 1.

Duże liczebnie próby powodują większą wrażliwość testu na błędy o charakterze nielosowym, w tym w szczególności na błędy systematyczne.

Wniosek 2.

Dla bardzo dużych prób typowe poziomy istotności (5% lub 10%) będą determinowały na tyle duże obszary krytyczne, że prawie zawsze podjęta zostanie decyzja o odrzuceniu hipotezy zerowej.

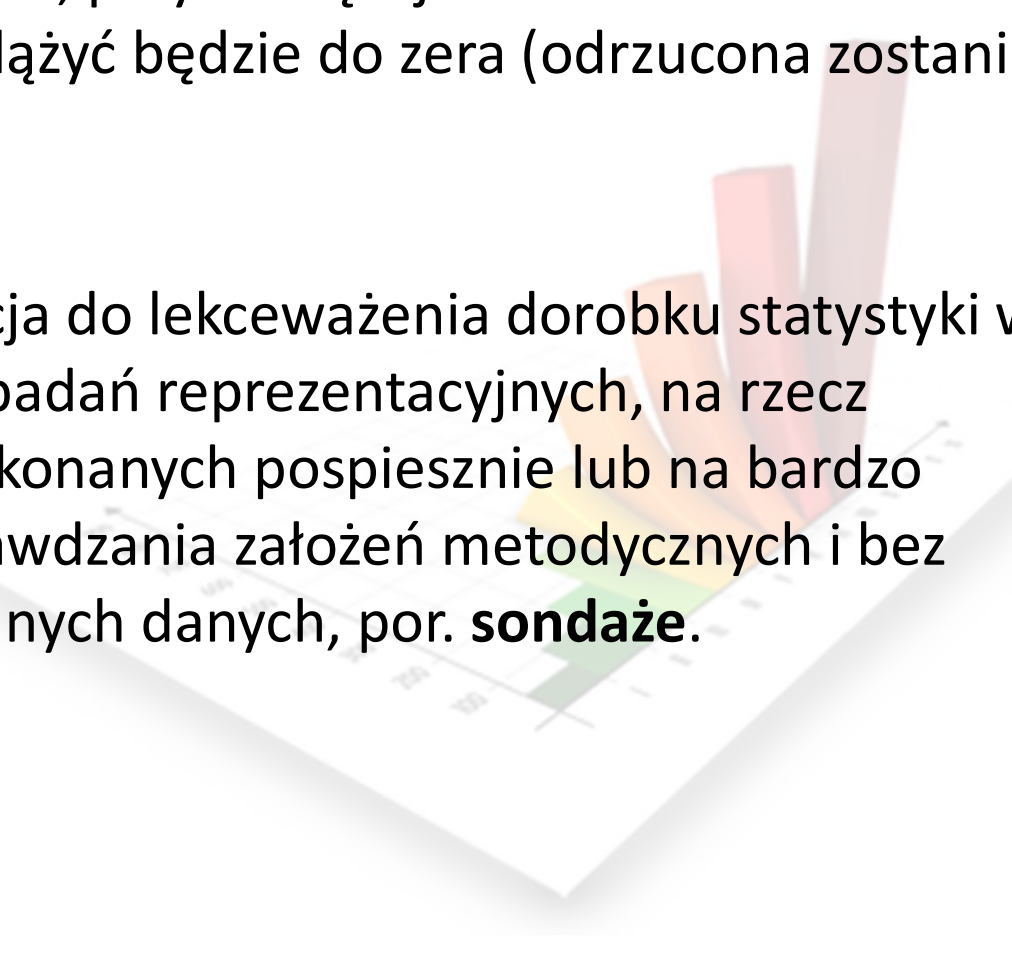
„One ad hoc remedy used by practitioners is to change the level of significance with the sample size, making it harder to reject the null hypothesis for large samples than for small ones”.

Kmenta J., Elements of econometrics, (1990), s. 128

Analogiczny wniosek (w kategoriach p-value):

Z wyjątkiem sytuacji, kiedy parametr populacji jest dokładnie równy wartości ujętej w H_0 , przy rosnącej do nieskończoności wielkości próby p-value dążyć będzie do zera (odrzucona zostanie hipoteza zerowa).

Niepokojąca jest tendencja do lekceważenia dorobku statystyki w zakresie teorii i praktyki badań reprezentacyjnych, na rzecz większej liczby badań wykonanych pośpiesznie lub na bardzo dużych próbach, bez sprawdzania założeń metodycznych i bez dbałości o jakość pozyskanych danych, por. **sondaże**.



Sondaże przedwyborcze

O czym informują i czy można je traktować jako prognozę wyborczą?

Wielka Brytania – wybory parlamentarne 7 maja 2015 r.

