DOI 10.14746/ssp.2018.1.7

Michał Pierzgalski

Uniwersytet Łódzki

Odkrywanie fałszerstw wyborczych a „prawo" Benforda

Streszczenie: Czy jest możliwa identyfikacja fałszerstw wyborczych za pomocą narzędzi statystycznych? W kilku artykułach opublikowanych w okresie kilku ostatnich lat, m.in. Roukema (2009), Perrichi i Torres (2011) oraz Mebane (2013) użyli metody badawczej, która ich zdaniem może pomagać w wykrywaniu oszustw, które miały miejsce podczas elekcji. W tym artykule, odpowiadam na pytanie, czy tzw. „prawo" Benforda, które zostało wykorzystane np. do weryfikacji nieprawidłowości podczas wyborów prezydenckich w Iranie (2009), jest wiarygodną metodą badawczą. Niestety, przykłady empiryczne wskazują, że użyteczność wskazanej metody jest dyskusyjna.

Słowa kluczowe: „prawo" Benforda, wybory, wykrywanie fałszerstw wyborczych

W kilku artykułach opublikowanych na przestrzeni ostatnich kilku lat, m.in. Roukema (2009), Perrichi i Torres (2011) oraz Mebane (2013) użyli narzędzia badawczego, które ich zdaniem może pomagać w identyfikowaniu fałszerstw podczas elekcji urzędników państwowych, a w tym parlamentarzystów i prezydentów państw. Jest to metoda, która wykorzystuje własności tzw. *prawa Benforda* (czasami zwanego prawem *Newcomba-Benforda*), którego treść omówiona jest w następnej sekcji artykułu.

Roukema (2009), Perrichi i Torres (2011) oraz Mebane (2013) przyjęli dyskusyjne założenie, że rozkład empiryczny głosów oddanych na kandydatów/partie podczas elekcji może być modelowany z użyciem równania rozkładu *Benforda*. Podstawowym celem artykułu jest weryfikacja wskazanego założenia, które wydaje się błędne i powinno być poddane weryfikacji empirycznej z użyciem wielu przypadków elekcji, w których uczciwość trudno wątpić. Należy podkreślić, że używane często w literaturze sformułowanie *prawo Benforda* trzeba by właściwie wziąć w cudzysłów („prawo" *Benforda*), gdyż, jak wspomniano, nie ma przekonujących przesłanek świadczących o tym, że reguła ta działa bez wyjątku. Niemniej, w literaturze używa się określenia *prawo Benforda*

(*Benford Law*), zwykle bez cudzysłowu, dlatego w dalszej części tekstu takie nazewnictwo zostało utrzymane.

Prawo Benforda zastosował Roukema (2009), aby ustalić, czy w wyborach prezydenckich w Iranie w 2009 r. miały miejsce oszustwa wyborcze. W przypadku Polski, omawianą metodę badawczą wykorzystali Gawron, Pawela, Puchała, Szklarski i Życzkowski (2015) do sprawdzenia uczciwości wyborów samorządowych 2014. Wskazane badanie wykazało jednak, że nie ma podstaw, aby twierdzić, że ostatnie wybory samorządowe były poddane manipulacji. Niemniej, jak dotąd nie ustalono, czy użyte narzędzie badawcze (wykorzystujące jeden z rozkładów prawdopodobieństwa, a mianowicie tzw. rozkład Benforda) rzeczywiście pozwala na wykrywanie fałszerstw wyborczych. Wątpliwości, jakie wzbudza wykorzystanie tej metody płyną m.in. z faktu, że **dotychczas nie sformułowano teorii wyjaśniającej, z której wynikałoby, że tzw. prawo Benforda ma zastosowanie w kontekście wyników wyborów**, a jeśli tak to dlaczego tak jest (Gelman, 2009; Brock, 2014).

Celem badania, którego wyniki przedstawiono w tym artykule jest **weryfikacja prawdziwości stwierdzenia, że tzw. prawo Benforda jest skutecznym narzędziem badawczym umożliwiającym identyfikację występowania fałszerstw wyborczych**.

Aby zweryfikować słuszność powyższego stwierdzenia, przedstawiona zostanie analiza przypadków, która wykaże, czy postawioną hipotezę należy przyjąć, czy odrzucić. Tak zwany rozkład Benforda znajduje praktyczne zastosowanie np. w wykrywaniu oszustw finansowych (Geyer, Williamson, 2004; Tam Cho, Gaines, 2007), ale przykłady empiryczne (m.in. te, które przedstawiono w tym tekście) przeczą użyteczności tego narzędzia w wykrywaniu fałszerstw wyborczych.

O czym mówi prawo Benforda?

Aby lepiej zrozumieć prawo Benforda, przypomnijmy kilka podstawowych pojęć rachunku prawdopodobieństwa. Wyobraźmy sobie prosty eksperyment losowy polegający na jednokrotnym rzucie sześciościenną kością do gry. Wynikom tego eksperymentu (1 oczko, 2 oczka...) możemy przyporządkować liczby 1, 2, 3, …, 6. Przypisane wynikom doświadczenia losowego liczby nazywamy wartościami zmiennej losowej, np. zmiennej X = *liczba wyrzuconych oczek*. Każdej wartości zmiennej losowej X można jednoznacznie przypisać prawdopodobieństwo jej pojawienia się podczas eksperymentu. Na podstawie klasycznej definicji prawdopodobieństwa wiadomo, że, dla tego prostego przykładu, prawdopodobieństwo dla dowolnej wartości zmiennej jest jednakowe i wynosi 1/6.

Jeżeli wszystkim wartościom zmiennej losowej przypiszemy odpowiadające im prawdopodobieństwa, to możemy powiedzieć, że określiliśmy funkcję **rozkładu prawdopodobieństwa** zmiennej losowej. Tę funkcję możemy np. przedstawić w postaci równania, graficznie lub za pomocą tabeli. Taki rozkład prawdopodobieństwa, który został określony dla jednokrotnego rzutu sześciościenną kością do gry nazywamy rozkładem jednostajnym skokowym. Jeżeli wykonalibyśmy symulację polegającą na wielokrotnym (np. 10 000 razy) rzucaniu kością, to możemy się spodziewać, że każda z możliwych wartości zmiennej {1, 2, …, 6} pojawi się w proporcji mniej więcej 1/6. Jeżeli tę symulację powtórzymy wielokrotnie i okaże się, że, proporcja, np. występowania liczby 6, istotnie odbiega od ułamka 1/6 w przypadku większości spośród wykonanych symulacji, to można by podejrzewać, że kość do gry jest nieuczciwa.

