

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

Інститут інформаційних технологій в економіці

Кафедра штучного інтелекту, моделювання та статистики

Освітньо-професійна програма «Економічна кібернетика»

**Галузь знань 05 «Соціальні та поведінкові науки»
Спеціальність 051 «Економіка»**

Форма здобуття освіти: очна (денна)

КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА РОБОТА

**на тему «Аналіз та моделювання регіональних особливостей
народжуваності в Україні»**

(назва теми)

здобувача Коваленко Поліни Ігорівни
(ПІБ)

(підпис)

Науковий керівник: кандидат економічних наук,
доцент Ольга ОСИПОВА

(підпис)

**Робота допущена до захисту перед екзаменаційною
комісією з атестації здобувачів вищої освіти (ЕК)**

Завідувач кафедри: кандидат фізико-математичних наук,
професор Галина ВЕЛИКОІВАНЕНКО

(підпис)

Київ 2026

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

Інститут інформаційних технологій в економіці

Кафедра штучного інтелекту, моделювання та статистики

Освітньо-професійна програма «Економічна кібернетика»

Галузь знань 05 «Соціальні та поведінкові науки»

Спеціальність 051 «Економіка»

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми

Ольга ПРИТОМАНОВА

(підпис)

_____ 2026 р.

ЗАТВЕРДЖУЮ

Завідувач кафедри штучного
інтелекту, моделювання та
статистики

Галина ВЕЛИКОІВАНЕНКО

(підпис)

_____ 2026 р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувача вищої освіти

Коваленко Поліни Ігорівни

(прізвище, ім'я, по батькові)

очної (денної)

форми здобуття освіти

очної (денної), заочної, дистанційної

на підготовку кваліфікаційної бакалаврської роботи

на тему

**«Аналіз та моделювання регіональних особливостей народжуваності в
Україні»**

Тему затверджено наказом ректора Університету від «16» березня 2026 р. № 417-ст

**Кваліфікаційна бакалаврська робота виконується на матеріалах
сайтів Державної служби статистики <https://www.ukrstat.gov.ua/>**

План кваліфікаційної бакалаврської роботи

Розділ 1	Теоретичні та методичні основи дослідження народжуваності в регіонах України
Розділ 2	Економетричне моделювання та застосування методів інтелектуального аналізу даних у дослідженні народжуваності в регіонах України

Об'єкт дослідження:	процес регіональної диференціації народжуваності в Україні до та в умовах повномасштабної війни
Предмет дослідження:	кількісні закономірності та статистичні зв'язки між показниками народжуваності і соціально-економічними чинниками в регіональному розрізі, виявлені за допомогою методів багатовимірного статистичного аналізу та регресійного моделювання
Мета виконання кваліфікаційної бакалаврської роботи:	комплексний аналіз та моделювання регіональних особливостей народжуваності в Україні у період до повномасштабного вторгнення (2021 р.) та в умовах повномасштабної війни (2025 р.) на основі методів багатовимірного статистичного аналізу та регресійного моделювання

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1: розкрити економічну сутність народжуваності як складової демографічної ситуації; проаналізувати регіональні особливості демографічних процесів в Україні; розглянути статистичні підходи до аналізу соціально-економічних чинників; обґрунтувати застосування методу головних компонент, кластерного аналізу та регресійного моделювання в демографічних дослідженнях.

У розділі 2: сформулювати інформаційну базу дослідження та виконати підготовку даних (нормування, перевірка кореляцій, відбір змінних); здійснити виділення головних компонент соціально-економічних чинників та інтерпретувати їх економічний зміст; провести кластеризацію регіонів України; побудувати регресійну модель впливу соціально-економічних чинників на рівень народжуваності; оцінити статистичну значущість параметрів моделі та сформулювати аналітичні висновки щодо отриманих результатів.

Завдання підготував
науковий керівник

(підпис)

Ольга ОСИПОВА

(ім'я, прізвище)

«__» _____ 2026 р.

Завдання одержав
здобувач

(підпис)

Поліна КОВАЛЕНКО

(ім'я, прізвище)

«__» _____ 2026 р.

Реферат

Кваліфікаційна бакалаврська робота складається зі вступу, двох розділів, висновків, переліку використаних джерел, що містить 30 найменувань на 3 сторінках, та 2 додатків на 2 сторінках. Загальний обсяг роботи складає 63 сторінки друкованого тексту, з яких 56 сторінок основної частини. Робота містить 16 рисунків та 13 таблиць.

«Аналіз та моделювання регіональних особливостей народжуваності в Україні»

Об'єктом дослідження є процес регіональної диференціації народжуваності в Україні до та в умовах повномасштабної війни.

Предметом дослідження є кількісні закономірності та статистичні зв'язки між показниками народжуваності і соціально-економічними чинниками в регіональному розрізі, виявлені за допомогою методів багатовимірної статистичного аналізу та регресійного моделювання.

Мета та завдання виконання дослідження. Метою роботи є комплексний аналіз та моделювання регіональних особливостей народжуваності в Україні у період до початку повномасштабного вторгнення (2021 р.) та в умовах повномасштабної війни (2025 р.) на основі методів багатовимірної статистичного аналізу та регресійного моделювання.

Відповідно до поставленої мети визначені такі завдання:

- дослідити теоретичні засади народжуваності як складової демографічної ситуації та обґрунтувати методологію аналізу;
- сформувати інформаційну базу і підготувати дані;
- провести багатовимірний статистичний аналіз для двох часових зрізів із застосуванням методу головних компонент та кластерного аналізу;
- побудувати регресійну модель впливу соціально-економічних і воєнних чинників на народжуваність і оцінити статистичні зв'язки між показниками.

Теоретична, методична та практична значущість отриманих результатів. Теоретичне значення роботи полягає у систематизації підходів до кількісного аналізу регіональної диференціації народжуваності в умовах воєнного часу. Методичне – у розробленні алгоритму порівняльного аналізу регіональної структури народжуваності для двох часових зрізів на основі комплексу методів багатовимірної статистики. Практичне – у виявленні типологічних груп регіонів і кількісній оцінці впливу прифронтового статусу та притоку внутрішньо переміщених осіб на народжуваність, що може слугувати аналітичною основою для розробки диференційованої регіональної демографічної і соціальної політики.

Рік виконання кваліфікаційної бакалаврської роботи – 2026.

Рік захисту роботи – 2026.

Ключові слова: народжуваність, регіональна диференціація, соціально-економічні чинники, прифронтові регіони, внутрішньо переміщені особи, повномасштабна війна.

В і д г у к
про кваліфікаційну бакалаврську роботу
здобувача освітньо-професійної програми «Економічна кібернетика»
навчально-наукового інституту
«Інститут інформаційних технологій в економіці»

Коваленко Поліни Ігорівни

на тему «Аналіз та моделювання регіональних особливостей народжуваності в Україні»

1. Актуальність теми обумовлена поглибленням демографічної кризи в Україні, яка суттєво посилилася внаслідок повномасштабної війни. Дослідження регіональних особливостей народжуваності та чинників, що впливають на демографічні процеси, є важливим для формування ефективної державної та регіональної соціально-демографічної політики. Використання сучасних методів багатовимірного статистичного аналізу та економетричного моделювання дозволяє отримати обґрунтовані висновки щодо закономірностей розвитку досліджуваного явища.

2. Позитивні риси кваліфікаційної бакалаврської роботи: у роботі здійснено ґрунтовний аналіз теоретичних аспектів дослідження народжуваності, сформовано інформаційну базу на основі офіційних статистичних даних та виконано комплексне економіко-математичне дослідження із застосуванням методу головних компонент, кластерного аналізу та множинної регресії. Робота демонструє високий рівень підготовки здобувачки, її вміння працювати з науковими джерелами, аналізувати статистичні дані, інтерпретувати результати моделювання та формулювати обґрунтовані висновки.

3. Наявність самостійних розробок автора: кваліфікаційна бакалаврська робота виконана автором самостійно. Здобувачкою здійснено збір та підготовку даних, реалізовано всі етапи статистичного та економетричного аналізу, проведено інтерпретацію отриманих результатів та сформульовано практичні висновки. Робота відповідає затвердженому індивідуальному завданню та оформлена відповідно до встановлених вимог.

4. Цінність теоретичних висновків та практичних рекомендацій, отриманих у роботі, полягає у виявленні особливостей регіональної диференціації народжуваності в Україні до та в умовах повномасштабної війни, визначенні типологічних груп регіонів та кількісній оцінці впливу соціально-економічних і воєнних чинників на демографічні процеси. Отримані результати можуть бути використані як аналітична основа для розроблення заходів демографічної та регіональної політики.

5. Наявність недоліків: суттєвих недоліків у роботі не виявлено. Окремі положення дослідження могли б бути розширені шляхом використання більш тривалих часових рядів даних та порівняння результатів декількох альтернативних економетричних моделей, що може бути напрямом подальших наукових досліджень.

6. Загальна оцінка кваліфікаційної бакалаврської роботи та її допущення до захисту перед ЕК: кваліфікаційна бакалаврська робота Коваленко П.І. виконана на високому рівні та може бути допущена до захисту перед ЕК. Робота свідчить про відповідність набутих здобувачкою компетентностей вимогам освітньої програми «Економічна кібернетика», а її автор, за умови успішного захисту, заслуговує на присвоєння освітньої кваліфікації «Бакалавр з економіки».

Науковий керівник: доцент кафедри
штучного інтелекту,
моделювання та статистики,
к.е.н., доцент

_____ Ольга ОСИПОВА

ЗМІСТ

ВСТУП.....	3
РОЗДІЛ 1. ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ НАРОДЖУВАНOSTІ В РЕГІОНАХ УКРАЇНИ	6
1.1 Народжуваність як складова демографічної ситуації та її регіональні особливості	6
1.2 Методологія зменшення розмірності та кластерного аналізу у дослідженні регіональної диференціації	11
1.3 Регресійний аналіз як інструмент дослідження впливу соціально- економічних чинників на народжуваність	18
РОЗДІЛ 2. ЕКОНОМЕТРИЧНЕ МОДЕЛЮВАННЯ ТА ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ У ДОСЛІДЖЕННІ НАРОДЖУВАНOSTІ В РЕГІОНАХ УКРАЇНИ	23
2.1. Формування інформаційної бази та підготовка даних	23
2.1.1. Формування інформаційної бази за період до повномасштабного вторгнення.....	23
2.1.2. Формування інформаційної бази в умовах повномасштабної війни .	25
2.2. Багатовимірний статистичний аналіз народжуваності регіонів України у період до повномасштабного вторгнення	26
2.3. Багатовимірний статистичний аналіз народжуваності регіонів України в умовах повномасштабної війни	41
2.4. Регресійне моделювання впливу соціально-економічних чинників на народжуваність в умовах повномасштабної війни	50
ВИСНОВКИ.....	56
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	59
ДОДАТКИ.....	62

ВСТУП

Актуальність теми. Україна зараз переживає одну з найгостріших демографічних криз за всю історію незалежності. Рівень народжуваності поступово знижувався ще до 2022 року, але повномасштабне вторгнення Росії різко погіршило ситуацію. Якщо у 2021 році в Україні народилося 273,8 тис. дітей, то у 2025 році їх кількість зменшилася до 168,8 тис. При цьому між регіонами спостерігається значна різниця, яка під час війни стала ще більшою. Звична картина «вища народжуваність на заході, нижча – на сході» змінилася. Тепер на показники народжуваності впливатиме також те наскільки регіон наближений до лінії фронту. Розуміння цих нових закономірностей є надзвичайно важливим для розробки ефективної демографічної та регіональної політики – як під час воєнного стану, так і після його завершення.

Аналіз останніх досліджень і публікацій. Питання демографічної ситуації та народжуваності в Україні досліджуються переважно фахівцями Інституту демографії та соціальних досліджень ім. М.В. Птухи НАН України. Зокрема Лібанова Е.М. у своїх публікаціях [5] аналізує втрати людського капіталу внаслідок війни і наголошує на неможливості швидкого відновлення народжуваності. Черенько Л.М. та ін. [6] досліджують зміни якості життя населення як чинника репродуктивної поведінки. Динаміку народжуваності в умовах воєнного конфлікту аналізують Напорчук Д. [13] і Бакута Р. [14]. Разом з тим комплексний кількісний аналіз регіональної диференціації народжуваності із застосуванням методів багатовимірної статистики та регресійного моделювання для порівняння довоєнного і воєнного періодів залишається недостатньо дослідженим у науковій літературі.

Мета і завдання дослідження. Метою роботи є комплексний аналіз та моделювання регіональних особливостей народжуваності в Україні у період до повномасштабного вторгнення (2021 р.) та в умовах повномасштабної війни

(2025 р.) на основі методів багатовимірного статистичного аналізу та регресійного моделювання.

Для досягнення мети поставлено такі завдання:

- визначити теоретичні засади дослідження народжуваності як складової демографічної ситуації та обґрунтувати методологію багатовимірного статистичного аналізу;
- сформувати інформаційну базу дослідження та підготувати дані для аналізу;
- провести багатовимірний статистичний аналіз регіональних особливостей народжуваності у період до повномасштабного вторгнення (2021 р.) із застосуванням методу головних компонент та кластерного аналізу;
- провести аналогічний аналіз в умовах повномасштабної війни (2025 р.) і виявити трансформацію регіональної структури народжуваності;
- побудувати регресійну модель впливу соціально-економічних чинників та воєнних факторів на народжуваність в умовах 2025 року і оцінити статистичні зв'язки між показниками.

Об'єкт дослідження – процес регіональної диференціації народжуваності в Україні до та в умовах повномасштабної війни.

Предмет дослідження – кількісні закономірності та статистичні зв'язки між показниками народжуваності і соціально-економічними чинниками в регіональному розрізі, виявлені за допомогою методів багатовимірного статистичного аналізу та регресійного моделювання.

Методи дослідження. У роботі використано: методи описової статистики та кореляційного аналізу – для характеристики розподілу показників і виявлення взаємозв'язків між ними; метод головних компонент з обертанням Varimax – для зменшення розмірності даних і виявлення латентних факторів; ієрархічну кластеризацію методом Варда та метод k-середніх – для типологізації регіонів; множинну лінійну регресію з фіктивними змінними – для кількісної оцінки статистичних зв'язків між народжуваністю і соціально-економічними та воєнними чинниками. Усі розрахунки виконано у середовищі Jupyter Notebook мовою

програмування Python із використанням бібліотек pandas, numpy, scikit-learn, seaborn та matplotlib.

Теоретична, методична та практична значущість отриманих результатів. Теоретичне значення роботи полягає у систематизації підходів до кількісного аналізу регіональної диференціації народжуваності в умовах воєнного часу. Методичне значення полягає у розробленні алгоритму порівняльного аналізу демографічних процесів на основі комплексу методів багатовимірної статистики. Практичне значення – у виявленні типологічних груп регіонів за характером народжуваності і кількісній оцінці впливу прифронтового статусу та статусу реципієнта ВПО на кількість народжених, що може бути використано при розробці регіональної демографічної і соціальної політики.

Інформаційна база дослідження. Основу інформаційної бази складають офіційні дані Міністерства юстиції України [2], Державної служби статистики України [3], Міністерства розвитку громад та територій України [26], Міжнародної організації з міграції (МОМ, Раунд 22) [15]. Теоретичну базу формують праці вітчизняних і зарубіжних науковців у галузі демографії, економетрики та багатовимірного статистичного аналізу, а також Стратегія демографічного розвитку України до 2040 року [11].

Структура роботи. Робота складається зі вступу, двох розділів, висновків, переліку використаних джерел та додатків. У першому розділі розглядаються теоретичні засади дослідження народжуваності і методологія застосованих методів. Другий розділ містить результати комплексного статистичного і регресійного аналізу регіональних особливостей народжуваності для двох часових зрізів. У висновках узагальнено основні результати і сформульовано практичні рекомендації.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ НАРОДЖУВАНOSTІ В РЕГІОНАХ УКРАЇНИ

1.1 Народжуваність як складова демографічної ситуації та її регіональні особливості

Народжуваність є одним з ключових показників демографічної ситуації будь-якої країни і водночас найважливішим індикатором її відтворювального потенціалу. У статистиці населення під народжуваністю розуміють процес дітонародження у певній сукупності людей за визначений період часу [1]. Для кількісного вимірювання використовується загальний коефіцієнт народжуваності - відношення числа народжених до середньорічної чисельності населення виражений в проміле (‰) [1, с. 29]:

$$b = \frac{B}{\bar{S}} \cdot 1000, \quad (1.1)$$

де b - загальний коефіцієнт народжуваності, ‰;

B - кількість народжених дітей за рік, осіб;

$\bar{S}$ - середньорічна чисельність населення, осіб.

Не менш важливим у демографічному аналізі є сумарний коефіцієнт народжуваності, який характеризує кількість дітей, що народжує одна жінка протягом усього репродуктивного періоду (15-49 років). Для простого відтворення населення має становити не менше 2,2-2,3 дитини [1, с. 37].

Для розуміння сучасної демографічної ситуації варто звернутися до динаміки народжуваності за останні роки. Після певного поживлення у 2008-2012 роках, коли народжуваність перевищувала 500 тис. осіб на рік тенденція знову

стала спадною. У 2021 році народилося 273,8 тис. дітей, що вже тривожно мало. Повномасштабна війна різко погіршила ситуацію: 2022 рік – 206 тис., 2023 рік – 187,4 тис., 2024 рік – 176,7 тис., 2025 рік – 168,8 тис. [2]. В Україні зафіксовано історично найнижчі рівні народжуваності – за період з 2021 по 2023 рік кількість новонароджених зменшилася майже вдвічі [12].

Динаміку народжуваності в Україні за 2010-2025 роки наведено на рис. 1. 1.

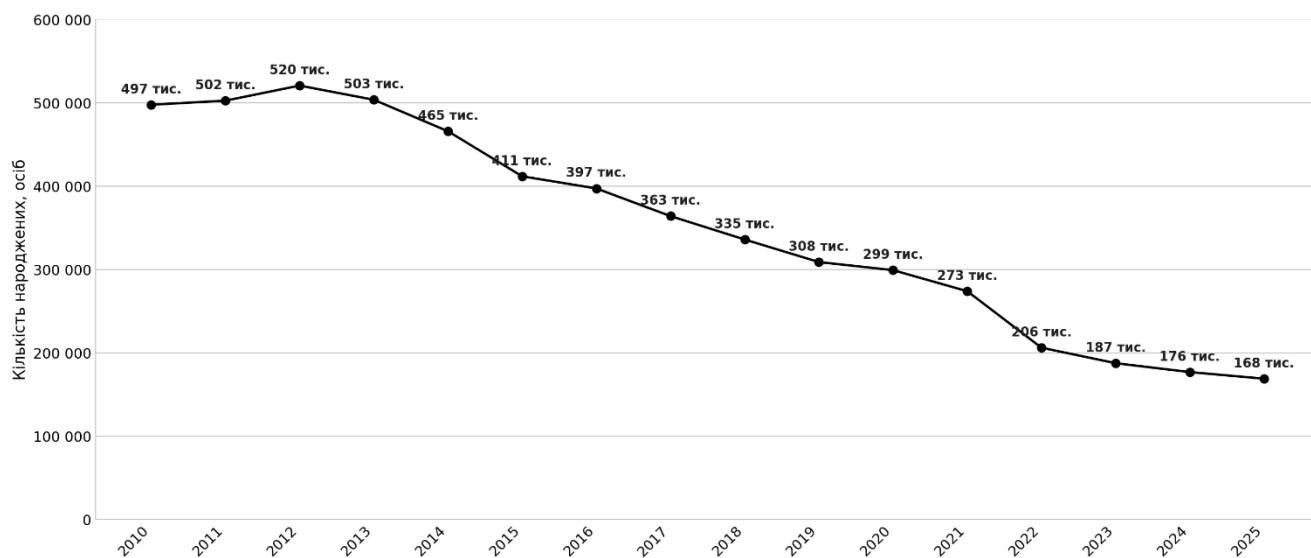


Рисунок 1. 1 – Динаміка кількості народжених в Україні, 2010-2025 рр., осіб

Джерело: розроблено автором на основі [2; 3]

Графік наочно відображає два різких спади – після 2014 року через початок збройного конфлікту на сході і у 2022 році через повномасштабне вторгнення. Сумарний коефіцієнт народжуваності знизився з 1,22 у 2021 році до орієнтовно 0,7-0,9 у 2023-2024 роках, що є одним з найнижчих показників у світі за даними ООН [4; 5].