	CON	LAB	LIB DEM	UKIP	GREEN	OTHER	METHOD	SAMPLE SIZE	FIELDWORK
	%	%	%	%	%	%		N	
Opinium	35	34	8	12	6	5	online	2 960	May 4-5
Survation	31	31	10	16	5	7	online	4 088	May 4-6
Ipsos MORI	36	35	8	11	5	5	telephone	1 186	May 5-6
ICM	34	35	9	11	4	7	telephone	2 023	May 3-6
ComRes	35	34	9	12	4	6	telephone	2 015	May 3-5
Populus	33	33	10	14	5	6	online	3 917	May 5-6
YouGov	34	34	10	12	4	6	online	10 307	May 4-6
Panelbase	31	33	8	16	5	7	online	3 019	May 4-6
<i>Average</i>	33.6	33.6	9	13	4.8	6.1			
<i>Result</i>	37.8	31.2	8.1	12.9	3.8	6.3			
<i>Difference</i>	-4.2	2.4	0.9	0.1	1	-0.2			

Prof. Patrick Sturgis, przewodniczący brytyjskiej komisji do zbadania źródeł błędnych wyników przewidywań rezultatu wyborczego:

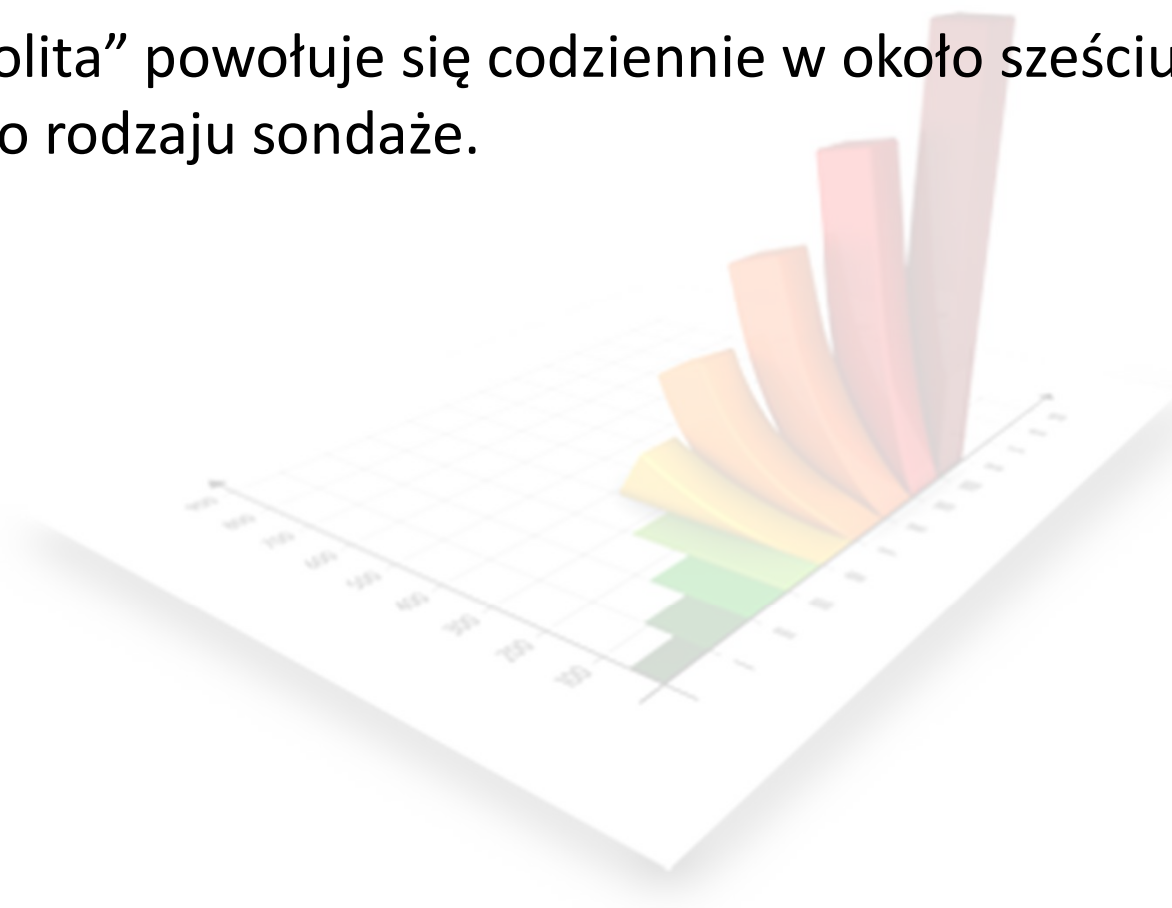
- nie wyklucza, że znaczenie miały obawy części ośrodków badawczych, które nie decydowały się na publikowanie wyników zasadniczo odbiegających od ogólnego trendu (ang. 'herding'),
- potrzebne jest przesunięcie punktu ciężkości z ilości na jakość („there needs to be a shift in emphasis away from quantity towards quality”).



