

Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту

Кафедра прикладної математики

Рівень вищої освіти другий (магістерський)

Спеціальність 113 Прикладна математика

(код і повна назва)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Прикладна математика

(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри ПМ _____

(підпис)

“ _____ ” _____ 2019 р.

ЗАВДАННЯ
НА АТЕСТАЦІЙНУ РОБОТУ

студентові Удовиці Даниїлі Валеріївні

(прізвище, ім'я, по батькові)

1. Тема роботи Застосування методів багатовимірної статистичного аналізу для дослідження динаміки розвитку економічних систем

затверджена наказом по університету від 31 жовтня 2019 р. № 1600 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 9 грудня 2019 р.

3. Вихідні дані до роботи статистичні дані економічних показників за регіонами України за 2000 р., 2005 р., 2010 р., 2015 р.

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Системний аналіз проблеми реалізації методів багатовимірної статистичного аналізу та постановка задач дослідження

2. Вибір і обґрунтування методу розв'язання

3. Програмна реалізація

4. Результати обчислювального експерименту

5. Аналіз можливих застосувань

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій _____

1. Актуальність теми роботи _____

2. Постановка задачі _____

3. Системний аналіз проблеми _____

4. Метод чисельного аналізу _____

5. Результати обчислювального експерименту _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Підбір та вивчення технічної літератури за темою роботи	вересень 2019 р.	виконано
2	Вибір та обґрунтування методу	жовтень – листопад 2019 р.	виконано
3	Розробка алгоритму і програми	листопад – грудень 2019 р.	виконано
4	Проведення аналітичних досліджень та розрахунків	листопад – грудень 2019 р.	виконано
5	Робота над текстом пояснювальної записки	грудень 2019 р.	виконано
6	Представлення роботи на рецензію в ЕК	грудень 2019 р.	виконано

Дата видачі завдання 2 вересня 2019 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Гибкіна Н.В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 98 с., 8 табл., 39 рис., 2 дод., 17 джерел.

КОМПОНЕНТНИЙ АНАЛІЗ, СТАТИСТИЧНІ ДАННІ, ФАКТОРНИЙ АНАЛІЗ, ГОЛОВНА КОМПОНЕНТА, ДИСПЕРСІЯ, КОВАРІАЦІЙНА МАТРИЦЯ, МАТРИЦЯ НАВАНТАЖЕНЬ.

Об'єкт дослідження – економічна система регіонів України.

Мета роботи – застосування методів компонентного аналізу та факторного аналізу для дослідження динаміки розвитку економічних систем.

Метод дослідження – метод компонентного аналізу, метод факторного аналізу.

У роботі проведено системний аналіз проблеми дослідження багатовимірних даних, що змінюються у часі. Побудовані морфологічна, інформаційна, функціональна моделі об'єкта дослідження.

Досліджено економічну систему, елементами якої є регіони України, що описуються набором соціально-економічних факторів. Методи компонентного та факторного аналізу застосовано для отримання інтегральних показників за кожним регіоном на основі сукупності вихідних факторів. Графічне подання результатів здійснено з використання методів кластеризації. Розроблено програмний продукт, який дозволяє відслідковувати зміни в економічному положенні регіонів України протягом часу. За допомогою розробленого продукту проведено серію обчислювальних експериментів та отримано ряд графіків, які наглядно відображають особливості та зміни економічного стану досліджуваних об'єктів у XXI столітті.

ABSTRACT

Introductory note: 98 pages, 8 tables, 39 figures, 2 appendixes, 17 sources.

COMPONENT ANALYSIS METHOD, STATISTICAL DATA, FACTOR METHOD, MAIN COMPONENT, DISPERSION, CORRELATION MATRIX.

Object of research – the economic system of Ukraine regions.

Purpose of work – application of methods of component analysis and factor analysis to study the dynamics of economic systems development.

Methods of research – method of component analysis, method of factor analysis.

The paper analyzes a systematic analysis of the problem of multivariate data changing over time. Morphological, informative, functional models of the object of study are constructed.

The economic system, elements of which are regions of Ukraine, described by a set of socio-economic factors, is investigated. Component and factor analysis methods were applied to obtain integral metrics for each region based on a set of baseline factors. Graphical presentation of results is carried out using clustering methods. A software product has been developed to track changes in the economic situation of Ukrainian regions over time. With the help of the developed product, a series of computational experiments were conducted and a series of graphs were obtained that clearly depicts the peculiarities and changes of the economic state of the monitored objects in the 21st century.

ЗМІСТ

	С.
Вступ	8
1 Системний аналіз проблеми реалізації методів багатовимірною статистичного аналізу та постановка задач дослідження	10
1.1 Системний аналіз проблеми реалізації методів багатовимірною статистичного аналізу	10
1.1.1 Вербальна модель системи	10
1.1.2 Морфологічний опис системи	10
1.1.3 Функціональна модель системи	13
1.1.4 Інформаційна модель системи	15
1.2 Аналіз сценаріїв вирішення проблеми реалізації методів багатовимірною статистичного аналізу	16
1.2.1 Модель аналізу проблеми	16
1.2.2 Оцінювання вектора пріоритетів незадоволеностей методом аналізу ієрархій	16
1.2.3 Модель вирішення проблеми	21
1.3 Змістовна та формальна постановка задач	21
1.3.1 Змістовна постановка задач	21
1.3.2 Формальна постановка задач	22
1.4 Постановка задач дослідження	25
2 Вибір та обґрунтування методу розв'язання	26
2.1 Огляд методів зниження вимірності	26
2.1.1 Багатовимірний нормальний розподіл як модель	27
2.1.2 Багатовимірне шкалювання	28
2.2 Метод головних компонент	30
2.3 Факторний аналіз	34
2.3.1 Основні положення факторного аналізу	34
2.3.2 Алгоритм факторного аналізу	35

	7
2.4 Кластерний аналіз	43
2.5 Алгоритм розв'язання задачі	46
2.5.1 Алгоритм розв'язання задачі дослідження динаміки розвитку економічних систем у часі.....	46
2.5.2 Аналіз початкових даних	47
2.5.3 Застосування методів багатовимірною статистичного аналізу для розв'язання задачі дослідження розвитку економічних систем у часі	49
3 Програмна реалізація	59
3.1 Mathematica 11 як система символної математики.....	59
3.2 Опис програми.....	62
4 Результати обчислювального експерименту	63
4.1 Аналіз економічного стану регіонів України методами головних компонент та факторного аналізу	63
4.2 Дослідження економічного стану регіонів України у динаміці	69
5 Аналіз можливих застосувань	77
Висновки	79
Перелік джерел посилання	80
Додаток А Лістинг програми	82
Додаток Б Результати роботи програми	87

ВСТУП

Набуття Україною незалежності у 1991 році спричинило перебудову всіх галузей господарювання. Створення збалансованого господарського комплексу на ринкових засадах передбачало не тільки зміну функціональних зв'язків, а й кардинальну реструктуризацію економіки, інше розміщення продуктивних сил.

З відновленням незалежності Україна отримала шанс перетворитися в розвинену країну Центральної та Східної Європи, але цей процес супроводжувався ризиками втрати переваг, властивих її економіці (достатньо розвинена промисловість, зокрема така сучасні виробництва, як авіа- і ракетобудування, системи космічного управління, високий рівень освіченості населення, високий науково-технічний потенціал, можливості для динамічного розвитку сільського господарства тощо).

Через низку об'єктивних і суб'єктивних факторів шанси України швидко наблизитись до розвинених країн були втрачені, а ризики нівелювання своїх переваг значною мірою проявилися. Протягом останніх трьох десятиліть періоди падіння та зростання економіки чергувалися, що значно утруднювало процес розвитку країни. При цьому окремі регіони України, завдяки історичним та іншим особливостям суттєво відрізняються за своїм економічним положенням, що ще збільшує нерівномірність розвитку.

За таких умов постає задача дослідження великої кількості показників їх соціально-економічного розвитку. Зазвичай для цього проводиться аналіз кожного з цих показників з урахуванням тенденцій економічного розвитку, історичних передумов, природних особливостей та робляться комплексні висновки щодо статусу аналізованих об'єктів [15, 16]. Зауважимо, що наявність великої кількості різнорідних показників, як правило, робить подібний аналіз надзвичайно складним. Отже, доцільним є застосування методів, що дозволяють зменшити кількість інформації, яку потрібно обробити для формулювання висновків, або інтегрувати велику кількість вихідних даних у масив значно меншої вимірності без втрати інформативності. Такий апарат надають математичні ме-

тоди багатовимірного статистичного аналізу, зокрема, методи зниження вимірності. Так, у роботах [6, 7, 11, 12] розглянуто застосування цих методів до аналізу соціально-економічного положення країн ЄС та окремих областей України.

Ціль атестаційної роботи полягає у дослідженні економічних систем на прикладі системи регіонів України за допомогою методів багатовимірного статистичного аналізу та отримання висновків про динаміку її розвитку.

1 СИСТЕМНИЙ АНАЛІЗ ПРОБЛЕМИ РЕАЛІЗАЦІЇ МЕТОДІВ БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Системний аналіз проблеми реалізації методів багатовимірною статистичного аналізу

1.1.1 Вербальна модель системи

Застосування методів багатовимірною статистичного аналізу дозволяє перейти в дослідженні системи об'єктів, які описуються багатьма показниками, часто є корельованими між собою, до системи нових, некорельованих показників, кількість яких є значно меншою, а інформативність у порівнянні з початковою системою знижена несуттєво. До статистичних методів, які дозволяють розв'язати цю задачу, належать методи зниження вимірності, зокрема, метод факторного аналізу та один з його різновидів – метод головних компонент.

Кожен з цих методів дозволяє визначити низку нових, інтегральних показників, кількість яких менша за кількість вихідних ознак системи. На їх основі можна проводити подальший аналіз стану досліджуваної системи об'єктів.

1.1.2 Морфологічний опис системи

Система, що досліджується – система «Методи багатовимірною статистичного аналізу».

Мета системи – вивчення можливості застосування методів багатовимірною статистичного аналізу в дослідженні змінення стану складних, зокрема, економічних систем.

Призначення системи – на основі отриманих даних давати оцінку стану економічних систем, спостерігати змінення його в часі та прогнозувати пода-

льший розвиток.

Зовнішнє середовище – сукупність всіх об’єктів поза межами системи, зміна властивостей яких впливає на систему, а також тих об’єктів, чії властивості змінюються в результаті функціонування системи. Модель зовнішнього середовища наведена на рис. 1.1.



Рисунок 1.1 – Модель зовнішнього середовища системи

Об’єкти зовнішнього середовища:

а) рівень складності реалізації того чи іншого методу залежить від вибору алгоритмічної мови;

б) розвиток техніки надає нові задачі, для яких можливо з’явиться необхідність розробити нові або вдосконалити вже існуючі методи;

в) розвиток ЕОМ визначає швидкодію виконання роботи наближених методів на машинах;

г) програмне забезпечення надає зручний інтерфейс для особи, яка досліджує отриманий за допомогою реалізованих методів наближений розв’язок.

Модель типу «чорний ящик» системи «Методи багатовимірного статистичного аналізу» наведено на рис. 1.2. Така модель є вихідною при побудові

моделі складної системи, акцентує увагу дослідника на взаємодії системи із зовнішнім середовищем. Тут виходами системи є цільові продукти, а входи – це дія середовища на систему. Зміст «чорного ящика» не розкривається, оскільки увага зосереджена лише на межах системи. Межі, в свою чергу, підкреслюють цілісність системи, відокремленість її від зовнішнього середовища і взаємодію системи та середовища.

В моделі типу «білий ящик» (рис. 1.3) описаний «вміст системи», вказані взаємозв'язки між елементами системи.



Рисунок 1.2 – Модель типу «чорний ящик»

Дана система складається з двох підсистем:

а) А1 – керуюча система, яка складається з:

- 1) А11 – дослідник;
- 2) А12 – засоби керування.

б) А2 – система, якою керують, складається з:

- 1) А21 – математичний апарат;
- 2) А22 – програмне забезпечення;
- 3) А23 – апаратне забезпечення.

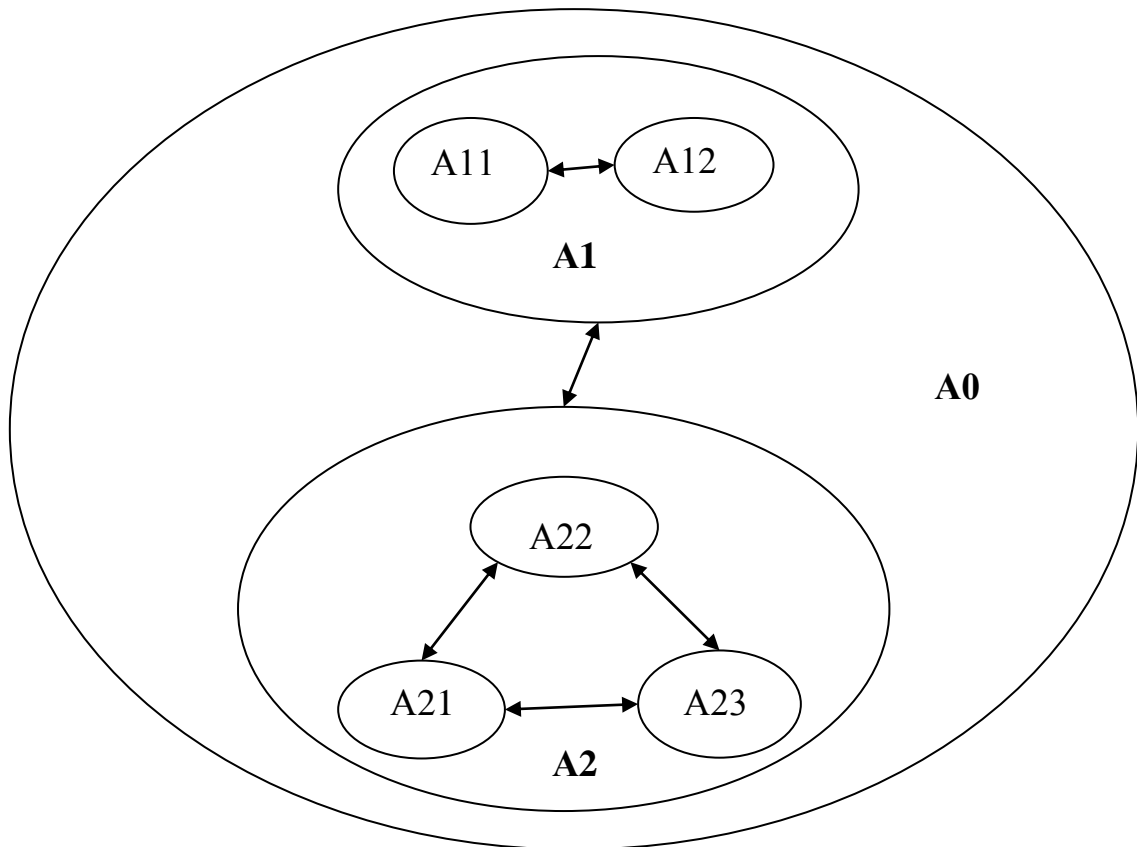


Рисунок 1.3 – Модель типу «білий ящик»

1.1.3 Функціональна модель системи

Перша діаграма в ієрархії діаграм – IDEF0 (рис. 1.4). Ця діаграма відображає функціонування системи в цілому і називається контекстною діаграмою.

Після того, як контекст описаний, проводиться побудова наступних діаграм в ієрархії. Кожна наступна діаграма є більш деталізованим описом однієї з робіт на діаграмі вищого рівня. Друга діаграма демонструє перший рівень декомпозиції (рис. 1.5).

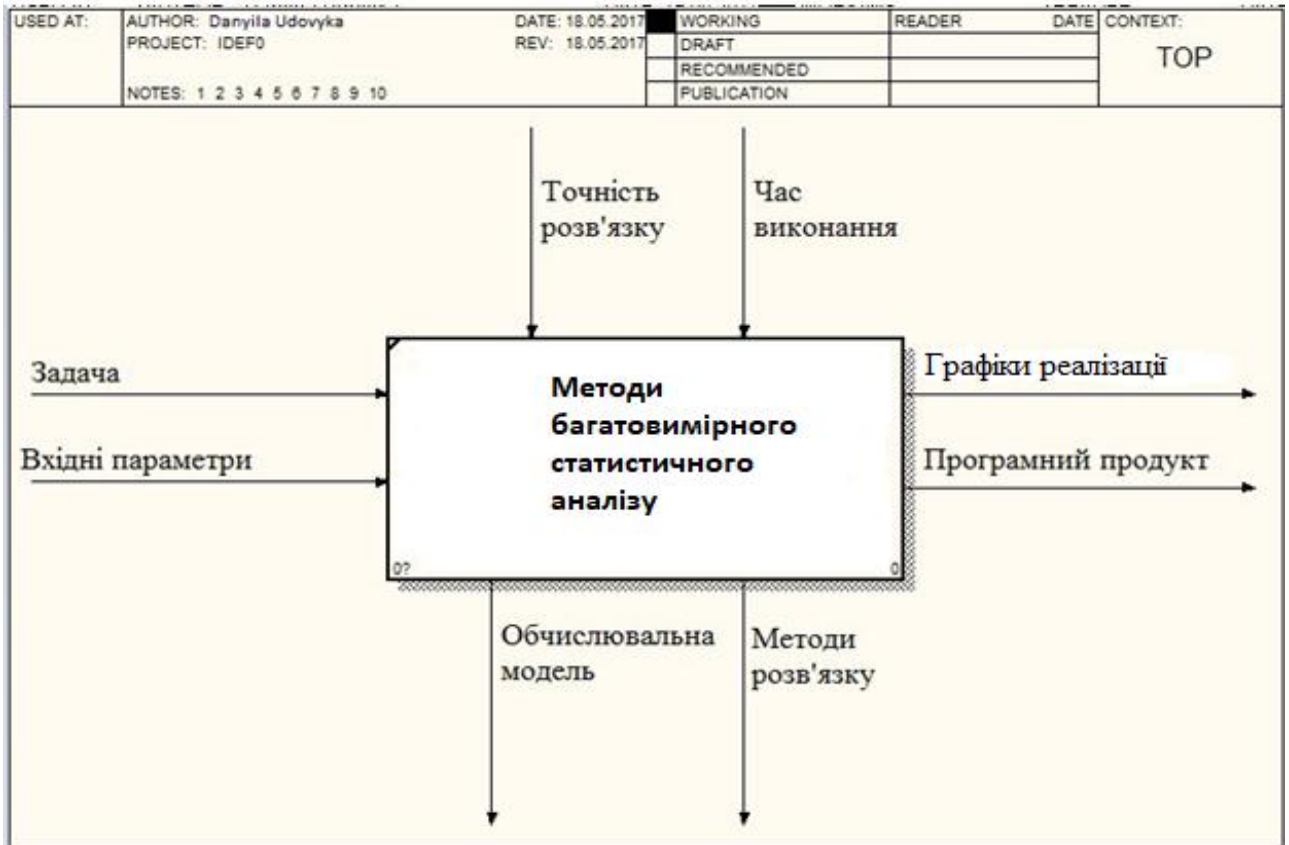


Рисунок 1.4 – Контекстна діаграма

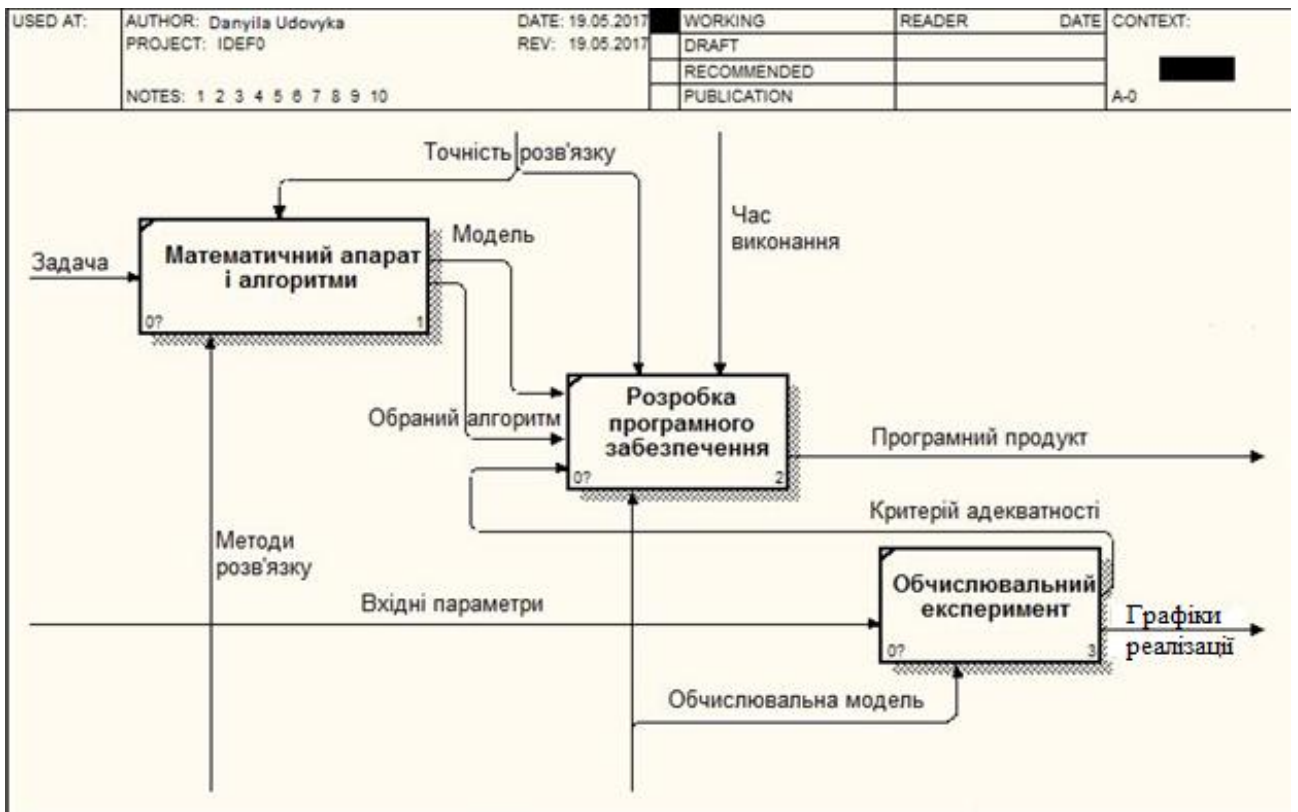


Рисунок 1.5 – IDEFO-діаграма першого рівня декомпозиції

1.1.4 Інформаційна модель

Інформаційні моделі відображають різні типи систем об'єктів, в яких реалізуються різні структури взаємодії і взаємозв'язку між елементами системи.

