

**Piotr Gawron\***  
**Łukasz Paweł\*\***  
**Zbigniew Puchała\*\*\***  
**Jacek Szklarski\*\*\*\***  
**Karol Życzkowski\*\*\*\*\***

## **WYBORY SAMORZĄDOWE 2014 W POSZUKIWANIU ANOMALII STATYSTYCZNYCH**

### **WSTĘP**

W roku 2014 wybory samorządowe w Polsce odbyły się w dniu 16 listopada. W tym dniu wybierano jednocześnie przedstawicieli do sejmików wojewódzkich, członków rad miast, dzielnic i gmin, a także wójtów, burmistrzów i prezydentów miast. Każdy wyborca otrzymywał w komisji kilka kart do głosowania. Niektóre z nich przygotowano na jednej karcie papieru, a w wyborach do sejmików, ze względu na dużą liczbę kandydatów, karty miały postać broszurki, na każdej stronie znajdowała się lista kandydatów danego komitetu wyborczego. Do objęcia urzędu wójta lub prezydenta miasta konieczna była bezwzględna większość głosów. Drugą turę głosowania przeprowadzono w dniu 30 listopada 2014 roku.

---

\* Dr hab. inż., Instytut Informatyki Teoretycznej i Stosowanej, Polska Akademia Nauk, Gliwice, gawron@iitis.pl.

\*\* Mgr inż., Instytut Informatyki Teoretycznej i Stosowanej, Polska Akademia Nauk, Gliwice, pawela@iitis.pl.

\*\*\* Dr hab., Instytut Informatyki Teoretycznej i Stosowanej, Polska Akademia Nauk, Gliwice, Instytut Fizyki, Uniwersytet Jagielloński, z.puchala@iitis.pl.

\*\*\*\* Dr, Instytut Podstawowych Problemów Techniki, Polska Akademia Nauk, Warszawa, jszklar@ippt.pan.pl.

\*\*\*\*\* Prof. dr hab., Instytut Fizyki, Uniwersytet Jagielloński, Centrum Fizyki Teoretycznej, Polska Akademia Nauk, Warszawa, karol@cft.edu.pl.

W niniejszej pracy koncentrujemy się na statystycznej analizie danych dotyczących wyborów do sejmików wojewódzkich. W nocy powyborczej z 16 na 17 listopada pojawiły się problemy z systemem informatycznym stosowanym przez Państwową Komisję Wyborczą, a oficjalne wyniki wyborów ogłoszono ze znacznym opóźnieniem. Spowodowało to poważne zaniepokojenie opinii publicznej i sprzyjało poddawaniu w wątpliwość rzetelności ogłoszonych wyników. Niektóre ugrupowania polityczne lansowały wprost tezę o sfalszowaniu wyborów do sejmików wojewódzkich. Zasadniczym celem tej pracy jest przedstawienie ilościowej analizy wyników wyborczych oraz weryfikacja hipotezy, dotyczącej zaistnienia systematycznych nieregularności wyborczych na dużą skalę. Dane na których pracowaliśmy zostały nam udostępnione przez Krajowe Biuro Wyborcze w dniu 5 marca 2015 roku. Dane zawierają kompletne wyniki wyborów do sejmików wojewódzkich ze wszystkich komisji obwodowych.

Na wstępie warto podkreślić, że matematyczna analiza danych wyborczych nie pozwala na dostarczenie prawnych dowodów popełnienia fałszerstw, lecz wykrycie istotnych nieregularności statystycznych może stanowić wiarygodny argument za tezą o możliwych nieuczciwościach lub nadużyciach<sup>1</sup>. Bez względu na to, czy po danych wyborach formułowane są bezpośrednio zarzuty o nieuczciwość, statystyczną analizę danych wyborczych prowadzi się rutynowo po każdym wyborach w krajach o długiej tradycji demokratycznej. Przykładem takiej analizy mogą być badania danych wyborczych z USA<sup>2</sup>, Niemiec<sup>3</sup> i Szwajcarii<sup>4</sup>. Niezależnie od możliwych podejrzeń o nieuczciwość, statystyczną analizę danych wyborczych prowadzi się, aby lepiej zrozumieć wszystkie aspekty całego procesu wyborczego, a także aby móc wykorzystać otrzymane rezultaty w celu ulepszenia systemu organizacji kolejnych wyborów. Badania statystyczne wyników prowadzone są niekiedy

---

<sup>1</sup> A. S i m p s e r, *Why governments and parties manipulate elections: theory, practice, and implications*. Cambridge 2013; C. B o r g h e s i, J.C. R a y n a l, J.P. B o u c h a u d, *Election turnout statistics in many countries: similarities, differences, and a diffusive field model for decision-making*, „PloS one” 2012, vol. 7, no. 5, s. e36289; B. B e b e r, A. S c a c c o, *What the numbers say: A digit-based test for election fraud*, „Political Analysis” 2012, vol. 20, no. 2, s. 211–234.

<sup>2</sup> R.M. A l v a r e z, T. E. H a l l, S.D. H y d e, *Election fraud: detecting and deterring electoral manipulation*, Washington, DC 2009.

<sup>3</sup> C. B r e u n i g, A. G o e r r e s, *Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany*, „Electoral Studies” 2011, vol. 30, no. 3, s. 534–545.

<sup>4</sup> L. L e e m a n n, D. B o c h s l e r, *A systematic approach to study electoral fraud*, „Electoral Studies” 2014, vol. 35, s. 33–47.

także przez instytucje organizujące wybory, np. w Wielkiej Brytanii<sup>5</sup> oraz w Niemczech<sup>6</sup>.

## STATYSTYCZNA ANALIZA WYNIKÓW

Do analizy wyników wyborów można posłużyć się wynikami z poszczególnych obwodowych komisji wyborczych. W wyborach do sejmików wojewódzkich w roku 2014 głosowanie odbyło się w 27 435 obwodowych komisjach wyborczych. Uprawnionych do głosowania było 30 613 691 osób, komisje wydały 14 466 824 kart do głosowania, oddano 11 928 736 głosów ważnych oraz 2 525 210 głosów nieważnych. Frekwencja wyborcza, liczona jako stosunek liczby wydanych kart do głosowania do liczby wyborców, wynosiła ok. 47,3%. Odsetek głosów nieważnych, liczony jako stosunek liczby głosów nieważnych do głosów wyjętych z urn, wyniósł 17,5%.

Duża liczba głosów nieważnych jest wysoce niepokojąca. Można próbować wytłumaczyć ją różnymi przyczynami. W porównaniu do wyborów do europarlamentu, przeprowadzonych w maju 2014, frekwencja w listopadzie była niemal dwukrotnie większa, a więc do urn wyborczych udali się także wyborcy niezainteresowani polityką europejską. Jest prawdopodobne, że część z nich udała się do lokalów wyborczych w celu wyboru wójta, burmistrza lub prezydenta miasta, a zlekceważyła wybory do sejmików wojewódzkich i oddała w nich głos w sposób przypadkowy. Inna część wyborców chciała oddać głos ważny, lecz na przykład wskutek błędnego postawienia więcej niż jednego krzyżyka ich głos został uznany za nieważny. Trzecia grupa zawiera wyborców, którzy celowo stawiali więcej krzyżyków na karcie i świadomie oddali głos nieważny, na przykład w celu wyrażenia protestu przeciwko klasie politycznej. Do jeszcze innej grupy należą wyborcy, którzy sami nie wiedzieli, na kogo chcą głosować i nie postawili na karcie żadnego krzyżyka, oddając głos zakwalifikowany jako nieważny.

Inną hipotetyczną, przyczyną zaliczenia głosu jako nieważny mogłoby być świadome działanie dowolnego członka komisji wyborczej, który celowo dodał na karcie jeden krzyżyk, aby uczynić głos nieważnym. Z drugiej strony trudno wykluczyć też inny scenariusz: nieważny głos z krzyżykiem na pierwszej i dowolnej innej stronie broszurki wyborczej zostaje omyłkowo uznany przez komisję za ważny głos oddany na listę z numerem 1. Na skutek zmian

---

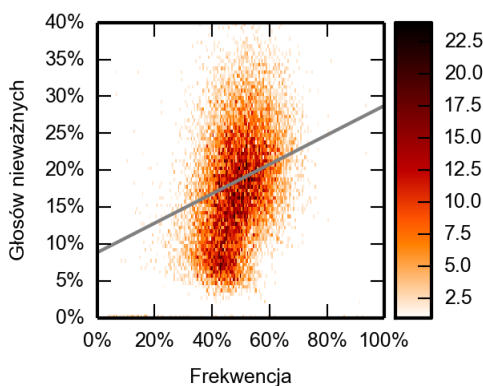
<sup>5</sup> *Electoral fraud in the UK: Final report and recommendations*, January 2014.

<sup>6</sup> *Wahlen zur Bürgerschaft und zu den Bezirksversammlungen am 20. Februar 2011*, Band 1: *Analyse*, 2011.

szczegółowych przepisów wyborczych wprowadzonych w roku 2011, które zniosły obowiązek osobnego zliczania przez komisje głosów z więcej niż jednym krzyżykiem oraz głosów pustych, dokładna analiza przyczyn powstania głosów nieważnych oraz ich interpretacja stała się praktycznie niemożliwa.