На рівень народжуваності впливає складний комплекс взаємопов'язаних чинників – економічних, соціальних, медико-демографічних та екологічних.

Серед економічних чинників ключову роль відіграє рівень доходів населення та стан ринку праці – в публікаціях Інституту демографії та соціальних досліджень імені М.В. Птухи НАН України показано, що сім'ї з вищим рівнем доходу частіше планують народження другої та третьої дитини [6; 7, с. 140].

Серед соціальних чинників важливу роль відіграє рівень професійної зайнятості жінок, відповідно до чого середній вік матері при народженні першої

дитини зріс з 23,5 років у 2002 році до 27,9 років у 2020 році [7]. Зокрема це може бути також пов'язано зі змінами в репродуктивних установках молоді. Дослідження Інституту соціології НАН України виявили, що протягом останніх 30-40 років відбувалася дуже стрімка зміна репродуктивної поведінки українців, які унаслідували західну модель сім'ї – коли жінки народжують дітей пізніше, і в основному менше. Також зростає частка свідомо бездітних пар і відбувається переорієнтація на одnodітну модель сім'ї [7, с. 34]. Разом з цим спостерігається зменшення традиційних сімейних структур і поширення співжиття без офіційного шлюбу. Ці зміни знижують стабільність родин і часто не сприяють прийняттю рішення про народження дитини.

Медико-демографічні чинники також суттєво впливають на рівень народжуваності. Стан репродуктивного здоров'я населення та доступність медичних послуг безпосередньо впливають на репродуктивні рішення сімей. В умовах повномасштабної війни доступ до медичної допомоги значно ускладнився через руйнування інфраструктури, переміщення населення та перевантаження закладів охорони здоров'я. За даними UNICEF, ця ситуація особливо негативно позначається на вагітних жінках та репродуктивному здоров'ї населення в цілому [8].

Стан екології є одним із ключових чинників, що визначають здоров'я населення, та тісно пов'язаний із демографічними процесами. За даними Міністерства захисту довкілля та природних ресурсів України, в регіонах з підвищеним рівнем забруднення повітря спостерігається вищий рівень захворюваності та смертності населення. Особливо гостро ця проблема стоїть у промислових регіонах України, де рівень забруднення повітря перевищує допустимі норми у 2-3 рази [16].

Регіональна диференціація народжуваності є однією з найбільш виражених просторових характеристик демографічної ситуації в Україні. Ще у період до повномасштабного вторгнення між регіонами існували суттєві відмінності, що мали стійкий географічний характер. Традиційно найвищі показники демонстрували західні регіони – Волинська, Рівненська, Закарпатська, Івано-

Франківська області – через збереження традиційних сімейних цінностей, вищу частку сільського населення і молодшу вікову структуру [9]. Натомість промислово розвинені регіони сходу – традиційно мали найнижчі показники через старішу структуру населення і вищий рівень урбанізації [5; 9].

Особливо помітним є зниження народжуваності у міських поселеннях. Згідно з дослідженнями Інституту демографії та соціальних досліджень, у великих містах України сумарний коефіцієнт народжуваності в середньому на 15-20% нижчий, ніж у сільській місцевості [7, с. 20].

Повномасштабна війна кардинально змінила регіональну картину народжуваності. Порівняння 2021 і 2025 років по регіонах наведено на рис. 1. 2.

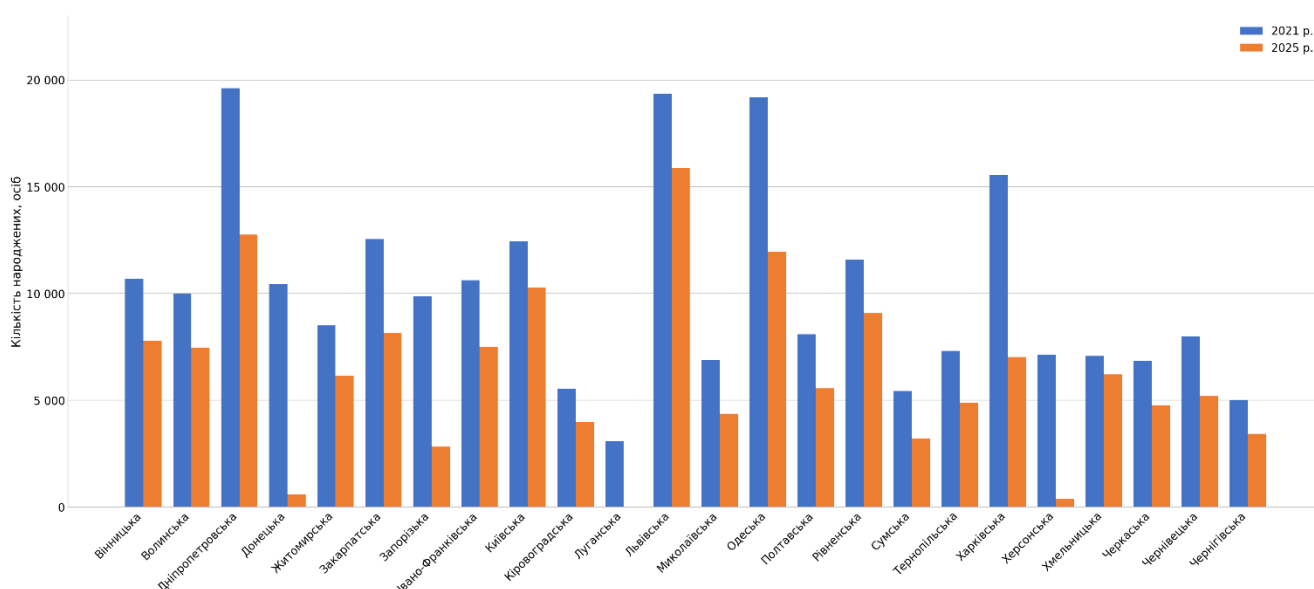


Рисунок 1. 2 – Кількість народжених по регіонах України у 2021 та 2025 рр., осіб

Джерело: розроблено автором на основі [2; 3]

Графік наочно демонструє два характерних явища. По-перше – різке падіння в прифронтових регіонах (Донецька, Луганська, Херсонська, Запорізька), де дані фіксуються лише з підконтрольних уряду територій [2; 3]. По-друге – відносна стійкість у великих регіонах-реципієнтах ВПО (Львівська, Одеська, Київська області), що, швидше за все, пояснюється збільшенням чисельності жінок репродуктивного віку через переміщення населення [15].

Директорка ІДСД ім. М.В. Птухи НАН України Елла Лібанова наголошує, що в умовах війни регіональна диференціація народжуваності визначається вже не традиційними соціально-демографічними чинниками, а переважно масштабом переміщення населення та інтенсивністю бойових дій [5; 10]. Окремою загрозою є те, що більшість емігрантів складають жінки репродуктивного віку з дітьми, зокрема 42% ВПО мають принаймні одну дитину [15]. За даними МОМ (Раунд 22, січень 2026 р.) найбільшими регіонами, які приймають ВПО з окупованих територій є Дніпропетровська область (18%, або 377000 осіб), м. Київ (10%, або 202000 осіб), Харківська (9%, або 186000 осіб), Запорізька (8%, або 163000 осіб) та Одеська (8%, або 159000 осіб) області [15].

Масова зовнішня міграція стала одним із головних чинників, що радикально змінили демографічну ситуацію в Україні. За даними UNHCR та Міжнародної організації з міграції (ІОМ), від початку повномасштабного вторгнення за кордон виїхало від 6 до 8 мільйонів українців, переважна більшість з яких оселилася в країнах Європи [13].

Війна спричинила масштабне внутрішнє переміщення населення. Понад п'ять мільйонів українців змінили місце проживання всередині країни. Це призвело до суттєвих змін у розселенні населення та соціально-економічному навантаженні на регіони. Західні області стали основними приймаючими регіонами, де тимчасово зросла чисельність населення, тоді як у східних і південних регіонах вона різко скоротилася. Значною проблемою залишається повернення мігрантів після війни. Багато жінок з дітьми, які виїхали за кордон, адаптуються до життя в європейських країнах, що створює ризик довгострокових демографічних втрат для України [14].

Прогнози щодо відновлення народжуваності залишаються стриманими. За Стратегією демографічного розвитку України до 2040 року затвердженою Кабінетом Міністрів 30 вересня 2024 року навіть за оптимістичним сценарієм сумарний коефіцієнт народжуваності до 2040 року може зрости лише до 1,5 [11]. Таким чином народжуваність в Україні перебуває під одночасним тиском структурних довгострокових чинників та гострих воєнних факторів, що суттєво

посилюють депопуляційні тенденції і трансформують традиційну регіональну диференціацію.

1.2 Методологія зменшення розмірності та кластерного аналізу у дослідженні регіональної диференціації

Дослідження регіональної диференціації народжуваності передбачає роботу з багатовимірними даними, що описують кожен регіон одночасно за кількома соціально-економічними показниками. Для виявлення прихованої структури таких даних застосовують методи зменшення розмірності та кластерного аналізу. У даному дослідженні використовується метод головних компонент для зменшення розмірності та ієрархічна кластеризація разом з методом k-середніх для типологізації регіонів.

Метод головних компонент є одним з основних інструментів факторного аналізу, що дозволяє виявити приховану структуру багатовимірних даних шляхом зменшення їх розмірності. Суть методу полягає у перетворенні набору взаємопов'язаних змінних на менше число некорельованих між собою узагальнених факторів – головних компонент, – які зберігають максимальну частку інформації вихідних даних [17].

Процес проведення факторного аналізу методом головних компонент включає кілька етапів [18]:

- 1) збір та підготовка даних: вибір змінних, які потенційно впливають на досліджуване явище, та підготовка їх до аналізу;
- 2) обчислення матриці кореляцій;
- 3) виділення факторів: знаходження власних чисел і власних векторів матриці кореляцій та відбір значущих компонент за встановленими критеріями;
- 4) обертання факторів: використання процедур, наприклад, варімакс-обертання, для спрощення інтерпретації факторів;

5) інтерпретація факторів: аналіз матриці навантажень та присвоєння виділеним компонентам змістовних назв відповідно до змінних, які вони об'єднують.

Нехай є набір m спостережуваних змінних $X_1, X_2, \dots, X_m$. Їх взаємодія передбачає існування прихованих r факторів $F_1, F_2, \dots, F_r$, які встановлюються у результаті узагальнення вихідних ознак [17]. Розв'язання задач методом головних компонент полягає у виокремленні таких лінійних комбінацій випадкових величин, що мають максимально можливу дисперсію, при цьому спираються на коваріаційну або кореляційну матрицю цих величин.

Метод головних компонент надає можливість за даними m ознаками виокремити m головних компонент, або узагальнених ознак. В основу методу покладено таку лінійну модель (формула 1.2) [17]:

$$z_{ij} = a_{j1}f_{1i} + a_{j2}f_{2i} + \dots + a_{jm}f_{mi}, \quad (1.2)$$

де a_{jr} – ваговий коефіцієнт r -ої головної компоненти F_r для j -ої змінної Z_j ;

f_{ri} – значення головних компонент F_r за i -им об'єктом.

Спочатку за вихідними значеннями ознак $X_1, X_2, \dots, X_m$ обчислюються їх нормовані значення $Z_1, Z_2, \dots, Z_m$ та будується матриця Z нормованих значень вихідних ознак:

$$Z = (z_{ij}) \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1j} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2j} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{i1} & z_{i2} & \dots & z_{ij} & \dots & z_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nj} & \dots & z_{nm} \end{bmatrix}, \quad (1.3)$$

де z_{ij} – нормоване значення j -ої ознаки для i -го об'єкта:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad (1.4)$$

де $\bar{x}_j$ – середнє значення ознаки X_j ;

σ_j – середньоквадратичне відхилення ознаки X_j .

Розраховується матриця R парних кореляцій за формулою:

$$R = (r_{ij}) \begin{bmatrix} 1 & r_{12} & \cdots & r_{1j} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2j} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{i1} & r_{i2} & \cdots & r_{ij} & \cdots & z_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nj} & \cdots & 1 \end{bmatrix} = \frac{1}{n} Z^T Z \quad (1.5)$$

Вектор шуканих перших r головних компонент $F = (F_1, F_2, \dots, F_r)^T$ є лінійною комбінацією нормованих значень вихідних показників $Z_1, Z_2, \dots, Z_m$:

$$F = \tilde{L}Z, \quad (1.6)$$

де $\tilde{L}$ – нормована матриця власних векторів.

Щоб обчислити першу головну компоненту F_1 потрібно розв'язати оптимізаційну задачу:

$$\begin{cases} D(l_1 Z) \rightarrow \max_{l_1}; \\ l_1 l_1^T = 1 \end{cases}, \quad (1.7)$$

де l_1 – перший рядок матриці L .

Для визначення l_1 є система рівнянь, що має такий вигляд:

$$(R - \lambda_1 I_m) * l_1^T = 0, \quad (1.8)$$

де λ_1 – найбільше власне значення матриці R ;

I_m – одинична матриця розмірності m .

Для знаходження λ розв'язується характеристичне рівняння:

$$|R - \lambda_1 I_m| = 0 \quad (1.9)$$

Головна компонента F_r обчислюється за формулою:

$$F_1 = (Z) = l_1 Z \quad (r = \overline{1p}), \quad (1.10)$$

де l_r – власний вектор матриці R , що відповідає r -му за величиною власному числу λ_r цієї матриці;

p – кількість узагальнених ознак.

Щоб інтерпретувати головні компоненти, будується матриця навантажень A головних компонент на вихідні ознаки за формулою:

$$A = L^T * \lambda^{\frac{1}{2}}, \quad (1.11)$$

де $\lambda^{\frac{1}{2}}$ – діагональна матриця, r -ий елемент головної діагоналі якої є $\sqrt{\lambda_r}$.

Елементи матриці навантажень a_{ir} визначають одночасно ступінь щільності парного лінійного зв'язку між Z_i та F_r і питому вагу впливу проноормованої r -ої головної компоненти на ознаку Z_i .

Для визначення оптимальної кількості компонент застосовують два основних критерії. Критерій Кайзера передбачає відбір компонент з власними числами $\lambda > 1$ – компоненти з меншим власним числом пояснюють менше дисперсії, ніж одна вихідна стандартизована змінна [19]. Графік кам'яного осипу (scree plot) будується як залежність власних чисел від номера компоненти – відбираються компоненти до точки «зламу» після якої крива вирівнюється [20]. Додатковим критерієм є поріг накопиченої поясненої дисперсії – рекомендовано, щоб відібрані компоненти пояснювали не менше 75% загальної дисперсії вихідних даних [17].

Для покращення інтерпретованості факторів застосовується ортогональне обертання факторного простору. В ортогональному перетворенні найбільш популярними є два критерії [17, с. 246].

Квартімакс-критерій обчислюється за формулою:

$$Q = \sum_{j=1}^m \frac{\sum_{k=1}^r a_{jk}^4 - \left(\sum_{k=1}^r a_{jk}^2 \right)^2}{r^2}, \quad (1.12)$$

де a_{jk} — факторне навантаження j -ої змінної на k -ий фактор.

Метод квартімакс передбачає обертання факторних осей так щоб величини факторних навантажень максимізували Q і одночасно враховується якість структури усіх загальних факторів що беруть участь в аналізі.

Варімакс-критерій обчислює показник Q_r якості структури кожного фактора окремо:

$$Q_r = \sum_{j=1}^m \frac{m \sum_{j=1}^m a_{jr}^4 - \left(\sum_{j=1}^m a_{jr}^2 \right)^2}{m^2} \quad (1.13)$$

Варімакс-критерій спрощує стовпці матриці навантажень – кожен фактор повинен мати високі навантаження лише для окремих змінних і близькі до нуля для решти. Це полегшує змістовну інтерпретацію факторів оскільки кожен з них чітко асоціюється з певною групою змінних. Оптимальним вважається розв'язок, при якому значення критерію є максимальним.

На основі факторних змінних, отриманих у результаті факторного аналізу, можна проводити подальші дослідження, зокрема кластеризацію. Факторні змінні виступають узагальненими показниками, водночас зберігаючи основну інформацію. Це створює оптимальні умови для кластерного аналізу, який дозволяє об'єднувати об'єкти або території в групи з подібними характеристиками.

Використання факторних змінних у кластеризації допомагає підвищити точність та ефективність групування, оскільки вони знижують вплив надлишкових або корельованих змінних. Завдяки цьому підходу можна не лише ідентифікувати регіони чи групи населення зі схожими демографічними характеристиками, а й поглибити розуміння закономірностей, які впливають на народжуваність та інші демографічні показники.

Методи кластерного аналізу дають змогу класифікувати багатовимірні спостереження, які описуються набором вихідних змінних $X_1, X_2, \dots, X_m$. Метою є створення груп схожих між собою об'єктів – кластерів – з урахуванням усіх групувальних ознак одночасно [21].

Схожість або розбіжність між класифікованими об'єктами встановлюється залежно від метричної відстані між ними. Якщо i -ий об'єкт описується m ознаками, то він може бути представлений як точка в m -вимірному просторі і схожість з іншим k -им об'єктом визначатиметься як відповідна відстань. У кластерному аналізі використовуються різноманітні міри відстані між об'єктами, найпоширенішою з них є евклідова відстань, що визначається за формулою:

$$\rho_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - x_{kj})^2}, \quad (1.14)$$

де x_{ij} - значення j -ої ознаки для i -го об'єкта;

x_{kj} - значення j -ої ознаки для k -го об'єкта.

За принципом побудови методи кластерного аналізу поділяються на дві основні групи – агломеративні та дивізимні. Агломеративні методи працюють за принципом об'єднання – починаючи з окремих об'єктів вони послідовно формують все більші групи. Дивізимні методи діють у зворотному напрямку – розбивають початкову сукупність на дедалі менші підгрупи. Окремо виділяють ітеративні методи – зокрема метод k -середніх, який не належить жодній з двох груп. Їх особливість полягає в тому що кластери формуються на основі заздалегідь заданих параметрів розбиття, які користувач може змінювати в процесі роботи алгоритму для досягнення кращої якості результату [17].

