ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA XXXVI (1993)

Z. Bondarczuk, A. Ciok, W. Szczesny

Warszawa

Statystyczna analiza odstępstwa od proporcjonalnego rozdziału miejsc w Sejmie RP w 1991 r.

(*Praca wpłynęła do Redakcji 27.10.1992*)

1. Wstęp. Opracowywaniu i przyjęciu ordynacji wyborczej do Sejmu RP z 28 czerwca 1991 roku[1] towarzyszyło wiele kontrowersji i poważnych napięć politycznych. Kontrowersje te dotyczyły nie tylko rozwiązań bezpośrednio odnoszących się do mechanizmu wyłaniania przedstawicieli i decydujących o charakterze systemu wyborczego: zasad zgłaszania kandydatów, techniki głosowania, struktury okręgów wyborczych i związanego z nią sposobu ustalania podziału mandatów. Podstawowy konflikt dotyczył bowiem samej zasadniczej koncepcji wyborów.

Ostatecznie przyjęta została przez posłów koncepcja oparcia ordynacji na zasadzie proporcjonalności. Było to uzasadniane koniecznością zapewnienia w Sejmie szerokiej reprezentacji politycznej, możliwie wiernie odwzorowywującej poglądy i postawy wyborców.

W analizach i prognozach przedwyborczych (patrz np. Janicki (1991)) wyrażane były poglądy, że "ordynacja nie będąc większościową, nie ma też zalet proporcjonalności, ma za to wszystkie jej wady". Na podstawie przykładów hipotetycznych sytuacji wyborczych dowodzono, ze marginalne partie odbiorą głosy większym ugrupowaniom. Prognozowano spłaszczenie wyników wyborów i zarzucano ordynacji, że doprowadzi do zwiększenia rozdrobnienia politycznego.

Nie jest naszym celem odgrzewanie dyskusji nad ordynacją. Interesuje nas wyłącznie ostateczny rezultat w postaci struktury politycznej Sejmu wyłonionego w wyniku wyborów, które odbyły się 27 października 1991 roku.

[1] Ustawa z dnia 28 czerwca 1991 r. – Ordynacja wyborcza do Sejmu Rzeczpospolitej Polskiej. Dz. U. Nr 59, poz. 252.

Przedmiotem prezentowanej pracy jest próba znalezienia odpowiedzi na następujące pytania:

- Jak silne są odstępstwa struktury Sejmu od proporcjonalnego odwzorowania woli wyborcow? Czym ją mierzyć?
- Jak dalece sposób pomiaru pozwala porównywać rezultaty wyborów przeprowadzanych w odmiennych warunkach?
- Jak zmierzyć siłę ewentualnej tendencji do nadreprezentowania partii, które zdobyły dużo (mało) głosów?
- Jaką rolę odegrały tzw. listy ogólnopolskie?

Spróbujemy odpowiedzieć na te pytania stosując techniki statystyczne, które zostały świeżo zaproponowane i szczegółowo opisane w artykule Bondarczuka, Kowalczyk, Pleszczyńskiej i Szczesnego (1992). Są to metody ilościowe oparte na pewnych modelach stochastycznych, stanowiące daleko idące rozwinięcie klasycznych rozważań zawartych np. w pracach Balińskiego i Younga (1985).

2. Podstawowe pojęcia i dane. Z założenia nie omawiamy całej ordynacji wyborczej, przedstawiamy jedynie niezbędne pojęcia i informacje. Krótko prezentujemy również wykorzystane przez nas dane.

W postępowaniu wyborczym brały udział wyłącznie tzw. komitety wyborcze. Miały one prawo do zgłaszania okręgowych i ogólnopolskich list kandydatów. Zasadniczy wybór dokonywał się w 37 okręgach wyborczych: tu rozdzielono w sumie 391 mandatów. Pozostałych 69 mandatów było dzielonych w skali ogólnokrajowej z list ogólnopolskich, proporcjonalnie do wyników głosowania na powiązane z nimi listy okręgowe.

Każdy wyborca dysponował w wyborach do Sejmu tylko jednym głosem. Głos ten miał jednak podwójne, a niekiedy potrójne znaczenie. Przede wszystkim był to głos na kandydata przy nazwisku którego wyborca postawił krzyżyk. W przyjętym w ordynacji mechanizmie głosowania głos ten był zarazem głosem na całą listę okręgową. W ten sposób liczba głosów oddanych na listę okregową jest sumą głosów ważnie oddanych na kandydatów na niej umieszczonych.

Zgodnie z zasadą proporcjonalności, mandaty w okręgu dzielono pomiędzy listy okręgowe odpowiednio do liczby uzyskanych przez nie głosów. W ordynacji przyjęto, że podział ten jest dokonywany za pomocą systemu "największych reszt" ("largest remainders"), zwanego też metodą Hamiltona od nazwiska rzecznika tej metody w słynnych debatach nad podziałem miejsc w senacie Stanów Zjednoczonych. Zgodnie z nim, sumę wszystkich ważnie oddanych głosów w danym okręgu dzieli się przez liczbę mandatów, jakie są w nim do obsadzenia. Przez uzyskaną w ten sposób średnią liczbę głosów przypadających na jeden mandat dzieli się następnie liczbę głosów uzyskanych

przez poszczególne listy. Części całkowite otrzymanych ilorazów określają liczby mandatów w okręgu, przypadających na każdą z list. Z reguły nie zostaną w ten sposób rozdzielone w okręgu wszystkie mandaty. Pozostałe przypadną tym listom, którym kolejno pozostały największe reszty niewykorzystanych głosów.

