

РОЗДІЛ 8. МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ
ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В ЕКОНОМІЦІКЛАСТЕРНИЙ АНАЛІЗ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ДЕТЕРМІНАНТ ЗДОРОВ'Я
CLUSTER ANALYSIS OF SOCIAL AND ECONOMIC DETERMINANTS OF HEALTH

Кластеризація даних є одним з найпопулярніших методів пошуку подібності статистичних даних в одному кластері даних та водночас відмінності даних в інших кластерах даних. Використання даного методу обумовлена кількістю статистичних даних які використовуються в процесі дослідження. Метою дослідження є здійснення кластерного аналізу для розподілу країн на групи за показниками: коефіцієнтом людської нерівності, коефіцієнтом Джині, нерівністю в очікуваній тривалості життя, очікуваній тривалості життя при народженні, валовим національним доходом, індексом очікуваної тривалості життя з поправкою на нерівність, загальною кількістю населення, міським населенням, індексом освіти, населенням віком від 15 до 64 років. Методика дослідження ґрунтується на застосуванні методів аналізу даних з відкритих джерел за допомогою програмного забезпечення такого як Statgraphics Centurion та Microsoft Excel. У процесі дослідження використовувались методи компаративного аналізу, систематизації, логічного узагальнення, бібліометричного аналізу (з використанням інструментарію SciencioPy), кластерного та дискримінаційного аналізу (з використанням інструментарію Statgraphics Centurion). Результати. Виявлено за допомогою формули Стерджеса необхідну кількість кластерів розподілу даних, також підтверджено оптимальну кількість кластерів за допомогою проміжків агломерації. Створено списки країн, що входять до кожного кластеру. Розроблено регресійну модель інструментарієм дискримінаційного аналізу, за допомогою якої визначено вплив кожного індикатора в процедурі кластеризації даних. Значущість дискримінаційних функцій обґрунтовано статистикою Лямбда Уїлкса та рівнем значущості P-value, розрахованими за допомогою інструментарію Statgraphics Centurion.

Ключові слова: кластерний аналіз, коефіцієнт людської нерівності, коефіцієнт Джині, метод Уорда, валовий національний дохід, загальна кількість населення, індекс освіти.

Data clustering is one of the most popular methods of search based on machine learning in the blind, the similarity of statistics in one data cluster and at the same time the differences in data in other data clusters. The use of this method is due to the amount of statistics used in the research process and the high speed of such analysis. Finding similar countries by type of development will make it possible to identify those statistics that give intra-cluster similarity of data, the difference between the data between clusters. The aim of the study is to find similar groups of countries that can be attributed to each other and determine the strength of the impact of each statistical indicator on the creation of a group of countries. The research methodology is based on the use of open-source data analysis techniques using software such as Statgraphics Centurion and Microsoft Excel. The research used methods of comparative analysis, systematization, logical generalization, bibliometric analysis (using SciencioPy tools), cluster and discriminant analysis (using Statgraphics Centurion tools). Results. Work on the analysis of recent publications on cluster analysis methods according to the Scopus scientometric database has generated a cloud of keywords that help to see the scope of cluster analysis methods in the scientific world. Analyzed the data of Human Development Index (HDI) statistical databases and took the 10 most relevant indicators in the opinion of the authors. The required number of clusters for the data was identified using the Sturges formula. Lists of countries included in each cluster have been created. The determined regression formula of the discriminant analysis function with its help is determined in the influence of each indicator on the created data cluster. The significance of discriminant functions is substantiated by Lambd Wilks and the significance level of P-value calculated using the Statgraphics Centurion toolkit. The results of cluster distribution can be used in the process of state development to find the optimal static values to which the development of the state should be directed. To make the transition of underdeveloped countries to more developed groups. The obtained data will be used for further in-depth analysis of data and finding new patterns in the development of the world.

Key words: Cluster analysis, human inequality rate, Gini coefficient, Ward method, gross national income, total population, education index.

