3. Cimpoi M., Maji S., Kokkinos I., Mohamed S., Vedaldi A. Describing textures in the wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014. P. 3606–3613.

4. van der Walt S., Schönberger J. L., Nunez-Iglesias J. et al. scikit-image: image processing in Python. PeerJ. 2014. Vol. 2. Art. e453. DOI: https://doi.org/10.7717/peerj.453.

5. Paszke A., Gross S., Massa F. et al. PyTorch: an imperative style, high-performance deep learning library. Advances in Neural Information Processing Systems 32 (NeurIPS 2019). 2019. P. 8024–8035.

УДК 519.237.8

КЛАСТЕРИЗАЦІЯ КРАЇН ЗА РІВНЕМ ЕКОНОМІЧНОГО РОЗВИТКУ ТА ОЧІКУВАНОЮ ТРИВАЛІСТЮ ЖИТТЯ МЕТОДОМ K-СЕРЕДНІХ

Олександр Хлівний

Миколаївський національний аграрний університет, м. Миколаїв, Україна

Науковий керівник: О. О. Жебко,
асистент Миколаївський національний аграрний університет

Соціально-економічний розвиток держав є багатовимірним явищем, яке неможливо достовірно охарактеризувати одним рейтингом або окремим абсолютним показником. Обсяг національного виробництва свідчить про економічні можливості, однак сам по собі не описує стан здоров’я населення, тривалість життя чи доступність суспільних благ. Країни зі схожими показниками виробництва на одну особу можуть відрізнятися за соціальними результатами, а держави з різним економічним потенціалом – демонструвати близькі демографічні характеристики. Тому актуальним є використання методів багатовимірного аналізу, які дозволяють виявляти групи подібних країн без попереднього призначення їм нормативних категорій.

Кластерний аналіз групує спостереження за близькістю обраних характеристик і, на відміну від ранжування лише за ВВП, враховує кілька ознак одночасно. Оскільки результат залежить від вибору показників, нормування та складу вибірки, важливо оцінювати не лише склад груп, а й якість розбиття. Теоретичну основу алгоритму k-середніх закладено Дж. Мак-Квіном, а силуетний критерій запропоновано П. Руссеєм [5; 6].

Метою дослідження є визначення можливості виокремити інтерпретовані групи країн за поєднанням економічного та соціально-демографічного показників на невеликій міжрегіональній вибірці за 2023 рік. Дослідницьке питання сформульовано так: чи дозволяє стандартизована двовимірна модель k-середніх розрізнити три групи країн із помітно відмінними профілями ВВП на одну особу та очікуваної тривалості життя? Об’єкт дослідження – відібрані країни світу; предмет – структура їх подібності у просторі двох кількісних ознак. Робота має характер відтворюваного навчально-аналітичного обчислювального експерименту, а не репрезентативної класифікації всіх держав.

Інформаційну основу становлять два відкриті масиви даних за календарний 2023 рік. Економічна змінна – ВВП на одну особу за паритетом купівельної спроможності (ПКС) у постійних міжнародних доларах 2021 року, код Світового банку NY.GDP.PCAP.PP.KD [1]. Урахування ПКС зменшує спотворення порівняння, пов’язане з відмінностями рівнів цін між країнами. Соціальна змінна – очікувана тривалість життя при народженні для обох статей, років, за оцінками Відділу народонаселення Департаменту ООН з економічних і соціальних питань у виданні World Population Prospects 2024 [2]. На рівні країни вона відображає узагальнену поточну структуру смертності, а не гарантовану тривалість життя конкретної людини.

Значення обох показників звірено з відкритими табличними виданнями за 2023 рік, що відтворюють дані відповідних першоджерел [3; 4]. Для зіставності використовуються 12 країн, відібраних цілеспрямовано з різних географічних регіонів і діапазонів доходу: Норвегія, США, Німеччина, Японія, Польща, Туреччина, Китай, Бразилія, Україна, Південно-Африканська

Республіка, Індія та Нігерія. Вибірка забезпечує наявність різних поєднань ознак, але не є випадковою. Економічні значення округлено до найближчого міжнародного долара, демографічні – до 0,001 року; усі розрахунки нижче виконано саме за наведеними в таблиці округленими числами. Пропущених значень у відібраних спостереженнях немає.