Z kolei liczbę głosów oddanych na listę ogólnopolską obliczono sumując głosy oddane na związane z nią listy okręgowe. Podział 69-ciu mandatów pomiędzy listy ogólnopolskie nastąpił zgodnie z systemem Saint-Lagüe. Polega on na podzieleniu liczby głosów oddanych na poszczególne listy przez 1.4, 3, 5, i dalej przez kolejne liczby nieparzyste. Spośród uzyskanych ilorazów wybiera się 69 największych. Każda lista otrzymuje tyle mandatów , ile "jej" ilorazów znalazło się wśród tych wybranych.

Istotny wpływ na ostateczny podział mandatów może mieć przewidziana w ordynacji możliwość tzw. blokowania list – zarówno okręgowych jak i ogólnopolskich. Polega ono na łącznym traktowaniu przy podziale mandatów głosów oddanych na zblokowane listy. W przypadku blokowania list okręgowych wewnętrzny podział mandatów może następować w sposób dowolnie ustalony i przedstawiony w oświadczeniu złożonym okręgowej komisji wyborczej.

W wyniku wyborów do Sejmu RP, przeprowadzonych 27 października 1991 roku, wybrano 460 posłów: 391 z list okręgowych (w 37 okręgach wyborczych) oraz 69 z ogólnopolskich list kandydatów na posłów. Głosów ważnych oddano łącznie 11 218 602.

Posłowie byli wybierani spośród 6 980 kandydatów, wystawionych przez 111 komitetów wyborczych. Spośród tych komitetów 69 zarejestrowało listy tylko w jednym okręgu wyborczym – nie miały więc one prawa do zgłaszania list ogólnopolskich.

Poszczególne komitety należące do tej grupy zebrały od 608 do 42 031 głosów. W skali całego kraju głosowało na nie 3.23% wyborców, zaś zdobyły one 9 (tj. 1.96%) spośród 460 mandatów. Na jedną spośród list okręgowych, z której kandydat wszedł do Sejmu, oddane zostały zaledwie 1 922 glosy. W sumie tylko 3 komitety wyborcze z tej grupy uzyskały mandaty zbierając w skali kraju ponad 0.22 (tj. 1/460) procent ważnie oddanych głosów.

W więcej niż jednym okręgu wyborczym zarejestrowały listy 42 komitety wyborcze, z których 20 wprowadziło do Sejmu co najmniej po jednym przedstawicielu. Po jednym mandacie uzyskały dwa z nich, gromadząc 0.46 i 1.42% wszystkich ważnie oddanych głosów. Z kolei wśród przegranych znalazły się 3 komitety wyborcze, które zdołały zgromadzić od 0.66 do 0.70% tych głosów.

Spośród 27 komitetów, które zarejestrowały listy ogólnopolskie, 11 spełniło warunki ordynacji wyborczej dotyczące udziału w podziale 69 mandatów niepodlegających rozdziałowi w okręgach wyborczych. Każdy z tych

komitetów zgromadził od 1.18 do 12.32% ważnie oddanych głosów.

Tab. 1. Wyniki wyborów do Sejmu RP: a – ostateczne, b – bez uwzględniania listy krajowej.

1	2	3		4	5	6	7
Grupa komitetów z procentem głosów:	Liczba komitetów w grupie	Liczba głosów oddanych na grupę		Podział według metody największych reszt	Uzyskana liczba miejsc w Sejmie	Oczekiwana liczba miejsc w Sejmie	Wskaźnik nadreprezentacji
0 – 1	95	1 166 562	a	48	19	47.83	0.397
			b	41	19	40.66	0.467
1 – 3	6	1 165 568	a	48	24	47.79	0.502
			b	41	23	40.62	0.566
3 – 5.5	3	1 547 285	a	63	71	63.44	1.119
			b	54	59	53.93	1.094
5.5 – 8	2	1 681 716	a	69	83	68.96	1.204
			b	58	69	58.61	1.177
8 – 11	3	2 930 600	a	120	141	120.16	1.173
			b	102	120	102.14	1.175
≥ 11	2	2 726 871	a	112	122	111.81	1.091
			b	95	101	95.04	1.063

W Tab. 1 prezentujemy wyniki wyborów po podzieleniu wszystkich komitetów wyborczych na 6 klas, w zależności od liczby zebranych przez każdy z nich, ważnie oddanych głosów. Wybrane przedziały klasowe są podane (w procentach) w pierwszej kolumnie. Liczby komitetów wyborczych należących do kolejnych klas sa podane w drugiej kolumnie. Następna kolumna zawiera sumy głosów zdobytych przez komitety z danej grupy. W czterech ostatnich kolumnach podane są:

- liczba miejsc jaka przypadłaby przy metodzie Hamiltona (zastosowanej w skali kraju a nie okręgu),
- liczba miejsc uzyskanych w Sejmie,
- oczekiwana liczba miejsc,
- tzw. wskaźniki nadreprezentacji.