Aspekt edukacyjny

Edukacja szkolna i pozaszkolna

Dziennik „Rzeczpospolita” powołuje się codziennie w około sześciu artykułach na różnego rodzaju sondaże.



Dnia 16.02.2016 r.

Ile zostało z exposé Beaty Szydło

*„Według ostatnich **sondaży** rządząca partia jest wciąż najpopularniejsza ...”*

Przyjazny 40-latek łatwiej znajdzie pracę po pięćdziesiątce

*„Potwierdza to **sondaż** agencji zatrudnienia Work Service ...”*

Lech Wałęsa: Gross powinien oddać medal

*„Tak wynika z **sondaży**”*

Kijów: żegnanie premiera

*„Obecne **sondaże** wskazują, że w nowym głosowaniu ...”*

Izrael traci w Niemczech

*„Jak wynika z **sondaży**, 48 proc. Niemców ma złą opinię o ...”*

PO chce dymisji Błaszczaka

*„... to reakcja na ostatnie **sondaże** ...”*

Sprawa Maja? Błaszczak: Problem jakości służb

*„Wystarczy spojrzeć na wczorajszy **sondaż** TNS Polska ...”*

Brytyjskie firmy szykują się do Brexit

*„Niedawne **sondaże** wskazują na niewielką, ale rosnącą przewagę zwolenników ...”*

Dnia 17.02.2016 r.

Katoliccy biskupi popierają Merkel ws. uchodźców

*„**Sondaże** pokazują, że z powodu reakcji rządu Niemiec ...”*

Rzecz o prawie: Nie wracajcie do Babilonu

*„... spektakularny sygnał do opinii publicznej dający wzrost poparcia w **sondażach** ...”*

Polacy podzieleni w ocenie rządu

*„2 punkty procentowe więcej niż w styczniowym **sondażu** ...”*

Obama nie wyraził poparcia dla Clinton

*„ ... faworyt republikańskich **sondaży** miliarder i ... „*

Szef KE chwali Merkel za politykę imigracyjną

*„**Sondaże** pokazują, że ... poparcie dla Merkel spadło ...,,*

Ilościowa narracja w PiS

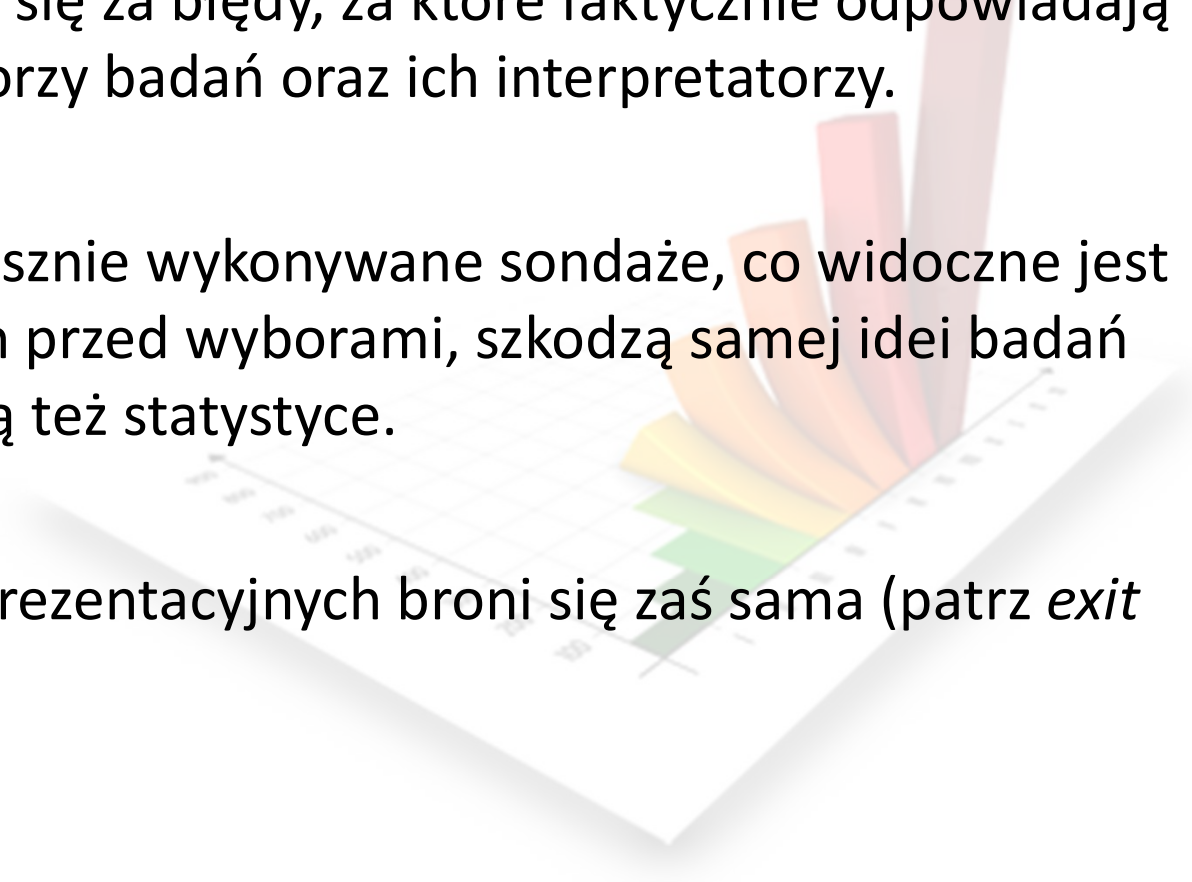
*„... bo w **sondażach**, nawet tych nieprzychylnych, PiS utrzymuje swoją przewagę”*