Istnieje wiele rozkładów prawdopodobieństwa, w tym np. rozkład dwumianowy i oczywiście dobrze znany badaczom zajmującym się społeczeństwem tzw. rozkład normalny. Jednym z rozkładów prawdopodobieństwa jest również **rozkład Benforda**. Jest to, podobnie jak rozkład dwumianowy, tzw. rozkład skokowy. Jeżeli założymy, że jakaś zmienna ma rozkład przynajmniej zbliżony do rozkładu Benforda, to można by oczekiwać, że zaobserwowany rozkład wartości takiej zmiennej, z dostatecznie licznej próby, będzie w przybliżeniu zgodny z rozkładem Benforda. Jeśli empiryczny rozkład wartości zmiennej wyraźnie odbiegałby od tego rozkładu, to trzeba by podważyć hipotezę, że dana zmienna faktycznie ma rozkład zgodny z teoretycznym rozkładem Benforda. I właśnie porównanie rozkładu empirycznego analizowanych zmiennych (w przypadku tego badania zmienną w centrum uwagi jest liczba głosów oddanych w obwodach głosowania) z rozkładem teoretycznym Benforda pozwoli na weryfikację przedstawionej wyżej hipotezy badawczej.

Jakie zmienne mają rozkład zbliżony do rozkładu teoretycznego Benforda? Prawo Benforda, które opiera się na funkcji rozkładu prawdopodobieństwa Benforda, mówi, że rozkład pierwszej cyfry znaczącej tworzącej liczby opisujące np.: długość rzek, populację państw, powierzchnię państw, ceny akcji, numery ulic itp., dany jest następującym równaniem:

$$P(X1 = x) = \log_{10}(1 + 1/x) \tag{1}$$

gdzie: x oznacza pierwszą znaczącą cyfrę (x = 1, 2, …, 9), a $P(X1 = x)$ oznacza prawdopodobieństwo, z jakim cyfra x będzie pierwszą cyfrą znaczącą zaobserwowanej liczby.

Na podstawie równania Benforda można wywnioskować, że:

– cyfra „1” pojawia się w 30,1%,
– cyfra „2” z częstotliwością 17,6%,
– cyfra „3” – 12,5%,
– cyfra „4” – 9,7%,
– cyfra „5” – 7,9%,
– cyfra „6” – 6,7%,
– cyfra „7” – 5,8%,
– cyfra „8” – 5,1%,
– cyfra „9” – 4,6%.

Ogólnie, im większa cyfra, tym mniejsza częstotliwość jej występowania.

Jeżeli badaniu poddamy powierzchnię państw świata w km^2, to okaże się, że rozkład pierwszych cyfr znaczących otrzymanych liczb oznaczających pola powierzchni wszystkich państw będzie przypominał rozkład Benforda. Innymi słowy, jeżeli wybierzemy losowo jakieś państwo, to, jeszcze przed sprawdzeniem jego powierzchni, możemy stwierdzić, że najbardziej prawdopodobne jest, że pierwsza cyfra liczby wyrażającej jego pole będzie równa jeden.

Niektórzy badacze (m.in. Gawron i in., 2015; Perrichi, Torres, 2011; Roukema, 2009) założyli, że do wyników wyborów w poszczególnych obwodach głosowania też można by zastosować prawo Benforda, a więc, dla przykładu, dana partia polityczna powinna w ok. 30% obwodów głosowania uzyskać wynik rozpoczynający się cyfrą 1, np. 15% głosów (0,15 głosów) lub ewentualnie np. 1200 głosów – jeżeli poddamy analizie bezwzględną liczbę głosów. Ci badacze, których zainteresował problem identyfikacji nieprawidłowości podczas elekcji, doszli do wniosku, że jeżeli rozkład pierwszej cyfry znaczącej tworzącej liczbę głosów oddanych na uczestników elekcji istotnie odróżnia się od rozkładu utworzonego zgodnie z prawem Benforda, i np. jakaś cyfra występuje jako pierwsza cyfra znacząca (tworząca liczbę/procent uzyskanych głosów) istotnie częściej/rzadziej, niż wynikałoby to z reguły Benforda, to jest to jeden z dowodów na występowanie fałszerstw wyborczych (Roukema, 2009). Roukema pokazał, że rozkład pierwszych cyfr liczb głosów oddanych na poziomie okręgów w wyborach prezydenckich w Iranie (2009) na kandydata partii opozycyjnej M. Karroubiego nie jest zgodny z rozkładem Benforda. Częściej niż powinna pojawia się liczba „siedem”.

Niestety, brakuje przekonujących argumentów, które tłumaczyłyby dlaczego właściwie rozkład pierwszych (lub też kolejnych) cyfr znaczących miałby być zgodny z rozkładem teoretycznym Benforda.

Zdarza się, że umiemy opisać pewne prawidłowości i potwierdzić ich faktyczne występowanie, ale nie potrafimy dokładnie wyjaśnić dlaczego one zachodzą. Gdyby, jednak, przyjąć, że tzw. prawo Benforda działa, to, na razie, trzeba by je traktować jako jedną z tych zależności, która potwierdza się empirycznie, ale jej mechanizm nie jest dobrze wyjaśniony. Jak zauważają Berger i Hill (2011), większość ekspertów jest zgodna, że wszechobecność prawa Benforda, w szczególności w przypadku danych rzeczywistych (*real-life*), nadal pozostaje tajemnicą („most experts seem to agree with [Fewster (2009)] that the ubiquity of BL, especially in real-life data, remains mysterious”). Jednakże, analiza przedstawiona niżej dowodzi, że **skuteczność stosowania prawa Benforda do odkrywania fałszerstw wyborczych budzi istotne wątpliwości**.