Таблиця 1 - Вихідні дані та результат кластеризації 12 країн у 2023 р.

Країна	ВВП на особу за ПКС, міжнар. дол.	Очікувана тривалість життя, років	Кластер
Норвегія	94 365	83,308	I
США	74 352	79,304	I
Німеччина	63 140	81,378	I
Японія	47 387	84,712	I
Польща	43 673	78,633	II
Туреччина	35 069	77,156	II
Китай	22 687	77,953	II
Бразилія	19 080	75,848	II
Україна	15 917	73,423	II
ПАР	13 695	66,139	III
Індія	8 871	72,003	II
Нігерія	7 843	54,462	III

Таблиця 1 демонструє суттєві відмінності як у матеріальному виробництві на одну особу, так і в демографічному результаті. Максимум економічного показника серед спостережуваних країн становить 94 365 міжнародних доларів у Норвегії, мінімум – 7 843 долари в Нігерії; відношення між ними дорівнює приблизно 12,03. Найвищу очікувану тривалість життя має Японія (84,712 року), найнижчу – Нігерія (54,462 року), різниця дорівнює 30,250 року. Уже ці зіставлення вказують на значну неоднорідність сукупності. Водночас вони не дають відповіді, які країни є найбільш подібними з урахуванням одночасно двох вимірів.

Перед групуванням показники стандартизовано, оскільки їхні одиниці та числові масштаби принципово різняться. Без попереднього перетворення Евклідова відстань майже цілком визначалася б ВВП, вимірюваним десятками тисяч доларів, тоді як тривалість життя характеризується десятками років. Для кожної ознаки обчислено середнє арифметичне та стандартне відхилення з дільником n = 12. Стандартизоване значення утворює безрозмірну координату, а одна одиниця після перетворення відповідає одному стандартному відхиленню в межах обраної вибірки.

$$z_{ij} = (x_{ij} - \mu_j) / \sigma_j \qquad (1)$$

Тут x_{ij} – вихідне значення показника j для країни i; μ_j – середнє значення показника за 12 країнами; σ_j – його стандартне відхилення. У підготовленому наборі даних μ для ВВП становить 37 173,25 міжнародного долара, σ – 26 894,82 долара; для тривалості життя відповідні значення дорівнюють 75,360 та 7,959 року. Наприклад, для України стандартизований ВВП дорівнює (15 917 − 37 173,25) / 26 894,82 = −0,790, а стандартизована тривалість життя – (73,423 − 75,360) / 7,959 = −0,243. Отже, обидва значення для України перебувають нижче середніх саме в досліджуваній сукупності.

Основним методом обрано k-середніх (k-means), який розподіляє спостереження між k групами за мінімізацією суми квадратів відстаней від об’єктів до центрів своїх кластерів. Для двох стандартизованих координат використано звичайну евклідову метрику. Вибір цього алгоритму обумовлено невеликим обсягом даних, прозорою цільовою функцією, можливістю безпосереднього відображення результату на площині та простотою відтворення обчислень [5]. Він не вимагає наявності еталонних міток класів або попередньої експертної шкали «розвинена / нерозвинена країна».

$$J = \sum_{r=1}^{k} \sum_{i} \in C_r \|z_i - m_r\|^2 \rightarrow min \quad (2)$$

У формулі (2) C_r означає множину країн r-го кластера, z_i – двовимірний стандартизований вектор країни, m_r – центроїд групи, k – кількість груп. Для відтворюваності обчислювального експерименту застосовано Python та реалізацію KMeans із параметрами n_init = 50, random_state = 42 і початковим вибором центрів k-means++. Багаторазова ініціалізація зменшує ймовірність зупинки на невдалому локальному розбитті; серед результатів 50 запусків збережено варіант із найменшим значенням цільової функції. Оскільки назви кластерів програмно довільні, після обчислення їх пронумеровано від I до III за спаданням середнього ВВП на одну особу.