Z braku umiejętności krytycznej oceny sondaży powstaje wiele nieporozumień, bezproduktywnych dyskusji i ogólnego szumu informacyjnego.

Naukę naszą obwinia się za błędy, za które faktycznie odpowiadają projektanci i realizatorzy badań oraz ich interpretatorzy.

Niestarannie i pośpiesznie wykonywane sondaże, co widoczne jest zwłaszcza w okresach przed wyborami, szkodzą samej idei badań próbkowych i szkodzą też statystyce.

Metodyka badań reprezentacyjnych broni się zaś sama (patrz *exit poll*).



Exit poll

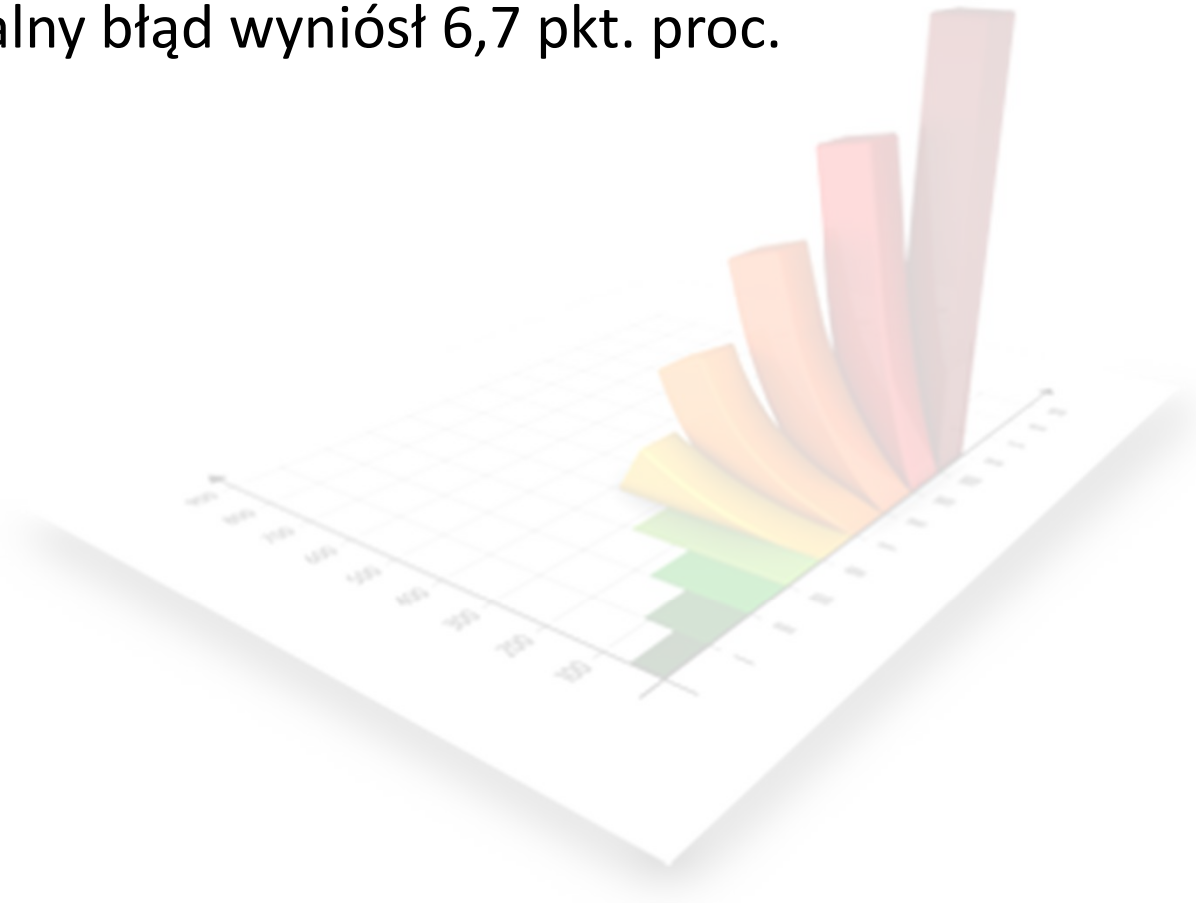
W minionych 10 latach znacznie poprawiła się precyzja badań *exit poll* w Polsce

