

Міністерство освіти і науки України  
Харківський національний університет імені В.Н. Каразіна

Кафедра статистики, обліку та аудиту

**«ЗАТВЕРДЖУЮ»**

Завідувач кафедри



Оксана НЕСТЕРЕНКО

Протокол № 1 від “26” серпня 2024 р.

## **НАВЧАЛЬНО-МЕТОДИЧНИЙ КОМПЛЕКС**

**дисципліни «Методи класифікації даних в пакеті Statistica»**

для студентів \_\_\_\_\_ денної (заочної) форми навчання

рівень вищої освіти \_\_\_\_\_ другий (магістерський)

галузь знань \_\_\_\_\_ 05 «Соціальні та поведінкові науки»

спеціальність \_\_\_\_\_ 051 «Економіка»

освітня програма \_\_\_\_\_ «Бізнес-аналітика та міжнародна статистика»  
\_\_\_\_\_ «Економічна аналітика та статистика»

Розроблено:

д.е.н., професор, професор ЗВО кафедри статистики, обліку та аудиту

**Корепанов Олексій Сергійович**

2024/2025 навчальний рік

## **ЗМІСТ**

1. Робоча програма навчальної дисципліни;
2. Методичні рекомендації для виконання курсових, лабораторних, практичних, самостійних робіт тощо;
3. Приклади завдань семестрових екзаменів (письмових залікових робіт).

# **1. РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Міністерство освіти і науки України  
Харківський національний університет імені В. Н. Каразіна

Кафедра статистики, обліку та аудиту

“ЗАТВЕРДЖУЮ”

Декан економічного факультету

Віталій ДЯЧЕК



2024 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

“Методи класифікації даних в пакеті Statistica”

(шифр і назва навчальної дисципліни)

рівень вищої освіти другий (магістерський)  
галузь знань 05 «Соціальні та поведінкові науки»  
спеціальність 051 «Економіка»  
освітня програма «Бізнес-аналітика та міжнародна статистика»  
вид дисципліни за вибором  
факультет економічний

2024/2025 навчальний рік



Програму рекомендовано до затвердження вченою радою факультету (інституту, центру)  
“27” серпня 2024 року, протокол № 11

РОЗРОБНИКИ ПРОГРАМИ:

Олексій КОРЕПАНОВ, професор закладу вищої освіти кафедри статистики, обліку та аудиту  
ХНУ імені В. Н. Каразіна, докт. екон. наук, професор.  
(вказати авторів, їхні посади, наукові ступені та вчені звання)

Програму схвалено на засіданні кафедри статистики, обліку та аудиту

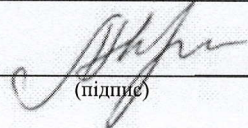
Протокол від “26” серпня 2024 року № 1

Завідувач кафедри статистики, обліку та аудиту

  
(підпис) Оксана НЕСТЕРЕНКО

Програму погоджено з гарантом освітньо-професійної програми (керівником проектної групи)  
«Бізнес-аналітика та міжнародна статистика»  
назва освітньої програми

Гарант освітньо-професійної програми  
(керівник проектної групи) «Бізнес-аналітика та міжнародна статистика»


  
(підпис) Юлія ЛАЗЕБНИК

Програму погоджено науково-методичною комісією  
економічного факультету

назва факультету, для здобувачів вищої освіти якого викладається навчальна дисципліна

Протокол від “27” серпня 2024 року № 1

Голова науково-методичної комісії

  
(підпис) Дар'я ЗАГОРСЬКА

## ВСТУП

Програма навчальної дисципліни “**Методи класифікації даних в пакеті Statistica**” складена відповідно до освітньо-професійної програми підготовки **магістра** спеціальності (напряму) 051 «Економіка» освітня програма «Бізнес-аналітика та міжнародна статистика»

### 1. Опис навчальної дисципліни

**1.1. Метою** викладання навчальної дисципліни є формування системи знань і практичних навичок застосування багатовимірної класифікації об'єктів для аналізу складних соціально-економічних явищ та процесів.