Для відображення системи з різними структурами використовують різні типи інформаційних моделей: табличні, ієрархічні і мережні.

DFD (Data Flow Diagrams) – діаграми потоків даних. Інструменти методології DFD дозволяють відображати джерела і адресати даних, ідентифікувати процеси і групи даних, що зв'язують в потоки одну функцію з іншою і ефективно використовуються для опису процесів при впровадженні процесного підходу до управління організацією, оскільки дозволяють максимально знизити суб'єктивність опису бізнес-процесів.

Крім того, нотація DFD дозволяє описувати потоки документів і потоки ресурсів.

Діаграма DFD інформаційної моделі системи представлена на рис. 1.6.

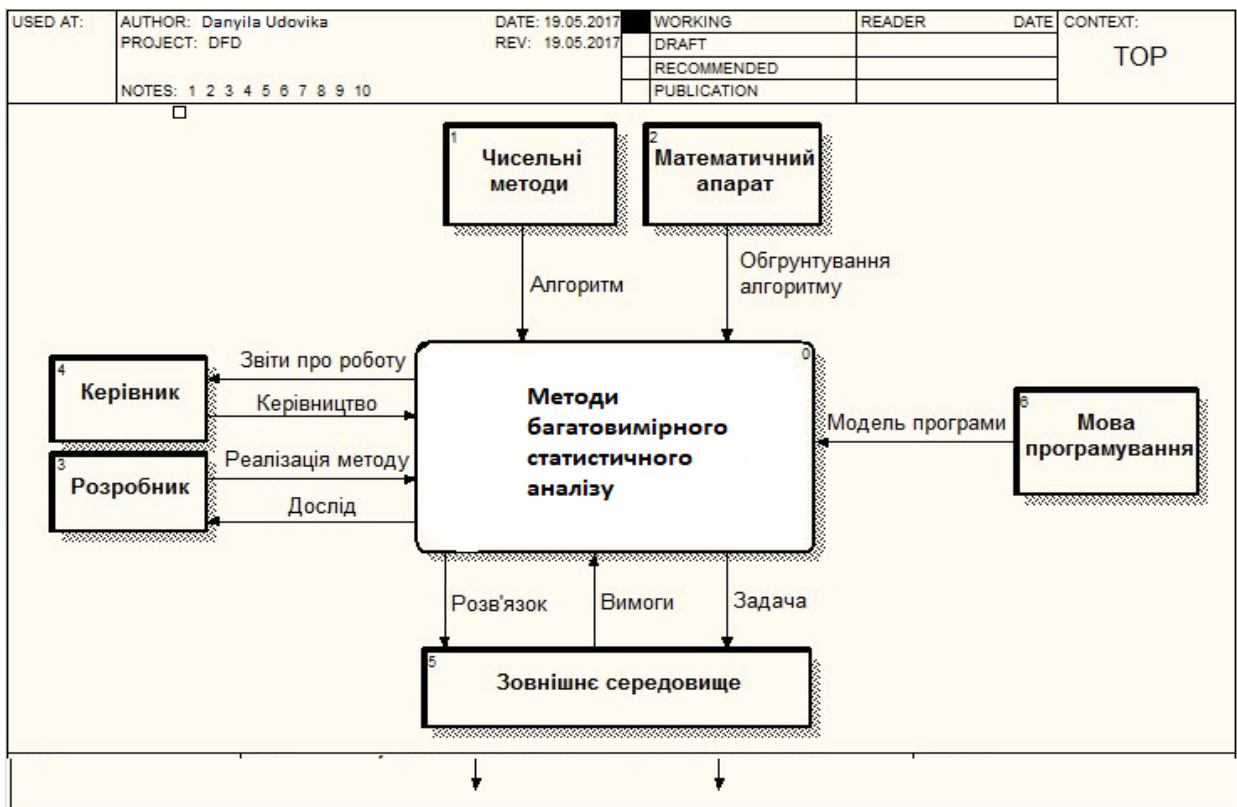


Рисунок 1.6 – DFD-діаграма

1.2 Аналіз сценаріїв вирішення проблеми реалізації методів багатовимірного статистичного аналізу

1.2.1 Модель аналізу проблеми

Розглянемо задачу про вибір методу, яким буде досліджуватися економічна система. Проблема буде вирішуватися за наступними критеріями:

- критерій 1: час роботи програми;
- критерій 2: складність алгоритму при програмуванні;
- критерій 3: необхідні для проведення розрахунків ресурси;
- критерій 4: точність отриманого розв'язку;
- критерій 5: універсальність;
- критерій 6: вартість програмного забезпечення.

Обирати будемо з множини альтернатив:

- альтернатива 1: методи експериментного оцінювання;
- альтернатива 2: методи багатовимірного статистичного аналізу;
- альтернатива 3: класичні методи статистики.

Побудуємо ієрархічну структуру, використовуючи метод парних порівнянь. Ієрархічна структура приведена на рис. 1.7.

1.2.2 Оцінювання вектора пріоритетів незадоволеностей методом аналізу ієрархій

Для продовження аналізу необхідно побудувати матриці парних порівнянь за критеріями та за альтернативами.

Після цього слід побудувати вектор локальних та глобальних пріоритетів та знайти найбільш оптимальну альтернативу для розв'язування поставленої у атестаційній роботі задачі.

Матриця попарних порівнянь критеріїв наведена у таблиці 1.1.

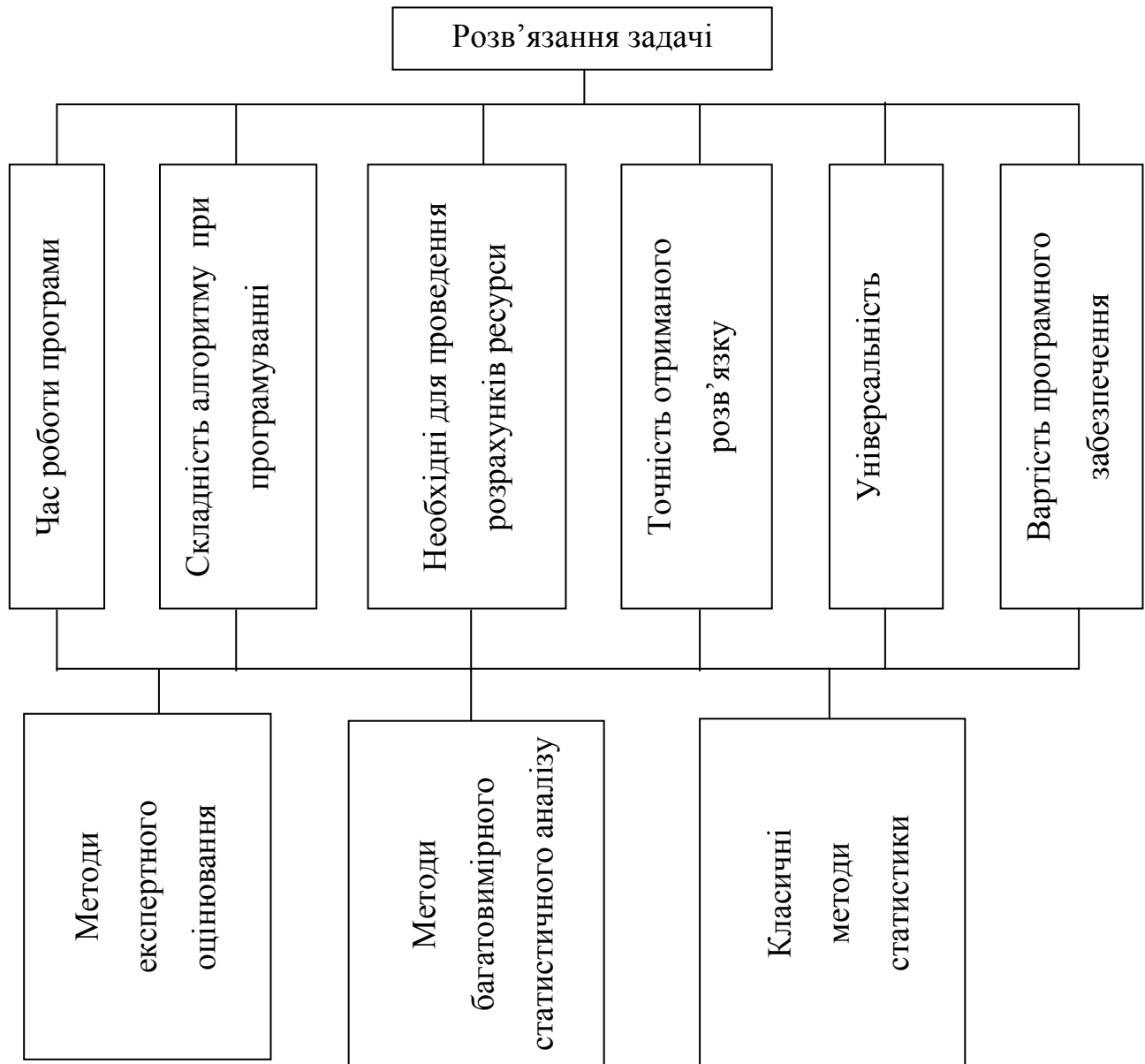


Рисунок 1.7 – Ієрархічна модель процесу аналізу розв'язання задачі

Для таблиці 1.1 індекс узгодженості $(IY) = \frac{6,635 - 6}{6 - 1} = 0,127$, узгодженість $= 1,24$, відносна узгодженість $(BY) = \frac{0,127}{1,24} = 0,1024 = 10,24\%$.

Для прийняття рішення про використання методу необхідно провести порівняльний аналіз альтернатив.

Оцінивши кожну з альтернатив, отримаємо дані, які представлені в таблицях 1.2 – 1.7.

Узгодженість для матриць дорівнює 0,58.

Для таблиці 1.2 індекс узгодженості (ІУ) = 0,0595, відносна узгодженість (ВУ) = 0,1025 = 10,25%, а для таблиці 1.3 індекс узгодженості (ІУ) = 0,027, відносна узгодженість (ВУ) = 0,0465 = 4,65%.

Для таблиці 1.4 індекс узгодженості (ІУ) = 0,038, відносна узгодженість (ВУ) = 0,0655 = 6,55%.

Таблиця 1.1 – Матриця попарних порівнянь

Критерії оцінювання	K1	K2	K3	K4	K5	K6	Оцінки компонент	Вектор пріоритетів	Величина значущості
K1	1	3	9	8	4	8	4,364	0,472	0,918
K2	$\frac{1}{3}$	1	5	8	3	5	2,418	0,261	1,270
K3	$\frac{1}{9}$	$\frac{1}{5}$	1	$\frac{1}{2}$	2	4	0,668	0,072	1,282
K4	$\frac{1}{8}$	$\frac{1}{8}$	2	1	$\frac{1}{6}$	$\frac{1}{2}$	0,371	0,040	1,023
K5	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	6	1	6	1,070	0,116	1,195
K6	$\frac{1}{8}$	$\frac{1}{5}$	$\frac{1}{4}$	2	$\frac{1}{6}$	1	0,357	0,039	0,947
Всього							9,249		6,635

Таблиця 1.2 – Порівняння за першим критерієм

Критерій 1	A1	A2	A3	Власний вектор	Вектор пріоритетів	Величина значущості
A1	1	$\frac{1}{5}$	$\frac{1}{2}$	0,464	0,106	0,845
A2	5	1	7	3,271	0,744	1
A3	2	$\frac{1}{7}$	1	0,659	0,150	1,274
Всього				4,394		3,119

Таблиця 1.3 – Порівняння за другим критерієм

Критерій 2	A1	A2	A3	Власний вектор	Вектор пріоритетів	Величина значущості
A1	1	$\frac{1}{2}$	4	1,260	0,345	1,120
A2	2	1	4	2,0	0,547	0,957
A3	$\frac{1}{4}$	$\frac{1}{4}$	1	0,397	0,109	0,977
Всього				3,657		3,054

Таблиця 1.4 – Порівняння за третім критерієм

Критерій 3	A1	A2	A3	Власний вектор	Вектор пріоритетів	Величина значущості
A1	1	2	$\frac{1}{8}$	0,630	0,129	1,229
A2	$\frac{1}{2}$	1	$\frac{1}{7}$	0,415	0,085	0,852
A3	8	7	1	3,826	0,785	0,996
Всього				4,871		3,076

Таблиця 1.5 – Порівняння за четвертим критерієм

Критерій 4	A1	A2	A3	Власний вектор	Вектор пріоритетів	Величина значущості
A1	1	3	7	2,759	0,649	0,9582
A2	$\frac{1}{3}$	1	5	1,186	0,279	1,1716
A3	$\frac{1}{7}$	$\frac{1}{5}$	1	0,306	0,072	0,9351
Всього				4,25		3,065

Для таблиці 1.5 індекс узгодженості (ІУ) = 0,0324, відносна узгодженість (ВУ) = 0,0559 = 5,59%.

Для таблиці 1.6 індекс узгодженості (ІУ) = 0,068, відносна узгодженість (ВУ) = 0,1172 = 11,72%.

Для таблиці 1.7 індекс узгодженості (ІУ) = 0,027, відносна узгодженість (ВУ) = 0,0466 = 4,66%.

Таблиця 1.6 – Порівняння за п'ятим критерієм

Критерій 5	A1	A2	A3	Власний вектор	Вектор пріоритетів	Величина значущості
A1	1	$\frac{1}{3}$	5	1,186	0,297	1,247
A2	3	1	5	2,466	0,618	0,947
A3	$\frac{1}{5}$	$\frac{1}{5}$	1	0,342	0,086	0,942
Всього				3,994		3,136

Таблиця 1.7 – Порівняння за шостим критерієм

Критерій 6	A1	A2	A3	Власний вектор	Вектор пріоритетів	Величина значущості
A1	1	$\frac{1}{4}$	$\frac{1}{2}$	0,500	0,131	0,918
A2	4	1	4	2,520	0,661	0,991
A3	2	$\frac{1}{4}$	1	0,794	0,208	1,145
Всього				3,814		3,054

1.2.3 Модель вирішення проблеми

З усіх отриманих результатів ми, як особа, що приймає рішення, можемо зробити кінцеві підрахунки та підсумувати висновки.

У таблиці 1.8 наведені результати, які дозволяють нам сказати, що кращою для нас буде друга альтернатива, а саме методи багатовимірною статистичного аналізу.

Таблиця 1.8 – Кінцеві дані

Критерій /Альтернатива	K1	K2	K3	K4	K5	K6	Узагальнені пріоритети
A1	0,106	0,345	0,129	0,649	0,297	0,131	0,317
A2	0,744	0,547	0,085	0,279	0,618	0,661	0,385
A3	0,150	0,109	0,785	0,072	0,086	0,208	0,298

1.3 Змістовна та формальна постановка задачі

1.3.1 Змістовна постановка задачі

Розв'язується задача класифікації економічних об'єктів за низкою соціально-економічних факторів та дослідження змінення стану цих об'єктів з плином часу. Як приклад розглядається економічна система України, представлена своїми регіонами. Для оцінювання рівня розвитку кожного з регіонів використовуються різноманітні показники, що характеризують валовий регіональний продукт (ВРП) та ВРП на душу населення, ефективність економіки, рівень та якість життя населення тощо.

Значення цих показників залежать від особливостей географічного та соціально-економічного положення регіонів, а також змінюються з плином часу залежно від тенденцій їх економічного розвитку. Тому задача класифікації та

порівняння регіонів за рівнем соціально-економічного стану стає складною з точки зору урахування інформації, поданої за всіма показниками, та наочного подання результатів.

Для розв'язання задачі класифікації та зменшення вимірності у атестаційній роботі запропоновано використовувати такі методи багатовимірної статистичного аналізу даних, як метод головних компонент та метод факторного аналізу.

1.3.2 Формальна постановка задачі

Досліджувана система складається з n об'єктів, стан кожного з яких описується набором з p параметрів $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, отже, вихідною інформацією є набір з n p -вимірних даних:

$$X_1 = (x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(p)})^T, X_2 = (x_2^{(1)}, x_2^{(2)}, \dots, x_2^{(p)})^T, \dots, X_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(p)})^T.$$

Задача зниження вимірності полягає у побудові для кожного спостереження X_i нового вектору ознак $Y_i = (y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(m)})$, де m є набагато меншим за p , причому таке стискання інформації повинне забезпечувати її мінімальну втрату.

Оскільки кожен з p показників має свої одиниці виміру, то перед розв'язанням поставленої задачі слід перейти до нормованих величин $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(p)}$ за формулами:

$$\tilde{X}_1 = (\tilde{x}_1^{(1)}, \tilde{x}_1^{(2)}, \dots, \tilde{x}_1^{(p)})^T, \tilde{X}_2 = (\tilde{x}_2^{(1)}, \tilde{x}_2^{(2)}, \dots, \tilde{x}_2^{(p)})^T, \dots, \tilde{X}_n = (\tilde{x}_n^{(1)}, \tilde{x}_n^{(2)}, \dots, \tilde{x}_n^{(p)})^T,$$

де

$$\tilde{x}_i^{(j)} = \frac{x_i^{(j)} - \bar{x}^{(j)}}{\hat{\sigma}^{(j)}}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p.$$

Тут $\bar{x}^{(j)} = \frac{1}{n} \sum_{k=1}^n x_k^{(j)}$ та $\hat{\sigma}^{(j)} = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k^{(j)} - \bar{x}^{(j)})^2}$ – вибіркове середнє та вибіркове середньоквадратичне відхилення для j -го показника відповідно, $j = 1, 2, \dots, p$.

Відповідно до методу головних компонент для розв'язання задачі зниження вимірності необхідно знайти таке лінійне перетворення вихідних змінних:

$$y^{(j)} = \sum_{i=1}^p l_{ji} \tilde{x}^{(i)}, \quad j = 1, 2, \dots, p,$$

для якого

$$\sum_{i=1}^p l_{ji}^2 = 1, \quad j = 1, 2, \dots, p,$$

$$\text{cov}(y^{(i)}, y^{(j)}) = 0, \quad \text{якщо } j < i,$$

$$\mathbf{D}y^{(1)} \geq \mathbf{D}y^{(2)} \geq \dots \geq \mathbf{D}y^{(p)}.$$

Величини l_{ji} , $i, j = 1, 2, \dots, p$, є невідомими і визначаються з умови, що перша вибіркова головна компонента $y^{(1)}$ як нормована лінійна комбінація показників $\tilde{x}^{(1)}$, $\tilde{x}^{(2)}$, ..., $\tilde{x}^{(p)}$, серед інших нормованих лінійних комбінацій цих показників має найбільшу дисперсію; друга вибіркова головна компонента $y^{(2)}$, як нормована лінійна комбінація показників $\tilde{x}^{(1)}$, $\tilde{x}^{(2)}$, ..., $\tilde{x}^{(p)}$, серед інших нормованих лінійних комбінацій цих показників, некорельованих з $y^{(1)}$, має найбільшу дисперсію і так далі.

За таких умов підмножина перших m змінних $y^{(1)}$, $y^{(2)}$, ..., $y^{(p)}$ ($m < p$)

пояснюватиме більшу частину загальної дисперсії і, таким чином, отримаємо стиснуте описання структури залежності вихідних змінних.

Відповідно до методу факторного аналізу для розв'язання задачі зниження вимірності необхідно визначити коефіцієнти f_i , $i = 1, 2, \dots, m$, лінійної моделі факторного аналізу

$$x^{(j)} = \sum_{i=1}^m l_{ji} f_i + e_j, \quad j = 1, 2, \dots, p,$$

де $\vec{f} = (f_1, f_2, \dots, f_m)^T$ – вектор спільних факторів, що відображають ті фактори, які обумовлюють залежність між $x^{(1)}$, $x^{(2)}$, ..., $x^{(p)}$, причому f_i є незалежними і $f_i \sim N(0, 1)$, $i = 1, 2, \dots, m$;

$\vec{e} = (e_1, e_2, \dots, e_p)^T$ – вектор специфічних факторів, що відображають локальні причини-фактори, $e_j \sim N(0, v_j)$, $j = 1, 2, \dots, p$;

l_{ij} – навантаження i -ї змінної на j -й фактор;

$L = [l_{ij}]_{p \times m}$ – матриця факторних навантажень.

Коваріаційна матриця C системи випадкових величин $x^{(1)}$, $x^{(2)}$, ..., $x^{(p)}$ дорівнює

$$C = LL^T + V,$$

де $V = \text{diag}\{v_1, v_2, \dots, v_p\}$ – діагональна матриця залишкових дисперсій v_j , $j = 1, 2, \dots, p$.