## FREKWENCJA WYBORCZA ORAZ ODSETEK GŁOSÓW NIEWAŻNYCH

Ze względu na dużą liczbę głosów nieważnych, interesuje nas związek pomiędzy frekwencją wyborczą a odsetkiem głosów nieważnych w poszczególnych komisjach wyborczych. W tym celu wykonaliśmy test istotności współczynnika korelacji  $r$ -Pearsona oraz test istotności współczynnika korelacji rang Spearmana<sup>7</sup>. Oba testy badają istotność zależności monotonicznej, jednakże pierwszy mierzy zależność liniową i wymaga spełnienia założeń o normalności, które w przypadku naszych danych nie zawsze są spełnione. Biorąc pod uwagę powyższy fakt, przy badaniu istotności korelacji większą wagę przywiązujemy do wyniku testu drugiego.



Szerokość przedziału w histogramie ustalono na 1%, a dla większej czytelności rysunku dobrano różne skale na obu osiach. Na wykresie wykreślono prostą regresji liniowej.

Rys. 1. Rozkład liczby obwodowych komisji wyborczych w funkcji frekwencji oraz odsetka głosów nieważnych

<sup>7</sup> J. Koronacki, J. Mielniczuk, *Statystyka: dla studentów kierunków technicznych i przyrodniczych*, Warszawa 2004.

Na rys. 1 przedstawiono histogram liczby obwodowych komisji wyborczych w funkcji frekwencji oraz procenta głosów nieważnych. Prosta regresji jest zadana przez zależność  $y = 0,199x + 0,088$ , współczynnik  $r$ -Pearsona wynosi 0,255. Współczynnik korelacji  $\rho$ -Spearmana wynosi 0,317, poziom istotności obu testów wynosi  $p < 10^{-3}$ . Zatem istnieje słaba dodatnia istotna statystycznie korelacja pomiędzy frekwencją wyborczą a odsetkiem głosów nieważnych.

Dodatnia korelacja może sugerować, że w komisjach, w których oddano więcej głosów, wyborcy popełniali więcej błędów podczas głosowania. Taka korelacja może też być potwierdzeniem tezy, że jeśli w danej gminie dochodzi do szczególnie ciekawych wyborów wójta (lub – w mieście – burmistrza czy prezydenta) i uczestniczy w nich więcej wyborców, to jednocześnie spada odsetek osób zainteresowanych wyborami do sejmików wojewódzkich, przez co rośnie odsetek głosów nieważnych.

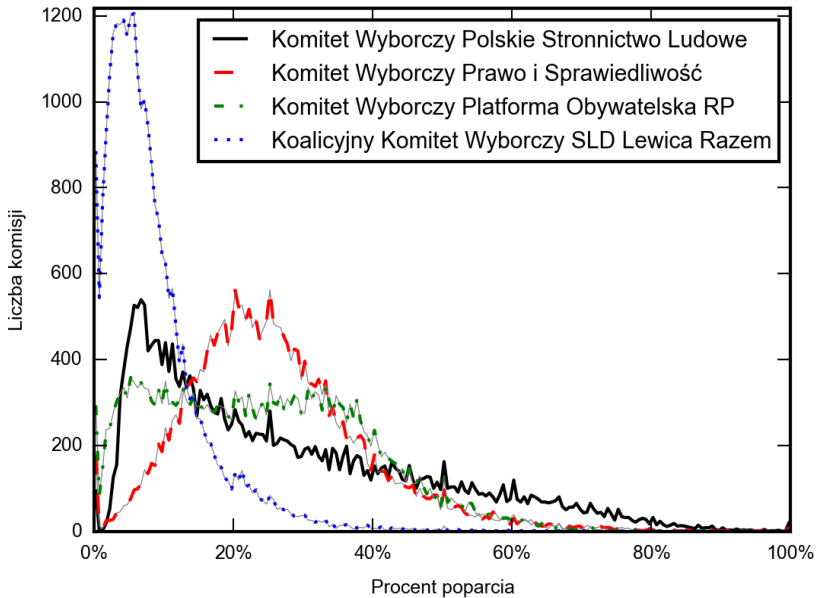
## ROZKŁAD POPARCIA DLA KOMITETÓW WYBORCZYCH

Jedną ze statystycznych metod analizy danych wyborczych jest badanie rozkładu liczby obwodowych komisji wyborczych w funkcji procentowego poparcia danego komitetu wyborczego. Histogram taki przedstawiono na rys. 2. Rozkład nie wykazuje ewidentnych statystycznych aberracji. W szczególności nie obserwujemy charakterystycznych ostrych wierzchołków histogramu, które były przedmiotem kontrowersji w wynikach wyborczych Jednej Rosji w wyborach do Dumy Federacji Rosyjskiej w roku 2011 oraz w wynikach Władimira Putina podczas wyborów prezydenta Rosji w roku 2012. Efekt ten został ujawniony w pracach D. Kobaka, S. Shpilkina i M. Pshenichnikova<sup>8</sup>, a następnie opisany przez K. Rządewskiego, W. Słomczyńskiego, K. Życzkowskiego<sup>9</sup>. Niemniej warto zauważyć, że dla PSL od pewnego momentu, liczba komisji maleje praktycznie liniowo wraz z procentem poparcia – taki efekt nie był dotychczas omawiany w literaturze przedmiotu.

---

<sup>8</sup> D. Kobak, S. Shpilkin, M. Pshenichnikov, „Statistical anomalies in 2011-2012 Russian elections revealed by 2D correlation analysis”, preprint arXiv:1205.0741 2012; D. Kobak, S. Shpilkin, M. Pshenichnikov, *Integer percentages as electoral falsification fingerprints*, preprint arXiv:1410.6059, 2014.

<sup>9</sup> K. Rządewski, W. Słomczyński, K. Życzkowski, *Każdy głos się liczy! Wędrowka przez krainę wyborów*, Warszawa 2014.



Rys. 2. Histogram liczby obwodowych komisji wyborczych rejestrujących dany odsetek poparcia dla wybranego komitetu wyborczego  
Szerokość przedziału w histogramie wynosi 0,5%

## ODSETEK GŁOSÓW NIEWAŻNYCH ORAZ FREKWENCJA WYBORCZA A POPARCIE DLA KOMITETÓW WYBORCZYCH

Interesującą cechą uzyskanych danych wyborczych są korelacje pomiędzy odsetkiem głosów nieważnych (oraz frekwencją wyborczą) a poparciem dla poszczególnych komitetów wyborczych. Korelacja dodatnia między parą obserwowanych zmiennych istnieje wtedy, gdy dużym wartościom jednej zmiennej towarzyszą średnio większe wartości drugiej zmiennej. Podobnie mówimy, że obserwujemy korelację ujemną, gdy dużym wartościom jednej zmiennej towarzyszą średnio mniejsze wartości drugiej zmiennej. W przypadku braku korelacji dane na wykresie rozrzutu przypominają elipsę o osiach równoległych do osi układu współrzędnych, natomiast wyraźne odstępstwo od równoległości osi sugeruje istnienie istotnej statystycznie korelacji. Analizę zależności przeprowadziliśmy dla czterech komitetów wyborczych, które uzyskały największe poparcie, czyli w kolejności numerów komitetów: Komitet Wyborczy Polskiego Stronnictwa Ludowego (PSL), Komitet Wyborczy

Prawa i Sprawiedliwości (PIS), Komitet Wyborczy Platformy Obywatelskiej (PO) oraz Komitet Wyborczy Sojusz Lewicy Demokratycznej Lewica Razem (SLD LR).

Zależności pomiędzy odsetkiem głosów nieważnych, frekwencją a poparciem prezentujemy w formie wykresów rozrzutu na rys. 3 oraz rys. 4. Istnienie niezerowej korelacji jest w większości przypadków widoczne na wykresach, a obserwowane korelacje są statystycznie istotne, co wykazaliśmy, przeprowadzając odpowiednie testy.

Rozważmy najpierw zależność pomiędzy odsetkiem głosów nieważnych a poparciem dla danego komitetu wyborczego. Na podstawie testu korelacji rang Spearmana zależność monotoniczna jest istotna dla wszystkich rozważanych przez nas komitetów wyborczych. Wynika to z wartości obliczonego współczynnika *p-wartość*, który we wszystkich przypadkach jest mniejszy niż  $10^{-12}$ .

Interesującą właściwością danych wyborczych jest znak korelacji. Okazuje się, że w znakomitej większości obserwuje się korelację ujemną, co oznacza, iż duży odsetek głosów nieważnych towarzyszy niskiemu względnemu poparciu. Wyjątkiem od tej reguły jest komitet wyborczy PSL, w którego przypadku obserwujemy korelację dodatnią, co oznacza, że duży odsetek głosów nieważnych towarzyszy wysokiemu poparciu tej partii.