Як допоміжний інструмент оцінювання взято середній коефіцієнт силуету [6]. Для кожного об'єкта порівнюють $a(i)$ – середню відстань до інших країн свого кластера – та $b(i)$ – мінімальну середню відстань до країн іншого кластера. Додатні значення коефіцієнта свідчать про відносно вдале розміщення країни у групі, значення поблизу нуля – про межове положення, а від'ємні – про те, що вона в середньому ближча до іншої групи. Середній силует не оцінює соціально-економічну «якість» країни, а характеризує геометрію отриманого розбиття.

$$s(i) = [b(i) - a(i)] / max\{a(i), b(i)\} \quad (3)$$

Для кількості кластерів k = 2, 3, 4, 5 середні значення силуету становили відповідно 0,446; 0,431; 0,332; 0,256. Строго за максимумом цього критерію кращим є двокластерне розбиття. Проте результат із трьома групами поступається йому лише на 0,015 пункту та дозволяє змістовно відділити країни з особливо низькою очікуваною тривалістю життя від решти країн із помірним або низьким ВВП. Тому для головного дослідницького питання обрано k = 3 як компроміс між геометричною компактністю і деталізацією. Це аналітичне рішення, а не твердження про існування єдиного об'єктивно правильного числа соціально-економічних груп.

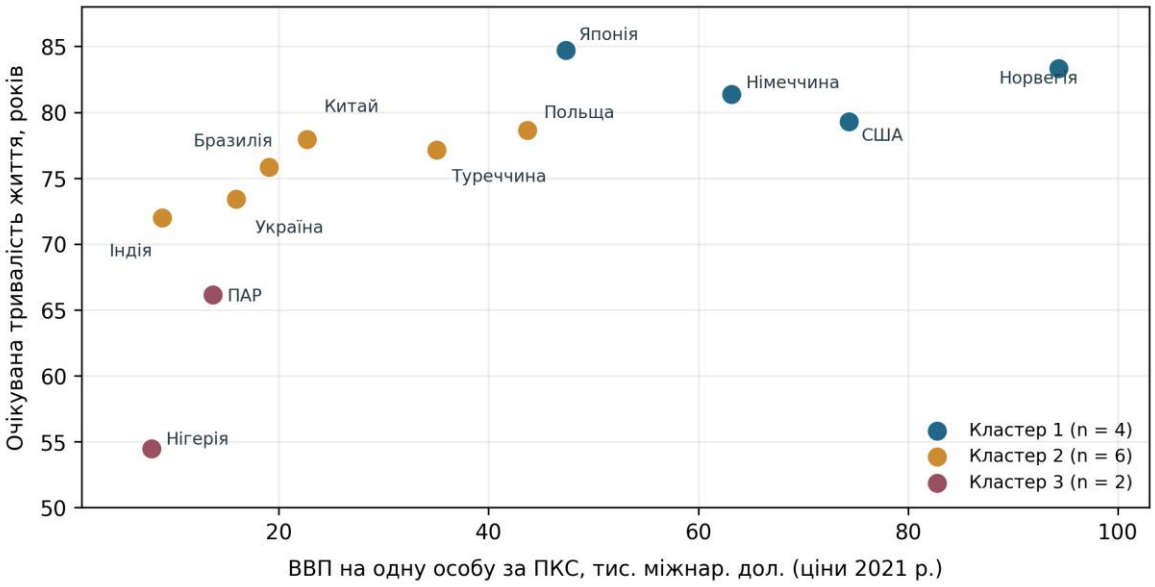


Рис. 1 - Розподіл 12 країн у просторі вихідних показників за результатами k-середніх, 2023 р.

Графічний результат на рис. 1 підтверджує, що метод розрізняє три профілі, але не створює абсолютно відокремлених блоків. Перший кластер об'єднує чотири країни – Норвегію, США, Німеччину та Японію – з відносно високим ВВП і високою тривалістю життя. Середній ВВП у цій групі становить 69 811 міжнародних доларів, а середня очікувана тривалість життя – 82,176 року. Водночас усередині групи є відмінності: Норвегія має найбільший у вибірці показник ВВП, тоді як Японія – найбільшу очікувану тривалість життя за помітно нижчого ВВП. Це ілюструє, чому двовимірна типологія інформативніша від простого ранжування за економічною ознакою.

