

УДК 519.22

## СТАТИСТИЧНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ І ТУРУ ПРЕЗИДЕНТСЬКИХ ВИБОРІВ В УКРАЇНІ 2019 РОКУ

Подрушняк В., Акбаш К.С.

*Анотація.* В статті проілюстровано статистичне дослідження щодо обробки результатів виборів в Україні 2019 року за допомогою статистичних методів обробки. Проведено перевірку нормальної розподіленості голосів по округам для кандидатів, побудовано кореляційні поля, проведений кластерний аналіз територіального розподілу голосів та аналіз узгодженості голосів між кандидатами методами кореляційного та кластерного аналізів.

**Ключові слова:** статистичний метод, кластерний аналіз, кореляційне поле, кореляційний аналіз

### STATISTICAL ANALYSIS OF THE RESULTS OF THE 1ST TOUR OF PRESIDENTIAL ELECTIONS IN UKRAINE, 2019

*Abstract.* The article illustrates the statistical survey on the processing of the results of elections in Ukraine in 2019 using statistical methods of processing. The analysis of the normal distribution of votes in constituencies for candidates is illustrated, correlation fields are constructed, cluster analysis of territorial distribution of votes is conducted and the analysis of the consistency of votes between candidates is carried out using correlation and cluster analysis methods.

**Keywords:** statistical method, cluster analysis, correlation field, correlation analysis.

**Постановка проблеми.** Для оцінки результатів виборів широко застосовуються методи математичної статистики. Статистика являє собою строгий науковий метод, що дозволяє зрозуміти дані, дійти до їх суті. Однак статистика може дати миттєву загальну картину даних — на основі доступної для сприйняття візуалізації або числового узагальнення — незалежно від кількості спостережень чи одиниць даних. Окрім узагальнення даних, важливим завданням статистики є формулювання логічних висновків та передбачення залежностей між змінними. Застосування нормального розподілу дозволяє виявляти різні аномалії і в суспільному житті – наприклад фальсифікації на виборах.

**Аналіз досліджень і публікацій.** Президентські вибори в Україні проходять кожні 5 років. Результати розрахунків голосів у %, а також

аналітичні результати виборів висвітлюються на сайті Центральної виборчої комісії [1].

**Метою роботи** є аналіз статистичними методами результатів I туру президентських виборів 2019 року.

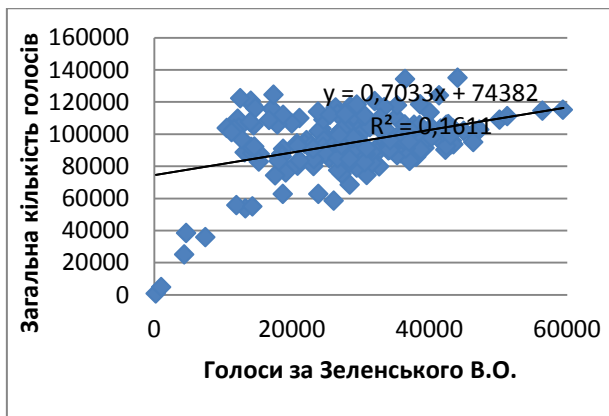
Вибори - одна з найдавніших форм прямого народовладдя, відома ще з часів так званої «вічової демократії». Нині вибори на загальнодержавному та місцевому рівні проводяться у більшості країн світу, за винятком теократичних монархій [3].

У даному дослідженні розглядаються результати I туру президентських виборів України. За результатами першого туру до другого туру вийшли Володимир Зеленський та Петро Порошенко. 7 квітня 2019 року ЦВК офіційно оголосила результати першого туру президентських виборів: Зеленський — 30,24 % голосів, Порошенко — 15,95 %.

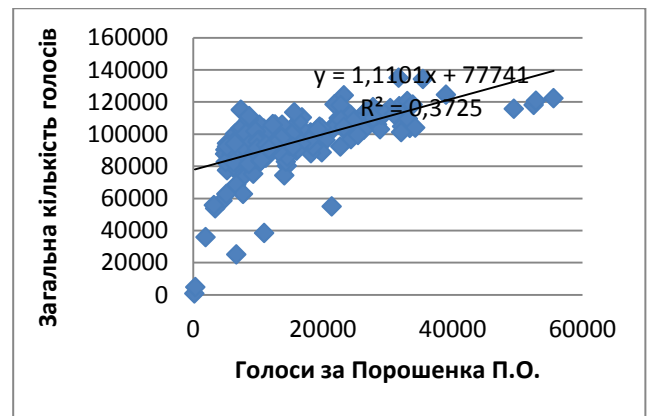
Розглянемо результати виборів (кількість голосів подана у %) кожного кандидата по округам. Загальна кількість округів – 200. Орієнтовна кількість виборців – 30225048.