- TNS Polska (TNS OBOP) podał w obu turach wyborów prezydenckich 2010 roku oraz w wyborach parlamentarnych 2011 roku wyniki prognoz *exit poll* różniące się od ostatecznych wyników o **mniej niż 1 pkt. proc.**
- W referendum w Warszawie 13 października 2013 r. błąd był **mniejszy niż 1,1 pkt. proc.**
- IPSOS podał wyniki I tury wyborów prezydenckich br. z błędem **mniejszym niż 1,6 pkt. proc.**, a w II turze z błędem **mniejszym niż 1,5 pkt. proc.**
- W wyborach parlamentarnych br. błąd nie przekraczał **1,6 pkt. proc.**

Exit poll

W minionych 10 latach znacznie poprawiła się precyzja badań *exit poll* w Polsce

- Wyjątek – wybory samorządowe 16 listopada 2014 r., w których maksymalny błąd wyniósł 6,7 pkt. proc.



Inne badania statystyczne budzące największe zainteresowanie:
płace i dochody.



Sławomir Sierakowski w swoim artykule pt.
„Sprzedana demokracja?” pisze, że:

Główny Urząd Statystyczny oszukuje nas podając comiesięczny komunikat o średniej płacy w sektorze przedsiębiorstw, bo dwie trzecie Polaków zarabia poniżej średniej krajowej.

Pensja w portfelu i w Excelu

Zawsze chcemy wiedzieć ile zarabiają inni.
Czy dowiemy się tego od statystyków



MIROSŁAW SZREDER

Sławomir Sierakowski w tekście „Sprzedana demokracja?” (POLITYKA 50/15) twierdzi, że GUS wprowadza w błąd, podając comiesięczny komunikat o średniej płacy w sektorze przedsiębiorstw albo co kwartał o średniej płacy w gospodarce, i sugeruje, aby przeciętne zarobki charakteryzować nie średnią arytmetyczną, lecz medianą lub dominantą. Zanim bliżej przyjrzymy się tej propozycji, warto rozważyć, po co w ogóle potrzebna jest informacja o średnich zarobkach.

Pierwszy powód wydaje się oczywisty – każdy pragnie wiedzieć, jak się ma jego comiesięczna pensja do zarobków innych osób. Naturalne jest więc poszukiwanie jakiegoś punktu odniesienia, bez którego 3 tys. zł może się wydawać albo dużo, albo mało. Na przykład 20 lat temu pensja tej wysokości uważana była za wyjątkowo wysoką – stanowiła bowiem cztero- lub pięciokrotność średniej krajowej. Dzisiaj z kolei stanowi zaledwie ok. 3/4 średniej płacy, a dodatkowo jest mniejsza od mediany, co oznacza, że ponad połowa zatrudnionych zarabia więcej niż 3 tys. zł miesięcznie. Dla wielu osób jest mniej ważne, ile za te 3 tys. zł mogą kupić towarów i usług,

czyli jaka jest siła nabywcza ich płacy, a ważniejsze to, czy płaca ich osiąga przynajmniej poziom średniej krajowej.

Statystyka bez porównań jest uboga, co ilustruje anegdota o tym, jak się wita statystycy. Jeden z nich pyta: „Jak się miewa twoja żona?”. Na co drugi, po chwili zawałowania, odpowiada: „W porównaniu z kim?”. Wyniki statystyczne zyskują właściwy i pełniejszy kontekst dopiero po skonfrontowaniu ich z innymi jednostkami zbiorowości lub innymi jej charakterystykami. Tego typu odniesienia, na przykład dynamiki średniej płacy do zmiany w czasie PKB, dają nowe możliwości interpretacyjne, na przykład: w jakim zakresie uczestniczymy w owo- cach wzrostu gospodarczego.

Innym powodem comiesięcznych komunikatów GUS o średniej płacy w przedsiębiorstwach jest dążenie do przekazania informacji o poziomie płacy pomocą jednego syntetycznego wskaźnika. Problem tkwi jednak w tym – jakiego wskaźnika? Najlepiej byłoby oczywiście zaprezentować cały rozkład płac w populacji zatrudnionych, a więc procentowe udziały pracowników w poszczególnych przedziałach zarobków, od najniższych do najwyższych. Nie jest to jednak ani

możliwe, ani – jak się wydaje – pożądane. Oznaczałoby to bowiem konieczność znacznie bardziej szczegółowego raportowania przez przedsiębiorstwa o wysokości pracowniczych wynagrodzeń, i to co miesiąc. Tak szczegółowe dane są zbierane przez GUS jedynie raz na dwa lata w badaniu reprezentacyjnym, obejmującym tylko wylosowane do próby przedsiębiorstwa. Badanie to jest i czasochłonne, i kosztowne. Ostatnie dane na ten temat, które ogłosił GUS na początku grudnia 2015 r., dotyczą struktury wynagrodzeń w październiku 2014 r.

