



FUNDACJA
IM. STEFANA BATOREGO

forumIdei

Fałszerstwa czy fałszywe alarmy? Statystyczna kontrola wyników II tury wyborów prezydenckich 2025

Dominik Batorski, Jarosław Flis, Piotr Szulc

Wprowadzenie

Podnoszone publicznie wątpliwości dotyczące uczciwości wyborów na urząd Prezydenta RP skłoniły nas do szczegółowego przeanalizowania wyników II tury. Naszym celem jest sprawdzenie, które z przedstawianych hipotez znajdują potwierdzenie w pogłębionych badaniach i jaka jest skala nieprawidłowości, które mogą być wykryte poprzez analizę oficjalnie podanych wyników. Nasze podejście opiera się na kilku założeniach:

- W wydarzeniu o tej skali nieprawidłowości są nieuchronne – zarówno te wynikające z ludzkich błędów, jak i te mające swe źródło w intencjonalnym naruszeniu reguł.
- Wyniki wyborów w lokalnej skali są w znacznym stopniu przewidywalne, lecz nieuchronna – choć także przewidywalna – jest skala odchylenia od ogólnego, dominującego wzorca.
- Na podstawie modeli ujawniających standardowe wzory można wskazać miejsca, gdzie mogły nastąpić błędy lub celowe oszustwa.
- Może się zdarzyć, że pojawiają się zjawiska systematycznie odkształcające intencje głosujących, czego przykładem było zastosowanie zbroszurowanej karty wyborczej w wyborach

samorządowych 2014 roku. Wykrycie takich problemów wymaga dodatkowych źródeł informacji¹. Nie ma jednak żadnych sygnałów, że takie zjawisko – masowe nieintencjonalne odkształcenie – mogłoby mieć miejsce w tegorocznych wyborach.

- Teoretycznie można też sobie wyobrazić, że pojawia się zorganizowana akcja, która w równomierny sposób obejmuje wyraźną większość komisji, polegająca na celowym fałszowaniu wyników. Taka akcja także nie byłaby do wykrycia za pomocą analizy oficjalnych wyników. Zważywszy jednak na skalę wydarzenia, jakim jest obliczanie wyników przez ponad ćwierć miliona osób, taki scenariusz wydaje się skrajnie nieprawdopodobny.

Pozostając przy hipotezie, że omyłkowe i intencjonalne zjawiska są wyjątkami a nie regułami, oszacowanie ich skali, kierunku i ostatecznego efektu ma szczególne znaczenie w sytuacji, gdy wynik wyborów był tak wyrównany. Temu właśnie służy nasza analiza. Dodatkowo, w dalszej części opracowania, odnosimy się do wyników podobnych przedsięwzięć, które prezentowane były w debacie publicznej.

Model

Aby znaleźć potencjalne nieprawidłowości w wynikach wyborów, stworzyliśmy zaawansowany model statystyczny. Jego głównym celem jest ustalenie, jak powinien wyglądać „typowy” wynik w danej komisji, a następnie wyłapanie tych miejsc, które od tego wzorca znacząco odbiegają. Podejście to jest dobrą metodą wykrywania potencjalnych nieprawidłowości, ponieważ wynik kandydata w II turze powinien być silnie skorelowany z sumą głosów oddanych w I turze na niego i na kandydatów o zbliżonym profilu ideologicznym². Poniżej wyjaśniamy ten proces krok po kroku, a dodatkowe, bardziej szczegółowe informacje zamieszczamy w załączniku.

Sposób konstrukcji modelu

Krok 1: Oczyszczanie danych – skupiamy się na tym, co typowe

Zanim zaczęliśmy budować model, musieliśmy odpowiednio przygotować dane. Postanowiliśmy skupić się na standardowych, stałych komisjach wyborczych, dlatego z analizy wykluczaliśmy:

- **Komisje o specjalnej charakterystyce:** szpitale, domy pomocy społecznej, areszty śledcze, zakłady karne czy komisje zagraniczne. Głosuje się w nich inaczej niż w „zwykłych” komisjach, więc ich wyniki zaburzyłyby ogólny obraz.
- **Najmniejsze komisje:** odrzuciliśmy też te komisje, w których oddano mniej niż 100 głosów. Przy tak małej liczbie głosujących wyniki są często bardzo losowe i nie odzwierciedlają szerszych trendów.

1 Por. G. Adam i in., *Nieważne głosy, ważny problem. Wyniki badania kart do głosowania z wyborów do sejmików województw 2014*, Fundacja im. Stefana Batorego, Warszawa 2016, s. 169; M. Pierzgalski, M.A. Górecki, P. Stępień, *Ballot paper design and vote spoiling at Polish local elections of 2014: Establishing a causal link*, „East European Politics and Societies” 2020, nr 34(3), s. 611–636; J. Flis, M.M. Kaminski, *Party-related primacy effects in proportional representation systems: Evidence from a natural experiment in Polish local elections*, „Public Choice” 2022, nr 190(3), s. 345–363.

2 Wykorzystywaliśmy analogiczne podejścia we wcześniej publikowanych analizach, por. P. Szulc, *Model detekcji anomalii wyborczych w wyborach prezydenckich*, 26 czerwca 2025, <https://danetyka.com/model-detekcji-anomalii/>; D. Batorski, *Anomalie wyborcze? Raczej nie*, 18 czerwca 2025, <https://mamprawowiedziec.pl/czytelnia/arttykul/anomalie-wyborcze-raczej-nie>. Dostęp w przypisach: 27 czerwca 2025.

Dzięki tej selekcji uzyskaliśmy czysty i spójny zbiór danych, który pozwolił na zbudowanie bardziej wiarygodnego modelu.

Krok 2: Kalibracja modelu – usuwanie statystycznych „zakłóceń”

Każdy model statystyczny, aby był precyzyjny, musi zostać odpowiednio „skalibrowany”. W naszym przypadku polegało to na zidentyfikowaniu i czasowym usunięciu z analizy komisji, które były na tyle nietypowe, że mogłyby zaburzyć działanie całego modelu.

Wykorzystaliśmy do tego standardowe techniki statystyczne, które pozwoliły nam odpowiedzieć na dwa pytania:

- Które komisje najbardziej odstają od wstępnych przewidywań?
- Które pojedyncze komisje mają nieproporcjonalnie duży wpływ na kształt całego modelu?

Po zidentyfikowaniu takich statystycznych „odstających” obserwacji (tzw. outlierów), zbudowaliśmy ostateczną wersję modelu na czystym zbiorze prawie 30 tysięcy komisji. Dzięki temu upewniliśmy się, że nasz model odzwierciedla ogólnokrajowe, stabilne tendencje, a nie jest zaburzony przez pojedyncze, ekstremalne przypadki.

Krok 3: Budowa i wykorzystanie modeli predykcyjnych

Na przygotowanym w poprzednich krokach zbiorze danych zbudowaliśmy modele regresji liniowej, przewidujące liczbę głosów uzyskanych przez Rafała Trzaskowskiego i Karola Nawrockiego w poszczególnych komisjach. Model ten opiera się na informacjach o rezultatach z I tury oraz dodatkowych zmiennych, takich jak liczba nowych wyborców czy liczba wydanych zaświadczeń. Dzięki temu dla każdej komisji uzyskujemy przewidywany – czyli typowy – wynik II tury.

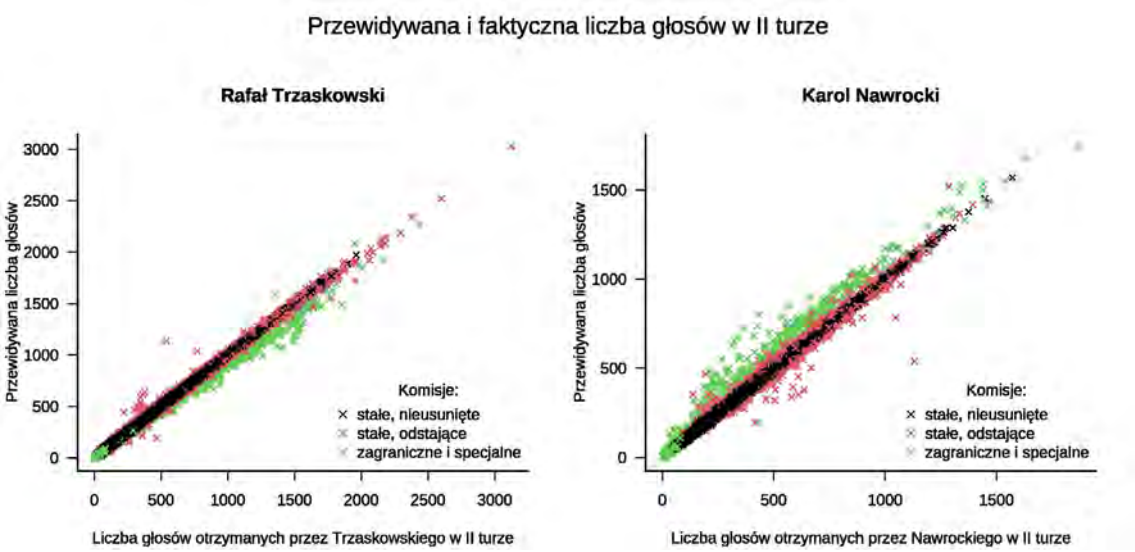
Tak skonstruowany model stanowi punkt odniesienia pozwalający na wykrywanie miejsc, gdzie wyniki odbiegają od ogólnopolskich wzorców. W dalszej analizie prezentujemy wyniki modeli i porównujemy rzeczywiste rezultaty z wartościami przewidywanymi, by zidentyfikować komisje o nietypowych wynikach, wymagających dodatkowej weryfikacji.

Wyniki modeli statystycznych

Aby zrozumieć, jak kształtują się wyniki wyborów w poszczególnych komisjach i wychwycić potencjalne nieprawidłowości, skonstruowaliśmy dwa zaawansowane modele statystyczne. Ich zadaniem było przewidzenie liczby głosów oddanych w II turze na każdego z kandydatów (Rafała Trzaskowskiego i Karola Nawrockiego), w oparciu o szczegółowe dane z I tury oraz dodatkowe czynniki, takie jak liczba nowych wyborców czy liczba osób głosujących na podstawie zaświadczeń. Dzięki temu mogliśmy zobaczyć, które komisje odbiegały od przewidywań – i tym samym wyłowić miejsca potencjalnie wartę szczegółowej kontroli.

Oba modele okazały się niezwykle trafne. Zdecydowana większość wyników komisji była bardzo dobrze przewidywana przez nasz algorytm, a współczynnik dopasowania modelu (R^2) wyniósł aż 99,86% dla Trzaskowskiego i 99,68% dla Nawrockiego. Oznacza to, że analizując dane z I tury oraz uwzględniając profile kandydatów, można niemal dokładnie przewidzieć, ile głosów dany kandydat uzyskał w II turze w każdej z typowych komisji w Polsce. Pokazuje to poniższy wykres, na którym czarnymi punktami zaznaczone są komisje stałe, które były podstawą do konstrukcji modelu. Układają się one

prawie idealnie wzdłuż przekątnej wykresu, dla której wartości przewidywane zgadzają się z liczbą faktycznie otrzymanych głosów.



Poniżej przedstawiamy podsumowanie kluczowych wyników modeli:

Tabela 1. Wyniki modeli regresji przewidujących liczbę głosów oddanych na kandydatów w II turze

Zmienna	Trzaskowski2	(SE)	Nawrocki2	(SE)
(stała)	-1,76	0,13	1,76	0,13
Trzaskowski (wynik w I turze)	1,10	0,0010	-0,10	0,0010
Nawrocki (wynik w I turze)	-0,11	0,0012	1,11	0,0012
zmiana liczby głosujących (względem I tury)	0,49	0,0031	0,51	0,0031
Hołownia	0,96	0,0069	0,04	0,0069
Zandberg	1,01	0,0050	-0,01	0,0050
Biejat	0,87	0,0080	0,13	0,0080
Senyszyn	0,96	0,021	0,04	0,021
Mentzen	0,13	0,0029	0,87	0,0029
Braun	0,04	0,0053	0,96	0,0053
Jakubiak	0,11	0,023	0,89	0,023

inni kandydaci	0,48	0,014	0,52	0,014
nowe zaświadczenia	0,17	0,0073	-0,17	0,0073

Jak interpretować te liczby?

- Wysokie współczynniki przy zmiennych odpowiadających liczbie głosów, jakie w I turze zdobywali Trzaskowski i Nawrocki pokazują, że wyniki z I tury są bardzo mocnym predyktorem liczby głosów na tych kandydatów w II turze – co potwierdza intuicję, że wyborcy są bardzo konsekwentni.
- Wartości przy innych kandydatach z I tury pokazują, jak głosy ich wyborców „przechodziły” na kandydatów w II turze. Przykładowo, im więcej głosów na Hołownię, tym więcej w II turze na Trzaskowskiego, ale już niekoniecznie na Nawrockiego.
- Dodatnie współczynniki przy zmiennej „zmiana liczby wyborców” oznaczają, że wszędzie tam, gdzie pojawiło się więcej nowych wyborców, korzystali na tym obaj kandydaci, choć nieco bardziej Nawrocki. Z drugiej strony, zmienna „nowe zaświadczenia” ma przeciwny znak dla obu kandydatów: wzrost liczby osób głosujących na podstawie zaświadczenia wiązał się z większą liczbą głosów na Trzaskowskiego, ale mniejszą na Nawrockiego.

Tak zbudowane modele stanowią bardzo precyzyjny punkt odniesienia dla dalszej analizy i umożliwiają skuteczne wyłapywanie komisji z nietypowymi, potencjalnie nieprawidłowymi wynikami.

Identyfikacja anomalii

Mając tak precyzyjne narzędzie, możemy je teraz wykorzystać jako punkt odniesienia. Dla każdej stałej komisji w kraju (a więc również tych, które nie były uwzględniane w tworzeniu modeli) porównujemy jej **rzeczywisty wynik** z wynikiem, jaki **przewidział dla niej nasz model**.

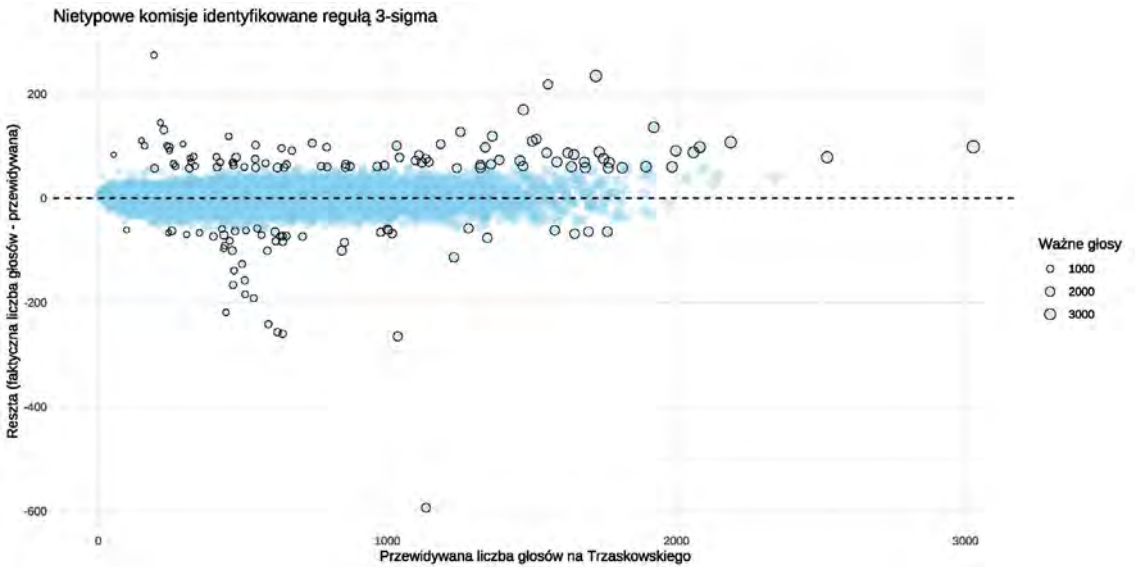
Jeżeli różnica między rzeczywistością a przewidywaniem jest znacząca, oznaczamy taką komisję jako **anomalie statystyczną**.