УДК 330.4+338.1

DOI: <https://doi.org/10.32843/bses.75-16>**Койбічук В.В.**¹к.е.н., доцент,
доцент кафедри економічної
кібернетики
Сумський державний університет**Дрозд С.А.**²аспірант
Сумський державний університет**Koibichuk Vitaliia**

Sumy State University

Drozd Serhii

Sumy State University

Постановка проблеми. На нашій планеті існує багато країн, кожна країна відрізняється за рівнем розвитку економіки, суспільства, екологічної безпеки, системи охорони здоров'я та багатьох інших підсистем, які функціонують в державах. Всі країни обрали свій шлях розвитку, але всі вони прагнуть покращення свого стану в порівнянні з тим станом, який вже є. Завдяки відкритим статистич-

ним даним маємо можливість сформулювати схожі за розвитком країн в групі та визначити, які фактори та з якою силою впливають на їх об'єднання в групі.

Аналіз останніх досліджень і публікацій. Актуальність проблематики дослідження підтверджується великою кількістю вітчизняних та закордонних публікацій науковців. Так за запитом

¹ ORCID: <http://orcid.org/0000-0002-3540-7922>² ORCID: <https://orcid.org/0000-0002-0716-3078>

TITLE-ABS-KEY (cluster AND data AND analysis) пошуковою системою наукометричної бази даних Скопус отримано перелік статей з 197 752 документів з 1932 по 2022 рік.

Велика кількість науковців проводить дослідження з використанням методів кластерного аналізу. При огляді досліджень про кластерний аналіз вітчизняних вчених є декілька важливих праць, які потрібно зазначити. Яловенко О.В., Фетісов В.С. [1] провели роботу з розкриття поняття та виділення переваг кластерного аналізу. Жигайло О.М. та Борис В.В. [2] проаналізували та реалізували за допомогою інтелектуального аналізу даних методи кластерного аналізу та створили авторську класифікацію. Такі дослідники як Тюрин А.Г., Зуев І.О. [3] провели аналіз методів кластерного аналізу та виявили переваги та недоліки кожного з методів. Серга Є.Н. [4] описує новий метод універсальний ітераційний метод кластеризації даних. Дронов С.В., Шеларь А.Ю. [5] зробили детальне пояснення методів латентної кластеризації даних. Монс О.А., Янов Ю.О., Безпалій І.О [6] модифікували підхід до кластеризації даних з невідомою наперед кількістю кластерів. Серед іноземних науковців виділяємо наступних вчених та їх праці. Ельгаміфар Е., Відаль Р. [7] вивчають алгоритм, що називається кластеризацією розрідженого підпростору. Мерве Д., Енгельбрехт А.

[8] запропонували два нових підходи для використання кластерних даних в методі К-середніх для посіву початкового рою. Сандіп Р., Санджай Дж., Раджеш К. [9] виконали роботу з огляду методу та алгоритму методу рою частинок та визначили переваги методу.

Постановка завдання. Метою дослідження є отримання з модельованого розподілу країн по групах, за допомогою соціально-економічних показників. Та проведення аналітичної роботи по виявленню впливаючих факторів розподілу кластерів на групи. Виявлення сили впливу кожного окремого індикатора за допомогою проведення дискримінантного аналізу.

Виклад основного матеріалу дослідження. Інформаційною базою для кластерного аналізу використано офіційні звіти про рівень людського розвитку (Human Development Index (HDI)) програми Організації Об'єднаних Націй (ООН) [10] за 2019 рік для 137 країн.

З метою проведення кластерного аналізу використано такі показники, як: коефіцієнт людської нерівності (K1), коефіцієнт Джині (K2), нерівність в очікуваній тривалості життя (K3), очікувана тривалість життя при народженні (K4), валовий національний дохід (K5), індекс очікуваної тривалості життя з поправкою на нерівність (K6), загальна кількість населення (K7), міське населення (K8),

