

STUDIA METODOLOGICZNE

Mirosław SZREDER

O niektórych nowych wyzwaniach i oczekiwaniach wobec statystyki¹

Streszczenie. *W artykule omówiono najnowsze wyzwania badawcze i dydaktyczne, jakie stoją przed statystyką w obliczu szybko rosnących możliwości gromadzenia i przetwarzania danych (Big Data) oraz coraz bardziej dominującego w życiu społecznym i ekonomicznym liczbowego opisu rzeczywistości. Te nowe zjawiska i tendencje powinny skłonić środowisko statystyków do podejmowania działań na rzecz podniesienia w społeczeństwie wiedzy statystycznej, niezbędnej do krytycznej oceny i poprawnej interpretacji wyników badań statystycznych.*

Słowa kluczowe: Big Data, weryfikacja hipotez statystycznych, wielkość próby, sondaż.

W świecie, w którym zasoby dostępnych informacji liczbowych rosną w tempie nigdy wcześniej niespotykanym, a w analizach rzeczywistości dominować zaczyna opis ilościowy, rośnie też rola statystyki, zarówno w jej wymiarze badawczym (metodycznym) jak i edukacyjnym. W ostatnich latach jesteśmy

¹ Artykuł opracowany na podstawie referatu pt. *Spoleczne wyzwania edukacji statystycznej* wygłoszonego podczas ogólnopolskiej konferencji naukowej z okazji obchodów Dnia Statystyki Polskiej, Gdynia, 17 i 18 marca 2016 r.

świadkami zderzenia się wielkich możliwości przetwarzania i analizowania danych liczbowych z ogólnie słabo przygotowanym do krytycznego odbioru tych analiz społeczeństwem. O tym, jakie z tego wynikają nowe obowiązki dla statystyków, zwłaszcza w sferze edukacyjnej, ale także badawczej, traktuje ten artykuł. Wyzwań w tym zakresie jest wiele, bo każdy z nas staje się w coraz większym stopniu odbiorcą wielu komunikatów o charakterze liczbowym, ale i ich twórcą, z racji łatwego dostępu do wielu danych statystycznych, różnej zresztą jakości.

WYZWANIA SFERY BADAWCZEJ

Dla teorii statystyki, a zwłaszcza dla wnioskowania statystycznego, sytuacja dostępu do bardzo dużych zbiorów danych oraz do narzędzi służących ich przetwarzaniu i analizowaniu, a więc do wszystkiego, co kryje się pod pojęciem Big Data, stwarza nową sytuację. O ile bowiem w przeszłości najczęstszym problemem w praktycznym zastosowaniu wnioskowania statystycznego była mała liczebność próby czy trudność ze zdobyciem danych w ogóle, o tyle obecnie w wielu dziedzinach życia znajdujemy raczej potwierdzenie tego, że „świat jest zanurzony w danych” (*the world awash with data*). Dla niektórych teoretyków statystyki ta nowa sytuacja stanowi wystarczający powód, by odrzucić wszelkie badania próbkowe, w tym w szczególności badania reprezentacyjne, oparte na niedużych, lecz starannie wyselekcjonowanych próbach losowych. Osoby te ulegają złudzeniu, że duże liczby obserwacji, bez względu na ich pochodzenie lub mechanizm generowania, są w stanie zrekomensować niedostatki jakości. *Jesteśmy gotowi do poświęcenia odrobiny dokładności w zamian za poznanie ogólnego trendu*² — piszą w tym kontekście autorzy głośnej książki pt. *BIG DATA. Rewolucja, która zmieni nasze myślenie, pracę i życie*. Tego typu bezkrytyczna wiara w to, że więcej znaczy zawsze lepiej jest być może pierwszym ważnym wyzwaniem dla statystyków, związanym z samym fundamentem badawczym statystyki jako nauki. Wśród manifestujących taką postawę dostrzec można skłonność do lekceważenia ważnego atrybutu danych statystycznych, jakim jest ich jakość. Pozostają oni pod wrażeniem dużej ilości danych, co ma być wystarczającym usprawiedliwieniem, by nie martwić się o ocenę ich jakości. Zapominają, że najczęściej ilość nie jest w stanie zrekomensować słabej jakości. Z kolei nieprzywiązywanie wagi do jakości idzie często w parze z brakiem respektu dla założeń metodycznych, na których oparte są metody wnioskowania statystycznego. Stąd nawiązywanie przez niektórych badaczy do praw wielkich liczb, aby uzasadnić, że duża liczba obserwacji zawsze prowadzi do określenia rzeczywistego trendu. W argumentacji takiej zapomina się jednak, że prawa wielkich liczb są zespołem ścisłych matematycznych twierdzeń o określonych założeniach, których niespełnienie czyni nieuzasadnionym odwoływanie się do tych twierdzeń. Innymi słowy, brak rygoru powtarzalności zdarzeń w jednakowych warunkach albo utożsamianie przypadkowych obserwacji z ob-