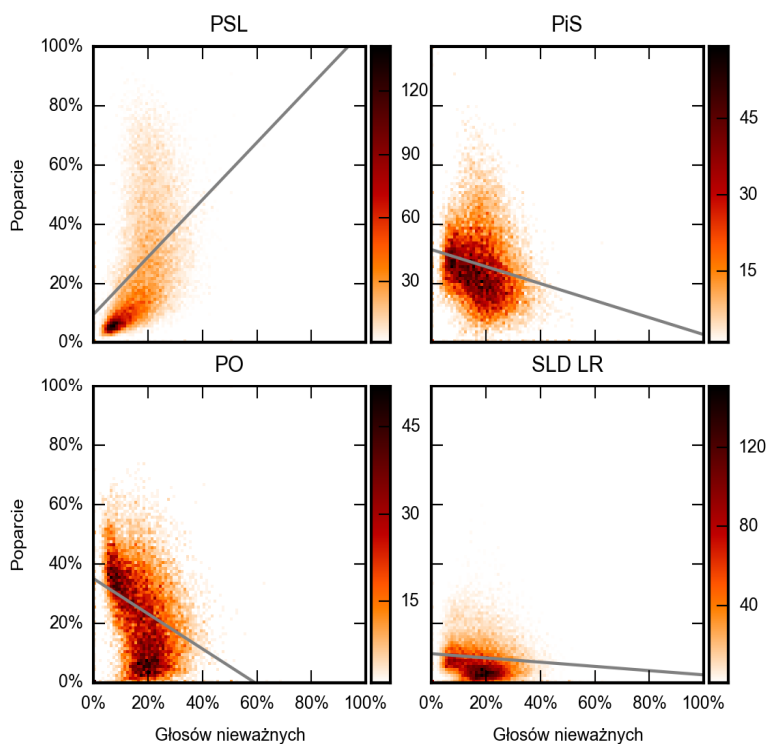
Wart podkreślenia jest fakt, że istnienie korelacji nie implikuje istnienia zależności przyczynowo skutkowej, wskazuje natomiast, iż potencjalnie istnieje czynnik zewnętrzny wpływający na wartości obu obserwowanych wartości. W przypadku zależności pomiędzy odsetkiem głosów nieważnych a poparciem dla danego komitetu wyborczego, czynnikiem tym może być efekt pierwszej strony opisany poniżej lub uwarunkowania społeczne.

Podobną sytuację obserwujemy rozważając zależność pomiędzy frekwencją wyborczą a poparciem dla danego komitetu wyborczego. W przypadku komitetu wyborczego PSL obserwujemy istotną korelację dodatnią pomiędzy frekwencją a poparciem, natomiast w przypadku pozostałych komitetów zależność jest ujemna lub statystycznie nieistotna.

Na rys. 3 pokazano dwuwymiarowe rozkłady liczby obwodowych komisji wyborczych w funkcji głosów nieważnych oraz poparcia dla czterech komitetów wyborczych, które uzyskały największe poparcie w skali kraju. Takie rozkłady mogłyby służyć w celu identyfikacji ewentualnego systematycznego unieważniania głosów przez komisje wyborcze. Na wykresach nie widać, aby istniały komisje, w których masowo unieważniano by głosy w celu zwiększenia względnego poparcia dla któregoś z komitetów.

Dane PSL wykazują dodatnią korelację pomiędzy procentem głosów nieważnych a poparciem dla tego komitetu wyborczego. Prosta regresji jest

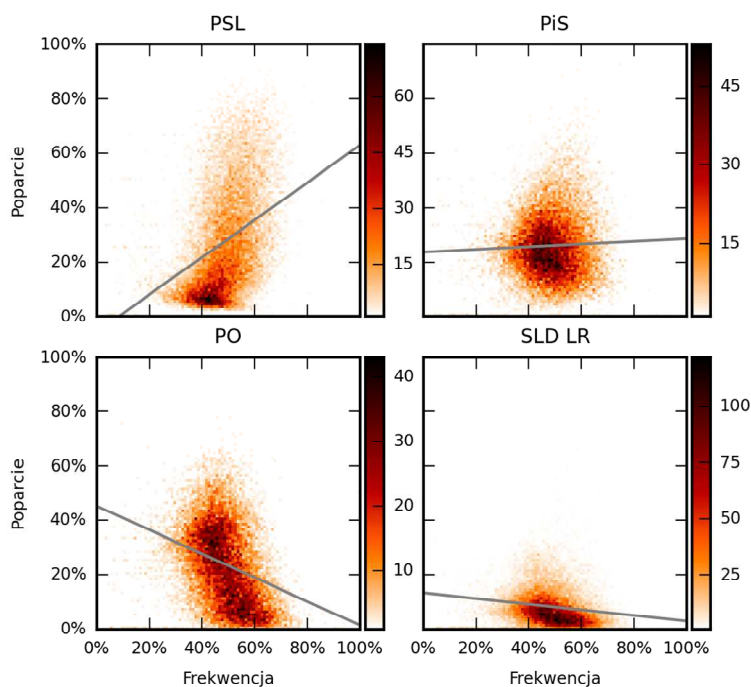
zadana przez zależność  $y_{PSL} = 0,969x + 0,095$ , współczynnik korelacji  $r$ -Pearsona wynosi 0,417, a współczynnik korelacji  $\rho$ -Spearmana wynosi 0,539. W przypadku PiS i SLD LR występują śladowe korelacje ujemne: proste regresji zadane są wzorami odpowiednio  $y_{PiS} = -0,288x + 0,314$  oraz  $y_{SLD} = -0,071x + 0,097$ , a współczynniki korelacji wynoszą  $r_{PiS} = -0,195$ ,  $\rho_{PiS} = -0,223$  oraz  $r_{SLD} = -0,086$ ,  $\rho_{SLD} = -0,149$ . W przypadku PO korelacja jest ujemna. Prosta regresji ma postać  $y_{PO} = -0,592x + 0,35$ , a ujemne współczynniki korelacji wynoszą  $r_{PO} = -0,345$ ,  $\rho_{PO} = -0,381$ . Wszystkie wyliczone korelacje są statystycznie istotne: poziom istotności dla testów Pearsona i Spearmana dla wszystkich czterech korelacji wynosi  $p < 10^{-12}$ .



Rys. 3. Rozkłady liczby obwodowych komisji wyborczych w funkcji odsetka głosów nieważnych oraz poparcia  
Szerokość przedziału na obu osiach ustalono na 1%

Korelacja dodatnia pomiędzy odsetkiem głosów nieważnych a poparciem dla PSL mogłaby wskazywać na, to iż w obwodach, w których naturalnie występuje większe poparcie dla tej partii, wyborcy popełniają więcej błędów

podczas głosowania. Możliwa jest też komplementarna interpretacja, że na terenach wiejskich, gdzie wysoką pozycję utrzymuje PSL, wyborcy chętnie biorą udział w wyborach wójta, lecz mniej interesują się wyborami do sejmików wojewódzkich i częściej oddają w nich głosy nieważne.



Rys. 4. Rozkłady liczby obwodowych komisji wyborczych w funkcji frekwencji oraz poparcia danej listy  
Szerokość przedziału na obu osiach ustalono na 1%

Na rys. 4 pokazano rozkłady liczby obwodowych komisji wyborczych w funkcji frekwencji wyborczej oraz poparcia dla czterech komitetów wyborczych, które uzyskały największą liczbę głosów w skali kraju. Wykresy te można porównać z rysunkami w pracy Kobaka i wsp.<sup>10</sup> oraz zauważyć, że obserwowane rozkłady zgodnie z oczekiwaniami statystyki posiadają pojedynczy wierzchołek. Co więcej, w obserwowanych rozkładach nie występuje efekt zwany „kometą”, czyli obszar większych wartości, który w przypadku danych z Rosji ciągnie się w kierunku większej frekwencji i większego

<sup>10</sup> D. Kobak, S. Shpilkin, M. Pshenichnikov, *Statistical anomalies...*

poparcia dla zwyczajnego komitetu wyborczego. Te obserwacje pozwalają wnioskować, że w analizowanych wyborach nie dochodziło do zjawiska dosypywania głosów przez komisje wyborcze w celu zwiększenia poparcia dla określonego komitetu wyborczego.