Таблиця 1

Вхідні дані

HDI rank	Країна	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
1	Норвегія	6,0	27,0	3,0	82,4	66 494	0,931	5,4	82,6	0,93	3,50
2	Ірландія	7,2	32,8	3,4	82,3	68 371	0,926	4,9	63,4	0,922	3,20
2	Швейцарія	6,8	32,7	3,5	83,8	69 394	0,947	8,6	73,8	0,9	5,70
4	Ісландія	5,6	26,8	2,4	83,0	54 682	0,946	0,3	93,9	0,926	0,20
6	Німеччина	7,9	31,9	3,8	81,3	55 314	0,908	83,5	77,4	0,943	54,00
7	Швеція	6,5	28,8	2,9	82,8	54 508	0,938	10	87,7	0,918	6,20
8	Австралія	7,9	34,4	3,7	83,4	48 085	0,940	25,2	86,1	0,924	16,30
8	Нідерланди	6,9	28,5	3,1	82,3	57 707	0,928	17,1	91,9	0,914	11,00
10	Данія	6,0	28,7	3,6	80,9	58 662	0,903	5,8	88	0,92	3,70
...
179	Ємен	30,9	36,7	24,7	66,1	1 594	0,534	29,2	37,3	0,35	16,90
181	Мозамбик	30,7	54,0	29,8	60,9	1 250	0,441	30,4	36,5	0,395	16,00
182	Буркіна-Фасо	29,5	35,3	32,0	61,6	2 133	0,435	20,3	30	0,312	10,80
182	Сьєрра-Леоне	34,5	35,7	39,0	54,7	1 668	0,326	7,8	42,5	0,406	4,40
184	Малі	32,4	33,0	36,7	59,3	2 269	0,383	19,7	43,1	0,286	9,90
185	Бурунді	29,6	38,6	28,5	61,6	754	0,457	11,5	13,4	0,417	6,00
185	Південний Судан	36,0	46,3	36,2	57,9	2 003	0,372	11,1	19,9	0,307	6,10
187	Чад	37,4	43,3	40,9	54,2	1 555	0,311	15,9	23,3	0,288	8,10
188	Центральноафриканська Республіка	41,3	56,2	40,1	53,3	993	0,307	4,7	41,8	0,353	2,50
189	Нігер	27,4	34,3	30,9	62,4	1 201	0,451	23,3	16,5	0,249	11,10

Джерело: побудовано авторами

індекс освіти (K9), населення віком від 15 до 64 років (K10). Фрагмент початкових вхідних даних подано в таблиці 1, де значення індикатора HDI відповідає рангу країни за розподілом за індексом людського розвитку [10].

Вхідні дані містять показники, що виміряні в різних шкалах, є як відносні показники (коефіцієнти та індекси), так і абсолютні, тому для коректності подальшого дослідження та високої якості та адекватності розрахунків, їх потрібно стандартизувати. Процедура стандартизації реалізована інструментарієм програмного забезпечення Statgraphics Centurion.

Для визначення оптимальної кількості кластерів використано формулу Стерджеса:

$$k = 1 + [3.322 \lg N], \quad (1)$$

де k – кількість кластерів, де N – обсяг загальної кількості країн (дорівнює 137). За результатами обчислення $k = 8$, що визначає оптимальну кількість розподілу країн на кластери.

Крім того, оптимальну кількість кластерів підтверджено протоколом агломерації (рис. 1). Відсутність характерних ривків між точками графіку дає можливість стверджувати, що обрана кількість кластерів для набору даних є правильною.

На графіку відстані агломерації видно нормальний розподіл проміжків (рис. 1).

Методом побудови кластерів обрано метод Уорда. Саме цей метод використано, оскільки він, на відміну від інших методів кластерного аналізу (метода одиничного зв'язку, метода повного зв'язку, центроїдного методу, методу медіан), використовує методи дисперсійного аналізу. В якості відстані між кластерами використовується при-

рошення суми квадратів відстаней об'єктів до центра кластера, що отримуються в результаті їх об'єднання (2). На рисунку 2 представлено кінцевий варіант розподілу кластерів за 137 країнами.

$$V_k = \sum_i \sum_j^p (x_{ij} - \bar{x}_{jk})^2, \quad (2)$$

де k – номер кластера, i – номер об'єкта, j – номер ознаки, p – кількість ознак, що характеризують кожен об'єкт, n_k – кількість об'єктів у k -му кластері.

Кількісній та процентній розподіл країн представлено в таблиці 2.

Таблиця 2

Числова характеристика кластерного розподілу

Кластер	Кількість	Процент
1	25	18,25
2	11	8,03
3	25	18,25
4	22	16,06
5	2	1,46
6	10	7,30
7	28	20,44
8	14	10,22

Джерело: розраховано авторами засобами програмного інструментарію Statgraphics Centurion

Таким чином, до першого кластеру увійшли країни Європи та країни з розвинутою економікою, а саме: Норвегія, Ірландія, Швейцарія, Ісландія, Німеччина, Швеція, Австралія, Нідерланди,