O ile wskazane byłoby wykonywanie takiego badania raz do roku, o tyle wątpliwe, czy istnieje rzeczywista potrzeba prezentowania szczegółowej struktury wynagrodzeń z częstotliwością miesięczną. Zwrócić bowiem uwagę, że w wielu dziedzinach życia gospodarczego i społecznego, gdzie istnieje łatwe dostępnienie danych szczegółowych na temat pewnych złożonych zjawisk, przywołujemy się najczęściej i tak do jednego tylko wskaźnika. Tak jest z oceną dynamiki wzrostu gospodarczego lub poziomu życia społecznego, gdzie powszechnie jest stosowany produkt krajowy brutto (PKB), mimo że posiada on sporo wad jako miernik syntetyczny. Na przykład

nie mierzy degradacji środowiska i wyczerpywania się zasobów naturalnych, a usługi publiczne są w nim wyrażone nie wynikami, lecz nakładami (w postaci wydatków rządowych).

W ostatnich kilkunastu latach nasiliła się tendencja do poszukiwania syntetycznych liczbowych miar wielu złożonych zjawisk. Odbiorcy tego typu informacji, a więc niemal my wszyscy, pragniemy jednej konkretnej liczby: wartości wskaźnika, procentu w sondażu, miejsca w rankingu. Mniej nas obchodzi treść, jakie się za tymi liczbami kryją, a pewnie jeszcze mniej konstrukcja tych zagregowanych wskaźników. Zamiast opisu rozkładu płac wolimy jedną jego charakterystykę.

Także i w tym dążeniu, podobnie jak w skłonności do porównań, wcale nie odbiegamy od postawy statystyków, od których oczekujemy, że tego typu adekwatne charakterystyki liczbowe będą w stanie zaproponować. W historii statystyki znanych jest wiele prób krótkiego liczbowego opisu pewnych regularności, które same w sobie były zawsze dość złożone.

Jeden z najstojniejszych statystyków XIX w. Adolf Quetelet obliczył na przykład, że „by mają największe szanse skwitować w Brukseli wtedy, gdy suma kwadratów średnich temperatur dziennych od ostatniego mrozu osiągnie 4264 stopnie”. Współczesnym odpowiednikiem tendencji do liczbowego opisu jakościowych cech zbiorowości jest mnogość różnorodnych indeksów (indeks wolności gospodarczej, ang. *index of economic freedom*; wskaźnik rozwoju społecznego, ang. *human development index*; indeks postrzegania korupcji, ang. *corruption perceptions index*), popularność technik ankietowych (punktowych) czy badań ankietowych. W każdej takiej sytuacji, o czym warto pamiętać, ceną za jednoznaczny liczbowy charakterystykę danej cechy (ilościowej bądź jakościowej) jest utrata wielu ważnych informacji. Nie inaczej rzecz się ma z rozkładem płac.

Zastąpienie przez jeden parametr informacji o wysokości, zróżnicowaniu czy koncentracji płac musi prowadzić do daleko idących uproszczeń. Chodzi jednak o to, aby te konieczne uproszczenia nie wykrzywiały prawdziwego obrazu opisywanego rozkładu. Odpowiedź więc na pytanie: czy średnia arytmetyczna na płac w stosunku do innych miar poziomu przeciętnego – w szczególności mediany i dominanty – fałszuje ten obraz czy nie?

Średniej arytmetycznej wytyka się najczęściej jej główną słabość, mianowicie wrażliwość na wartości skrajne. W przypadku płac w gospodarce Polski oznacza to, że bardzo wysokie płace – mimo że niewielkiej części zatrudnionych (6,4 proc. zarabia powyżej dwukrotnej

średniej) – podwyższają istotnie płacę średnią. Trzeba przyznać, że w tego typu rozkładach, mówiąc językiem statystyki – o silnej asymetrii prawostronnej – średnia arytmetyczna nie stanowi dobrej miary poziomu średniego. Stąd bierze się dyskomfort wielu pracujących, którzy za przeciętną płacę są skłonni uważać taką jej wysokość (medianę), poniżej i powyżej której lokuje się po 50 proc. zatrudnionych. W rozkładach płac jest to często wykorzystywany miernik, aczkolwiek również niepozbawiony wad.