Ми знайшли загальну кількість виборців по кожному округу та кількість голосів за певного кандидата по кожному округу. Таким чином, побудуємо кореляційні поля для кожного кандидата, де по осі  $y$  відкладемо загальну кількість виборців по округу, а по осі  $x$  - кількість голосів за певного кандидата по цьому округу. Кореляційний аналіз дає змогу встановити відповідність або взаємозв'язок між ознаками, що виявляється при таких великих спостереженнях зміни середньої величини однієї ознаки залежно від значення іншої ознаки.

У нашому випадку кореляційні поля будуть вказувати на рівномірність розподілу голосів по території України. Далі на графіках (рис. 1-2) наведені приклади кореляційних полів кандидатів, які пройшли до другого туру виборів.



**Рис.1.** Кореляційне поле результатів Зеленського В.О.



**Рис.2.** Кореляційне поле результатів Порошенка П.О.

Аналізуючи кореляційні поля голосів за кожного кандидата можна сказати, що умовно можна розділити ці поля на дві групи: 1 група – група з малою ймовірністю, яка вказує на малу залежність між двома величинами (кореляція  $R^2 \approx$  від 0 до 0,2, тобто  $R \approx$  від 0 до 0,44); 2 група – група з великою ймовірністю, яка вказує на залежність між двома величинами (кореляція  $R^2 \approx$  від 0,2 до 0,5, тобто  $R \approx$  від 0,45 до 0,7).

Отже, до 1 групи відносяться Богословська, Бондар, Ващенко, Вілкул, Габер, Журавльов, Каплін, Кармазін, Кива, Корнацький, Кривенко, Кошулинський, Купрій, Литвиненко, Ляшко, Морозов, Насіров, Риганов, Скоцик, Соловійов, Тарута, Шевченко І., Шевченко О., Бойко, Тимошенко, Зеленський. До 2 групи відносяться Безсмертний, Балашов, Богомолець, Данилюк, Наливайченко, Новак, Носенко, Петров, Смешко, Гриценко, Порошенко. Так як в нас є випадки, коли кореляція дуже близька до 0, проте це може свідчити лише про відсутність лінійної залежності, а не взагалі про відсутність будь-якого статистичного зв'язку.

Взагалі візуальний аналіз кореляційних полів дуже важливий. Він дає можливість виявити аномальні спостереження (округи, які різко відрізняються своїми результатами від інших) та проаналізувати рівномірність розподілу голосів по округам.

Шляхом аналізу даних нами було розглянуто результати виборів п'яти кандидатів, які отримали найбільшу кількість голосів за результатами

виборів, а саме Зеленського В.О., Порошенка П.О., Тимошенко Ю.В., Бойка Ю.А., Гриценка А.С..

Так як нормальний закон розподілу найбільш розповсюджений у практичних задачах обробки експериментальних даних, ми перевірили результати виборів кожного з п'яти кандидатів на нормальність.

Для цього використаємо модифікований критерій  $\chi^2$ . Спочатку сукупність вибірових змінних розбиваємо на 10 рівномірних інтервалів. Оптимальну кількість інтервалів розбиття визначаємо за формулою Стерджеса. Після визначення кількості інтервалів підраховуємо статистику критерія за формулою:

$$\chi^2 = \frac{k}{n} \sum_{i=1}^n m_i^2 - n.$$

де  $n$  – обсяг вибірки (в нашому випадку  $n = 200$ );  $m_i$  – кількість членів вибірки, які попали в  $i$  – й інтервал. Для кожного кандидата критичне значення береться з таблиці. Отже, для  $k = 10, \alpha = 0.01 \Rightarrow d_k(\alpha) = 18,852$ .

Потім будуємо границі інтервалів для кожного кандидата для відсоткових показників результатів голосування по округах. Запишемо отримані границі інтервалів в таблицю (табл.1.) і порахуємо кількість елементів вибірки для кожного інтервалу:

**Таблиця 1.** Границі інтервалів за результатами голосів за Зеленського В.О.

Класів 10		i	Границі інтервалів	$m_i$	$m_i^2$
Знач. коефіцієнтів		1	0-18,06	25	625
$c_1$	-1,2816	2	18,06-22,27	13	169
$c_2$	-0,8416	3	22,27-25,3	19	361
$c_3$	-0,5244	4	25,3-27,89	26	676
$c_4$	-0,2533	5	27,89-30,32	23	529
$c_5$	0	6	30,32-32,74	10	100
$c_6$	0,2533	7	32,74-35,34	15	225
$c_7$	0,5244	8	35,34-38,37	28	784
$c_8$	0,8416	9	38,37-52,58	20	400
$c_9$	1,2816	10	42,58-100	21	441
				<b>200</b>	<b>4310</b>

Аналогічні інтервальні розподіли знаходимо для інших чотирьох кандидатів, які по рейтингу голосів увійшли до п'ятірки найкращих.