² Mayer-Schönberger, Cukier (2014), s. 55.

serwacjami w próbie losowej wykluczają możliwość stosowania tych twierdzeń. Gdyby tak nie było, to już obecnie wszelkie badania reprezentacyjne, w tym badania opinii publicznej i wiele badań społecznych, zostałyby zastąpione przez sondy internetowe, w których zwykle bierze udział kilkadziesiąt lub kilkaset razy więcej respondentów niż wynosi liczebność typowej próby badawczej.

Wydaje się więc, że same ułatwienia w dotarciu do dużej liczby obserwacji w populacji poddanej badaniu nie spowodują marginalizacji badań reprezentacyjnych. Jednym z ważnych wyzwań badawczych w tym zakresie jest natomiast kwestia umiejętności wykorzystania owych łatwo dostępnych danych o populacji w celu poprawy jakości wnioskowania. Chodzi przede wszystkim o większą kontrolę i redukcję błędów nielosowych w badaniach reprezentacyjnych³, ta kategoria błędów bowiem, a nie błąd losowania, przesądza coraz częściej o wiarygodności wyników wnioskowania. Błędy losowe mają — jak wiadomo — tendencję do zmniejszania się wraz ze wzrostem liczby obserwacji w próbie. Właściwości tej nie posiada natomiast większość błędów o charakterze nielosowym, m.in. błędy pokrycia, błędy spowodowane brakiem odpowiedzi czy błędy przetwarzania. Współczesne zasoby informacji o wielu populacjach podlegających badaniom, a także rosnące możliwości zdobycia informacji o innych cechach populacji niebędących celem badania czynią wręcz koniecznym łączenie informacji z próby losowej z informacjami *a priori*. Coraz lepiej funkcjonujące mechanizmy ważenia, imputacji i kalibracji danych są przykładami praktycznego wykorzystania szans, jakie w tym zakresie przynoszą osiągnięcia informatyki i statystyki.

W początkach XX w., kiedy powstawała teoria klasycznego wnioskowania statystycznego, a pewnie i dziesiątki lat później, badacze nie stawiali sobie zbyt wielu pytań o efektywność metod estymacji czy testowania hipotez w warunkach bardzo dużej liczebności prób. Prawdziwym problemem była mała liczba obserwacji dla zastosowań poszczególnych metod wnioskowania. Wydawać by się więc mogło, że obecna sytuacja, w której nietrudno o dużej wielkości próby, nie powinna rodzić wśród statystyków obaw. I rzeczywiście, w możliwościach uzyskania dużej liczby obserwacji widzimy przede wszystkim szanse, a rzadziej zagrożenia. Warto zdać sobie jednak sprawę, że nowa sytuacja stwarza pewne niebezpieczeństwa. Omówimy je krótko w odniesieniu do ważnego działu wnioskowania, jakim jest weryfikacja hipotez, a ściślej testy istotności.