W kolumnach tych dla każdej klasy podane są po dwie liczby. Te z nich, które znajdują się w wierszach oznaczonych symbolem "a",odnoszą się do podziału 460 mandatów. W wierszach "b" podane są te same wielkości, ale odniesione do 391 mandatów obsadzanych z list okręgowych.

Tzw. oczekiwana liczba miejsc została wyznaczona proporcjonalnie do liczby głosów podanej w trzeciej kolumnie Tab. 1. Uzyskano w ten sposob

możliwość porównania faktycznego podziału mandatów wśród wybranych klas komitetów wyborczych z niemożliwym do zrealizowania podziałem "ściśle proporcjonalnym", dopuszczającym dzielenie miejsca na części. Z kolei liczby z ostatniej kolumny tablicy są ilorazami liczb umieszczonych w tych samych wierszach w dwóch poprzedzających ją kolumnach. Ilorazy te zostały obliczone w celu zobrazowania nadreprezentowania (iloraz > 1) lub niedoreprezentowania (iloraz < 1) poszczególnych klas komitetów wyborczych.

Porównanie liczby miejsc uzyskanych w Sejmie z liczbą wyznaczoną metodą największych reszt wykazuje, że komitety o większej liczbie głosów są uprzywilejowane. Również uzyskane wartości wskaźnika nadreprezentacji świadczą o tym, że wyniki wyborów odbiegają od proporcjonalnego podziału mandatów; widoczna jest przy tym tendencja do nadreprezentowania komitetów wyborczych, które zdobyły dużo głosów. Wektor wskaźników nadreprezentacji nie wystarcza jednak do obiektywnej oceny wielkości odstępstw od "sprawiedliwej" proporcjonalnosci. Potrzebne są do tego inne narzędzia, które prezentujemy w następnym punkcie.

3. Narzędzia pomiarowe. Przedstawmy na początek model ogólny. Załóżmy, że badana populacja składa się z s podpopulacji i jest charakteryzowana parą cech (X, Y) o wartościach naturalnych. Jest więc opisana ciągiem liczb (x_i, y_i), $i = 1, \ldots s$, przy czym $\sum_{i=1}^{s} x_i = N$, $\sum_{i=1}^{s} y_i = M$, $x_1 \leq \ldots \leq x_s$.

W naszym przypadku (wyborów do Sejmu RP) x_i jest liczbą głosów ważnie oddanych na i-ty komitet wyborczy, a y_i – liczbą mandatów uzyskanych przez ten komitet.

Rozważmy teraz parę rozkładów prawdopodobieństwa na zbiorze $\{1, \ldots, s\}$ zdefiniowanych prawdopodobieństwami:

$$f_i = \frac{x_i}{N}, \quad g_i = \frac{y_i}{M}, \quad i = 1, \ldots, s \tag{1}$$

lub wartościami dystrybuant:

$$F_i = \sum_{j=1}^{i} \frac{x_j}{N}, \quad G_i = \sum_{j=1}^{i} \frac{y_j}{M}, \quad i = 1, \ldots, s. \tag{2}$$

Krzywą łamaną o wierzchołkach $(0,0), (F_1, G_1), \ldots, (F_{s-1}, G_{s-1}), (1,1)$, oznaczamy przez C i nazywamy krzywą proporcjonalności. Krzywa ta jest niemalejąca i zawiera się w kwadracie jednostkowym.

Zauważmy, że krzywa C byłaby identyczna z przekątną kwadratu jednostkowego wtedy i tylko wtedy, gdyby wszystkie pary (x_i, y_i) spełniały warunek:

$$\frac{y_i}{x_i} = \frac{M}{N}, \quad i = 1, \ldots, s. \tag{3}$$

czyli gdyby liczba uzyskanych mandatów była proporcjonalna do liczby uzyskanych głosów. Jednakże dla zadanych liczb $x_1, \ldots, x_s$ oraz M wyrażenia $(x_i M)/N$ nie są na ogół liczbami naturalnymi, czyli ściśle proporcjonalna reprezentacja elektoratów nie istnieje; można jedynie starać się o to, żeby odchylenie od zasady propocjonalności było niewielkie.

Wyróżnijmy dwa skrajne typy odchyleń:

(i) — gdy ilorazy y_i/x_i są niemalejącą funkcją i,

(ii) — gdy ilorazy y_i/x_i są nierosnącą funkcją i.

W przypadku (i) krzywa C jest położona pod przekątną i jest wypukła, w przypadku (ii) krzywa C leży nad przekątną i jest wklęsła. Jeśli ilorazy y_i/x_i rosną do pewnego $t < s$, a następnie maleją, to krzywa C przecina przekątną.