Знаходимо статистику критерію та порівнюємо її з критичним значенням:

$\chi^2 = 15,5 < d_k \alpha = 18,852$  – за результатами голосів за Зеленського В.О.

$\chi^2 = 76,4 > d_k \alpha = 18,852$  – за результатами голосів за Порошенка П.О.

$\chi^2 = 23,3 > d_k \alpha = 18,852$  – за результатами голосів за Тимошенко Ю.В.

$\chi^2 = 130,4 > d_k \alpha = 18,852$  – за результатами голосів за Бойка Ю.А.

$\chi^2 = 80,8 > d_k \alpha = 18,852$  – за результатами голосів за Гриценка А.С.

Можна зробити висновок, що нульова гіпотеза про нормальність приймається лише за результатами розподілу голосів за Зеленського В.О. Аналогічний ефект спостерігався на минулих виборах для переможця Петра Порошенка.

Далі проведемо кластерний аналіз областей України за результатами виборів. Об'єктами класифікації виступатимуть області, а ознаками виступатимуть відсоткові результати голосів відданих за кожного претендента. Кластерний аналіз буде проведений з метою виявлення схожих регіонів за результатами голосування.

Для цього необхідно порахувати середні відсотки по кожному кандидату. Кандидати, які отримали менше, ніж 3% голосів ми не враховували. Тобто кластерний аналіз будемо проводити по наступним кандидатам: Зеленський – 30,24%, Порошенко – 15,95%, Тимошенко – 13,4%, Бойко – 11,68%, Гриценко – 6,91%, Смешко – 6,04%, Ляшко – 5,48%, Вілкул – 4,15%.

Утворилась таблиця по рядках якої будуть області ( 24 області + м.Київ + закордонні дільниці), а по стовпцях – кандидати (8 кандидатів), у комірках – відсотки голосів. Проведемо Ієрархічний кластерний аналіз за допомогою програми *SPSS*. Отримавши таблицю належності до кластеру, можна вибрати оптимальну кількість кластерів. В нашому випадку – 5 кластерів. В результаті кластеризації, всі дані було поділено на 5 кластерів, перший з яких

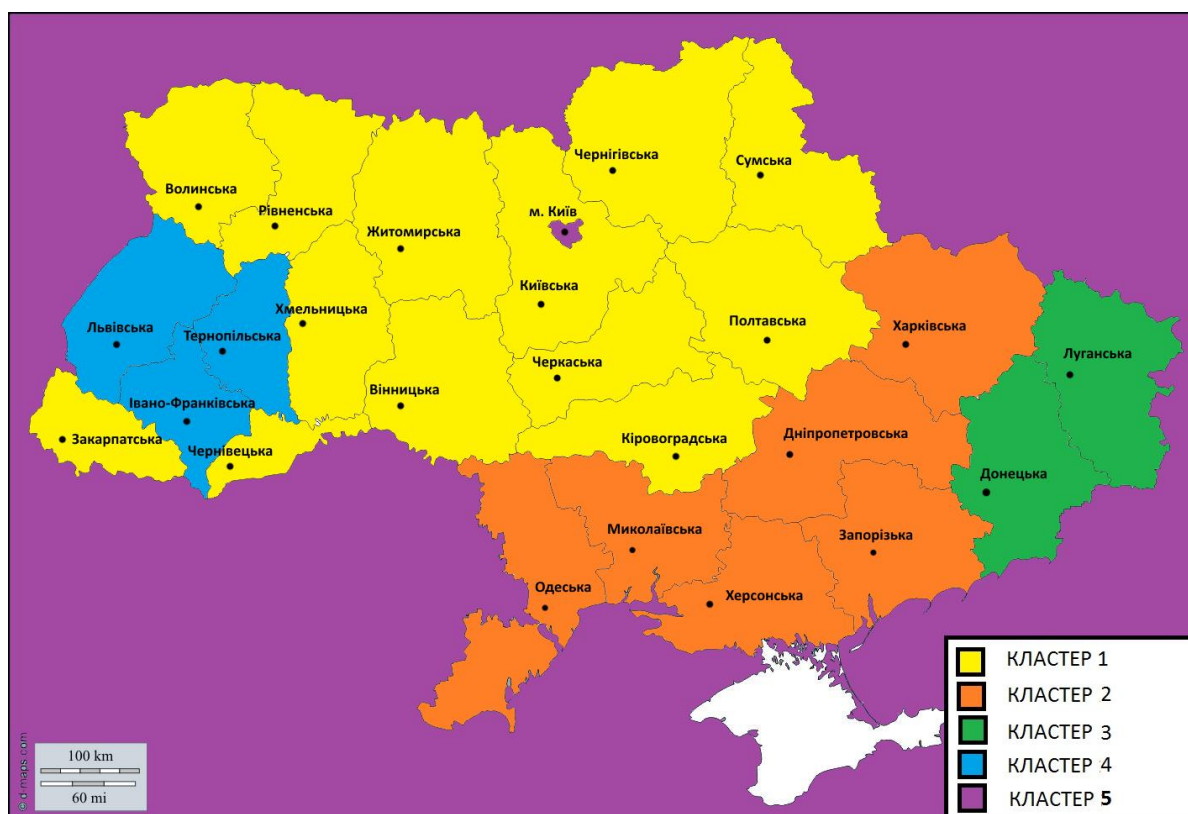
містив 50% даних, другий – 23,08%, третій – 7,69%, четвертий – 11,54%, п'ятий – 7,69% даних. Це можна побачити в таблиці 2.