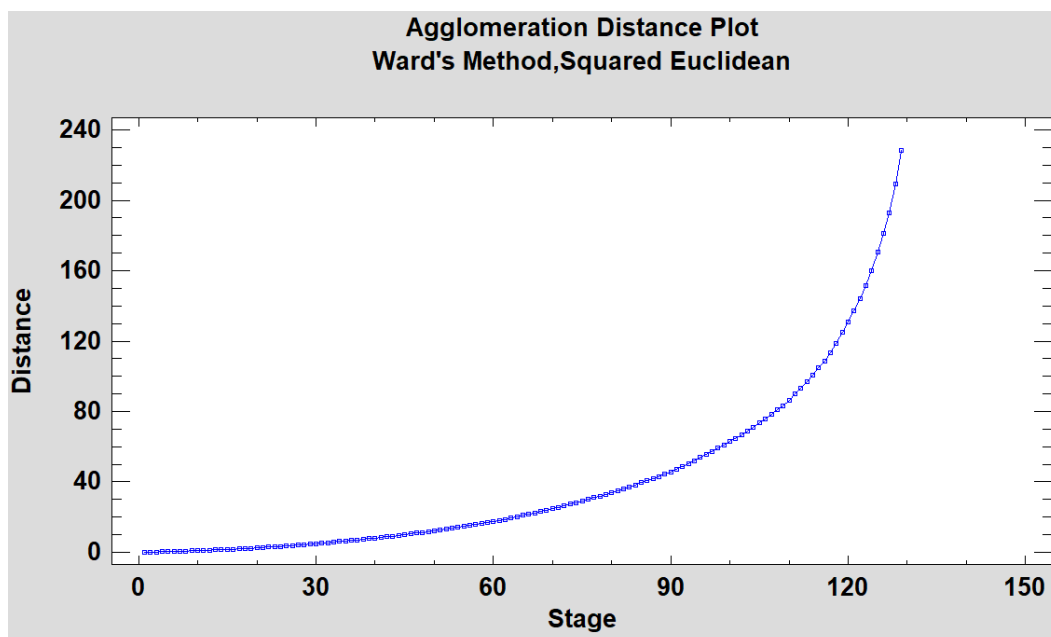


Рис. 1. Проміжки агломерації

Джерело: побудовано авторами засобами програмного інструментарію Statgraphics Centurion