Należy podkreślić, że „anomalia” nie jest równoznaczna z „fałszerstwem”. Jest to jednak ważny sygnał, że w danej komisji wydarzyło się coś nietypowego, co odbiega od ogólnopolskiego wzorca. Taki przypadek może być efektem błędu w protokole, specyficznych lokalnych uwarunkowań, ale może również wskazywać na nieprawidłowości. Każda taka komisja staje się więc kandydatem do dalszej, szczegółowej weryfikacji.

Poniższy wykres przedstawia odchylenia wartości przewidywanych od wartości podanych przez Okręgowe Komisje Wyborcze. Komisje zaznaczone obwódką są identyfikowane jako odstające według metody 3 sigma³. Wykres przedstawia wyłącznie komisje stałe. Analizy zawierające również inne typy komisji znajdują się w załączniku. Pokazują one znaczną liczbę odchyleń dla komisji zagranicznych, które nie podlegają takim samym zależnościom, jak komisje krajowe, i niejako z definicji są nietypowe.

³ Metoda bazuje na tym, że przy rozkładzie normalnym jedynie ok. 0,27% wartości jest odległych od średniej o więcej niż trzy odchylenia standardowe. Rozkład reszt jest mniej więcej normalny o średniej zero, dlatego to te, które są większe od trzech odchyleń (co do wartości bezwzględnej), traktuje się jako odstające. Warto tu jednak zauważyć, że część z nich „powinna” być tak odległa (0,27%, czyli ok. 86 komisji), a przez to część wskazanych przez nas anomalii to najprawdopodobniej tzw. wyniki fałszywie dodatnie. Jeśli chcemy uniknąć takiej sytuacji (podać listę komisji, spośród których wyników fałszywie dodatnich powinno być bardzo mało), należy być bardziej rygorystycznym. Jednym z rozwiązań jest zastosowanie progu 4s zamiast 3s, wtedy oczekiwana liczba niepoprawnie wskazanych przez nasz algorytm komisji wynosi jedynie 2.

W ich przypadku ten brak typowości nie może być traktowany jako wskaźnik występowania nieprawidłowości.



Jak można zaobserwować, różnice są w obie strony. W tabeli 2 przedstawiamy komisje, w których zaobserwowaliśmy różnice przynajmniej 120 głosów między wartościami ogłoszonymi a przewidywanymi przez nasze modele. Pełna lista 138 komisji, których wyniki można uznać za odstające, znajduje się w załączniku.

Tabela 2. Skrócona lista komisji z wynikami odstającymi od przewidywanych – wyłącznie komisje, gdzie różnica między faktyczną a oczekiwaną liczbą głosów na kandydata wynosi minimum 120. Kolorem tła wyróżniono dotychczas zweryfikowane komisje. Pełna lista komisji dostępna w załączniku.

			Wyniki wg OKW		Przewidywana liczba głosów		różnica
			Trzaskowski	Nawrocki	Trzaskowski	Nawrocki	
1	95	m. Kraków	540	1132	1133	539	-593
2	61	m. Bielsko-Biała	771	1048	1036	783	-265
3	3	gm. Olesno	378	637	638	377	-260
4	13	m. Mińsk Mazowiecki	363	611	620	354	-257
5	35	m. Tychy	347	581	588	340	-241
6	9	gm. Strzelce Opolskie	223	416	442	197	-219

7	17	m. Gdańsk	346	585	537	394	-191
8	25	m. Grudziądz	324	504	508	320	-184
9	403	m. Kraków	300	668	466	502	-166
10	30	m. Bielsko-Biała	349	610	506	453	-157
11	4	gm. Brześć Kujawski	331	466	469	328	-138
12	12	m. Kobyłka	371	507	497	381	-126
	...	...					
131	706	Ursus	1380	459	1253	586	127
132	13	m. Zakopane	358	834	227	965	131
133	73	m. Gdańsk	2058	671	1922	807	136
134	4	gm. Staszów	360	209	215	354	145
135	10	gm. Trzebiatów	1640	852	1470	1022	170
136	113	Mokotów	1774	136	1555	355	219
137	1	gm. Jastarnia	1956	1286	1721	1521	235
138	1	gm. Magnuszew	467	193	192	468	275

Znaczna liczba spośród komisji, w których są największe odchylenia, była weryfikowana przez Sąd Najwyższy i faktycznie stwierdzono tam że doszło do błędów w liczeniu głosów. Wątek ten poruszamy szerzej w części dotyczącej odwróconych wyników.

Oszacowanie skali pomyłek i ich wpływu na wynik wyborów

Wśród 138 komisji, w których odnotowaliśmy nietypowe wyniki, korzystając z reguły 3 sigma, w 89 Rafał Trzaskowski miał w wynikach ogłoszonych przez OKW wyższe wyniki od przewidywanych, a w 49 różnice były na korzyść Karola Nawrockiego. W tych pierwszych, Trzaskowski zdobył łącznie o 7727 głosów więcej niż wynika z przewidywania modelu. Z kolei w pozostałych, Karol Nawrocki dostał o 5512 głosów więcej.

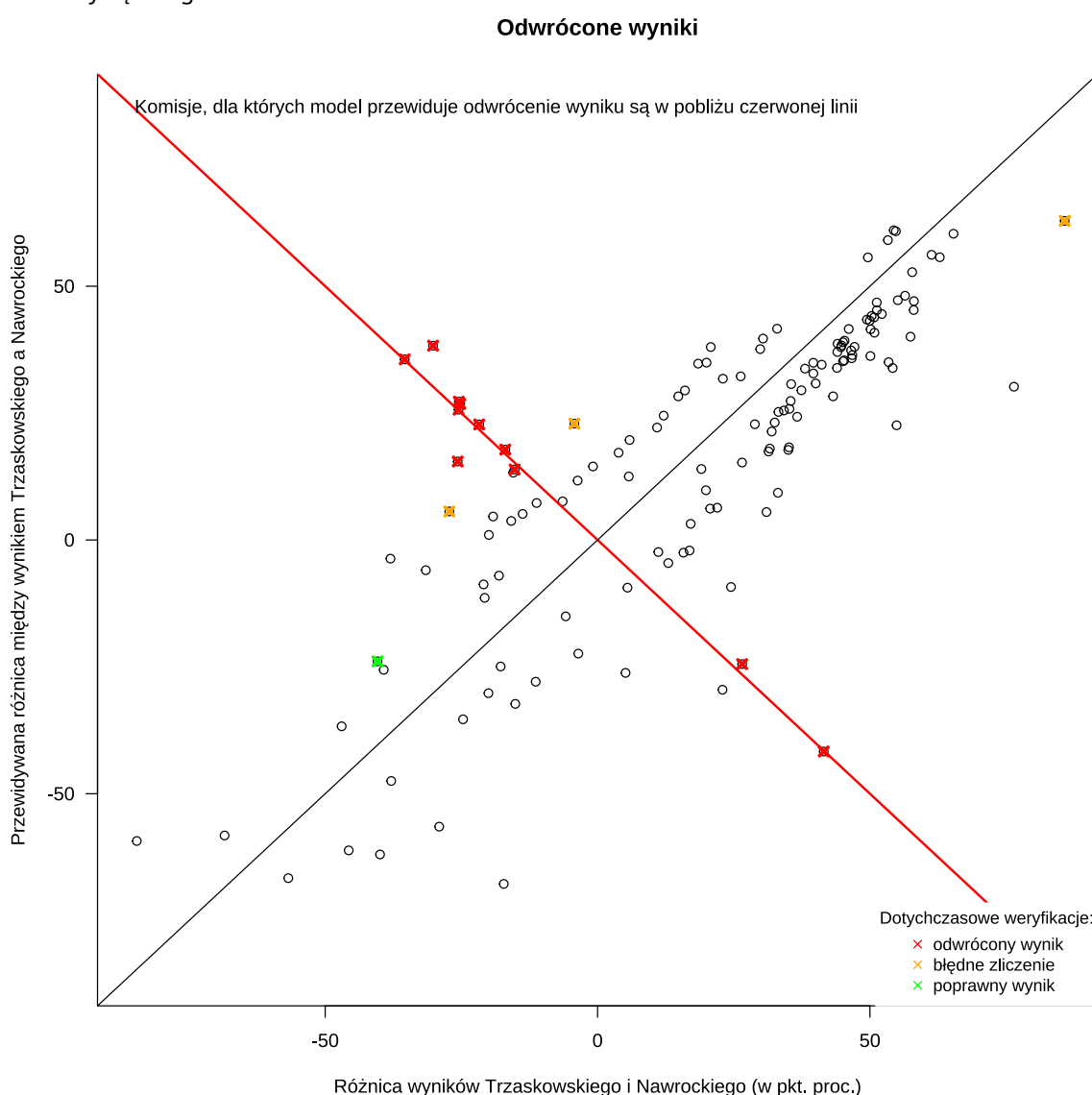
Podsumowując, powyższe dane nie pokazują występowania istotnych anomalii wyników wyborczych, które mogłyby mieć wpływ na końcowy wynik wyborów. Z kilkoma wyjątkami, które powinny być szczegółowo wyjaśnione, rozkład wyników wykazuje prawidłowe zależności, a wyjątki te mają niewielkie znaczenie.

Wątpliwości podnoszone w przestrzeni publicznej

W debacie publicznej pojawiło się wiele opracowań, które analizują wyniki wyborów i przedstawiają na ich podstawie konkretne hipotezy. W przypadku powtarzających się problemów będziemy je omawiali bez odnoszenia się do autorów konkretnych opracowań. Problem lokalnych anomalii, będący podstawą głośnego medialnie opracowania dr Krzysztofa Kontka, będzie omówiony szerzej, także ze względu na szereg jego metodologicznych aspektów.

Odwrócone wyniki

Dotychczas stwierdzono odwrócenie wyniku w kilku komisjach. Przypadki takie wskazywane są przez nasz model, co pokazuje poniższy wykres. Odwrócenie wyników jest szczególnie ewidentne w przypadku, gdy pomiędzy kandydatami jest wyraźna różnica. Gdy wynik jest wyrównany, nie może ustalić tego z jakąś pewnością, lecz jednocześnie oznacza to, że strata/zyski kandydatów na skutek takiej zamiany są marginalne.



Potwierdzone przypadki odwróconych wyników dostarczają też dodatkowej informacji. We wszystkich tych przypadkach same głosy zostały policzone jednoznacznie poprawnie. Nie podważa to znaczenia

przypadków, gdzie doszło do nieprawidłowego przypisania głosów. Lecz jednocześnie sugeruje, że to poprawne przeliczenie głosów jest regułą.

Wśród wykrytych przez model odstępstw są też przypadki trzech komisji, w których co prawda nie doszło do odwrócenia wyniku, ale błędnie zaklasyfikowano większe liczby głosów. Nie ma natomiast komisji nr 53 z Katowic, w której wykryto omyłkowe przypisanie do wyniku obu kandydatów również trzynastu głosów nieważnych. Nie ma też komisji nr 20 z Zabrze, gdzie podczas ponownego liczenia wśród głosów oddanych na kandydata PiS znaleziono jeden głos, który został oddany na Trzaskowskiego. Jako wynik odstający zidentyfikowaliśmy również komisję nr 10 z Tarnowa, gdzie jednak po przeliczeniu okazało się, że głosy zostały dobrze policzone. Omówione tu przypadki pokazują, że model dobrze identyfikuje przypadki odwrócenia wyniku, a także większych przesunięć. Ewentualne bardzo małe zmiany są trudne do wykrycia, ale też część zidentyfikowanych odstających wyników jest zupełnie naturalna.

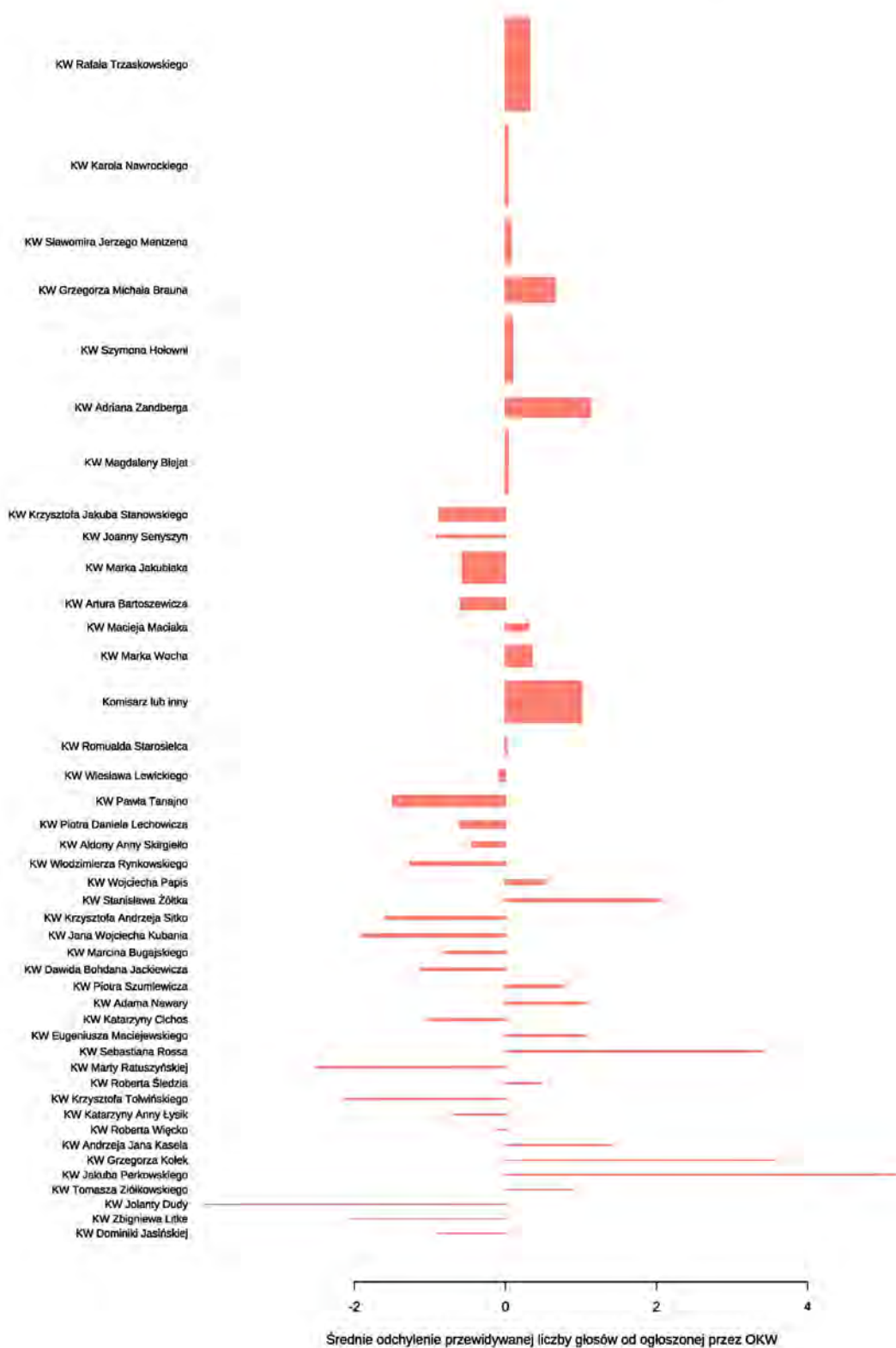
Problem odwróconych wpisów do protokołu jest ważnym wnioskiem z debaty wokół tegorocznych wyborów. Bez wątplenia ten newralgiczny punkt procedury wyborczej wymaga większej uwagi i systemowej kontroli jeszcze przed upublicznieniem podliczonych wyników. Miejmy nadzieję, że problem ten zostanie rozwiązany przez Krajowe Biuro Wyborcze przed kolejnymi wyborami. Jednocześnie nasze wyniki wskazują, że obawy o to, że w znacznie większej liczbie komisji mogło dojść do odwrócenia wyników, nie znajdują podstaw w danych.