Серед методів кластерного аналізу найширше застосування отримали ієрархічні агломеративні методи. На початку кожен об'єкт утворює окремий кластер – таким чином початкова кількість кластерів дорівнює кількості спостережень. Далі на кожному кроці два найближчі кластери об'єднуються в один на основі матриці відстаней або схожості. Процес продовжується до тих пір поки всі об'єкти не увійдуть до єдиного кластера – для вибірки розміром m це відбувається за $m-1$ кроків. Результати ієрархічної кластеризації наочно відображаються у вигляді дендрограми – деревоподібного графа, де вказуються об'єкти, що об'єднуються та рівень відстані при якому відбулось злиття [17].

Ієрархічні методи кластерного аналізу відрізняються між собою як мірами схожості, що використовуються так і алгоритмами об'єднання об'єктів. Найпоширенішими є чотири методи [17].

Метод одиночного зв'язку визначає відстань між кластерами як відстань між їх найближчими об'єктами. Спочатку об'єднуються два найближчих об'єкти вибірки, а на кожному наступному кроці до кластера приєднується той об'єкт, який є найближчим хоча б до одного з вже включених.

Метод повних зв'язків навпаки визначає відстань між кластерами як відстань між їх найвіддаленішими об'єктами. Новий об'єкт включається до кластера лише, якщо його відстань до кожного об'єкта кластера не перевищує заданого порогового рівня.

Метод середнього зв'язку використовує як міру відстані між кластерами середнє значення відстаней між усіма парами об'єктів, що належать різним кластерам. Об'єднання відбувається, якщо це середнє значення не перевищує встановленого порогу.

Метод Варда має за собою те, що на першому кроці кожний кластер складається з одного об'єкта. Спочатку об'єднуються два найближчих кластери. Для них визначається середнє значення кожної ознаки і обчислюється сума квадратів відхилень Q_k :

$$Q^k = \sum_{i=1}^{n_k} \sum_{j=1}^m (x_{ik} - \bar{x}_j^k)^2, \quad (1.15)$$

де k -номер кластера;

i - номер об'єкта;

j -номер ознаки;

m -кількість ознак, що характеризують кожний об'єкт;

n_k – кількість об'єктів у k -му кластері.

На кожному кроці алгоритму об'єднуються ті об'єкти або кластери, що дають найменший приріст Q^k .

Метод k -середніх є ітеративним методом, який розбиває вибірку на задану кількість k кластерів мінімізуючи цільову функцію:

$$W = \sum_k \sum_{i \in C_k} \|x_i - \mu_k\|^2, \quad (1.16)$$

де μ_k - центроїд (середнє значення) k -го кластера;

C_k - множина об'єктів, що належать k -му кластеру.

Алгоритм на кожній ітерації призначає кожен об'єкт до найближчого центроїда і перераховує центроїди доки розбиття не стабілізується.

Для визначення оптимального числа кластерів k застосовуються різні методи. Метод ліктя аналізує залежність внутрішньокластерної суми квадратів W від числа кластерів k – оптимальним вважається значення k у точці «зламу» графіка після якої подальше збільшення k дає незначне зменшення W [22].

Силуетний коефіцієнт запропонований Rousseeuw P. J. оцінює якість розбиття для кожного об'єкта [22]:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad (1.17)$$

де $a(i)$ - середня відстань від i -го об'єкта до всіх інших об'єктів його кластера;

$b(i)$ - середня відстань від i -го об'єкта до об'єктів найближчого сусіднього кластера.

Коефіцієнт приймає значення від -1 до +1 - чим вище значення, тим краще об'єкт відповідає своєму кластеру. Оптимальна кількість кластерів відповідає максимальному середньому значенню силуетного коефіцієнта по всій вибірці.

1.3 Регресійний аналіз як інструмент дослідження впливу соціально-економічних чинників на народжуваність

Регресійний аналіз дозволяє не лише зафіксувати сам факт зв'язку між економічними показниками, а й кількісно оцінити його силу та напрям. Коли на

результативну ознаку одночасно діє кілька чинників, застосовують множинну лінійну регресію, загальне рівняння якої має вигляд [17, с. 144]:

$$y_i = a_0 + a_1x_{1i} + a_2x_{2i} + \dots + a_mx_{mi} + u_i, \quad (1.18)$$

де y_i - фактичне значення залежної змінної для i -го спостереження;

a_0 - константа;

$a_1, \dots, a_m$ - коефіцієнти регресії;

$x_{1i}, \dots, x_{mi}$ - значення пояснювальних змінних;

m - кількість чинників;

u_i - випадкова похибка.

Коефіцієнт a_m показує, на скільки одиниць змінюється y при зростанні x_k на одиницю за незмінності всіх інших чинників – тобто відображає *ceteris paribus* ефект. Вибір переліку чинників і форми залежності називають специфікацією моделі. Розрахункові значення залежної змінної визначаються підстановкою оцінених коефіцієнтів у рівняння (1.18) без похибки.

Оцінювання параметрів регресії здійснюється за допомогою методу найменших квадратів (МНК), який мінімізує суму квадратів відхилень фактичних значень від розрахункових [17, с. 143; 25, с. 131]:

$$\sum (y_i - \hat{y}_i)^2 \rightarrow \min \quad (1.19)$$

У матричному записі вектор оцінок знаходиться за формулою [1, с. 144]:

$$A = (X^T X)^{-1} X^T Y \quad (1.20)$$

де A - вектор оцінок параметрів;

X - матриця спостережень (перший стовпець - одиниці, решта - значення чинників);

Y - вектор значень залежної змінної.

Якщо виконуються класичні умови (відсутність мультиколінеарності, нульове сподівання похибок, стала дисперсія, відсутність автокореляції), то згідно з теоремою Гаусса-Маркова МНК – оцінки є BLUE (Best Linear Unbiased Estimators) – тобто лінійними, незміщеними й ефективними [25 с. 132; 23].

Після побудови рівняння важливо оцінити, наскільки воно добре описує дані. Головний показник – коефіцієнт детермінації R^2 , який показує частку варіації залежної змінної, що пояснює модель [17, с. 146]:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad (1.21)$$

Значення $R^2 \in [0; 1]$: що ближче до 1, то краща модель. Проте додавання нових змінних механічно підвищує R^2 , тому для порівняння моделей з різною кількістю регресорів використовують скоригований коефіцієнт [24, с. 82]:

$$\bar{R}_2 = 1 - \frac{n-1}{n-m-1} (1 - R^2) \quad (1.22)$$

Загальна значущість моделі перевіряється за F -критерієм Фішера [17, с. 146]:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m} \quad (1.23)$$

Модель визнається статистично значущою, якщо розраховане значення F -критерію перевищує критичне табличне при обраному рівні значущості α . Значущість окремих коефіцієнтів оцінюють за t -критерієм Стюдента [17, с. 147]:

$$t_{a_j} = \frac{a_j}{s_{a_j}} \quad (1.24)$$

де a_j - оцінка коефіцієнта;

s_{a_j} - його стандартна помилка.

Якщо модуль розрахованого t -значення для коефіцієнта перевищує критичне табличне значення при обраному рівні значущості, відповідний коефіцієнт регресії вважається статистично значущим, тобто його вплив на залежну змінну не є випадковим.

Коректність МНК-оцінок залежить від виконання низки передумов. Їх порушення робить оцінки зміщеними або неефективними. Основні умови такі:

- нормальний розподіл залишків: $E(u_i) = 0$;
- гомоскедастичність – стала дисперсія похибок: $D(u_i) = \sigma^2 = \text{const}$;

- відсутність автокореляції: $\text{cov}(u_i, u_j) = 0$ при $i \neq j$;
- відсутність мультиколінеарності між пояснювальними змінними;
- екзогенність – незалежність чинників від похибки.

Нормальність залишків перевіряється тестом Шапіро-Вілکا: якщо p -значення $> 0,05$, підстав відхиляти нормальність немає [17].

Гетероскедастичність виявляється тестом Бройша-Пагана: нульова гіпотеза – гомоскедастичність. При $p < 0,05$ потрібні корективи, наприклад, у вигляді зваженого МНК [17].

Автокореляція залишків діагностується критерієм Дарбіна-Уотсона [17, с. 153]:

$$DW = \frac{\sum_{i=2}^n (u_i - u_{i-1})^2}{\sum_{i=1}^n u_i^2} \quad (1.25)$$

Показник приймає значення від 0 до 4. При відсутності автокореляції він наближається до 2. Фактичні значення критерію порівнюються з табличними – нижньою і верхньою межею (DW_1 і DW_2 відповідно). Якщо $DW < DW_1$ – є додатна автокореляція; $DW > DW_2$ – автокореляція відсутня; у проміжку $[DW_1; DW_2]$ – зона невизначеності [17, с. 153].

Мультиколінеарність кількісно оцінюється через показник VIF – коефіцієнт росту дисперсії [23, р 92]:

$$VIF_j = \frac{1}{(1 - R_j^2)} \quad (1.26)$$

де R_j^2 - коефіцієнт детермінації регресії x_j на всі інші пояснювальні змінні моделі.

Орієнтири інтерпретації: $VIF < 5$ – мультиколінеарність незначна; $VIF \in [5; 10]$ – помірна, варто звернути увагу; $VIF > 10$ – критична, що суттєво знижує надійність оцінок [23, р 92].

Мультиколінеарність – це наявність тісного лінійного зв'язку між двома чи більше пояснювальними змінними [17, с. 148]. Розрізняють функціональну (точна лінійна залежність, матриця $(X^T X)$ вироджена, визначник дорівнює нулю – МНК-оцінки взагалі не можна знайти) та стохастичну форму (тісний, але не

функціональний зв'язок). Остання є типовою для соціально-економічних даних: оцінки технічно знаходяться, але їхні стандартні похибки різко зростають. Наслідками є нестабільність коефіцієнтів, можлива зміна їхніх знаків, суперечність між значущою загальною статистикою F і незначущими t -статистиками окремих коефіцієнтів. Простим попереднім індикатором наявності мультиколінеарності є матриця парних кореляцій між регресорами: при $|r_{ij}| \geq 0,8$ мультиколінеарність присутня [17, с. 150].

Помилки специфікації виникають, коли структура моделі сформована неправильно. На практиці розрізняють три основних типи [23]:

- пропуск суттєвої змінної – оцінки параметрів інших чинників стають зміщеними;
- включення зайвої змінної – оцінки залишаються незміщеними, але втрачають ефективність;
- неправильна функціональна форма – якщо реальна залежність нелінійна, а специфікована лінійна, модель не здатна адекватно відтворити процес.

Для врахування якісних характеристик у регресійну модель вводять фіктивні (дамі) змінні, які набувають значень 0 або 1 [23]. Якщо якісна ознака має k категорій, до моделі включають $k - 1$ фіктивну змінну – щоб уникнути ідеальної мультиколінеарності з константою (так звана «пастка фіктивних змінних»). Коефіцієнт при фіктивній змінній інтерпретується як відхилення середнього рівня залежної змінної у відповідній категорії від базової за інших рівних умов.

РОЗДІЛ 2

ЕКОНОМЕТРИЧНЕ МОДЕЛЮВАННЯ ТА ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ У ДОСЛІДЖЕННІ НАРОДЖУВАНOSTІ В РЕГІОНАХ УКРАЇНИ

2.1. Формування інформаційної бази та підготовка даних

Дослідження регіональних особливостей народжуваності в Україні проводиться на основі двох інформаційних масивів: даних за 2021 рік – до початку повномасштабного вторгнення – та даних за 2025 рік, що відображають ситуацію в умовах тривалої війни. Така структура інформаційної бази зумовлена об'єктивними обставинами: повномасштабне вторгнення Росії в Україну у лютому 2022 року суттєво змінило як демографічну ситуацію в регіонах, так і доступність статистичних даних. Використання даних періоду до повномасштабного вторгнення дасть змогу визначити базові закономірності регіональних відмінностей у народжуваності, натомість аналіз сучасних даних покаже зміни, що відбулися під впливом повномасштабної війни.

2.1.1. Формування інформаційної бази за період до повномасштабного вторгнення. Для першого блоку дослідження сформована база даних за 2021 рік по 24 областях України на основі офіційних матеріалів Державної служби статистики України, зокрема статистичного збірника «Регіони України» 2021 року [3;9]. Місто Київ як столиця не включене до дослідження через його специфічні соціально-економічні умови, які значно відрізняються від інших регіонів і можуть створювати аномалії в розрахунках.

Дані для аналізу за 2021 рік містять такі показники:

- `birth_rate` – коефіцієнт народжуваності на 1000 осіб постійного населення. Є основним результативним показником дослідження та характеризує рівень відтворення населення у регіоні;
- `women_share` – частка жінок репродуктивного віку (15-49 років) у чисельності постійного населення (%). Відображає демографічну структуру регіону та безпосередньо визначає демографічний потенціал народжуваності;
- `unemployment` – рівень безробіття населення віком 15-70 років за методологією МОП (%). Характеризує ситуацію на ринку праці та економічну стабільність домогосподарств;
- `preschool` – забезпеченість закладами дошкільної освіти на 10000 осіб постійного населення. Відображає рівень розвитку соціальної інфраструктури регіону та доступність послуг догляду за дітьми;
- `doctors` – забезпеченість лікарями усіх спеціальностей на 10000 осіб постійного населення. Характеризує рівень розвитку медичної інфраструктури;
- `cpi` – індекс споживчих цін (%). Показує інфляційні процеси та зміну вартості життя;
- `wage` – середньомісячна номінальна заробітна плата (грн). Відображає рівень доходів населення та економічні можливості домогосподарств;
- `ecology` – викиди забруднюючих речовин і парникових газів в атмосферне повітря від стаціонарних джерел забруднення у розрахунку на одну особу (кг). Показує рівень штучного забруднення природи в регіоні – тобто той екологічний фактор, який безпосередньо визначає здоров'я людей та загальний комфорт їхнього проживання.

Використання відносних показників дає можливість мінімізувати вплив масштабу регіону та забезпечує коректність подальшого багатовимірного аналізу.

Зібрані дані були структуровані у вигляді єдиної таблиці, що наведена в додатку А – табл. А. 1.

2.1.2. Формування інформаційної бази в умовах повномасштабної війни. Для другого блоку дослідження сформовано базу даних за 2025 роком по тих самих 24 областях України на основі даних Державної служби статистики України та Міністерства юстиції України [2; 3]. Виокремлення даних за актуальний період пов'язане з суттєвими обмеженнями доступності статистичних даних зумовленими повномасштабним вторгненням. Зокрема Державна служба статистики України після 2021 року припинила публікацію регіональних даних про рівень безробіття за методологією МОП, а також даних про забезпеченість лікарями та чисельність населення по областях. У зв'язку з цим набір змінних для аналізу за 2025 рік є меншим порівняно з 2021 роком.

До складу дослідження за 2025 рік увійшли такі показники:

- birth – кількість народжених по регіонах (осіб). Є основним результативним показником. Використовується в абсолютному вираженні у зв'язку з відсутністю актуальних даних про чисельність населення по регіонах. Для Херсонської, Запорізької, Донецької та Луганської області наведені дані з підконтрольних українській владі територій [2];

- wage – середньомісячна номінальна заробітна плата (грн);
- cpi – індекс споживчих цін (%);
- preschool – кількість закладів дошкільної освіти (одиниць). Оскільки офіційні дані Державної служби статистики доступні лише до 2024 року включно, то значення за 2025 рік розраховані методом екстраполяції лінійного тренду на основі даних 2019-2024 років. Для Запорізької та Херсонської областей тренд розраховувався за 2022-2024 роками у зв'язку з аномальними значеннями попередніх років через активні бойові дії;

- wage_debt – заборгованість із виплати заробітної плати в середньому на одного працівника (грн). Використовується як індикатор стану ринку праці замість показника рівня безробіття за методологією МОП, публікацію якого по регіонах було припинено після 2021 року.

Окрім кількісних змінних для побудови регресійної моделі введено дві фіктивні змінні, які відображають вплив повномасштабної війни на регіональні особливості народжуваності:

- *frontline* – фіктивна змінна, що приймає значення 1 для регіонів, які станом на 2025 рік містять території активних бойових дій або тимчасово окупованих територій: Дніпропетровська, Запорізька, Миколаївська, Сумська, Харківська та Херсонська області. Регіони віднесені до цієї категорії відповідно до Переліку територій на яких ведуться (велися) бойові дії або тимчасово окупованих Російською Федерацією затвердженого Наказом Міністерства розвитку громад та територій України від 28.02.2025 № 376 [26];

- *recipient* – фіктивна змінна, що приймає значення 1 для регіонів, які прийняли понад 100000 внутрішньо переміщених осіб станом на грудень 2025 року та не віднесені до прифронтових: Київська, Одеська, Полтавська, Львівська, Черкаська, Вінницька, Кіровоградська. Дані про кількість ВПО по регіонах отримано з Звіту МОМ про внутрішнє переміщення населення в Україні, Раунд 22 (січень 2026 року) [15].

Зведені дані за всіма показниками за 2025 рік наведено в додатку Б – табл. Б. 1.

2.2. Багатовимірний статистичний аналіз народжуваності регіонів України у період до повномасштабного вторгнення

Для виявлення регіональних закономірностей народжуваності в Україні за період до повномасштабного вторгнення застосовується декілька методів багатовимірного статистичного аналізу. Дослідження проводиться у кілька послідовних етапів: описова статистика для характеристики розподілу показників та виявлення аномальних значень, кореляційний аналіз для визначення сили та напрямку зв'язків між змінними, метод головних компонент для зменшення

розмірності даних та усунення мультиколінеарності, а також кластерний аналіз для типологізації регіонів за сукупністю соціально-економічних умов. Всі розрахунки виконуватимуться у середовищі Jupyter Notebook на мові програмування Python із використанням бібліотек pandas, numpy, scikit-learn, seaborn та matplotlib.

Завантажимо дані до зазначеного середовища та проаналізуємо описову статистику всіх показників, що наведено в табл. 2. 1.

Таблиця 2. 1 – Результати описової статистики для показників (2021 р.)

	birth rate	women share	unemployment	preschool	doctors	cpi	wage	ecology
count	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00
mean	6.80	23.22	10.72	4.18	35.54	110.00	12412.46	40.03
std	2.02	0.85	2.38	1.21	10.52	0.83	1016.21	51.03
min	1.60	21.47	6.70	1.24	8.80	108.40	11180.00	1.87
25%	5.97	22.71	9.20	3.40	31.08	109.57	11617.50	10.81
50%	7.03	23.09	11.05	4.61	35.53	110.10	12205.50	17.30
75%	7.84	24.01	12.02	4.98	41.95	110.43	12818.50	39.06
max	10.25	24.59	16.00	5.79	52.43	111.70	14853.00	182.39

Джерело: розроблено автором на основі [3; 9]

Описова статистика показує значні регіональні відмінності у демографічних, економічних та соціальних показниках між областями. Середній коефіцієнт народжуваності становить 6,80 на 1000 осіб постійного населення, а стандартне відхилення – 2,02 – вказує на суттєву варіацію між регіонами.

Частка жінок репродуктивного віку варіюється від 21,47% до 24,59% при середньому 23,22%, тобто розкид між регіонами є незначним порівняно з іншими показниками.