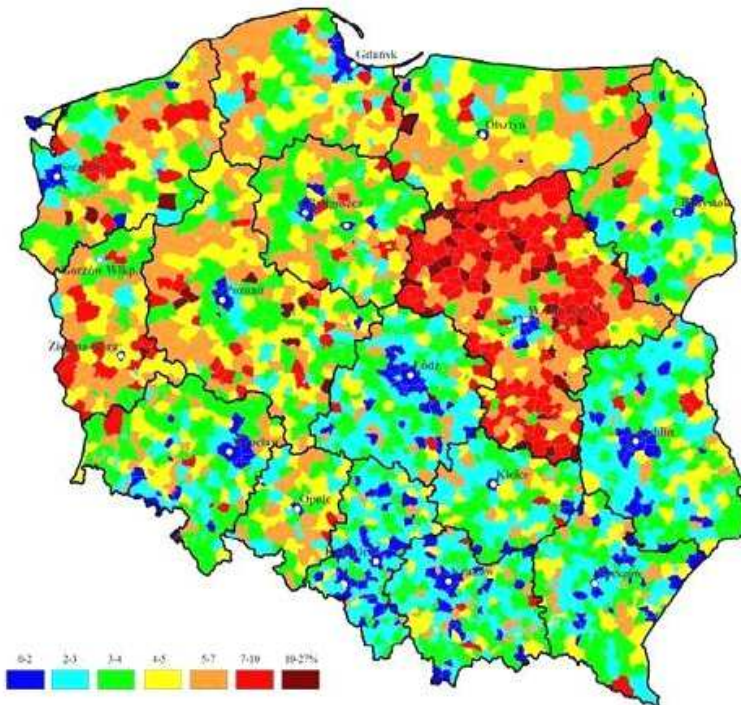
Badanie korelacji pomiędzy frekwencją wyborczą a poparciem dla danego komitetu wyborczego wykazało, że w przypadku PSL istnieje umiarkowana korelacja dodatnia pomiędzy frekwencją a poparciem dla tego komitetu. Prosta regresji jest zadana przez zależność  $y_{PSL} = 0,688x - 0,057$ , współczynnik korelacji  $r$ -Pearsona wynosi  $r_{PSL} = 0,38$ , a współczynnik korelacji  $\rho$ -Spearmana wynosi,  $\rho_{PSL} = 0,439$ . W przypadku PiS zależność korelacyjna jest bardzo słaba i nie jest ona istotna statystycznie. Prosta regresji jest zadana przez zależność  $y_{PiS} = 0,05x + 0,237$ , współczynnik korelacji  $r_{PiS} = 0,044$ . Współczynnik korelacji  $\rho_{PiS} = 0,007$ . W przypadku PO korelacja jest słabo ujemna. Prosta regresji jest zadana przez zależność  $y_{PO} = -0,437x + 0,452$ , współczynnik korelacji  $r_{PO} = -0,328$ , a współczynnik korelacji  $\rho_{PO} = -0,385$ . W przypadku SLD LR korelacja jest bardzo słabo ujemna. Prosta regresji jest zadana przez zależność  $y_{SLD} = -0,105x + 0,134$ , współczynnik korelacji  $r_{SLD} = -0,162$ , a współczynnik korelacji  $\rho_{SLD} = -0,261$ . Poziom istotności dla testów Pearsona i Spearmana dla korelacji wynosi  $p < 10^{-12}$  z wyjątkiem poziomu istotności testu Spearmana dla PiS, który wynosi  $p_{PiS} = 0,262$ . Korelacje te, oprócz przypadku PiS, są statystycznie istotne.

Dodatnia korelacja pomiędzy frekwencją a poparciem dla PSL może sugerować dwa zjawiska. Pierwsze, że wielu głosujących oddawało głos, świadomie bądź nieświadomie, na kandydata na pierwszej stronie karty do głosowania. Drugie – takie, że na wielu kartach do głosowania było postawionych kilka krzyżyków, ale komisja, widząc krzyżyk na pierwszej stronie, nie poszukiwała krzyżyków na stronach pozostałych, a zatem przypisała głos nieważny komitetowi wyborczemu oznaczonemu numerem jeden (PSL).

Na koniec omawiania wspomnianych statystycznych zależności dodajmy dwie uwagi. Po pierwsze, współczynniki korelacji są tak duże, gdyż brane były pod uwagę wszystkie obwody. Sytuacja zmieni się, gdy podzielimy zbiór na dwie grupy: obwody wiejskie i obwody należące do powiatów miejskich. Szczególne wyraźnie widoczne będą różnice na przykładzie PSL, które posiada większe poparcie na wsiach, gdzie odsetek głosów nieważnych jest większy. Niemniej jednak, rozważając jedynie obwody wiejskie, wspomniane korelacje – chociaż mniejsze – mają podobny i statystycznie istotny charakter. Po drugie, tego rodzaju zależności nie są czymś nowym i bardzo podobne wartości uzyskamy dla analogicznych wyborów z lat 2010 i 2006. Należy podkreślić, że w poprzednich wyborach samorządowych liczba głosów nieważnych była znacząco niższa.

## GEOGRAFIA GŁOSÓW NIEWAŻNYCH

Po wyborach do sejmików wojewódzkich w roku 2010 Przemysław Śleszyński opublikował wyniki badań dotyczących rozkładu przestrzennego głosów nieważnych<sup>11</sup>. Najbardziej niepokojące było stwierdzenie zadziwiających różnic pomiędzy poszczególnymi województwami. Na rys. 5 przedstawiającym odsetek głosów nieważnych z powodu postawienia na karcie do głosowania więcej niż jednego znaku, zdecydowanie wyróżnia się województwo mazowieckie (poza Warszawą), w którym średni odsetek takich głosów był istotnie większy niż w innych województwach. Obserwacje Śleszyńskiego



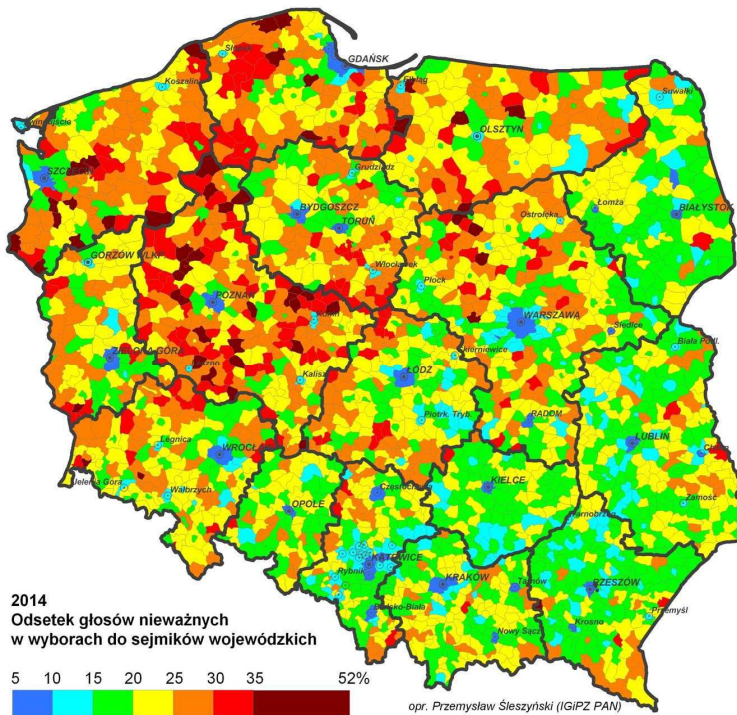
Rys. 5. Rozkład przestrzenny odsetka głosów uznanych za nieważne z powodu postawienia na karcie więcej niż jednego znaku w czasie wyborów do sejmików wojewódzkich w roku 2010

Źródło: P. Śleszyński, *Analiza rozkładu przestrzennego głosów nieważnych w wyborach samorządowych*, Warszawa 2011

<sup>11</sup> P. Śleszyński, *Analiza rozkładu przestrzennego głosów nieważnych w wyborach samorządowych*, Warszawa 2011.

były wielokrotnie cytowane w prasie oraz omówione w książce K. Rządewskiego i wsp.<sup>12</sup>

Jednakże Państwowa Komisja Wyborcza nie zainicjowała przeprowadzenia szczegółowych badań, które mogłyby wyjaśnić efekty obserwowane na Mazowszu jeszcze przed kolejnymi wyborami samorządowymi. Co gorsza, nieprzemysłaną decyzją ustawodawcy z roku 2011 społeczeństwo zostało pozbawione informacji o liczbie głosów, które podczas wyborów 2014 zostały uznane za nieważne z powodu postawienia na karcie więcej niż jednego krzyżyka. Na podstawie oficjalnych danych nie da się obecnie sporządzić mapki dotyczącej wyborów w roku 2014, ściśle odpowiadającej mapce przedstawionej na rys. 5. Dla porównania na rys. 6 przedstawiamy rozkład **wszystkich** głosów nieważnych oddanych podczas wyborów do sejmików w roku 2014.

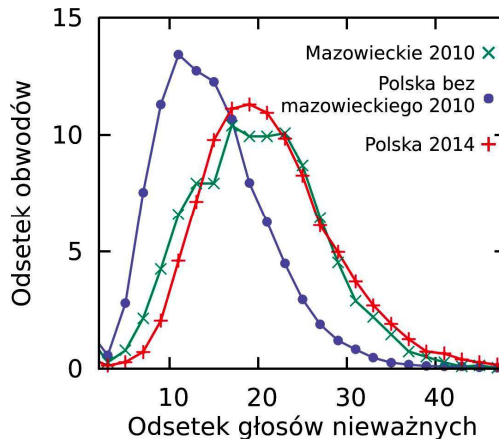


Rys. 6. Rozkład przestrzenny odsetka *wszystkich* głosów nieważnych oddanych podczas wyborów do sejmików wojewódzkich w roku 2014

Źródło: P. Śleszyński, *Analiza rozkładu przestrzennego głosów nieważnych w wyborach samorządowych*, Warszawa 2011

<sup>12</sup> K. Rządewski, W. Słomczyński, K. Życzkowski, *Każdy głos...*

W danych z roku 2014 uwagę zwraca zwiększony odsetek głosów nieważnych, jakie padły w północno-zachodnich województwach kraju. Województwo mazowieckie nie jest tym razem szczególnie wyróżnione, ale też w ostatnich wyborach karta do głosowania miała formę broszury na terenie całego kraju, podczas gdy w roku 2010 była stosowana w województwie mazowieckim.