За умови, що $m \ll p$, випадкові величини $x^{(1)}$, $x^{(2)}$, ..., $x^{(p)}$ пов'язані одна з одною за допомогою лише невеликої кількості випадкових величин f_1, f_2, \dots, f_m , що відображають ті причини-фактори, які обумовлюють залежність між $x^{(1)}$, $x^{(2)}$, ..., $x^{(p)}$.

1.4 Постановка задач дослідження

Виходячи з проведеного системного аналізу проблеми вибору ефективного методу розв'язання задачі зниження вимірності під час дослідження складних економічних систем, сформулюємо задачі дослідження даної атестаційної роботи:

- сформулювати задачу дослідження динаміки розвитку економічних систем з плином часу на прикладі економічної системи України, представленої своїми регіонами;
- методами компонентного та факторного аналізу розв'язати задачу дослідження динаміки розвитку економічних систем, що описуються низкою соціально-економічних факторів, з плином часу;
- розробити програмний продукт, який реалізує задіяні методи;
- використовуючи отримані результати, в тому числі і графічні, проаналізувати динаміку економічного розвитку регіонів України.

2 ВИБІР ТА ОБҐРУНТУВАННЯ МЕТОДУ РОЗВ'ЯЗАННЯ

2.1 Огляд методів зниження вимірності

Метою методів зниження вимірності є стиснення множини ознак, що описують досліджувану систему випадкових величин, без істотної втрати інформації, яку вони містять. Таке стиснення полягає у поданні великої кількості вихідних факторів, безпосередньо виміряних на об'єктах, через меншу кількість максимально інформативних, але безпосередньо не відслідковуваних внутрішніх характеристик об'єктів, які відображають найбільш істотні властивості досліджуваних об'єктів.

Метод головних компонент виявляє деяку кількість компонент, що пояснюють всю дисперсію і кореляції вихідних випадкових величин. При цьому компоненти розташовуються в порядку спадання, пояснюючи частки сумарної дисперсії вихідних величин, що дозволяє часто обмежитися декількома першими компонентами.

Факторний аналіз виявляє задану кількість загальних для всіх вихідних величин факторів, пояснюючи дисперсію, що залишилася після цього, впливом специфічних факторів.

Серед прикладних задач, що вирішуються зазначеними методами, належать наступні:

- пошук прихованих, але об'єктивно існуючих взаємозв'язків між показниками, перевірка гіпотез про взаємозв'язки між ними, виявлення природи відмінностей між об'єктами;

- опис досліджуваної системи кількістю ознак, значно меншою кількості вихідних факторів, що містять в середньому більше інформації, ніж безпосередньо зафіксовані на об'єктах значення вихідних факторів;

- візуалізація вихідних багатовимірних спостережень шляхом їх проектування на спеціально підібрану пряму, площину або тривимірний простір;

- побудова регресійних моделей за головними компонентами;

- класифікація за узагальненими показниками;
- стиснення вихідної інформації, значне зменшення обсягів інформації, що зберігається в базах даних, без істотних втрат в інформативності.

2.1.1 Багатовимірний нормальний розподіл як модель

Багато методів багатовимірного статистичного аналізу (в тому числі і методи факторного і компонентного аналізу) ґрунтуються на припущенні, що досліджувані величини мають багатовимірний нормальний розподіл. Зокрема, цей розподіл часто є найкращим наближенням до дійсного розподілу в прикладних економічних задачах. В даний час багатовимірні методи, засновані на нормальному розподілі, знайшли широке застосування при вивченні різних процесів в економіці [2].

Серед математичних методів багатовимірного аналізу виділяють наступні.

1. Аналіз кореляції. При дослідженні кореляції обчислюються:

- вибіркові коефіцієнти кореляції, що використовуються для оцінки відповідних параметрів розподілу;
- частинні коефіцієнти кореляції, які вимірюють залежність між випадковими величинами, коли дія інших корельованих випадкових величин виключена;
- множинний коефіцієнт кореляції, що поширює поняття коефіцієнта кореляції на вимірювання залежності між однією випадковою величиною і багатьма іншими випадковими величинами.

2. Аналоги одновимірних статистичних методів в багатовимірному аналізі. Застосування аналогів пов'язано з тим, що багато проблем, які виникають під час вивчення багатовимірних сукупностей, базуються на подібних задачах для одновимірних сукупностей. Для цих проблем вибір системи координат пов'язаний з лінійним перетворенням змінних.

3. Проблеми системи координат. Вдалий вибір нової системи координат може найбільш економним способом виявити деякі важливі для дослідника влас-

тивості багатовимірної випадкової сукупності. Прикладом може служити виявлення головних компонент, тобто відшукування такої нормалізованої лінійної комбінації випадкових величин, щоб її дисперсія була максимальною або мінімальною. Це рівноцінно повороту осей, який призводить коваріаційну матрицю до діагонального виду. Інший приклад – знаходження канонічних кореляцій. Для вирішення подібних задач потрібно визначення характеристичних векторів і характеристичних коренів різних систем лінійних алгебраїчних рівнянь.

4. Проблеми класифікації, тобто розбиття множини випадкових величин на підмножини. Виникає важливе питання перевірки гіпотези про незалежність підмножин. Факторний аналіз, метод головних компонент і кластерний аналіз зазвичай використовують в задачах багатовимірної класифікації.

5. Залежність спостережень. Якщо в економічних дослідженнях здійснюється аналіз часових рядів, то дослідник стикається зі спостереженнями над рядами випадкових величин, що є послідовними в часі. Спостереження в даний момент часу можуть залежати від раніше зроблених спостережень. Це вимагає, наприклад, вивчення внутрірядової кореляції.

2.1.2 Багатовимірне шкалювання

Розглянемо застосування методу багатовимірного шкалювання до розв'язання задачі зниження вимірності простору ознак [3]. Нехай вихідна інформація про об'єкти задана в формі матриці їх попарних порівнянь

$$\gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} \\ \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{pmatrix}, \quad (2.1)$$

де γ_{ij} інтерпретується як евклідова відстань між i -м та j -м об'єктами,

$i, j = 1, 2, \dots, n :$

$$\gamma_{ij} = \sqrt{\sum_{k=1}^p (x_i^{(k)} - x_j^{(k)})^2}. \quad (2.2)$$

При цьому ні координати об'єктів $X_i = (x_i^{(1)}, \dots, x_i^{(p)})^T$, ні вимірність їх простору ознак p невідомі. Потрібно на підставі відомих даних (2.1) відновити невідому вимірність p аналізованого простору ознак і приписати кожному i -му об'єкту координати $X_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})^T$ так, щоб обчислені за формулою (2.2) попарні евклідові відстані по можливості співпали б з заданими матрицею γ .

Відновлення координат X_1, X_2, \dots, X_n при цьому відбувається з точністю до ортогонального перетворення, оскільки при ортогональному перетворенні системи координат попарні відстані між об'єктами не змінюються. Отже, мета методів метричного багатовимірного шкалювання полягає в тому, щоб відобразити інформацію про конфігурацію вихідних багатовимірних даних, задану матрицею відстаней γ , у вигляді геометричної конфігурації n точок у відповідному багатовимірному просторі.

В неметричному методі шкалювання передбачається, що відмінності (близькості) γ_{ij} виміряні в ординальній шкалі, так що важливий тільки ранговий порядок відмінностей, а не самі їх чисельні значення. Процедури неметричного методу шкалювання дозволяють побудувати таку геометричну конфігурацію точок в просторі заданої вимірності p' , щоб ранговий порядок попарних відстаней між ними, по можливості, мінімально відрізнявся від того порядку, який заданий матрицею γ . Процедури неметричного методу шкалювання відрізняються одна від одної вибором критерію відмінності двох різних упорядкувань.

2.2 Метод головних компонент

Суть компонентного аналізу полягає у наступному [1, 2, 3]. Розглядається система з n об'єктів, стан кожного з яких описується p показниками $x^{(1)}$, $x^{(2)}$, ..., $x^{(p)}$. Це означає, що після проведення спостережень дослідник матимемо набір з n p -вимірних даних

$$X_1 = (x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(p)})^T, X_2 = (x_2^{(1)}, x_2^{(2)}, \dots, x_2^{(p)})^T, \dots, X_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(p)})^T.$$

Зауважимо, що кожен з даних показників може мати свої одиниці вимірювання, тому для використання методу головних компонент необхідно провести нормування вихідної інформації $\tilde{x}^{(1)}$, $\tilde{x}^{(2)}$, ..., $\tilde{x}^{(p)}$ за формулами

$$\tilde{X}_1 = (\tilde{x}_1^{(1)}, \tilde{x}_1^{(2)}, \dots, \tilde{x}_1^{(p)})^T, \tilde{X}_2 = (\tilde{x}_2^{(1)}, \tilde{x}_2^{(2)}, \dots, \tilde{x}_2^{(p)})^T, \dots, \tilde{X}_n = (\tilde{x}_n^{(1)}, \tilde{x}_n^{(2)}, \dots, \tilde{x}_n^{(p)})^T,$$

де

$$\tilde{x}_i^{(j)} = \frac{x_i^{(j)} - \bar{x}^{(j)}}{\hat{\sigma}^{(j)}}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p. \quad (2.3)$$

Тут $\bar{x}^{(j)} = \frac{1}{n} \sum_{k=1}^n x_k^{(j)}$ та $\hat{\sigma}^{(j)} = \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k^{(j)} - \bar{x}^{(j)})^2}$ – вибіркове середнє та вибіркове середньоквадратичне відхилення для j -го показника відповідно, $j = 1, 2, \dots, p$.

На основі значень $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ обчислюємо вибірккову коваріаційну матрицю показників $\Sigma = [\hat{\sigma}_{ij}]_{p \times p}$, кожний з елементів якої визначається за формулою:

$$\hat{\sigma}_{ij} = \frac{1}{n} \sum_{k=1}^n \tilde{x}_k^{(i)} \tilde{x}_k^{(j)}, \quad i, j = 1, 2, \dots, p. \quad (2.4)$$

Зауважимо, що оскільки величини $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(p)}$ є нормованими, то побудована вибіркова коваріаційна матриця (2.4) співпадає з вибірковою кореляційною матрицею, побудованою для цих же величин.

Метод компонентного аналізу дозволяє побудувати таке лінійне перетворення:

$$\begin{aligned} y^{(1)} &= l_{11} \tilde{x}^{(1)} + l_{12} \tilde{x}^{(2)} + \dots + l_{1p} \tilde{x}^{(p)}, \\ y^{(2)} &= l_{21} \tilde{x}^{(1)} + l_{22} \tilde{x}^{(2)} + \dots + l_{2p} \tilde{x}^{(p)}, \\ &\dots \dots \dots \dots \dots \dots \dots \dots \dots \\ y^{(p)} &= l_{p1} \tilde{x}^{(1)} + l_{p2} \tilde{x}^{(2)} + \dots + l_{pp} \tilde{x}^{(p)}, \end{aligned}$$

для якого

$$\begin{aligned} l_{j1}^2 + l_{j2}^2 + \dots + l_{jp}^2 &= 1, \quad j = 1, 2, \dots, p, \\ \text{cov}(y^{(i)}, y^{(j)}) &= 0, \quad \text{якщо } j < i, \\ Dy^{(1)} &\geq Dy^{(2)} \geq \dots \geq Dy^{(p)}. \end{aligned}$$

Вибіркові головні компоненти $y^{(1)}, y^{(2)}, \dots, y^{(p)}$ є новими узагальненими інтегральними показниками, що утворені як лінійні комбінації вихідних нормованих ознак $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(p)}$, які безпосередньо описують стан досліджуваної системи.

Метод головних компонент полягає в знаходженні послідовності ортогональних осей координат, уздовж яких кожен раз в порядку спадання визначається максимум повної дисперсії. Отже, вибіркові головні компоненти впорядковано за ступенем розсіювання. Першою вибірковою головною компонентою

$y^{(1)}$ є така нормована лінійна комбінація показників $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(p)}$, яка серед інших нормованих лінійних комбінацій показників $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(p)}$ має найбільшу дисперсію. Другою вибірковою головною компонентою $y^{(2)}$ є така нормована лінійна комбінація показників $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(p)}$, яка серед інших нормованих лінійних комбінацій показників $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \dots, \tilde{x}^{(p)}$, некорельованих з $y^{(1)}$, має найбільшу дисперсію. І так далі [1, 3].

Можна показати, що вектор $L^{(j)} = (l_{j1}, l_{j2}, \dots, l_{jp})^T$ є j -м нормованим власним вектором вибіркової коваріаційної матриці Σ , який відповідає j -му за величиною власному значенню λ_j .

Позначимо λ_j – j -й за величиною корінь характеристичного рівняння

$$\det(\Sigma - \lambda E) = 0,$$

де E – одинична матриця порядку p . Також позначимо $L^{(j)}$ – розв'язок однорідної системи

$$(\Sigma - \lambda_j E)L^{(j)} = 0, \quad j = 1, 2, \dots, p,$$

такий, що $l_{j1}^2 + l_{j2}^2 + \dots + l_{jp}^2 = 1$. Тоді

$$Dy^{(j)} = \lambda_j,$$

причому

$$D\tilde{x}^{(1)} + D\tilde{x}^{(2)} + \dots + D\tilde{x}^{(p)} = Dy^{(1)} + Dy^{(2)} + \dots + Dy^{(p)} = \lambda_1 + \lambda_2 + \dots + \lambda_p.$$

За таких умов внесок перших p' головних компонент ($1 \leq p' \leq p$) у зага-

льну дисперсію розраховується за формулою:

$$q(p') = \frac{Dy^{(1)} + Dy^{(2)} + \dots + Dy^{(p')}}{D\tilde{x}^{(1)} + D\tilde{x}^{(2)} + \dots + D\tilde{x}^{(p')}} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_{p'}}{\lambda_1 + \lambda_2 + \dots + \lambda_p}. \quad (2.5)$$

З цієї формули випливає, що обираючи певну кількість p' головних компонент з їх загальної кількості p , можна зменшити вимірність факторного простору (простору показників) задачі, зробивши його більш зручним для подальшого аналізу.

Отже, метод компонентного аналізу дозволяє отримати наочну інформацію про структуру досліджуваних об'єктів, що описуються вихідними спостереженнями $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$, заданими у p -вимірному просторі. Зокрема, за допомогою методу головних компонент ці спостереження можуть бути спроектовані на одновимірний, двовимірний або тривимірний підпростір, що описується головною компонентою $y^{(1)}$ (або головними компонентами $y^{(1)}, y^{(2)}$ чи $y^{(1)}, y^{(2)}, y^{(3)}$). Координати $(y_k^{(1)})$ ($(y_k^{(1)}, y_k^{(2)})$ чи $(y_k^{(1)}, y_k^{(2)}, y_k^{(3)})$), $k = 1, 2, \dots, n$, кожного з n досліджуваних об'єктів у новому підпросторі розраховуються за формулами:

$$\begin{aligned} y_k^{(1)} &= l_{11}\tilde{x}_k^{(1)} + l_{12}\tilde{x}_k^{(2)} + \dots + l_{1p}\tilde{x}_k^{(p)}, \\ y_k^{(2)} &= l_{21}\tilde{x}_k^{(1)} + l_{22}\tilde{x}_k^{(2)} + \dots + l_{2p}\tilde{x}_k^{(p)}, \\ y_k^{(3)} &= l_{31}\tilde{x}_k^{(1)} + l_{32}\tilde{x}_k^{(2)} + \dots + l_{3p}\tilde{x}_k^{(p)}, \quad k = 1, 2, \dots, n. \end{aligned} \quad (2.6)$$

Кількість головних компонент у даному поданні обирається виходячи з аналізу значення величини $q(p')$, що обчислюється за формулою (2.5) для різної кількості головних компонент p' .

Одинична дисперсія всіх змінних в компонентному аналізі розглядається в сукупності як загальна дисперсія. Хоча це призводить до однозначного

розв'язку, однак є нереальним майже у всіх практичних ситуаціях. Мабуть, ніколи не проводиться аналіз змінних, при якому мала б місце гіпотеза, що повна дисперсія дорівнює загальній, тобто що вона визначається тільки змінними, що спостерігаються. Лише в тому випадку, коли ця гіпотеза підтверджується критерієм, можна застосовувати модель компонентного аналізу.

2.3 Факторний аналіз

2.3.1 Основні положення факторного аналізу

Методи факторного аналізу виходять із загальної ідеї, відповідно до якої структура зв'язків між p аналізованими ознаками $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ може бути пояснена тим, що всі ці змінні залежать (лінійно або якимось іншим чином) від меншого числа інших, безпосередньо не вимірюваних («прихованих», «латентних») факторів $f^{(1)}, f^{(2)}, \dots, f^{(p')}$, $p' < p$, які прийнято називати загальними і які в більшості моделей конструюються так, щоб бути взаємно некорельованими [3, 4, 5].

При цьому в загальному випадку не постулюється можливість однозначного (детермінованого) відновлення значень кожного з спостережуваних ознак $x^{(j)}$ за відповідними значеннями загальних факторів $f^{(1)}, \dots, f^{(p')}$ (в припущенні, що ми їх вміємо обчислювати); допускається, що кожна з вихідних ознак $x^{(j)}$ залежить також від деякої своєї («специфічної» тільки для неї) залишкової випадкової компоненти $u^{(j)}$, яка і обумовлює статистичний характер зв'язку між $x^{(j)}$ з одного боку, та $f^{(1)}, \dots, f^{(p')}$ – з іншого.

Кінцева мета статистичного дослідження, проведеного з залученням апарату факторного аналізу, як правило, полягає у виявленні та інтерпретації латентних загальних факторів з одночасним протидіючим прагненням мінімізувати як їх число, так і ступінь залежності $x^{(j)}$ від своїх специфічних залишкових випадкових компонент $u^{(j)}$.

У певному сенсі шукані загальні чинники $f^{(1)}, \dots, f^{(p')}$ можна вважати причинами, а досліджувані ознаки $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ – наслідками. Прийнято вважати статистичний аналіз такого роду успішним, якщо велике число наслідків вдалося пояснити малим числом причин.

Таким чином, методи і моделі факторного аналізу націлені, так само як і метод головних компонент, на стиснення інформації або на зниження вимірності вихідного простору ознак.

Виникнення схеми і моделей факторного аналізу пов'язане з іменами Ч. Спірмена, Л. Терстоуна, Г. Томсона і розв'язанням задач психології на початку ХХ століття. Однак через специфічні наукові інтереси перших дослідників, які працювали в цій галузі, дослідження цього розділу багатовимірного статистичного аналізу довгий час залишалися практично неопрацьованими, а інтерпретації і аналізу різних моделей факторного аналізу була притаманна деяка невизначеність. Лише з середини 50-х років ХХ століття починають з'являтися цікаві результати саме ймовірнісно-статистичних досліджень цього апарату.

2.3.2 Алгоритм факторного аналізу

Розглянемо процедуру пошуку головних факторів [4]. Припустимо, що x_i мають багатовимірний нормальний розподіл, їх дисперсії та коваріації утворюють матрицю $C = [c_{ij}]_{p \times p}$. Прості фактори f_r є ортогональними (некорельованими). Тоді з рівняння

$$x_i = \sum_{r=1}^k l_{ir} f_r + e_i \quad (i = 1, 2, \dots, p), \quad (2.7)$$

отримаємо, що c_{ij} через навантаження та остаточні дисперсії розраховуються за формулами

$$c_{ii} = \sum_{r=1}^k l_{ij}^2 + v_i, \quad (2.8)$$

$$c_{ij} = \sum_{r=1}^k l_{ir} l_{jr} \quad (i \neq j).$$

В матричній формі це має вигляд

$$C = LL^T + V, \quad (2.9)$$

де $L = [l_{ir}]_{p \times k}$ – матриця навантажень;

V – діагональна матриця з елементами v_i .

Таким чином, основна модель полягає у виборі гіпотези H_0 відносно коваріаційної матриці C : вона може бути представлена у вигляді суми діагональної матриці з додатними елементами та матриці рангу k з додатними власними значеннями. Значення, яке візьмемо за k , повинно бути не занадто великим, оскільки гіпотеза буде тривіальною. За умови відомих v_i , можна було б обмежитися $k < p$, але в більш загальній ситуації, коли v_i не відомі, умови набувають вигляду $(p+k) < (p-k)^2$.

Нехай $A = [a_{ij}]$ – вибіркова коваріаційна матриця, елементами якої є вибіркові оцінки дисперсії та коваріацій x_i з n ступенями свободи (відповідна вибірці розміру $n+1$). Використовуючи інформацію, яку містить матриця A , побудуємо ефективні оцінки параметрів l_{ir} та v_i , вважаючи їх невідомими.

Для оцінки невідомих параметрів застосуємо метод максимальної правдоподібності. Оскільки x_i розподілені нормально, то a_{ij} мають розподіл Уїшарта, а функція максимальної правдоподібності має вигляд

$$L = -\frac{1}{2} n \ln |C| - \frac{1}{2} n \sum_{i,j} a_{ij} c^{ij}, \quad (2.10)$$

де c^{ij} – елемент i -го рядка та j -го стовпця матриці C^{-1} . Сума в (2.10) може бути записана як $tr(AC^{-1})$.