Składy komisji wyborczych

Jedną z hipotez wywołujących szczególne emocje była sprawa znaczenia dla wyniku wyborów obsady komisji. Wątpliwości zostały tu spotęgowane przez fakt pojawiania się szeregu komitetów, które po zarejestrowaniu uzyskały prawo wskazywania członków obwodowych komisji wyborczych, lecz nie zarejestrowały już potem kandydatów w wyborach⁴. Znaczenie obsady komisji przetestowaliśmy w dwóch krokach. W pierwszym sprawdziliśmy, jak wygląda różnica między raportowaną liczbą głosów a tą przewidywaną przez nasz model w zależności od tego, który z komitetów zgłosił osobę, która została przewodniczącą/-ym komisji. Znaczenie tego czynnika przedstawiliśmy na wykresie, na którym jednocześnie pokazano, w jakiej części komisji przewodniczącymi byli przedstawiciele danego komitetu (szerokość słupka jest proporcjonalna do liczby nominowanych przewodniczących). W pierwszej kolejności pokazaliśmy komitety kandydatów startujących w I turze (według wielkości poparcia), następnie przewodniczących zgłaszanych przez komisarzy wyborczych lub inne podmioty, następnie zaś komitety bez zarejestrowanego kandydata.

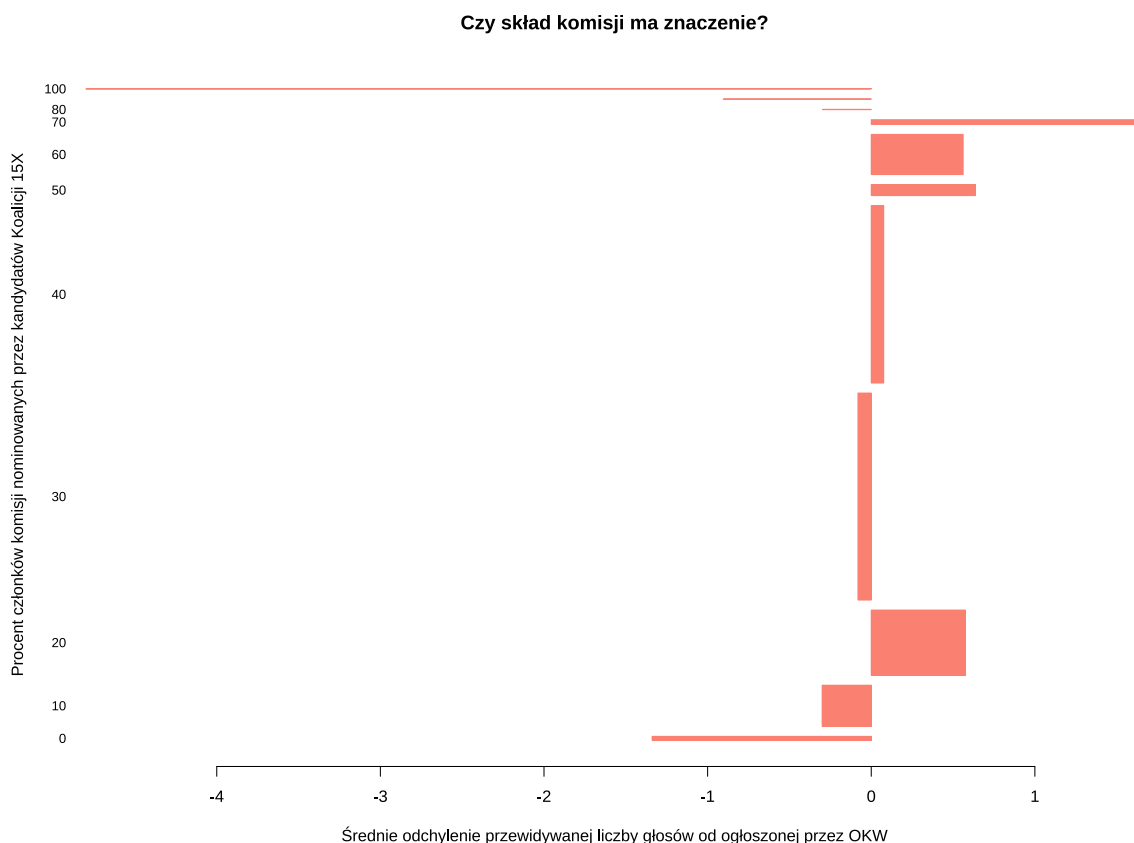
4 Por. A. Porębski, J. Figura, *Przeanalizowaliśmy ponad 250 *tysięcy* członków obwodowych komisji wyborczych*, <https://www.linkedin.com/feed/update/urn:li:activity:7340433718559322113/>.

Czy to kto nominował przewodniczących komisji ma znaczenie?



Taka analiza pokazuje, że wpływ przynależności przewodniczącego jest śladowy (w granicach średnio 1–2 głosy na komisję). Takie odchylenia są zupełnie naturalne i wynikają z niedokładności samego modelu predykcyjnego. Nie sposób tu dostrzec żadnych niepokojących zjawisk – wszystkie odchylenia mieszczą się w granicach logiki.

W drugim kroku sprawdziliśmy, jak wyglądają odchylenia od modelu w zależności od tego, jaki procent członków komisji to osoby zgłoszone przez komitety kandydatów wchodzących w skład „Koalicji 15 Października” (w pierwotnym składzie, czyli z Adrianem Zanbergiem). Efekty takiej analizy pokazuje wykres:



Także tutaj nie widać istotnej, systematycznej różnicy, w szczególności na niekorzyść przegranego kandydata.

Podobnie jak w przypadku odwróconych wyników, problem obsady komisji i roli komitetów niezarejestrowanych kandydatów warty jest rozwiązania po uważnym namyśle.

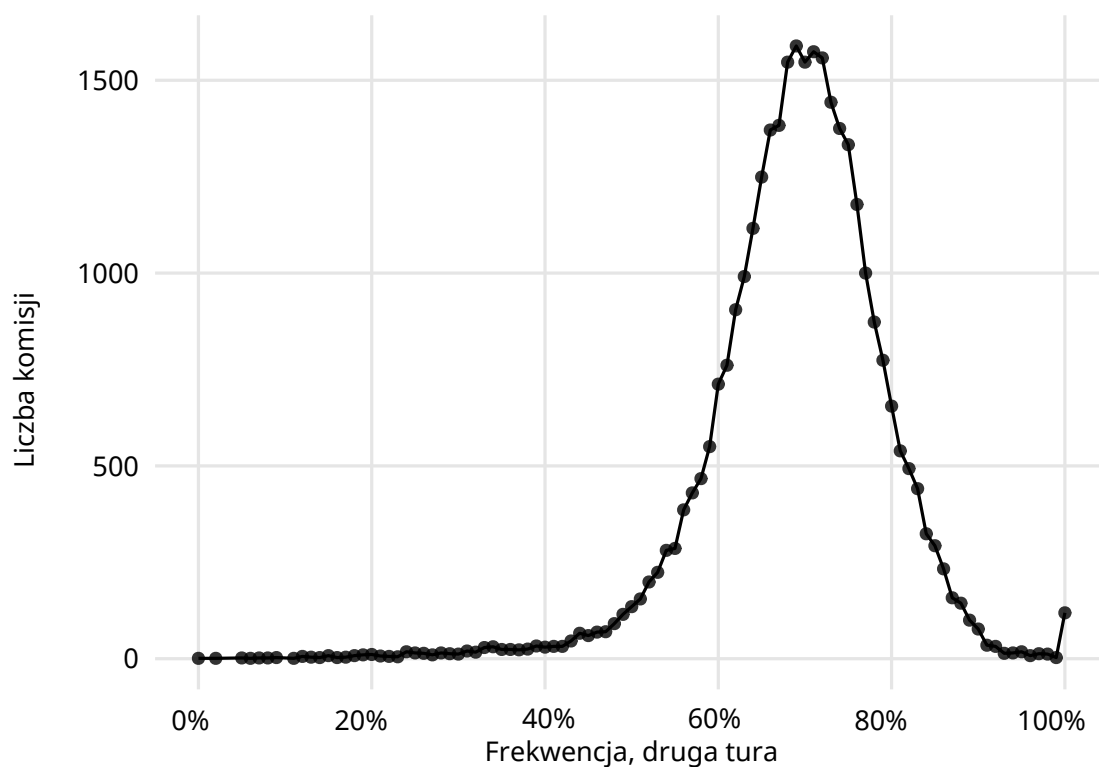
Dosypywanie głosów i nienaturalne wzrosty frekwencji

Część obaw związanych z potencjalnymi nieprawidłowościami wyborczymi dotyczy „dosypywania głosów” i sztucznego zawyżania frekwencji. Warto w związku z tym w pierwszej kolejności przeanalizować, jak zmieniała się frekwencja w komisjach, w których wygrywali poszczególni kandydaci.

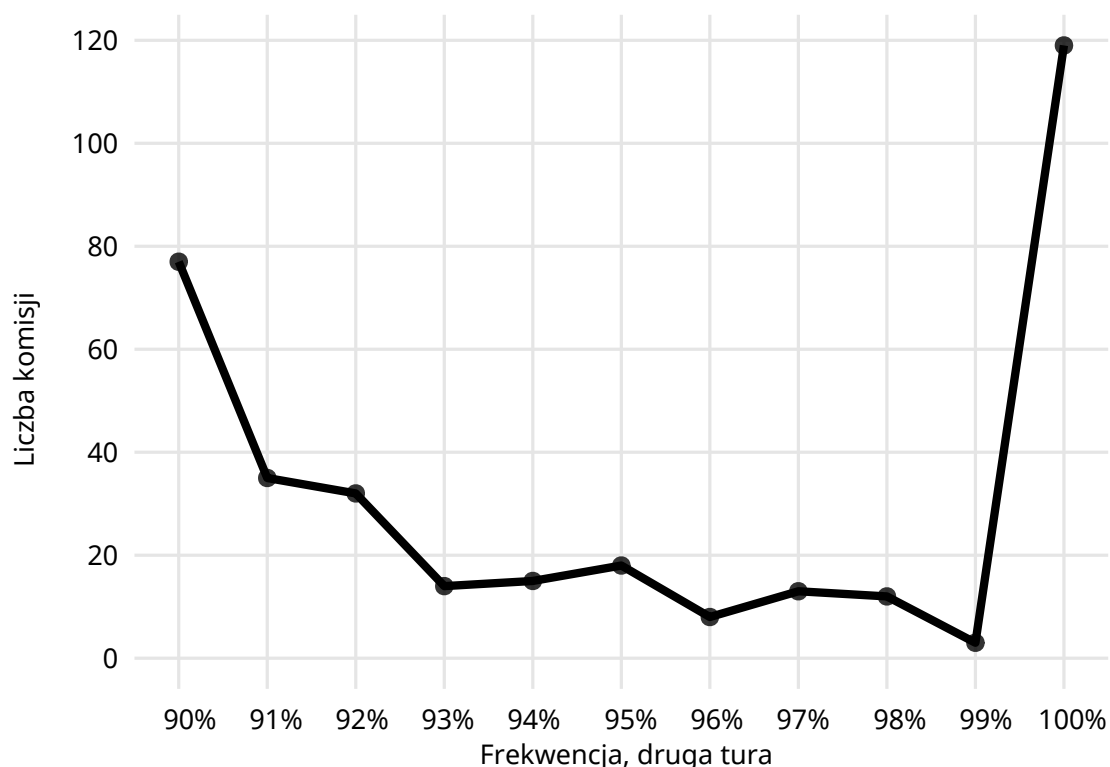
Nie obserwujemy znaczącego przyrostu frekwencji między I a II turą w komisjach, w których wygrał Nawrocki. Wręcz przeciwnie, nieco większe wzrosty frekwencji miały miejsce w dużych miastach, a znacznie większe w komisjach zagranicznych, gdzie Trzaskowski miał lepsze wyniki. W komisjach

gdzie w II turze wygrywał Trzaskowski, frekwencja wzrosła średnio o 8,2% a w tych gdzie wygrywał Nawrocki o 6,7%.

Według autorów niektórych protestów wyborczych, na wykresie prezentującym rozkład frekwencji w II turze widać tzw. „ruski ogon”, czyli nadwyżkę komisji z bardzo wysoką frekwencją, bliską 100%. Ma być to „sygnałem możliwych korekt kart lub innych anomalii w procesie wyborczym”. Pokazuje to poniższy wykres przedstawiający, w ilu komisjach odnotowano daną frekwencję, po zaokrągleniu do jednego procentu. W większości komisji zagłosowało od 60% do 80% uprawnionych.



Przybliżmy wykres, skupiając się na jego prawym ogonie (frekwencja co najmniej 90%).

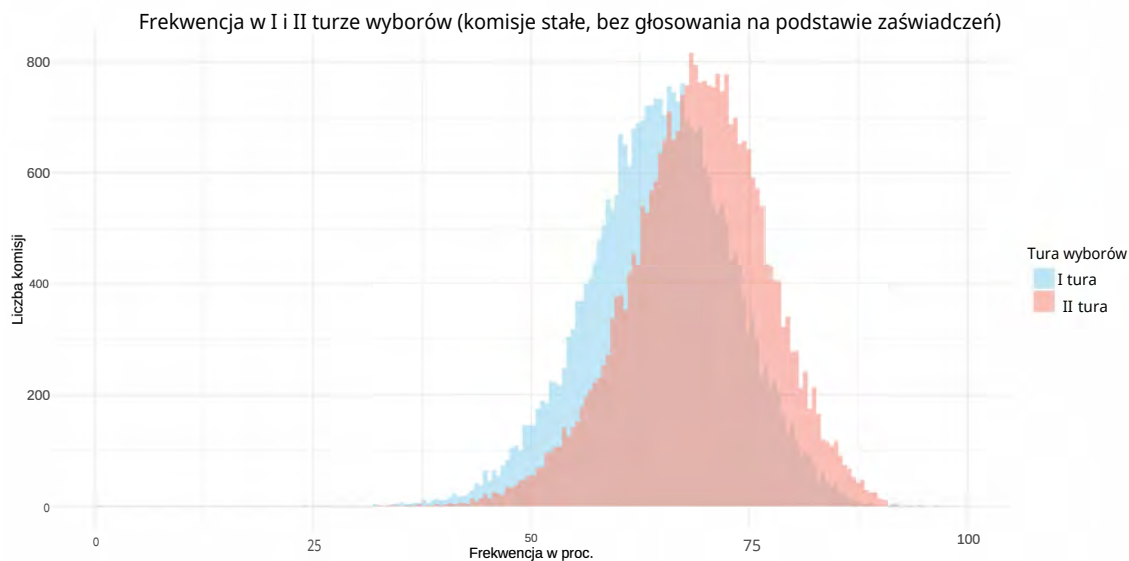


Rzeczywiście, wydaje się, że jest pewna nadreprezentacja komisji z frekwencją 100% po zaokrągleniu (liczba komisji dla 99% odpowiada tempu spadku). Jest 119 takich komisji.