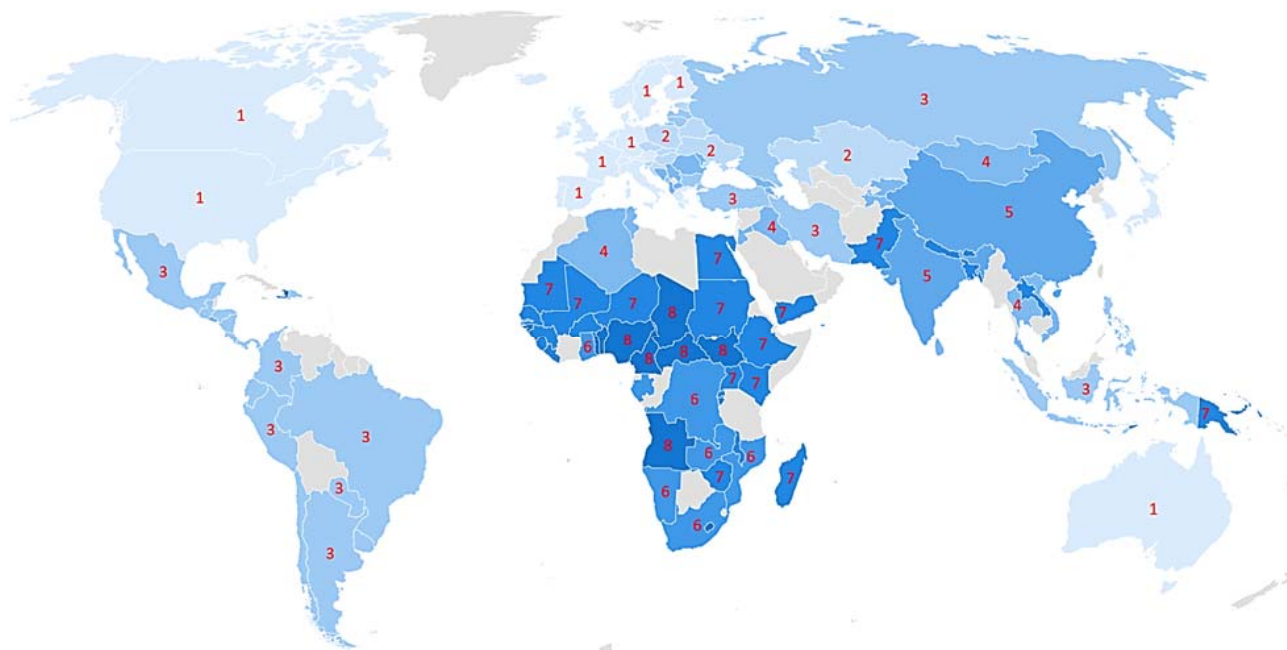


Рис. 2. Кластерний розподіл на географічній карті

Джерело: побудовано авторами засобами програмного інструментарію Microsoft Excel

Данія, Фінляндія, Об'єднане Королівство, Бельгія, Канада, Сполучені Штати, Австрія, Ізраїль, Японія, Корея (Республіка), Люксембург, Іспанія, Франція, Мальта, Італія, Греція, Португалія.

До другого кластеру увійшли такі країни Європи: Словенія, Чеська Республіка, Естонія, Кіпр, Польща, Словаччина, Угорщина, Хорватія, Казахстан, Білорусь, Україна.

До третього кластеру увійшли країни з різних частин світу: Литва, Латвія, Чилі, Аргентина, Чорногорія, Російська Федерація, Туреччина, Уругвай, Болгарія, Панама, Коста-Ріка, Сейшельські острови, Іран, Мексика, Перу, Колумбія, Бразилія, Еквадор, Домініканська республіка, Парагвай, Індонезія, Філіппіни, Гватемала, Нікарагуа, Гондурас.

До четвертого кластеру увійшли теж країни з різних частин світу: Румунія, Грузія, Сербія, Маврикій, Албанія, Шрі Ланка, Боснія і Герцеговина, Таїланд, Вірменія, Північна Македонія, Сент-Люсія, Молдова, Алжир, Мальдіви, Туніс, Монголія, Йорданія, В'єтнам, Киргизстан, Ірак, Сальвадор, Таджикистан.

До п'ятого кластеру увійшли дві країни з найбільшим населенням в світі: Китай та Індія.

До шостого кластеру увійшли такі країни Африки: Болівія, Південна Африка, Габон, Намібія, Сан-Томе і Принсіпі, Есватіні, Гана, Замбія, Конго, Мозамбік.

До сьомого кластеру увійшли наступні країни Африки: Єгипет, Бутан, Бангладеш, Лаоська Народно-Демократична Республіка, Східний Тимор, Непал, Кенія, Зімбабве, Пакистан, Папуа-

Нова Гвінея, Мавританії, Уганда, Руанда, Танзанія, Мадагаскар, Сенегал, Судан, Гамбія, Ефіопія, Малаві, Ліберія, Гвінея, Ємен, Буркіна-Фасо, Сьєрра-Леоне, Малі, Бурунді, Нігер.

До восьмого кластеру увійшли теж країни Африки: Ангола, Камерун, Коморські острови, Бенін, Нігерія, Берег Слонової Кістки, Лесото, Йти, Гаїті, Конго, Гвінея-Бісау, Південний Судан, Чад, Центральнаафриканська Республіка.

Такий розподіл країн за кластерами даних відбувся не тільки за допомогою географічного розподілу за регіонами, але й у розрізі впливу кожного індикатора на створення кластеру. Так, до першого та другого кластеру увійшли країни Європи, що розділилися за рівнем економічного розвитку на дві групи. До третього та четвертого кластеру увійшли країни з різних частин світу, в цих кластерах досить важко визначити, які індикатори впливають найбільше на їх формування. В п'ятий кластер увійшло всього дві країни, Китай та Індія, для створення кластеру найбільший вплив зробив індикатор загального населення. До кластерів шість сім вісім увійшли країни Африки та деякі країни Азії. Їх розподіл зумовлений економічними факторами, а також факторами населення країн.

Дискримінантний аналіз є методом який використовується для аналізу даних дослідження, коли критерій або залежна змінна є категоричними, а предиктор або незалежна змінна мають інтервальний характер. Метою дискримінантного аналізу є розробка дискримінантних функцій, які є не що інше, як лінійна комбінація незалежних змінних, яка ідеально розрізняє категорії залеж-

ної змінної. Це дає змогу досліднику перевірити, чи існують суттєві відмінності між групами з точки зору змінних-провісників [11].