Деяким з показників притаманні значні відмінності між регіонами. Наприклад, забезпеченість лікарями варіюється від 8,80 до 52,43 на 10000 осіб, що демонструє нерівність у доступі до медичних послуг. Забезпеченість закладами дошкільної освіти коливається від 1,24 до 5,79 на 10000 осіб. Екологічний стан регіонів у вигляді викидів забруднюючих речовин і парникових газів в атмосферне повітря знаходиться у значному діапазоні – від 1,87 до 182,39 кг на особу.

Індекс споживчих цін демонструє найменший розкид серед усіх показників: значення варіюються від 108,4% до 111,7%, що становить лише 3,3 відсоткових пункти. Середня заробітна плата також характеризується помірною

диференціацією.

Для візуального відображення розподілу значень показників побудуємо діаграми розмаху (рис. 2. 1).

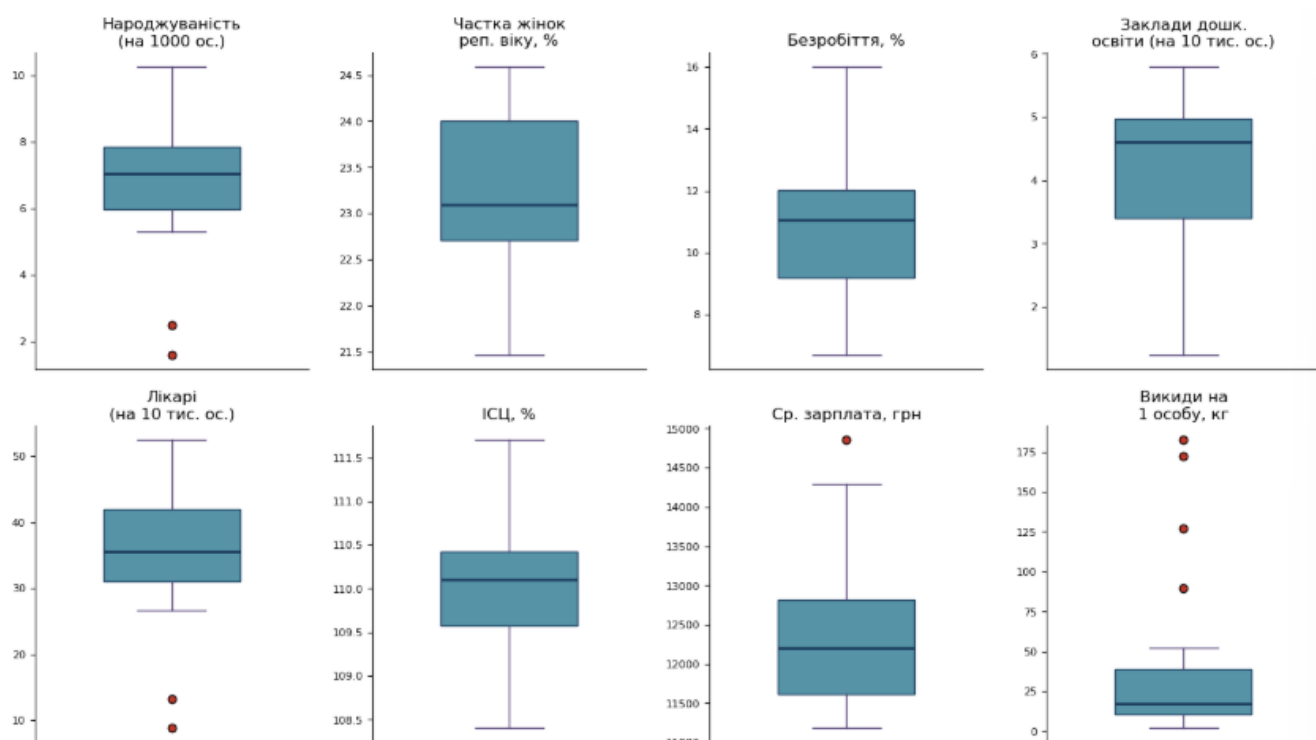


Рисунок 2. 1 – Діаграми розмаху показників по регіонах України (2021 р.)

Джерело: розроблено автором на основі [3; 9]

На діаграмах розмаху помітні окремі викиди (позначені червоними крапками), що відповідають регіонам із критично заниженими або завищеними значеннями показників, зумовленими специфічними умовами цих областей. Для показників народжуваності, доступності медицини та середньої заробітної плати такими викидами є Донецька та Луганська області, для екологічного показника викиди вгору на діаграмі розмаху відповідають Донецькій, Дніпропетровській, Івано-Франківській та Запорізькій областям, де сконцентровані великі металургійні, хімічні та енергетичні підприємства.

Окремо розглянемо питання регіональної вибірки. Донецька та Луганська області мають критично занижені значення народжуваності (1,60 та 2,50 на 1000 осіб при середньому 6,80) та аномальні значення деяких інших показників, що є наслідком можливих неповних даних унаслідок збройного конфлікту, як

зазначається в Державній службі статистики України [3], а не відображенням реальних соціально-економічних закономірностей. Їх включення до методу головних компонент та кластеризації призвело б до того, що перша компонента фактично описувала б лише різницю між цими двома аномальними регіонами та рештою вибірки, що спотворило б результати факторного аналізу. Аналогічно, їх включення до кластеризації призвело б до того, що ці два регіони утворили б окремий кластер, який відображав би лише наслідки збройного конфлікту, а не реальні відмінності між областями. Тому Донецька та Луганська області будуть вилучені з подальшого аналізу, а всі наступні розрахунки проводитимуться на вибірці з 22 областей.

Відмінності між регіонами підтверджують необхідність багатовимірного аналізу для виявлення основних чинників впливу на рівень народжуваності. Першим кроком такого аналізу є кореляційний аналіз, який дозволяє визначити силу та напрям зв'язку між показниками народжуваності та потенційними факторами впливу [27, с. 78].

Стандартизуємо дані для їх подальшого дослідження – значення змінних після цього матимуть середнє значення 0 і стандартне відхилення 1, що дозволить порівнювати змінні з різними одиницями виміру.

Для початку потрібно побачити, які змінні добре корелюють із залежною змінною – коефіцієнтом народжуваності – і чи є змінні, що значно корелюють між собою. Подивимося на кореляцію між змінними за допомогою кольорової мапи кореляцій (рис. 2. 2).

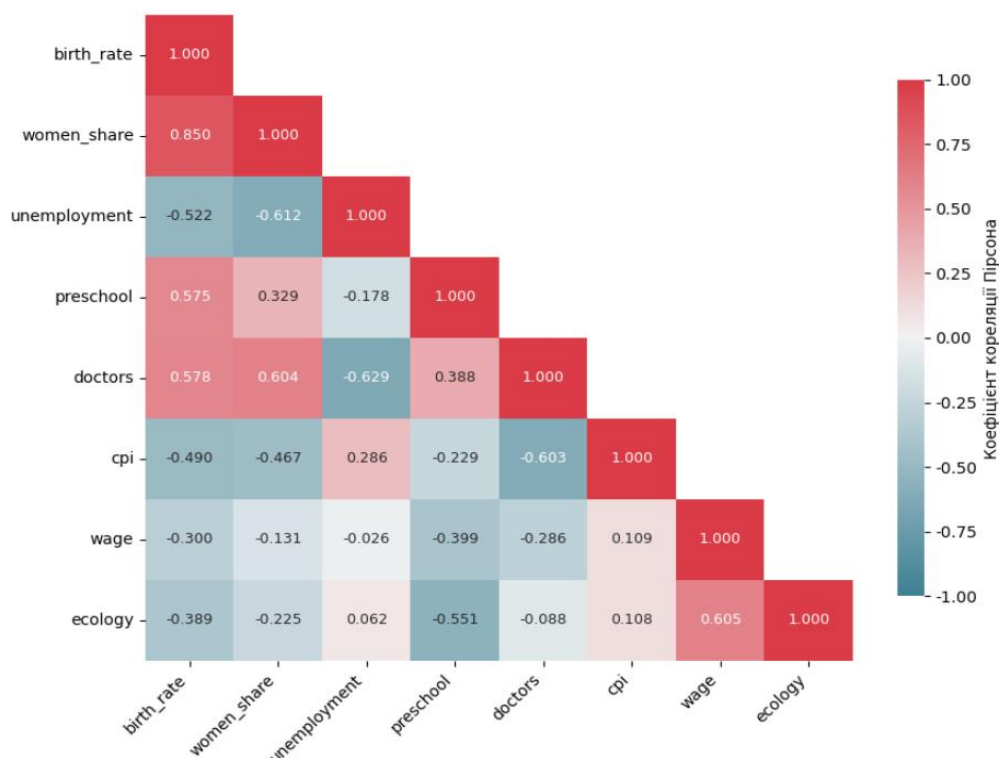


Рисунок 2. 2 – Кольорова мапа кореляцій (22 регіони, 2021 р.)

Джерело: розроблено автором на основі [3; 9]

Використаємо таку оцінку сили кореляційного зв'язку за шкалою Чеддока: практично відсутній ($|r| < 0,3$), слабкий ($0,3 \leq |r| < 0,5$), помірний ($0,5 < |r| \leq 0,7$), високий ($0,7 \leq |r| < 0,9$), дуже високий ($0,9 \leq |r| < 1,0$) [27, с 80].

Сильну кореляцію з цільовою змінною «коефіцієнт народжуваності» має лише частка жінок репродуктивного віку ($r = +0,850$). Помірну кореляцію демонструють забезпеченість лікарями ($r = +0,578$), забезпеченість закладами дошкільної освіти ($r = +0,575$) та рівень безробіття ($r = -0,522$). Слабкий зв'язок з народжуваністю мають індекс споживчих цін ($r = -0,490$), екологічний показник регіону ($r = +0,389$) та середня заробітна плата ($r = -0,300$). Жоден із показників не має практично відсутнього зв'язку з народжуваністю, тому всі сім факторних змінних зберігаються для подальшого аналізу.

Під час кореляційного аналізу також виявлено пари змінних з помірним взаємозв'язком між собою ($|r| > 0,5$):

- частка жінок репродуктивного віку та рівень безробіття ($r = -0,612$);
- частка жінок репродуктивного віку та забезпеченість лікарями ($r = +0,604$);

- рівень безробіття та забезпеченість лікарями ($r = -0,629$);
- забезпеченість лікарями та індекс споживчих цін ($r = -0,603$);
- забезпеченість закладами дошкільної освіти та екологічний показник регіону ($r = -0,551$);
- середня заробітна плата та екологічний показник регіону ($r = 0,605$).

Це є додатковою підставою до застосування методу головних компонент, який усуне кореляцію факторів між собою [17].

Щоб стиснути розмірність даних і при цьому не втратити важливу інформацію, скористаємося методом головних компонент (PCA – Principal Component Analysis).

Суть методу головних компонент полягає у зменшенні кількості змінних, перетворенні набору корельованих змінних на нові некорельовані між собою компоненти. Кожна компонента є лінійною комбінацією вихідних змінних і впорядкована за кількістю дисперсії, яку вона пояснює [17]. Оскільки у кореляційному аналізі було виявлено значні зв'язки між факторними змінними, застосування PCA дозволяє усунути мультиколінеарність та звести 7 незалежних показників до меншої кількості компонент без суттєвої втрати інформації.

У табл. 2.2 показані основні характеристики можливих головних компонент.

Таблиця 2.2 – Власні числа та пояснена дисперсія головних компонент (2021 р.)

Компонента	Власне число (λ)	Частка дисперсії, %	Накопичена дисперсія, %
PC1	2.6515	36.2	36.2
PC2	1.6861	23.0	59.1
PC3	1.0454	14.3	73.4
PC4	0.8689	11.8	85.3
PC5	0.4408	6.0	91.3
PC6	0.4119	5.6	96.9
PC7	0.2288	3.1	100.0

Джерело: розроблено автором на основі [3; 9]

Для обґрунтування вибору кількості компонент застосуємо два критерії: критерій Кайзера ($\lambda > 1$) та графік «кам'яного осипу». За критерієм Кайзера відбираємо компоненти, власне число яких перевищує 1, оскільки лише такі

компоненти пояснюють більше дисперсії, ніж одна вихідна стандартизована змінна [19]. Графік «кам'яного осипу» дозволяє візуально визначити точку «зламу», після якої додавання нових компонент не дає суттєвого приросту поясненої дисперсії [20] (рис. 2. 3).

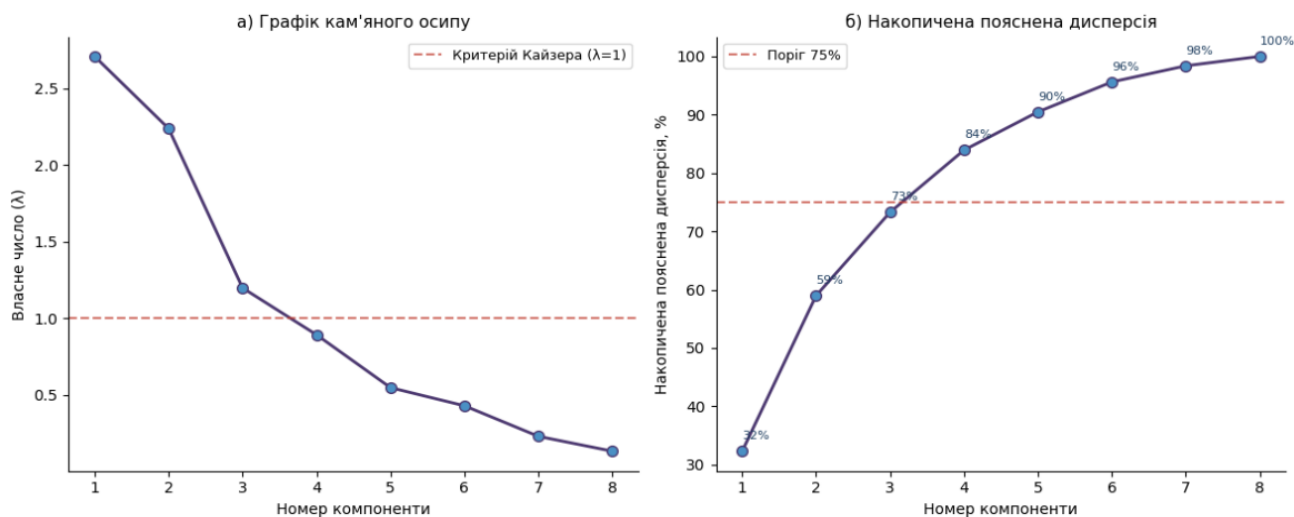


Рисунок 2. 3 – Графік «кам'яного осипу» та накопиченої дисперсії (2021 р.)

Джерело: розроблено автором на основі [3; 9]

Відповідно до критерію Кайзера відібрано 3 головні компоненти з власними числами $\lambda_1 = 2,6515$, $\lambda_2 = 1,6861$ та $\lambda_3 = 1,0454$. На графіку «кам'яного осипу» після третьої компоненти спостерігається виразний злам – подальші компоненти мають власні числа нижче 1. Три відібрані компоненти у сукупності пояснюють 73,4% загальної дисперсії вихідних даних, що є близьким до рекомендованого порогу 75% [17, с. 239].

Для інтерпретації змісту кожної компоненти розглянемо матрицю факторних навантажень (рис. 2. 4 і табл. 2. 3), до якої було застосовано обертання Varimax normalized, яке максимізує варіацію навантажень у кожному факторі: великі навантаження стають ще більшими, а малі – ще меншими [17, с. 246]. Обертання не змінює загальну пояснену дисперсію (73,4%), але перерозподіляє її між факторами рівномірніше та робить інтерпретацію кожного фактора значно чіткішою. Навантаження відображає кореляцію між вихідною змінною та головною компонентою – що більше за абсолютним значенням, то більший внесок змінної у компоненту [28, с. 9].

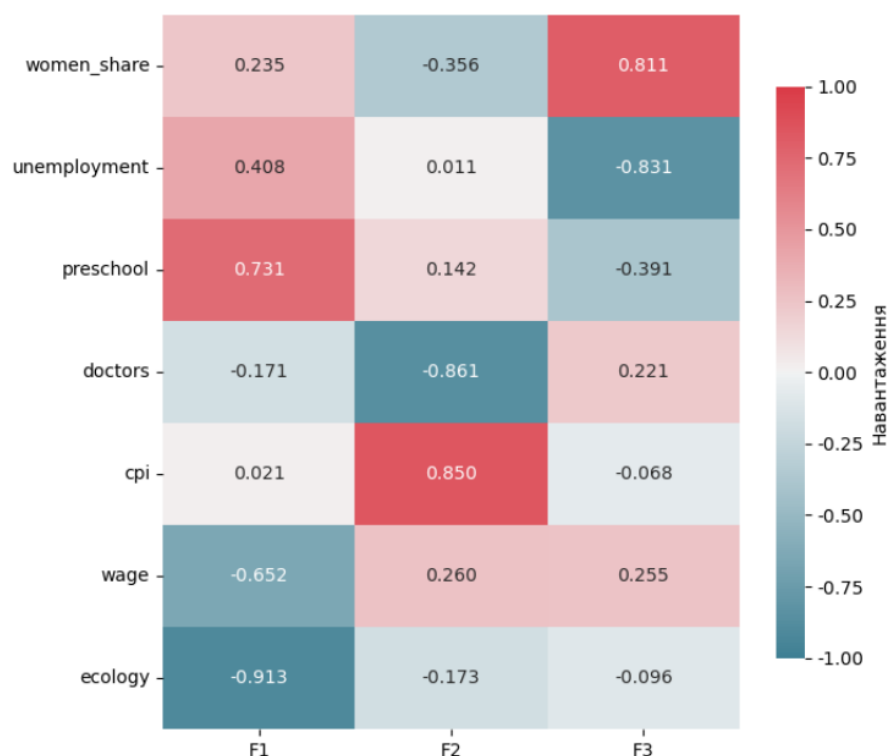


Рисунок 2. 4 – Теплова карта матриці факторних навантажень (Varimax, 2021 р.)

Джерело: розроблено автором на основі [3; 9]

Таблиця 2. 3 – Матриця факторних навантажень після обертання Varimax (2021 р.)

Показник	F1	F2	F3
women_share	+0.235	-0.356	+0.811
unemployment	+0.408	+0.011	-0.831
preschool	+0.731	+0.142	-0.391
doctors	-0.171	-0.861	+0.221
cpi	+0.021	+0.850	-0.068
wage	-0.652	+0.260	+0.255
ecology	-0.913	-0.173	-0.096

Джерело: розроблено автором на основі [3; 9]

Аналіз матриці факторних навантажень після обертання Varimax дозволяє дати змістовне пояснення кожній з виділених компонент.

Перший фактор F1 можна охарактеризувати як «Соціально-екологічний». На нього найбільше впливають викиди забруднюючих речовин на одну особу з від'ємним знаком (-0,913), забезпеченість закладами дошкільної освіти з позитивним (+0,731) та рівень заробітної плати (-0,652). Від'ємний знак при викидах означає, що регіони з вищими значеннями цього фактора мають менше промислове навантаження на довкілля. Водночас висока забезпеченість

дошкільними закладами свідчить про розвинену інфраструктуру для сімей. Такий профіль притаманний переважно аграрним регіонам з чистішим природним середовищем.

Другий фактор F2 – «Медична інфраструктура та цінова стабільність». Його визначають забезпеченість лікарями з від'ємним знаком (-0,861) та індекс споживчих цін з позитивним (+0,850). Низькі значення цього фактора відповідають регіонам, де краще розвинена медична мережа і нижча інфляція – тобто загалом кращі умови для проживання.