**1.2. Основним завданням** вивчення дисципліни є вивчення методів рішення задач класифікації або дискримінації об'єктів спостереження згідно визначеним ознакам.

Після завершення курсу студент має набути такі **загальні, фахові та спеціалізовані компетентності**:

- ЗК2. Здатність до абстрактного мислення, аналізу та синтезу.
- ЗК8. Здатність проводити дослідження на відповідному рівні.
- СК1. Здатність застосовувати науковий, аналітичний, методичний інструментарій для обґрунтування стратегії розвитку економічних суб'єктів та пов'язаних з цим управлінських рішень.
- СК3. Здатність збирати, аналізувати та обробляти статистичні дані, науково-аналітичні матеріали, які необхідні для розв'язання комплексних економічних проблем, робити на їх основі обґрунтовані висновки.
- СК4. Здатність використовувати сучасні інформаційні технології, методи та прийоми дослідження економічних та соціальних процесів, адекватні встановленим потребам дослідження.
- ДСК2. Здатність до бізнес-аналітики, статистичного забезпечення та обґрунтування управлінських рішень щодо ведення бізнесу (здатність проводити системний аналіз зовнішнього та внутрішнього бізнес-середовища діяльності суб'єктів господарювання; здатність визначати вплив макро- і мікро-економічних елементів на бізнес; здатність використовувати відповідні інструменти для аналізу бізнес-середовища; здатність проводити фінансовий та кредитний аналіз).
- ДСК3. Здатність використовувати професійно-профільовані знання й уміння щодо практичного використання комп'ютерних технологій в галузі бізнес-аналітики та міжнародної статистики. Уміння розробляти і впроваджувати інформаційні системи.

**1.3.** Кількість кредитів – 5.

**1.4.** Загальна кількість годин – 150.

**1.5.** Характеристика навчальної дисципліни

За вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	1-й
Семестр	
2-й	2-й
Лекції	
32 год.	6 год.

Практичні, семінарські заняття	
32 год.	4 год.
Лабораторні заняття	
-	-
Самостійна робота	
86 год.	140 год.
у тому числі, індивідуальні завдання	
-	-

### 1.6. Заплановані результати навчання.

Згідно з вимогами освітньо-професійної (освітньо-наукової) програми студенти повинні досягти таких **результатів навчання**:

– РН1. Формулювати, аналізувати та синтезувати рішення науково-практичних проблем.

– РН8. Збирати, обробляти та аналізувати статистичні дані, науково-аналітичні матеріали, необхідні для вирішення комплексних економічних завдань.

– РН9. Приймати ефективні рішення за невизначених умов і вимог, що потребують застосування нових підходів, методів та інструментарію соціально-економічних досліджень.

– РН10. Застосовувати сучасні інформаційні технології та спеціалізоване програмне забезпечення у соціально-економічних дослідженнях та в управлінні соціально-економічними системами.

– ДРН2. Демонструвати навички проведення бізнес-аналізу, оцінювання стану та тенденцій розвитку економічних явищ і процесів, факторів формування зовнішнього (та внутрішнього) бізнес-середовища функціонування підприємств.

– ДРН3. Демонструвати спроможність до статистичного забезпечення та обґрунтування управлінських рішень у бізнесі на основі системного аналізу, багатовимірного статистичного аналізу, моделювання та прогнозування бізнес-процесів.

– ДРН4. Розв'язувати складні спеціалізовані задачі та практичні проблеми у галузі економіки, бізнес-аналітики та міжнародної статистики. Володіти методикою проведення фінансового та кредитного аналізу, оперувати аналітичною інформацією, здійснювати зіставлення параметрів економічних процесів, виявлення змін і аналізу причин їх виникнення, діагностування підприємства і позиціонування його у бізнес-середовищі.