Łatwo daje się analitycznie udowodnić pozytywny wpływ rosnącej liczebności próby na moc testu, czyli na jego zdolność do prawidłowego rozstrzygnięcia o prawdziwości lub nieprawdziwości testowanej hipotezy. Zauważmy jednak, że oba błędy (pierwszego i drugiego rodzaju), które bierze się pod uwagę w ocenie jakości testu statystycznego, związane są z samą naturą losowości, a dokładniej z niedoskonałością mechanizmu generującego obserwacje losowe z danego rozkładu (z danej populacji). Nie są w tej ocenie brane pod uwagę jakiegokolwiek inne błędy, tj. błędy o charakterze nielosowym. Mamy więc następującą sytuację: z jednej strony wraz ze wzrostem liczebności próby maleje dyspersja roz-

³ Szerzej na ten temat w pracach — Szreder (2015); Stefanowicz, Cierpień-Wolan (2015).

kładu statystyki testowej, co w warunkach braku zakłóceń spowodowanych czynnikami nielosowymi oznacza lepszą moc testu; z drugiej zaś strony, przy bardzo małej dyspersji rozkładu statystyki z próby (a w konsekwencji małym zakresie nieodrzućenia hipotezy), każde nawet niewielkie uchybienie pomiarowe albo inny błąd systematyczny prowadzić będą do odrzucenia hipotezy zerowej, gdyż wartość statystyki łatwo znajdzie się w obszarze krytycznym o znacznych rozmiarach. **Duże liczebnie próby powodują więc większą wrażliwość testu na błędy o charakterze nielosowym, w tym w szczególności na błędy systematyczne.** Te ostatnie zaś wcale nie są w praktyce rzadkie, wystarczy bowiem niejasno lub niejednoznacznie sformułowane pytanie w kwestionariuszu badania, uchybienie w kodowaniu danych albo błąd w algorytmie przetwarzania informacji uzyskanych z próby.

Problem ten jest jednak głębszy i nie dotyczy wyłącznie błędów nielosowych w testowaniu hipotez. Gdyby abstrahować od tej kategorii błędów, a skoncentrować się tylko na błędzie losowania, to nawet wówczas duża liczebność prób stanowi wyzwanie dla podjęcia prawidłowej decyzji na podstawie testu statystycznego. W szczególności trudny do utrzymania jest w tych okolicznościach typowy poziom istotności 0,1 lub 0,05. Przy bardzo małym rozproszeniu statystyki testowej towarzyszącej dużej liczebności próby, taki poziom istotności będzie determinował na tyle duży obszar krytyczny, że prawie zawsze podjęta będzie decyzja o odrzuceniu hipotezy zerowej. Świadomość konieczności wzięcia pod uwagę wielkości próby przy ustalaniu poziomu istotności istniała wśród statystyków od dawna. Zwracał na nią uwagę m.in. Kmenta⁴, proponując najprostsze, chociaż niedoskonałe rozwiązanie — *zmianę poziomu istotności wraz z wielkością próby, tak aby trudniej było odrzucić hipotezę zerową dla dużych prób niż dla małych (change the level of significance with the sample size, making it harder to reject the null hypothesis for large samples than for small ones)*. Obecnie zaś badacze rozpatrują ten sam problem, tyle że raczej w kategoriach *p-value* aniżeli poziomu istotności — Goldfarb i Lu (2006), Greene (2003), Hubbard i Armstrong (2006), Lin i in. (2013). Autorzy tej ostatniej z wymienionych pozycji uzasadniają, że z wyjątkiem sytuacji, kiedy parametr populacji jest dokładnie równy wartości ujętej w hipotezie zerowej (co w praktyce zdarza się bardzo rzadko), przy rosnącej do nieskończoności wielkości próby *p-value* dążyć będzie do zera (co oznaczać będzie, że prawie zawsze hipoteza zerowa będzie odrzucona). Wynika to wprost z asymptotycznej własności estymatorów wykorzystywanych w statystyce testowej, jaką jest zgodność. W konsekwencji dla bardzo dużych prób klasyczne testy istotności przestają być skutecznym narzędziem rozstrzygającym o prawdziwości lub nieprawdziwości weryfikowanej hipotezy.

Nowych wyzwań badawczych, których wspólnym mianownikiem jest rzeczywistość określana umownie terminem Big Data, jest więcej⁵. W najbliższych

⁴ Kmenta (1990), s. 128.