Rozkład Benforda można uogólnić i odnieść również do kolejnych cyfr znaczących. Mebane (2013) wykorzystał do poszukiwania nieprawidłowości podczas elekcji w Rosji uogólnioną wersję rozkładu Benforda dla przypadku drugiej cyfry znaczącej (zamiast cyfry pierwszej). Autor jednak przyznaje, że rezultaty zastosowania prawa Benforda w badaniu nieprawidłowości podczas wyborów są niepewne. Jak zauważa Mebane[1] (2013), użycie prawa Benforda do identyfikacji fałszerstw wyborczych generuje zaskakujące i jednocześnie nieprawdopodobne wyniki. Na przykład, badanie dotyczące wyborów w Rosji sugeruje, że wyniki Putina w wyborach 2004 i 2012, a także wynik Zjednoczonej Rosji (*Jednaja Rossija*) w 2011 roku, nie były sfałszowane, ale wynik Miedwiediewa w 2008 roku, był poddany manipula-

[1] „The digit tests produce surprising and on balance implausible results. For example, they suggest that none of the votes for Putin in 2004 and 2012 or for United Russia in 2011 were fraudulent, while votes for Medvedev in 2008 were fraudulent. The usefulness of simple and direct application of either kind of digit tests for fraud detection seems questionable, although in connection with more nuanced interpretations they may be useful” (Mebane, 2013, Abstract).

cji (Mebane, 2011, Abstract). Mebane (2011) zauważa, że, jak dotąd, pytanie o wiarygodność metody wykrywania fałszerstw wyborczych, opierającej się na własnościach rozkładu Benforda, pozostaje bez odpowiedzi[2].

Rozkład Benforda w wersji dla drugiej cyfry znaczącej (w skrócie **2BL**) jest dany następującym równaniem:

$$P(X2 = x) = \sum_{0 \le k \le 9} \log_{10}(1 + \frac{1}{10k + x}) \tag{2}$$

Podstawiając za x do równania (2) kolejne cyfry, uzyskujemy rozkład częstości względnych dla drugich cyfr:
- cyfra „0" pojawia się z częstotliwością 0,1197 (dla drugiej cyfry znaczącej bierzemy pod uwagę także „zero"),
- cyfra „1" – 0,1139,
- cyfra „2" – 0,1088,
- cyfra „3" – 0,1043,
- cyfra „4" – 0,1003,
- cyfra „5" – 0,0967,
- cyfra „6" – 0,0934,
- cyfra „7" – 0,0904,
- cyfra „8" – 0,0876,
- cyfra „9" – 0,0850.

* * *

Rozkład teoretyczny Benforda dla pierwszej cyfry znaczącej (w skrócie **1BL**) przedstawiony został graficznie na rysunku 1 (Benford). Na tym samym rysunku przedstawiono, dla porównania, empiryczny rozkład populacji stanów (USA) wg spisu powszechnego z 2000 r. (PopUSA). Jest to przykład zmiennej, którą cechuje bardzo dobra zbieżność z rozkładem teoretycznym Benforda. Można powiedzieć, że ta zmienna podlega prawu Benforda, co potwierdza też wynik testu zgodności *chi*-kwadrat (por. tabela 1). Wartość prawdopodobieństwa testowego (p-value) jest wystarczająco wysoka, aby stwierdzić, że nie można odrzucić hipotezy zerowej

[2] „Whether the tests are useful for detecting fraud remains an open question, but approaching this question requires an approach more nuanced and tied to careful analysis of real election data […]" (Mebane, 2011, Abstract).

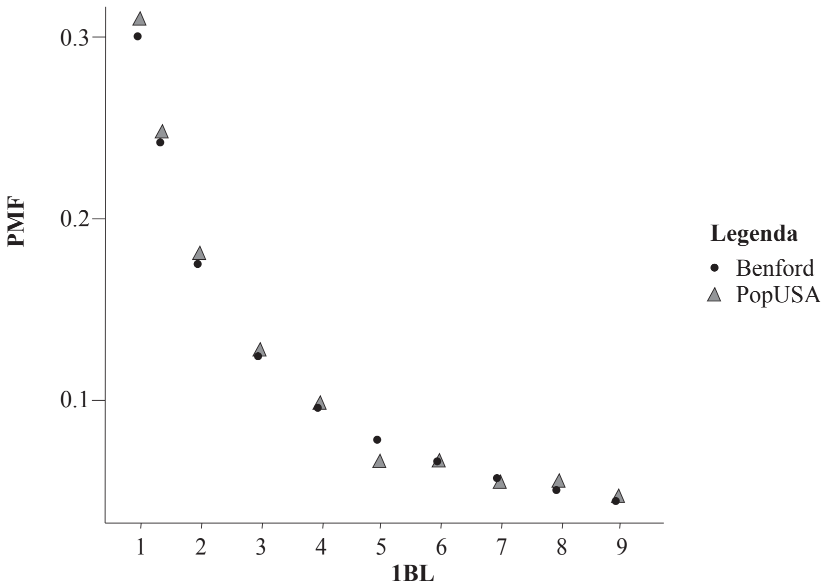
– H0: Empiryczny rozkład danych jest zbieżny z rozkładem teoretycznym Benforda.

Tabela 1

Wynik testu zgodności *chi*-kwadrat, rozkładu Benforda z rozkładem empirycznym populacji stanów wg spisu z 2000 r.

Test wykonany z użyciem pakietu *benford.analysis*, w środowisku języka R
– ***chisq*** = 9.9973 – ***p-value*** = 0.2679

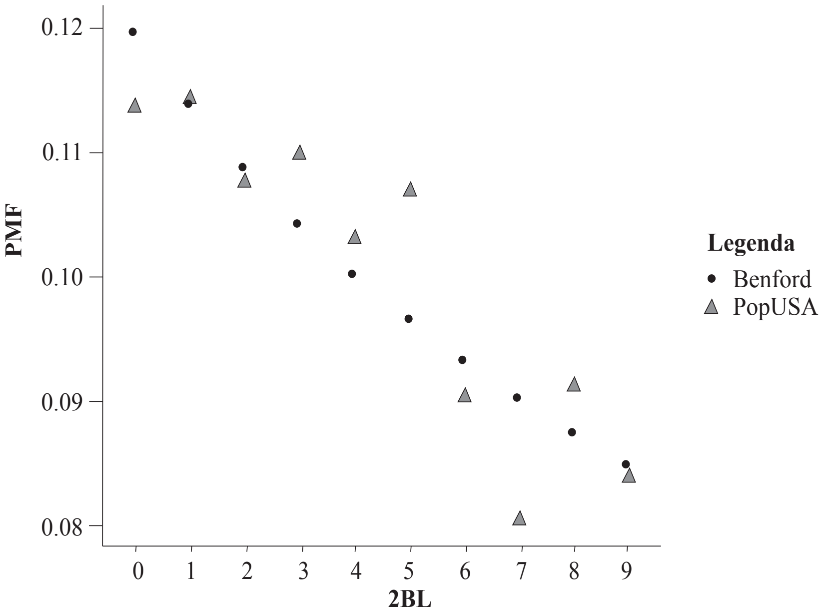
Źródło: Opracowanie własne.



Rys. 1. Prawo Benforda dla pierwszej cyfry znaczącej (**1BL**)

Źródło: Opracowanie własne na podstawie bazy danych zawierającej wyniki spisu powszechnego w USA (2000).