Rys. 7. Odsetek obwodów wyborczych, w których rejestrowano podany odsetek głosów nieważnych  
Wyniki z roku 2014 dla całego kraju (czerwone znaki +) przypominają wyniki z województwa mazowieckiego z roku 2010 (zielone znaki x). W poprzednich wyborach w województwie mazowieckim karta do głosowania miała postać „broszury wyborczej”

Na rys. 7 przedstawiono porównanie rozkładu wszystkich głosów nieważnych, jakie padły w poszczególnych obwodach wyborczych w roku 2014, z danymi z roku 2010, z których osobno potraktowano dane z województwa mazowieckiego. Zauważmy, że rozkład prawdopodobieństwa opisujący dane z roku 2014 jest istotnie przesunięty w prawo w porównaniu do danych z roku 2010, gdyż wzrósł całkowity odsetek głosów nieważnych. Podobieństwo rozkładu z roku 2014 oraz rozkładu przedstawiającego dane z województwa mazowieckiego z roku 2010 dostarcza argumentów za tezą, że wprowadzenie broszury wyborczej zwiększyło odsetek głosów nieważnych na terenie całego kraju, a więc istotnie wpłynęło na przebieg wyborów.

## PRAWO BENFORDA

Prawo Benforda opisuje rozkład częstości występowania cyfr na poszczególnych miejscach w liczbie w danych empirycznych. Przykładowo, w tym rozkładzie cyfra 1 występuje około 30% na pierwszym miejscu w liczbie. Dokładniej prawdopodobieństwo wystąpienia cyfry  $d$  na pierwszym miejscu dane jest wzorem:

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right).$$

Prawdopodobieństwo, że cyfra  $d$  znajduje się na miejscu  $n$ , dane jest wzorem:

$$P(d; n) = \sum_{k=10^{n-2}}^{10^n - 1} \log_{10}\left(1 + \frac{1}{10k + d}\right).$$

Prawo Benforda jest spełnione przez dane służące opisaniu wielu rodzajów zjawisk empirycznych, jak na przykład kwoty na rachunkach za energię, numery domów, ceny akcji, populacje krajów, długości rzek. Prawo Benforda jest też wykorzystywane w analizie wyników wyborów<sup>13</sup>. Wśród najbardziej znanych przykładów wykorzystania tego prawa w badaniu wyników wyborów prezydenckich są wybory w Iranie, w 2009 roku, gdzie przy jego użyciu znaleziono nieścisłości w wynikach<sup>14</sup>.

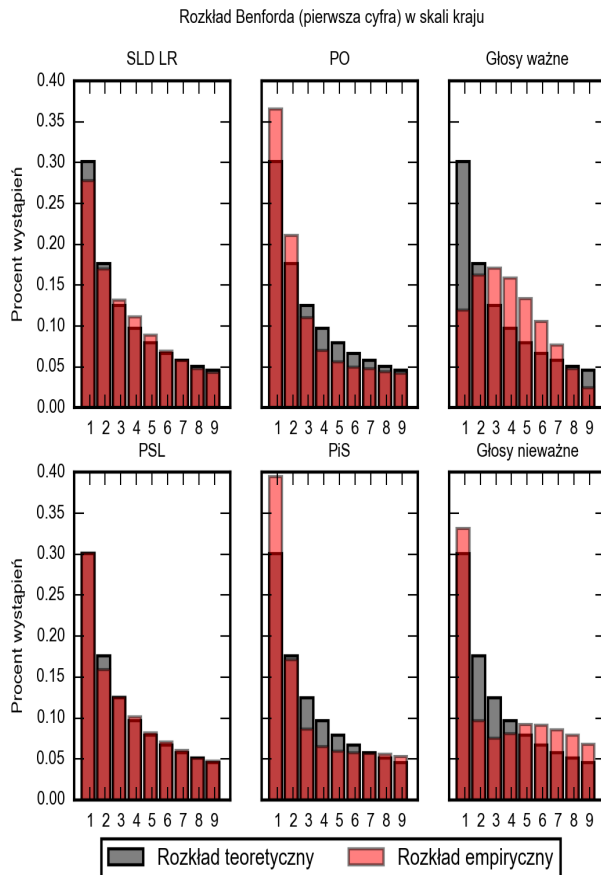
Na rys. 8 zostało zilustrowane porównanie empirycznej oraz teoretycznej częstości występowania poszczególnych cyfr na pierwszym miejscu liczby osób głosujących na daną partię, jak również liczby głosów ważnych i nieważnych, zaś na rys. 9 przedstawiono częstości występowania cyfr na drugim miejscu dla tych samych danych.

Dla wszystkich danych wykonaliśmy test  $\chi^2$  zgodności rozkładów. Przyjęliśmy poziom istotności testu  $\alpha = 0.01$ . Dla takiej wartości  $\alpha$  druga cyfra w liczby głosów oddanych na PO, PSL oraz SLD LR jest zgodna z rozkładem Benforda. Reszta danych, w szczególności dotyczących liczby głosów ważnych

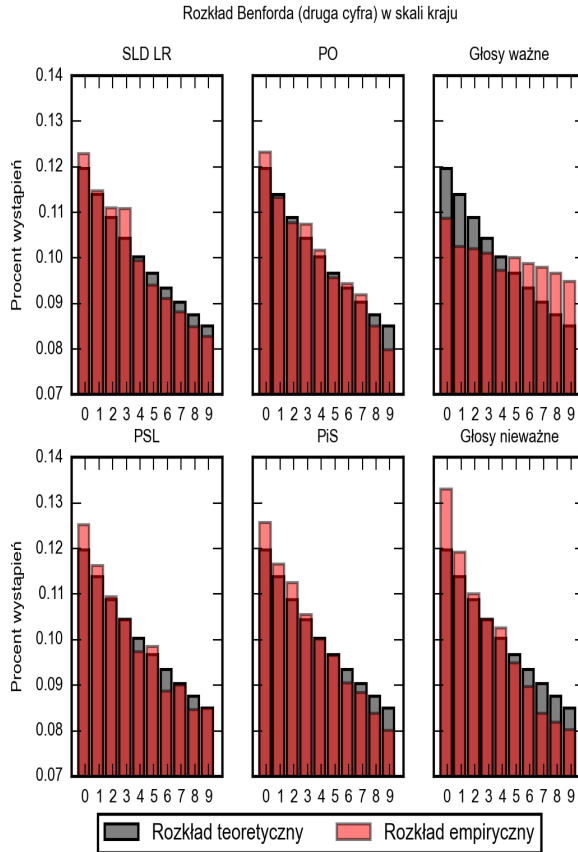
<sup>13</sup> C. Breunig, A. Goerres, *Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany*, „Electoral Studies” 2011, vol. 30, no. 3, s. 534–545; W.R. Mebane Jr, „Election forensics: Vote counts and Benford's law”, [w:] *Summer Meeting of the Political Methodology Society*, Davis, CA 2006, s. 20–22. J. Deckert, M. Myagkov, P.C. Ordeshook, *Benford's Law and the detection of election fraud*, „Political Analysis” 2011, vol. 19, no. 3, s. 245–268; W.R. Mebane, *Comment on „Benford's Law and the detection of election fraud”*, „Political Analysis” 2011, vol. 19, no. 3, s. 269–272.

<sup>14</sup> B.F. Roukema, *A first-digit anomaly in the 2009 Iranian presidential election*, „Journal of Applied Statistics” 2014, vol. 41, no. 1, s. 164–199.

i nieważnych, nie pochodzi z rozkładu Benforda. Odstępstwa pierwszej cyfry od rozkładu Benforda można tłumaczyć liczbą uprawnionych do głosowania w poszczególnych komisjach obwodowych. Liczba osób uprawnionych do głosowania w większości komisji obwodowych mieści się pomiędzy 1000 a 2000 osób. W podobny sposób można wytłumaczyć odstępstwa od prawa Benforda dla drugiej cyfry liczby głosów ważnych i nieważnych. Na podstawie tych faktów nie możemy wnioskować, iż przy liczeniu głosów miały miejsca istotne nieprawidłowości na dużą skalę. Gdyby takie nieprawidłowości miały miejsce, można by oczekiwać znacznych odstępstw od rozkładu Benforda drugiej cyfry w liczbie głosów na poszczególne komitety we wszystkich przypadkach.