Zacznijmy od tego, że w ponad połowie z nich frekwencja w I turze również była bardzo wysoka (powyżej 90%). Po drugie, 63% z nich to komisje z liczbą uprawnionych poniżej 100. Po trzecie, spójrzmy, co to za komisje: 64 z nich to zakłady lecznicze, 25 to domy pomocy społecznej, 15 to komisje zagraniczne, a tylko 10 to komisje stałe (plus pojedyncze przypadki: zakłady karne, dom studencki, statek).

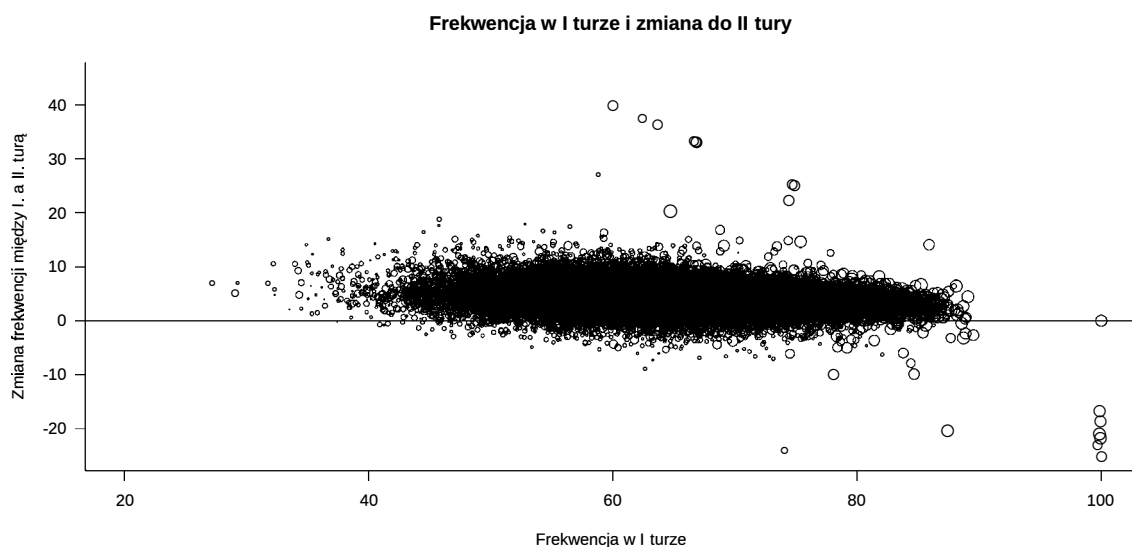
Oczywiście nie ma w tym nic dziwnego, że w zakładach leczniczych i domach pomocy społecznej frekwencja może być bardzo wysoka (powodów jest wiele, między innymi wszystkie osoby są często na miejscu). W przypadku komisji zagranicznych wyborcy specjalnie muszą się zapisać, więc wysoka frekwencja też jest naturalna (we wszystkich tych komisjach w I turze też taka była). Zostało 10 stałych komisji, ale zauważmy, że tylko część komisji z bardzo wysoką frekwencją można uznać za „nadmiarowe”. Poza tym, gdyby wybory odbyły się tylko w tych „podejrzanych” stałych komisjach, Trzaskowski otrzymałby około 62% głosów (a 68% we wszystkich komisjach z frekwencją co najmniej 90%).

Rozkład frekwencji w I i II turze wyborów w komisjach stałych (bez uwzględnienia osób głosujących na podstawie zaświadczenia) pokazuje poniższy wykres. Oba rozkłady mają charakter normalny i nie występują w nich istotne odchylenia.



Dane te wskazują, że jest mało prawdopodobne, aby występowały fałszerstwa polegające na dosypywaniu głosów.

Jednocześnie jest kilkanaście przypadków komisji, które wymagają większego sprawdzenia. Są to komisje stałe, w których frekwencja w I turze wynosiła prawie 100%. A także takie, w których doszło do znaczących zmian frekwencji, co pokazuje poniższy wykres. Należy jednocześnie zaznaczyć, że przynajmniej część z tych nietypowych przypadków to komisje, które błędnie podały liczbę osób głosujących w oparciu o zaświadczenia.

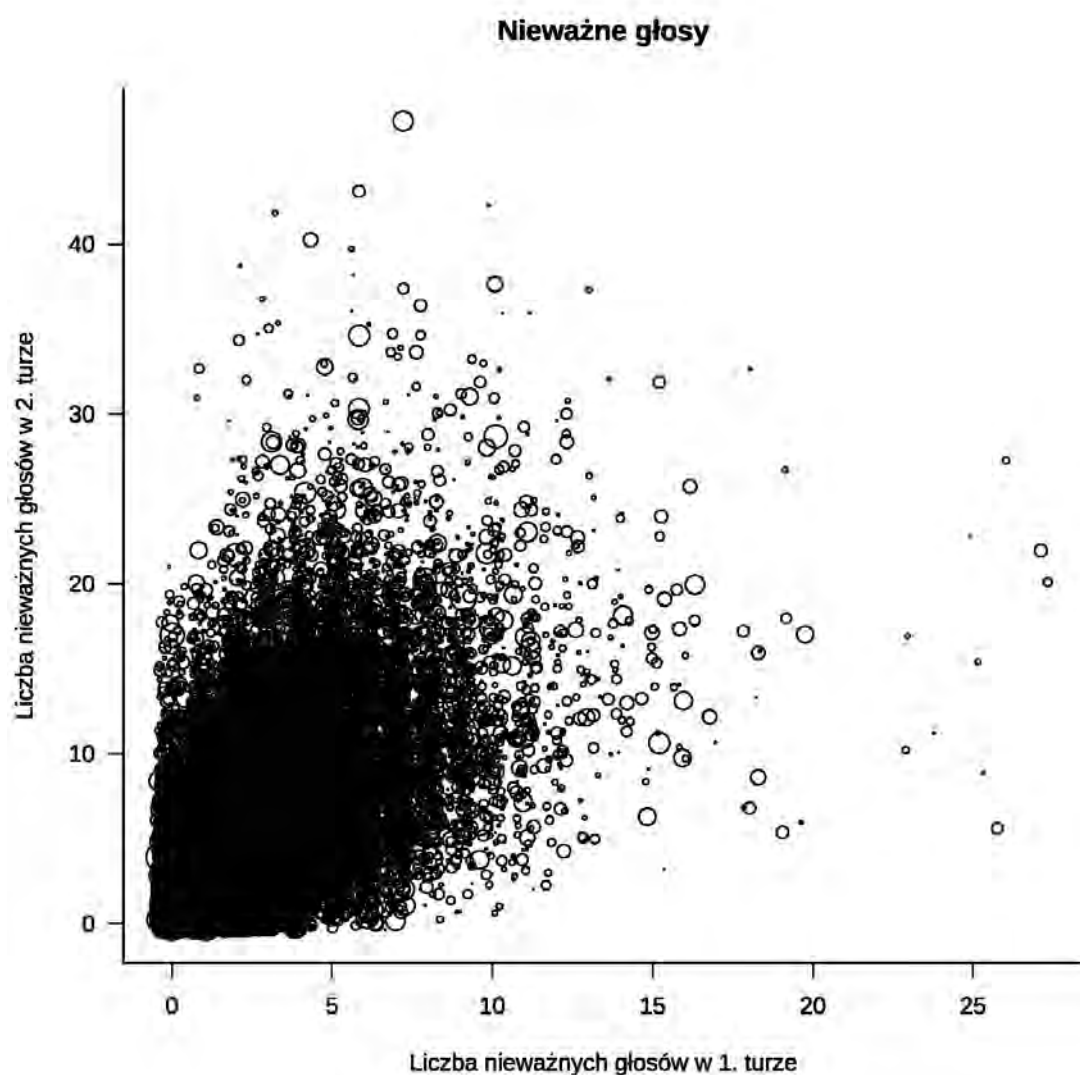


Unieważnianie głosów

W II turze wyborów było prawie 190 tysięcy głosów nieważnych (189 294), dodatkowo 1423 nieważne karty do głosowania (bez pieczęci). Liczby te nie odbiegają od wielkości z poprzednich wyborów. W II turze w 2020 roku było prawie 178 tysięcy głosów nieważnych (przy niższej frekwencji), a liczba kart nieważnych była wyższa i wynosiła 2269. Z kolei w roku 2015 głosów nieważnych było aż 250 tysięcy. Tym samym w obecnych wyborach, podobnie jak pięć lat temu, głosy nieważne stanowiły około

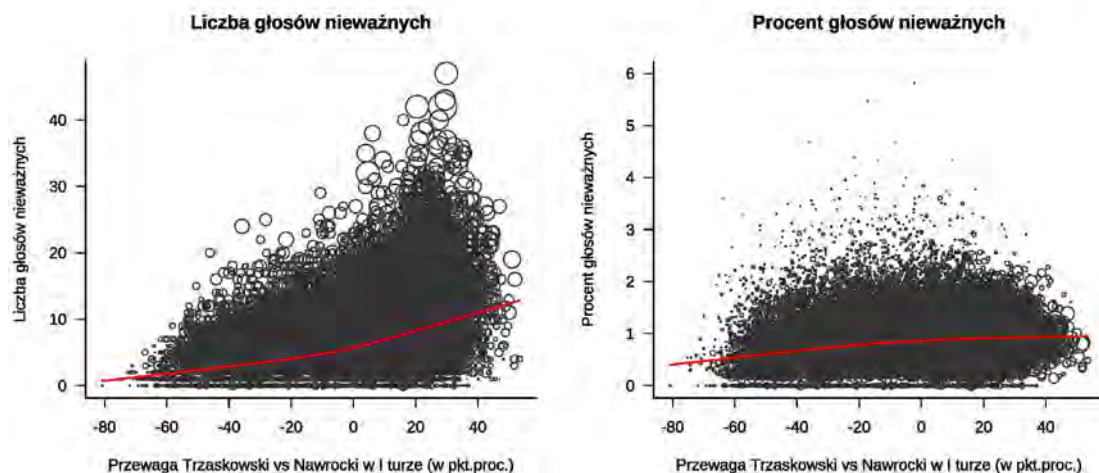
0,9% wszystkich głosów. W 2015 głosów nieważnych było 1,5%. Teza o wzroście liczby nieważnych głosów w stosunku do poprzednich wyborów nie znajduje potwierdzenia.

Również na poziomie głosów nieważnych w poszczególnych komisjach – co przedstawia poniższy wykres – możemy zobaczyć, że brak znacząco odstających przypadków i stosunkowo niewiele jest komisji, w których głosów nieważnych było w II turze więcej niż 20. Przy tak małych liczbach analizy procentowego wzrostu liczby głosów nieważnych, które są dla niektórych argumentem do składania protestów wyborczych, nie mają sensu. Obserwowany tam wyższy wzrost głosów nieważnych w większych komisjach, gdzie wygrywał Trzaskowski, jest prostym efektem niskiej bazy i tego, że te komisje są większe.



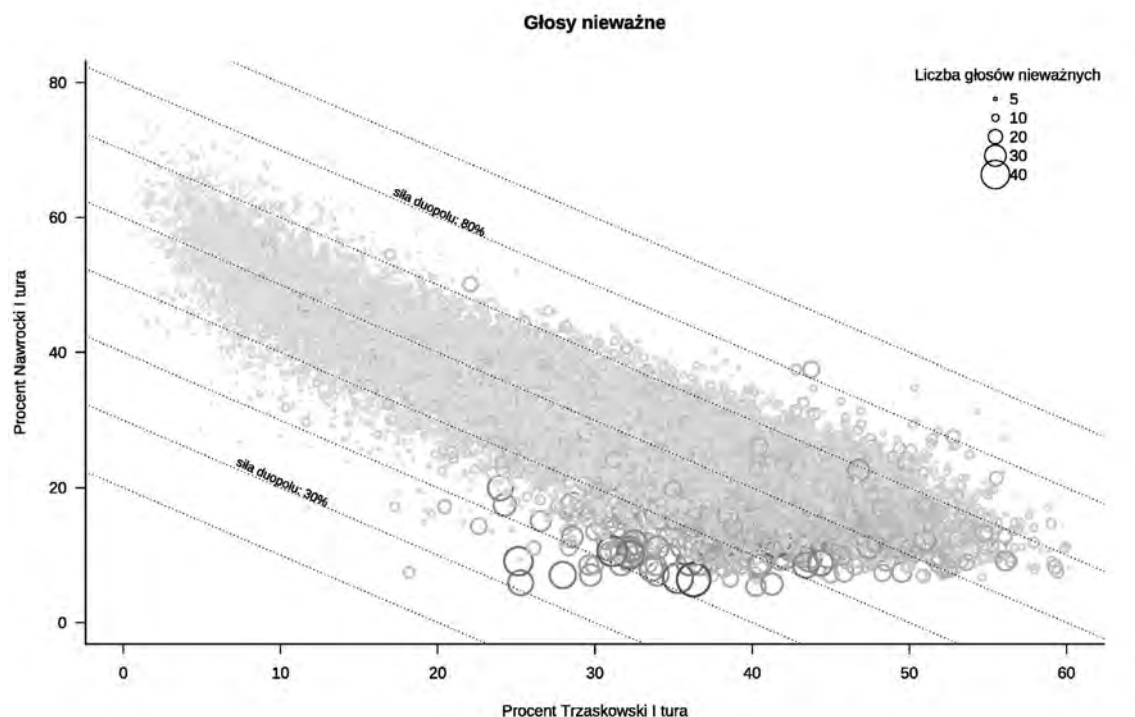
Nieco więcej anomalii dostrzeżemy, przyglądając się procentowemu udziałowi głosów nieważnych. W niektórych przypadkach głosów nieważnych jest bardzo dużo – nawet ponad 20%. Jednak analiza 405 komisji, w których głosów nieważnych było powyżej 5%, pokazuje że są to praktycznie wyłącznie domy pomocy społecznej, szpitale i hospicja. W 1684 komisjach, w których udział głosów nieważnych przekraczał 2%, było łącznie niecałe 9 tysięcy nieważnych głosów. Tym samym zjawisko to ma niewielką skalę i nie odbiega od wartości z wcześniejszych wyborów prezydenckich.

Głosy nieważne a przewaga z I. tury



Analiza rozkładu liczby głosów nieważnych pokazuje, że ich absolutna liczba jest wyższa w komisjach, w których wyższe poparcie uzyskał Trzaskowski. Wynika to jednak przede wszystkim z faktu, że są to generalnie większe komisje – z wyższą liczbą wszystkich oddanych głosów. Komisje miejskie są z reguły większe i większa jest też w nich frekwencja. Po przeliczeniu liczby głosów nieważnych na procent wszystkich głosów w komisji zależność ta przestaje być widoczna: odsetek głosów nieważnych nie różni się istotnie w zależności od dominującego kandydata. Oznacza to, że wyższa liczba głosów nieważnych w „silnych komisjach Trzaskowskiego” jest pochodną ich wielkości, a nie wyrazem szczególnych nieprawidłowości czy specyficznych zachowań wyborców w tych lokalizacjach.