Odstępstwo krzywej C od przekątnej może być opisane za pomocą podwojonej różnicy między polem obszaru położonego pod przekątną a nad krzywą, oznaczonego A na Rys. 1, oraz polem obszaru położonego nad przekątną a pod krzywą, oznaczonego B. Różnicę tę będziemy nazywać wskaźnikiem przemieszczenia i oznaczać przez a_C. Zatem

$$a_C = \frac{1}{NM} \sum_{i=1}^{s} \left(y_i \sum_{j=1}^{i-1} x_j - x_i \sum_{j=1}^{i-1} y_j \right) \tag{4}$$

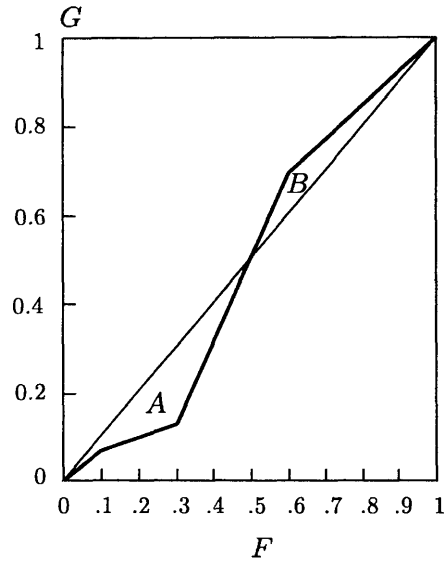
W przypadku (i) wskaźnik przemieszczenia a_C jest dodatni, w przypadku (ii) – ujemny.

Tab. 2. Dane wykorzystywane w Rys. 1 i 2.

x_i	y_i	f_i	g_i	$\frac{g_i}{f_i}$
1000	5	0.1	0.05	0.50
2000	5	0.2	0.05	0.25
3000	60	0.3	0.60	2.00
4000	30	0.4	0.30	0.75

Rys. 1 przedstawia krzywą C dla $s = 4$ i danych przedstawionych w Tab. 2. Dla tej krzywej wskaźnik przemieszczenia a_C ma wartość równą 0.045.

Wprowadźmy teraz drugą krzywą – oznaczmy ją przez L – która powstaje tak jak krzywa C po uprzednim dokonaniu permutacji punktów (x_i, y_i) w taki sposób, żeby ilorazy $h_i = (y_i N)/(x_i M)$ były w następstwie tego ustawione niemalejąco. Jeśli ilorazy h_i są uporządkowane niemalejąco, czyli zgodnie z uporządkowaniem $x_1, \ldots, x_s$, to krzywe L i C pokrywają się. Krzywa L jest wypukła; pokrywa się z przekątną kwadratu wtedy i tylko wtedy, gdy spełniony jest warunek (3); w pozostałych przypadkach przebiega poniżej przekątnej. Wskaźnik przemieszczenia dla krzywej L ozna-



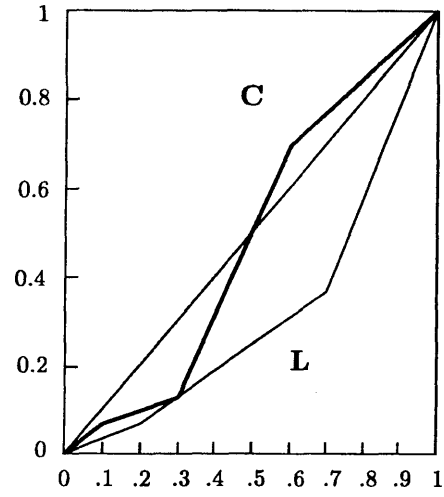
Rys. 1. Krzywa C dla danych przedstawionych w Tab. 2.

czamy tu symbolem a_L. Jest on zawsze nieujemny. Oczywiście dla każdego ciągu $((x_i, y_i), i = 1, \ldots, s)$ jest spełniona nierówność $|a_C| \leq a_L$.

Tak więc a_L jest wskaźnikiem calkowitego odstępstwa od proporcjonalności, natomiast a_C określa, jak silna jest tendencja do nadreprezentacji dużych komitetów. Gdy $a_C = a_L$, całkowite odstępstwo jest tożsame z nadreprezentacją dużych komitetów; gdy $a_c = -a_L$, jest przeciwnie: całkowite odstępstwo jest tożsame z nadreprezentacją małych komitetów. Ogólnie, para wskaźników a_C i a_L stanowi narzędzie pomiaru wielkości i typu odstępstw od proporcjonalności.

Przyjrzyjmy się parom (a_C, a_L) dla danych (a) z Tab. 1 oraz z Tab. 2. Dla danych (a) z Tab. 1, przyjmując jako x_i liczbę głosów oddanych na grupę i, a jako y_i liczbę miejsc uzyskanych przez tę grupę w Sejmie przy podziale 460 mandatów, otrzymujemy $a_C = 0.116$ i $a_L = 0.131$. Ponieważ wartość wskaźnika a_L jest niewielka, a wskaźnik a_C ma wartość zbliżoną do a_L, jesteśmy skłonni sądzić, że mamy do czynienia z małym odstępstwem od proporcjonalności spowodowanym głównie nadreprezentowaniem grup o większych x'ach.