Rys. 8. Rozkłady pierwszych cyfr w poparciu partii, liczby głosów ważnych oraz nieważnych



Rys. 9. Rozkłady drugich cyfr w poparciu partii, liczby głosów ważnych oraz nieważnych

## EFEKT PIERWSZEJ STRONY

Hipoteza zakładająca, że kolejność kandydatów na karcie do głosowania może mieć znaczący wpływ na jego wynik rozważana jest od dawna. Zwykle badania koncentrują się na statystyce głosów oddanych na nazwiska wymienione na pojedynczym arkuszu. Przykładem może być tutaj praca: A. King, A. Leigh<sup>15</sup>, w której autorzy dokonali analizy 1187 australijskich plebiscytów

<sup>15</sup> A. King, A. Leigh, *Are Ballot Order Effects Heterogeneous?*, „Social Science Quarterly” 2009, vol. 90, no. 1, s. 71–87.

wyborczych z lat 1984–2004. Jednym z wniosków płynących z tej pracy jest to, że kandydaci wymienieni jako pierwsi mogą liczyć na premię w postaci – średnio rzecz ujmując – jednego punktu procentowego, bez względu na typ wyborów. Czynnikiem mającym znaczenie, ale niewielkie, jest wiek i wykształcenie wyborców. Tego typu obserwacje można mnożyć: J.M. Miller i J.A. Krosnick<sup>16</sup> pokazali, że w wyborach w stanie Ohio w USA kandydaci, których nazwiska znajdowały się na pierwszym miejscu, mieli dodatkowe 2–3 punkty procentowe (p.p.) w stosunku do tych, którym przypadło miejsce ostatnie; w wyborach w Kalifornii różnice te mieściły się w przedziale 2–4 punktów procentowych<sup>17</sup>. Wnioski o wpływie miejsca na liście można wyciągnąć na przykład porównując wybory, w których kandydaci wymienieni byli w kolejności alfabetycznej, z analogicznymi plebiscytami, gdzie nazwiska zostały posortowane w sposób losowy.

W przypadku ostatnich wyborów do sejmików wojewódzkich mieliśmy do czynienia z książeczką, a na jej poszczególnych stronach znajdowały się nazwiska kandydatów należących tylko do jednego komitetu wyborczego. Głos uznawany był za ważny, gdy tylko na jednej ze stron, przy jednym kandydacie postawiony został krzyżyk. W związku z nadspodziewanie dobrym wynikiem komitetu wyborczego, któremu przypadła pierwsza strona, w przestrzeni publicznej pojawiły się opinie, iż to właśnie fakt wylosowania numeru 1 znacząco przyczynił się do takiego rezultatu (w tym przypadku w całej Polsce było to PSL). Wprawdzie istnieje fizyczna różnica między kolejnymi kartkami broszury, a jedną kartą z listą nazwisk (co miało miejsce w przytoczonych wyżej badaniach), ale racjonalne wydaje się być założenie, że efekt pierwszej strony broszury wyborczej istnieje.

Być może jest nawet wyraźniejszy, gdyż z psychologicznego punktu widzenia różnica między partią na pierwszej stronie i na pozostałych może być większa niż między pierwszym i kolejnymi nazwiskami na tej samej karcie. Niestety organy nadzorujące wybory w Polsce nie prowadzą ewidencji informacji o wyglądzie kart do głosowania w różnych województwach i wyborach. W związku z tym dokładna analiza statystyczna jest w tym momencie w zasadzie niemożliwa.

Racjonalnie jest przypuszczać, że w ogólności efekt pierwszej strony jest zjawiskiem złożonym. Przykładowo można rozważać jakąś formę sugestii (np. znany z psychologii mechanizm zwany torowaniem), wpływającą na

---

<sup>16</sup> J.M. Miller, J.A. Krosnick, *The impact of candidate name order on election outcomes*, „Public Opinion Quarterly” 1998, s. 291–330.

<sup>17</sup> D.E. Ho, K. Imai, *Estimating causal effects of ballot order from a randomized natural experiment the California alphabet lottery, 1978–2002*, „Public Opinion Quarterly” 2008, vol. 72, no. 2, s. 216–240.

zmianę sympatii wyborcy, który dopiero w procesie głosowania decyduje się oddać swój głos na partię widniejącą na pierwszej stronie, gdyż uzna ją z jakichś powodów za bardziej atrakcyjną. Ocena skali tego rodzaju zjawisk w kontekście wyborów w Polsce wymagałaby przeprowadzenia konkretnych eksperymentów socjologicznych i leży poza zakresem niniejszej pracy.

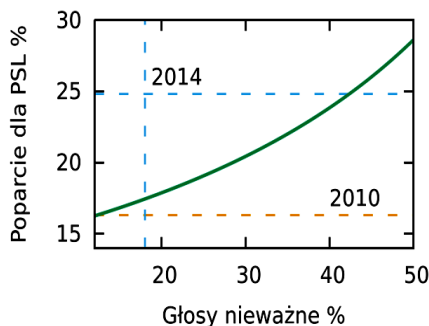
Niemniej możemy rozważyć prosty model wyborcy, który na skutek zastosowania książeczki nieświadomie popełnia błąd, który faworyzuje partię wymienioną na pierwszej stronie. Przyjmijmy tutaj za punkt wyjścia wynik wyborów do sejmików wojewódzkich z roku 2010 dla całego kraju. Zastanówmy się jak te wyniki mogłyby się zmienić, gdyby wówczas w całym kraju zastosowano książeczkę i gdyby wyborcy – z prawdopodobieństwem będącym parametrem modelu – popełniali błędy, których skutkiem jest wynik wyborczy komitetu wyborczego z pierwszej strony większy niż jego rzeczywiste społeczne poparcie. Rozważmy jedynie dwa rodzaje nieświadomie popełnionych błędów, które mogą mieć miejsce w momencie, gdy wyborca oddaje swój głos: (A) i (B).

**(A)** Wyborca partii, która wymieniona jest na stronie innej niż pierwsza, z jakimś prawdopodobieństwem omyłkowo stawia krzyżyk na liście nr 1, po czym stawia krzyżyk na stronie kolejnej, czyniąc swój głos nieważnym.

Wprawdzie w takiej sytuacji bezwzględna liczba głosów na listę nr 1 zwiększy się, jednak procentowe poparcie już tak, gdyż to tylko wyborcy pozostałych partii oddają głosy nieważne, zmniejszając całkowitą liczbę głosów ważnych. Zauważmy, że przy takim założeniu nie ma żadnego powodu, żeby wyborcy innych partii w jakiś szczególny sposób się od siebie różnili. Oznacza to, że prawdopodobieństwo popełnienia błędu tego rodzaju jest niezależne od preferencji wyborcy i głosy z partii wypisanych na kolejnych stronach są transferowane do głosów nieważnych w sposób równomierny – żadna partia, poza tą z pierwszej strony, nie jest tutaj wyróżniona. To oszacowanie nie uwzględnia możliwości, że głos z dwoma krzyżykami został uznany za ważny głos na listę nr 1 w przypadku gdy drugi krzyżyk nie został zauważony przez komisję (takie zjawisko mogłoby mieć znaczenie; formalnie zawiera się w drugim rozważanym scenariuszu błędnego głosu).

Biorąc za punkt wyjścia wyniki wyborów do sejmików z roku 2010, oraz zakładając, że liczba głosów nieważnych w 2014 w stosunku do liczby z roku 2010 zwiększa się tylko dzięki (A), można oszacować górną granicę przyrostu względnego poparcia na skutek zdarzenia (A) dla partii wymienionej na pierwszej stronie książeczki. Oczywiście jest to możliwe dlatego, że liczba głosów nieważnych wygenerowana w ten sposób jest ograniczona. Na rys. 10 przedstawiono, jak zmieniałoby się względne poparcie dla PSL przy zwięks-

szaniu liczby głosów nieważnych wraz ze zwiększaniem prawdopodobieństwa zajścia (A). Z naszych wyliczeń wynika, że przy powyższych założeniach, z dużym prawdopodobieństwem PSL nie zyskałoby więcej niż ok. 1,5 do 2 p.p. w skali całego kraju na skutek tego rodzaju pomyłek. Aby zwiększyć wynik PSL z roku 2010 do poziomu z roku 2014 tylko za pomocą takiego mechanizmu, liczba głosów nieważnych musiałaby przekroczyć 40%.



Rys. 10. Teoretyczny przyrost względnego poparcia dla PSL poprzez zwiększanie liczby głosów nieważnych z powodu postawienia drugiego krzyżyka na liście partii na drugiej lub dalszych stronach broszury wyborczej

Warto tutaj podkreślić, że błędy rodzaju (A) nie wpłyną na wynik testu rozkładu Benforda dla liczby głosów ważnych oddanych na partię. Jest tak dlatego, że rozkłady zmiennych, które spełniają to prawo, są niezmiennicze względem przeskalowania. Innymi słowy, jeśli zbiory liczb reprezentujące liczby głosów ważnych spełniają prawo Benforda, przekształcając wyniki za pomocą omawianego mechanizmu, otrzymamy zbiory również je spełniające (z wyjątkiem liczby głosów nieważnych).

Na koniec dodajmy, że określenie rzeczywistej skali takiego zjawiska wymaga analizy oddanych kart do głosowania. Na skutek wspomianej już wyżej niedawnej modyfikacji przepisów nie zarejestrowano danych dotyczących liczby głosów, które podczas wyborów w roku 2014 zostały uznane za nieważne z powodu postawienia na karcie więcej niż jednego krzyżyka. Brak takiej informacji utrudnia lub wręcz uniemożliwia wyciągnięcie definitywnych wniosków.