– ДРН5. Професійно застосовувати статистичні прикладні пакети програм для розв'язання науково-дослідних та прикладних аналітичних завдань у бізнес-сфері.

Після завершення курсу студент отримує наступні

*знання*:

- ідентифікувати мету та задачі, які вирішують методи кластерного аналізу;
- визначати та описувати типи задач класифікації та варіанти кінцевих прикладних цілей дослідження для даного типу задач класифікації;
- демонструвати розуміння математичного апарату методів класифікації;
- усвідомлювати доцільність і ефективність застосування тих або інших методів класифікації;

*вміння*:

- застосовувати навички проведення систематизації задач класифікації відповідно до кінцевих прикладних цілей дослідження;
- визначати критерії якості, цільову функцію, значення якої дозволять зіставити різні схеми класифікації та здійснити вибір алгоритму класифікації;
- демонструвати навички використання різних методів багатомірної класифікації для вирішення практичних завдань.

## 2. Тематичний план навчальної дисципліни

### Розділ 1. Теоретичні засади багатомірної класифікації об'єктів та методика проведення ієрархічних та ітеративних методів класифікації

#### Тема 1. Основні поняття методів багатомірної класифікації.

Поняття класифікації. Формалізація загальної задачі класифікації. Комбінаційні угруповання та їх безперервне узагальнення. Проста типологізація. Зв'язна упорядкована типологізація. Структурна типологізація. Класифікація динамічних траєкторій розвитку систем. Типологізація математичних ставлень задач класифікації.

#### Тема 2. Кластерний аналіз. Ієрархічні методи класифікації.

Загальні поняття про кластерний аналіз. Характеристика методів кластерного аналізу. Відстань між об'єктами та міри близькості: евклідова відстань, зважена евклідова відстань, Хемінгова відстань, відстань Чебишева, ступінна відстань.

Міра схожості для об'єднання двох кластерів. Метод “найближчого сусіда”. Метод “дальнього сусіда”. Метод середнього зв'язку. Метод медіанного зв'язку.

Методи ієрархічного кластерного аналізу: метод одиночного зв'язку, метод повних зв'язків, метод середнього зв'язку, метод Уорда.

#### Тема 3. Кластерний аналіз. Ітеративні методи класифікації.

Загальні поняття про ітеративні методи кластерного аналізу. Математичне описання методу k-середніх. Модифікації методу k-середніх.

Алгоритм обчислювальних процедур ітеративних методів класифікації.

Модифікації методу розшуку згущень. Схема ітеративного алгоритму типу “форель”. Способи добору радіусу сфери. Оцінка сталості угруповань об'єктів за методом розшуку згущень.

Поняття про функціонали або критерії якості. Найбільш поширені функціонали якості: сума квадратів відстаней до центрів класів, сума внутрішньокласових відстаней між об'єктами, сумарна внутрішньокласова дисперсія.

### Розділ 2. Основи дискримінантного аналізу та аналіз даних методами нечіткої кластеризації

#### Тема 4. Дискримінантний аналіз.

Загальні положення дискримінантного аналізу. Дискримінантні змінні та шкала відношень. Геометрична інтерпретація дискримінантної функції та дискримінантних змінних. Центри множин, які розділяються, та константа дискримінації.

Розрахунок коефіцієнтів дискримінантної функції. Алгоритм класифікації при наявності двох та більше навчаючих вибірок. Взаємозв'язок між дискримінантними змінними та дискримінантними функціями.

#### Тема 5. Аналіз даних методами нечіткої кластеризації.

Постановка задачі нечіткої кластеризації. Концептуальний зв'язок між кластерним аналізом і теорією нечітких множин. Методи нечіткої кластеризації

Алгоритм розв'язування задачі нечіткої кластеризації.

Виконання алгоритму FCM в системі MATLAB.

Приклади реалізації алгоритму FCM.