Więcej głosów nieważnych jest także w komisjach, w których dwaj główni kandydaci (Trzaskowski i Nawrocki) zdobywali łącznie mniej głosów, a więc tam, gdzie siła duopolu PO-PiS jest słabsza. W komisjach tych jest więcej wyborców, którzy nie mieli swojej reprezentacji w II turze wyborów. W dużych komisjach, w których w I turze Trzaskowski zdobywał 25–40% głosów, a Nawrocki miał niewielkie poparcie, więcej było osób, które nie wybierały w II turze żadnego z nich. Potwierdzają to model wyjaśniający liczbę głosów nieważnych, który stworzyliśmy. Analiza pokazuje silną zależność od liczby głosów oddanych na innych kandydatów, szczególnie Jakubiaka, Brauna, Hołownię i Zandberga.



Obserwowane liczby głosów nieważnych nie odstają od normy i wydają się naturalnym efektem decyzji wyborców, którzy nie czuli się reprezentowani przez kandydatów, którzy dostali się do II tury wyborów.

Lokalne anomalie

Bardzo dużo uwagi mediów przyciągnęła analiza, jaką przeprowadził dr Krzysztof Kontek. Stawia on sobie za cel szukanie anomalii na poziomie lokalnym i opiera się na założeniu, że w geograficznie bliskich komisjach wyborcy głosują podobnie, a przynajmniej brak komisji, które odstawałyby od reszty w sposób naturalny. Jeśli któraś z komisji nie jest podobna pod względem pewnych przyjętych charakterystyk, jest kwalifikowana jako anomalia. Już samo to założenie wydaje się mocno problematyczne, co pokazujemy niżej.

Autor zidentyfikował aż 3679 komisji, w których wystąpiły anomalie (a w mniej konserwatywnej wersji nawet 5453). Już te liczby budzą wątpliwości, bo ta druga stanowi 17% wszystkich analizowanych komisji. Metody detekcji anomalii mają swoje ograniczenia. Opierają się na parametrach, które wyznaczamy na podstawie danych⁵, ale skoro są w nich anomalie, to jaką mamy pewność, że te parametry są prawidłowe? Takie założenie można przyjąć, jeśli anomalie są czymś wyjątkowym – wtedy powinny mieć niewielki wpływ na szacowane parametry. Jeśli anomalie są typowe, nie jesteśmy w stanie wiarygodnie oszacować parametrów, które mają posłużyć do wyznaczenia anomalii... Innymi słowy, wskazując tak dużą ich liczbę, zaprzeczamy założeniu, na którym opiera się detekcja anomalii. Problemów z analizą dr. Kontka jest jednak znacznie więcej, poniżej podamy kilka najważniejszych.

Algorytm dr. Kontka wyznacza jedynie anomalie na korzyść Nawrockiego. Nie ma żadnej wzmianki o tym, jakie wyniki otrzymamy, gdyby przeprowadzić podobną analizę na korzyść Trzaskowskiego. Jest to niezbędne z dwóch powodów. Po pierwsze, w ten sposób można ocenić, czy potencjalne pomyłki są symetryczne, tzn. zaszyły na korzyść jednego i drugiego kandydata (i nawet po korekcie nie zmieniałyby ogólnego wyniku). Po drugie, pomogłoby to w ustaleniu, czy przyjęte przez autora progi detekcji anomalii (do których odnieśliśmy się w dalszej części) nie są zbyt liberalne. Jeśli otrzymalibyśmy porównywalną liczbę anomalii, ale na korzyść Nawrockiego, świadczyłoby to raczej nie tyle o symetrii w pomyłkach, ale o tym, że przyjęto złe progi.

Przekonanie, że terytorialna bliskość komisji jest podstawą do przypuszczenia, iż wzory głosowania w tych komisjach są podobne, jest sprzeczna z podstawową wiedzą o zachowaniach wyborczych polskiego społeczeństwa. Takie założenie ignoruje dwa fundamentalne fakty – w miastach sąsiadują ze sobą obszary o różnej zabudowie (stare bloki, nowe osiedla wielorodzinne, nowa zabudowa jednorodzinna, wsie włączone w obręb miasta), gdzie mieszkają ludzie o różnej zamożności, wieku, wykształceniu i zakorzenieniu. Każdy z tych czynników jest powszechnie uznaną składową polskich podziałów politycznych. Można to sprawdzić, np. porównując wyniki wyborów parlamentarnych 2023 w dwóch komisjach mających siedzibę w jednej szkole na krakowskich Skotnikach. Komisja nr 166 obejmuje starą zabudowę i Koalicja Obywatelska wygrywa z Prawem i Sprawiedliwością tylko 347 do 332. W komisji nr 167, obejmującej nowe osiedla „szeregówek”, KO zdobywa już prawie dwukrotnie więcej głosów (307 do 161). Wzór ten potwierdził się w tegorocznych wyborach i przypuszczalnie w analizie dr. Krzysztofa Kontka komisja nr 166 została zakwalifikowana jako „nienaturalny” wzrost poparcia dla Karola Nawrockiego.

5 Por. P. Szulc, *Model detekcji anomalii wyborczych w wyborach prezydenckich*, <https://danetyka.com/model-detekcji-anomalii/>.

Jeszcze większe różnice występują w gminach miejsko-wiejskich pomiędzy ich składowymi, jak też na przedmieściach metropolii. W podkrakowskiej gminie Liszki w miejscach najbliższych miastu, gdzie od lat osiedlają się nowi mieszkańcy – w Kryspinowie i Budzowie – dwaj główni kandydaci toczą wyrównany pojedynek. Na drugim końcu gminy, gdzie dopiero zaczyna się nowe osadnictwo – w Kaszowie i Czułowie – Rafał Trzaskowski przegrywa 1:2.

Obiecujące wydawało się wykorzystanie kodu pocztowego jako identyfikatora terytorialnej bliskości. Nie jest to podejście stosowane w dotychczasowych analizach zachowań wyborczych. Jednak całkiem pobieżny test pokazał, dlaczego to jest ślepa ścieżka. Podział według kodów pocztowych jest bardzo chaotyczny. W kilkuset przypadkach obejmuje terytorium dwóch gmin o bardzo różnym charakterze. Na przykład pod jednym kodem są Gorlice – średniej wielkości małopolskie miasto z wyrównanymi siłami dwóch głównych ugrupowań – i trzy wsie z sąsiadującej gminy wiejskiej, ze zdecydowaną przewagą PiS. W analizie dr. Kontka wyniki w tych trzech wsiach najpewniej zostały uznane za „nienaturalne”, choć są bardzo zbliżone do pozostałych wsi w tej samej gminie, mających jednak już inny kod pocztowy.

Wielkość obszaru objętego jednym kodem różni się w nietypowy sposób – w średnich miastach Kongresówki obszary takie są dwu-trzykrotnie większe od większych miast i od gmin wiejskich. Jednak nawet gdyby przyjąć, że jednostki podziału terytorialnego są jednolite pod względem zachowań wyborczych, zastosowana metoda nie jest przekonująca, co wyjaśniamy w ramce.

Okiem statystyka

Grupy bliskich komisji zostały tak zdefiniowane, że ich wielkość w większości przypadków zawiera się w przedziale 10–16 (są to więc małe grupy). Dla każdej komisji jest liczona pewna cecha, np. „poparcie dla Nawrockiego”. Jeśli to poparcie zwiększyło się „nadmiernie” (w porównaniu do pobliskich komisjach), jest to anomalia.

Według autora, nadmierne poparcie jest wtedy, kiedy po pewnej standaryzacji jest większe od ustalonego progu: 2, 2,5 lub 3 (ten ostatni próg autor traktuje jako „konserwatywne” podejście). Standaryzacja polega na tym, że od wyniku w danej komisji odejmuje się medianę z grupy, a następnie dzieli przez rozrzut w tej grupie, liczony jako MAD (mediana odchyłeń bezwzględnych). Jest to alternatywa dla odchylenia standardowego, stosowana czasem dla niesymetrycznych rozkładów. Procedura przypomina zastosowanie reguły trzech sigm: w standardowym rozkładzie normalnym mamy jedynie około 0,13% wartości, które są większe od 3, przez co w niektórych zastosowaniach można takie wartości potraktować jako anomalie.

Pierwszy problem polega na tym, że ta reguła nie działa dla MAD. Należy przynajmniej uwzględnić, że np. dla rozkładu normalnego MAD jest około 1,5 razy mniejsze od SD⁶. Innymi słowy, rzekomo konserwatywne podejście z progiem 3 wcale takie nie jest, bo de facto jest on równy 2. Takich rzekomo odstających obserwacji jest w rozkładzie normalnym około 2%, co daje prawie 700 komisji uznanych za anomalie, mimo że takimi nie są. Co więcej, w rzeczywistości rozkład wyników na pewno nie jest normalny, a np. w rozkładzie lognormalnym takich „odstających” komisji (typowanych algorytmem autora) będzie 21,5%, czyli prawie 7000.

W pracy nie pojawia się analiza rozkładu rozważanych przez autora cech (przy założeniu braku nieprawidłowości), a jest to niezbędne do wyznaczenia progów.

Jeśli spojrzymy na ten problem inaczej, mamy tu do czynienia z tzw. **problemem porównań wielokrotnych**⁷. Im więcej komisji badamy przy pomocy zaproponowanej reguły, tym więcej będzie kwalifikowanych jako anomalie, mimo że takimi nie są (wyniki fałszywie dodatnie). Rozwiązaniem tego problemu są odpowiednie korekty – i nieuwzględnienie ich jest błędem. Na dodatek autor popełnia go podwójnie. Po pierwsze, dokonuje ponad 30 tysięcy porównań (każda komisja). Po drugie, sprawdza niezależnie cztery cechy mające świadczyć o anomalii i ostatecznie raportuje wszystkie przypadki, dla których choć jedna z nich miałaby świadczyć o anomalii. Jest to główny powód, dla którego otrzymał tak dużą ich liczbę.

Jedną z cech, jaką bada autor, jest nazwana „flip”. Nie ma ona nic wspólnego z wyżej podaną standaryzacją i progami, ale autor za anomalię uznaje każdy przypadek, gdzie „Nawrocki wygrywa lokalnie, mimo że mediana wyników w grupie wskazuje przewagę Trzaskowskiego”. Zobrazujmy to przykładem: założmy, że procenty poparcia dla Nawrockiego w danej grupie wynoszą: 45, 46, 47, 48, 49, 51, 52, 53, 54. Mediana wynosi 49%, więc „mediana wyników w grupie wskazuje przewagę Trzaskowskiego”, a zatem te cztery (!) komisje, gdzie Nawrockich otrzymał ponad 50%, to według autora anomalie.

Gdyby za anomalie traktować komisje, dla których np. wartości co najmniej trzech cech są odstające, wtedy uwzględnienie powyższej miałoby sens – natomiast wystarczy jedna. Co więcej, ta opisywana jest odpowiedzialna za ponad połowę (!) wskazań.

Podsumowując, praca autora zawiera fundamentalne błędy i nie może stanowić argumentu za nieprawidłowościami w liczeniu głosów. Przeciwnie: podane wyniki (przy tak liberalnej definicji anomalii) są zgodne z tezą, że w zdecydowanej większości komisji podano poprawne wyniki.

Nienaturalne przepływy głosów

Analiza anomalii umieszczona na stronie defoliator.pl⁸ opiera się na tym samym założeniu, które przyjęliśmy w zaprezentowanym powyżej modelu, że wyniki II tury wyborów powinny silnie korelować z wynikami I tury, uwzględniając dodatkowo przepływy głosów. W odróżnieniu jednak od naszych

6 https://en.wikipedia.org/wiki/Median_absolute_deviation#Relation_to_standard_deviation.

7 https://pl.wikipedia.org/wiki/Problem_por%C3%B3wna%C5%84_wielokrotnych.

8 *Matematyczne spojrzenie na II turę wyborów prezydenckich*, <https://defoliator.pl/blog/matematyczne-spojrzenie-na-ii-ture-wyborow-prezydenckich-2025>.

modeli⁹, autor opiera się na szacowanym przepływie głosów z sondażu UCE Research¹⁰. Podobne podejście zostało wykorzystane w raporcie Obywatelskiej Kontroli Wyborów¹¹.

Autor przyjmuje pewien próg, przy którym uznaje, że wyniku II tury nie da się wyjaśnić wynikiem I tury. Na tej podstawie kwalifikuje około 1000 komisji jako anomalie. Co ważne, nie ma w nich symetrii: w 80% z nich potencjalne pomyłki wystąpiły na korzyść Nawrockiego, a w 20% Trzaskowskiego. Podobnie jak w przypadku modelu dr. Kontka, brak uzasadnienia statystycznego dla przyjętego progu. Uznajmy jednak, że – jak twierdzi autor – próg jest na tyle rygorystyczny, że rzeczywiście oficjalnych wyników II tury nie da się wyjaśnić wynikami I tury, tzn. różnice są zbyt duże. Zgodnie ze sztuką, pierwszym wnioskiem, jaki powinno się w takich wypadkach wyciągnąć, jest to, że przyjęto błędne założenia, np. dotyczące szacowanych przepływów.

Autor stwierdza, że nawet jeśli wyniki sondażowe są w jakimś stopniu błędne, to ten błąd nie może być tak duży, by skutkował zaobserwowanymi przez niego różnicami. Nie jest to prawdą. Gdy podaje się tzw. „błąd statystyczny” dla sondaży, zwykle nie uwzględnia się jego potencjalnego obciążenia, czyli systematycznego niedoszacowania lub przeszacowywania (taka sytuacja najprawdopodobniej miała miejsce w przypadku poparcia Grzegorza Brauna, które było niedoszacowane w większości sondaży).

Tego typu obciążenia mogą, po pierwsze, wyjaśniać większość różnic między prognozą autora a oficjalnym wynikiem, jaki również – co ma znacznie poważniejsze konsekwencje – brak symetrii w tych różnicach.

Podobnie jak w modelu dr. Kontka, aby tego typu analizy miały sens, musimy wiedzieć, jakiego rozkładu wyników powinniśmy się spodziewać przy założeniu, że nie doszło do fałszerstw (hipoteza zerowa, H_0). Skąd mamy wiedzieć, że konkretna różnica jest anomalią, jeśli nie wiemy, ile one mogą wynosić, gdy wyniki są zgodne z H_0 ? Może powinny być właśnie takie, jakie otrzymuje autor omawianej analizy (dodatkowo biorąc pod uwagę potencjalne błędy w przyjętych przepływach głosów).

W zaproponowanych przez nas modelach analizujemy, jak takie różnice (tzw. reszty) się zachowują, w szczególności jaki jest ich rozkład przed usunięciem anomalii i po ich usunięciu. Używamy powszechnie przyjętych metod detekcji anomalii: reguły trzech sigm i testów statystycznych (p-wartości). Wiemy, czego powinniśmy się spodziewać przy założeniu, że głosy policzono poprawnie i porównujemy otrzymane wyniki z taką sytuacją.

Ograniczenia zastosowania prawa Benforda w analizie wyników wyborczych

W publicznej debacie na temat uczciwości wyborów często pojawia się odwołanie do prawa Benforda¹² – statystycznej obserwacji, zgodnie z którą w wielu naturalnie występujących zbiorach danych liczbowych, obejmujących wiele rzędów wielkości i pozbawionych sztucznych ograniczeń, rozkład pierwszych cyfr będzie zgodny z charakterystycznym logarytmicznym wzorcem. Cyfra 1 pojawia się jako pierwsza cyfra znacznie częściej (ok. 30% przypadków) niż cyfry kolejne, aż do cyfry 9 (ok. 5%

9 Zaprezentowanego powyżej i tych publikowanych gdzie indziej, w szczególności P. Szulc, *Model detekcji anomalii wyborczych w wyborach prezydenckich*, <https://danetyka.com/model-detekcji-anomalii/>.