Третій фактор F3 – «Демографічний потенціал та зайнятість». Найбільший вплив мають частка жінок репродуктивного віку (+0,811) та рівень безробіття з від'ємним знаком (-0,831). Вищі значення фактора відповідають регіонам, де більша частка жінок репродуктивного віку поєднується з нижчим безробіттям, тобто демографічно активним регіонам з розвиненим ринком праці.

Найбільший внесок у загальну дисперсію робить F1 (37,9%), дещо менший F2 (24,1%) і F3 (11,5%).

Факторні оцінки регіонів за трьома компонентами будуть використані як вхідні дані для кластерного аналізу.

Для типологізації регіонів України за рівнем народжуваності та соціально-економічними умовами застосуємо кластерний аналіз. Вхідною матрицею слугують три фактори PCA (F1, F2, F3) та стандартизований коефіцієнт народжуваності. Кластеризація виконується у два етапи: спочатку ієрархічна кластеризація для візуальної оцінки природної структури даних та попереднього визначення кількості кластерів, після чого метод k-середніх для отримання фінального розбиття.

Ієрархічна кластеризація дозволяє візуально оцінити природну структуру даних та зробити попереднє припущення про кількість кластерів до застосування k-середніх. Застосуємо метод Варда, який мінімізує внутрішньокластерну дисперсію на кожному кроці злиття та є найбільш ефективним для виявлення компактних однорідних груп [17, с. 268]. Результат відображено у вигляді дендрограми (рис. 2. 5).

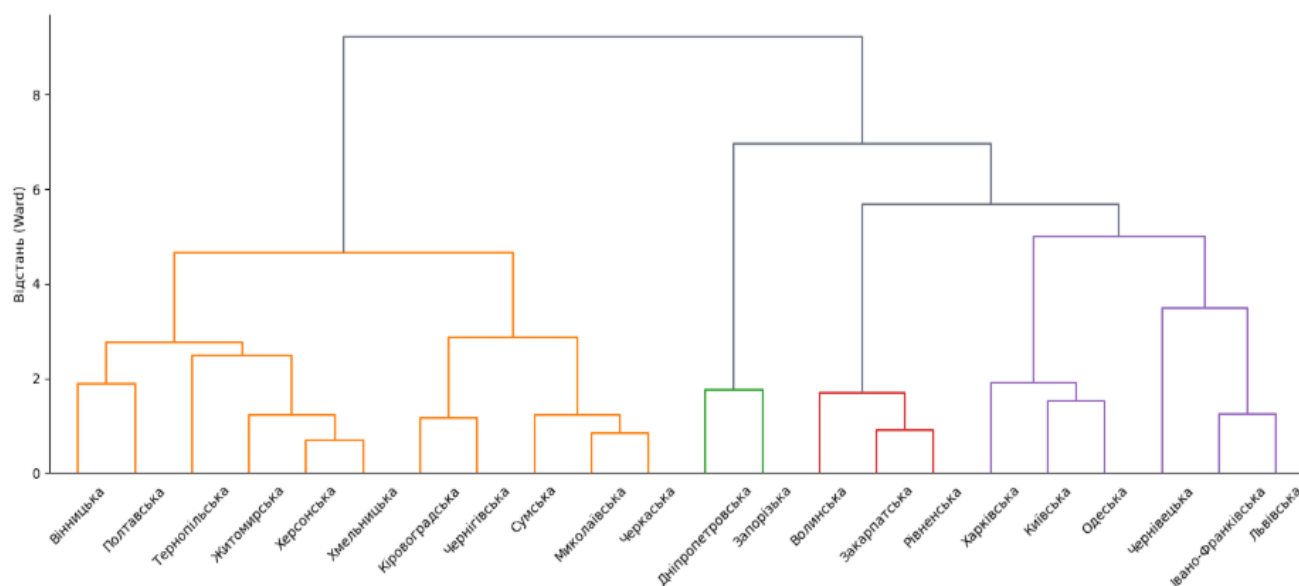


Рисунок 2. 5 – Дендрограма ієрархічної кластеризації (метод Варда, 2021 р.)
 Джерело: розроблено автором на основі [3; 9]

Дендрограма відображає природну структуру злиттів регіонів і дозволяє попередньо виділити чотири кластери. Найбільша група охоплює центральні та південні регіони і формує єдину гілку на рівні відстані близько 5, що свідчить про помірну схожість між ними. Дніпропетровська і Запорізька утворюють окрему пару – два великих промислових регіони зі схожим соціально-економічним профілем. Волинська, Закарпатська і Рівненська формують групу з традиційно вищою народжуваністю. Четверта гілка об'єднує великі міські центри (Харківська, Київська, Одеська області) та західноукраїнські регіони (Чернівецька, Івано-Франківська, Львівська області). Відстань на вертикальній осі відображає ступінь схожості між регіонами – чим більша відстань, тим менш схожі регіони між собою. Наприклад Дніпропетровська і Запорізька приєднуються до решти вибірки лише на рівні відстані близько 7, що підкреслює їх відмінність від інших кластерів.

Для додаткового підтвердження оптимальної кількості кластерів застосуємо метод «ліктя» та силуетний аналіз (рис. 2.6).

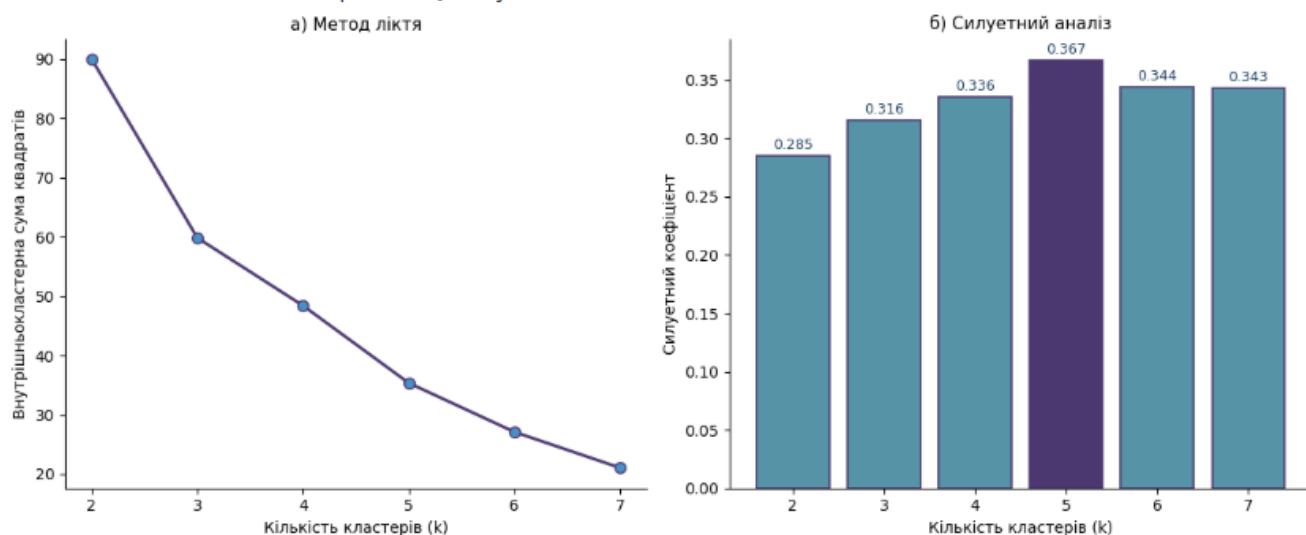


Рисунок 2. 6 – Вибір оптимальної кількості кластерів (метод «ліктя» та силуетний аналіз, 2021 р.)

Джерело: розроблено автором на основі [3; 9]

На графіку методу «ліктя» найбільш виражений злам спостерігається при переході від $k=2$ до $k=3$. Після цього темп зменшення суми квадратів поступово сповільнюється – між $k=4$ і $k=5$ ще помітний спад на 15 одиниць, тоді як починаючи з $k=5$ крива вирівнюється і спадає рівномірно. Це свідчить що додавання шостого і наступних кластерів вже не дає суттєвого приросту якості розбиття.

Силуетний аналіз вказує на оптимальність $k=5$ – силуетний коефіцієнт досягає максимального значення 0,367 і помітно знижується при $k=6$ (0,344). Ієрархічна кластеризація натомість виокремлює чотири чи шість груп залежно від обраного рівня розрізу (на відстані 4 і 5 відповідно). Оскільки метод ліктя і силуетний аналіз узгоджено вказують на $k=5$ – обрано саме цю кількість кластерів для подальшого аналізу.

Для отримання фінального розбиття регіонів застосовано метод k -середніх з $k=5$. Результати розподілу 22 регіонів за п'ятьма кластерами наведено в табл. 2. 4.

Таблиця 2. 4 – Розподіл регіонів України за кластерами (k -середніх, 2021 р.)

Кластер 1	Кластер 2	Кластер 3	Кластер 4	Кластер 5
Волинська	Вінницька	Івано-Франківська	Кіровоградська	Дніпропетровська
Закарпатська	Житомирська	Київська	Миколаївська	Запорізька
Рівненська	Тернопільська	Львівська	Полтавська	
	Чернівецька	Одеська	Сумська	

Продовження таблиці 2.4

Кластер 1	Кластер 2	Кластер 3	Кластер 4	Кластер 5
		Харківська	Херсонська	
			Хмельницька	
			Черкаська	
			Чернігівська	

Джерело: розроблено автором на основі [3; 9]

Як видно з табл. 2.4 розподіл регіонів між кластерами є нерівномірним – найбільша група налічує 8 регіонів, тоді як найменша – 2. Для оцінки демографічних відмінностей між групами розглянемо середній коефіцієнт народжуваності для кожного з отриманих кластерів (рис. 2. 7).

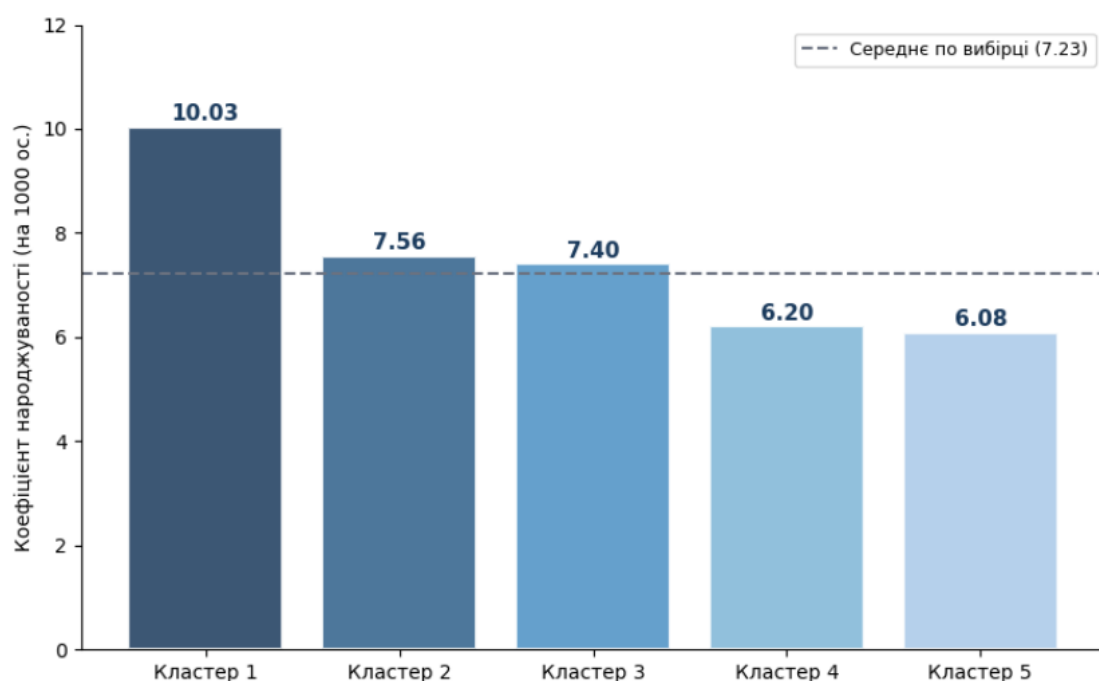


Рисунок 2. 7 – Середній коефіцієнт народжуваності по кластерах (2021 р.)

Джерело: розроблено автором на основі [3; 9]

Найвище значення має кластер 1 – 10,03 народжених на 1000 осіб, що значно перевищує середнє по вибірці (7,23). Кластери 2 і 3 мають близькі між собою показники – 7,56 і 7,40 відповідно, що є трохи вищим за середнє. Кластери 4 і 5 демонструють найнижчі значення народжуваності – 6,20 і 6,08 відповідно, що свідчить про найменшу демографічну активність у цих групах регіонів. Таким чином результати кластеризації підтверджують суттєву диференціацію регіонів України за рівнем народжуваності.

Детальна характеристика кожного кластера наведена в табл. 2. 5 та

табл. 2. 6, на основі яких розглянемо їх змістовну інтерпретацію.

Таблиця 2. 5 – Середні значення показників по кластерах (2021 р.)

Показник	Кластер 1	Кластер 2	Кластер 3	Кластер 4	Кластер 5
birth_rate	10.03	7.56	7.40	6.20	6.08
women_share	24.28	23.65	23.85	22.67	22.97
unemployment	11.13	10.85	7.52	11.45	10.00
preschool	4.67	5.16	3.59	4.86	3.15
doctors	33.50	44.50	42.11	33.12	38.55
cpi	109.87	108.90	110.00	110.54	109.20
wage	12147.33	11722.75	12637.00	12108.88	13740.50
ecology	5.30	18.04	46.76	22.22	131.02

Джерело: розроблено автором на основі [3; 9]

Таблиця 2. 6 – Середні значення факторних оцінок по кластерах (2021 р.)

Кластер	F1 «Соціально- екологічний»	F2 «Медична інфраструктура та цінова стабільність»	F3 «Демографічний потенціал та зайнятість»	birth_rate
K1	+1.166	+0.133	+0.562	10.03
K2	+1.060	-1.612	-0.173	7.56
K3	-1.012	-0.281	+1.685	7.40
K4	+0.462	+1.074	-1.135	6.20
K5	-3.185	-0.571	-0.170	6.08

Джерело: розроблено автором на основі [3; 9]

Аналіз середніх значень факторних оцінок по кластерах (табл. 2. 6) дозволяє змістовно інтерпретувати кожен з отриманих груп регіонів.

Кластер 1 (3 регіони: Волинська, Закарпатська, Рівненська області) – «Демографічно активні західні регіони». Найвище значення F1 (+1,166) і високе F3 (+0,562) свідчать про сприятливий екологічний стан і вищу частку жінок репродуктивного віку при помірному безробітті. Зокрема середній рівень викидів становить лише 5,30 кг на особу – найнижчий серед усіх кластерів, а частка жінок репродуктивного віку 24,28% – найвища. Це визначає найвищий коефіцієнт народжуваності серед усіх груп (10,03). Також за фактором F2 (+0.133) даним регіонам притаманна низько розвинена забезпеченість лікарями та середній рівень ІСЦ.

Кластер 2 (4 регіони: Вінницька, Житомирська, Тернопільська, Чернівецька області) – «Регіони з помірним демографічним потенціалом». Поєднує високий F1 (+1,060) з різко від'ємним F2 (-1,612) – найкраща медична забезпеченість серед усіх

кластерів (44,50 лікарів на 10 тис. осіб) і найнижчий ІСЦ (108,90%), хоча заробітна плата в регіонах цього кластеру є найнижчою. Також регіонам притаманна гарна забезпеченість садочками (5,16 на 10 тис. осіб) при добрих екологічних умовах. Народжуваність дещо вища за середню по вибірці (7,56).

Кластер 3 (5 регіонів: Івано-Франківська, Київська, Львівська, Одеська, Харківська області) – «Урбанізовані регіони з розвиненою інфраструктурою». Від'ємний F1 (-1,012) відображає вищі промислові викиди (46,76 кг на особу), друге місце в лідерстві за рівнем заробітної плати (в середньому 12637 грн) і меншу кількість дошкільних закладів на 10 тисяч жителів (3,59), натомість найвище значення F3 (+1,685) серед усіх кластерів свідчить про найнижче безробіття (7,52%) – характеристика великих міських центрів з розвиненим ринком праці. Середня народжуваність – коефіцієнт на рівні 7,40 на 1000 осіб.

Кластер 4 (8 регіонів: Кіровоградська, Миколаївська, Полтавська, Сумська, Херсонська, Хмельницька, Черкаська, Чернігівська) – «Регіони з несприятливими демографічними тенденціями». Найвищий F2 (+1,074) серед усіх кластерів вказує на меншу забезпеченість лікарями і найвищий ІСЦ (110,54%), різко від'ємний F3 (-1,135) – найнижча частка жінок репродуктивного віку (22,67%) і підвищене безробіття (11,45%). Народжуваність на рівні 6,20 – нижче середнього.

Кластер 5 (Дніпропетровська та Запорізька області) – «Промислові регіони з критичним екологічним навантаженням». Різко від'ємний F1 (-3,185) – найбільші промислові викиди серед усіх регіонів (131,02 кг на особу) і найменша кількість дошкільних закладів на 10 тис. осіб. Попри найвищу середню заробітну плату (13741 грн) народжуваність є найнижчою (6,08), що підкреслює визначальну роль екологічного фактора.

Для наочного порівняння профілів кластерів за нормованими значеннями оригінальних показників побудовано радарну діаграму (рис. 2. 8).

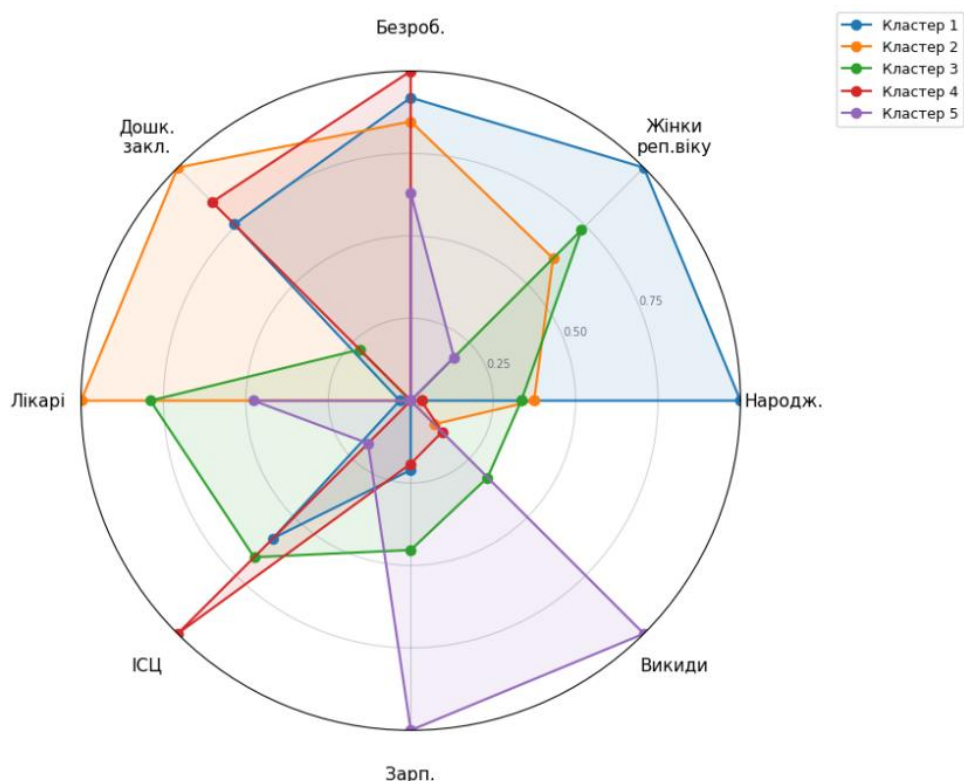


Рисунок 2. 8 - Радарна діаграма профілів кластерів (нормовані значення, 2021 р.)