Słabą stroną mediany jest to, że może się ona w ogóle nie zmieniać przy znacznych zmianach wysokości płac. Na przykład, gdyby wspomnianym wyżej 6,4 proc. najlepiej zarabiających podnieść płace o dowolnie dużą kwotę, to mediana płac – jako miernik poziomu średniego – nie ulegnie zmianie. Otrzymamy więc lepsze samopoczucie: średnia w gospodarce nie powinna rosnąć, skoro poprawiło się jedynie tym z najwyższymi zarobkami, ale z drugiej strony na własne życzenie otrzymamy nieprawdziwy obraz rzeczywistości. Trudno bowiem uznać za prawdziwe stwierdzenie, że przeciętnie w rozkładzie płac nie się nie zmieniło, skoro płace najlepiej zarabiających istotnie wzrosły. Medianę uznać więc można za dobre uzupełnienie informacji o średniej arytmetycznej, ale nie jako jej lepszą alternatywę. W niektórych krajach, gdzie zróżnicowanie płac jest większe niż w Polsce (np. w Wielkiej Brytanii), częściej operuje się jednak medianą płac niż średnią arytmetyczną.

Za jeszcze bardziej kontrowersyjną można uznać propozycję Sławomira Sierakowskiego o przywołaniu wcześniej artykułu, aby dominującą w gospodarce płacę (dominantę) traktować jako uogólnienie wysokości zarobków („to są realne zarobki milionów Polaków” – pisze autor). Fakt, że największy odsetek zatrudnionych stanowią ci, których zarobki oscylują wokół dominanty, nie mówi prawie nic o przeciętnym poziomie zarobków. Jeżeli 26 proc. osób zarabowało w 2014 r. pomiędzy 50 proc. a 75 proc. średniego wynagrodzenia (wg danych GUS), to z samej tej informacji nie wynika, jak rozłożone są płace pozostałych ponad trzech czwartych zatrudnionych. A na dodatek płace tej większości mogą się istotnie zmieniać, nie wpływając w ogóle na wartość dominanty. Podobne słabości mediany i dominanty biorą się stąd, że obie one należą do tzw. miar pozycyjnych poziomu przeciętnego, stąd nie reagują na różne zmiany w wartościach analizowanej cechy u poszczególnych jednostek. Miary pozycyjne uważa się za mniej precyzyjne charakterystyki syntetyczne rozkładu od miar klasycznych, do których zalicza się m.in. średnią arytmetyczną.

Gdyby comiesięczną informację o średniej płacy w przedsiębiorstwach wzbogacić o medianę i dominantę płac, to obraz rozkładu płac byłby pełniejszy. Są to miary względem siebie komplementarne. Przeszkodą uniemożliwiającą obecnie podawanie wszystkich tych trzech wielkości jest inny zakres informacji statystycznych potrzebnych do ich obliczenia.

Średnią arytmetyczną płac daje się łatwo wyliczyć, znając jedynie sumę wykorzystanego w danym miesiącu funduszu płac w przedsiębiorstwach i liczbę zatrudnionych. Z kolei wyznaczenie mediany i dominanty wymagałoby dokładniejszych informacji o wysokości zarobków wszystkich zatrudnionych w firmach.

Należy postulować zwiększenie częstotliwości wykonywania przez GUS reprezentacyjnego badania struktury płac do przynajmniej raz w roku. Przy udoskonalonym planie próbkowania i zmniejszeniu dość dużej obecnie próby (29,7 tys. przedsiębiorstw zatrudniających powyżej 9 osób, tj. 13 proc. populacji) nie musiałoby się to wiązać z istotnym zwiększeniem nakładów finansowych na badanie. W dłuższej zaś perspektywie informacje o płacach mogłyby być możliwe do pozyskania od przedsiębiorstw w zintegrowanych komputerowych systemach sprawozdawczych. Skoro dane o wypłatach kwotach wynagrodzeń są i tak rejestrowane co miesiąc w bazach danych przedsiębiorstw, to dlaczego nie miałyby się one stać elementem zasobów informacyjnych statystyki publicznej? Wówczas rozkład płac można byłoby charakteryzować nie tylko średnią arytmetyczną, ale dowolną miarą poziomu przeciętnego – co ważne – zróżnicowania (dyspersji). Równie bowiem gorzej jak o średniej płacy bywają dyskusje o zróżnicowaniu i koncentracji rozkładu płac.

Autor jest statystykiem, obecnie pełni funkcję prorektora Uniwersytetu Gdańskiego.

REKLAMA



„Odmiany biedy, czyli Gini nie mówi wszystkiego” (s. 181-194).

Autor wskazuje, że w szczególności wskaźnik ten niewiele mówi o dochodach najbiedniejszej części społeczeństwa.

Pisząc o zagrożeniu ubóstwem, odsyła czytelników do głębszych badań statystyki publicznej: „Obszary polskiej nędzy z dużą starannością zostały przebadane przez GUS” (s. 188)

„Ubóstwo ekonomiczne w Polsce w 2014 roku (na podstawie badania budżetów gospodarstw domowych)”, GUS, 2015 r.