#### Тема 6. Практика застосування методів класифікації з використання ППП.

Приклади реалізації різних алгоритмів класифікації у ППП Statistica.



### 3. Структура навчальної дисципліни\*

Назви розділів і тем	Кількість годин											
	денна форма						заочна форма					
	усьо- го	у тому числі					усьо- го	у тому числі				
		л	п	лаб.	інд.	с.р.		л	п	дист	інд.	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13
<b>Розділ 1. Теоретичні засади багатомірної класифікації об'єктів та методика проведення ієрархічних та ітеративних методів класифікації</b>												
Тема 1. Основні поняття методів багатомірної класифікації	18	4	4	-	-	10	22	1	-	-	-	21
Тема 2. Кластерний аналіз. Ієрархічні методи класифікації	27	6	6	-	-	15	22	0,5	0,5	-	-	21
Тема 3. Кластерний аналіз. Ітеративні методи класифікації	27	6	6	-	-	15	22	0,5	0,5	-	-	21
Разом за розділом 1	72	16	16	-	-	40	66	2	1	-	-	63
<b>Розділ 2. Основи дискримінантного аналізу та аналіз даних методами нечіткої кластеризації</b>												
Тема 4. Дискримінантний аналіз	27	6	6	-	-	15	28,0	1	1		-	25
Тема 5. Аналіз даних методами нечіткої кластеризації	27	6	6	-	-	15	28,5	1,5	1		-	26
Тема 6. Практика застосування методів класифікації з використання ППП	24	4	4	-	-	16	28,5	1,5	1		-	26
Разом за розділом 2	78	16	16	-	-	46	84	4	3		-	77
<b>Усього годин</b>	<b>150</b>	<b>32</b>	<b>32</b>	-	-	<b>86</b>	<b>150</b>	<b>6</b>	<b>4</b>		-	<b>140</b>

\* структура навчальної дисципліни (години на вивчення) може змінюватись у зв'язку із змінами законодавства, ступенем засвоєння матеріалу студентами, коригуванням структури навчального року (графіку навчального процесу) тощо. Можлива зміна послідовності розкриття навчального матеріалу в межах окремого розділу або теми, але так, щоб не порушувалась загальна логіка його викладу.

#### 4. Теми практичних занять

№ з/п	Назва теми	Зміст практичного заняття	денна	заочна
1	Основні поняття методів багатомірної класифікації	Визначити сутність класифікації об'єктів за допомогою методів багатомірної класифікації.	4	-
2	Кластерний аналіз. Ієрархічні методи класифікації	За вихідними даними здійснити класифікацію підприємств згідно ієрархічного агломеративного алгоритму з використанням звичайної та зваженої евклідової відстані, а також принципів: „найближчого” і „дальнього” сусіда, центру ваги і середнього зв'язку.	6	0,5
3	Кластерний аналіз. Ітеративні методи класифікації	За вихідними даними здійснити класифікацію об'єктів за допомогою методу к-середніх. За допомогою одного з функціоналів якості дати оцінку якості розподілу на угруповання.	6	0,5
4	Дискримінантний аналіз	За вихідними даними, що характеризують сільськогосподарське виробництво, для регіонів України, які підлягають дискримінації, обчислити значення дискримінантної функції та провести класифікацію регіонів.	6	1
5	Аналіз даних методами нечіткої кластеризації	Розв'язування задач нечіткої кластеризації за алгоритмом FCM.	6	1
6	Практика застосування методів класифікації з використання ППП	Застосування методів класифікації з використання ППП	4	1
	<b>Разом</b>		<b>32</b>	<b>4</b>

#### 5. Завдання для самостійної роботи

Самостійна робота студента включає: опрацювання навчального матеріалу, підготовку до практичних занять, виконання індивідуальних завдань, підготовку до усіх видів контролю.