Z kolei Rys. 2 przedstawia obie krzywe C i L dla danych z Tab. 2. Krzywa C przecina przekątną i silnie odbiega od krzywej L. Wartość wskaźnika a_L jest w tym przypadku równa 0.355, jest więc znacznie większa niż wartość wskaźnika a_C wynosząca 0.045. Wydaje się więc, że dość znaczne odstępstwo od proporcjonalności w niewielkim tylko stopniu jest spowodowane słabą tendencją do nadreprezentowania jednostek o dużych wartościach x_i.



Rys. 2. Krzywe C i L dla danych przedstawionych w Tab. 2.

Jednakże takie wnioskowanie nie jest jeszcze dostatecznie uzasadnione. Wyznaczanie wartości parametru a_L i porównywanie a_C z a_L dla oceny nadreprezentowania dużych (małych) jednostek nie może być wiarygodne tak długo, dopóki nie stworzy się odpowiedniej skali dla par wskaźników (a_C, a_L). Bez takiej skali nawet porównywanie między sobą wskaźników a_L (a tym bardziej par (a_C, a_L)) dla dwóch różnych zbiorów danych jest prawomocne tylko wtedy, gdy wektory $(x_1, \ldots, x_s, M)$ są w obu tych zbiorach jednakowe. Chodzi więc o takie wyskalowanie par (a_C, a_L), które uwolni od wpływu wektora $(x_1, \ldots, x_s, M)$.

W rozpatrywanej przez nas sytuacji wyborczej liczby $x_1, \ldots, x_s$ zależą od specyfiki populacji wyborców i od zachowań tych wyborców w ramach ustalonej ordynacji; liczby $y_1, \ldots, y_s$ przy danych $x_1, \ldots, x_s$ zależą już tylko od ordynacji wyborczej. W takich przypadkach chcemy oceniać (porównywać) wielkość i rodzaj odstępstw ordynacji wyborczej od proporcjonalności, eliminując wpływ wywierany przez strukturę populacji i zachowania wyborców.

4. Cechowanie narzędzia. W pomiarach fizycznych wyróżnia się punkty wzorcowe, które mają mniej lub bardziej uniwersalną interpretację ("punkt wrzenia", "punkt topnienia" itp.) i za ich pomocą cechuje się narzędzia pomiarowe. Narzędzia statystyczne cechuje się analogicznie, wyróżniając pewne punkty oraz stochastyczne "modele odniesienia" o przejrzystej i uniwersalnej interpretacji.

Stosowanymi przez nas narzędziami statystycznymi są wskaźniki a_C i a_L. Punktem wyróżnionym na skali wskaźnika a_C jest wartość tego wskaźnika dla reprezentacji otrzymanej metodą największych reszt. Oznaczamy

go a_C^H (od nazwiska Hamitona, który był rzecznikiem metody największych reszt w słynnych debatach nad podziałem miejsc w Senacie Stanów Zjednoczonych) i nazywamy punktem Hamiltona dla wskaźnika a_C. Analogicznie wprowadzamy punkt Hamiltona a_L^H dla wskaźnika a_L.

Skale dla wskaźników a_C i a_L cechujemy ponadto za pomocą ich rozkładów w pewnej specjalnej rodzinie modeli stochastycznych. Rodzina ta została zaproponowana w pracy Bondarczuka i in. (1992).

Każdy z modeli rodziny generuje podziały $y_1, \ldots, y_s$ mandatów przy ustalonym podziale $x_1, \ldots, x_s$ uzyskanych głosów. W pierwszym z nich, podziałem mandatów rządzi pewien rozkład wielomianowy $W(p, M)$, nazwany tu modelem losowej proporcjonalności. Mandaty losuje się M-krotnie ze zwracaniem według wektora prawdopodobieństw $p = \frac{1}{N}(x_1, \ldots, x_s)$, tworząc realizację $y_1, \ldots, y_s$. Po jej wyznaczeniu obliczane są wartości wskaźników a_C i a_L. Powtarzając wielokrotnie całą opisaną wyżej procedurę otrzymujemy symulacyjny rozkład każdego z tych wskaźników. Co więcej (Bondarczuk i in. (1992)), wartość oczekiwana i wariancja są wyrażone explicite za pomocą $(x_1, \ldots, x_s)$ i M:

$$E(a_C) = 0, \tag{5}$$

$$D^2(a_C) = \frac{1}{M}\left(\sum_{i=1}^{s} \frac{x_i}{N}\alpha_i^2 - 1\right), \tag{6}$$

gdzie

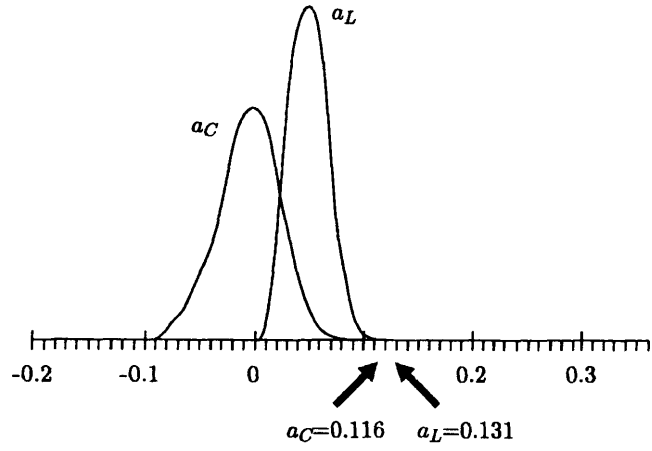
$$\alpha_i = 2 - 2\frac{1}{N}\left(\sum_{j=1}^{i-1} x_j - x_i\right), \quad i = 1, \ldots, s, \tag{7}$$

a rozkład wyrażenia $\sqrt{M}a_C$ dąży wraz z M do rozkładu normalnego.