Максимізуємо співвідношення (2.10) по l_{ir} і v_i . У (2.7) фактори можуть бути замінені будь-яким їх ортогональним перетворенням. Це відобразиться на навантаженнях за рахунок множення L справа на ортогональну матрицю розміру $k \times k$. При будь-якому такому множенні матриця LL^T залишиться незмінною. Це означає, що метод максимальної правдоподібності приводить до єдиних оцінок c_{ij} , але для оцінок l_{ir} веде до рівнянь, яким задовольняє нескінченна кількість розв'язків, однаково добрих зі статистичних міркувань.

Виберемо L так, щоб матриця $J = L^T V^{-1} L$ розміру $k \times k$ була діагональною. Для максимізації L прирівняємо до нуля частинні похідні по l_{ir} і v_i . Щоб отримати їх, зауважимо, що частинна похідна від $\ln|C|$ по l_{ir} дорівнює

$$2 \sum_j l_{jr} C_{ij} / |C| = 2 \sum_j l_{jr} c^{ji},$$

де C_{ij} є кофактором c_{ij} в $|C|$.

Частинна похідна $\ln|C|$ по v_i дорівнює $C_{ii} / |C| = c^{ii}$.

Сума в (2.10) може бути записана як

$$\sum_{u,w} a_{uw} C_{uw} / |C|.$$

Частинна похідна від цього співвідношення по l_{ir} має вигляд

$$2 \sum_{u,w,j} a_{uw} l_{jr} C_{uw,ij} / |C| - \sum_{u,w} a_{uw} C_{uw} \cdot 2 \sum_i l_{jr} c_{ij} / |C|^2,$$

де у першій сумі $u \neq i$, $w \neq j$ і $C_{uw,ij}$ позначає кофактор c_{ij} в C_{uw} . Скориставшись

рівністю

$$C_{uw,ij} |C| = C_{uw} C_{ij} - C_{uj} C_{iw}$$

приведемо це співвідношення до вигляду

$$-2 \sum_{u,w,j} a_{uw} l_{jr} C_{uj} C_{iw} / |C|^2 = -2 \sum_{j,w,u} l_{jr} c^{ju} a_{uw} c^{wi}.$$

Частинна похідна суми в (2.10) по v_i дорівнює

$$-\sum_{u,w} c^{iu} a_{uw} c^{wi}.$$

За таких умов

$$\frac{\partial L}{\partial l_{ir}} = \sum_j l_{jr} c^{ji} - \sum_{j,u,w} l_{jr} c^{ju} a_{uw} c^{wi},$$

що є елементом r -го рядка та i -го стовпця матриці

$$L^T C^{-1} - L^T C^{-1} A C^{-1},$$

помноженого на $-n$, в той час як

$$\frac{\partial L}{\partial v_i} = c^{ii} - \sum_{u,w} c^{iu} a_{uw} c^{wi},$$

що є i -м діагональним елементом матриці

$$C^{-1} - C^{-1}AC^{-1},$$

помноженої на $-n/2$.

За таких умов оцінки $\hat{L} = [\hat{l}_{ir}]$ матриці навантажень L , \hat{V} – матриці остаточної дисперсій, $\hat{C} = [\hat{c}_{ij}]$ – коваріаційної матриці задаються співвідношеннями

$$\begin{aligned}\hat{L}^T \hat{C}^{-1} - \hat{L}^T \hat{C}^{-1} A \hat{C}^{-1} &= 0, \\ \text{diag}(\hat{C}^{-1} - \hat{C}^{-1} A \hat{C}^{-1}) &= 0, \\ \hat{C} &= \hat{L} \hat{L}^T + \hat{V},\end{aligned}$$

де $\text{diag}(A)$ позначає матрицю, що містить лише діагональні елементи матриці A .

Матриця

$$\hat{J} = \hat{L}^T \hat{V}^{-1} \hat{L}$$

має бути діагональною. У подальшому знак кришки опускаємо, тобто вважатимемо, що матриці L , V , C та окремі їх елементи є оціненими значеннями.

Домноживши першу в останній групі рівностей справа на C , дістанемо

$$L^T - L^T C^{-1} A = 0.$$

Домноживши другу рівність зліва на $V = C - LL^T$ і використавши останнє співвідношення, дістанемо

$$\text{diag}(I - AC^{-1}) = 0.$$

Домноживши це співвідношення справа на V , з урахуванням передостаннього співвідношення дістанемо

$$\text{diag}(C - A) = 0.$$

Тоді $c_{ii} = a_{ii}$ для всіх i або

$$v_i = a_{ii} - \sum_{r=1}^k l_{ir}^2, \quad i = 1, 2, \dots, n. \quad (2.11)$$

Використаємо очевидну тотожність

$$C^{-1} = V^{-1} - V^{-1}L(I + J)^{-1}L^T V^{-1}.$$

Її множення зліва на L^T дає тотожність

$$L^T C^{-1} = (I + J)^{-1} L^T V^{-1}. \quad (2.12)$$

Підставляючи цю тотожність у $L^T - L^T C^{-1} A = 0$, дістанемо

$$L^T = (I + J)^{-1} L^T V^{-1} A,$$

або

$$L^T = J^{-1} L^T V^{-1} (A - V), \quad (2.13)$$

З останнього рівняння випливає, що матриця

$$H = L^T V^{-1} (A - V) V^{-1} L$$

є діагональною, оскільки $H = J^2$. Елементи J або квадратні корні елементів H є власними значеннями для $V^{-1}(A - V)$. Знайдемо перші k значень у порядку

спадання їх величин. Рядки L^T є відповідними власними вектор-рядками. Позначимо r -й рядок матриці $L_{(1)}^T$ через $l_{r(1)}^T$, де $L_{(1)}$ та $V_{(1)}$ – є наближеннями матриць L та V . Обчислимо спочатку вектор-рядки

$$w_1^T = l_{1(1)}^T V_{(1)}^{-1},$$

$$u_1^T = w_1^T A - l_{1(1)}^T$$

і додатне число $h_1 = u_1^T w_1$. Краще наближення до l_1^T задається рівнянням

$$l_{1(2)}^T = \frac{1}{\sqrt{h_1}} u_1^T.$$

Якщо є другий простий фактор, тобто $k \geq 2$, то необхідно послідовно обчислити вектор-рядок

$$w_2^T = l_{2(1)}^T V_{(1)}^{-1},$$

число

$$j_{21} = w_2^T l_{1(2)},$$

вектор-рядок

$$u_2^T = w_2^T A - l_{2(1)}^T - j_{21} l_{1(2)}^T$$

та число

$$h_2 = u_2^T w_2.$$

Кращим наближенням до l_2^T буде

$$l_{2(2)}^T = \frac{1}{\sqrt{h_2}} u_2^T.$$

Для третього фактору послідовно визначимо:

$$\begin{aligned} w_3^T &= l_{3(1)}^T V_{(1)}^{-1}, \\ j_{31} &= w_3^T l_{1(2)}, \quad j_{32} = w_3^T l_{2(2)}, \\ u_3^T &= w_3^T A - l_{3(1)}^T - j_{31} l_{1(2)}^T - j_{32} l_{2(2)}^T, \\ h_3 &= u_3^T w_3. \end{aligned}$$

Кращим наближенням до l_3^T буде

$$l_{3(2)}^T = \frac{1}{\sqrt{h_3}} u_3^T.$$

Аналогічно продовжують для будь-якого значення k .

Кращі оцінки для v_i , які утворюють матрицю $V_{(2)}$, визначаються з рівняння (2.11). Нові матриці оцінок $L_{(2)}$ і $V_{(2)}$ використовуються як початкові на новій ітерації.

Ітераційний процес триває до тих пір, поки не матиме місця умова

$$\max_{\substack{1 \leq i \leq p \\ 1 \leq j \leq m}} |l_{ij(k+1)} - l_{ij(k)}| < \delta$$

де $\delta > 0$ – задана точність.

Після того, як будуть визначені оцінки матриць L та V , можна обчислити

оцінки факторів для кожного i -го об'єкту з n досліджуваних об'єктів:

$$\vec{f}_i = (L^T V^{-1} L)^{-1} L^T V^{-1} \vec{x}_i, \quad t = 1, 2, \dots, n.$$

Дані вектори є координатами розташування досліджуваних об'єктів у просторі \mathbf{R}^k [1].

2.4 Кластерний аналіз

Дуже часто при дослідженні великих сукупностей об'єктів необхідно виявити об'єкти, близькі між собою в певному сенсі.

Метою методів класифікації є дослідження внутрішньої структури досліджуваної системи n об'єктів, кожен з яких описується m ознаками; «стиснення» вихідної інформації без істотних її втрат шляхом виявлення класів схожих між собою об'єктів і ототожнення об'єктів всередині кожного класу. Кожен об'єкт у класі схожий (однорідний) один на одного, а об'єкти між класами відрізняються (неоднорідні) один від одного. Кількісне оцінювання подібності між об'єктами базується на відстані: більш подібні об'єкти мають меншу відстань і навпаки [2, 3].

Основними серед задач, що вирішуються методами класифікації, є:

- проведення класифікації об'єктів, описуваних великою кількістю різноманітних показників;
- виявлення структури досліджуваної сукупності об'єктів;
- зниження обсягу вибірки шляхом ототожнення кожного класу об'єктів з його типовим представником;
- зниження вимірності простору шляхом ототожнення близьких ознак (в цьому випадку класифікуються не об'єкти, а ознаки);
- побудова окремих регресійних моделей в кожному класі об'єктів.

Практика показує, що класифікація об'єктів, проведена за факторами або

головними компонентами, розрахованими в результаті зниження вимірності вихідного простору ознак, є більш об'єктивною, ніж класифікація тих же об'єктів за вихідними ознаками.

Кластерний аналіз – важливий інструмент у дослідженні структури даних [8]. Методи кластерного аналізу дозволяють розбити досліджувану сукупність об'єктів на групи однорідних в певному сенсі, що мають назву кластерів або класів. Найбільшого поширення отримали два підходи до задачі класифікації:

- евристичний, який реалізує деяку схему поділу об'єктів на класи, виходячи з інтуїтивних міркувань;
- екстремальний, який реалізує схему розподілу на основі заданого критерію оптимальності. Найбільш важким в задачі класифікації є визначення міри однорідності об'єктів.

Нехай кожен з вихідних n об'єктів з номерами $1, 2, \dots, n$ задається як точка $x_k = (x_{k1}, x_{k2}, \dots, x_{km})$, $k = 1, 2, \dots, n$, в m -вимірному просторі ознак X_1, X_2, \dots, X_m . Сукупність цих точок можна трактувати як вибірку обсягу n з багатовимірної генеральної сукупності.

Зауважимо, що результати класифікації залежатимуть від вибору масштабу і одиниць вимірювання ознак. Щоб уникнути цього, від вихідних ознак стандартно переходять до їх нормованих значень.

У разі залежних ознак та їх різної значущості при класифікації об'єктів за міру однорідності об'єктів обирають відстань Махалонобіса:

$$l_{ij} = \sqrt{(x_i - x_j) \Lambda \Sigma^{-1} (x_i - x_j)^T}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, n,$$

де $\Lambda \in \mathbf{R}^{m \times m}$ – симетрична (найчастіше діагональна) невід'ємно визначена матриця «вагових» коефіцієнтів ознак, Σ – коваріаційна матриця генеральної сукупності, з якої обираються об'єкти. Окремими випадками цієї формули є евклідова відстань:

$$l_{ij} = \sqrt{(x_i - x_j)(x_i - x_j)^T} = \sum_{k=1}^m (x_{ik} - x_{jk})^2, i = 1, 2, \dots, n, j = 1, 2, \dots, n,$$

використання якої виправдано, якщо генеральна сукупність розподілена за багатовимірним нормальним законом з коваріаційною матрицею $\Sigma = \sigma^2 I$, а ознаки однорідні за своїм фізичним змістом і однаково «вагомі» з точки зору вирішення питання про віднесення об'єкта до того чи іншого кластеру.

Ієрархічні і паралельні кластер-процедури є практично реалізуємі лише в задачах класифікації з кількома десятками спостережень. Для розв'язання задачі з більшою кількістю спостережень застосовують послідовні кластер-процедури – ітераційні алгоритми, на кожному кроці яких оброблюється одне спостереження (або невелика частина результативних спостережень) і результати розбиття на попередньому кроці.

Ідею цих процедур пояснимо на прикладі методу k -середніх (k -Means Clustering) з наперед заданим числом k класів:

- на нульовому кроці за центри шуканих k кластерів приймають випадкові вибрані k спостережень – точки x_1, x_2, \dots, x_k , кожному кластеру назначають одиничну вагу;

- на першому кроці знаходять відстані від точки x_{k+1} до центрів кластерів, побудованих на попередньому кроці, і точку x_{k+1} відносять до кластеру, відстань до якого мінімальна. Після цього розраховують новий центр ваги цього кластера (як зважене середнє по кожному показнику) і вагу кластера збільшують на одиницю. Всі інші кластери залишаються незмінними (з колишніми центрами і вагами);

- на другому етапі аналогічну процедуру виконують для точки x_{k+2} й так далі.

При досить великому числі n об'єктів, що підлягають класифікації, або достатньо великій кількості ітерацій перерахунок центрів тяжіння практично не призводить до їх зміни.

Якщо в якійсь точці не вдається, розглянувши всі $n = k + (n - k)$ точок,

досягти практично не змінюваних центрів тяжіння, то, використовуючи отримані розбиття n точок на k кластерів в якості початкового, застосовують викладену процедуру k точкам x_1, x_2, \dots, x_k , або в якості початкового розбиття приймають різні комбінації k точок з вихідних n , а в якості остаточного результату беруть найбільш часто повторювані фінальні розбиття.

2.5 Алгоритм розв'язання задачі

2.5.1 Алгоритм розв'язання задачі дослідження динаміки розвитку економічних систем у часі

Наведемо послідовність дій обраних методів зниження вимірності для розв'язання поставленої у атестаційній роботі задачі аналізу складної системи, що складається з m об'єктів, і кожен з яких описується p показниками.

Етап 1 (підготовчий).

Крок 1. Віднормувати вихідні дані $X_1 = (x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(p)})^T, \dots, X_m = (x_m^{(1)}, x_m^{(2)}, \dots, x_m^{(p)})^T$.

Крок 2. За віднормованими спостереженнями $\tilde{X}_1 = (\tilde{x}_1^{(1)}, \tilde{x}_1^{(2)}, \dots, \tilde{x}_1^{(p)})^T, \dots, \tilde{X}_m = (\tilde{x}_m^{(1)}, \tilde{x}_m^{(2)}, \dots, \tilde{x}_m^{(p)})^T$ обчислити вибірккову коваріаційну матрицю показників $\Sigma = [\hat{\sigma}_{ij}]_{p \times p}$.

Крок 3. Знайти власні значення $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ та відповідні їм нормовані власні вектори $\mathbf{L}^{(j)} = (l_{j1}, l_{j2}, \dots, l_{jp})^T, j = 1, 2, \dots, p$, коваріаційної матриці Σ .

Етап 2 (метод головних компонент).

Крок 1. Обрати кількість головних компонент p' ($p' = 1, 2, 3$), аналізуючи величину

$$q(p') = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_{p'}}{\lambda_1 + \lambda_2 + \dots + \lambda_p}.$$

Крок 2. Для кожного з об'єктів, що підлягають класифікації, обчислити

$$y_k^{(s)} = l_{s1} \tilde{x}_k^{(1)} + l_{s2} \tilde{x}_k^{(2)} + \dots + l_{sp} \tilde{x}_k^{(p)}, \quad k = 1, 2, \dots, m, \quad s = 1, \dots, p'.$$

Крок 3. У просторі $\mathbb{R}^{p'}$ із застосуванням процедури кластеризації (методи k -середніх та k -медіан) побудувати точки з координатами, обчисленими у попередньому пункті, що геометрично відображають досліджувані об'єкти.

Етап 3 (метод факторного аналізу).

Крок 1. Задавши початкові наближення для матриць L та V для обраної кількості k факторів з використанням ітераційної процедури, описаної у п. 2.3.2, обчислити з заданою точністю оцінки векторів l_r^T , $r = \overline{1, k}$, що є рядками шуканої матриці навантажень L .

Крок 2. Обчислити оцінки факторів для кожного i -го об'єкту з n досліджуваних об'єктів:

$$\vec{f}_i = (L^T V^{-1} L)^{-1} L^T V^{-1} \vec{x}_i, \quad t = 1, 2, \dots, n.$$

Крок 3. У просторі \mathbf{R}^k із застосуванням процедури кластеризації (методи k -середніх та k -медіан) побудувати точки з координатами, обчисленими у попередньому пункті, що геометрично відображають досліджувані об'єкти.

2.5.2 Аналіз початкових даних

Розглянемо застосування методу головних компонент та методу факторного аналізу для аналізу економічного стану регіонів України. Для опису соціа-

льно-економічного положення кожного з регіонів було використано наступні параметри:

$x^{(1)}$ – валовий регіональний продукт, у фактичних цінах, млн грн.;

$x^{(2)}$ – валовий регіональний продукт у розрахунку на одну особу, у фактичних цінах, грн.;

$x^{(3)}$ – середньомісячна номінальна заробітна плата працівника за регіонами, грн.;

$x^{(4)}$ – індекс споживчих цін за регіонами, грудень до грудня попереднього року %;

$x^{(5)}$ – індекс промислової продукції за регіонами, до попереднього року %;

$x^{(6)}$ – індекс сільськогосподарської продукції за регіонами, до попереднього року %;

$x^{(7)}$ – індекси будівельної продукції за регіонами, до попереднього року %;

$x^{(8)}$ – обсяг експорту товарів за регіонами України, вартість, млн. дол. США;

$x^{(9)}$ – обсяг імпорту товарів за регіонами України, вартість, млн. дол. США;

$x^{(10)}$ – оптовий товарооборот за регіонами, млн. грн.;

$x^{(11)}$ – роздрібний товарооборот за регіонами, млн. грн.;

$x^{(12)}$ – роздрібний товарооборот на одну особу за регіонами, грн.;

$x^{(13)}$ – кількість зареєстрованих безробітних за регіонами, на кінець року тис. осіб;

$x^{(14)}$ – рівень зареєстрованого безробіття за регіонами, на кінець року %;

$x^{(15)}$ – чисельність наявного населення за регіонами, на кінець року тис. осіб.

Оскільки маємо велику кількість параметрів, які характеризують кожен із регіонів, причому деякі з них можуть бути похідними або вкладеними в інші, то безпосередня класифікація за всією сукупністю цих параметрів є надзвичайно ускладненою. Тому для розв'язання поставленої задачі класифікації застосуємо метод головних компонент та метод факторного аналізу. Значення обраних показників для регіонів України взято з сайтів Державної служби статистики

України [9] та статистичних щорічників за 2000-2015 рр.

Позначимо n – кількість регіонів (24 області України, м. Київ, до 2014 року м. Севастополь та Автономна республіка Крим), p – кількість параметрів, за якими проводиться аналіз.

Оскільки обрані показники мають різні одиниці виміру (долари США, грн., відсотки тощо), то для застосування методів зниження розмірності необхідно провести нормування їх значень за формулою (2.3). Вихідна інформація за регіонами за 2000, 2005, 2010, 2015 роки та відповідні нормовані значення наведено у додатку Б.