Джерело: розроблено автором на основі [3; 9]

Радарна діаграма унаочнює відмінності між кластерами – чітко видно, що кластер 1 домінує за народжуваністю і часткою жінок репродуктивного віку, кластер 5 різко виокремлюється за викидами, а кластери 3 і 4 протилежні за безробіттям.

Для оцінки ступеня відмінності між кластерами розраховано матрицю евклідових відстаней між центроїдами кластерів у просторі факторних оцінок (табл. 2. 7).

Таблиця 2. 7 – Матриця евклідових відстаней між центроїдами кластерів (2021 р.)

	K1	K2	K3	K4	K5
K1	0.000	1.896	2.485	2.064	4.467
K2	1.896	0.000	3.085	2.914	4.370
K3	2.485	3.085	0.000	3.458	2.871
K4	2.064	2.914	3.458	0.000	4.115
K5	4.467	4.370	2.871	4.115	0.000

Джерело: розроблено автором на основі [3; 9]

Матриця підтверджує, що кластер 5 (Дніпропетровська, Запорізька області) є найбільш відокремленим – відстані від нього до інших кластерів становлять від

2,871 до 4,467. Найближчими між собою є кластери 1 і 2 (відстань 1,896), що відображає їх схожість за соціально-екологічним фактором F1. Найбільш різними є кластери 3 і 4 (3,458), які розрізняються насамперед за F3 – фактором демографічного потенціалу та зайнятості.

Кластерний аналіз підтвердив суттєву регіональну диференціацію народжуваності в Україні. Ієрархічна кластеризація виявила чотири природні групи регіонів, тоді як метод k-середніх з урахуванням силуетного аналізу деталізував розбиття до п'яти кластерів. Найвища народжуваність зосереджена у західних регіонах з молодшою демографічною структурою і чистим природним середовищем, тоді як промислові регіони з високим екологічним навантаженням і регіони з відтоком молоді демонструють найнижчі показники. Результати кластеризації підтверджують висновки кореляційного та факторного аналізу щодо ключової ролі демографічної структури та соціально-економічних умов у формуванні рівня народжуваності. Типологія регіонів за даними до початку повномасштабного вторгнення виступатиме базою для порівняльного аналізу змін, що відбулися в умовах війни.

2.3. Багатовимірний статистичний аналіз народжуваності регіонів України в умовах повномасштабної війни

Другий блок дослідження охоплює дані за 2025 рік, сформовані по тих самих 24 областях без міста Київ. Відповідно до загальної методології дослідження аналіз проводиться у тій самій послідовності, що і для 2021 року – описова статистика, кореляційний аналіз та кластеризація регіонів, – але без зменшення розмірності даних шляхом виділення головних компонент. Проте є принципова відмінність у природі самих даних: через відсутність актуальних відомостей про чисельність населення по регіонах, а саме бази порівняння, показники у цьому блоці є абсолютними, а не відносними. Це означає, що вони не нормовані на

кількість жителів і відображають загальний обсяг явищ у кожному регіоні, тому при інтерпретації результатів слід враховувати, що регіони з більшим населенням матимуть вищі абсолютні значення незалежно від інтенсивності досліджуваних процесів [1, с. 11]. Зменшення впливу масштабу регіону забезпечується стандартизацією змінних перед кластеризацією.

Завантажимо дані до середовища мови програмування Python та проаналізуємо описову статистику всіх показників, що наведено в табл. 2. 8

Таблиця 2. 8 – Результати описової статистики для показників (2025 р).

	birth	wage	cpi	preschool	wage_debt
count	24	24	24	24	24
mean	6223.67	22337.29	107.97	499.17	53079.54
std	3902.42	2815.01	0.73	238.73	32089.45
min	0	18604	106.80	11	0
25%	3845.25	21064	107.48	386.25	32106
50%	5858	21719	108.00	505.50	52012.50
75%	7857.75	23362	108.32	605	68064
max	15872	31638	110.10	976	124297

Джерело: розроблено автором на основі [2; 3]

Описова статистика показує значну варіацію між регіонами за більшістю показників. Середня кількість народжених по вибірці становить 6224 особи, проте стандартне відхилення 3902 вказує на суттєву нерівномірність між регіонами.

Середня заробітна плата варіює від 18604 до 31638 грн при середньому значенні 22337 грн. Розкид є помірним – міжквартильний діапазон становить від 21064 до 23362 грн.

Індекс споживчих цін є найстабільнішим показником: діапазон від 106,8% до 110,1%, стандартне відхилення лише 0,73. Це найвужча варіація серед усіх змінних.

Кількість закладів дошкільної освіти варіює від 11 (Луганська – аномально низьке значення через окупацію) до 976 (Львівська), середнє значення 499 закладів. Показник тісно пов'язаний з масштабом регіону: більші регіони мають більше закладів.

Заборгованість із заробітної плати на одного працівника має найбільший відносний розкид серед усіх змінних – від 0 у регіонах, де явище не зафіксовано (Чернівецька, Луганська), до 124297 грн у Івано-Франківській області. Медіана

становить 52013 грн, що свідчить про значне навантаження на домогосподарства в більшості регіонів.

Для візуального відображення розподілу значень показників наведено діаграми розмаху (рис. 2. 9).

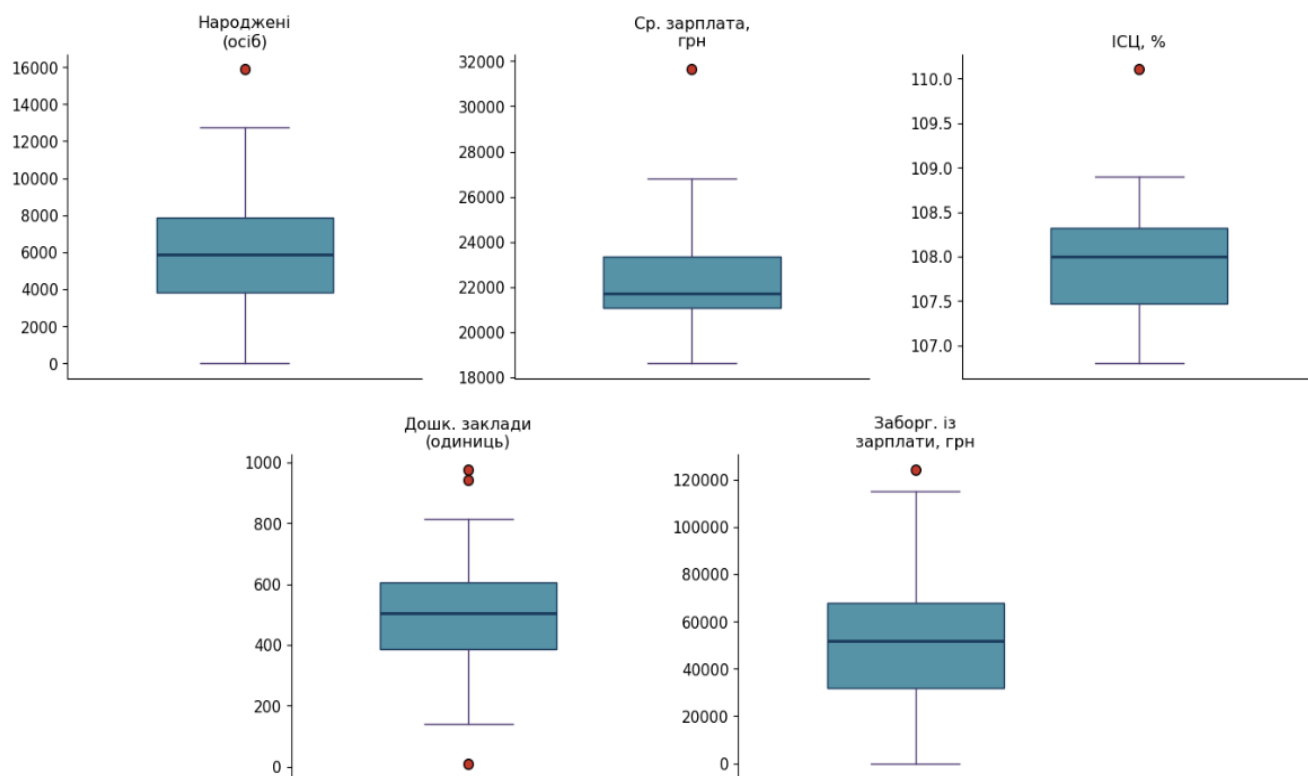


Рисунок 2. 9 – Діаграми розмаху показників по регіонах України (2025 р.)
Джерело: розроблено автором на основі [2; 3]

На діаграмах розмаху помітні характерні викиди по кожному з показників. За кількістю народжених спостерігається один викид вгору – Львівська область (15872 особи), що відображає значний приток ВПО та збільшення фактичного населення регіону.

За показником заробітної плати виявлено один викид вгору – Луганська область (31638 грн), що пояснюється специфічною структурою зайнятості на підконтрольній українській владі частині. За ІСЦ виокремлюється Волинська область (110,1%) з найвищим рівнем інфляції серед усіх регіонів. За кількістю дошкільних закладів наявні три викиди: два вгору – Дніпропетровська (943) і Львівська (976), і один вниз – Луганська область (11) через майже повну окупацію території. За заборгованістю із зарплати виявлено один значний викид вгору – Івано-Франківська область (124297 грн).

Щодо Донецької та Луганської областей, то вони мають аномально занижені значення народжуваності – 602 та 0 осіб відповідно – оскільки дані фіксуються лише з підконтрольних уряду територій, без урахування тимчасово окупованих та частини територій, де ведуться бойові дії [2]. Ці значення відображають наслідки окупації, а не реальні демографічні тенденції регіонів. Аналогічно до блоку 2021 року обидві області виключено з подальших розрахунків – всі наступні етапи проводитимуться на вибірці з 22 областей.

Виявлений розкид значень та наявність викидів свідчать про суттєву неоднорідність регіонів за всіма досліджуваними показниками. Першим кроком є кореляційний аналіз на 22 регіонах після виключення аномальних спостережень. Подивимося на кольорову мапу кореляцій (рис. 2. 10).

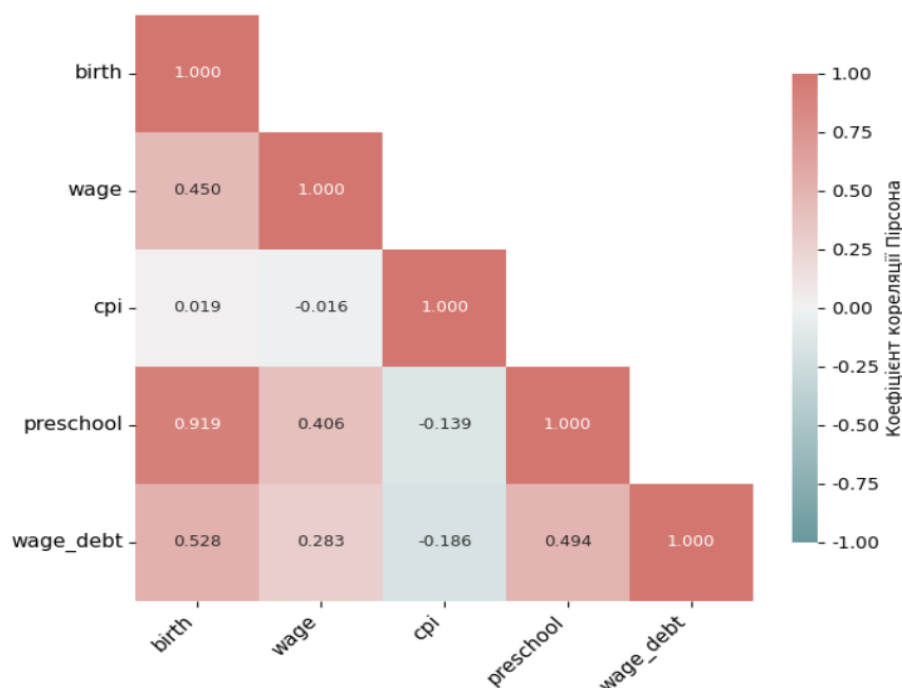


Рисунок 2. 10 – Кольорова мапа кореляцій (22 регіони, 2025 р.)

Джерело: розроблено автором на основі [2; 3]

Для оцінки сили зв'язку використано шкалу Чеддока [27, с. 80]. Дуже високу кореляцію з кількістю народжених має кількість закладів дошкільної освіти ($r = +0,919$). Це пояснюється тим, що обидва показники є абсолютними і тісно пов'язані з розміром регіону – більші регіони мають і більше народжень, і більше дошкільних закладів. Це є об'єктивним методологічним обмеженням даного блоку дослідження зумовленим відсутністю актуальних даних про чисельність населення по регіонах,

щоб звести абсолютні показники до спільної основи. Помірну кореляцію демонструють заборгованість із заробітної плати ($r = +0,528$) та середня заробітна плата ($r = +0,450$). Натомість ІСЦ практично не має лінійного зв'язку з народжуваністю ($r = +0,019$) і характеризується найменшою варіацією між регіонами серед усіх змінних, проте залишається у наборі як характеристика цінової стабільності регіону для подальшого регресійного аналізу в підрозділі 2.4. Стандартизація змінних перед кластеризацією нівелює різницю в масштабах і дозволяє порівнювати регіони за структурою показників.

Для типологізації регіонів застосовано кластерний аналіз на стандартизованих даних 22 областей. Спочатку виконано ієрархічну кластеризацію методом Варда для попередньої оцінки природної структури даних (рис. 2. 11).

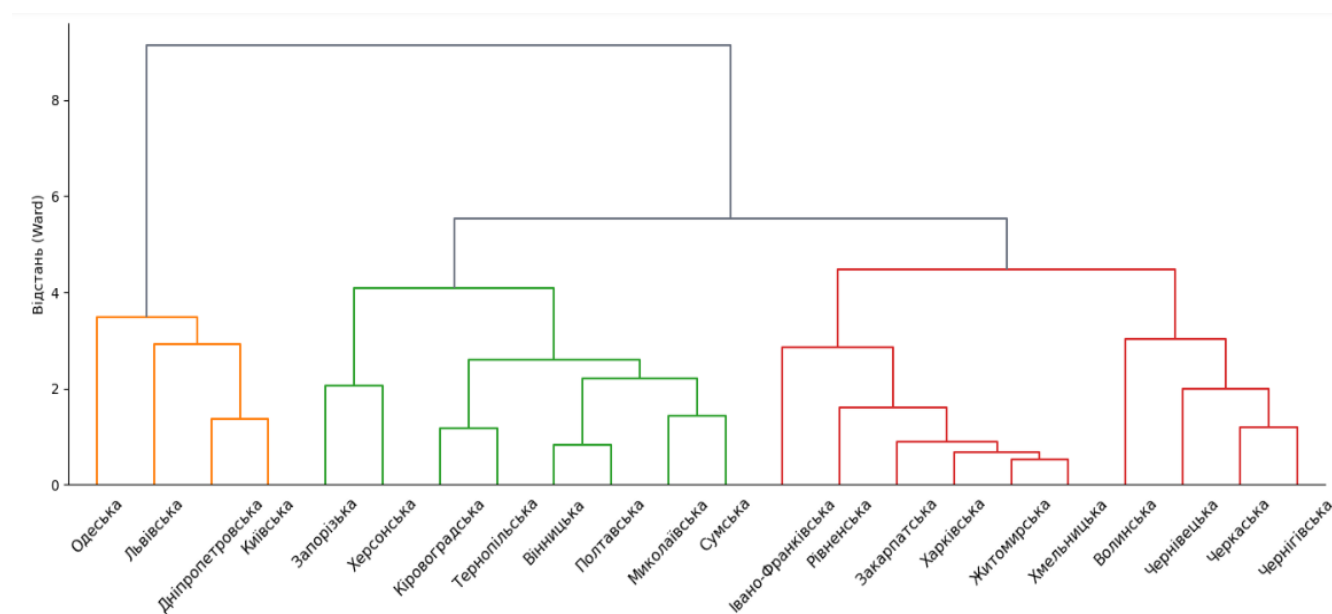


Рисунок 2. 11 – Дендрограма ієрархічної кластеризації (метод Варда, 2025 р.)

Джерело: розроблено автором на основі [2; 3]

Дендрограма виокремлює три групи регіонів. Перша – група з чотирьох великих регіонів (Одеська, Львівська, Дніпропетровська, Київська) – зливається з рештою вибірки лише на найвищому рівні ієрархії (близько 9), що свідчить про їх суттєву відмінність від інших областей. Друга група об'єднує переважно центральні та південні регіони включно з прифронтовими (Запорізька, Херсонська, Миколаївська та ін.), третя – переважно північні та західні регіони. Обидві менші групи зливаються між собою на рівні відстані близько 5, що свідчить про помірну

схожість між ними за сукупністю показників – вони відрізняються від першої групи але не настільки різко між собою.

Для підтвердження оптимальної кількості кластерів застосовано метод ліктя та силуетний аналіз (рис. 2. 12).

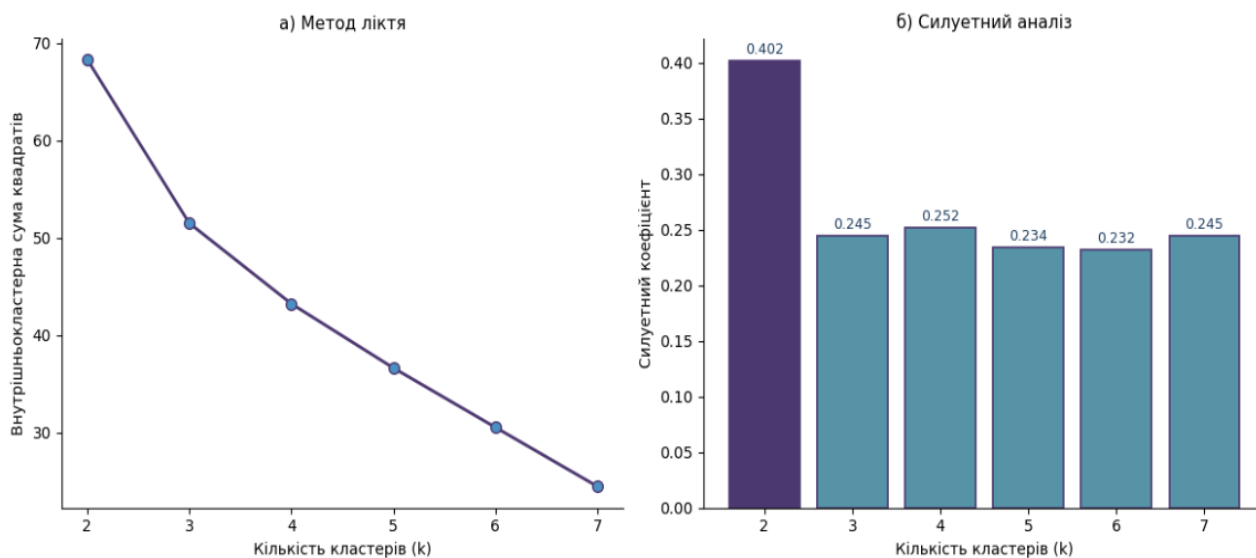


Рисунок 2. 12 – Вибір оптимальної кількості кластерів (2025 р.)