⁵ Jednym z nich jest malejąca wraz ze wzrostem liczebności próby waga informacji *a priori* we wnioskowaniu, w szczególności we wnioskowaniu bayesowskim.

latach najważniejsze jednak będzie to, czy kwestie wnioskowania na podstawie dużych zbiorów danych pozostawione zostaną specjalistom od sztucznej inteligencji i informatyki czy też statystycy będą mieli w tym ważny udział. Niepokojąca jest obecnie 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 uzyskiwanych danych. Ilustracją tej tendencji może być współczesne podejście do badań sondażowych. Coraz częściej wyniki takiego pojedynczego badania uznaje się za mało wartościowe, niewiarogodne i nieprecyzyjne, jak też postuluje się uśrednianie wyników z kilku różnych badań albo jedynie obserwowanie tendencji ukazanej przez większą liczbę badań na ten sam temat. Jest to jeden z efektów wspomnianego błędnego przekonania, że zwiększona liczba sondaży jest w stanie zrekompensować ich niższą jakość.

Niedawne doświadczenia płynące z sondaży przedwyborczych w Polsce, zwłaszcza przed pierwszą turą wyborów prezydenckich w 2015 r., jak i w Wielkiej Brytanii przed wyborami parlamentarnymi w maju 2015 r. pokazują, że wątpliwe jest istnienie takiej prawidłowości. Powodem tego jest powielanie tych samych błędów, zwłaszcza nielosowych, przez kolejne pracownie sondażowe, co doprowadziło do przeszacowania jednego z kandydatów (w Polsce) i jednej z partii (w Wielkiej Brytanii), a w efekcie żadne uśrednienie nie zniwelowało tych błędów. Jak podkreśla prof. P. Sturgis, przewodniczący brytyjskiej komisji do zbadania źródeł błędnych wyników przewidywań rezultatu wyborczego w 2015 r., nie można wykluczyć, że część ośrodków badawczych nie decyduje się na publikowanie wyników, które zasadniczo odbiegają od ogólnego trendu (*herding*)⁶. Efektem tego było niedoszacowanie popularności Partii Konserwatywnej średnio aż o 4,2 p.proc. Wniosek prof. Sturgisa brzmi, jak się wydaje, bardzo przekonująco — *potrzebne jest przesunięcie punktu ciężkości z ilości na jakość (there needs to be a shift in emphasis away from quantity towards quality)*⁷.

WYZWANIA SFERY EDUKACYJNEJ

Gdyby odrębnie rozpatrywać wyzwania edukacji szkolnej i pozaszkolnej w zakresie statystyki to sądzę, że więcej uwagi należy poświęcić tej drugiej. Edukacją statystyczną w oświacie i szkolnictwie wyższym zajmują się: PTS, Komitet Statystyki i Ekonometrii PAN, GUS, a także inne organizacje i osoby. Programy kształcenia próbują, lepiej lub gorzej, nadążać za zmieniającą się rzeczywistością, wymagającą nowych treści w zakresie statystyki oraz nowych kompetencji absolwentów szkół. Stosunkowo zaniedbana pozostaje natomiast

⁶ *Opinion polls failure at 2015 election due to unrepresentative samples*, „The Telegraph”, 19.01.2016 r.

⁷ Clark, Parraudin (2016).

edukacja pozaszkolna, a więc działania prowadzące do właściwego rozumienia i poprawnej interpretacji wyników badań statystycznych przez społeczeństwo. Przy takim jak w ostatnich dekadach tempie przeobrażeń w otaczającej nas rzeczywistości trudno jest zakładać, że wiedza statystyczna nabyta kilkanaście lub kilkadziesiąt lat temu jest nadal wystarczająca. Potrzebne są długofalowe działania zmierzające do wyposażenia każdej osoby w umiejętność poprawnego odczytywania i krytycznej oceny komunikatów i analiz statystycznych. Chodzi zwłaszcza o te badania statystyczne, z którymi najczęściej mamy do czynienia. Co prawda zmienia się rodzaj i tematyka badań, o których najszerzej informują media, ale są w nich wspólne elementy, na które warto zwrócić uwagę w pierwszej kolejności.