Na rysunku 2 przedstawiono graficznie rozkład Benforda dla drugiej cyfry znaczącej (**2BL**). Również na wykresie 2 pokazano rozkład drugiej cyfry znaczącej dla liczb reprezentujących populacje stanów. W tym przypadku, zbieżność rozkładu 2BL i rozkładu empirycznego nie jest duża.



Rys. 2. Prawo Benforda dla drugiej cyfry znaczącej (**2BL**)

Źródło: Opracowanie własne na podstawie bazy danych zawierającej wyniki spisu powszechnego w USA (2000).

Dane i metoda

Aby zbadać wiarygodność prawa Benforda, przeprowadzono analizę przykładowych elekcji według następującego schematu:

1) wybrano kilka głosowań, w których przypadku nie ma podstaw, aby twierdzić, że były nieuczciwe. Analiza obejmuje wybory w Polsce i, dodatkowo, w Australii. Wybór tych przypadków był arbitralny, ale uzależniony od dostępności do danych na odpowiednim poziomie agregacji;
2) z baz danych zawierających rezultaty wyborów, wybrano obserwacje odnoszące się do wyników na poziomie obwodów (*precincts*);
3) na podstawie wybranych danych zbudowano empiryczne rozkłady częstości względnych dla pierwszych, a także dla drugich, cyfr znaczących liczb określających wyniki kilku najpopularniejszych partii politycznych lub kandydatów we wszystkich obwodach gło-

sowania. Analizę przeprowadzono w środowisku języka programowania R z wykorzystaniem m.in. pakietu funkcji graficznych ggplot2;

4) rozkłady empiryczne przedstawione zostały w postaci wykresów i następnie porównane z wykresami przedstawiającymi teoretyczny rozkład Benforda w wersji dla pierwszej i drugiej cyfry znaczącej.

Analiza podobieństwa rozkładów cyfr pozwoliła odpowiedzieć na pytanie, czy w przypadku analizowanych elekcji występuje zgodność rozkładów empirycznych z rozkładem Benforda. Jeżeli zaobserwowana zostałaby niezgodność rozkładów, to należałoby stwierdzić pojawienie się dowodu na występowanie nieprawidłowości podczas badanych elekcji, co jednak stałoby w sprzeczności z trudnym do podważenia założeniem o uczciwości badanych wyborów. W takiej sytuacji należałoby uznać, że wiarygodność metody identyfikowania fałszerstw wyborczych w oparciu o prawo Benforda jest wątpliwa.

Do badania wybrano elekcje z Polski i Australii. W przypadku Polski, analiza dotyczy wyborów prezydenckich 2015 i wyborów parlamentarnych do Sejmu 2015. Jeśli chodzi o Australię, to dane odnoszą się do wyborów do Izby Reprezentantów 2013 (izba niższa parlamentu). Wybór krajów jest, jak wspomniano, dość przypadkowy, natomiast chodziło o to, aby można było przyjąć założenie o uczciwości wyborów i następnie poszukiwać kontrprzykładów, które ewentualnie ujawnią słabości prawa Benforda pokazując, że sugeruje ono nieprawidłowości w wyborach, które były uczciwe. Co ważne, w przypadku Australii łatwo dostępne były dane dotyczące wyników głosowania na poziomie obwodów, co zdecydowało o wyborze tego kraju – do analizy wybrano dane o wynikach z najbardziej zaludnionego stanu *Nowa Południowa Walia*.

Warto wspomnieć, że w Australii, w wyborach do izby niższej wykorzystuje się jeden z większościowych systemów wyborczych, tzw. system głosowania alternatywnego (w języku angielskim często występujący pod nazwą *Instant Runoff Method*). Okręgi wyborcze są jednomandatowe, ale aby uzyskać mandat, należy zdobyć ponad 50% głosów ważnych w okręgu.

Dane umożliwiające replikację przedstawionej w artykule analizy pochodzą z oficjalnych i powszechnie dostępnych zbiorów:

– PKW: http://pkw.gov.pl/;
– Australian Electoral Commission: http://results.aec.gov.au/17496/Website/HouseResultsMenu-17496.htm.

Wyniki badania

W tej części przedstawiona zostanie analiza przykładów, które dowodzą, że do użyteczności prawa Benforda w kontekście analizy uczciwości wyborów należy podchodzić z dużą ostrożnością. W celu weryfikacji hipotezy badawczej przedstawionej na początku artykułu, sprawdzono zgodność z rozkładem Benforda wyników wyborów, w przypadku których można z prawdopodobieństwem graniczącym z pewnością stwierdzić, że nie były sfałszowane. Jako kontrprzykłady, kolejno analizujemy następujące przypadki.

Dla prawa Benforda odnoszącego się do pierwszej cyfry znaczącej:
– wybory prezydenckie 2015 w Polsce,
– wybory do Sejmu 2015 w Polsce,
– wybory do Izby Reprezentantów 2013 w Australii.

Następnie, dla prawa Benforda odnoszącego się do drugiej cyfry znaczącej:
– wybory prezydenckie 2015 w Polsce,
– wybory do Sejmu 2015 w Polsce,
– wybory do Izby Reprezentantów 2013 w Australii.

Tabela 2

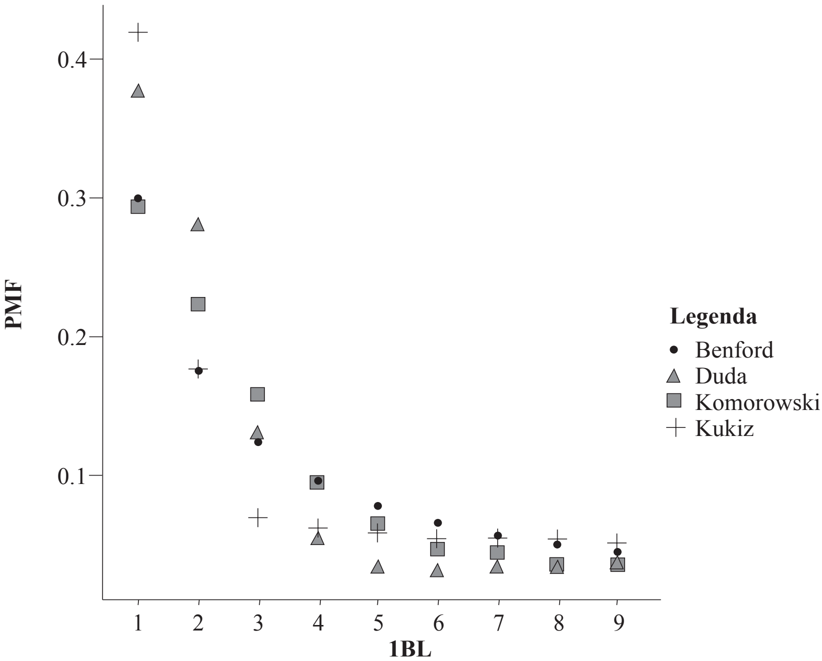
Podstawowe charakterystyki obwodów głosowania

	Wybory prezydenckie 2015	**Wybory do Sejmu 2015**	**Wybory do IR w Australii 2013***
Liczba obwodów	27 817	27 859	3 022
Średnia liczba uprawnionych do głosowania w obwodach	1 103	1 099	brak danych

* Wyniki wyborów dla stanu *Nowa Południowa Walia*.