2.5.3 Застосування методів багатовимірного статистичного аналізу для розв'язання задачі дослідження розвитку економічних систем у часі

На основі нормованих даних за 2000 рік, використовуючи формулу (2.4) будуємо коваріаційну матрицю Σ :

$$\begin{pmatrix} 1. & 0.83 & 0.82 & 0.13 & -0.12 & -0.17 & 0.3 & 0.94 & 0.77 & 0.92 & 0.93 & 0.6 & 0.48 & -0.09 & 0.91 \\ 0.83 & 1. & 0.89 & 0.13 & -0.12 & -0.39 & 0.35 & 0.73 & 0.87 & 0.82 & 0.86 & 0.81 & 0.13 & -0.07 & 0.56 \\ 0.82 & 0.89 & 1. & 0.3 & 0.02 & -0.59 & 0.13 & 0.76 & 0.82 & 0.81 & 0.83 & 0.76 & 0.05 & -0.12 & 0.58 \\ 0.13 & 0.13 & 0.3 & 1. & 0.09 & -0.34 & -0.23 & 0.21 & 0.12 & 0.2 & 0.03 & -0.02 & -0.22 & 0.02 & 0.06 \\ -0.12 & -0.12 & 0.02 & 0.09 & 1. & -0.38 & -0.26 & -0.02 & 0.06 & 0.11 & -0.02 & 0.17 & -0.15 & -0.12 & -0.2 \\ -0.17 & -0.39 & -0.59 & -0.34 & -0.38 & 1. & 0.11 & -0.18 & -0.56 & -0.38 & -0.32 & -0.62 & 0.49 & 0.14 & 0.13 \\ 0.3 & 0.35 & 0.13 & -0.23 & -0.26 & 0.11 & 1. & 0.17 & 0.45 & 0.3 & 0.39 & 0.37 & 0.15 & -0.08 & 0.23 \\ 0.94 & 0.73 & 0.76 & 0.21 & -0.02 & -0.18 & 0.17 & 1. & 0.66 & 0.9 & 0.8 & 0.46 & 0.47 & -0.1 & 0.85 \\ 0.77 & 0.87 & 0.82 & 0.12 & 0.06 & -0.56 & 0.45 & 0.66 & 1. & 0.88 & 0.9 & 0.9 & 0.01 & -0.08 & 0.49 \\ 0.92 & 0.82 & 0.81 & 0.2 & 0.11 & -0.38 & 0.3 & 0.9 & 0.88 & 1. & 0.92 & 0.72 & 0.31 & -0.1 & 0.74 \\ 0.93 & 0.86 & 0.83 & 0.03 & -0.02 & -0.32 & 0.39 & 0.8 & 0.9 & 0.92 & 1. & 0.81 & 0.36 & -0.1 & 0.78 \\ 0.6 & 0.81 & 0.76 & -0.02 & 0.17 & -0.62 & 0.37 & 0.46 & 0.9 & 0.72 & 0.81 & 1. & -0.06 & -0.08 & 0.28 \\ 0.48 & 0.13 & 0.05 & -0.22 & -0.15 & 0.49 & 0.15 & 0.47 & 0.01 & 0.31 & 0.36 & -0.06 & 1. & -0.08 & 0.69 \\ -0.09 & -0.07 & -0.12 & 0.02 & -0.12 & 0.14 & -0.08 & -0.1 & -0.08 & -0.1 & -0.1 & -0.08 & -0.08 & 1. & -0.08 \\ 0.91 & 0.56 & 0.58 & 0.06 & -0.2 & 0.13 & 0.23 & 0.85 & 0.49 & 0.74 & 0.78 & 0.28 & 0.69 & -0.08 & 1. \end{pmatrix}.$$

Відповідно до методу головних компонент обчислимо власні значення наведеної коваріаційної матриці та розташуємо їх у порядку спадання:

$$\lambda_1 = 7,86499, \lambda_2 = 2,81086, \lambda_3 = 1,52236, \lambda_4 = 1,10241, \lambda_5 = 0,61249,$$

$$\lambda_6 = 0,41995, \lambda_7 = 0,22250, \lambda_8 = 0,18747, \lambda_9 = 0,15124, \lambda_{10} = 0,04341, \\ \lambda_{11} = 0,03063, \lambda_{12} = 0,01414, \lambda_{13} = 0,01285, \lambda_{14} = 0,00294, \lambda_{15} = 0,00177.$$

Оскільки найбільш наочною є візуалізація у двовимірному просторі, то для класифікації використовуватимемо дві головні компоненти, що дають

$$q(2) = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \dots + \lambda_{18}} = 0,7114 = 71,14\%$$

внеску у загальну дисперсію. Власні вектори, що відповідають першим двом максимальним власним значенням $\lambda_1 = 7,86499$, $\lambda_2 = 2,81086$, дорівнюють:

$$L^{(1)} = (0,34; 0,33; 0,33; 0,07; 0,0; -0,16; 0,12; 0,31; \\ 0,33; 0,34; 0,34; 0,28; 0,09; -0,18; 0,27)^T; \\ L^{(2)} = (-0,16; 0,02; 0,14; 0,27; 0,26; -0,48; -0,2; -0,14; \\ 0,1; -0,03; -0,08; 0,16; -0,51; -0,36; -0,31)^T.$$

Значення перших двох головних компонент, розраховані за формулою (2.6), мають вигляд:

$$y^{(1)} = 0,34x_1 + 0,33x_2 + 0,33x_3 + 0,07x_4 + 0,0x_5 - 0,16x_6 + \\ + 0,12x_7 + 0,31x_8 + 0,33x_9 + 0,34x_{10} + 0,34x_{11} + 0,28x_{12} + \\ + 0,09x_{13} - 0,18x_{14} + 0,27x_{15}; \quad (2.15)$$

$$y^{(2)} = -0,16x_1 + 0,02x_2 + 0,14x_3 + 0,27x_4 + 0,26x_5 - 0,48x_6 - \\ - 0,02x_7 - 0,14x_8 + 0,1x_9 - 0,03x_{10} - 0,08x_{11} + 0,16x_{12} - \\ - 0,51x_{13} - 0,36x_{14} - 0,31x_{15}. \quad (2.16)$$

Далі розглянемо розв'язання даної задачі методом факторного аналізу. Використовуючи коваріаційну матрицю, отриману з нормованих даних для

2000 року, розраховуємо початкові наближення методу. Початкова матриця факторних навантажень:

$$L_{(0)} = \begin{pmatrix} 0,337 & 0,325 & 0,328 & 0,069 & -0,001 & -0,156 & 0,118 & 0,310 \\ -0,161 & 0,015 & 0,139 & 0,27 & 0,257 & -0,479 & -0,199 & -0,142 \\ 0,327 & 0,34 & 0,344 & 0,285 & 0,088 & -0,183 & 0,265 \\ 0,101 & -0,033 & -0,083 & 0,164 & -0,507 & -0,357 & -0,311 \end{pmatrix}.$$

Використовуючи $L_{(0)}$, розраховуємо $V_{(0)}$ – початкове наближення для матриці залишкових дисперсій

$$V_{(0)} = \text{diag} \{0,861 \quad 0,894 \quad 0,873 \quad 0,922 \quad 0,934 \quad 0,746 \quad 0,946 \quad 0,884 \\ 0,883 \quad 0,883 \quad 0,875 \quad 0,892 \quad 0,735 \quad 0,839 \quad 0,833\}.$$

Проводимо ітераційну процедуру відповідно до алгоритму факторного аналізу та за допомогою математичного пакету Wolfram Mathematica отримуємо (за 33 ітерації) оцінку матриці навантажень та матриці залишкових дисперсій з точністю $\delta = 10^{-3}$:

$$L = \begin{pmatrix} 0,998 & 0,844 & 0,829 & 0,131 & -0,104 & -0,196 & 0,306 & 0,934 \\ -0,034 & 0,379 & 0,356 & 0,043 & 0,286 & -0,732 & 0,203 & -0,123 \\ 0,789 & 0,931 & 0,940 & 0,626 & 0,460 & -0,401 & 0,893 \\ 0,566 & 0,211 & 0,248 & 0,723 & -0,596 & -0,284 & -0,389 \end{pmatrix},$$

$$V = \text{diag} \{0,002 \quad 0,144 \quad 0,186 \quad 0,981 \quad 0,907 \quad 0,425 \quad 0,865 \quad 0,112 \\ 0,056 \quad 0,089 \quad 0,054 \quad 0,085 \quad 0,434 \quad 0,759 \quad 0,05\}.$$

Далі наведено результати застосування методів зниження вимірності за статистичними даними 2005 року. Коваріційна матриця Σ нормованих даних

має вигляд:

$$\begin{pmatrix} 1. & 0.9 & 0.9 & 0.03 & -0.19 & -0.33 & 0.21 & 0.86 & 0.87 & 0.95 & 0.95 & 0.76 & 0.09 & -0.64 & 0.8 \\ 0.9 & 1. & 0.93 & 0.12 & -0.02 & -0.56 & 0.24 & 0.64 & 0.96 & 0.95 & 0.91 & 0.9 & -0.23 & -0.63 & 0.5 \\ 0.9 & 0.93 & 1. & 0.17 & -0.12 & -0.51 & 0.21 & 0.73 & 0.87 & 0.9 & 0.88 & 0.82 & -0.18 & -0.77 & 0.61 \\ 0.03 & 0.12 & 0.17 & 1. & -0.04 & 0.14 & 0.09 & 0.11 & -0.02 & -0.06 & -0.12 & -0.1 & -0.09 & -0.03 & -0.05 \\ -0.19 & -0.02 & -0.12 & -0.04 & 1. & -0.27 & 0. & -0.36 & 0.06 & -0.03 & -0.09 & 0.11 & -0.45 & 0.04 & -0.41 \\ -0.33 & -0.56 & -0.51 & 0.14 & -0.27 & 1. & -0.16 & -0.07 & -0.59 & -0.5 & -0.43 & -0.7 & 0.74 & 0.54 & 0.15 \\ 0.21 & 0.24 & 0.21 & 0.09 & 0. & -0.16 & 1. & 0.17 & 0.22 & 0.23 & 0.17 & 0.15 & -0.14 & -0.13 & 0.08 \\ 0.86 & 0.64 & 0.73 & 0.11 & -0.36 & -0.07 & 0.17 & 1. & 0.54 & 0.72 & 0.68 & 0.4 & 0.31 & -0.51 & 0.84 \\ 0.87 & 0.96 & 0.87 & -0.02 & 0.06 & -0.59 & 0.22 & 0.54 & 1. & 0.96 & 0.93 & 0.95 & -0.29 & -0.58 & 0.44 \\ 0.95 & 0.95 & 0.9 & -0.06 & -0.03 & -0.5 & 0.23 & 0.72 & 0.96 & 1. & 0.96 & 0.88 & -0.13 & -0.62 & 0.62 \\ 0.95 & 0.91 & 0.88 & -0.12 & -0.09 & -0.43 & 0.17 & 0.68 & 0.93 & 0.96 & 1. & 0.89 & -0.01 & -0.65 & 0.7 \\ 0.76 & 0.9 & 0.82 & -0.1 & 0.11 & -0.7 & 0.15 & 0.4 & 0.95 & 0.88 & 0.89 & 1. & -0.41 & -0.64 & 0.31 \\ 0.09 & -0.23 & -0.18 & -0.09 & -0.45 & 0.74 & -0.14 & 0.31 & -0.29 & -0.13 & -0.01 & -0.41 & 1. & 0.34 & 0.56 \\ -0.64 & -0.63 & -0.77 & -0.03 & 0.04 & 0.54 & -0.13 & -0.51 & -0.58 & -0.62 & -0.65 & -0.64 & 0.34 & 1. & -0.51 \\ 0.8 & 0.5 & 0.61 & -0.05 & -0.41 & 0.15 & 0.08 & 0.84 & 0.44 & 0.62 & 0.7 & 0.31 & 0.56 & -0.51 & 1. \end{pmatrix}$$

Відповідно до методу компонентного аналізу визначаємо власні значення наведеної коваріаційної матриці, розташувавши їх у порядку спадання:

$$\begin{aligned} \lambda_1 &= 8,2252, \lambda_2 = 2,8203, \lambda_3 = 1,1839, \lambda_4 = 0,9068, \lambda_5 = 0,7451, \lambda_6 = 0,5755, \\ \lambda_7 &= 0,2616, \lambda_8 = 0,1401, \lambda_9 = 0,06097, \lambda_{10} = 0,03998, \lambda_{11} = 0,02566, \\ \lambda_{12} &= 0,007246, \lambda_{13} = 0,00562, \lambda_{14} = 0,00121, \lambda_{15} = 0,000597. \end{aligned}$$

Дві головні компоненти дають 73,63 % внеску у загальну дисперсію. Власні вектори, що відповідають першим двом максимальним власним значенням $\lambda_1 = 8,2252$, $\lambda_2 = 2,8203$, дорівнюють:

$$\mathbf{L}^{(1)} = (0,33; 0,33; 0,33; 0,0; -0,03; -0,19; 0,08; 0,26; 0,33; 0,34; 0,34; 0,31; -0,06; -0,26; 0,23)^T;$$

$$\mathbf{L}^{(2)} = (-0,15; 0,06; 0,0; -0,03; 0,39; -0,43; 0,05; -0,31; 0,11; 0,0; -0,05; 0,2; -0,55; -0,08; -0,42)^T.$$

Розраховані за формулою (2.6) значення двох головних компонент мають вигляд:

$$\begin{aligned}
 y^{(1)} = & 0,33x_1 + 0,33x_2 + 0,33x_3 + 0,0x_4 - 0,03x_5 - 0,19x_6 + \\
 & + 0,08x_7 + 0,26x_8 + 0,33x_9 + 0,34x_{10} + 0,34x_{11} + 0,31x_{12} - \\
 & - 0,06x_{13} - 0,26x_{14} + 0,23x_{15};
 \end{aligned} \tag{2.17}$$

$$\begin{aligned}
 y^{(2)} = & -0,15x_1 + 0,06x_2 + 0,0x_3 - 0,03x_4 + 0,39x_5 - 0,49x_6 + \\
 & + 0,05x_7 - 0,31x_8 + 0,11x_9 + 0,0x_{10} - 0,05x_{11} + 0,2x_{12} - \\
 & - 0,55x_{13} - 0,08x_{14} - 0,42x_{15}.
 \end{aligned} \tag{2.18}$$

Далі наведемо обчислення за методом факторного аналізу за даними 2005 року. Початкова матриця факторних навантажень:

$$L_{(0)} = \begin{pmatrix} 0,334 & 0,335 & 0,334 & 0,003 & -0,031 & -0,189 & 0,084 & 0,258 \\ -0,15 & 0,061 & -0,004 & -0,026 & 0,389 & -0,426 & 0,055 & -0,314 \\ 0,327 & 0,34 & 0,335 & 0,31 & -0,055 & -0,26 & 0,226 & \\ 0,114 & 0,001 & -0,05 & 0,204 & -0,547 & -0,076 & -0,419 & \end{pmatrix}.$$

Використовуючи $L_{(0)}$, розрахуємо $V_{(0)}$ – початкове наближення для матриці залишкових дисперсій:

$$\begin{aligned}
 V_{(0)} = & \text{diag} \{0,866 \quad 0,884 \quad 0,888 \quad 0,999 \quad 0,848 \quad 0,783 \quad 0,989 \quad 0,835 \\
 & 0,879 \quad 0,884 \quad 0,885 \quad 0,862 \quad 0,698 \quad 0,926 \quad 0,772\}.
 \end{aligned}$$

Проводимо ітераційну процедуру відповідно до алгоритму факторного аналізу та за допомогою математичного пакету Wolfram Mathematica отримуємо (за 48 ітерацій) оцінку матриці факторного аналізу та матриці залишкових дисперсій з точністю $\delta = 10^{-3}$:

$$L = \begin{pmatrix} 0,998 & 0,920 & 0,913 & 0,017 & -0,169 & -0,364 & 0,214 & 0,832 \\ -0,053 & 0,318 & 0,157 & -0,079 & 0,488 & -0,675 & 0,074 & -0,424 \end{pmatrix}$$

$$\left. \begin{array}{cccccc} 0,891 & 0,966 & 0,959 & 0,789 & 0,048 & -0,646 & 0,775 \\ 0,438 & 0,222 & 0,163 & 0,558 & -0,810 & -0,096 & -0,574 \end{array} \right\},$$

$$V = \text{diag} \{ 0,001 \quad 0,052 \quad 0,141 \quad 0,993 \quad 0,734 \quad 0,411 \quad 0,949 \quad 0,128 \\ 0,015 \quad 0,017 \quad 0,054 \quad 0,066 \quad 0,341 \quad 0,573 \quad 0,069 \}.$$

Далі наведено результати застосування методів зниження вимірності за статистичними даними 2010 року. Коваріаційна матриця Σ нормованих даних має вигляд:

$$\left(\begin{array}{cccccccccccccccc} 1. & 0.92 & 0.91 & 0.18 & -0.09 & -0.4 & 0.39 & 0.84 & 0.85 & 0.96 & 0.97 & 0.8 & 0.18 & -0.54 & 0.81 \\ 0.92 & 1. & 0.94 & 0.21 & -0.11 & -0.61 & 0.38 & 0.64 & 0.93 & 0.93 & 0.92 & 0.92 & -0.03 & -0.48 & 0.54 \\ 0.91 & 0.94 & 1. & 0.31 & -0.05 & -0.6 & 0.4 & 0.72 & 0.86 & 0.91 & 0.89 & 0.87 & -0.06 & -0.62 & 0.61 \\ 0.18 & 0.21 & 0.31 & 1. & -0.18 & -0.5 & 0.11 & 0.23 & 0.12 & 0.19 & 0.19 & 0.32 & -0.3 & -0.57 & 0.14 \\ -0.09 & -0.11 & -0.05 & -0.18 & 1. & 0.07 & -0.07 & 0.07 & -0.1 & -0.07 & -0.11 & -0.07 & -0.07 & 0.02 & -0.09 \\ -0.4 & -0.61 & -0.6 & -0.5 & 0.07 & 1. & 0.05 & -0.14 & -0.62 & -0.53 & -0.45 & -0.73 & 0.54 & 0.54 & 0.04 \\ 0.39 & 0.38 & 0.4 & 0.11 & -0.07 & 0.05 & 1. & 0.51 & 0.12 & 0.31 & 0.28 & 0.15 & 0.48 & 0.04 & 0.44 \\ 0.84 & 0.64 & 0.72 & 0.23 & 0.07 & -0.14 & 0.51 & 1. & 0.5 & 0.77 & 0.73 & 0.45 & 0.34 & -0.43 & 0.86 \\ 0.85 & 0.93 & 0.86 & 0.12 & -0.1 & -0.62 & 0.12 & 0.5 & 1. & 0.92 & 0.91 & 0.94 & -0.22 & -0.52 & 0.43 \\ 0.96 & 0.93 & 0.91 & 0.19 & -0.07 & -0.53 & 0.31 & 0.77 & 0.92 & 1. & 0.95 & 0.87 & -0.03 & -0.56 & 0.66 \\ 0.97 & 0.92 & 0.89 & 0.19 & -0.11 & -0.45 & 0.28 & 0.73 & 0.91 & 0.95 & 1. & 0.88 & 0.07 & -0.61 & 0.75 \\ 0.8 & 0.92 & 0.87 & 0.32 & -0.07 & -0.73 & 0.15 & 0.45 & 0.94 & 0.87 & 0.88 & 1. & -0.3 & -0.66 & 0.39 \\ 0.18 & -0.03 & -0.06 & -0.3 & -0.07 & 0.54 & 0.48 & 0.34 & -0.22 & -0.03 & 0.07 & -0.3 & 1. & 0.43 & 0.55 \\ -0.54 & -0.48 & -0.62 & -0.57 & 0.02 & 0.54 & 0.04 & -0.43 & -0.52 & -0.56 & -0.61 & -0.66 & 0.43 & 1. & -0.44 \\ 0.81 & 0.54 & 0.61 & 0.14 & -0.09 & 0.04 & 0.44 & 0.86 & 0.43 & 0.66 & 0.75 & 0.39 & 0.55 & -0.44 & 1. \end{array} \right).$$

Відповідно до методу компонентного аналізу власні значення наведеної коваріаційної матриці, розташовані у порядку спадання, мають вигляд:

$$\lambda_1 = 8,3744, \lambda_2 = 2,7013, \lambda_3 = 1,2249, \lambda_4 = 1,0389, \lambda_5 = 0,7943, \lambda_6 = 0,3263, \\
 \lambda_7 = 0,2284, \lambda_8 = 0,1744, \lambda_9 = 0,0763, \lambda_{10} = 0,0314, \lambda_{11} = 0,01498, \lambda_{12} = 0,00838, \\
 \lambda_{13} = 0,002979, \lambda_{14} = 0,00219, \lambda_{15} = 0,0002756.$$

Дві головні компоненти дають 73,84 % внеску у загальну дисперсію. Власні вектори, що відповідають власним значенням $\lambda_1 = 8,3744$, $\lambda_2 = 2,7013$, дорівнюють:

$$\mathbf{L}^{(1)} = (0,33; 0,33; 0,33; 0,12; -0,03; -0,2; 0,12; 0,26; \\ 0,31; 0,34; 0,33; 0,31; -0,01; -0,23; 0,24)^T;$$

$$\mathbf{L}^{(2)} = (-0,14; 0,01; 0,01; 0,23; 0,0; -0,4; -0,35; \\ -0,28; 0,12; -0,02; -0,05; 0,19; -0,57; -0,24; -0,36)^T.$$

Тоді розраховані за формулою (2.6) значення двох головних компонент мають вигляд:

$$y^{(1)} = 0,33x_1 + 0,33x_2 + 0,33x_3 + 0,12x_4 - 0,03x_5 - 0,2x_6 - \\ + 0,12x_7 + 0,26x_8 + 0,31x_9 + 0,34x_{10} + 0,33x_{11} + 0,31x_{12} - \\ - 0,01x_{13} - 0,23x_{14} - 0,24x_{15}. \quad (2.19)$$

$$y^{(2)} = -0,14x_1 + 0,01x_2 + 0,01x_3 + 0,23x_4 + 0,0x_5 - 0,4x_6 - \\ - 0,35x_7 - 0,28x_8 + 0,12x_9 - 0,02x_{10} - 0,05x_{11} + 0,19x_{12} - \\ - 0,57x_{13} - 0,24x_{14} - 0,36x_{15}. \quad (2.20)$$

Далі наведемо обчислення методом факторного аналізу для 2010 року. Початкова матриця факторних навантажень:

$$L_{(0)} = \begin{pmatrix} 0,334 & 0,329 & 0,332 & 0,115 & -0,033 & -0,203 & 0,122 & 0,265 \\ -0,135 & 0,011 & 0,01 & 0,23 & -0,005 & -0,401 & -0,353 & -0,281 \\ 0,312 & 0,335 & 0,334 & 0,313 & -0,01 & -0,23 & 0,239 \\ 0,119 & -0,021 & -0,054 & 0,194 & -0,567 & -0,241 & -0,359 \end{pmatrix}.$$

Використовуючи $L_{(0)}$, розрахуємо $V_{(0)}$ – початкове наближення для матриці залишкових дисперсій

$$V_{(0)} = \text{diag} \{0,87 \quad 0,891 \quad 0,889 \quad 0,934 \quad 0,998 \quad 0,798 \quad 0,860 \quad 0,851 \\ 0,889 \quad 0,887 \quad 0,885 \quad 0,864 \quad 0,678 \quad 0,889 \quad 0,813\}.$$