Другий кластер включає шість країн: Польщу, Туреччину, Китай, Бразилію, Україну та Індію. Його центроїд у вихідних одиницях дорівнює 24 216 міжнародних доларів на одну особу та 75,836

року. Це найбільш численна і неоднорідна група. З одного боку, Польща з ВВП 43 673 долари та тривалістю життя 78,633 року розташована ближче до групи з високими значеннями, ніж більшість сусідів у кластері. З іншого боку, Індія має 8 871 долар на особу, але її очікувана тривалість життя дорівнює 72,003 року. Таким чином, кластер об'єднує не тотожні країни, а об'єкти з порівняно близькими двовимірними координатами після стандартизації.

Третій кластер утворюють Південно-Африканська Республіка та Нігерія. Їхній середній ВВП на одну особу дорівнює 10 769 міжнародним доларам, середня очікувана тривалість життя – 60,301 року. Важливо, що Нігерія відокремлюється переважно внаслідок дуже низького демографічного показника: 54,462 року порівняно з 66,139 року у ПАР. Відстань між ними за тривалістю життя становить 11,677 року, тобто ця група також не є внутрішньо однорідною. Тому її коректніше називати кластером з особливо низькими значеннями показника здоров'я населення, а не автоматично «найменш розвиненими державами» за сукупністю всіх можливих критеріїв.

Порівняння центроїдів показує, що середній ВВП першого кластера перевищує середній ВВП другого на 45 595 міжнародних доларів, а середня тривалість життя – на 6,340 року. Розрив між другим і третім кластерами за середньою тривалістю життя дорівнює 15,536 року; за ВВП – 13 447 доларів. Найбільше групи II та III відрізняються саме демографічною координатою. З огляду на стандартизацію обидві змінні мали однакову дисперсію на вході, отже перевага соціального показника в цьому конкретному розмежуванні виникла з фактичної конфігурації країн, а не через обрану одиницю вимірювання.

Для k = 3 значення суми внутрішньокластерних квадратів відстаней J становить 4,673 у стандартизованих одиницях у квадраті. Внески кластерів I, II та III дорівнюють відповідно 1,881; 1,692; 1,100. Середній силует усіх 12 спостережень дорівнює 0,431, що вказує на помірну виокремленість груп. Середні силуети першої, другої та третьої груп становлять 0,434; 0,512; 0,180. Особливо обережно слід трактувати третій кластер: для ПАР індивідуальний силует дорівнює −0,106. Отже, розподіл цієї країни перебуває поблизу межі двох груп, і під час зміни набору спостережень або ознак він може змінитися.

Окремої уваги заслуговує положення України. Її показники становлять 15 917 міжнародних доларів на одну особу та 73,423 року, що відповідає стандартизованим координатам приблизно (−0,790; −0,243). Україна віднесена до другого кластера разом із Бразилією, Китаєм, Туреччиною, Польщею та Індією. Це означає тільки відносну близькість за двома показниками за 2023 рік. Таке групування не дає підстав ототожнювати рівень розвитку перелічених економік, робити висновки про їхню інституційну подібність або приписувати виявлені відмінності окремим реформам, політичним рішенням чи подіям.

Практична цінність підходу – швидке формування порівняльних груп для наступного поглибленого дослідження. Замість довільного вибору країн-орієнтирів аналітик отримує перелік близьких за заданими координатами держав і надалі перевіряє їх за нерівністю доходів, освітою, доступом до медицини та зайнятістю. Такий алгоритм придатний для навчальних і прикладних міжнародних порівнянь.

Межі отриманого результату визначаються насамперед малим цілеспрямованим набором спостережень. Дванадцять країн становлять лише ілюстративний зріз різноманіття світу. Додавання інших держав може зрушити центроїди та змінити межі груп навіть без зміни самих показників раніше відібраних країн. З огляду на це некоректно поширювати склад отриманих кластерів на всі економіки або вважати три групи усталеною міжнародною класифікацією. Не проводилися ні статистичне вибіркове оцінювання параметрів генеральної сукупності, ні перевірка причинних гіпотез.