Źródło: Opracowanie własne na podstawie danych: PKW: http://pkw.gov.pl/ oraz Australian Electoral Commission: http://results.aec.gov.au/17496/Website/HouseResultsMenu-17496.htm.

Rysunek 3 przedstawia rozkłady empiryczne pierwszej cyfry znaczącej dla trzech kandydatów, którzy uzyskali najlepsze wyniki w wyborach prezydenckich w Polsce w 2015 (I tura) – dla Andrzeja Dudy, Bronisława Komorowskiego i Pawła Kukiza, w porównaniu z rozkładem teoretycznym Benforda. Na wykresach, koła ● oznaczają punkty teoretyczne rozkładu Benforda, a pozostałe symbole reprezentują rozkłady empiryczne – zob. legendę. Analiza wykresów dla kilku wybranych kandydatów pokazuje, że zgodność z rozkładem Benforda jest nieznaczna.



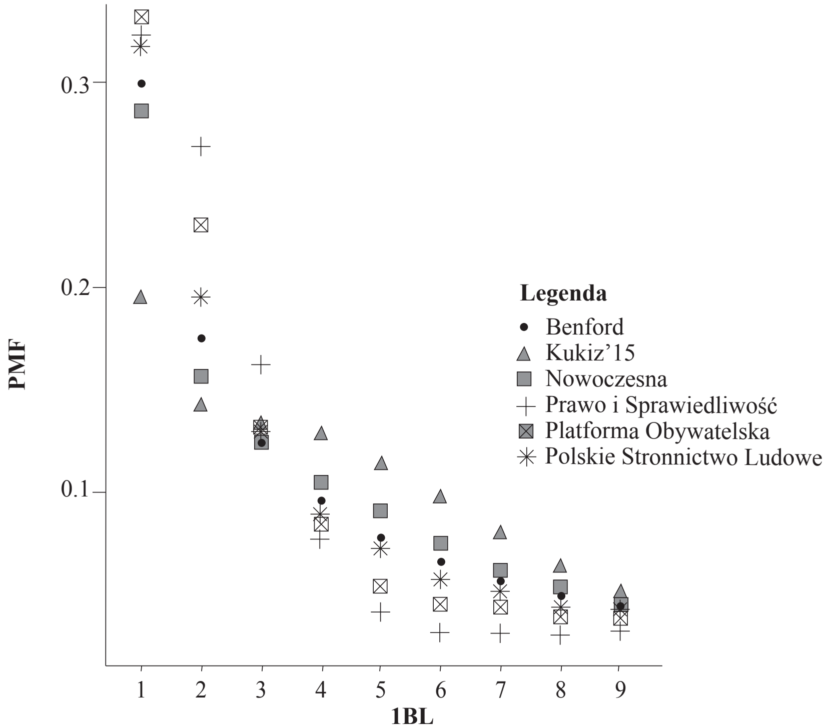
Rys. 3: Rozkład pierwszej cyfry znaczącej dla wyników wyborów prezydenckich 2015

Źródło: Opracowanie własne na podstawie danych: PKW: http://pkw.gov.pl/.

Dla przykładu, w przypadku pierwszej cyfry znaczącej, tylko wyniki Bronisława Komorowskiego wskazują na zgodność z wartością oczekiwaną w rozkładzie Benforda. W przypadku Andrzeja Dudy i Pawła Kukiza, rozkład pierwszych cyfr znaczących pokazuje, że cyfra „jeden” występuje zbyt często niż można by tego oczekiwać – oczywiście, przy założeniu, że prawo Benforda ma zastosowanie do wyników wyborów.

Jeżeli chodzi o wybory do Sejmu 2015, to w tym przypadku również obserwujemy brak zgodności rozkładów empirycznych pierwszych cyfr znaczących z rozkładem teoretycznym. Szczególnie duża rozbieżność występuje dla komitetu wyborczego Kukiz’15 (K), a także dla Prawa i Sprawiedliwości.

Czy na tej podstawie można by twierdzić, że doszło do nieprawidłowości podczas elekcji? Należy raczej stwierdzić, że rozkład pierwszych cyfr znaczących liczb głosów uzyskanych przez poszczególnych kandydatów/partie polityczne nie musi być zbieżny z rozkładem Benforda, również wtedy, gdy wybory są uczciwe. Pewne podobieństwo wspomnianych rozkładów występuje, ale różnice, widoczne wyraźnie na wykresach, są dość istotne, szczególnie w niektórych przypadkach.