При проведенні множинного дискримінантного аналізу в програмному забезпеченні Statgraphics Centurion розроблено регресійну модель F1, що описує внесок кожного індикатора в формування кластера 1, та має вигляд:

$$F1 = -0,180166 \cdot K9 - 0,0471193 \cdot K2 + 1,29466 \cdot K10 - 0,664787 \cdot K7 + 0,0394757 \cdot K8 + 0,219873 \cdot K5 + 2,55374 \cdot K6 - 0,41317 \cdot K1 + 1,06147 \cdot K3 - 1,13345 \cdot K4, \quad (3)$$

де K1 – коефіцієнт людської нерівності, K2 – коефіцієнт Джині, K3 – нерівність в очікуваній тривалості життя, K4 – очікувана тривалість життя при народженні, K5 – валовий національний дохід, K6 – індекс очікуваної тривалості життя з поправкою на нерівність, K7 – загальна кількість населення, K8 – міське населення, K9 – індекс освіти, K10 – населення віком від 15 до 64 років.

Вага кожного індикатора відображається за абсолютним значенням коефіцієнта, чим більше значення, тим більший внесок даного показника в формування кластера 1 (табл. 3).

Таблиця 3 показує коефіцієнти функцій, що використовуються для визначення впливу кожного індикатора на формування кластера.

З відносної величини коефіцієнтів у моделі (3) можна визначити, як незалежні змінні використовуються для розрізнення кластерів (табл. 4).

Отже, для розробки моделі було використано 137 країн розподілених на 8 кластерів. Для формування кластерів використано 10 предикторних змінних. В п'яти з семи дискримінаційних функцій значення рівня значущості (P-value) має нижче 0,05%, що свідчить про статистично значущі функції при довірчому рівні 95,0%. Значення Лямбда Уїлкса вимірюються в проміжку від 0 до 1, якщо значення прямує до нуля це означає, що значення краще відрізняються одне від одного. П'ять функцій демонструють достатній рівень дискримінації в використовуваних даних.

Висновки з проведеного дослідження. В статті проведена робота з аналізу публікацій в наукометричній базі даних Scopus та створено хмара ключових слів, що використовуються. Також проведена робота щодо бібліометричного аналізу окремих публікацій вітчизняних науковців та міжнародних робіт по тематиці в роботі з кластерним аналізом.

Таблиця 3

Стандартизовані коефіцієнти дискримінантних функцій

	F1	F2	F3	F4	F5	F6	F7
Індекс освіти	-0,180	0,022	0,503	0,186	-0,895	-0,010	0,543
Коефіцієнт Джині	-0,047	-0,099	0,412	0,875	0,200	0,046	0,250
Населення віком 15-64	1,295	3,525	0,013	0,407	-0,507	-1,407	3,292
Загальна кількість населення	-0,665	-2,829	0,061	-0,421	0,509	1,464	-3,486
Міське населення	0,039	-0,183	0,281	0,478	0,115	-0,052	-0,808
Валовий національний дохід (ВНД)	0,220	-0,237	0,491	-0,307	0,696	-0,612	0,045
Індекс очікуваної тривалості	2,554	-0,855	2,864	-2,523	-4,844	4,328	-0,390
Коефіцієнт людської нерівності	-0,413	0,234	0,363	-0,326	-0,127	0,347	-0,144
Нерівність в очікуваній	1,061	-0,273	2,137	-1,285	-2,027	2,110	0,674
Очікувана тривалість життя	-1,138	0,286	-1,564	1,478	3,466	-1,817	0,932

Джерело: розраховано авторами засобами програмного інструментарію Statgraphics Centurion

Таблиця 4

Значення дискримінантної функції

Дискримінантна функція	Власне значення	Відносний відсоток	Канонічна кореляція	Статистика Лямбда Уїлкса	Хі-квадрат	Ступені свободи	P-value
1	13,9593	42,72	0,966	0,0002	1068,61	70	0
2	12,98	39,72	0,9635	0,0033	725,03	54	0
3	2,9069	8,9	0,8625	0,0463	390,05	40	0
4	2,1649	6,63	0,8270	0,1811	216,98	28	0
5	0,5147	1,58	0,5829	0,5732	70,66	18	0
6	0,1434	0,44	0,3542	0,8683	17,93	10	0,0561
7	0,0071	0,02	0,0841	0,9929	0,90	4	0,9242