Dziennik „Rzeczpospolita” codziennie powołuje się na różnego rodzaju sondaże w ok. sześciu artykułach. Obok tego na portalach internetowych obecne są dość powszechnie różnego rodzaju sondy dla internautów. Należy się spodziewać, że tego rodzaju liczbowej charakterystyki opinii na rozmaite tematy będzie przybywać. Tymczasem problemy z interpretacją wyników tych badań, zwłaszcza błędów i ich źródeł, mają nie tylko zwykli odbiorcy tych przekazów, ale także politolodzy, politycy i dziennikarze. Z braku umiejętności krytycznej oceny sond internetowych z jednej strony oraz sondaży z drugiej, powstaje wiele nieporozumień, bezproduktywnych dyskusji i ogólnego szumu informacyjnego, w którym najczęściej za winną uznana zostaje statystyka. Naukę naszą obwinia się za błędy, za które faktycznie odpowiadają projektanci i realizatorzy badań oraz ich interpretatorzy. Wielokrotnie zdarza się, że sprzeczne sondaże lub nie trafione prognozy wyborcze stają się przesłanką do krytyki metod statystycznych. Trudno bowiem zwykłemu odbiorcy oddzielić kwestie metodyczne badania sondażowego od kwestii praktycznych, czyli realizacji badania. Niestarannie i pośpiesznie wykonywane sondaże, co widoczne jest zwłaszcza w okresach bezpośrednio przed wyborami, szkodzą samej idei badań próbkowych i szkodzą też statystyce. **Edukacji statystycznej potrzebują, jak się wydaje, zarówno wykonawcy badań sondażowych, jak i ich odbiorcy.**

Metoda reprezentacyjna z kolei, czy też ogólniej badania próbkowe, bronią się same. W szczególności dobrą weryfikację użyteczności i dokładności badań reprezentacyjnych stanowi badanie *exit poll* wykonywane w dniu wyborów. Uda-
ne *exit poll* daje każdorazowo dowód na to, że naukowe podstawy wnioskowania na podstawie prób losowych dobrze się sprawdzają w praktyce. W ostatnich pięciu latach realizatorzy badań *exit poll* w naszym kraju, w szczególności TNS Polska oraz Ipsos, uzyskali w wyborach prezydenckich, parlamentarnych i eurowyborach wyniki różniące się od oficjalnych rezultatów o mniej niż 1,6 p.proc. w odniesieniu do każdej partii lub kandydata. Wyjątkiem były wybory samorządowe z jesieni 2014 r., w których źródłem większego błędu była najprawdopodobniej znaczna frakcja głosów nieważnych. Duża precyzja wyników badań *exit poll* w Polsce i w wielu innych krajach dowodzi trwałej wartości fundamentów teorii wnioskowania statystycznego. Jest przekonującym dowodem na to, że starannie przygotowane i zrealizowane badanie reprezentacyjne na

próbie stanowiącej zwykle mniej niż 1% wielkości populacji może z dużą dokładnością tę populację charakteryzować.

O sondażach należało wspomnieć, gdyż nawet jeżeli nie budzą powszechnego zainteresowania lub są lekceważone, to są chyba najczęściej przywoływanym w mediach badaniem statystycznym. Z kolei rzadziej pojawiają się w publicznym obiegu dane statystyczne o dochodach ludności i płacach, ale budzą one jeszcze żywsze komentarze i dyskusje. Ostatnie miesiące przyniosły w naszym kraju nasilenie się dyskusji związanych z rzeczywistą sytuacją materialną rodzin. Pojawiło się niemało głosów wskazujących na dysonans między obrazem dochodów gospodarstw domowych wynikającym z danych statystyki publicznej a odczuciami dużej części społeczeństwa. S. Sierakowski (2015), w artykule pt. *Sprzedana demokracja?* („Polityka” z 8 grudnia 2015 r.), pisząc o średniej płacy w polskiej gospodarce, której nie osiąga ok. 2/3 pracujących, oskarża statystykę słowami — *Co miesiąc więc statystyka oszukuje trzy czwarte Polaków (tyle mniej więcej zarabia poniżej średniej) i to aż o połowę*⁸. Autor postuluje, aby przeciętne zarobki charakteryzować nie średnią arytmetyczną, lecz medianą lub dominantą. O słabościach obu tych miar pozycyjnych, a także o praktycznych trudnościach z dostępem do danych pozwalających na ich wyliczenie pisałem na łamach „Polityki”⁹, zwracając jednocześnie uwagę, że lepiej jest traktować wszystkie te trzy mierniki syntetyczne (średnią arytmetyczną, medianę i dominantę) jako komplementarne, a nie jako substytucyjne.