Problem ten jest bardziej złożony, na co wskazuje m.in. dr hab. Michał Brzeziński („Gazeta Wyborcza” z 24.07.2015 r.) w artykule pt. „Trzeba patrzeć w PIT-y”

M. Kośny i E. Mazurek, *Redistribution and Equity of Polish Personal Income Tax: Measurement Using Micro Data from Tax Returns*, "Statistica & Applicazioni", 7(2), 2009,

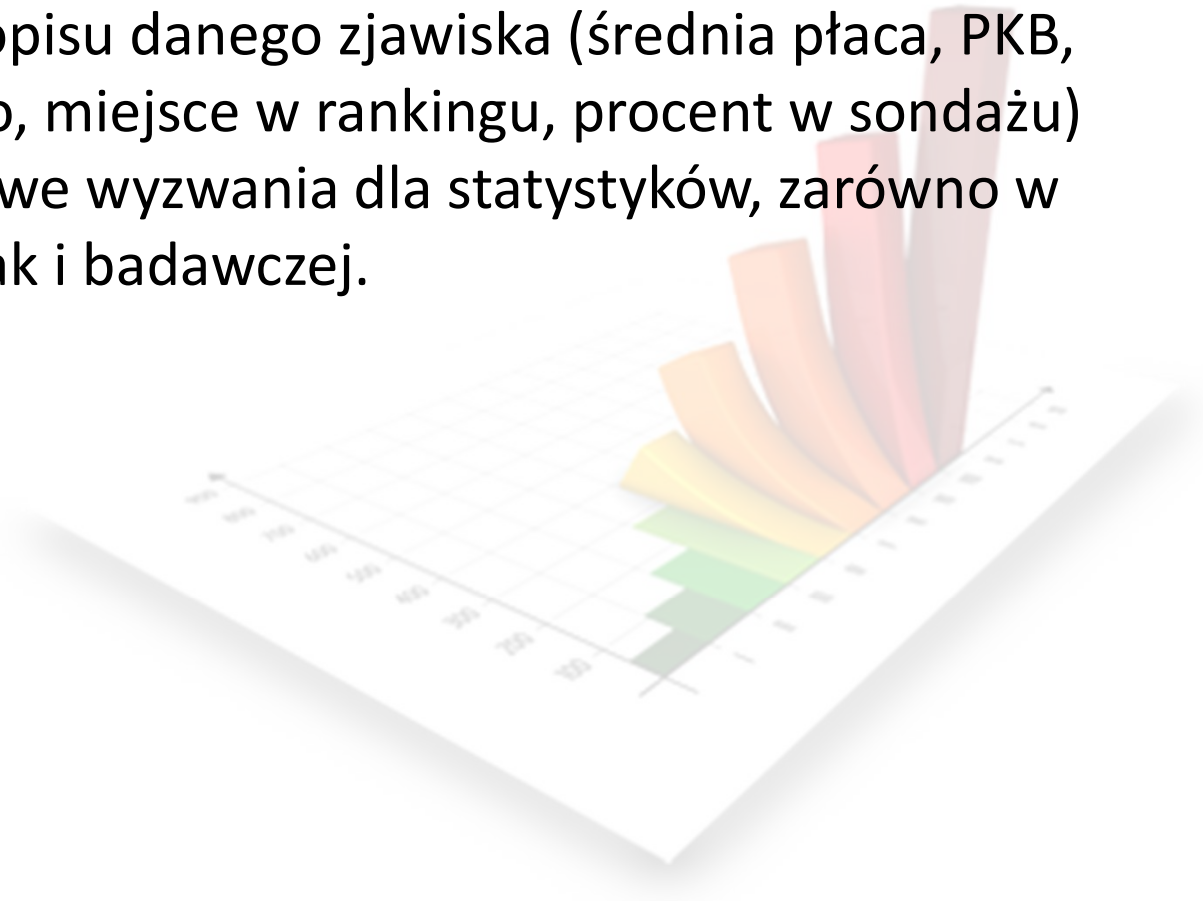
M. Kośny i E. Mazurek, *Influence of Child Tax Credit on Inequity of Personal Income Tax in Poland*, "International Research Journal of Finance and Economics", 50, 2010.

Analizy M. Kośnego pokazują, że według badań budżetów GUS w roku 2010 w województwie dolnośląskim udział dochodów 10 proc. najczęściej zarabiających podatników w całkowitym dochodzie podatników stanowił ok. 27 proc. Tymczasem z zeznań podatkowych wynika, że udział ten to aż ok. 40 proc.!

O ile według danych ankietowych indeks Giniego wynosi około 0,33, o tyle dane podatkowe sugerują wielkość zbliżoną do 0,44.

Konkluzja:

Tendencja do posługiwania się w środkach masowej komunikacji jedną, prostą miarą opisu danego zjawiska (średnia płaca, PKB, współczynnik Giniego, miejsce w rankingu, procent w sondażu) utrwała się i rodzi nowe wyzwania dla statystyków, zarówno w sferze edukacyjnej, jak i badawczej.



Dziękuję za uwagę!