Джерело: розраховано авторами засобами програмного інструментарію Statgraphics Centurion

Для розрахунку кластерного аналізу використані такі індикатори: коефіцієнт людської нерівності (К1), коефіцієнт Джині (К2), нерівність в очікуваній тривалості життя (К3), очікувана тривалість життя при народженні (К4), валовий національний дохід (К5), індекс очікуваної тривалості життя з поправкою на нерівність (К6), загальна кількість населення (К7), міське населення (К8), індекс освіти (К9), населення віком 15–64 (К10). Їхні значення мають різні величини виміру вони були стандартизовані засобами програмного інструментарію Statgraphics Centurion.

Аргументована кількість кластерів за допомогою формули Стерджеса кількість вісім кластерів даних. Створено списки країн які входять до кожного кластеру.

Визначену регресійну формулу функції з дискримінантного аналізу з її допомогою визначено вплив кожного індикатора на створений кластер даних. Значущість дискримінантних функцій обґрунтовано показниками Лямбда Уїлкса та рівня значущості P-value розрахованими за допомогою інструментарію Statgraphics Centurion.

Результати кластерного розподілу можуть бути використані в процесі розвитку держав для знаходження оптимальних статичних значень, до яких потрібно прямувати розвиток держави. Щоб зробити перехід недостатньо розвинених держав до більш розвинених угруповань. Отримані дані прослужать для подальшого глибокого аналізу даних та знаходженню нових патернів в розвитку країн світу.

БІБЛІОГРАФІЧНИЙ СПИСОК:

1. Яловенко О.В., Фетісов В.С. Аналіз даних анкетування за допомогою кластерного аналізу. 2020. С 1–8. URL: <https://phm.cuspu.edu.ua/ojs/index.php/SNYS/article/download/1778/pdf> (дата звернення: 6.05.2022).
2. Жигайло О.М., Борис В.В. Кластерний аналіз даних в автоматизованих системах простежуваності. *Автоматизація технологічних і бізнес-процесів*. 2018. С. 39–46. URL: https://www.researchgate.net/publication/324382544_klasternij_analiz_danih_v_avtomatizovanih_sistemah_prostezuvanosti/fulltext/5b1056c54585150a0a5e0a13/klasternij-analiz-danih-v-avtomatizovanih-sistemah-prostezuvanosti.pdf (дата звернення: 6.05.2022).
3. Тюрин А.Г., Зуев І.О., Кластерний аналіз, методи та алгоритми кластеризації. *Вісник МДТУ МИРЕА*. 2014. С. 86–97. URL: <https://rtj.mirea.ru/upload/medialibrary/fba/09-tyurin.pdf>.
4. Серга Э.Н. Універсальний ітераційний метод кластеризації даних. *Український гідрометеорологічний журнал*. 2013. С. 112–123, URL: <http://uhmj.odeku.edu.ua/wp-content/uploads/2013/05/17.pdf> (дата звернення: 6.05.2022).
5. Дронов С.В., Шеларь А.Ю. Латентний кластерний аналіз для двох кластерів. 2018. С. 23–26.

URL: <http://elibrary.asu.ru/xmlui/bitstream/handle/asu/6240/23-26.pdf?sequence=1&isAllowed=> (дата звернення: 6.05.2022).

6. Амонс О.А., Янов Ю.О., Безпалый І.О. Кластеризація документів на основі статистичної близькості термів. *Вісник НТУУ «КПІ» Інформатика, управління та обчислювальна техніка*. 2008. С. 56–62. URL: <https://core.ac.uk/download/pdf/47221436.pdf> (дата звернення: 6.05.2022).

7. Ельгаміфар Е., Відаль Р. Розріджена кластеризація підпростору: алгоритм, теорія та застосування. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2003. С. 2765–2781. URL: <https://ieeexplore.ieee.org/document/6482137> (дата звернення: 6.05.2022).

8. Мерве Д., Енгельбрехт А.П. Кластеризація даних за допомогою оптимізації роя частинок. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2004. URL: <https://ieeexplore.ieee.org/document/1299577/authors#authors> (дата звернення: 6.05.2022).