Podobnie wiele kontrowersji i emocji budzi pomiar zróżnicowania dochodów ludności w Polsce. Popularny w tych zagadnieniach współczynnik Giniego, mający swoje niewątpliwe zalety, nie jest w wielu przypadkach wystarczającą miarą opisu nierównomierności w rozkładzie dochodów. Interesująco pisze o tym wskaźniku prof. D. Filar w rozdziale zatytułowanym *Odmiany biedy, czyli Gini nie mówi wszystkiego*¹⁰. Autor słusznie wskazuje na niemożność opisu przez dość prosty współczynnik wielu aspektów zróżnicowanej z natury sytuacji materialnej rodzin. Współczynnik Giniego — zdaniem Autora — *pokazuje dochodowe zróżnicowanie wewnątrz konkretnego społeczeństwa, ale nie odpowiada na pytanie, co to praktycznie znaczy znaleźć się w tej jego części, która otrzymuje najmniejszą część dochodów*. Pisząc dalej o zagrożeniu ubóstwem, odsyła czytelników do głębszych badań statystyki publicznej na ten temat — *Obszary polskiej nędzy z dużą starannością zostały przebadane przez GUS (...)*¹¹. Tego typu przekaz albo edukacyjna inicjatywa, mające na celu uświadomienie społeczeństwu potrzeby sięgania do więcej niż jednego prostego wskaźnika przy opisie złożonych zjawisk społecznych, jest z pewnością zadaniem dla całego środowiska statystyków. Dotyczy to — jak obserwujemy — i gorącego tematu

⁸ Sierakowski (2015), s. 21 — należy zwrócić uwagę, że Autor myli się tu pisząc o 3/4 zamiast o 2/3 zarabiających mniej od średniej krajowej.

⁹ Szreder (2016).

¹⁰ Filar (2015), s. 181—194.

¹¹ *Ubóstwo...* (2015), s. 188 — publikację tę przywołuje D. Filar.

poziomu płac, i wielkości oraz zróżnicowania dochodów w gospodarstwach domowych. W rzeczywistości problem pomiaru zróżnicowania dochodów w gospodarstwach domowych jest bardziej złożony, bo nieco inny obraz uzyskuje się z badań budżetów gospodarstw domowych, a inny z analizy zeznań podatkowych. Zwracają na to uwagę m.in. Kośny i Mazurek (2009, 2010), a także Myck i Najsztub (2014), pisząc m.in. o niedoskonałości systemu wag statystycznych używanego przez GUS do zapewnienia reprezentatywności próby gospodarstw domowych¹². Nie da się, jak widzimy, traktować odrębnie kwestii metodycznych (badawczych) oraz kwestii edukacyjnych. Obie te sfery są ze sobą powiązane i wzajemnie się przenikają w codziennym obcowaniu ze statystyką.

Podsumowanie

Nowe oczekiwania i wyzwania wobec statystyki pojawiają się nie tylko na skutek naturalnego rozwoju rzeczywistości, którą statystyka stara się opisywać i analizować, ale także dlatego, że liczbowy opis zjawisk zaczyna dominować w społecznym przekazie obrazu współczesnego świata. Przybywa więc osób starających się dobrze rozumieć komunikaty statystyków i wnioski z ich badań, a równocześnie pragnących samodzielnie, przy pomocy łatwo dostępnych narzędzi informatycznych, analizować zbiory różnych danych liczbowych. Dla środowiska statystyków rodzi to nowe wyzwanie, którego istota sprowadza się do edukacji statystycznej ogółu społeczeństwa, aby było ono przygotowane do krytycznej oceny badań statystycznych, poprawnej interpretacji wniosków z tych badań oraz prawidłowego posługiwania się metodami statystycznymi, których narzędzia informatyczne zawarte są, jak nigdy wcześniej, w niemal każdym urzędzie służącym do codziennej komunikacji. W sferze badawczej natomiast jednym z ważnych wyzwań dla statystyków jest wypracowanie analitycznego podejścia do wnioskowania na podstawie bardzo dużych zbiorów danych. Big Data nie powinny pozostać jedynie w polu zainteresowań specjalistów od sztucznej inteligencji i informatyki.