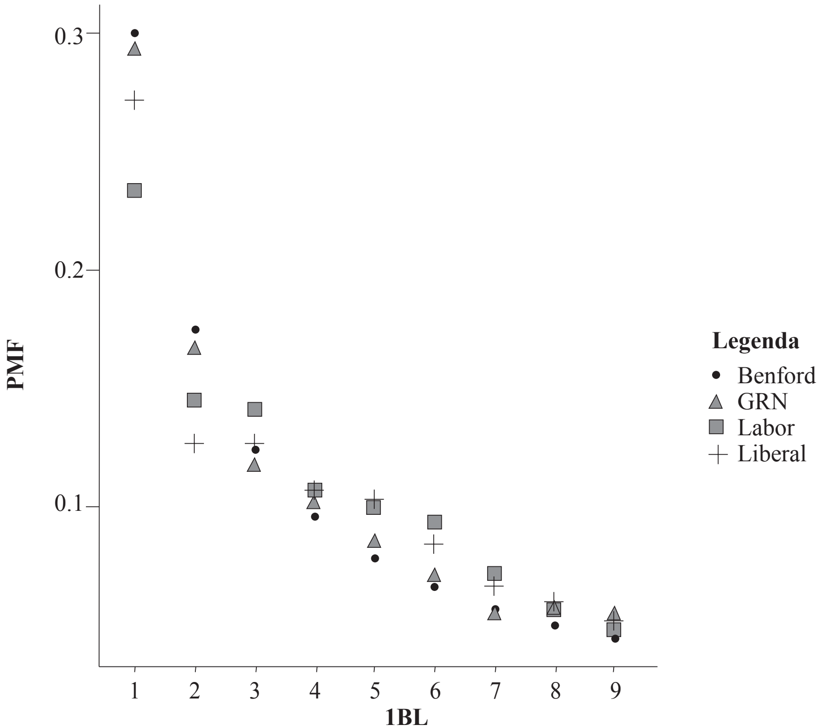


Rys. 4. Rozkład pierwszej cyfry znaczącej dla wyników wyborów do Sejmu 2015

Źródło: Opracowanie własne na podstawie danych: PKW: http://pkw.gov.pl/.

Ponadto, dla porównania, na rysunku 5 przedstawiono rozkłady empiryczne pierwszych cyfr liczb głosów w wyborach w Australii (analiza dotyczy tzw. pierwszych preferencji na kartach do głosowania w obwodach). Porównując otrzymane wyniki z rozkładem teoretycznym Benforda obserwujemy wprawdzie nieco lepsze dopasowanie w porównaniu do wyborów w Polsce, ale rozbieżności są nadal stosunkowo duże i, np. w przypadku Partii Pracy (*Labor Party*), zgodność z rozkładem teoretycznym jest mała.

Podsumowując, wyniki analizy zgodności rozkładów empirycznych pierwszej cyfry znaczącej liczb głosów z rozkładem teoretycznym Benforda (na poziomie obwodów do głosowania) ujawniają, że w przypadku niektórych partii politycznych/kandydatów, stopień zgodności rozkładów jest niski, chociaż nie ma podstaw, aby podejrzewać fałszerstwa wyborcze w analizowanych wyborach. Ta obserwacja pozwala wyciągnąć wniosek, że niezgodność rozkładu empirycznego z rozkładem teoretycznym Benforda nie może być interpretowana jako dowód na występowanie nieprawidłowości podczas elekcji.

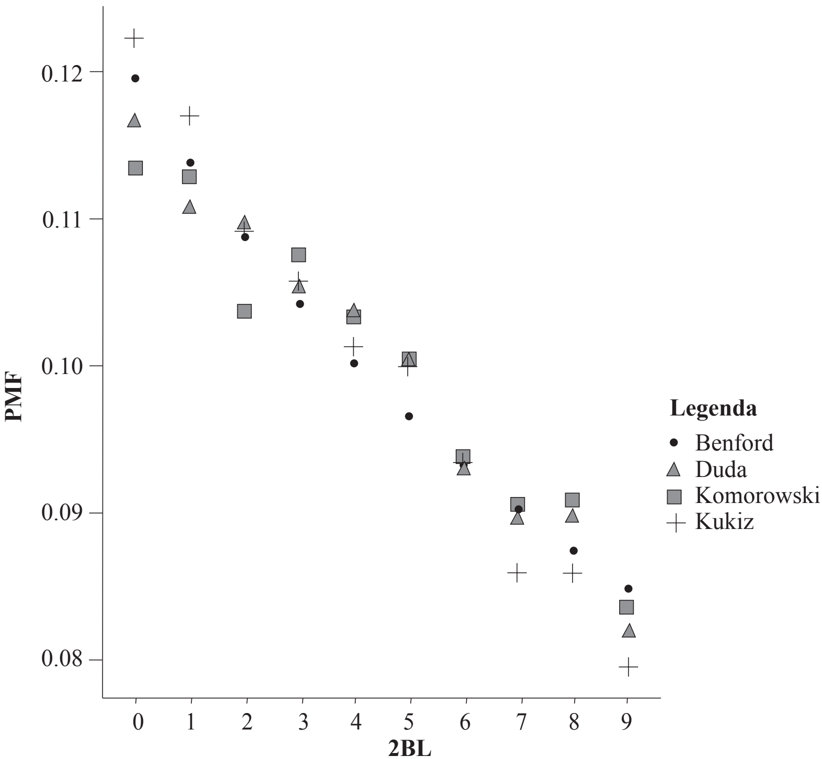


Rys. 5. Rozkład pierwszej cyfry znaczącej dla wyników wyborów do Izby Reprezentantów w Australii 2013

Źródło: Opracowanie własne na podstawie danych: Australian Electoral Commission: http://results.aec.gov.au/17496/Website/HouseResultsMenu-17496.htm

* * *

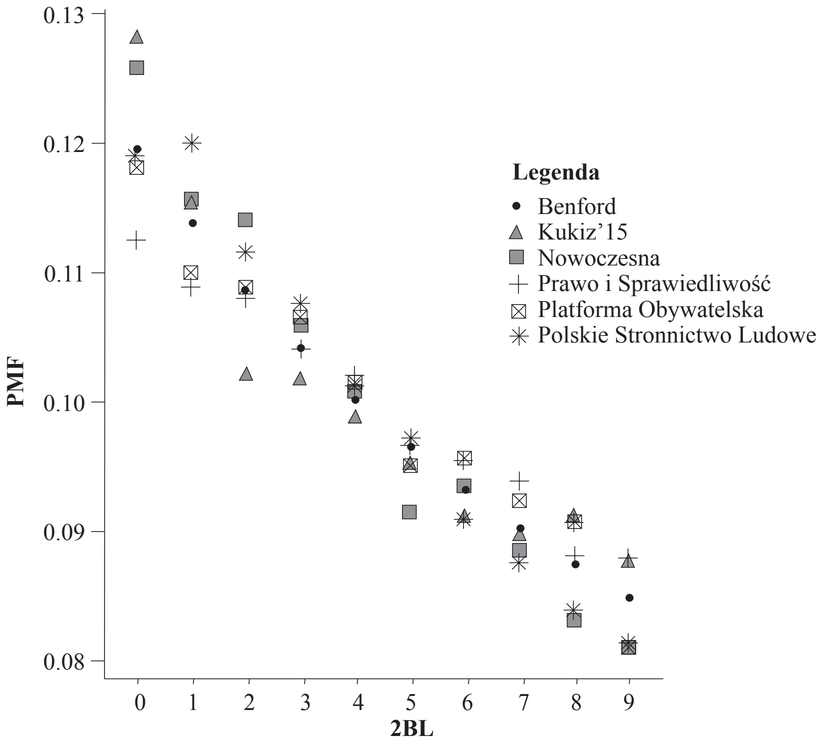
Powstaje pytanie, czy większą skuteczność w identyfikowaniu fałszerstw wyborczych można by osiągnąć stosując prawo Benforda dla drugiej cyfry znaczącej. Na rysunku 6 przedstawiono porównanie rozkładu teoretycznego Benforda dla drugiej cyfry znaczącej z rozkładami empirycznymi drugiej cyfry w przypadku wyborów **prezydenckich 2015**. Jak widać na wykresie, zbieżność rozkładów nie jest idealna, ale różnice nie są znaczne. Taki wynik może sugerować, że efektywność 2BL jest wyższa w porównaniu do 1BL. Czy w przypadku wyborów do Sejmu 2015 i wyborów do Izby Reprezentantów w Australii poziom zgodności rozkładów jest też tak duży? Jeśli tak by było, to mogłoby to sugerować, że 2BL można pomocniczo wykorzystać jako narzędzie do identyfikowania nieprawidłowości podczas wyborów. Wyniki badania przedstawiono na wykresach na rysunkach 7, 8 i 9.