**Таблиця 2.** Кластеризація областей України за результатами голосування

	Зеленський	Порошенко	Тимошенко	Бойко	Гриценко	Смешко	Ляшко	Вілкул
<b>1-Вінницька</b>	23,44	21,99	17,83	4,86	6,15	8,71	8,55	2,29
<b>1-Волинська</b>	21,74	18,84	19,9	3,64	9,03	6,89	9,9	1,09
<b>1-Житомирська</b>	27,75	15,95	17,48	7,21	5,67	8,35	9,84	1,72
<b>1-Закарпатська</b>	38,9	11,16	15,54	7,55	4,97	4,57	6,33	2,03
<b>1-Київська</b>	30,67	18,29	17,13	5,19	7,19	8,86	5,09	1,29
<b>1-Кіровоградська</b>	34,38	11,69	18,69	8,76	4,33	6,18	8,36	2
<b>1-Полтавська</b>	35,52	10,58	15,94	9,47	5,25	7,6	7,67	2,07
<b>1-Рівненська</b>	24,07	22,28	15,98	4,4	7,68	5,73	9,88	0,96
<b>1-Сумська</b>	32,45	10,39	16,33	11,67	5,33	8,1	8,41	2,31
<b>1-Хмельницька</b>	24,83	16,66	17,57	5,08	8,56	7,61	9,91	1,07
<b>1-Черкаська</b>	30,09	12,17	16,75	5,49	9,44	10,66	8,09	1,5
<b>1-Чернівецька</b>	31,08	13,95	19,75	9,08	5,09	5,15	7,66	1,11
<b>1-Чернігівська</b>	26,13	12,7	19,62	7,3	5,87	8,29	12,77	2,46
<b>2-Дніпропетровська</b>	45,21	8,37	8,6	12,48	3,31	4,77	2,87	10,71
<b>2-Запорізька</b>	39,77	8,74	8,25	18,84	2,96	4,57	3,73	9,37
<b>2-Миколаївська</b>	40,66	9,73	8,97	17,18	3,16	3,99	5,93	5,8
<b>2-Одеська</b>	40,82	9,11	8,18	22,19	2,82	2,91	3,76	5,9
<b>2-Харківська</b>	36,39	8,39	7,38	26,74	3,07	3,64	3,34	7,41
<b>2-Херсонська</b>	37,59	11,54	10,34	15,68	3,95	5,07	5,96	5,29
<b>3-Донецька</b>	23,85	15,61	3,54	36,66	1,2	2,21	2,45	10,66
<b>3-Луганська</b>	24,04	6,79	4,5	45,52	1,42	2,01	2,77	9,34
<b>4-Івано-Франківська</b>	16,13	21,05	22,66	1	15,09	3,83	4,14	0,3
<b>4-Львівська</b>	11,93	34,9	15,2	1,31	18,99	4,66	3,55	0,31
<b>4-Тернопільська</b>	14,72	24,34	18,97	1,54	16,87	5,08	7,49	0,37
<b>5-м. Київ</b>	27,11	25,53	12,9	6,26	9,1	9,62	1,42	2,19
<b>5-закордон</b>	26,01	38,93	4,21	3,91	10,64	6,62	0,33	1,8

Можна побачити, що до 1 кластеру відносяться такі області: Вінницька, Волинська, Житомирська, Закарпатська, Київська (але не м. Київ), Кіровоградська, Полтавська, Рівненська, Сумська, Хмельницька, Черкаська, Чернівецька, Чернігівська. До 2 кластеру – Дніпропетровська, Запорізька, Миколаївська, Одеська, Харківська, Херсонська. До 3 кластеру відноситься

лише 2 області – Донецька та Луганська. До 4 – Івано-Франківська, Львівська, Тернопільська. До 5 кластеру віднесли м. Київ та закордонні округи.