9. Сандіп Р., Санджай Дж., Раджеш К. Огляд алгоритмів оптимізації роя частинок та їх застосування до кластеризації даних. *Springer link*. 2011. URL: <https://link.springer.com/article/10.1007/s10462-010-9191-9> (дата звернення: 6.05.2022).

10. United nations development programme Human Development Reports. 2022. URL: <https://hdr.undp.org/en/content/human-development-index-hdi> (дата звернення: 6.05.2022).

11. Discriminant Analysis. 2020. URL: <https://www.statisticssolutions.com/discriminant-analysis> (дата звернення: 6.05.2022).

REFERENCES:

1. Yalovenko O.V., Fetisov V.S. (2020). Analiz danykh anketuvannia za dopomohoiu klasternoho analizu [Analysis of survey data using cluster analysis]. Available at: <https://phm.cuspu.edu.ua/ojs/index.php/SNYS/article/download/1778/pdf> (accessed 06 May 2022).
2. Zhigailo O.M., Boris V.V. (2018) Klasternyi analiz danykh v avtomatyzovanykh sistemakh prostezhuvanosti [Cluster data analysis in automated traceability systems]. *Avtomatyzatsiia tekhnolohichnykh i biznes-protsesiv*, pp. 39–46. Available at: https://www.researchgate.net/publication/324382544_klasternij_analiz_danih_v_avtomatizovanih_sistemah_prostezuvanosti/fulltext/5b1056c54585150a0a5e0a13/klasternij-analiz-danih-v-avtomatizovanih-sistemah-prostezuvanosti.pdf (accessed 06 May 2022).
3. Tyurin A.G., Zuev I.A., (2014) Klasternyi analiz, metody ta alhorytmy klasteryzatsii [Cluster analysis, methods and algorithms of clustering]. *Visnyk MDTU MYREA*, pp. 86–97. Available at: <https://rtj.mirea.ru/upload/medialibrary/fba/09-tyurin.pdf> (accessed 06 May 2022).
4. Serga E.N. (2013) Universalnyi iteratsiyni metod klasteryzatsii danykh [Universal iterative method of data clustering]. *Ukrainskyi hidrometeorolohichniy zhurnal*, pp. 112–123. Available at: <http://uhmj.odeku.edu.ua/wp-content/uploads/2013/05/17.pdf> (accessed 06 May 2022).

5. Dronov S.V., Shelar A.Yu. (2018) Latentnyi klasternyi analiz dlia dvokh klasteriv [Latent cluster analysis for two clusters], pp. 23–26. Available at: <http://elibrary.asu.ru/xmlui/bitstream/handle/asu/6240/23-26.pdf?sequence=1&isAllowed=> (accessed 06 May 2022).
6. Amons O.A., Yanov Yu.A., Bepaly I.A. (2008) Klasteryzatsiia dokumentiv na osnovi statystychnoi blyzkosti termiv [Clustering of documents based on statistical proximity of terms]. *Visnyk NTUU «KPI» Informatyka, upravlinnia ta obchysliuvanna tekhnika*, pp. 56–62. Available at: <https://core.ac.uk/download/pdf/47221436.pdf> (accessed 06 May 2022).
7. Elgamifar E., Vidal R. (2003) Rozridzhena klasteryzatsiia pidprostoru: alhorytm, teoriia ta zastosuvannia [Sparse subspace clustering: algorithm, theory, and application]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2765–2781. Available at: <https://ieeexplore.ieee.org/document/6482137> (accessed 06 May 2022).
8. Merwe D., Engelbrecht A. P.(2004) Klasteryzatsiia danykh za dopomohoiu optymizatsii roia chastynok [Data clustering using particle swarm optimization]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Available at: <https://ieeexplore.ieee.org/document/1299577/authors#authors> (accessed 06 May 2022).
9. Sandeep R., Sanjay J. Rajesh K.(2011) Ohliad alhorytmiv optymizatsii roia chastynok ta yikh zastosuvannia do klasteryzatsii danykh [Review of particle swarm optimization algorithms and their application to data clustering]. *Springer link*. Available at: <https://link.springer.com/article/10.1007/s10462-010-9191-9> (accessed 06 May 2022).
10. United nations development programme (2022). Retrieved May 6, 2022, from <https://hdr.undp.org/en/content/human-development-index-hdi>.
11. Discriminant Analysis (2020). Retrieved May 13, 2022. Available at: <https://www.statisticssolutions.com/discriminant-analysis>.