Rys. 6. Rozkład drugiej cyfry znaczącej dla wyników wyborów prezydenckich 2015

Źródło: Opracowanie własne na podstawie danych: PKW: http://pkw.gov.pl/.

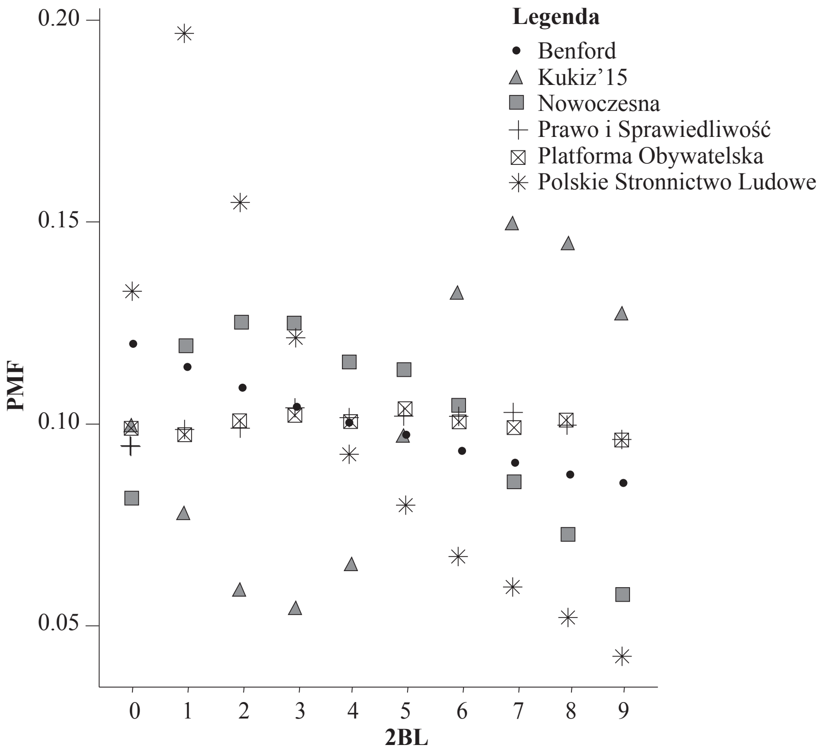
Rysunek 7 dowodzi, że zbieżność rozkładów empirycznych i teoretycznego jest na wyższym poziomie w porównaniu do wyników analizy zbieżności rozkładów dla prawa Benforda w przypadku pierwszej cyfry znaczącej. Niemniej, podobnie jak w przypadku wyborów prezydenckich 2015, podobieństwo rozkładów nie jest doskonałe. Widać to np. w przypadku PSL-u, dla cyfry „jeden”, lub w przypadku Nowoczesnej (N) i Kukiz’15, dla cyfry 2, a także cyfry „0”. Jednak, nadal można powiedzieć, że różnice nie są znaczne, aczkolwiek trudno określić jak duże miałyby być odstępstwa od rozkładu teoretycznego, aby można było podejrzewać występowanie nieprawidłowości podczas elekcji.



Rys. 7. Rozkład drugiej cyfry znaczącej dla wyników wyborów do Sejmu RP 2015

Źródło: Opracowanie własne na podstawie danych: PKW: http://pkw.gov.pl/.

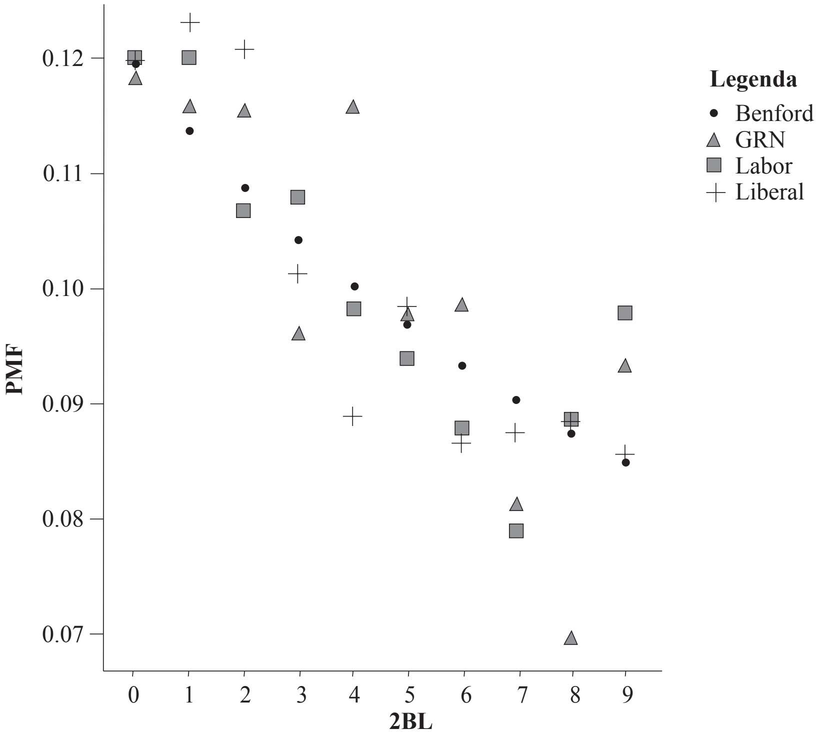
Na rysunku 8 pokazano dodatkowo rozkład drugich cyfr liczb głosów oddanych w obwodach do głosowani na poszczególne komitety wyborcze po przeliczeniu ich na procenty. Jak się okazuje, w tym przypadku rozbieżność rozkładów empirycznych i 2BL jest znaczna. Okazuje się, że zastąpienie liczb bezwzględnych przez wartości procentowe nie poprawia zbieżności z rozkładem Benforda w przypadku wyników wyborów. Wynika to z faktu, że **prawo Benforda działa najlepiej, wtedy, gdy spełnione jest założenie, że analizujemy liczby, które mogą przyjmować różne rzędy wielkości**, co nie jest prawdą w przypadku procentów.