**Рис. 3.** Розбиття областей на кластери

На рис. 3. зображена розфарбована карта областей України ( що дає можливість краще оцінити результат роботи), в яких схожі показники голосів по всіх кандидатах.

Знайдемо середні значення у відсотках по кожному кластеру для кожного кандидата (Таблиця 3). Червоним кольором позначено найменший відсоток голосів кожного кандидата в певному кластері. Блакитним кольором – найбільший відсоток.

**Таблиця 3.** Середні значення голосів по кластерам для кандидатів

	Зелен-ський	Поро-шенко	Тимо-шенко	Бойко	Грице-нко	Смешко	Ляшко	Віл-кул
Кластер 1	29,31	15,12	17,57	6,90	6,50	7,43	8,65	1,68
Кластер 2	40,07	9,31	8,62	18,85	3,21	4,15	4,26	7,41
Кластер 3	23,94	11,20	4,02	41,09	1,31	2,11	2,61	10,00
Кластер 4	14,26	26,76	18,94	1,28	16,98	4,52	5,06	0,32
Кластер 5	26,56	32,23	8,55	5,08	9,87	8,12	0,87	1,99

Побудуємо таблицю по рядках якої будуть кандидати ( вісім кандидатів), а по стовпцях – назви областей (26 областей), у комірках – відсотки. За допомогою аналізу даних та інструменту аналізу «Кореляція» ми отримали таблицю узгодженості голосів між кандидатами ( Таблиця 4).

**Таблиця 4.** Кореляційний аналіз узгодженості голосів

	Зеле- нський	Поро- шенко	Тимо- шенко	Бойко	Гри- ценко	Сме- шко	Ляш- ко	Віл- кул
Зеленський	1							
Порошенко	-0,69	1,00						
Тимошенко	-0,38	0,14	1,00					
Бойко	0,30	-0,55	-0,75	1,00				
Гриценко	-0,75	0,77	0,49	-0,70	1,00			
Смешко	-0,10	0,22	0,53	-0,63	0,23	1,00		
Ляшко	-0,12	-0,20	0,72	-0,41	0,03	0,47	1,00	
Вілкул	0,51	-0,54	-0,81	0,84	-0,69	-0,58	-0,50	1,00

Розглянувши таблицю 4., можна зробити такі висновки: наявність прямого і достатньо великого зв'язку між результатами виборів кандидата Зеленського та іншими кандидатами не спостерігається, але існує обернений зв'язок з результатами голосів за Порошенка та Гриценка. Кореляційний аналіз результатів голосів за Порошенка вказує на прямий зв'язок з результатами кандидата Вілкула. Дивлячись на кореляцію результатів голосів за Тимошенко, можна сказати, що в неї дуже схожі результати з Ляшком, але абсолютно різні з Бойком і Вілкулом. Бойко має схожі результати з Вілкулом та геть різні з Гриценком. Таким чином матриця парних кореляцій дає уявлення про рівень підтримки між кандидатами.

**Висновки.** У статті проаналізовані методи статистичної обробки результатів виборів, а також продемонстрований аналіз кореляційних полів, кластерний аналіз територіального розподілу голосів та аналіз узгодженості голосів між кандидатами методами кореляційного та кластерного аналізів. В цілому можна сказати, що при перевірці нормальності було виявлено нормальну розподіленість голосів лише в Зеленського, що свідчить про те,



що існує певна регіональна специфіка та рівномірність розподілу голосів за цього кандидата, а нині вже президента України, на всій території країни.

#### **Список використаної літератури**

1. Вибори президента України. Центральна виборча комісія. [Електронний ресурс] - Режим доступу: <https://www.cvk.gov.ua/pls/vp2019/wp404.html>
2. Поняття та види виборів в Україні. [Електронний ресурс] - Режим доступу: <https://studfiles.net/preview/5424043/page:17/>
3. Основні види виборів. [Електронний ресурс] - Режим доступу: <http://ru.osvita.ua/vnz/reports/politolog/16237/>