10 Wykorzystanie danych sondażowych o przepływach, a nie danych z badania exit-poll wynika z chęci uwzględnienia również informacji o osobach, które nie wzięły udziału w powtórnym głosowaniu.

11 *Zaufanie do wyborów – raporty OKW*, 17 czerwca 2025, <https://ruchkod.pl/zaufanie-do-wyborow-raporty-okw/>.

12 https://pl.wikipedia.org/wiki/Rozk%C5%82ad_Benforda.

przypadków). Zwolennicy tej metody argumentują, że dane sfabrykowane przez człowieka często nie podążają za tym wzorcem, co pozwala używać prawa Benforda jako narzędzia do wykrywania fałszerstw, m.in. w danych finansowych czy wyborczych. W praktyce jednak możliwość zastosowania prawa Benforda jest ściśle uzależniona od spełnienia szeregu warunków dotyczących charakteru i struktury analizowanych danych¹³.

W przypadku wyników wyborów, szczególnie na poziomie obwodowych komisji wyborczych, powyższe warunki nie są spełnione. Po pierwsze, liczba ważnych głosów oddanych w każdej komisji nie jest wynikiem spontanicznego procesu generowania liczb, lecz wypadkową organizacyjnych reguł wyznaczania wielkości komisji oraz ograniczeń prawnych (np. minimalnej i maksymalnej liczby uprawnionych wyborców przypadających na jedną komisję). Prowadzi to do silnego „sklejenia” wartości wokół jednego lub dwóch rzędów wielkości (np. najczęściej 500–1500 głosów), co drastycznie ogranicza rozpiętość analizowanych liczb. W przeciwieństwie do założeń prawa Benforda, rozkład głosów nie jest więc rozciągnięty na kilka rzędów wielkości, lecz skoncentrowany wokół wąskiego przedziału, ze sporadycznymi wyjątkami.

Po drugie, rozkład liczby głosów w komisjach jest kształtowany przez dodatkowe czynniki demograficzne, geograficzne i logistyczne, takie jak frekwencja wyborcza czy liczba wyborców zarejestrowanych w danej jednostce administracyjnej. Te czynniki wprowadzają do rozkładu liczb systematyczne, przewidywalne odchylenia od wzorca Benforda, które nie mają związku z intencjonalnymi manipulacjami czy nieprawidłowościami.

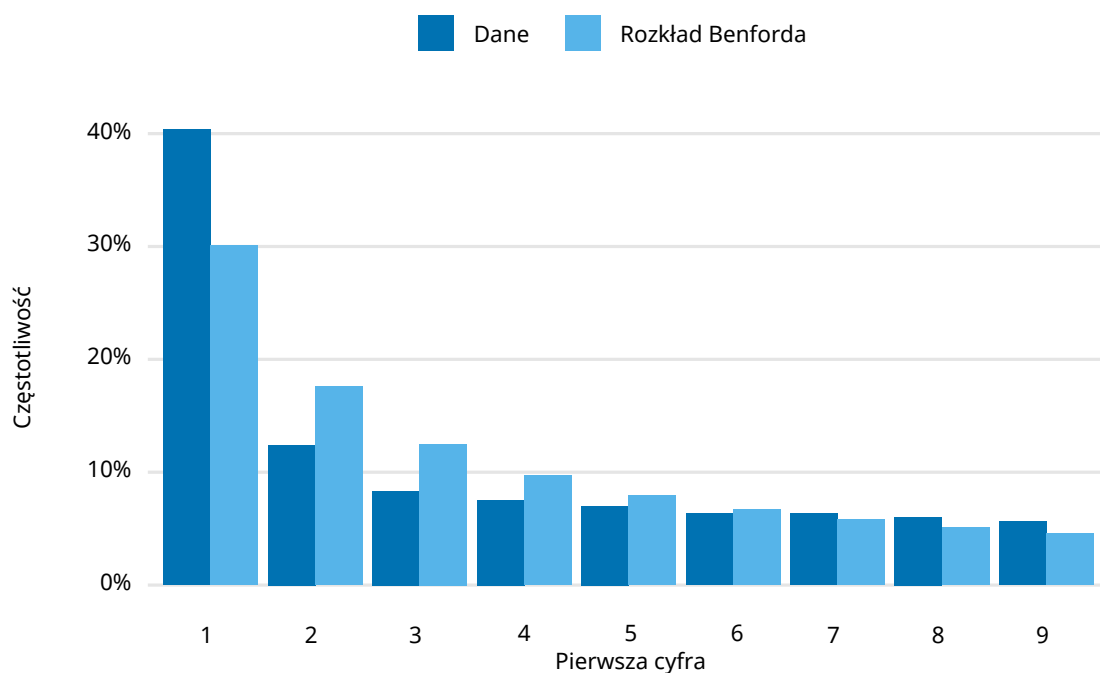
Empiryczna analiza liczby ważnych głosów w poszczególnych komisjach wyborczych dla omawianych danych potwierdza powyższe zależności: zdecydowana większość wartości mieści się w przedziale 100–1500, a ich rozkład jest silnie skupiony wokół mediany około 558 głosów. Brak szerokiego zakresu wartości sprawia, że rozkład pierwszych cyfr jest zdominowany przez cyfry odpowiadające tym przedziałom, co skutkuje istotnymi odchyleniami od rozkładu przewidzianego przez prawo Benforda. Odchylenia te nie mogą być jednak interpretowane jako dowód występowania anomalii czy fałszerstw, lecz jako konsekwencja specyfiki danych wynikającej z konstrukcji procesu wyborczego.

Podsumowując, prawo Benforda nie jest adekwatnym narzędziem do wykrywania nieprawidłowości w danych dotyczących liczby ważnych głosów oddanych w obwodowych komisjach wyborczych, gdyż podstawowe założenia tej metody są w tym przypadku naruszone. Próba zastosowania tego prawa w takim kontekście prowadziłaby do fałszywych alarmów i błędnych interpretacji, maskując rzeczywiste procesy i wzorce generujące dane wyborcze.

Mimo że powyższa argumentacja jest całkowicie wystarczająca, sprawdźmy, jak wygląda rozkład pierwszej cyfry dla liczby uprawnionych (ważne: nie dla liczby głosów) w komisjach, oraz porównajmy go z rozkładem Benforda.

13 J. Deckert, M. Myagkov, P.C. Ordeshook, *Benford's Law and the Detection of Election Fraud*, „Political Analysis” 2011, nr 19, s. 245–268, <https://doi.org/10.1093/pan/mpi014>.

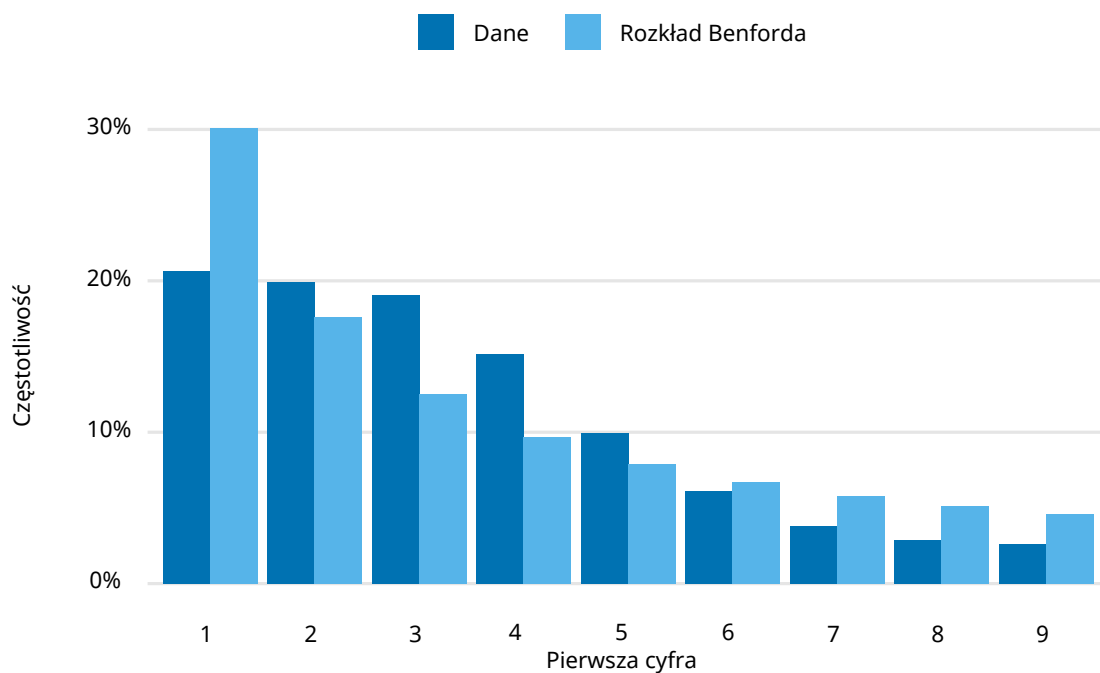
Liczba uprawnionych, druga tura 2025



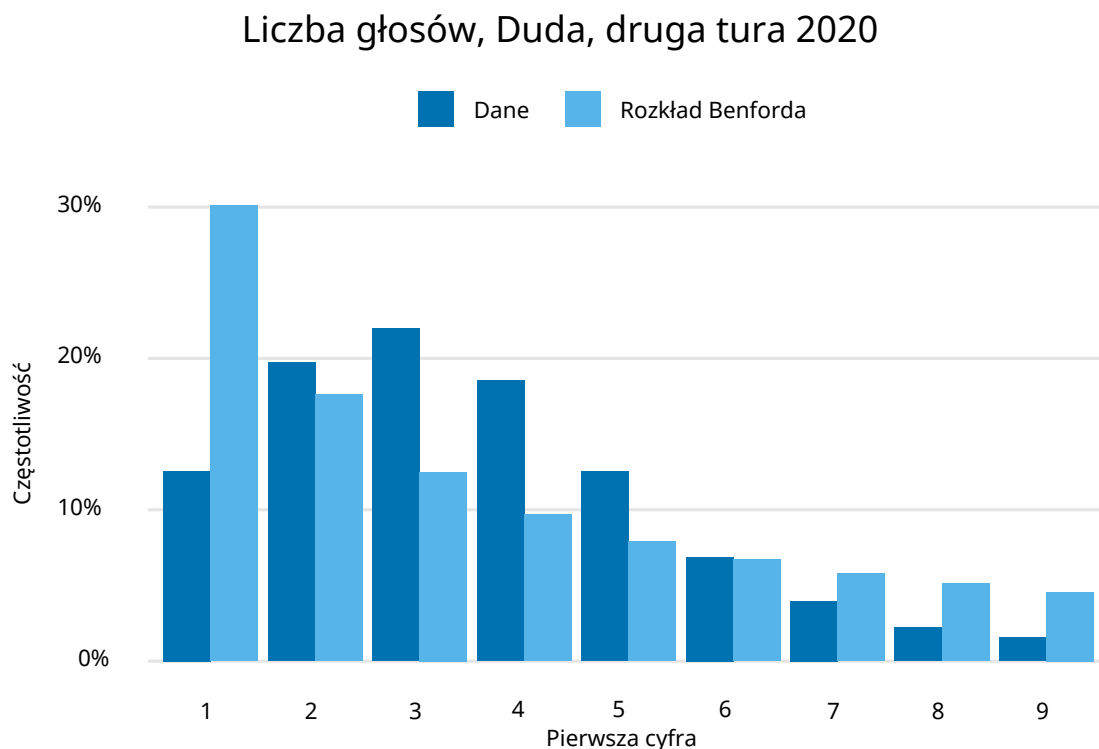
Odstępstwa są duże (szczególnie widoczne w przypadku cyfry 1), a zatem jeśli odstępstwo od prawa Benforda miałyby świadczyć o nieprawidłowościach, właśnie udało nam się pokazać, że sfałszowana została... liczba uprawnionych do głosowania.

Na kolejnym wykresie rozkład pierwszej cyfry dla liczby głosów oddanych na Nawrockiego.

Liczba głosów, Nawrocki, druga tura 2025



Możemy porównać ten wykres z rozkładem wyników dla Dudy w poprzednich wyborach.



Odstępstwo od rozkładu Benforda jest jeszcze większe, w takim razie musielibyśmy uznać, że dokonano tam jeszcze poważniejszych fałszerstw. Co ważne, gdyby były to pojedyncze przypadki oraz założenia prawa Benforda byłyby spełnione, odstępstwo od niego powinno być minimalne (moglibyśmy go nawet nie zauważyć). A różnice są tak wielkie, że fałszerstwa musiałyby mieć miejsce w większości komisji.

Czy da się wytłumaczyć, dlaczego rozkład pierwszych cyfr dla Dudy (granatowe słupki) tak wygląda? Wynika to z tego, że liczby oddanych na niego głosów w poszczególnych komisjach zawierają się prawie wyłącznie w przedziale 100–999 (ten sam rząd wielkości). W takim wypadku rozkład pierwszej cyfry jest po prostu przybliżeniem histogramu dla rozkładu głosów, który dla Dudy jest mniej więcej normalny, z maksimum w przedziale 300–399 (stąd cyfra 3 jest najpopularniejsza).

Podnoszone wątpliwości – podsumowanie

Rozgorączkowanie wynikami zaowocowało szczególnym zainteresowaniem problemem uczciwości wyborów. Zaniepokojeni obywatele i eksperci podejmowali liczne próby przetestowania hipotezy o masowych nieprawidłowościach. Nie wszystkie z nich spełniały standardy naukowe, a część uzyskiwała alarmujące wyniki właśnie za sprawą naruszenia takich standardów. Warto same te próby docenić jako wyrazy obywatelskiej troski. Nie trzeba jednak im dawać wiary tylko dlatego, że były szlachetnie motywowane. Żadne z tych podejść nie dostarcza faktów podważających nasz model. Jednocześnie startowy problem – odwrócenia wyników w trakcie ich raportowania – jest dokładnie pokazywany przez nasze podejście. Najbardziej spektakularne opracowania są – w świetle naszych analiz – przykładem podejścia nieadekwatnego do analizowanego problemu.

Zakończenie

Obie ścieżki naszych analiz – stworzenie własnego modelu oraz weryfikacja opracowań o dużym oddźwięku społecznym – prowadzą do jednego wniosku. W II turze tegorocznych wyborów prezydenckich pojawiły się przekłamania w oficjalnie raportowanych wynikach. Nie jest to zjawisko nowe: ani co do charakteru, ani skali. Co jednak najważniejsze, pomimo wyjątkowo wyrównanej rywalizacji, nic nie wskazuje na to, by te przekłamania doprowadziły do zmiany ostatecznego rezultatu wyborów. Ani ich skala, ani ukierunkowanie nie dają podstaw do podważania rozstrzygnięcia, które w wyniku wyborów zapadło.