Джерело: розроблено автором на основі [2; 3]

На графіку методу «ліктя» найбільш виражений злам спостерігається при переході від $k=2$ до $k=3$ – внутрішньокластерна сума квадратів зменшується з 68,3 до 51,5, тобто на 16,8 одиниць. Після $k=3$ темп зменшення суттєво сповільнюється: між $k=3$ і $k=4$ спад становить лише 8,3 одиниці, між $k=4$ і $k=5$ – 6,5 одиниць. Це свідчить, що додавання четвертого і наступних кластерів не дає суттєвого приросту якості розбиття.

Силуетний аналіз показує найвищий коефіцієнт при $k=2$ (0,402), що свідчить про найчіткіше розділення на дві групи. При $k=3$ силуетний коефіцієнт знижується до 0,245, при $k=4$ дещо зростає до 0,252, далі поступово зменшується. Попри те що формально $k=2$ є оптимальним за силуетним критерієм, таке розбиття надто загальне – воно об'єднує в одну групу регіони, які на дендрограмі чітко поділяються на дві різні підгрупи та мають різні характеристики. Вибір $k=3$ дозволяє виявити ці змістовні відмінності і відповідає природній структурі даних виявленій ієрархічною кластеризацією. З урахуванням результатів та необхідності змістовної типологізації регіонів обрано $k=3$.

Для отримання фінального розбиття застосовано метод k-means з $k=3$. Результати розподілу 22 регіонів наведено в табл. 2. 9.

Таблиця 2. 9 – Розподіл регіонів України за кластерами (k-середніх, 2025р.)

Кластер 1	Кластер 2	Кластер 3
Дніпропетровська	Волинська	Вінницька
Київська	Житомирська	Запорізька
Львівська	Закарпатська	Кіровоградська
Одеська	Івано-Франківська	Миколаївська
	Рівненська	Полтавська
	Харківська	Сумська
	Хмельницька	Тернопільська
	Черкаська	Херсонська
	Чернігівська	Чернівецька

Джерело: розроблено автором на основі [2; 3]

Розподіл регіонів між кластерами є нерівномірним: перший налічує лише 4 регіони, другий і третій по 9. Для оцінки відмінностей між групами розглянемо середню кількість народжених по кластерах (рис. 2. 13).

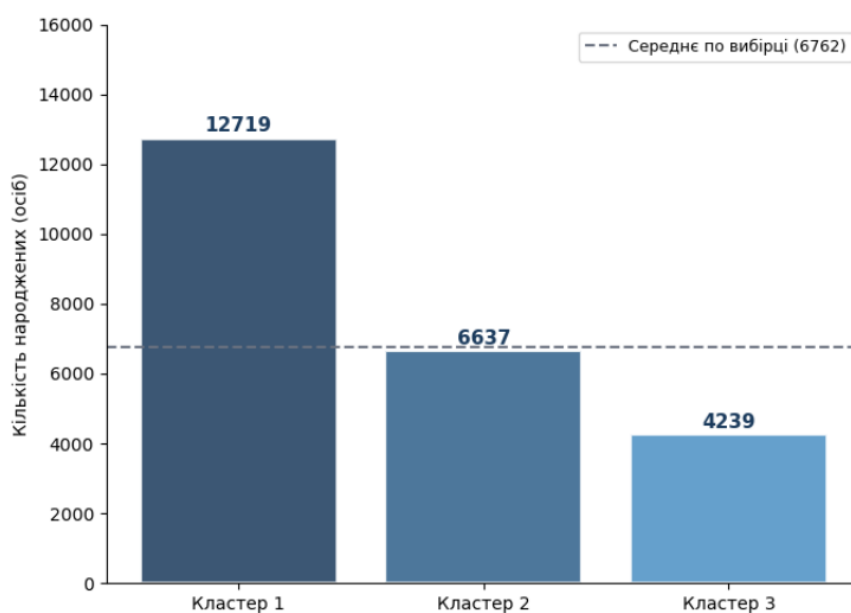


Рисунок 2. 13 – Середня кількість народжених по кластерах (2025 р.)

Джерело: розроблено автором на основі [2; 3]

Кластер 1 суттєво вирізняється: середня кількість народжених становить 12719 осіб, що майже вдвічі перевищує середнє по вибірці (6762). Кластери 2 і 3 демонструють помірне розмежування – 6637 і 4239 осіб відповідно. Це свідчить, що в умовах воєнного часу відбулась значна концентрація народжень у кількох великих регіонах України, що водночас є в лідерах реципієнтів ВПО.

Детальна характеристика кожного кластера на основі середніх значень усіх показників наведена в табл. 2. 10 та рис. 2. 14.

Таблиця 2. 10 – Середні значення показників по кластерах (2025 р.)

Показник	Кластер 1	Кластер 2	Кластер 3
birth	12719	6637.2	4239.4
wage	24314.5	21223.9	21400.3
cpi	107.6	108.7	107.5
preschool	870.8	522.8	401.2
wage_debt	89340	60130.3	36557.1

Джерело: розроблено автором на основі [2; 3]

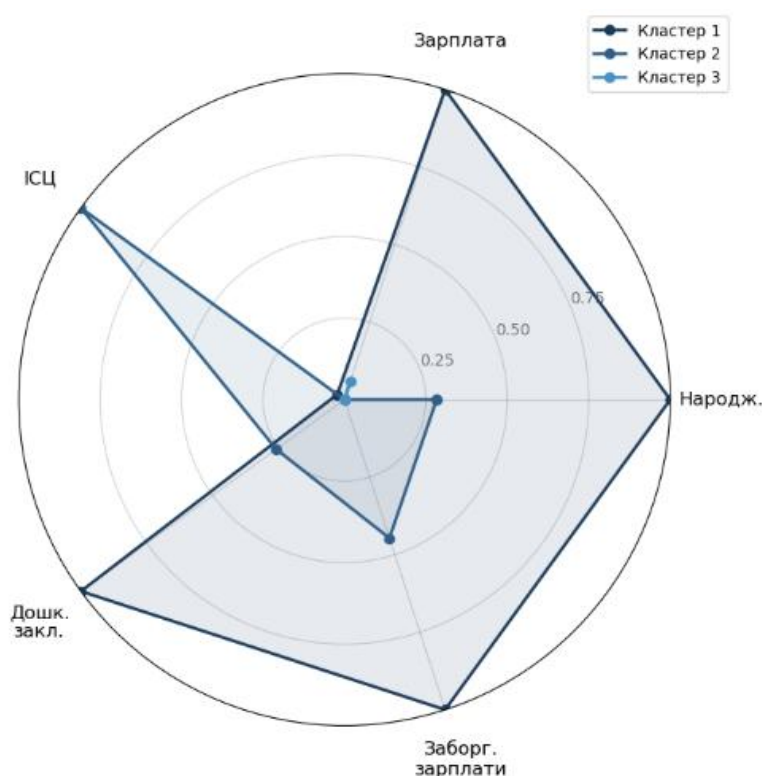


Рисунок 2. 14 – Радарна діаграма профілів кластерів (нормовані значення, 2025 р.)

Джерело: розроблено автором на основі [2; 3]

Кластер 1 (4 регіони: Дніпропетровська, Київська, Львівська, Одеська області) – «Великі регіони-реципієнти з концентрацією народжень». Найвища кількість народжених (12719 осіб у середньому) поєднується з найвищою заробітною платою (24315 грн) і найбільшою кількістю дошкільних закладів (871). Всі чотири регіони є великими урбанізованими центрами, які прийняли значну кількість ВПО, що суттєво збільшило їх фактичне населення. Водночас заборгованість із заробітної плати тут найвища серед кластерів (89340 грн), що відображає концентрацію великих підприємств.

Кластер 2 (9 регіонів: Волинська, Житомирська, Закарпатська, Івано-Франківська, Рівненська, Харківська, Хмельницька, Черкаська, Чернігівська області) – «Регіони з помірною народжуваністю». Середній рівень народжених (6637 осіб), що близьке до загального середнього значення по вибірці. Заробітна плата порівняно з кластером 1 та 3 є найнижчою, а ІСЦ найвищим (108,7%). Це різнорідна група, яка поєднує західні, центральні та прифронтний (Харківська область) регіони.

Кластер 3 (9 регіонів: Вінницька, Запорізька, Кіровоградська, Миколаївська, Полтавська, Сумська, Тернопільська, Херсонська, Чернівецька) – «Регіони з низькою народжуваністю». Найнижча кількість народжених (4239 осіб) поєднується з найнижчою заборгованістю із зарплати (36557 грн) і найменшою кількістю дошкільних закладів (401). Середня заробітна плата (21400 грн) є подібною до кластеру 2, натомість ІСЦ (107,5%) найнижчий серед усіх кластерів, що свідчить про дещо нижчу інфляцію. Група є неоднорідною за своїм складом – включає як аграрні регіони центру (Вінницька, Полтавська, Кіровоградська), так і прифронтні території (Запорізька, Херсонська, Сумська, Миколаївська).

Для оцінки ступеня відмінності між кластерами розраховано матрицю евклідових відстаней між центроїдами кластерів у просторі стандартизованих показників (табл. 2. 11).

Таблиця 2. 11 – Матриця евклідових відстаней між центроїдами кластерів (2025 р.)

	K1	K2	K3
K1	0.000	3.360	4.005
K2	3.360	0.000	1.933
K3	4.005	1.933	0.000

Джерело: розроблено автором на основі [2; 3]

Матриця відстаней підтверджує, що кластер 1 (Дніпропетровська, Київська, Львівська, Одеська області) є найбільш відокремленим від решти – відстань від нього до кластера 2 становить 3,360 і до кластера 3 – 4,005. Це відображає принципову відмінність великих регіонів від інших областей за сукупністю абсолютних показників. Кластери 2 і 3 є відносно ближчими між собою (1,933), що

свідчить про помірніші відмінності між регіонами з середньою і низькою народжуваністю.

Кластерний аналіз даних за 2025 рік виявив суттєву трансформацію регіональної картини народжуваності порівняно з періодом до повномасштабного вторгнення. Якщо раніше виокремлювались демографічно активні західні регіони з традиційно вищою народжуваністю, то тепер структура визначається переважно масштабом регіону та притоком ВПО – регіони з найбільшою кількістю народжень – це великі урбанізовані центри, що стали головними реципієнтами переміщеного населення, тоді як прифронтові регіони демонструють найнижчі показники народжуваності. Слід зазначити, що використання абсолютних показників через відсутність актуальних даних про чисельність населення є методологічним обмеженням цього блоку аналізу та прямо вплинуло на отримані результати. Кількісна оцінка впливу соціально-економічних чинників на народжуваність буде здійснена у наступному розділі за допомогою регресійного аналізу.

2.4. Регресійне моделювання впливу соціально-економічних чинників на народжуваність в умовах повномасштабної війни

Метою даного підрозділу є кількісна оцінка статистичних зв'язків між соціально-економічними чинниками і кількістю народжених по регіонах України в умовах повномасштабної війни, а також визначення того, які з цих факторів є статистично значущими.

Для цього було застосовано метод множинної лінійної регресії. На відміну від парного кореляційного аналізу, регресійна модель дає можливість оцінити чистий вплив кожного фактора на залежну змінну при контролі всіх інших – тобто зрозуміти, на скільки зміниться кількість народжень при зміні певного показника за умови, що решта факторів залишаються незмінними [17]. Крім того, метод

дозволяє включати фіктивні змінні, що важливо для врахування воєнного контексту через показники *frontline* та *recipient*.

Серед обмежень методу слід зазначити припущення про лінійність зв'язків між змінними – тобто що вплив кожної змінної є постійним незалежно від її рівня, а також порівняно невелику вибірку (22 регіони), що знижує статистичну потужність аналізу та ускладнює виявлення слабких ефектів. Крім того через відсутність актуальних даних про чисельність населення залежна змінна є абсолютною, що частково обмежує порівнянність між регіонами різного розміру.

Залежна змінна – кількість народжених по регіону (*birth*, осіб). Незалежні змінні обрано відповідно до сформованої інформаційної бази (підрозділ 2.1) та результатів кореляційного аналізу. Перед побудовою моделей сформульовано гіпотези щодо очікуваного напрямку впливу кожної змінної:

- *wage* (середня заробітна плата, грн): очікується позитивний вплив – вищий рівень доходів створює кращі економічні умови для народження дітей;

- *cpi* (індекс споживчих цін, %): очікується від'ємний вплив – вища інфляція знижує реальні доходи і відкладає рішення про народження дітей;

- *preschool* (кількість закладів дошкільної освіти, од.): очікується позитивний вплив – більша кількість освітніх закладів відображає розвиненішу соціальну інфраструктуру для сімей з дітьми, що може сприяти вищій народжуваності;

- *wage_debt* (заборгованість із зарплати на одного працівника, грн): очікується від'ємний вплив – вища заборгованість свідчить про економічну нестабільність і знижує впевненість домогосподарств у майбутньому;

- *frontline* (фіктивна змінна, 1 – прифронтовий регіон): очікується від'ємний вплив – активні бойові дії і окупація частини території критично знижують народжуваність;

- *recipient* (фіктивна змінна, 1 – регіон-реципієнт ВПО): вплив неоднозначний – приток ВПО збільшує кількість жінок репродуктивного віку, але водночас супроводжується стресом і нестабільністю.

Побудовано дві моделі. Модель 1 включає всі 6 змінних, а Модель 2 виключає змінну preschool.

Рівняння множинної лінійної регресії Моделі 1 наведено за формулою:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 + u, \quad (2.1)$$

де y - кількість народжених по регіону (осіб);

a_0 - вільний член;

a_1 - a_6 - коефіцієнти регресії;

x_1 - середньомісячна заробітна плата (грн);

x_2 - індекс споживчих цін (%);

x_3 - кількість закладів дошкільної освіти (одиниць);

x_4 - заборгованість із виплати заробітної плати на одного працівника (грн);

x_5 - фіктивна змінна для прифронтових регіонів;

x_6 - фіктивна змінна для регіонів-реципієнтів ВПО;

u - випадкова похибка

Рівняння множинної лінійної регресії Моделі 2 (основна) наведено за формулою:

$$y = a_0 + a_1x_1 + a_2x_2 + a_4x_4 + a_5x_5 + a_6x_6 + u \quad (2.2)$$

Всі позначення у формулі (2.2) відповідають формулі (2.1) за винятком змінної x_3 , яка у даній моделі відсутня.

Окремо варто пояснити рішення щодо змінної x_3 . Кореляційний аналіз виявив дуже тісний зв'язок між preschool і birth ($r = +0,919$) – обидва показники є абсолютними і відображають загальний масштаб регіону, тому включення x_3 до моделі призводить, так би мовити, до "поглинання" більшої частини пояснювальної сили і робить інші змінні статистично незначущими. Це є помилкою специфікації [17, с. 151]. Модель 1 будується для демонстрації цього ефекту, а основною для інтерпретації є Модель 2.

Зведені показники якості обох моделей наведено в табл. 2. 12.

Таблиця 2. 12 – Показники якості регресійних моделей

Показник	Модель 1	Модель 2
R^2	0.891	0.597
Adj. R^2	0.848	0.471
F-статистика	20.47	4.74
p-значення (F)	<0.001	0.008

Джерело: розроблено автором на основі [2; 3; 15; 26]

Результати оцінювання коефіцієнтів обох моделей наведено в табл. 2. 13.

Таблиця 2. 13 – Оцінки коефіцієнтів регресійних моделей (залежна змінна: birth)

Змінна	a (M1)	se (M1)	t (M1)	p (M1)	VIF ₁	a (M2)	se (M2)	t (M2)	p (M2)	VIF ₂
const	-65674	52959	-1.240	0.234	-	-30848	98142	-0.314	0.757	-
wage	0.28	0.22	1.308	0.211	2.110	0.93	0.36	2.594	0.020 **	1.652
cpi	543.06	498.92	1.089	0.294	1.494	146.09	922.27	0.158	0.876	1.471
preschool	13.52	2.12	6.367	0.000 ***	2.121	-	-	-	-	-
wage debt	0.01	0.01	1.115	0.283	1.417	0.04	0.02	2.095	0.052 *	1.202
frontline	-1149	1060	-1.084	0.296	2.469	-3623	1837	-1.972	0.066 *	2.137
recipient	-246	875	-0.282	0.782	1.840	698	1606	0.435	0.670	1.787

Примітка: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Джерело: розроблено автором на основі [2; 3; 15; 26]

Перед інтерпретацією результатів перевірено виконання передумов МНК для основної Моделі 2. Тест Шапіро-Вілکا підтвердив нормальний розподіл залишків ($W=0,921$, $p=0,079 > 0,05$). Тест Бройша-Пагана не виявив гетероскедастичності залишків ($LM=1,952$, $p=0,162 > 0,05$) – дисперсія залишків є однорідною по всьому діапазону значень. Статистика Дарбіна-Уотсона становить $DW=2,272$. При рівні значущості 0,05 і параметрах $n=22$, $k=5$ критичні значення становлять $DW_1=1,013$ і $DW_2=1,776$ ($4-DW_2=2,224$, $4-DW_1=2,987$). Оскільки $DW=2,272$ знаходиться в зоні невизначеності ($2,224 < 2,272 < 2,987$), однозначний висновок про відсутність автокореляції на рівні 0,05 зробити не можна [17, с. 153]. Однак при більш суворому рівні значущості 0,01 ($DW_2=1,583$, $4-DW_2=2,417$) статистика потрапляє в зону відсутності автокореляції.

Значення VIF для всіх змінних не перевищують 2,2 – значно нижче критичного порогу 5, що підтверджує відсутність мультиколінеарності між незалежними змінними [30, с 350]. Таким чином усі основні передумови МНК виконуються і результати оцінювання є статистично коректними.

Порівняння двох моделей наочно демонструє ефект включення *preschool*. Модель 1 пояснює 89,1% варіації народжуваності між регіонами ($R^2=0,891$) проте єдиною статистично значущою змінною є *preschool* ($a=13,52$, $p<0,001$). Всі інші змінні включно з *frontline* і *recipient* – статистично незначущі. Це підтверджує, що *preschool* як абсолютний показник розміру регіону витісняє решту чинників і унеможливорює змістовну оцінку їх зв'язку з народжуваністю. З цієї причини Модель 1 не може використовуватись для досягнення мети дослідження незважаючи на формально вищий R^2 .

Модель 2 пояснює 59,7% варіації ($R^2 = 0,597$, $adj.R^2 = 0,471$) і є загалом статистично значущою ($F=4,74$, $p=0,008$). Нижчий R^2 порівняно з Моделлю 1 є очікуваним і свідчить лише про те, що масштаб регіону дійсно пояснює значну частину варіації абсолютної кількості народжених.