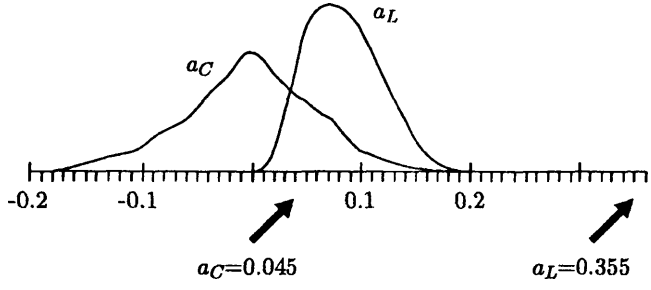
Kwantyle rozkładów a_C i a_L w modelu losowej proporcjonalności służą oczywiście do cechowania skali dla tych wskaźników. Jeśli a_L przekracza kwantyl rzędu 0.95 (lub $|a_C|$ przekracza kwantyl rzędu 0.975), można odrzucać hipotezę "losowej proporcjonalności" na poziomie istotności 0.05 dla każdego wektora $(x_1, \ldots, x_s, M)$.

Na Rys. 3 pokazano rozkłady wskaźników a_C i a_L w modelu losowej proporcjonalności dla $x_1, \ldots, x_s, M$ dla danych (a) z Tab. 1. Okazuje się, że rozkład a_L w modelu losowej proporcjonalności skupia się na lewo od zaobserwowanej wartości $a_L = 0.131$, a rozkład wskaźnika a_C w tym modelu – na lewo od $a_C = 0.116$. Widzimy więc, że dane (a) z Tab. 1, wbrew pierwszemu wrażeniu, wykazują jednak dość znaczne odstępstwo od proporcjonalności, wyraźnie przewyższające to co bywa osiągane przy proporcjonalności losowej.

Rys. 4 przedstawia rozkłady wskaźników a_C i a_L w modelu $W(p, M)$ dla danych z Tab. 2. Kwantyl rzędu 0.95 w rozkładzie wskaźnika a_L jest tu blisko dwukrotnie większy niż taki sam kwantyl dla danych (a) z Tab. 1, ale



Rys. 3. Rozkłady wskaźników a_C i a_L dla danych (a) z Tab. 1.



Rys. 4. Rozkłady wskaźników a_C i a_L dla danych z Tab. 2.

zaobserwowana wartość wskaźnika $a_L = 0.355$ zdecydowanie go przewyższa. Rozkład wskaźnika a_C jest znacznie słabiej skupiony wokół zera niż w przypadku danych (a) z Tab. 1; zaobserwowana wartość wskaźnika $a_C = 0.045$ jest na tle tego rozkładu szczególnie bliska zeru. Podsumowując, odstępstwo od proporcjonalności jest więc silne (wyraźnie większe niż to co bywa osiągane przy losowej proporcjonalności), przy czym nie ma ono nic wspólnego z tendencją nadreprezentowania dużego (małego) elektoratu.

Dalsze wzorcowanie skali wskaźników a_C i a_L posłuży do bardziej precyzyjnej oceny, czy występuje tendencja do nadreprezentowania dużego (małego) elektoratu. W tym celu wprowadzamy takie modele stochastyczne, w których szanse zdobycia mandatu przez dany komitet wyborczy wzrastają wraz z x_i silniej niż proporcjonalnie. Są to rozkłady wielomianowe $W(p^{(k)}, M)$, gdzie k jest obraną liczbą nieujemną, przy czym

$$(8) \qquad p^{(k)} = (p_1^{(k)}, \ldots, p_s^{(k)}) = \left(\frac{x_i^{k+1}}{\sum_{j=1}^{s} x_j^{k+1}} \right), \quad i = 1, \ldots, s.$$

Przy k równym 0 uzyskamy w ten sposób model losowej proporcjonalności. Możemy też dopuścić $k < 0$; mamy wtedy do czynienia z tendencją odwrotną: szansa zdobycia mandatu maleje wraz z x_i. Przy $k = -1$ mamy $p_i^{(k)} = 1/s$, co oznacza, że poszczególne komitety wyborcze mają te same szanse na uzyskanie mandatów bez względu na liczby faktycznie uzyskanych głosów.