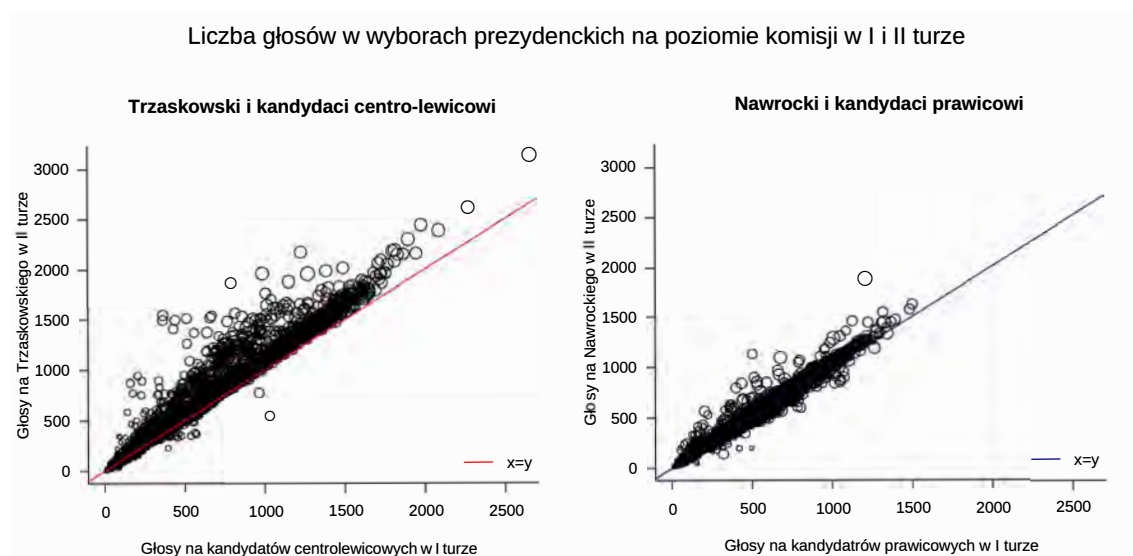
Nie oznacza to jednak w żadnym stopniu, że te przekłamania można lekceważyć. Ich ujawnienie jest ważne dla jakości polskiej demokracji. Fakt, że wszystkie anomalie poddawane są szczegółowej analizie, jest bardzo silnym sygnałem dla wszystkich tych, którzy w przyszłych wyborach będą zliczać głosy. Niezależnie, czy do przekłamań doszło na skutek niedbalstwa, czy są to przestępstwa popełnione pod wpływem politycznego zaciętrzewienia, przypadki takie zostawiają ślad i są ujawniane.

Obecny kryzys powinien mieć ten skutek, że udoskonalone zostaną procedury ustalania i raportowania wyników wyborów, w tym także procedury weryfikacji, czy w trakcie całego procesu nie doszło do zaniedbań lub nadużyć. Wierzimy, że nasze opracowanie pomoże w przygotowaniu takich procedur i że w przyszłości będą one wcześniej przemyślane, nie zaś tworzone na poczekaniu, często pochopnie. Rzetelność weryfikacji wyników wyborów powinna być jeszcze większa niż rzetelność samego ich zliczania. Tylko dzięki temu możliwe jest łagodzenie społecznych napięć wywoływanych przez wybory, a nie ich eskalacja. Chcielibyśmy, żeby nasze opracowania do takiego łagodzenia się przyczyniało.

Załącznik

Charakterystyka danych

Wyniki wyborów analizowane na poziomie obwodowych komisji wyborczych charakteryzują się szeregiem prawidłowości. Przede wszystkim, wykazują mocną zależność wyniku w obu turach. Poniższe wykresy porównują wyniki kandydatów z II tury z liczbą głosów, jaką łącznie zdobywali kandydaci o zbliżonych poglądach w I turze (z jednej strony kandydaci centrowi i lewicowi: Rafał Trzaskowski, Szymon Hołownia, Adrian Zandberg, Magdalena Biejat, Joanna Senyszyn, a z drugiej prawicowi: Karol Nawrocki, Sławomir Mentzen, Grzegorz Braun i Marek Jakubiak). Zależności są bardzo silne.



W większości komisji Trzaskowski zdobywał nieco więcej głosów (113%) niż wszyscy kandydaci demokratyczni razem wzięci, a w niektórych nawet istotnie więcej. W stosunkowo niewielkiej liczbie komisji (wyniki poniżej czerwonej linii, wyznaczającej tyle samo głosów w I i II turze) widzimy wyniki odbiegające od trendu – tj. Trzaskowski dostał wyraźnie mniej głosów niż kandydaci demokratyczni łącznie. Te odstające przypadki mogą być wynikiem nieprawidłowości lub błędów komisji wyborczych.

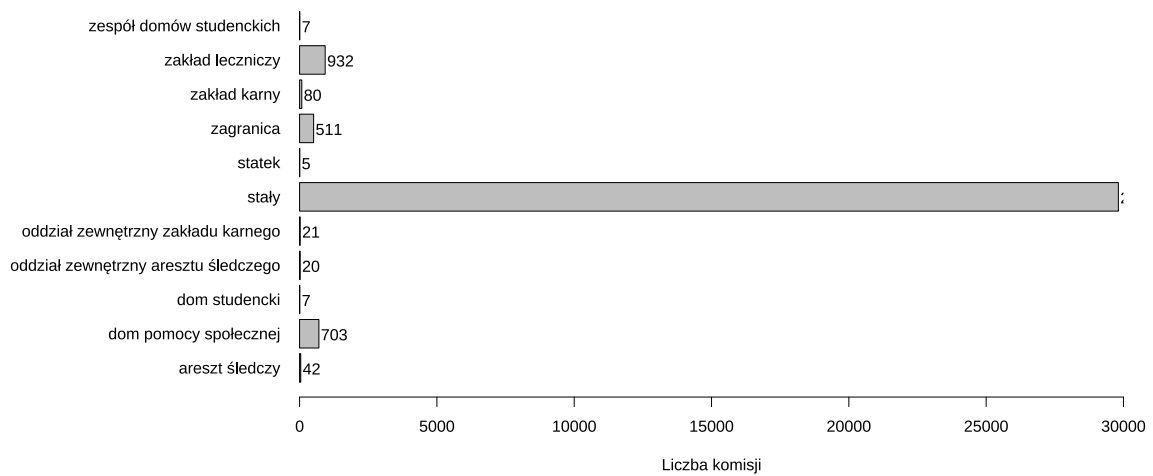
Przypadków mocno odstających jest stosunkowo niewiele, nawet gdybyśmy wzięli pod uwagę wszystkie 3398 komisji wyborczych, w których Trzaskowski miał choć jeden głos mniej niż dostali wszyscy kandydaci demokratyczni łącznie w II turze, skala tych różnic nie jest duża. Łącznie w tych komisjach kandydaci demokratyczni zdobyli w I turze o niecałe 20 tysięcy głosów więcej niż zdobył Trzaskowski w II turze.

Dla porównania, warto spojrzeć, jak wygląda podobna analiza dla Nawrockiego i kandydatów prawicowych w I turze (Nawrocki, Mentzen, Braun, Jakubiak). Tu również w wielu komisjach wynik Nawrockiego jest gorszy niż łączny wynik grupy kandydatów w I turze. Co więcej, w 114 komisjach zdobył on mniej głosów w II niż w I turze. W 5546 komisjach zdobył mniej głosów niż kandydaci prawicowi w I turze.

Przygotowanie danych do modelu

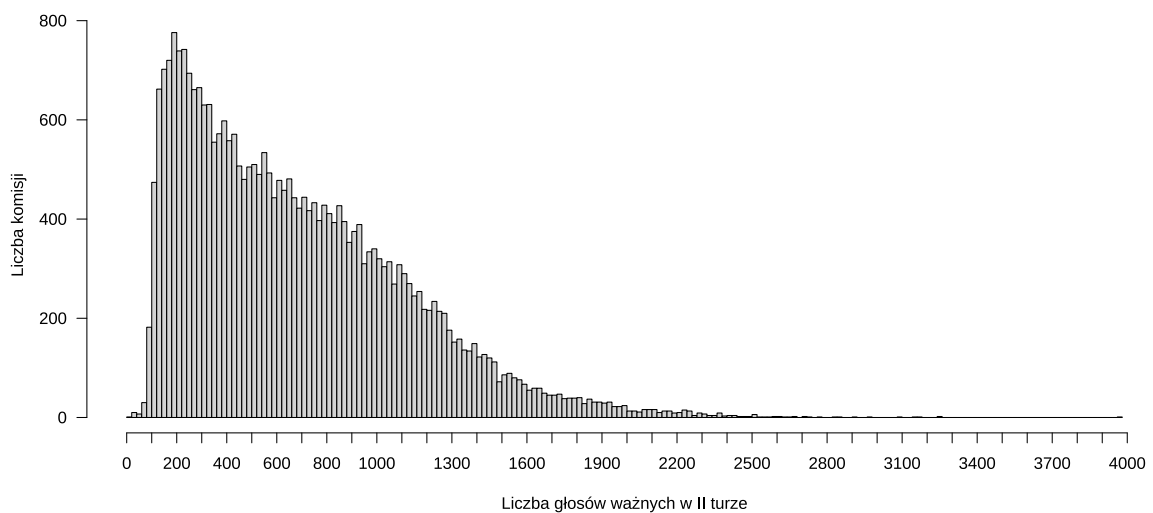
W pierwszym kroku przygotowaliśmy dane do analizy, usuwając z całego zbioru komisje o szczególnej specyfice, czyli działające w domach pomocy społecznej (DPS), szpitalach, za granicą, w akademikach, aresztach śledczych i zakładach karnych (łącznie wykluczaliśmy w ten sposób 2328 komisji).

Typ obwodowych komisji wyborczych



Następnie, spośród pozostałych komisji stałych wykluczaliśmy te, w których w II turze oddano mniej niż 100 ważnych głosów. Ta operacja pozwoliła wyeliminować najmniejsze komisje, często charakteryzujące się dużą losowością wyników oraz podwyższoną wariancją reszt; w rezultacie wykluczono 216 komisji.

Rozkład liczby głosów oddanych w stałych komisjach wyborczych



W celu zwiększenia trafności i stabilności oszacowań modelu regresji liniowej przeprowadzono systematyczną identyfikację oraz usuwanie obserwacji odstających (outlierów) według dwóch klasycznych kryteriów diagnostycznych: reszt studentyzowanych i dystansu Cooka.

Pierwszym krokiem była analiza reszt studentyzowanych (*studentized residuals*), które mierzą odchylenie wartości obserwowanej od wartości przewidywanej przez model, wyrażone w jednostkach odchylenia standardowego reszt dla danej obserwacji. Za przypadki odstające uznano obserwacje, dla których wartość bezwzględna reszty studentyzowanej przekraczała wartość progową ± 3 . Takie obserwacje mogą wskazywać na nienormalność rozkładu reszt lub błędy pomiarowe.

Drugim kryterium była analiza dystansu Cooka (*Cook's distance*), który ocenia wpływ pojedynczej obserwacji na wartości dopasowane w modelu. Obserwacje o dużym dystansie Cooka mają potencjał do istotnego zniekształcenia estymacji współczynników regresji. Zgodnie z zaleceniami w literaturze

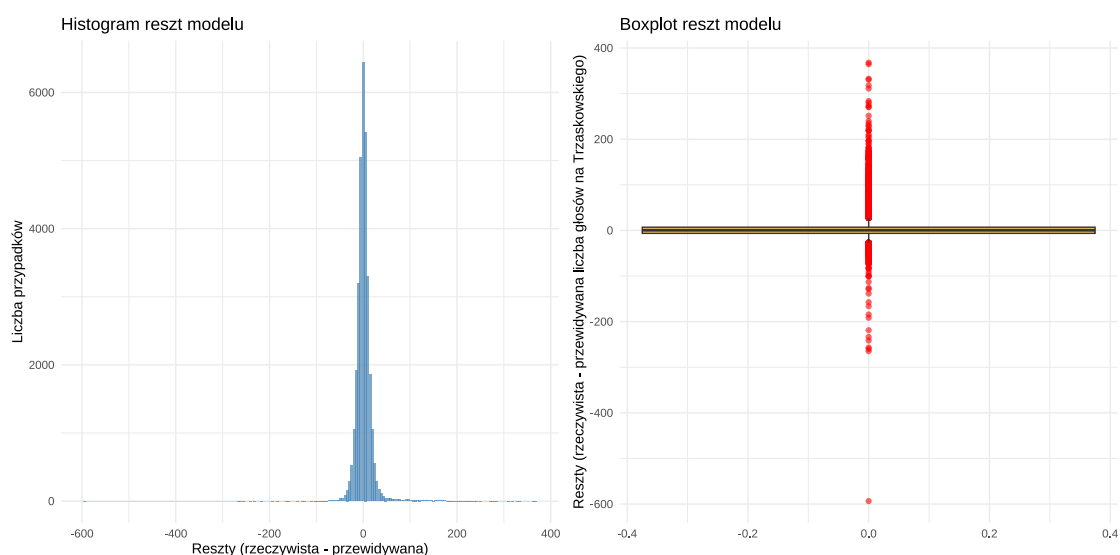
przedmiotu¹⁴, za obserwacje wpływowe uznano te, dla których dystans Cooka przekraczał wartość progową $4/n$, gdzie n oznacza liczbę obserwacji w próbie.

Obserwacje spełniające przynajmniej jedno z powyższych kryteriów zostały oznaczone jako outliers i wykluczone z dalszej analizy. Model regresji został następnie oszacowany ponownie na zredukowanym zbiorze danych, obejmującym 29 599 komisji stałych, w celu uzyskania stabilnych zależności.

Model i identyfikacja nietypowych komisji

Uzyskany model wykazuje bardzo wysokie dopasowanie ($R^2 = 0,9986$), co świadczy o dużej sile predykcyjnej przyjętego zestawu zmiennych objaśniających. Model przewidujący liczbę głosów uzyskanych przez Nawrockiego ma tylko nieznacznie niższą wartość R^2 (0,9968). Tak skonstruowane modele wykorzystaliśmy do identyfikacji komisji wyborczych, których wyniki odbiegają istotnie od wzorca statystycznego. Odstępstwa te, analizowane w rozkładzie reszt modelu oraz z użyciem różnych kryteriów identyfikacji anomalii, stanowią sygnał do dalszej weryfikacji – mogą wskazywać zarówno na specyficzne lokalne uwarunkowania, jak i potencjalne nieprawidłowości lub błędy w przebiegu głosowania.

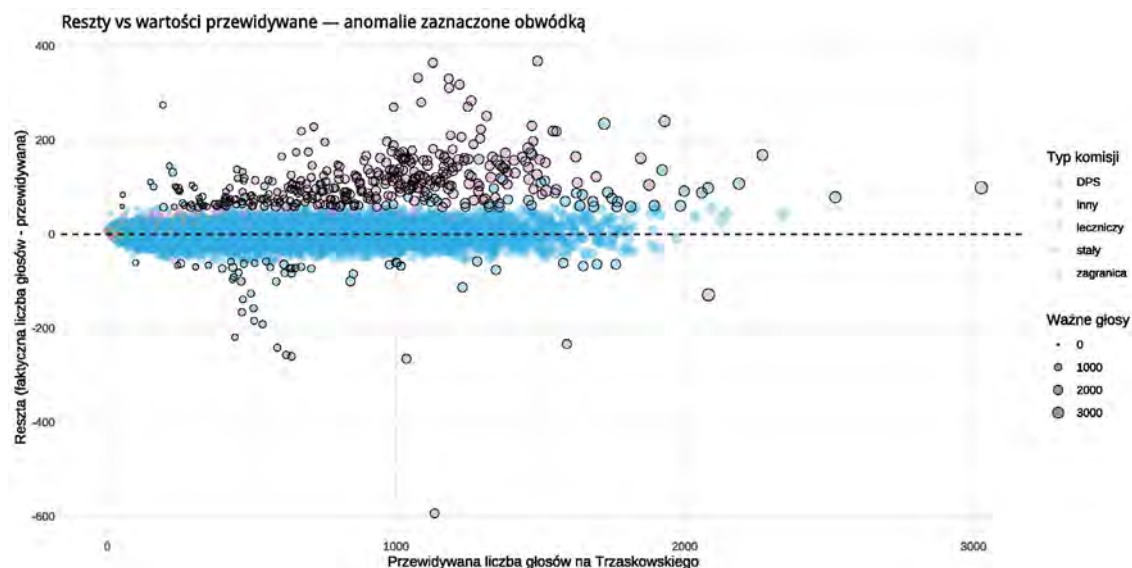
W celu identyfikacji komisji wyborczych o potencjalnie nieprawidłowych lub nietypowych wynikach wykorzystaliśmy model predykcyjny II tury wyborów oraz analizę rozkładu reszt. Zastosowaliśmy dwa komplementarne podejścia: regułę 3-sigma oraz regułę Tukeya. Pozwoliło to na identyfikację zarówno umiarkowanych, jak i skrajnych odstępstw od modelu predykcyjnego. Za anomalię uznaliśmy komisje, których wartość reszty przekraczała trzykrotność odchylenia standardowego lub wykraczała poza $[Q1 - 1,5 \times IQR; Q3 + 1,5 \times IQR]$ (reguła Tukeya). Komisje takie uznano za wymagające pogłębionej weryfikacji – zarówno pod kątem możliwości błędów, jak i potencjalnych nieprawidłowości. Wynikające z modelu odstępstwa nie stanowią jednak bezpośredniego dowodu nieprawidłowości, lecz wskazują przypadki o nietypowym zachowaniu na tle wzorca statystycznego.