Проводимо ітераційну процедуру відповідно до алгоритму факторного аналізу та за допомогою математичного пакету Wolfram Mathematica отримуємо (за 27 ітерацій) оцінку матриці факторного аналізу та матриці залишкових дисперсій з точністю $\delta = 10^{-3}$:

$$L = \begin{pmatrix} 0,999 & 0,924 & 0,913 & 0,188 & -0,086 & -0,419 & 0,381 & 0,834 \\ -0,025 & 0,296 & 0,214 & 0,166 & -0,025 & -0,684 & -0,334 & 0,403 \\ 0,865 & 0,966 & 0,977 & 0,819 & 0,157 & -0,55 & 0,796 \\ 0,451 & 0,169 & 0,117 & 0,529 & -0,798 & -0,25 & -0,531 \end{pmatrix},$$

$$V = \text{diag}\{0,001 \quad 0,058 \quad 0,121 \quad 0,937 \quad 0,992 \quad 0,356 \quad 0,743 \quad 0,142$$

$$0,048 \quad 0,038 \quad 0,031 \quad 0,049 \quad 0,338 \quad 0,635 \quad 0,084\}.$$

Далі наведено результати застосування методів зниження вимірності за статистичними даними 2015 року. Коваріційна матриця Σ нормованих даних має вигляд:

$$\begin{pmatrix} 1. & 0.94 & 0.92 & -0.34 & 0.06 & -0.74 & -0.1 & 0.94 & 0.95 & 0.94 & 0.97 & 0.86 & 0.32 & -0.32 & 0.61 \\ 0.94 & 1. & 0.85 & -0.34 & 0.23 & -0.66 & 0.04 & 0.83 & 0.93 & 0.91 & 0.91 & 0.91 & 0.32 & -0.06 & 0.34 \\ 0.92 & 0.85 & 1. & -0.25 & -0.09 & -0.77 & -0.23 & 0.92 & 0.89 & 0.88 & 0.86 & 0.75 & 0.24 & -0.38 & 0.68 \\ -0.34 & -0.34 & -0.25 & 1. & 0.22 & 0.42 & -0.09 & -0.28 & -0.41 & -0.4 & -0.31 & -0.27 & 0. & 0.07 & -0.02 \\ 0.06 & 0.23 & -0.09 & 0.22 & 1. & 0.19 & 0.48 & -0.01 & 0.04 & 0.04 & 0.13 & 0.33 & 0.25 & 0.46 & -0.36 \\ -0.74 & -0.66 & -0.77 & 0.42 & 0.19 & 1. & 0.28 & -0.63 & -0.85 & -0.89 & -0.68 & -0.6 & 0.25 & 0.5 & -0.42 \\ -0.1 & 0.04 & -0.23 & -0.09 & 0.48 & 0.28 & 1. & -0.14 & -0.05 & -0.08 & 0.01 & 0.2 & 0.05 & 0.24 & -0.4 \\ 0.94 & 0.83 & 0.92 & -0.28 & -0.01 & -0.63 & -0.14 & 1. & 0.86 & 0.83 & 0.88 & 0.75 & 0.37 & -0.34 & 0.68 \\ 0.95 & 0.93 & 0.89 & -0.41 & 0.04 & -0.85 & -0.05 & 0.86 & 1. & 0.99 & 0.93 & 0.89 & 0.06 & -0.36 & 0.43 \\ 0.94 & 0.91 & 0.88 & -0.4 & 0.04 & -0.89 & -0.08 & 0.83 & 0.99 & 1. & 0.91 & 0.87 & 0.03 & -0.35 & 0.43 \\ 0.97 & 0.91 & 0.86 & -0.31 & 0.13 & -0.68 & 0.01 & 0.88 & 0.93 & 0.91 & 1. & 0.93 & 0.26 & -0.38 & 0.56 \\ 0.86 & 0.91 & 0.75 & -0.27 & 0.33 & -0.6 & 0.2 & 0.75 & 0.89 & 0.87 & 0.93 & 1. & 0.14 & -0.21 & 0.27 \\ 0.32 & 0.32 & 0.24 & 0. & 0.25 & 0.25 & 0.05 & 0.37 & 0.06 & 0.03 & 0.26 & 0.14 & 1. & 0.43 & 0.4 \\ -0.32 & -0.06 & -0.38 & 0.07 & 0.46 & 0.5 & 0.24 & -0.34 & -0.36 & -0.35 & -0.38 & -0.21 & 0.43 & 1. & -0.57 \\ 0.61 & 0.34 & 0.68 & -0.02 & -0.36 & -0.42 & -0.4 & 0.68 & 0.43 & 0.43 & 0.56 & 0.27 & 0.4 & -0.57 & 1. \end{pmatrix}.$$

Відповідно до методу компонентного аналізу власні значення наведеної коваріаційної матриці, розташовані у порядку спадання, мають вигляд:

$$\lambda_1 = 8,5085, \lambda_2 = 2,4236, \lambda_3 = 1,5939, \lambda_4 = 0,99097, \lambda_5 = 0,7417, \lambda_6 = 0,2841,$$

$$\lambda_7 = 0,2129, \lambda_8 = 0,12815, \lambda_9 = 0,06409, \lambda_{10} = 0,02908, \lambda_{11} = 0,0109, \\ \lambda_{12} = 0,005698, \lambda_{13} = 0,003105, \lambda_{14} = 0,00259, \lambda_{15} = 0,0004848.$$

Дві головні компоненти дають 72,88 % внеску у загальну дисперсію. Власні вектори, що відповідають власним значенням $\lambda_1 = 8,5085$, $\lambda_2 = 2,4236$, дорівнюють:

$$\mathbf{L}^{(1)} = (0,34; 0,31; 0,32; -0,13; 0,0; -0,28; -0,04; 0,32; \\ 0,33; 0,33; 0,33; 0,3; 0,07; -0,14; 0,21)^T;$$

$$\mathbf{L}^{(2)} = (-0,05; -0,21; 0,06; -0,02; -0,54; -0,2; -0,43; \\ 0,0; -0,03; -0,02; -0,09; -0,22; -0,29; -0,45; 0,28)^T.$$

Тоді розраховані за формулою (2.6) значення двох головних компонент мають вигляд:

$$y^{(1)} = 0,34x_1 + 0,31x_2 + 0,32x_3 - 0,13x_4 + 0,0x_5 - 0,28x_6 - \\ -0,04x_7 + 0,32x_8 + 0,33x_9 + 0,33x_{10} + 0,33x_{11} + 0,3x_{12} + \\ +0,07x_{13} - 0,14x_{14} + 0,21x_{15}; \quad (2.21)$$

$$y^{(2)} = -0,05x_1 - 0,21x_2 + 0,06x_3 - 0,02x_4 - 0,54x_5 - 0,2x_6 - \\ -0,43x_7 + 0,0x_8 - 0,03x_9 - 0,02x_{10} - 0,09x_{11} - 0,22x_{12} - \\ -0,29x_{13} - 0,45x_{14} + 0,28x_{15}. \quad (2.22)$$

Далі наведемо обчислення за методом факторного аналізу. Початкова матриця факторних навантажень:

$$L_{(0)} = \begin{pmatrix} 0,339 & 0,315 & 0,324 & -0,129 & 0,001 & -0,278 & -0,039 & 0,318 \\ -0,05 & -0,208 & 0,062 & -0,016 & -0,544 & -0,2 & -0,431 & -0,005 \\ 0,334 & 0,331 & 0,33 & 0,298 & 0,068 & -0,137 & 0,206 \\ -0,03 & -0,017 & -0,085 & -0,223 & -0,294 & -0,452 & 0,282 \end{pmatrix}.$$

Використовуючи $L_{(0)}$, розрахуємо $V_{(0)}$ – початкове наближення для матриці залишкових дисперсій

$$V_{(0)} = \text{diag} \{0,883 \quad 0,858 \quad 0,891 \quad 0,983 \quad 0,704 \quad 0,882 \quad 0,812 \quad 0,899 \\ 0,887 \quad 0,890 \quad 0,884 \quad 0,861 \quad 0,909 \quad 0,777 \quad 0,878\}.$$

Проводимо ітераційну процедуру відповідно до алгоритму факторного аналізу та за допомогою математичного пакету Wolfram Mathematica отримуємо (за 308 ітерацій) оцінку матриці факторного аналізу та матриці залишкових дисперсій з точністю $\delta = 10^{-3}$.

$$L = \begin{pmatrix} 0,998 & 0,844 & 0,829 & 0,131 & -0,104 & -0,196 & 0,306 & 0,934 \\ -0,034 & 0,379 & 0,356 & 0,043 & 0,286 & -0,732 & 0,203 & -0,123 \\ 0,789 & 0,931 & 0,940 & 0,626 & 0,460 & -0,401 & 0,893 \\ 0,566 & 0,211 & 0,248 & 0,723 & -0,596 & -0,284 & -0,389 \end{pmatrix},$$

$$V = \text{diag} \{0,002 \quad 0,144 \quad 0,186 \quad 0,981 \quad 0,907 \quad 0,425 \quad 0,865 \quad 0,112 \\ 0,056 \quad 0,089 \quad 0,054 \quad 0,085 \quad 0,434 \quad 0,759 \quad 0,05\}.$$

Графічна візуалізація та аналіз отриманих результатів наведено у розділі 4 атестаційної роботи.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Mathematica 11 як система символної математики

Wolfram Mathematica (WM) є пакетом символної математики. Велика кількість закладених розробниками функцій, а також відкрите середовище, яке дозволяє доповнювати пакет своїми власними розширеннями, робить його можливості майже безмежними. Wolfram Mathematica має високу швидкість та практично не обмежену точність розрахунків, що дозволяє їй працювати як на дуже потужних комп'ютерах, так і на слабких персональних комп'ютерах. На основі ядра пакету створений Web-сервер, який дозволяє використовувати її можливості необмеженій кількості людей.

Основними конкурентами пакету Wolfram Mathematica називають Maple, MathCAD та MatLab. Якщо з першим складно сперечатися, то відносно інших можна зробити такі зауваження. Пакети MathCAD та MatLab займають зовсім іншу нішу ніж Wolfram Mathematica. Обидва вони для розрахунків використовують чисельні, а не символні алгоритми. Символьні розрахунки в них є слабо розвиненими додатками.

Інтерфейс пакету будується з декількох базових понять: Зошит (Notebooks), Осередок (Cell) і Палітра (Palletes). Зошитом називається файл, з яким працює користувач. У ньому створюються і обчислюються формули, будуються графіки і таблиці. За бажанням в зошиті можна навіть програти звуковий файл або фільм.

Зошит структуровано за комірками. Грубо комірки можна порівняти з параграфом в текстовому редакторі. Всі інформація, яка є в зошиті, зберігається в його комірках. Як тільки ви в порожньому новому файлі наберете хоча б один символ, Wolfram Mathematica створить для нього комірку. Комірка також є мінімальною одиницею, яку можна обчислити. Тобто, якщо у вас в комірці є дві формули, обчислити їх окремо не вийде. Всі комірки можна розділити на три типи:

- комірки введення – в них задаються команди (формули), які будуть обчислені;
- комірки результату – в них Wolfram Mathematica виводить результат обчислень;
- інші комірки – комірки з текстом, заголовки і все інше, що вводить користувач і обчислювати не треба (можна було б назвати їх необчислювані комірки).

Будь-які комірки можна об'єднувати і розбивати з допомогою команд меню Cell: Divide Cell (розбити комірки) і Merge Cells (об'єднати комірки).

Для швидкого доступу до функцій розробники Wolfram Mathematica ввели спеціальні типи вікон, які називаються палітрами. Палітри містять вікна з кнопками, які виконують дії. Дії можуть бути зовсім різними – від додавання грецької букви до розкриття дужок в алгебраїчному вираз. Різні палітри доступні через меню File-Palletes.

Пакет підтримує кирилицю і грецькі літери нарівні з англійським алфавітом. У той же час, ідентифікатори розрізняються за регістром, тобто змінна A не те саме, що змінна a . Для такого введення індексу можна скористатися палітрою Basic Input. За допомогою неї також можна ввести шаблони операцій (таких як сума, множення, корінь) і грецькі літери. Інші символи можна знайти на палітрі Basic Typesettings. Повний список всіх символів, які знає Wolfram Mathematica, можна знайти на палітрі Complete Charecters. Щоб ввести символ, потрібно натиснути на кнопку з його зображенням.

Однією з особливостей програми є назва стандартних функцій повними іменами без скорочень. Це дозволяє (при певному рівні знання математичної англійської мови) дуже швидко знаходити потрібні функції.

Wolfram Mathematica не тільки може виконувати необхідні обчислення, але й у багатьох випадках вона вибере оптимальний спосіб проведення обчислень. Все що потрібно зробити – це визначити завдання; Wolfram Mathematica ховає всі складні механічні аспекти процесу обчислення, дозволяючи концентруватися безпосередньо на завданні.

Величезною перевагою програми Wolfram Mathematica є потужна довідкова система, яка дозволяє уточнити призначення будь-якої функції, оператора або службового слова системи і поступово знайомить з її можливостями. Однак вона включає в себе не тільки дуже якісний опис функцій з прикладами, а також підручник. У ній є всі матеріали для тих, хто тільки починає роботу з програмою, і для тих, хто працює з нею дуже давно. Але вся програма і довідкова система написані виключно англійською мовою. Тому ця довідкова система не претендує на роль навчальної системи і незручна для знайомства з системою Wolfram Mathematica.

У відношенні графіки система Wolfram Mathematica є лідером серед систем комп'ютерної алгебри. Велика кількість опцій дозволяє оформляти графічні образи практично в будь-якому бажаному вигляді. Графіки в системі Wolfram Mathematica є об'єктами і тому вони можуть бути значеннями змінних.

Для побудови двовимірних графіків функцій виду $f(x)$ використовується вбудована в ядро функція Plot.

Також часто виникає необхідність побудови графіка по точках. Це забезпечує вбудована в ядро графічна функція ListPlot.

ListPlot [{y1, y2, ...}] виводить графік списку величин. Координати x приймають значення 1, 2,

ListPlot [{x1, y1}, {x2, y2}, ...] виводить графік списку величин з зазначеними x і y координатами.

Система Wolfram Mathematica також дозволяє будувати графіки функцій в полярній системі координат, параметрично або неявно заданих функцій.

Для розв'язання поставленої задачі після аналізу можливостей різних середовищ програмування і математичних пакетів був обраний пакет Wolfram Mathematica 11.2, оскільки він є найбільш оптимальним для проведення теоретичного і чисельного аналізу.

3.2 Опис програми

Програма виконана з використанням можливостей пакету Wolfram Mathematica 11.2, який дозволяє створювати програмні продукти у формі електронних документів. При відкритті програми перед користувачем з'являється вікно, в якому показано меню можливостей програми. В кожному із пунктів меню є комірочки, що дозволяють редагувати параметри величин, які використовуються при розв'язанні даної задачі.

В першому пункті програми імпортуються статистичні данні для обробки програмою. В наступному пункті всі данні, що були імпортовані, збираються у матрицю та нормуються. Після цього формуються основні співвідношення для розв'язання задачі обраними методами та задається процедура графічного відображення результатів.

На наступному етапі безпосередньо реалізуються алгоритми методу компонентного аналізу та факторного аналізу, а саме:

- а) розраховується коваріаційна матриця;
- б) визначаються власні числа та відповідні їм власні вектори матриці з п. а);
- в) отримані в п б) значення упорядковуються у порядку спадання;
- г) обчислюються значення координат точок-проекцій вихідних даних відповідно до обраних методів;
- д) здійснюється графічне подання отриманих результатів.

На графіках наочно відображаються отримані результати, за якими можна прослідкувати зміни у стані досліджуваних об'єктів з плином часу.

Лістинг програми додається в додатку А.

4 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ

4.1 Аналіз економічного стану регіонів України методами головних компонент та факторного аналізу

За допомогою програми, написаної в Wolfram Mathematica, в атестаційній роботі були проведені обчислювальні експерименти за статистичними даними, що описують соціально-економічне положення регіонів України у 2000, 2005, 2010, 2015 роках.

Обчислені за формулами (2.15) – (2.16) координати регіонів України у системі координат $(y^{(1)}, y^{(2)})$ у 2000 році, мають вигляд:

Вінницька: $(-1,241 \quad 0,0076)$;	Миколаївська: $(-0,6699 \quad 0,9288)$;
Волинська: $(-1,6283 \quad 0,6326)$;	Одеська: $(1,2326 \quad 0,4306)$;
Дніпропетровська: $(4,5253 \quad -2,1296)$;	Полтавська: $(0,1617 \quad -0,6894)$;
Донецька: $(5,6454 \quad -3,1394)$;	Рівненська: $(-2,1507 \quad -0,5601)$;
Житомирська: $(-1,8419 \quad -1,2377)$;	Сумська: $(-0,9319 \quad -1,0478)$;
Закарпатська: $(-1,6833 \quad 0,0512)$	Тернопільська: $(-2,6430 \quad -0,9981)$;
Запорізька: $(1,6706 \quad 0,5067)$;	Харківська: $(1,3694 \quad -1,5178)$;
Івано-Франківська: $(-1,1255 \quad -0,6438)$;	Херсонська: $(-1,6293 \quad 1,7433)$;
Київська: $(0,3196 \quad -0,3059)$;	Хмельницька: $(-1,8479 \quad 0,2074)$;
Кіровоградська: $(-1,9367 \quad 0,8348)$;	Черкаська: $(-1,4998 \quad -0,0387)$;
Луганська: $(0,5277 \quad -0,1309)$;	Чернівецька: $(-2,2708 \quad 0,3058)$;
Львівська: $(0,5625 \quad -2,2247)$	Чернігівська: $(-1,3865 \quad -0,8744)$;
м. Київ: $(10,3666 \quad 3,3606)$;	м. Севастополь: $(-1,1392 \quad 5,5399)$;
Республіка Крим: $(-0,7553 \quad 0,9889)$.	

Ці числові значення представляють собою деякі інтегральні показники, що характеризують соціально-економічний стан кожного регіону України у цілому.

Графічне розташування країн за головними компонентами $(y^{(1)}, y^{(2)})$ наведено на рис. 4.1. Для виділення подібних за досліджуваними показниками об'єктів використано процедуру кластеризації методами k -середніх та k -медіан із кількістю кластерів 6 та 7 (рис. 4.2).

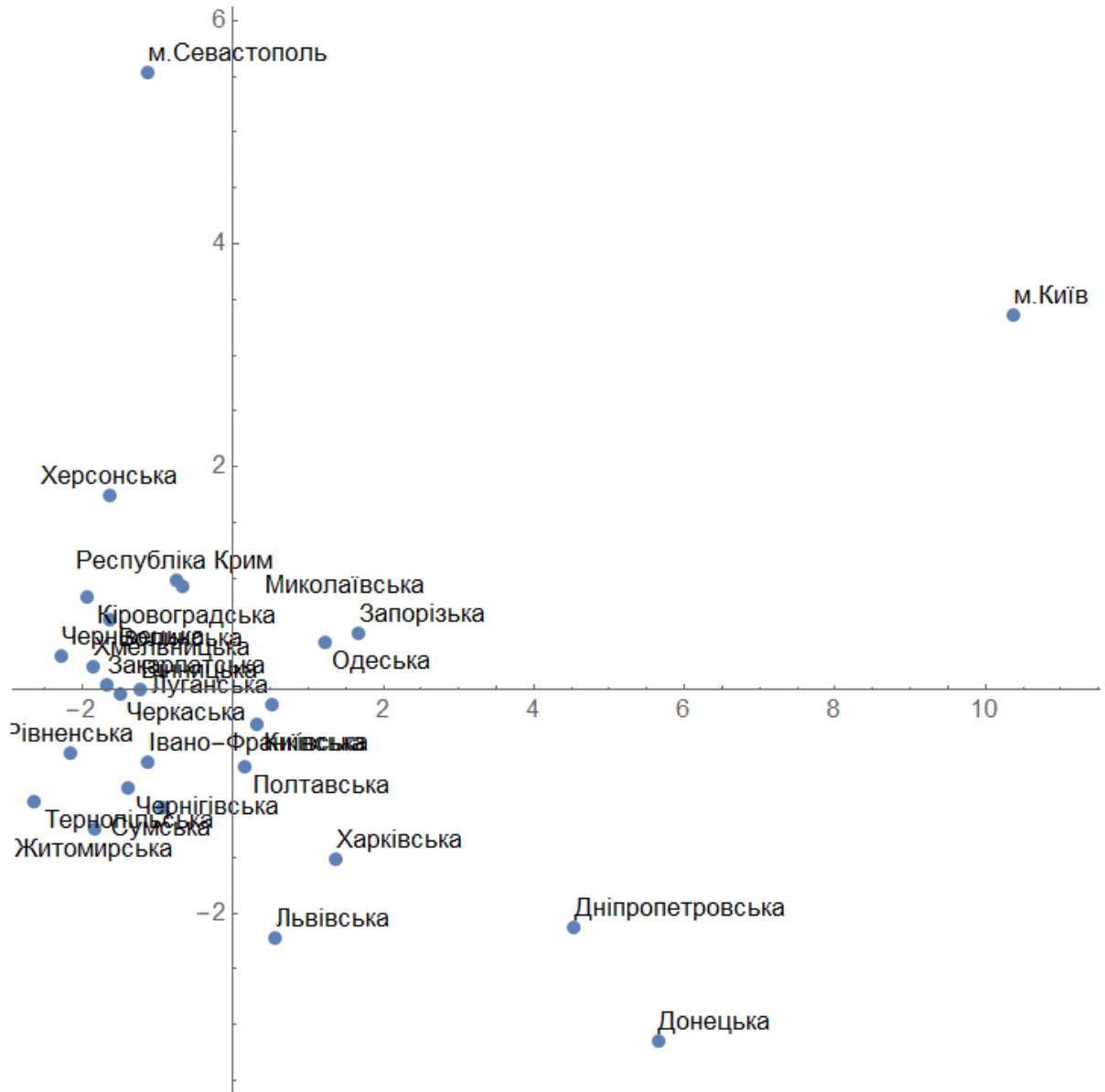


Рисунок 4.1 – Розташування регіонів у 2000 році у відповідності з методом компонентного аналізу

Як видно, із рис. 4.1 регіони України можна поділити на декілька груп за їх розташуванням. А саме, окреме положення займають міста Київ та Севастополь, що пов'язано з особливостями їх політичного статусу; окремо виділяють-

ся південно-східні області, які мають розвинену промисловість та високо розвинену соціальну сферу; а також західні та центральні регіони, які, можливо, є подібними за досліджуваними соціально-економічними показниками (невелика площа, превалювання розвитку певних галузей економіки тощо).