Każdy z modeli $W(p^{(k)}, M)$ generuje pewien łączny rozkład wskaźników a_C i a_L, które będziemy tu oznaczać $a_C^{(k)}$ i $a_L^{(k)}$. Jak udowodniono w pracy Bondarczuka i in. (1992), wskaźniki $a_C^{(k)}$ są asymptotycznie normalne, przy czym

$$E(a_C^{(k)}) = \sum_{i=1}^{s} p_i^{(k)} \alpha_i, \tag{9}$$

$$D^2(a_C^{(k)}) = \frac{1}{M}\left(\sum_{i=1}^{s} p_i^{(k)} \alpha_i^2 - (1 - E(a_C^{(k)})^2\right). \tag{10}$$

Ponadto w pracy tej pokazano, że wskaźniki $a_C^{(k)}$ są stochastycznie uporządkowane według wartości k. Mediany tych rozkładów są naturalnymi punktami wzorcowymi: zaobserwowanej wartości a_C możemy przyporządkować tę potęgę k, której trzebaby użyć, żeby otrzymać rozkład $a_C^{(k)}$ o takiej właśnie medianie. Posługując się wskaźnikiem k uzyskujemy pewną "uniwersalną" interpretację wyników wyborów, która abstrahuje od wartości $x_1, \ldots, x_s, M$.

Oprócz omówionej wyżej oceny punktowej parametru k, proponujemy również jego ocenę przedziałową w postaci przedziału ufności na obranym poziomie ufności α. W tym celu ustalamy takie k_G i k_D , dla których zaobserwowana wartość wskaźnika a_C stanowi, odpowiednio, kwantyl rzędu $\alpha/2$ lub $1 - \alpha/2$ w rozkładzie wskaźnika $a_C^{(k)}$ dla $k = k_G$ i $k = k_D$. Im bardziej skupione są rozkłady wskaźników $a_C^{(k)}$, tym węższe uzyskuje się przedziały ufności, czyli tym większa jest dokładność naszego wzorcowania.

Posługiwanie się parametrem k pozwala na porównywanie wyników wyborów przy różnych ciągach $(x_1, \ldots, x_s, M)$ pod warunkiem, że w obu przypadkach mamy do czynienia z tendencją do nadreprezentowania większych (mniejszych) elektoratów. Rozłączność przedziałów ufności dla dwóch porównywanych zestawów danych $\{(x_i, y_i)\}$ pozwoliłaby wskazać zestaw o silniejszej tendencji do faworyzowania komitetów wyborczych, na które została oddana większa liczba głosów.

W przypadku danych (a) z Tab. 1 na podstawie $a_C = 0.116$ otrzymujemy $k = 0.60$ i $(k_D, k_G) = (0.35, 0.85)$, a na podstawie a_L mamy $k = 0.63$ oraz $(k_D, k_G) = (0.35, 0.90)$. Potwierdza to zdecydowanie poprzednią diagnozę:

istnieje odstępstwo całkowicie wyjaśniane tendencją do nadreprezentowania dużych komitetów, gdyż zgodność wyników uzyskanych na podstawie a_C i a_L jest wyjątkowo dobra.

W przypadku danych z Tab. 2 na podstawie a_C otrzymujemy $k = 0.22$ i $(k_D, k_G) = (-0.30, 0.50)$, a na podstawie a_L mamy $k = 2.5$ oraz $(k_D, k_G) = (1.00, 4.00)$. Rozłączność przedziałów (-0.30,0.50) i (1.00,4.00) stanowi kolejny bardzo silny argument, że nie ma tendencji do nadreprezentowania. Zatem odstępstwo od proporcjonalności jest duże, ale ma zupełnie inny charakter niż nadreprezentowanie dużych (małych) elektoratów.

5. Analiza proporcjonalności wyników wyborów do Sejmu. Nasze badania prowadziliśmy na pięciu zestawach danych. Na pierwszy z nich składa się 111 par liczb – po jednej parze dla każdego komitetu wyborczego biorącego udział w wyborach do Sejmu. Pierwsza liczba z pary jest równa liczbie wszystkich ważnie oddanych głosów zdobytych przez dany komitet w całym kraju. Druga liczba – liczbie zdobytych miejsc w Sejmie.

Drugi zestaw danych różni się od pierwszego jedynie tym, że liczby zdobytych miejsc w Sejmie zostały pomniejszone o liczby mandatów uzyskanych z list ogólnopolskich.

Zestaw numer 3 powstał z pierwszego w wyniku agregacji danych opisanej w p. 2. Dane te dotyczyły sześciu klas komitetów wyborczych. Pierwszy element z pary był brany z 3-iej kolumny Tab. 1, zaś drugi element – z odpowiedniego wiersza "a" piątej kolumny. Ten właśnie zestaw służył jako przykład w poprzednich rozdziałach tej pracy (dane (a) z Tab. 1).

W zestawie numer 4 w miejsce elementu z wiersza "a" brano element z wiersza "b".

Ostatni, piąty zestaw danych powstał z zestawu nr 1 przez odrzucenie tych komitetów wyborczych, które w skali kraju nie uzyskały co najmniej 5% wszystkich ważnie oddanych głosów. Zestaw ten składał się z 9-ciu par liczb.