**(B)** Drugim rozważanym tutaj efektem ubocznym związanym z pierwszą stroną karty do głosowania w formie książeczki jest oddanie ważnego głosu na listę nr 1, mimo że intencją głosującego było wyrażenie poparcia dla innej partii.

Dokładna przyczyna, czemu zaszło zdarzenie (B), jest w naszych rozważaniach drugorzędna. Mogą tu odegrać rolę czynniki psychologiczne; możliwe, że głoszący omyłkowo postawił krzyżyk na pierwszej stronie, ale jednak postanowił nie unieważniać swojego głosu i wrzucił formalnie poprawnie wypełnioną kartę do urny; być może głoszący wykazał się ignorancją i po prostu zakreślił cokolwiek na pierwszej stronie. Podstawowym założeniem tutaj jest, iż zdarzenie takie zachodzi niezależnie od oryginalnych preferencji wyborcy. Formalnie jest to nierozróżnialne od faktu, że wszyscy wyborcy innych partii w takim samym stopniu zmienili swoje preferencje i zaczęli wybierać partię ze strony pierwszej (co w rzeczywistości jest jednak mało prawdopodobne). Tak więc i tutaj należy mówić o górnej granicy, czyli będziemy szacować maksymalną liczbę głosów, którą przy tak postawionych założeniach, mogło w roku 2014 zyskać PSL względem wyniku z roku 2010.

Przeprowadziliśmy symulacje, dopasowując prawdopodobieństwa wystąpienia zdarzeń (A) i (B), tak aby przekształcić wyniki wyborów do sejmików z roku 2010 i jak najbardziej zbliżyć do tych z roku 2014. Należy zaznaczyć, że pod pojęciem „zbliżenia się” rozumiemy nie tylko np. dopasowanie całkowitej liczby głosów oddanych na partie i liczby głosów nieważnych, ale także zgodność innych cech statystycznych, np. rozkładów wystąpienia obwodów z danym poparciem. Okazuje się, że samym efektem pierwszej strony książeczki nie można uzyskać przejścia z danych z roku 2010 do wyników z 2014, co oznacza, że w tym czasie najpewniej znacząco zmieniły się preferencje wyborcze. Rzecz jasna, nie jest to stwierdzenie odkrywcze; bardzo zaskakujące byłoby, gdyby było inaczej.

Korzystając z takich badań można określić maksymalny zysk, jaki PSL mogło uzyskać w porównaniu z rokiem 2010. Dla obwodów typu wiejskiego górna granica prawdopodobieństw zdarzenia (A) wynosi 10%, a (B) 2,5%. Dzięki *efektowi* (A) PSL zyskał nie więcej niż 2,4 punktu procentowego, a dzięki (A) oraz (B) łącznie nie więcej niż 4,4 p.p., czyli w sumie nie więcej niż ok. 160 tys. głosów. To wciąż bardzo dużo, jak na premię w postaci pozycji na karcie do głosowania, choć mniej niż szacował ostatnio Jarosław Flis<sup>18</sup> porównując wyniki wyborcze partii startującej z nr 1 w dwóch ostatnich wyborach do sejmików i rad powiatów.

W stosowanej procedurze dopasowania średnia górna granica prawdopodobieństwa zdarzenia (A) w woj. mazowieckim wynosi ok. 5,5%, a poza nim aż 13%. Ta różnica jest w zgodzie z faktem, iż w woj. mazowieckim w roku 2010 zaobserwowano ponadprzeciętną liczbę głosów nieważnych, w szcze-

---

<sup>18</sup> J. Flis – wywiad w portalu [www.onet.pl](http://www.onet.pl) oraz analiza w blogu, 21.01. 2015, <http://jaroslawflis.salon24.pl/627326,efekt-ksiazeczki-pierwsza-strona-demokracji> [odczyt 21.01.2015].

gólności – na co zwrócił uwagę Przemysław Śleszyński<sup>19</sup> – znacząco wyższą liczbę kart nieważnych z powodu postawienia więcej niż jednego krzyżyka. Konsekwentnie, przyrost głosów nieważnych na skutek (A) musi być mniejszy w tym województwie. Zasadne jest też przypuszczenie, że tak istotny przyrost głosów nieważnych w porównaniu do poprzednich wyborów związany jest z postawieniem więcej niż jednego krzyżyka na karcie do głosowania.

## UWAGI KOŃCOWE

Praca zawiera wstępne wyniki statystycznej analizy kompletnych danych wyborczych dotyczących wyborów samorządowych z roku 2014, udostępnionych w marcu 2015 przez Państwową Komisję Wyborczą. Wykorzystując techniki zaproponowane ostatnio do analizy różnych danych wyborczych na świecie<sup>20</sup>, poszukiwaliśmy istotnych nieregularności statystycznych, które mogłyby świadczyć o nadużyciach wyborczych na dużą skalę.

Przeprowadzona analiza nie pozwoliła jednak stwierdzić wyraźnych oznak patologii: rozkłady statystyczne poparcia danej partii w obwodach wyborczych nie wykazują charakterystycznych rezonansów (ostrych wierzchołków na wykresie), obserwowanych w danych z wyborów do Dumy Federacji Rosyjskiej w roku 2011<sup>21</sup>. Zgodnie z przewidywaniami statystycznymi<sup>22</sup> rozkłady prawdopodobieństwa poparcia każdej partii w zależności od frekwencji wyborczej lub odsetka głosów nieważnych opisywane są rozkładem jednowierzchołkowym. Analizowane częstości występowania poszczególnych cyfr w danych wyborczych (wyniki wyborów, frekwencja wyborcza, odsetek głosów nieważnych) nie wykazują ewidentnych odchyień od przewidywań prawa Benforda<sup>23</sup>. Wszystkie przeprowadzone badania nie ujawniają oczywistych

---

<sup>19</sup> P. Śleszyński, *Analiza rozkładu...*

<sup>20</sup> A. King, A. Leigh, *Are Ballot Order...*; B.F. Roukema, *A first-digit anomaly in the 2009 Iranian presidential election*, „Journal of Applied Statistics” 2014, vol. 41, no. 1, s. 164–199; W.R. Mebane Jr, *Election forensics...*; D. Kobak, S. Shpilkin, M. Pshenichnikov, *Statistical anomalies...*; D. Kobak, S. Shpilkin, M. Pshenichnikov, *Integer percentages...*; P. Klimek et. al., *Statistical detection of systematic election irregularities*, „Proceedings of the National Academy of Sciences” 2012, vol. 109, no. 41, s. 16469–16473.

<sup>21</sup> D. Kobak, S. Shpilkin, M. Pshenichnikov, *Statistical anomalies...*

<sup>22</sup> P. Klimek et. al., *Statistical detection...*

<sup>23</sup> F. Benford, *The law of anomalous numbers*, „Proceedings of the American Philosophical Society” 1938, s. 551–572; W.R. Mebane Jr, *lection forensics...*; W.R. Mebane, *Comment on...*; s. 269–272; J. Deckert, M. Myagkov, P.C. Ordeshook, *Benford's Law...*; s. 245–268.

anomalii statystycznych i nie dostarczają żadnych argumentów za formułowaną niekiedy hipotezą o fałszerstwach wyborczych w skali całego kraju.

Z drugiej strony wiele obserwowanych własności danych wyborczych jest zastanawiających i wymaga dalszych badań. Za najistotniejszą niepokojącą cechą omawianych wyborów uznać można bardzo wysoki odsetek głosów nieważnych w całej Polsce, który istotnie wzrósł w porównaniu z analogicznymi wyborami w roku 2010. Dalszej analizy wymagają przykładowo istotne różnice pomiędzy rozkładami poparcia konkretnych partii w obwodach wyborczych, skośny rozkład odsetka głosów nieważnych odnotowanego w poszczególnych obwodach oraz zróżnicowanie geograficzne własności statystyki głosów nieważnych w różnych regionach kraju.

Nasze badania pokazują zadziwiające podobieństwo rozkładu głosów nieważnych rejestrowanego w całym kraju w roku 2014 do danych otrzymanych w województwie mazowieckim podczas wyborów w roku 2010. Ponieważ właśnie w tym województwie użyto wtedy kart do głosowania w formie wielostronicowej broszury, które w ostatnich wyborach zostały zastosowane już na terenie całego kraju, otrzymane rezultaty uwiadcniają wady takiego rozwiązania. Przy równoległym przeprowadzaniu innych wyborów (wójta gminy lub burmistrza miasta, rady gminy lub rady dzielnicy), w których karta do głosowania ma formę pojedynczej kartki, karta w postaci broszury istotnie zwiększa odsetek głosów nieważnych. Ponadto pojawia się „efekt pierwszej strony”, gdyż partia, która wylosowała numer jeden i pierwszą stronę na karcie do głosowania, jest uprzywilejowana w porównaniu do innych partii. Wyniki przeprowadzonych symulacji pozwalają oszacować, że efekt ten mógł zwiększyć wynik wyborczy PSL, kandydaci której figurowali na pierwszej stronie karty do głosowania, o ok. 2 punkty procentowe.