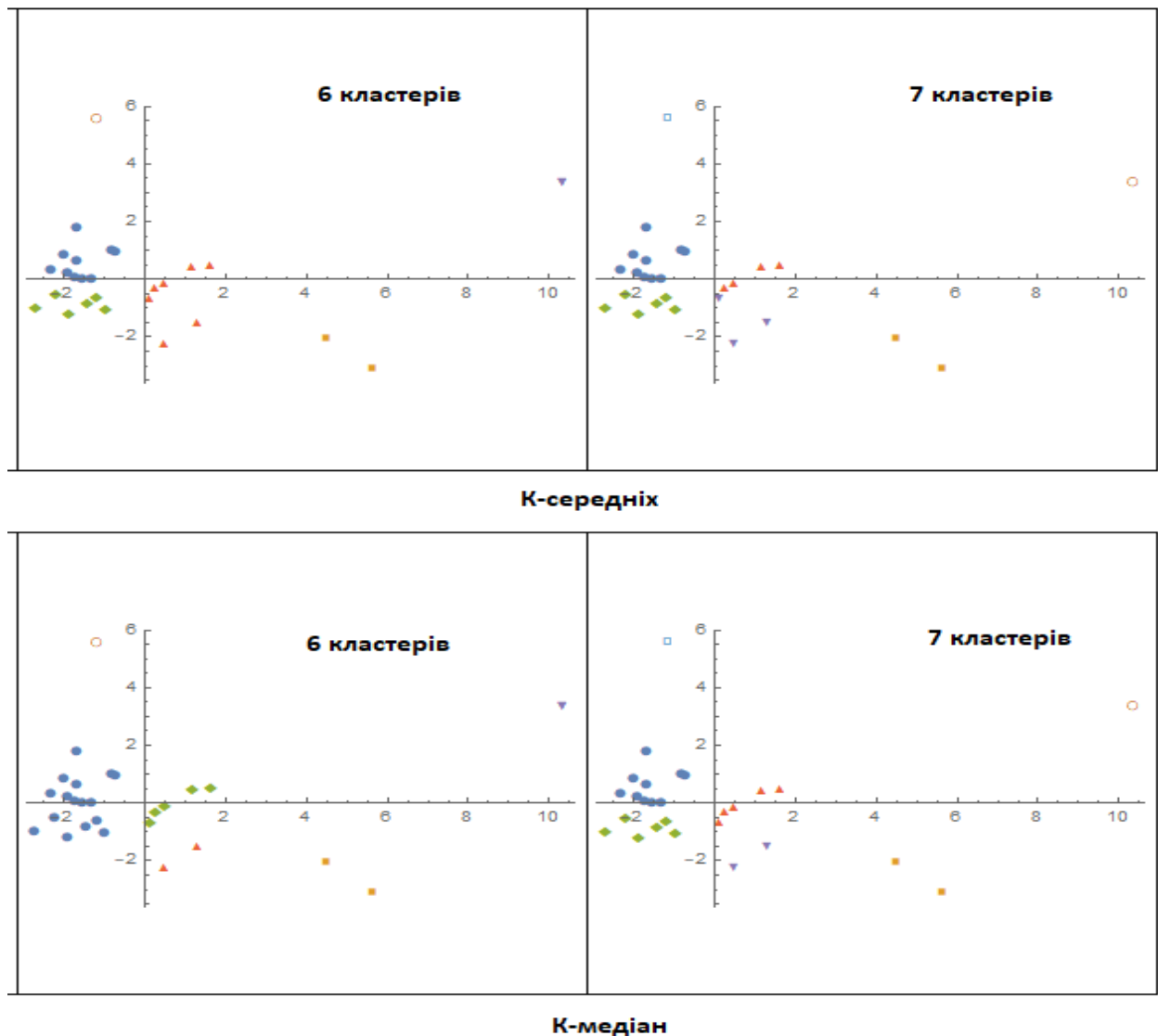


Рисунок 4.2 – Групування регіонів у 2000 році у відповідності з методом компонентного аналізу

Аналізуючи рис. 4.2 можна виділити особливості, що є спільними для обох застосованих методів кластеризації. Зокрема, у випадку 6-ти кластерів області Запорізька, Одеська, Луганська, Київська, Полтавська, Харківська та Львівська формують окремий кластер як методом k -середніх так і методом

k -медіан; аналогічно як і області Житомирська, Тернопільська, Чернігівська, Івано-Франківська, Рівненська та Сумська. Також окремими кластерами є м. Київ, м. Севастополь, об'єднання Донецької та Дніпропетровської областей.

Продовжуючи аналіз результатів, розглянемо координати регіонів України за 2000 рік, обчислені за формулою (2.14) відповідно до методу факторного аналізу:

Вінницька: $(-0,3289 \quad -0,3925)$;	Миколаївська: $(-0,4311 \quad 0,1664)$;
Волинська: $(-0,6882 \quad 0,3886)$;	Одеська: $(0,4542 \quad -0,3664)$;
Дніпропетровська: $(1,9281 \quad -0,9326)$;	Полтавська: $(0,1249 \quad -0,0156)$;
Донецька: $(2,8414 \quad -2,5049)$;	Рівненська: $(-0,6371 \quad -0,123)$;
Житомирська: $(-0,5477 \quad -0,1133)$;	Сумська: $(-0,3881 \quad 0,1423)$;
Закарпатська: $(-0,6859 \quad 0,4048)$	Тернопільська: $(-0,7981 \quad -0,3001)$;
Запорізька: $(0,576 \quad 0,1803)$;	Харківська: $(0,7533 \quad -0,4034)$;
Івано-Франківська: $(-0,4812 \quad -0,1226)$;	Херсонська: $(-0,6675 \quad 0,0573)$;
Київська: $(0,1751 \quad -0,0544)$;	Хмельницька: $(-0,5384 \quad -0,2992)$;
Кіровоградська: $(-0,7154 \quad 0,0002)$;	Черкаська: $(-0,4745 \quad -0,133)$;
Луганська: $(0,2992 \quad -0,9802)$;	Чернівецька: $(-0,9033 \quad 0,3295)$;
Львівська: $(0,2138 \quad -0,1874)$	Чернігівська: $(-0,4912 \quad 0,2133)$;
м. Київ: $(2,7057 \quad 4,1375)$;	м. Севастополь: $(-1,0473 \quad 1,2782)$;
Республіка Крим: $(-0,2479 \quad -0,3699)$.	

На рис. 4.3. продемонстровано графічне розташування отриманих точок. Результати кластеризації отриманих результатів методами k -середніх та k -медіан із кількістю кластерів 6 та 7 наведено на рис. 4.4.

Як видно з рисунку 4.4, у випадку 6-ти кластерів області Запорізька, Харківська, Луганська, Одеська, Львівська, Київська, Дніпропетровська формують окремий кластер як методом k -середніх так і методом k -медіан, також згруповані області Рівненська, Житомирська, Хмельницька, Тернопільська, Івано-Франківська, Вінницька, Республіка Крим, Черкаська. Окремими кластерами,

як і раніше, є м. Київ та м. Севастополь.

Порівнюючи результати обчислень за методами компонентного та факторного аналізу, наведені на рис. 4.1 та 4.3 відповідно, можна зробити висновок, що структури розташування областей після застосування обох методів зниження вимірності є подібними і відповідають загальноприйнятним уявленням про особливості регіональної економіки України. Це свідчить про те, що обидва метода є перспективними для використання їх у задачах аналізу економічних об'єктів.

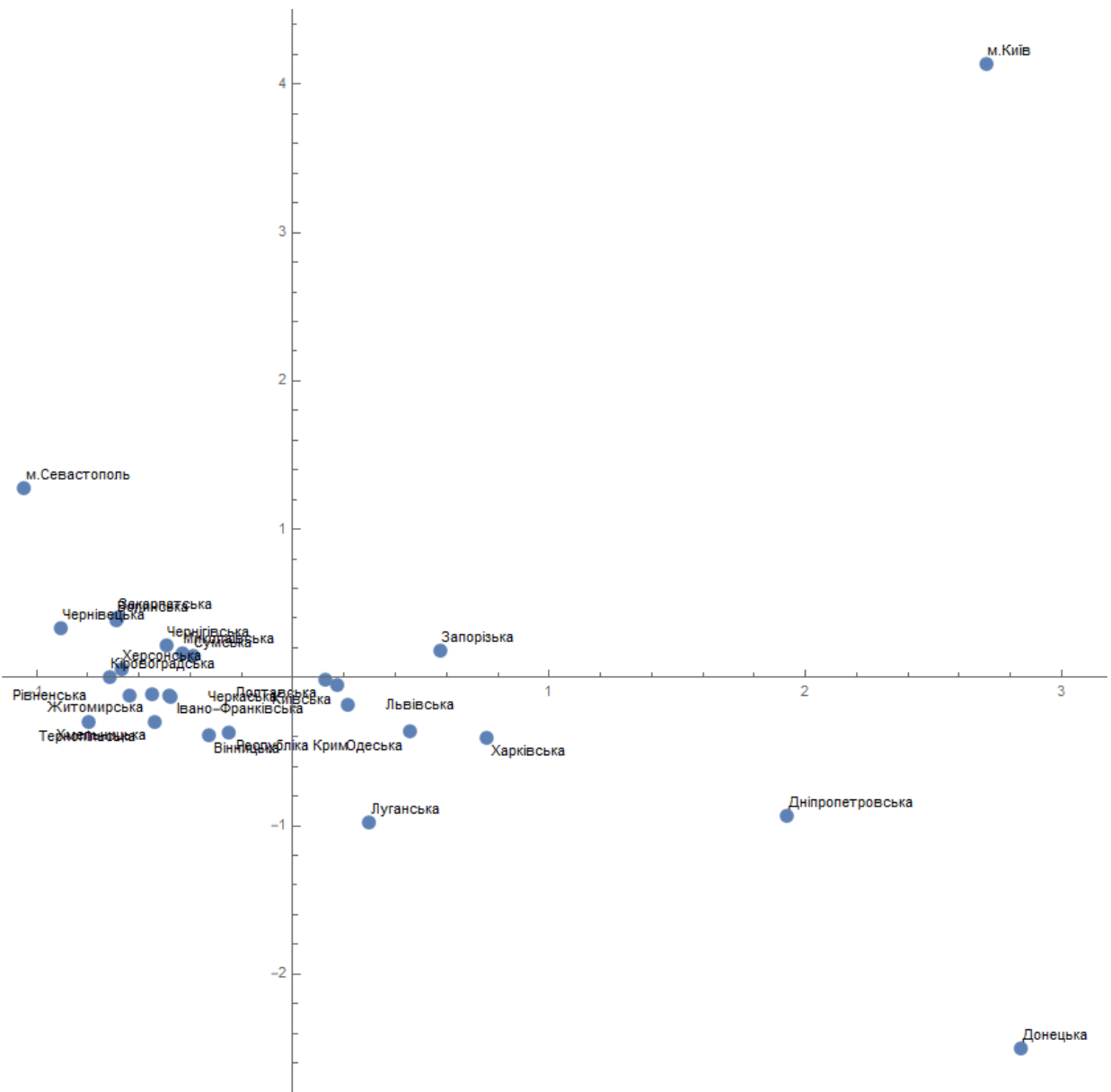


Рисунок 4.3 – Розташування регіонів у 2000 році у відповідності з методом факторного аналізу

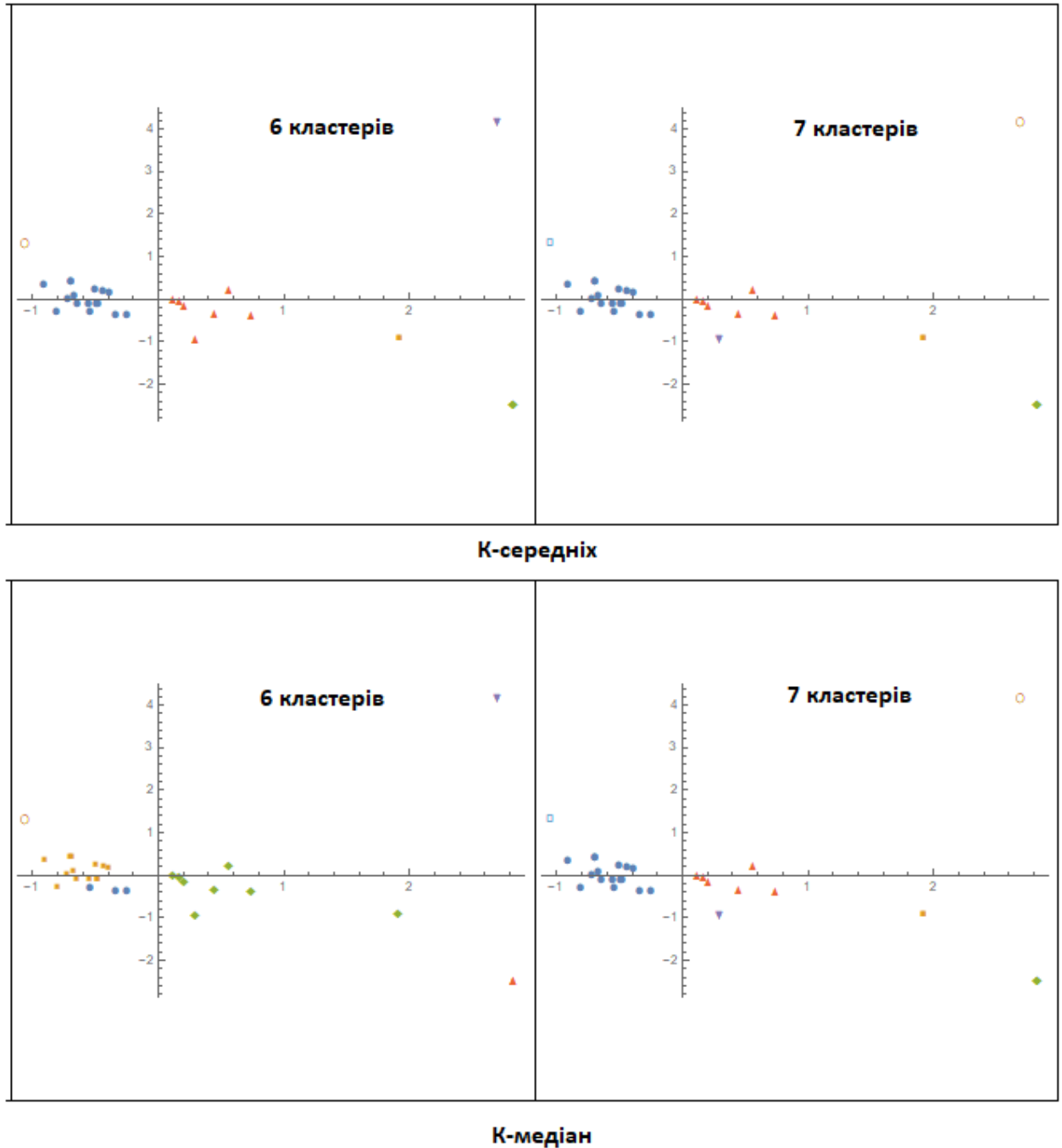


Рисунок 4.4 – Групування регіонів у 2000 році у відповідності з методом факторного аналізу

Результати аналогічних обчислень за даними 2005, 2010, 2015 років наведено у додатку Б. Аналіз цих результатів дозволяє зробити висновки, аналогічні висновкам, отриманим за даними 2000 року і свідчить про ефективність обраних методів для розв'язання поставленої у атестаційній роботі задачі.