Dla każdego z tych zestawów danych policzono wartości wskaźników a_C i a_L, a także ich punkty Hamiltona oraz punktowe i przedziałowe oceny potęgi k. W przypadku wskaźnika a_L oceny wyznaczono w tych przypadkach, w których było to możliwe, to jest pominięto k_D wyznaczone na podstawie a_L w zestawach 1 i 2 oraz k i k_D w zestawie 5. Wartości k_D w zestawach 1, 2 i 5 nie mogły być wyznaczone, gdyż przy losowej proporcjonalności prawdopodobieństwo uzyskania wartości a_L przekraczającej wielkość zaobserwowaną tego parametru było w tych przypadkach większe niż 0.05. Z kolei wartość k w zestawie 5 nie mogła być wyznaczona, gdyż zaobserwowana wartość a_L była mniejsza od mediany przy losowej proporcjonalności.

Tab. 3. Wartości wskaźników odstępstwa od proporcjonalności oraz oceny potęg k.

Nr. zestawu	Wskaźniki odstępstwa od proporcjonalności	Punkt Hamiltona	k	k_D	k_G
1	a_C=0.108	0.0350	0.26	0.10	0.45
	a_L=0.175	0.0440	0.30	–	0.45
2	a_C=0.093	0.0010	0.20	0.12	0.35
	a_L=0.175	0.0480	0.20	–	0.40
3	a_C=0.116	-0.0004	0.28	0.18	0.41
	a_L=0.131	0.0020	0.28	0.16	0.41
4	a_C=0.094	-0.0020	0.26	0.14	0.39
	a_L=0.119	0.0030	0.23	0.10	0.37
5	a_C=-0.012	-0.0010	-0.10	-0.35	0.25
	a_L=0.035	0.0040	—	—	0.30

Porównując ze sobą wiersze 1 i 2, a następnie wiersze 3 i 4, widzimy, że listy ogólnopolskie miały znikomy wpływ na ocenę proporcjonalności. W dalszych rozważaniach możemy więc brać pod uwagę tylko zestawy 1, 3 i 5, gdyż zestaw 2 jest prawie identyczny z zestawem 1, a zestaw 4 z zestawem 3.

Największe i stosunkowo duże odstępstwo od proporcjonalności osiągane jest w zestawie 3. Zatem, ordynacja była raczej odległa od proporcjonalnej z punktu widzenia rozpatrywanych sześciu grup komitetów wyborczych; uprzywilejowując dość wyraźnie grupy złożone z komitetów o większym procencie głosów. Jeśli rozpatruje się zbiór wszystkich komitetów (zestaw 1) tendencja uprzywilejowania większych z nich występuje nadal, lecz ostępstwo od proporcjonalności jest mniejsze. Natomiast największe komitety rozpatrywane odrębnie (zestaw 5) mają reprezentację bardzo zbliżoną do idealnie proporcjonalnej, czyli w tym zakresie ordynacja pozwoliła dobrze odwzorować wolę wyborców. Zaobserwowana w zestawie 5 wartość $a_L = 0.035$ stanowi zaledwie kwantyl rzędu 0.02 rozkładu tego parametru przy losowej proporcjonalności (chociaż jest większa niż wartość 0.004 osiągana dla punktu Hamiltona).

Przeprowadzając podobną analizę dla innych wyborów (w innym kraju, w innym czasie etc.) możnaby dokonać porównań zastosowanych ordynacji wyborczych. Proponowane narzędzie mogłoby też być użyteczne przy opracowywaniu różnych wariantów przyszłej ordynacji wyborczej.

Bibliografia

[1] M. L. Balinski, H. P. Young, (1985), *The apportionment of representation*, W: H.P. Young, ed., *Fair Allocation*, American Mathematical Society Proceedings of Symposium in Applied Mathematics, Vol. 33, pp. 1–29.

[2] Z. Bondarczuk, T. Kowalczyk, E. Pleszczyńska, W. Szczesny, (1992), *Evaluating departures from fair representation*, Prace IPI PAN, Nr 716.

[3] A. Ciok, T. Kowalczyk, E. Pleszczyńska, W. Szczęsny, (1992), *Comparing methods of fair representation*, Prace IPI PAN, Nr 718.

[4] M. Janicki, (1991), *Ordynacja jest litościwa – Jak głosować żeby wygrać*, Polityka Nr 41, (12.10.1991).

Statistical analysis of the departure from proportionality in the 1991 election in Poland

Abstract

New statistical measures of departure from proportional representation, introduced in Bondarczuk, Kowalczyk, Pleszczyńska and Szczesny (1992), are applied to the results of the 1991 election to the Polish Parliament. A pair of indices is used to evaluate departures from proportionality between number of votes and number of elects. One index indicates general strength of the departures, the other one indicates to which extent they are due to overrepresentation increasing (or decreasing) with number of votes. The indices are corrected in order to eliminate the influence of the distribution of votes , number of votes and size of representation. The correction is based on intensive computer simulation.

The analysis of the election data showed significant overrepresentation of six parties (or group of parties forming election committees) with maximal numbers of votes. On the other hand, in the subpopulation of voters formed by these six parties only, the departure from proportionality is very small.

INSTYTUT PODSTAW INFORMATYKI PAN
UL. ORDONA 21
01-237 WARSZAWA
POLSKA