Rys. 8: Rozkład drugiej cyfry znaczącej dla wyników wyborów do Sejmu RP 2015 (dla wyników w procentach)

Źródło: Opracowanie własne na podstawie danych: PKW: http://pkw.gov.pl/.

Ostatni przypadek to wybory do Izby Reprezentantów w Australii. Rysunek 9 dowodzi, że również w przypadku rozkładu drugich cyfr znaczących, różnice między rozkładem teoretycznym a empirycznymi mogą być znaczne, chociaż analiza dotyczy wyborów, które nie były sfałszowane. Na rysunku 9 widać szczególnie duże rozbieżności w przypadku liczby głosów The Greens (GRN).



Rys. 9. Rozkład drugiej cyfry znaczącej dla wyników wyborów do Izby Reprezentantów w Australii 2013

Źródło: Opracowanie własne na podstawie danych: Australian Electoral Commission: http://results.aec.gov.au/17496/Website/HouseResultsMenu-17496.htm.

Podsumowanie

Wskazane w artykule przykłady zastosowania tzw. rozkładu Benforda nie mogą być wprawdzie traktowane jako jednoznaczny dowód na bezużyteczność tego narzędzia w kontekście badania nieprawidłowości podczas elekcji, ale na pewno poddają w wątpliwość jego skuteczność. Nie jest to wcale zaskakujące, gdyż trudno wskazać przesłanki pozwalające sądzić, że tzw. prawo Benforda dobrze opisuje rozkład cyfr tworzących liczby głosów oddanych np. w obwodach/okręgach na poszczególnych kandydatów/partie polityczne. Co istotne, brakuje przekonującej teorii wyjaśniającej, która uzasadniałaby jego użycie.

Podsumowując, prawo Benforda nie jest skutecznym narzędziem pozwalającym na efektywne tropienie nieprawidłowości podczas wyborów. Wprawdzie, jest możliwe, że sfałszowane wyniki wyborów spowodują, że analiza wykaże istnienie odchylenia od rozkładu Benforda dla pewnych cyfr, ale samo odchylenie wcale nie musi wynikać z faktu, że wybory zostały poddane manipulacji.

W świetle przedstawionych przykładów, zarówno w przypadku rozkładu pierwszej cyfry znaczącej (1BL), jak i w przypadku rozkładu dla cyfry drugiej (2BL), należy z ostrożnością podchodzić do wiarygodności analizowanej metody w kontekście badania nieprawidłowości podczas wyborów, albowiem jej użycie może skutkować wskazaniem występowania fałszerstw wyborczych, w przypadku elekcji, w których uczciwość trudno wątpić.

Bibilografia

Berger A., Hill T. P. (2011), *A basic theory of Benford's Law*, „Probability Surveys”, nr 8, s. 1–126.

Berger A., Hill T. P. (2011), *Benford's Law strikes back: No simple explanation in sight for mathematical gem*, „The Mathematical Intelligencer”, nr 33(1), s. 85–91.

Breunig C., Goerres A. (2011), *Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in Unified Germany*, „Electoral Studies”, nr 30(3), s. 534–545.

Brock T. (2014), *Benford's law and elections – part 2*, http://datatodisplay.com/blog/politics/benfords-law-elections-2/.

Deckert J., Myagkov M., Ordeshook P. C. (2011), *Benford's Law and the detection of election fraud*, „Political Analysis”, nr 19(3), s. 245–268.

Fewster R. (2009), *A Simple Explanation of Benford's Law*, „American Statistician”, nr 63(1), s. 20–25.

Gawron P., Pawela Ł., Puchała Z., Szklarski J., Życzkowski K. (2015) *Wybory samorządowe 2014 w poszukiwaniu anomalii statystycznych*, „Studia wyborcze”, nr 30(3), s. 534–545.

Gelman A. (2009) *Unconvincing (to me) Use of Benford's Law to Demonstrate Election Fraud in Iran*, http://fivethirtyeight.com/features/unconvincing-to-me-use-of-benfords-law/.

Geyer C. L., Williamson P. P. (2004), *Detecting fraud in data sets using Benford's Law*, „Communications in Statistics-Simulation and Computation”, nr 33(1), s. 229–246.

Mebane W. R. (2011), *Comment on „Benford's Law and the detection of election fraud"*, „Political Analysis”, nr 19(3), s. 269–272.

Mebane W. R. (2006), September, *Election forensics: the second-digit Benford's law test and recent American presidential elections*, Election Fraud Conference.

Pericchi L., Torres D. (2011), *Quick Anomaly Detection by the Newcomb-Benford Law, with Applications to Electoral Processes Data from the USA*, *Puerto Rico and Venezuela*, „Statistical Science”, nr 26(4), s. 502–516.

Roukema B. F. (2009), *Benford's Law anomalies in the 2009 Iranian presidential election*, „Journal of Applied Statistics”, nr 41(1), s. 164–199.

Tam Cho W. K., Gaines B. J. (2007), *Breaking the (Benford) law: Statistical fraud detection in campaign finance*, „The American Statistician”, nr 61(3), s. 218–223.

Źródła danych:

PKW, http://pkw.gov.pl/.

Australian Electoral Commission, http://results.aec.gov.au/17496/Website/HouseResultsMenu-17496.htm.

Revealing Election Fraud and Benford's „Law”

Summary

Is it possible to identify election fraud by using statistical tools? The authors of several articles published in the last few years, e.g. Roukema (2009), Perrichi and Torres (2011) and Mebane (2013), used a research method which they believe can help in detecting fraud during elections. This article addresses the question of whether the so-called Benford's „Law,” which was used to verify irregularities during the presidential elections in Iran (2009), among others, is a reliable research tool. Unfortunately, empirical examples indicate that the usefulness of this method is questionable.

Key words: Benford's „Law”, elections, revealing election fraud

Data przekazania tekstu: 21.11.2017; data zaakceptowania tekstu: 8.12.2017.