Другим обмеженням є вузький набір змінних. ВВП за ПКС на одну особу характеризує середній обсяг виробництва, але не розподіл доходів, економічну безпеку домогосподарств чи екологічні наслідки розвитку. Очікувана тривалість життя акумулює вплив багатьох факторів, проте сама собою не дозволяє встановити якість медичної допомоги або безпосередню роль доходів у смертності. Крім того, в розрахунку кожна країна отримала однакову вагу незалежно від чисельності населення. Це

відповідне рішення для кластеризації держав як окремих об’єктів, але не для аналізу становища «середньої людини» у світі.

Третє обмеження стосується датування та ревізій статистики. ВВП для 2023 року взято з оприлюдненого у 2026 році перерахованого ряду Світового банку, а демографічний показник – із редакції оцінок ООН 2024 року, тому значення можуть відрізнятися від архівних публікацій та від показників інших міжнародних організацій. Порівняння зроблено саме за роком, до якого належать спостереження, а не за роком публікації статистичних файлів. Усі вихідні числа, одиниці вимірювання, процедуру округлення й параметри алгоритму наведено явно, що дає можливість перевірити та повторити обчислення.

Порівняння обмежене одним роком: воно не показує, чи є склад кластерів стійким у часі. Інший спосіб масштабування, інші ваги показників або ширша вибірка можуть змінити групування. Наведене розбиття слід трактувати як результат конкретно визначеної аналітичної процедури, а його часову стабільність перевіряти окремо.

Отже, за результатами обчислювального експерименту на даних 2023 року для 12 країн підтверджено можливість отримати зрозумілу трьохкластерну типологію у двовимірному просторі ВВП за ПКС на одну особу та очікуваної тривалості життя. Перший кластер характеризується високими середніми значеннями обох ознак (69 811 доларів; 82,176 року), другий – проміжним профілем (24 216 доларів; 75,836 року), третій – низькими значеннями, особливо тривалості життя (10 769 доларів; 60,301 року). Ці числа є описом конкретної вибірки, а не універсальними порогами рівня розвитку.

Оцінка середнього силуету 0,431 показала помірну чіткість розбиття $k = 3$, причому силует для $k = 2$ є дещо вищим (0,446). Тому обрана кількість груп виправдана насамперед аналітичною інтерпретацією, а не формальним максимумом критерію. Виявлена межовість ПАР, неоднорідність проміжної групи та залежність від відбору країн підкреслюють необхідність обережних висновків. Подальшим логічним кроком є перевірка результату на ширшій вибірці та кількох послідовних роках, однак у межах цієї роботи поставлене вузьке дослідницьке питання вирішено на основі реальних перевірюваних даних і прозорих розрахунків.

Список використаних джерел

1. World Bank. World Development Indicators: GDP per capita, PPP (constant 2021 international $), indicator NY.GDP.PCAP.PP.KD. URL: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD (дата звернення: 14.10.2026).

2. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2024. New York: United Nations, 2024. URL: https://www.un.org/development/desa/pd/world-population-prospects-2024 (дата звернення: 14.10.2026).

3. Nations Inquiry. GDP per Capita (PPP), 2023: World Bank Group data [електронний набір даних; редакція 2026 р.]. URL: https://nationsinquiry.com/datasets/gdp-per-capita/2023 (дата звернення: 14.10.2026).

4. Nations Inquiry. Life Expectancy at Birth, 2023: UN DESA Population Division data [електронний набір даних]. URL: https://nationsinquiry.com/datasets/life-expectancy/2023 (дата звернення: 14.10.2026).

5. MacQueen J. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967. Vol. 1. P. 281–297. URL: https://digicoll.lib.berkeley.edu/record/113015 (дата звернення: 14.10.2026).

6. Rousseeuw P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987. Vol. 20. P. 53–65. DOI: 10.1016/0377-0427(87)90125-7.