Середня заробітна плата (*wage*) є статистично значущою ($a=0,927$, $p=0,020$). При збільшенні середньої заробітної плати на 1 гривню кількість народжених в регіоні пов'язана у середньому з додатковими новонародженими 0,93 особами за незмінності інших показників. Це узгоджується з висунутою гіпотезою – вищий рівень доходів статистично пов'язаний з вищою народжуваністю навіть в умовах воєнного часу.

Заборгованість із заробітної плати (*wage_debt*) демонструє майже значущий позитивний зв'язок ($a=0,042$, $p=0,052$). Знак коефіцієнта є протилежним до очікуваного. Це може пояснюватись тим, що висока заборгованість концентрується у великих промислових регіонах з більшим населенням і відповідно більшою абсолютною кількістю народжень, що є наслідком використання абсолютних показників.

Фіктивна змінна *frontline* є майже значущою ($a=-3623$, $p=0,066$). Прифронтові регіони пов'язані в середньому з на 3623 народженими менше

порівняно з іншими регіонами за однакових значень економічних показників. Це підтверджує гіпотезу про від'ємний зв'язок між активними бойовими діями і народжуваністю і є одним з ключових результатів дослідження.

Змінна *recipient* не є статистично значущою ($a=698$, $p=0,670$). Статус регіону-реципієнта ВПО не пов'язаний значущо з кількістю народжених при контролі економічних показників. Це може свідчити про те, що вплив притоку ВПО на народжуваність є неоднозначним і нейтралізується збільшенням кількості жінок репродуктивного віку з одного боку і стресом та нестабільністю з іншого.

Індекс споживчих цін (*cpi*) не продемонстрував статистично значущого зв'язку з народжуваністю ($p=0,876$), що пояснюється низькою варіацією цього показника між регіонами.

Результати регресійного аналізу узгоджуються з висновками кластерного аналізу підрозділу 2.3. Кластер 1, що об'єднує великі регіони-реципієнти ВПО з найвищою народжуваністю характеризується найвищою середньою заробітною платою (24315 грн) – регресія підтверджує, що *wage* статистично значущо пов'язана з народжуваністю. Кластер 3 з найнижчою народжуваністю включає переважно прифронтові регіони і коефіцієнт при *frontline* кількісно підтверджує цей від'ємний зв'язок. Отже, регресійний аналіз підтверджує кількісно ті закономірності, які були виявлені методами описового та кластерного аналізу.

Основним обмеженням побудованої моделі є використання абсолютних показників через відсутність актуальних даних про чисельність населення по регіонах після 2021 року. Крос-секційна природа даних одного року дозволяє виявляти статистичні асоціації між показниками, але не причинно-наслідкові зв'язки – для їх встановлення потрібні панельні дані за кілька років. Невелика вибірка (22 регіони) знижує статистичну потужність тестів – зокрема змінні з p на рівні 0,05-0,10 могли б стати значущими при більшій вибірці. Перспективним напрямом подальшої роботи є усунення цих обмежень – насамперед шляхом використання панельних даних і відносних показників.

ВИСНОВКИ

У роботі проведено комплексний статистичний аналіз регіональних особливостей народжуваності в Україні для двох часових зрізів – 2021 і 2025 років – із застосуванням методів багатовимірної статистики та регресійного моделювання. За результатами дослідження зроблено певні висновки.

Народжуваність в Україні відчуває вплив як довгострокових структурних проблем, так і гострих наслідків війни. Кількість новонароджених значно зменшилася – з 273,8 тис. у 2021 році до 168,8 тис. у 2025 році. Навіть за оптимістичним сценарієм Стратегії демографічного розвитку до 2040 року сумарний коефіцієнт народжуваності не зможе досягти рівня простого відтворення населення (2,13). Це свідчить про глибокий і системний характер демографічної кризи.

Інформаційна база дослідження включає дані по 24 регіонах України за два періоди. Для 2021 року були використані відносні показники (народжуваність, частка жінок репродуктивного віку, безробіття, забезпеченість лікарями та садочками, індекс споживчих цін, заробітна плата та екологічне навантаження). Для 2025 року через відсутність актуальних даних про чисельність населення довелося використовувати абсолютні показники. Це є методологічним обмеженням даного дослідження.

Багатовимірний статистичний аналіз за 2021 р. виявив три латентні фактори, що пояснюють 73,4% загальної дисперсії: соціально-екологічний (F1), фактор медичної інфраструктури та цінової стабільності (F2) і демографічно-трудова (F3). На основі факторних оцінок і коефіцієнта народжуваності регіони розподілено на п'ять типологічних груп. Найвищу народжуваність демонструє кластер демографічно активних західних регіонів (Волинська, Закарпатська, Рівненська) з коефіцієнтом 10,03 на 1000 осіб, найнижчу – кластер промислових регіонів з

критичним екологічним навантаженням (Дніпропетровська, Запорізька області) з показником народжуваності 6,08.

Аналіз даних 2025 року показав суттєву трансформацію регіональної картини народжуваності. Кластеризація на стандартизованих абсолютних показниках виділила три групи регіонів. Кластер великих регіонів-реципієнтів ВПО (Дніпропетровська, Київська, Львівська, Одеська) має середню кількість народжених 12719 осіб – удвічі більше за середнє по вибірці. Прифронтові регіони (кластер 3) демонструють найнижчі показники (4239 осіб). Традиційна закономірність вищої народжуваності на заході поступилась новій структурі яка визначається переважно масштабом та наближеністю регіона до лінії фронту.

Регресійний аналіз підтвердив статистично значущий позитивний зв'язок між середньою заробітною платою і кількістю народжених ($\beta=0,927$, $p=0,020$) – регіони з вищими доходами мають більше народжень навіть в умовах воєнного часу. Фіктивна змінна прифронтового статусу регіону є майже значущою ($\beta=-3623$, $p=0,066$) і фіксує від'ємний зв'язок з народжуваністю: прифронтові регіони мають в середньому на 3623 народжених менше порівняно з іншими при однакових економічних умовах. Статус регіону-реципієнта ВПО не виявив значущого самостійного зв'язку з народжуваністю ($p=0,670$). Модель пояснює 59,7% варіації ($R^2=0,597$) і є статистично значущою ($F=4,74$, $p=0,008$).

Усі три методи аналізу – кластеризація 2021, кластеризація 2025 і регресійна модель – узгоджено вказують на визначальну роль воєнного контексту у трансформації демографічних закономірностей. Кількісне підтвердження негативного впливу прифронтового статусу регіону є ключовим результатом дослідження.

На основі отриманих результатів сформульовано такі практичні рекомендації.

По-перше, при розробці регіональної демографічної політики доцільно диференціювати заходи підтримки народжуваності залежно від типологічної групи регіону – стимулюючі заходи для прифронтових і постраждалих регіонів мають

бути більш інтенсивними і передбачати компенсацію не лише фінансових а й безпекових ризиків для молодих сімей.

По-друге, збереження і розвиток дошкільної інфраструктури у регіонах з великим притоком ВПО є важливим напрямом оскільки доступність цих послуг опосередковано підтримує народжуваність через зниження навантаження на сім'ї з дітьми.

По-третє, відновлення повноцінної статистичної звітності по регіонах – включно з даними про чисельність населення – є необхідною умовою для проведення коректного демографічного аналізу в умовах та після завершення воєнного стану.

Перспективами подальших досліджень є застосування панельних даних за кілька років для виявлення причинно-наслідкових зв'язків, а також аналіз демографічної відбудови регіонів у післявоєнний період.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Гладун О. М., Гончар І. А. Методологічні положення зі статистичного аналізу природного руху населення. Київ : Державна служба статистики України, 2013. 71 с.
2. Статистична інформація щодо державної реєстрації актів цивільного стану. Сайт Міністерства юстиції України. URL: https://minjust.gov.ua/actual-info/stat_info (дата звернення: 15.03.2026)
3. Населення та міграція. Сайт Державної служби статистики України. URL: <https://stat.gov.ua/index.php/uk/topics/naselennya-ta-mihratsiya> (дата звернення: 15.03.2026)
4. World Population Prospects 2024. United Nations : official website. URL: <https://population.un.org/wpp/> (дата звернення: 21.03.2026)
5. Лібанова Е. М. Про бебібум давайте не мріяти: інтерв'ю. Українська правда. 29.05.2023. URL: <https://www.pravda.com.ua/articles/2023/05/29/7404353/> (дата звернення: 21.03.2026)
6. Якість життя населення України та перші наслідки війни : монографія / Черенько Л. М., Полякова С. В., Шишкін В. С. та ін. Київ : НАН України, Ін-т демографії та соціальних досліджень ім. М. В. Птухи, 2023. 191 с.
7. Населення України. Демографічні тенденції в Україні у 2002-2019 рр.: кол. моногр. / О.М. Гладун, С.Ю. Аксьонова, О.К. Гаврилюк / за ред. О.М. Гладуна. Київ : НАН України, Ін-т демографії та соціальних досліджень імені М.В. Птухи, 2020. 174 с.
8. Ситуаційний аналіз становища дітей в Україні 2024. Сайт ЮНІСЕФ в Україні. URL: <https://www.unicef.org/ukraine/> (дата звернення: 21.03.2026)
9. Статистичний збірник «Регіони України» 2021. Частина 1 / ред. Вернер І. Є. Київ : Державна служба статистики України, 2019. 276 с.
10. Шулікін Д. Наш людський капітал: втрати й перспективи відродження. Газета «Світ» – Науково-популярне періодичне видання. URL:

<https://svit.kpi.ua/2025/03/08/наш-людський-капітал-втрати-й-перспек/> (дата звернення: 30.03.2026)

11. Про схвалення Стратегії демографічного розвитку України на період до 2040 року : Розпорядж. Каб. Міністрів України від 30.09.2024 № 922-р. URL: <https://zakon.rada.gov.ua/go/922-2024-p> (дата звернення: 05.04.2026)

12. Демографічна криза в Україні: оцінки Інституту демографії та шляхи розв’язання проблем. UNN. 2024. URL: <https://unn.ua/en/news/demographic-crisis-in-ukraine-estimates-from-the-institute-of-demography-and-solutions-to-the-problems/> (дата звернення: 10.05.2026)

13. Напорчук Д. Демографічна ситуація в Україні в умовах війни. Моделювання розвитку економічних систем. 2024. №1. С. 235–241. URL: <https://mdes.khmnu.edu.ua/index.php/mdes/article/view/314> (дата звернення: 10.05.2026)

14. Самодай В., Літвінова Л., Бакута Р. Вплив війни на демографічну ситуацію у вимірах статистичних показників. *Економіка та суспільство*, 2026. Вип. 83. С. 181-185.

15. Звіт про внутрішнє переміщення населення в Україні. Опитування загального населення. Раунд 22. Січень 2026 року. Міжнародна організація з міграції (МОМ), 2026. URL: <https://dtm.iom.int/ukraine>

16. Міністерство захисту довкілля та природних ресурсів України. Екологічна загроза: офіційний портал. URL: <https://ecozagroza.gov.ua> (дата звернення: 10.05.2026)

17. Кількісні методи в економіці : навч. посіб. / Великоіваненко Г.І., Піскунова О.В., Ващаєв С.С. та ін. Київ : КНЕУ, 2024. 392 с.

18. Бізнес-аналітика багатовимірних процесів: навч. посіб. / Борисенко В. І., Горбенко Л. Л., Матющенко Т. І. та ін. / за ред. В. І. Борисенка. Харків: ХНЕУ ім. С. Кузнеця, 2020. 30 с.

19. Kaiser H. F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*. University of Illinois, 1960. Vol. 20(1). P. 141-151.

20. Cattell R. B. The scree test for the number of factors. *Multivariate Behavioral Research*. University of Illinois, 1966. Vol. 1(2). P. 245-276.
21. Бегун С. Застосування кластерного аналізу для дослідження демографічної ситуації в регіоні. *Економічний часопис Східноєвропейського національного університету імені Лесі Українки*, 2016. Вип. 2. С. 122-128.
22. 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.
23. Wooldridge J. M. *Introductory Econometrics: A Modern Approach*. 7th ed. Boston : Cengage Learning, 2020. 789 p.
24. Економетрія (економетрика): навч. посіб. / Єрмоєнко В. О., Алілуйко А. М., Мартинюк О. М., Попіна С. Ю. Тернопіль: Підручники і посібники, 2012. 116 с.
25. Волошин О.Р., Галайко Н.В. Економетрія. Ч. 1: навч. посіб. Львів: Львівський державний університет внутрішніх справ, 2012. 192 с.
26. Про затвердження Переліку територій, на яких ведуться (велися) бойові дії або тимчасово окупованих Російською Федерацією: наказ Міністерства розвитку громад та територій України від 28.02.2025 № 376. Зареєстровано в Міністерстві юстиції України 11.03.2025 № 380/43786. URL: <https://zakon.rada.gov.ua/laws/show/z0380-25>
27. Коломієць С. В. Кількісні методи в економіці : конспект лекцій : у 2 ч. Суми : Сумський державний університет, 2023. Ч. 2. 155 с.
28. Нечітайло І. С. Аналіз психологічних даних в STATISTICA, SPSS: факторний аналіз : текст лекції з навч. дисц. «Комп'ютерні методи практичної психології». Харків : Харківський національний університет внутрішніх справ, 2023. 10 с.
29. Голоскоков О.Є. Основи регресійного аналізу: навч. посіб. Харків: НТУ «ХПІ», 2008. 90 с.
30. Gujarati D. N., Porter D. C. *Basic Econometrics*. 5th ed. New York : McGraw-Hill, 2009. 922 p.

ДОДАТКИ

Додаток А

Таблиця А. 1 – Загальні дані для дослідження народжуваності за 2021 р.

region	birth rate	women share	unemployment	preschool	doctors	cpi	wage	ecology
Вінницька	7,01	22,95	11	5,03	44,85	108,7	12488	52,171
Волинська	9,67	24,27	12,7	4,76	33,66	109,9	11445	5,456
Дніпропетровська	6,31	23,16	8,9	3,18	35,54	109,6	13699	172,36
Донецька	2,50	21,91	15,4	1,36	13,20	111,2	14853	182,39
Житомирська	7,24	23,06	11,2	5,44	33,16	109	11625	10,017
Закарпатська	10,17	24,54	11,1	4,61	31,10	110,1	12235	2,266
Запорізька	5,84	22,77	11,1	3,12	41,57	108,8	13782	89,68
Івано-Франківська	7,82	24,21	8,7	3,39	51,95	110,2	11988	127,114
Київська	7,06	24,20	7,2	4,61	31,00	110,4	14290	33,101
Кіровоградська	6,17	22,34	13,2	5,10	29,53	110,5	11180	12,163
Луганська	1,60	21,47	16	1,24	8,80	110,9	12176	16,611
Львівська	7,90	23,74	7,7	3,66	47,81	108,9	12530	30,319
Миколаївська	6,44	23,13	11,3	4,96	26,67	110,3	13807	11,077
Одеська	8,24	23,47	7,3	3,41	38,98	110,4	12326	15,216
Полтавська	6,00	22,78	12,4	4,61	43,10	110,4	12988	38,501
Рівненська	10,25	24,03	9,6	4,64	35,72	109,6	12762	8,19
Сумська	5,31	22,72	10,1	4,11	35,53	111	11844	17,495
Тернопільська	7,14	24,00	11,9	5,79	47,57	109,5	11595	8,087
Харківська	5,99	23,61	6,7	2,88	40,80	110,1	12051	28,043
Херсонська	7,11	23,02	11,8	4,68	26,66	110	11398	16,919
Хмельницька	7,08	22,69	10,3	5,79	37,32	109,8	12326	17,104
Черкаська	5,90	22,57	10,1	5,39	31,83	111,7	11965	40,745
Чернівецька	8,83	24,59	9,3	4,37	52,43	108,4	11183	1,866
Чернігівська	5,61	22,08	12,4	4,25	34,29	110,6	11363	23,732

Джерело: розроблено автором на основі [3; 9]

Додаток Б

Таблиця Б. 1 – Загальні дані для дослідження народжуваності за 2025 р.

region	birth	wage	cpi	preschool	wage debt	frontline	recipient
Вінницька	7765	22054	107.4	623	39730	0	1
Волинська	7466	21504	110.1	474	29163	0	0
Дніпропетровська	12754	26803	107.5	943	93004	1	0
Донецька	602	23581	107.5	170	46362	1	0
Житомирська	6156	21066	108.4	554	60994	0	0
Закарпатська	8136	22171	108.3	591	53288	0	0
Запорізька	2823	24599	108.3	166	32121	1	0
Івано-Франківська	7501	20271	108.3	453	124297	0	0
Київська	10294	25897	107.9	750	98532	0	1
Кіровоградська	3985	18737	107.3	422	54938	0	1
Луганська	0	31638	107.5	11	0	1	0
Львівська	15872	23500	108.1	976	50737	0	1
Миколаївська	4365	23316	107.7	434	53830	1	0
Одеська	11956	21058	106.8	814	115087	0	1
Полтавська	5560	22469	107.5	552	29754	0	1
Рівненська	9078	22421	108.9	495	69849	0	0
Сумська	3199	21181	107.3	348	72019	1	0
Тернопільська	4899	19796	106.9	527	34189	0	0
Харківська	7019	21591	108.4	672	64703	1	0
Херсонська	366	21847	107.4	141	12433	1	0
Хмельницька	6211	21373	108.1	599	67469	0	0
Черкаська	4742	21074	108.8	516	32061	0	1
Чернівецька	5193	18604	108.1	398	0	0	0
Чернігівська	3426	19544	108.7	351	39349	0	0

Джерело: розроблено автором на основі [2; 3; 15; 26]

Метадані

ДОКУМЕНТ

Заголовок

КоваленкоП_Осипова_плагіат

Автор

Коваленко П

Науковий керівник / Експерт

Осипова

ІД документа

334098355

ОРГАНІЗАЦІЯ

Назва організації

Kyiv National Economic University named after Vadym Hetman
KNEU

підрозділ

кафедра штучного інтелекту, моделювання та статистики

ЗВІТ

Дата звіту

5/30/2026

Дата редагування

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



КП 1

25

Довжина фрази для коефіцієнта подібності 2



КП 2

11924

Кількість слів



КЦ

89396

Кількість символів

Тривога

У цьому розділі ви знайдете інформацію щодо текстових спотворень. Ці спотворення в тексті можуть говорити про МОЖЛИВІ маніпуляції в тексті. Спотворення в тексті можуть мати навмисний характер, але частіше характер технічних помилок при конвертації документа та його збереженні, тому ми рекомендуємо вам підходити до аналізу цього модуля відповідально. У разі виникнення запитань, просимо звертатися до нашої служби підтримки.

Заміна букв		4
Інтервали		0
Мікропробіли		3
Білі знаки		0
Парафрази (SmartMarks)		41

Джерела

Нижче наведений список джерел. В цьому списку є джерела із різних баз даних. Копію тексту означає в якому джерелі він був знайдений. Ці джерела і значення Коефіцієнту Подібності не відображають прямого плагіату. Необхідно відкрити кожне джерело і проаналізувати зміст і правильність оформлення джерела.

10 найдовших фраз

Копію тексту

#	НАЗВА ТА АДРЕСА ДЖЕРЕЛА URL (НАЗВА БАЗИ)	КІЛЬКІСТЬ ІДЕНТИЧНИХ СЛІВ (ФРАГМЕНТІВ)
1	https://brojects.dune-hd.com/bitstream/handle/2010/27587/matem_14.pdf?	54 (0.45 %)