W przeprowadzonych badaniach trudno jest odnieść się bezpośrednio do wyników *exit polls* przeprowadzonych w dniu wyborów i opisać przyczyny rozbieżności obserwowanych w porównaniu do oficjalnych wyników wyborów. Taka analiza byłaby możliwa, gdyby instytucja wykonująca badania ujawniła wszystkie zebrane dane i przedstawiła jasno stosowaną metodologię badawczą. Niepełne dane przedstawione przez wykonawców badań z listopada 2014 roku pozwalają przypuszczać, że niewłaściwie zdiagnozowano i zaklasyfikowano przypadki, w których wyborca oddawał głos nieważny, oraz użyto próbki badawczej, która nie była w pełni statystycznie reprezentatywna. Ponieważ jakość rezultatów badań *exit polls* w istotny sposób wpływa na wiarygodność całego systemu wyborczego, w przyszłości byłoby wskazane, aby badania prowadziło kilka niezależnych instytucji. Ponadto wykonawca badań opłacany z budżetu państwa powinien w umowie mieć narzuc-

ny obowiązek rzetelnego przedstawienia wszystkich szczegółów badawczych, bez możliwości zasłaniania się tajemnicą handlową.

Statystyczna analiza danych z wyborów samorządowych w roku 2014 prowadzi do wniosku, że zostały one niewłaściwie zaplanowane i przeprowadzone. Zauważmy, że pewne symptomy choroby systemu zostały ujawnione w analizie danych wyborczych w województwie mazowieckim<sup>24</sup> w roku 2010, ale nie zostały one uwzględnione i wykorzystane przy planowaniu kolejnych wyborów. Efektu „pierwszej strony” można było uniknąć na wiele sposobów, przykładowo przez stosowanie karty do głosowania mieszczącej się na jednej stronie. Ponieważ duży odsetek głosów nieważnych może być interpretowany jako brak zaufania do całego systemu wyborczego, należy dążyć do jego zmniejszenia. W tym celu warto poprawić edukację wyborców oraz rozważyć możliwość wstrzymania się od głosu przy wyborach do sejmików wojewódzkich.

Biorąc przykład z analogicznych instytucji działających w innych krajach, Państwowa Komisja Wyborcza winna inicjować prowadzenie systematycznych badań statystycznych wyników wszystkich przeprowadzanych wyborów, aby przy ich pomocy lepiej zaplanować kolejne wybory i odzyskać zaufanie wyborców. Należy dążyć do pełnej otwartości wszystkich dokumentów wyborczych, a pełne wyniki wszystkich wyborów, wzory kart do głosowania w każdym województwie oraz protokoły komisji wyborczych na każdym szczeblu powinny być ogólnie dostępne i umieszczane w Biuletynie Informacji Publicznej Państwowej Komisji Wyborczej, co umożliwi przeprowadzenie badań wszystkim zainteresowanym.

## PODZIĘKOWANIA

Autorzy chcieliby podziękować Mirosławowi Lechowi Bogdanowiczowi z Krajowego Biura Wyborczego za wysiłek włożony w udostępnienie danych wyborczych. Jesteśmy głęboko wdzięczni Przemysławowi Śleszyńskiemu za wyrażenie zgody na opublikowanie jego rysunków. Dziękujemy także Jarosławowi Flisowi, Kazimierzowi Rzążewskiemu oraz Wojciechowi Słomczyńskiemu za liczne dyskusje oraz uwagi. Praca nie była finansowana z żadnych projektów badawczych. Autorzy deklarują brak konfliktu interesów.

---

<sup>24</sup> P. Śleszyński, *Analiza rozkładu...*; K. Rzążewski, W. Słomczyński, K. Życzkowski, *Każdy głos...*

## BIBLIOGRAFIA

- Alvarez R.M., Hall T.E., S. D. Hyde, *Election fraud: detecting and deterring electoral manipulation*, Washington, DC 2009.
- Beber B., Scacco A., *What the numbers say: A digit-based test for election fraud*, „Political Analysis” 2012, vol. 20, no. 2.
- Benford F., „*The law of anomalous numbers*”, „Proceedings of the American Philosophical Society” 1938.
- Borghesi C., Raynal J.-C., Bouchaud J.-P., *Election turnout statistics in many countries: similarities, differences, and a diffusive field model for decision-making*, „PloS one” 2012, vol. 7, no. 5.
- Breunig C., Goerres A., *Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany*, „Electoral Studies” 2011, vol. 30, no. 3.
- Deckert J., Myagkov M., Ordeshook P.C., *Benford's Law and the detection of election fraud*, „Political Analysis” 2011, vol. 19, no. 3.
- Electoral fraud in the UK: Final report and recommendations*, January 2014.
- Flis J. – wywiad w portalu [www.onet.pl](http://www.onet.pl) oraz analiza w blogu, 21.01.2015, <http://jaroslawflis.salon24.pl/627326,efekt-ksiazeczki-pierwsza-strona-demokracji>.
- Ho D.E., Imai K., *Estimating causal effects of ballot order from a randomized natural experiment the California alphabet lottery, 1978–2002*, „Public Opinion Quarterly” 2008, vol. 72, no. 2.
- King A., Leigh A., *Are Ballot Order Effects Heterogeneous?*, „Social Science Quarterly” 2009, vol. 90, no. 1.
- Klimek P., Yegorov Y., Hanel R., Thurner S., *Statistical detection of systematic election irregularities*, „Proceedings of the National Academy of Sciences” 2012, vol. 109, no. 41.
- Kobak D., Shpilkin S., Pshenichnikov M., *Integer percentages as electoral falsification fingerprints*, preprint arXiv:1410.6059, 2014.
- Kobak D., Shpilkin S., Pshenichnikov M., *Statistical anomalies in 2011–2012 Russian elections revealed by 2D correlation analysis*, preprint arXiv:1205.0741, 2012.
- Koronacki J., Mielniczuk J., *Statystyka: dla studentów kierunków technicznych i przyrodniczych*, Warszawa 2004.
- Leemann L., Bochsler D., *A systematic approach to study electoral fraud*, „Electoral Studies” 2014, vol. 35.
- Mebane Jr W.R., *Election forensics: The meanings of precinct vote counts' second digits*, [w:] *Summer Meeting of the Political Methodology Society*, Charlottesville 2013.
- Mebane Jr W.R., *Election forensics: Vote counts and Benford's law*, [w:] *Summer Meeting of the Political Methodology Society*, Davis, CA 2006.
- Mebane W.R., *Comment on „Benford's Law and the detection of election fraud”*, „Political Analysis” 2011, vol. 19, no. 3.
- Miller J.M., Krosnick J.A., *The impact of candidate name order on election outcomes*, „Public Opinion Quarterly” 1998.
- Roukema B.F., *A first-digit anomaly in the 2009 Iranian presidential election*, „Journal of Applied Statistics” 2014, vol. 41, no. 1.
- Rzążewski K., Słomczyński W., Życzkowski K., *Każdy głos się liczy! Wędrówka przez krainę wyborów*, Warszawa 2014.
- Simpser A., *Why governments and parties manipulate elections: theory, practice, and implications*, Cambridge 2013.

Śleszyński P., *Analiza rozkładu przestrzennego głosów nieważnych w wyborach samorządowych*, Warszawa 2011.

*Wahlen zur Bürgerschaft und zu den Bezirksversammlungen am 20. Februar 2011 Band 1: Analyse*, 2011.

## Streszczenie

W pracy przeanalizowano dane liczbowe dotyczące wyników wyborów do sejmików wojewódzkich, które odbyły się w Polsce w listopadzie 2014. Badania korelacji pomiędzy frekwencją wyborczą, odsetkiem głosów nieważnych oraz wynikami danego komitetu wyborczego, a także porównanie częstości występowania poszczególnych cyfr w wynikach wyborczych z prawem Benforda nie wykazały oczywistych anomalii statystycznych, które mogłyby dostarczyć argumentów za tezą o masowych nieprawidłowościach lub oszustwach. Z drugiej strony porównanie wyników wyborów z lat 2010 i 2014 oraz analiza modelu matematycznego symulującego zarejestrowane wyniki pozwala potwierdzić „efekt pierwszej strony”: głosowanie za pomocą broszury wyborczej mogło przynieść wymierne korzyści komitetowi wyborczemu, który w wyborach wylosował numer 1.

**Słowa kluczowe:** wybory samorządowe, analiza statystyczna, prawo Benforda, efekt pierwszej strony.

## LOCAL ELECTIONS 2014 IN POLAND: IN SEARCH FOR SYSTEMATIC STATISTICAL IRREGULARITIES (summary)

We perform statistical analysis of results of local elections, which took place in Poland in November 2014. Investigating data concerning voivodeship sejmik (provincial-level assembly) we studied correlations between participation in elections, percentage of invalid votes and results of a given party. Furthermore, the frequency of appearance of a given digit in numbers provided by the electoral commission was analyzed and compared with the law of Benford. Our results did not reveal significant statistical irregularities, which could possibly confirm an alleged electoral fraud. However, our investigations do confirm the “first page” effect: as in these elections several-page brochures were used as ballot papers, the party whose candidates were enlisted in the first page of the ballot had a substantial advantage over the competitors.

**Keywords:** local elections, statistical analysis, Benford law, “first page” effect.