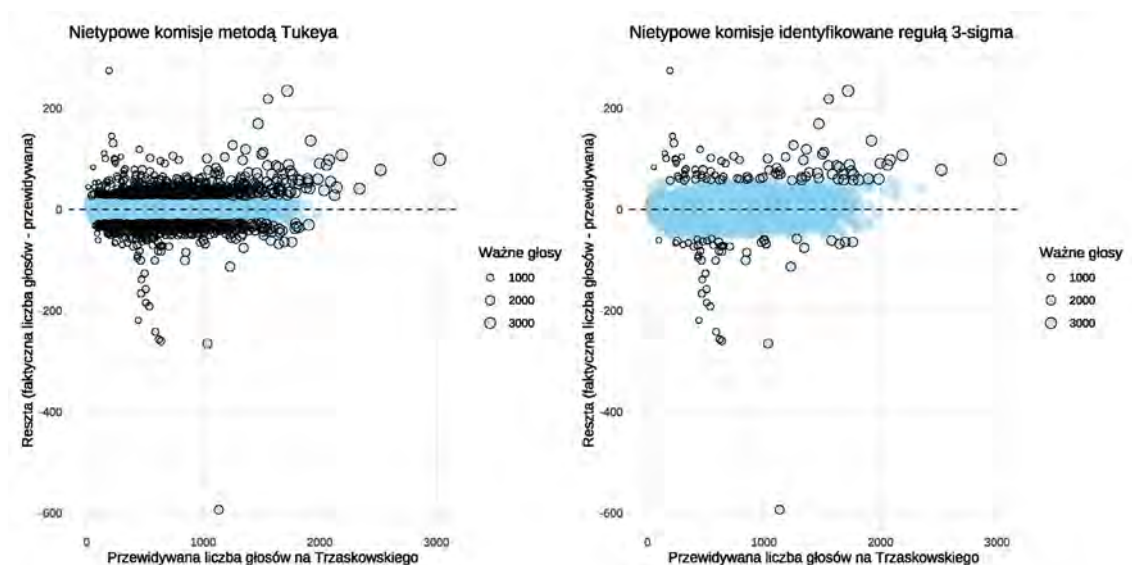
Poniższy wykres pokazuje odstające komisje (zaznaczone czarną obwódką) zidentyfikowane metodą 3-sigma. W analizie tej zostały uwzględnione wszystkie komisje wyborcze i – jak możemy zauważyć – większość odstających wyników stanowią komisje zagraniczne. Są one mocno nietypowe

14 Np. K.A. Bollen, R.W. Jackman, *Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases*, „Sociological Methods & Research” 1985, 13 (4), s. 510–542.

i zastosowanie standardowych metod do szukania potencjalnych nieprawidłowości nie jest w ich przypadku możliwe. Dlatego też w dalszych analizach skupiamy się na prawie 30 tysiącach komisji stałych.



Identyfikacja nietypowych komisji metodą Tukeya wydaje się nazbyt restrykcyjna, ponieważ klasyfikuje jako odstające nawet niewielkie odstępstwa od przewidywania, z których zdecydowana większość jest efektem normalnego zróżnicowania wyników na poziomie lokalnym. Nawet jeśli przyjmiemy taką bardziej restrykcyjną metodę identyfikacji podejrzanych przypadków, to okaże się, że liczba komisji, dla których model przewiduje zawyżenie wyniku Nawrockiego, jest praktycznie identyczna (661), jak dla komisji, w których mógł być zawyżony wynik Trzaskowskiego (668). Jeśli porówna się wartości przewidywanych z tymi podanymi przez okręgowe komisje wyborcze, okaże się, że Nawrocki mógł uzyskać 26 442 głosów, a Trzaskowski 28 583. Wyniki te wskazują, że ewentualne nieprawidłowości w liczeniu wyników wyborów nie miały dużej skali (podkreślimy: część z tych odstępstw jest zupełnie naturalna) i nie mogły mieć wpływu na końcowy efekt wyborów.



W głównym tekście skupiamy się na analizie odstających wyników przy użyciu reguły 3-sigma. Poniżej zamieszczamy listę wszystkich komisji, które zostały na tej podstawie zakwalifikowane jako warte sprawdzenia.

Odstające komisje

			Wyniki wg OKW		Przewidywana liczba głosów		
			Trzaskowski	Nawrocki	Trzaskowski	Nawrocki	
1	95	m. Kraków	540	1132	1133	539	-593
2	61	m. Bielsko-Biała	771	1048	1036	783	-265
3	3	gm. Olesno	378	637	638	377	-260
4	13	m. Mińsk Mazowiecki	363	611	620	354	-257
5	35	m. Tychy	347	581	588	340	-241
6	9	gm. Strzelce Opolskie	223	416	442	197	-219
7	17	m. Gdańsk	346	585	537	394	-191
8	25	m. Grudziądz	324	504	508	320	-184
9	403	m. Kraków	300	668	466	502	-166
10	30	m. Bielsko-Biała	349	610	506	453	-157
11	4	gm. Brześć Kujawski	331	466	469	328	-138
12	12	m. Kobyłka	371	507	497	381	-126
13	6	gm. Postomino	1117	897	1230	784	-113
14	58	m. Katowice	484	606	585	505	-101
15	10	gm. Tarnów	363	855	463	755	-100
16	1	gm. Iwonicz-Zdrój	742	1070	842	970	-100
17	6	m. Kamienna Góra	338	368	434	272	-96

18	13	m. Radom	347	520	438	429	-91
19	3	gm. Międzyzdroje	767	601	852	516	-85
20	1	gm. Lesznówola	555	631	638	548	-83
21	20	m. Łęborg	532	541	614	459	-82
22	20	m. Żyrardów	373	492	455	410	-82
23	3	gm. Mielno	1270	686	1346	610	-76
24	11	m. Gdynia	633	458	706	385	-73
25	9	gm. Ślesin	325	745	398	672	-73
26	9	m. Słupsk	558	496	631	423	-73
27	3	gm. Brzeg Dolny	565	523	638	450	-73
28	4	m. Wrocław	579	386	651	314	-72
29	16	gm. Oborniki	494	324	565	253	-71
30	6	gm. Nowa Dęba	364	1009	434	939	-70
31	47	m. Katowice	236	348	305	279	-69
32	624	Ursynów	1580	467	1648	399	-68
33	162	m. Bydgoszcz	949	594	1017	526	-68
34	7	gm. Osiecznica	284	391	350	325	-66
35	1	gm. Braniewo	177	340	243	274	-66
36	4	m. Ustroń	912	487	977	422	-65
37	38	m. Siedlce	547	833	611	769	-64
38	714	Bemowo	1697	517	1761	453	-64
39	715	Bemowo	1632	477	1696	413	-64
40	3	m. Ostrołęka	410	627	473	564	-63
41	6	gm. Poronin	192	1026	254	964	-62
42	1	gm. Kołbaskowo	449	309	511	247	-62

43	11	m. Kołobrzeg	942	475	1003	414	-61
44	617	Ursynów	1518	511	1579	450	-61
45	2	gm. Biszcz	37	443	98	382	-61
46	70	m. Rzeszów	940	838	1000	778	-60
47	2	gm. Jerzmanowa	369	397	428	338	-59
48	20	m. Ostrów Wielkopolski	492	365	550	307	-58
49	718	Bemowo	1223	714	1281	656	-58
50	9	gm. Grybów	252	914	195	971	57
51	1	gm. Bukowina Tatrzańska	373	828	315	886	58
52	117	m. Poznań	1298	495	1240	553	58
53	108	m. Poznań	1822	437	1764	495	58
54	20	Śródmieście	1871	839	1812	898	59
55	164	m. Gdańsk	1743	754	1684	813	59
56	4	gm. Czernica	1381	939	1322	998	59
57	2	gm. Kościelisko	603	678	544	737	59
58	1	gm. Kościelisko	679	973	620	1032	59
59	2	gm. Sobótka	702	469	643	528	59
60	4	gm. Międzyzdroje	914	448	855	507	59
61	2	gm. Warka	471	707	411	767	60
62	2	gm. Trzebnica	565	295	505	355	60
63	73	m. Zielona Góra	852	309	792	369	60
64	294	Wola	2045	659	1985	719	60
65	1	gm. Kosakowo	1697	807	1637	867	60

66	71	m. Gdańsk	1955	721	1894	782	61
67	410	m. Kraków	1026	489	965	550	61
68	228	m. Wrocław	932	399	871	460	61
69	9	gm. Łęczna	328	412	267	473	61
70	18	gm. Środa Wielkopolska	396	305	335	366	61
71	51	m. Wrocław	833	303	771	365	62
72	305	Wola	1531	518	1469	580	62
73	26	m. Świnoujście	1053	528	990	591	63
74	10	m. Zabrze	531	376	468	439	63
75	168	m. Kraków	1386	530	1322	594	64
76	37	m. Słupsk	716	332	651	397	65
77	368	Praga-Północ	921	469	856	534	65
78	13	Śródmieście	1424	465	1359	530	65
79	1	gm. Szypliszki	326	442	260	508	66
80	4	m. Karpacz	646	256	579	323	67
81	3	gm. Gniezno	533	341	465	409	68
82	2	gm. Myślibórz	536	258	467	327	69
83	273	m. Wrocław	1188	541	1119	610	69
84	6	gm. Kosakowo	1836	713	1767	782	69
85	10	gm. Busko-Zdrój	490	439	421	508	69
86	42	m. Wrocław	1751	564	1682	633	69
87	128	m. Wrocław	1213	403	1143	473	70
88	2	gm. Jastarnia	390	196	320	266	70
89	676	Włochy	1656	627	1586	697	70
90	3	gm. Rewal	1531	845	1459	917	72

91	292	m. Wrocław	1170	420	1097	493	73
92	6	gm. Pruszcz Gdański	1461	459	1387	533	74
93	7	gm. Wejherowo	618	494	543	569	75
94	124	m. Kraków	1209	441	1133	517	76
95	75	m. Gdańsk	1824	686	1748	762	76
96	8	gm. Barlinek	395	208	318	285	77
97	484	Praga-Południe	1120	436	1042	514	78
98	1	gm. Rzekuń	555	919	477	997	78
99	36	m. Poznań	2600	545	2521	624	79
100	10	gm. Żnin	488	347	409	426	79
101	5	gm. Lubomia	409	439	329	519	80
102	18	m. Kołobrzeg	1192	571	1109	654	83
103	4	gm. Orzysz	137	194	53	278	84
104	30	m. Kraków	1729	672	1645	756	84
105	292	m. Kraków	1711	495	1624	582	87
106	52	m. Wrocław	1638	456	1551	543	87
107	101	Mokotów	2147	710	2059	798	88
108	703	Ursus	1822	788	1733	877	89
109	3	gm. Jastarnia	761	231	670	322	91
110	186	m. Poznań	2089	477	1998	568	91
111	7	m. Ostrów Mazowiecka	338	205	246	297	92
112	2	m. Wisła	730	217	634	313	96
113	186	m. Gdańsk	1436	541	1338	639	98
114	5	gm. Nowy Targ	343	920	245	1018	98

115	23	m. Dąbrowa Górnicza	888	240	790	338	98
116	280	Wola	2179	727	2081	825	98
117	109	m. Poznań	3126	837	3027	936	99
118	3	m. Grybów	338	305	237	406	101
119	94	m. Sosnowiec	1133	658	1032	759	101
120	1	gm. Stanin	261	475	160	576	101
121	8	m. Mława	646	470	544	572	102
122	607	Ursynów	1288	342	1184	446	104
123	157	m. Łódź	397	53	293	157	104
124	99	m. Łódź	846	406	740	512	106
125	9	m. Gdańsk	2295	957	2188	1064	107
126	291	m. Wrocław	1611	610	1502	719	109
127	4	gm. Bychawa	260	163	149	274	111
128	287	m. Wrocław	1629	432	1515	546	114
129	6	gm. Władysławowo	570	166	451	285	119
130	1	gm. Rewal	1482	764	1363	883	119
131	706	Ursus	1380	459	1253	586	127
132	13	m. Zakopane	358	834	227	965	131
133	73	m. Gdańsk	2058	671	1922	807	136
134	4	gm. Staszów	360	209	215	354	145
135	10	gm. Trzebiatów	1640	852	1470	1022	170
136	113	Mokotów	1774	136	1555	355	219
137	1	gm. Jastarnia	1956	1286	1721	1521	235
138	1	gm. Magnuszew	467	193	192	468	275

Autorzy

Dominik Batorski – dr hab. w dziedzinie nauk społecznych i ekspert w dziedzinie *data science*. Pracuje w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego (ICM) na Uniwersytecie Warszawskim. Jest członkiem Rady Polskiego Instytutu Ekonomicznego i przewodniczącym Rady Centrum Badania Opinii Społecznej. Kieruje Radą Programową konferencji Data Science Summit oraz DSS AI/ML Edition.

Jarosław Flis – dr hab., socjolog, profesor Uniwersytetu Jagiellońskiego. Wykładowca na Wydziale Zarządzania i Komunikacji Społecznej Uniwersytetu Jagiellońskiego. Komentator polityczny. Współpracuje z „Tygodnikiem Powszechnym”. Członek Zespołu Ekspertów Wyborczych forumIdei Fundacji im. Stefana Batorego.

Piotr Szulc – dr, specjalizuje się w statystyce i uczeniu maszynowym, łącząc pracę badawczą z zastosowaniami komercyjnymi. Z wykształcenia matematyk, uzyskał stopień doktora na Wydziale Matematyki Politechniki Wrocławskiej. W pracy badawczej koncentrował się na zastosowaniach modeli regresyjnych w analizie danych genetycznych, współpracując m.in. z Politechniką Wrocławską oraz Uniwersytetem Stanforda. Jest autorem serwisu danetyka.com, gdzie publikuje artykuły dydaktyczne oraz analizy statystyczne dotyczące istotnych problemów społecznych.

Fundacja im. Stefana Batorego

Sapieżyńska 10a
00-215 Warszawa
tel. (48-22) 536 02 00
fax (48-22) 536 02 20
batory@batory.org.pl
www.batory.org.pl

Teksty udostępniane na licencji
Creative Commons. Uznanie autorstwa
na tych samych warunkach
3.0 Polska (CC BY SA 3.0 PL)



Redakcja: Izabella Sariusz-Skąpska
Korekta: Joanna Liczner
Warszawa 2025
ISBN 978-83-68235-75-3