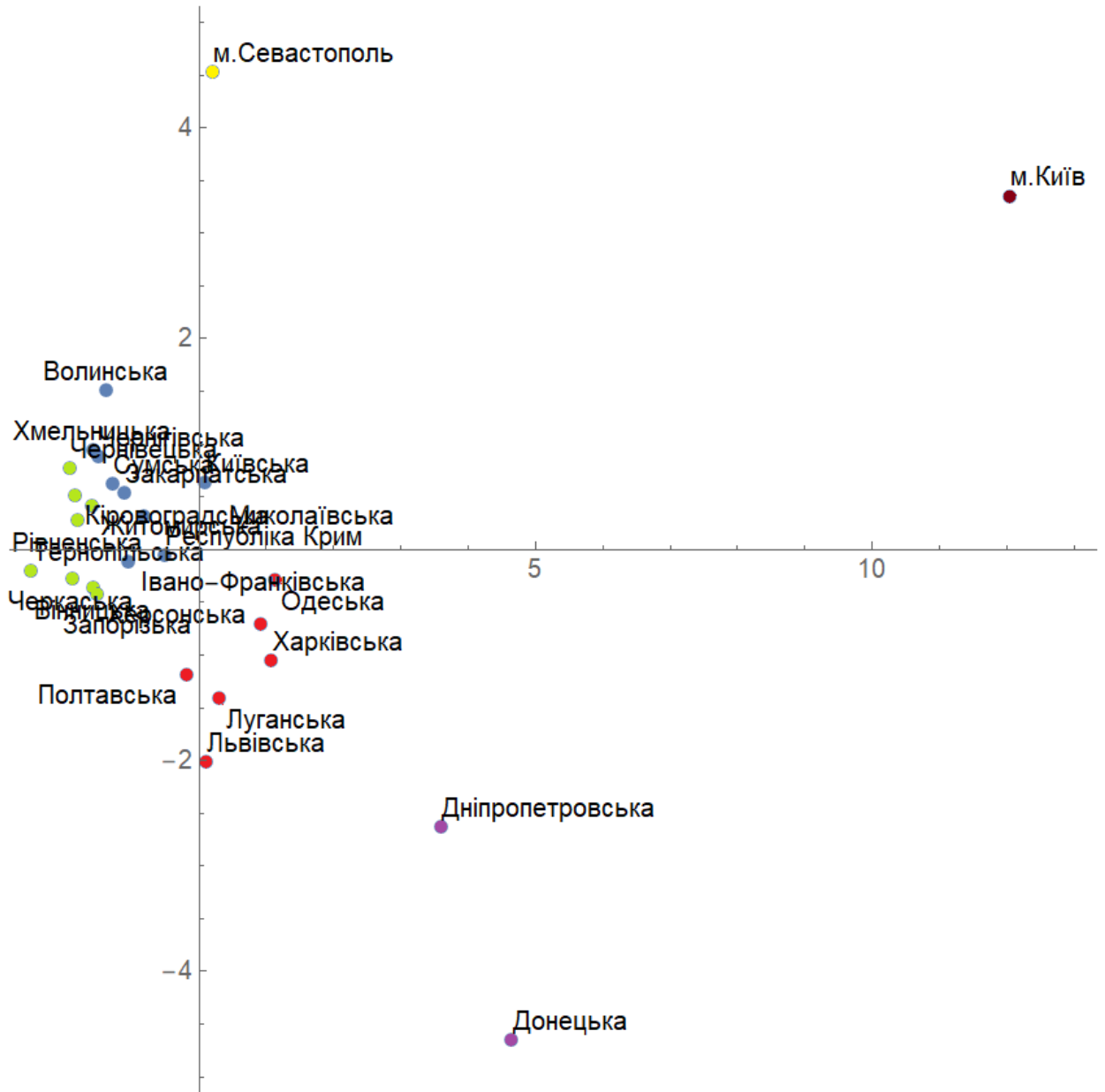


Рисунок 4.6 – Розташування регіонів України у 2005 році

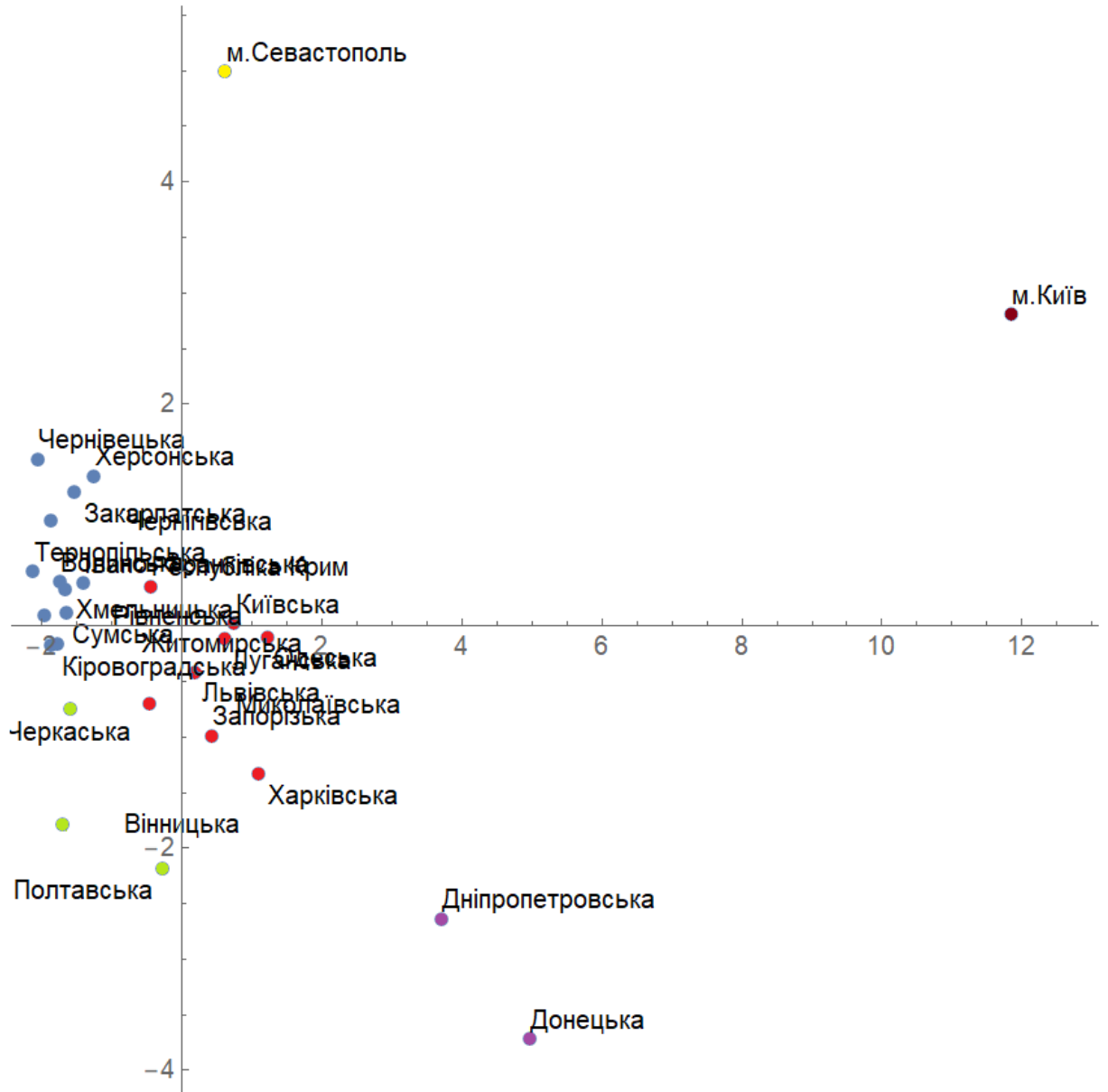


Рисунок 4.7 – Розташування регіонів України у 2010 році

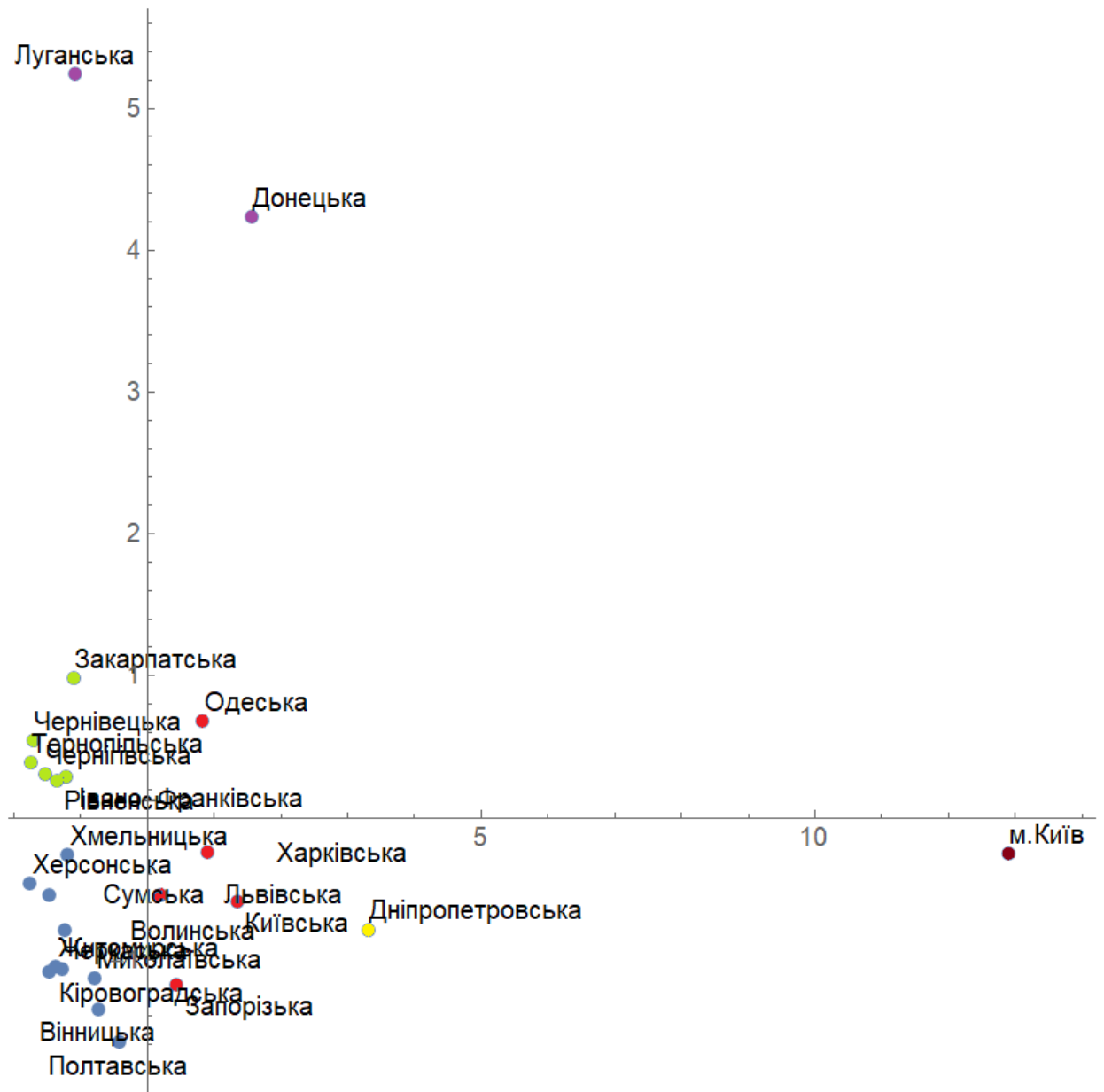


Рисунок 4.8 – Розташування регіонів України у 2015 році

Розглянувши результати компонентного аналізу у динаміці з 2000 по 2015 роки, бачимо, що деякі області змінюють своє розташування в кластерах, це можна пояснити тим що з часом змінювалось їх соціально-економічне положення. Інші області з плином часу не змінювали свою роль основоутворюючих для відповідних кластерів.

Аналогічно проаналізуємо змінення економічного стану регіонів України за обрані роки XXI століття за результатами застосування методу факторного аналізу. Нижче наведено графіки, що характеризують положення досліджува-

них об'єктів після застосування факторного аналізу та подальшої кластеризації методом k -середніх у 2000, 2005, 2010, 2015 роках (рис. 4.9 – 4.12).

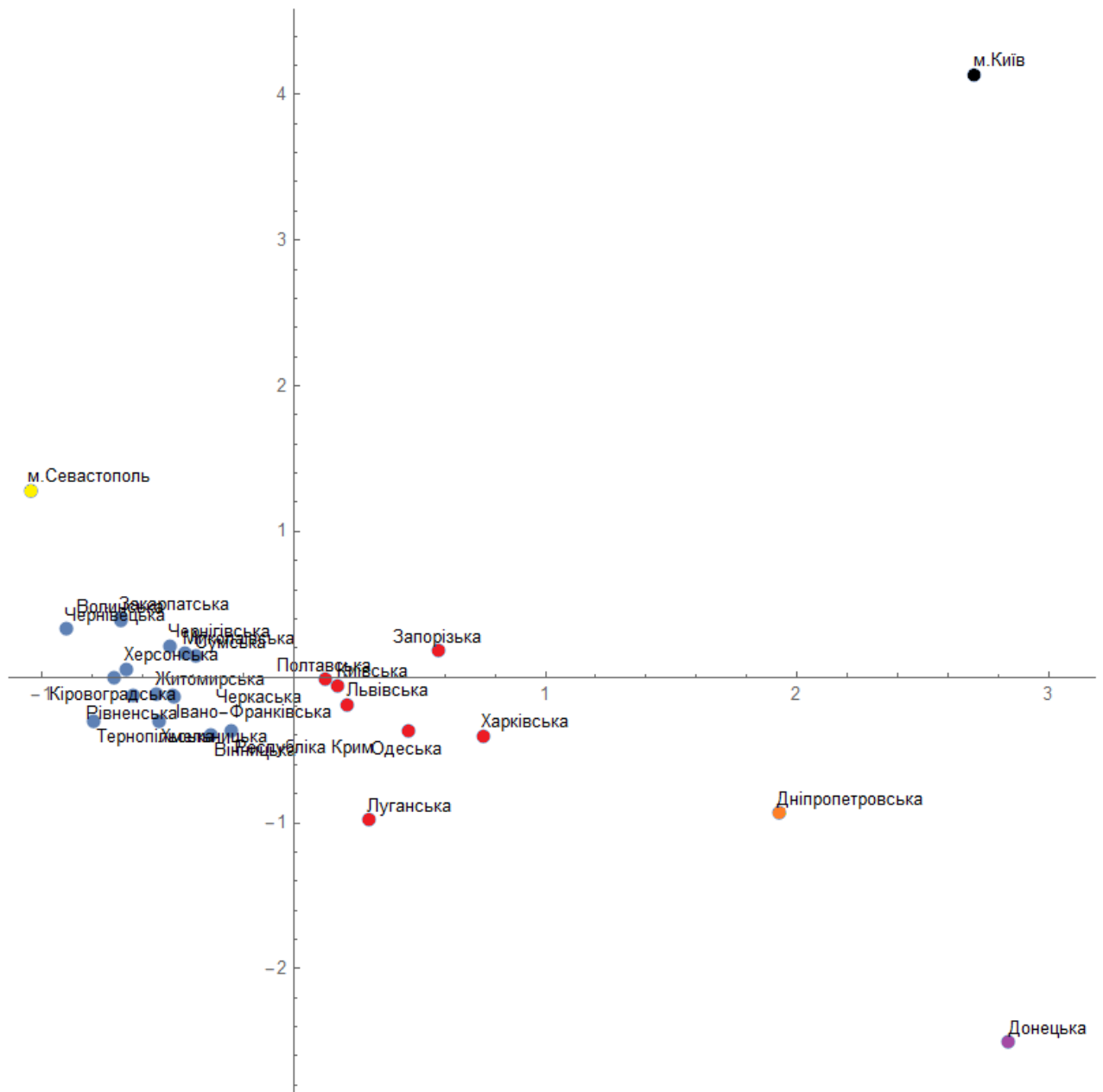


Рисунок 4.9 – Результати факторного аналізу за 2000 рік

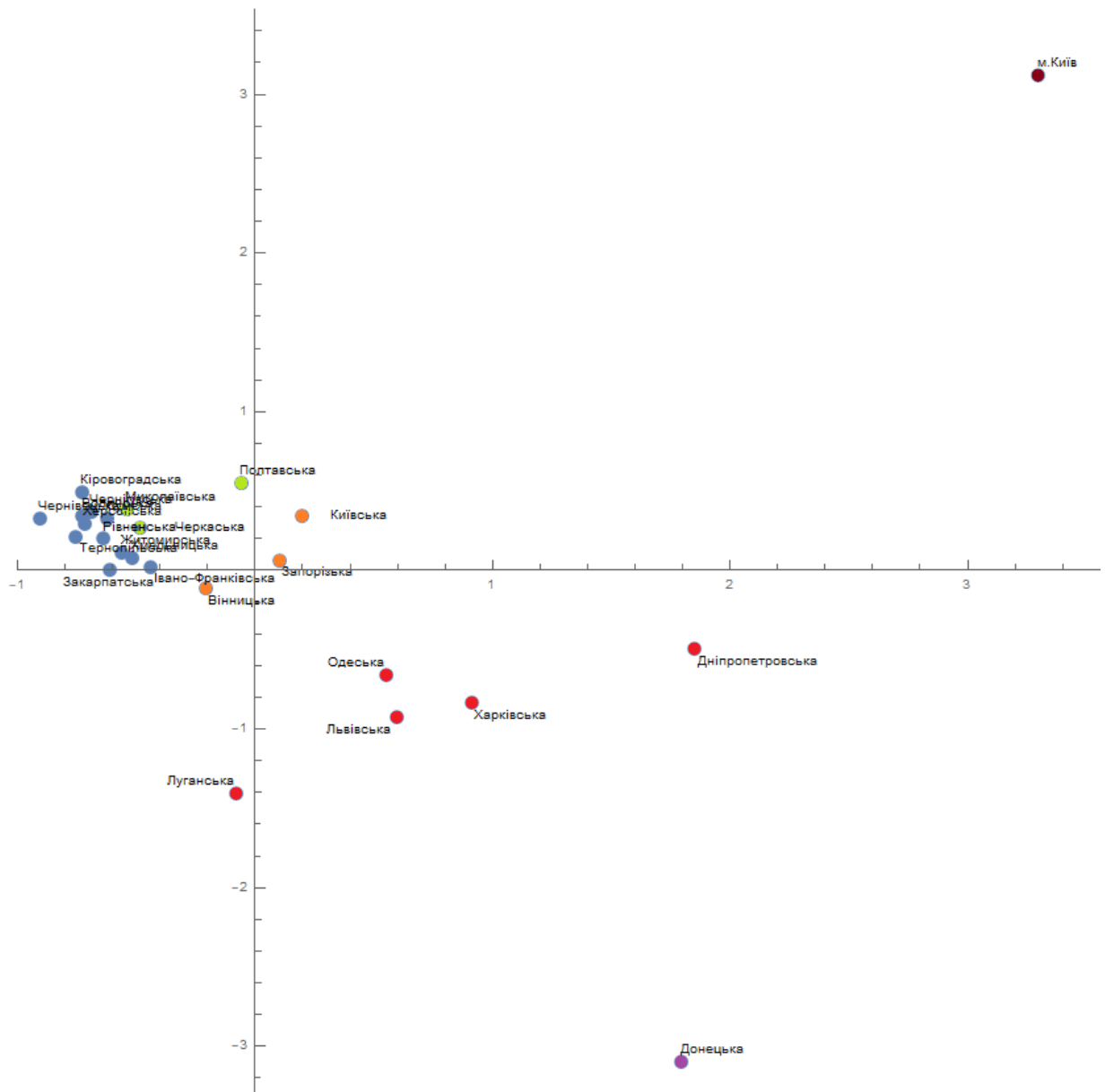


Рисунок 4.12 – Результати факторного аналізу за 2015 рік

Аналізуючи наведені графіки, можна помітити деякі зміни у положенні регіонів України, що пояснюється змінням їх соціально-економічних показників з часом. Методи головних компонент і факторного аналізу утворюють схожі кластери, а це означає що їх можна використовувати не лише для більш наочного подання вхідних даних, а й для дослідження змін у динаміці. Це робить їх перспективними методами багатовимірного статистичного аналізу складних об'єктів (не тільки економічної природи), що описуються великою кількістю різноманітних показників.

5 АНАЛІЗ МОЖЛИВИХ ЗАСТОСУВАНЬ

Актуальність задачі, розв'язаної у атестаційній роботі, пояснюється тим, що методи зниження вимірності можуть застосовуватися не лише в економіці, а також в психодіагностиці, соціології, політології, тощо.

Розглянемо коротко застосування методів багатовимірною статистичного аналізу у найбільш перспективних дослідженнях. Психодіагностика є однією з найбільш розроблених областей застосування цих методів. Стратегія використання ґрунтується на гіпотезі про автоінформативність експериментальних даних, яка має на увазі, що діагностичну модель можна створити шляхом апроксимації геометричної структури множини об'єктів в просторі вихідних ознак. Добру лінійну діагностичну модель вдається побудувати, коли значна частина вихідних ознак внутрішньо узгоджена. Якщо ця внутрішня узгодженість відбиває шуканий психологічний конструкт, то параметри лінійної діагностичної моделі (ваги ознак) дає метод головних компонент.

У соціології методи зниження вимірності можуть бути застосовані для вирішення двох основних задач: перша – аналіз даних (опис результатів опитувань або інших досліджень, представлених у вигляді масивів числових даних); друга – опис соціальних явищ (побудова моделей явищ, в тому числі і математичних моделей).

Метод головних компонент є одним з основних методів в хемометриці, де він дозволяє розділити матрицю вихідних даних X на дві частини: «змістовну» і «шум». Хемометрика – це хімічна дисципліна, яка застосовує математичні, статистичні і інші методи, засновані на формальній логіці, для побудови або відбору оптимальних методів вимірювання і планів експерименту, а також для отримання найбільш важливої інформації при аналізі експериментальних даних.

Факторний аналіз також застосовується для вивчення взаємозв'язків між значеннями змінних, що впливають на технологічні процеси, апарати.

Програма проста у використанні і не потребує спеціальних навиків робо-

ти з середовищем програмування чи математичними пакетами, тому її може використовувати будь-який дослідник для аналізу зібраних статистичних даних та отримання графіків, що роблять наявну інформацію більш наочною для висновків.

Вхідними даними для проведення розрахунків є статистичний матеріал, що містить велику кількість даних, які характеризують однорідні об'єкти та об'єднані за багатьма параметрами. Всі ці дані є статистичною інформацією конкретних об'єктів і можуть бути використані для аналізу динаміки їх розвитку.

ВИСНОВКИ

В результаті виконання атестаційної роботи була розглянута проблема застосування методів багатовимірного статистичного аналізу для дослідження динаміки розвитку економічних систем на прикладі регіонів України.

В першому розділі була розглянута система «Реалізація методів багатовимірного статистичного аналізу» та розроблені морфологічна, інформаційна, функціональна моделі системи. За допомогою методу аналізу ієрархій було обґрунтовано вибір методів багатовимірного статистичного аналізу для реалізації задачі дослідження динаміки економічних систем.

У другому розділі сформульовані основні теоретичні відомості для обраних методів, детально описано та алгоритми розв'язання задачі.

Для обчислень та графічного подання результатів використовувався математичний пакет Wolfram Mathematica 11.2. У даному програмному середовищі було реалізовано програмний продукт, який дозволив полегшити та прискорити обчислення та приділити основну увагу аналізу отриманих графічних результатів. Аналізуючи отримані графіки було зроблено висновки щодо динаміки розвитку регіонів України.

Результати, отримані в роботі, можуть знайти своє застосування в різноманітних економічних, соціальних, психологічних та інших дослідженнях, де виникає потреба зниження вимірності та класифікації об'єктів, що описуються великою кількістю різноманітних параметрів.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений. Москва : Статистика, 1974. 240 с.
2. Дубров А. М., Мхитарян В. С., Трошин Л.И. Многомерные статистические методы. Москва : Финансы и статистика, 2003. 352 с.
3. Прикладная статистика: Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; Под ред. С. А. Айвазяна. Москва : Финансы и статистика, 1989. 471 с.
4. Лоули Д., Максвелл А. Факторный анализ как статистический метод. Москва : Мир, 1967. 145 с.
5. Иберла К. Факторный анализ. Пер. с нем. В.М. Ивановой. Москва : Статистика, 1980. 398 с.
6. Гибкіна Н. В., Сидоров М. В, Стороженко О. В. Класифікація країн Європейського Союзу за основними соціально-економічними показниками методом головних компонент // Математичні моделі та новітні технології управління економічними та технічними системами: монографія. За заг. ред. В. О. Тимофєєва, І. В. Чумаченко. Харків : ФОП Мезіна В.В., 2017. С. 116 – 133.
7. Гибкіна Н. В., Сидоров М. В., Стороженко О. В. Застосування факторного аналізу до класифікації країн Європейського Союзу за показниками соціально-економічного розвитку // Математичні моделі та новітні технології управління економічними та технічними системами: монографія. За заг. ред. В. О. Тимофєєва, І. В. Чумаченко. Харків : ФОП Панов А.М., 2018. С. 125 – 143.
8. Мандель И. Д. Кластерный анализ. Москва : Финансы и статистика, 1988. 176 с.
9. Прикладная статистика: Основы моделирования и первичная обработка данных / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин; под ред. С. А. Айвазяна. Москва : Финансы и статистика, 1983. 607 с.
10. Андерсон Т. Введение в многомерный статистический анализ. Москва : ГИФМЛ, 1963. 500 с.

11. Gybkina N., Sidorov M., Storozhenko O. To classification of the countries of the European Union for some social indicators // Big Data processing: methods, models and information technologies : monograph. Edited by Oleg I. Pursky. Shioda GmbH, Steyr, Austria, 2019. P. 49 – 58.

12. Gybkina N., Sidorov M., Storozhenko O. Research of the economic development structure of the regions of Ukraine with the methods of multi-dimensional statistical analysis // Information systems and innovative technologies in project and program management: Collective monograph edited by I. Linde, I. Chumachenko, V. Timofeyev. Riga : ISMA, 2019. P. 280 293.

13. Швайка Л. А. Державне регулювання економіки. Київ : Знання, 2008. 463 с.

14. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. Москва : Мир, 1982. 488 с.

15. Регіональна економіка: підручник / За ред. проф. Є.П. Качана. Тернопіль : Тернопіль. нац. екон. ун-т, 2008. 800 с.

16. Шаблій О. І. Україна – об'єкт дослідження політичної географії // Суспільна географія. Книга третя. Проблеми теоретичних досліджень, геогуманістики, географічної освіти [Вибрані праці] / Олег Шаблій. Львів : ЛНУ імені Івана Франка, 2017. С. 154-164.

17. Державна служба статистики України [Електронний ресурс] : <http://www.ukrstat.gov.ua> (23.09.2019)