prof. dr hab. Mirosław Szreder — *Uniwersytet Gdański*

LITERATURA

- Brzeziński M. (2015), *Trzeba patrzeć w PIT-y*, „Gazeta Wyborcza”, 24.07.2015 r.: s. 23.
- Clark T., Parraudin F. (2016), *General election opinion poll failure down to not reaching Tory voters*, „The Guardian”, 19.01.2016 r.
- Filar D. (2015), *Między zieloną wyspą a dryfującą krą. Gospodarka Polski w latach 2007—2015*, Wydawnictwo ARCHE, Sopot.
- Goldfarb A., Lu Q. (2006), *Household-specific regressions using clickstream data*, „Statistical Science”, Vol. 21, No. 2: s. 247—255.

¹² Brzeziński (2015).

- Greene W. (2003), *Econometric Analysis* (5th ed.), Prentice Hall, New York.
- Hubbard R., Armstrong J. (2006), *Why we don't really know what statistical significance means: A major educational failure*, „Journal of Marketing Education”, Vol. 28, No. 2: s. 114—120.
- Kmenta J. (1990), *Elements of Econometrics* (2nd ed.), „Macmillan”, New York.
- Kośny M., Mazurek E. (2009), *Redistribution and Equity of Polish Personal Income Tax: Measurement Using Micro Data from Tax Returns*, „Statistica & Applicazioni”, Vol. 7, No. 2: s. 211—221.
- Kośny M., Mazurek E. (2010), *Influence of Child Tax Credit on Inequity of Personal Income Tax in Poland*, „International Research Journal of Finance and Economics”, No. 50: s. 45—50.
- Lin M., Lucas Jr. H. C., Shmueli G. (2013), *Too Big to Fail: Large Samples and the p-Value Problem*, „Information Systems Research”, Vol. 24, No. 4, December 2013: s. 906—917.
- Mayer-Schönberger V., Cukier K. (2014), *BIG DATA. Rewolucja, która zmieni nasze myślenie, pracę i życie*, Wydawnictwo MT Biznes, Warszawa.
- Myck M., Najsztub M. (2014), *Data and Model Cross-Validation to Improve Accuracy of Microsimulation Results: Estimates for the Polish Household Budget Survey*, „Discussion Papers”, No. 1368, DIW Berlin, German Institute for Economic Research.
- Sierakowski S. (2015), *Sprzedana demokracja?*, „Polityka”, nr 50 (3039): s. 20 i 21.
- Stefanowicz B., Cierpiął-Wolan M. (2015), *Błędy przetwarzania danych*, „Wiadomości Statystyczne”, nr 9: s. 23—29.
- Szreder M. (2015), *Zmiany w strukturze całkowitego błędu badania próbkowego*, „Wiadomości Statystyczne”, nr 1: s. 4—12.
- Szreder M. (2016), *Pensja w portfelu i w Excelu*, „Polityka”, nr 9 (3048): s. 66 i 67.
- Ubóstwo ekonomiczne w Polsce w 2014 r. (na podstawie badania budżetów gospodarstw domowych)* (2015), opracowanie sygnałne, GUS, 9.06.2015 r.

Summary. *The article discusses recent research and educational challenges faced by the statistics in the world of fast-growing possibilities of collecting and processing data (Big Data) and the increasingly dominant in social and economic life numerical description of reality. These new phenomena and trends should encourage statisticians' environment to take action of improving society statistical knowledge needed to critically evaluate and correct interpretation of the survey results.*

Keywords: Big Data, testing statistical hypotheses, a sample size, opinion poll.

Резюме. *В статье обсуждаются последние исследовательские и дидактические вызовы, которые стоят перед статистикой в ситуации быстро повышающихся возможностей сборки и обработки данных (Big Data), а также более доминирующего в социальной и экономической жизни численного описания реальности. Эти новые явления и тенденции должны поощрять статистиков принимать меры по повышению статистических знаний общества, необходимых для критической оценки и правильной интерпретации результатов статистических обследований.*

Ключевые слова: Big Data, проверка статистической гипотезы, размер выборки, зондирование.