№ з/п	Види, зміст самостійної роботи (назва теми)	Кількість годин	
		денне	заочне
1	Проробка конспекту лекції, вивчення рекомендованої літератури по темах, які викладаються на лекціях	20	35
2	Підготовка до практичних занять	22	41
3	Підготовка до поточного контролю	22	25
4	Підготовка до підсумкового контролю (екзамену)	22	39
	<b>Разом</b>	<b>86</b>	<b>140</b>

## 6. Індивідуальні завдання

Не передбачено.

## 7. Методи навчання

Основними відмінностями активних та інтерактивних методів навчання від традиційних вважаються не лише методика і техніка викладання, але і висока ефективність навчального процесу, яка проявляється через:

- високу мотивацію студентів;
- закріплення теоретичних знань на практиці;
- підвищення самосвідомості студентів;
- вироблення здатності ухвалювати самостійні рішення;
- вироблення здатності приймати колективні рішення;
- вироблення здатності до соціальної інтеграції;
- придбання навичок вирішення конфліктів;
- розвиток здатності до компромісів.

При викладанні дисципліни для активізації навчального процесу передбачено використання таких сучасних навчальних технологій як кейс-метод, проблемні лекції, міні-лекції, робота в малих групах.

**Кейс-метод** – метод аналізу конкретних ситуацій, який дає змогу наблизити процес навчання до реальної практичної діяльності спеціалістів і передбачає розгляд виробничих, управлінських та інших ситуацій, інцидентів у процесі вивчення навчального матеріалу.

**Проблемні лекції** спрямовані на розвиток логічного мислення студентів. Коло питань теми лекції обмежується кількома ключовими моментами і увага студентів має бути сконцентрована на матеріалі, що не знайшов відображення в підручниках. При проведенні проблемної лекції має бути розданий друкований матеріал і виділені головні висновки з питань, що розглядаються.

При викладанні лекційного матеріалу студентам пропонуються питання для самостійного розмірковування. При цьому лектор ставить питання, які спонукають студента шукати вирішення проблемної ситуації. Така ситуація змушує студентів сконцентруватися і почати активно мислити в пошуках правильної відповіді.

На початку проведення проблемної лекції необхідно чітко сформулювати проблему, яку необхідно вирішити студентам. При викладанні лекційного матеріалу слід уникати прямої відповіді на поставлені питання, а висвітлювати матеріал таким чином, щоб отриману інформацію студент міг використати при вирішенні проблеми.

**Міні-лекції** передбачають викладення навчального матеріалу за короткий проміжок часу й характеризуються значною ємністю, складністю логічних побудов, образів, доведень та узагальнень.

На початку проведення міні-лекції за обраними темами лектор акцентує увагу студентів на необхідності представити викладений лекційний матеріал у так званому структурно-логічному вигляді. На розгляд виносять питання, які зафіксовані у плані лекцій, але викладають їх стисло. Лекційне заняття, проведене у такий спосіб, пробуджує у студента активність та увагу до сприйняття матеріалу, а також спрямовує його на використання системного підходу при відтворенні інформації, яку він отримав під час лекції.

Проблемні лекції та міні-лекції доцільно поєднувати з такою формою активізації навчального процесу, як робота в малих групах.

**Дискусії** передбачають обмін думками та поглядами учасників щодо певної теми, а також розвивають мислення, допомагають формувати погляди і переконання, виробляють вміння формулювати думки й висловлювати їх, вчать оцінювати пропозиції інших людей, критично підходити до власних поглядів.

Розподіл форм та методів активізації процесу навчання навчальної дисципліни представлено в табл.

### Використання навчальних технологій для активізації процесу навчання

Тема	Практичне застосування навчальних технологій
<b>Розділ 1. Теоретичні засади багатомірної класифікації об'єктів та методика проведення ієрархічних та ітеративних методів класифікації</b>	
Тема 1. Основні поняття методів багатомірної класифікації	Лекція, дискусія
Тема 2. Кластерний аналіз. Ієрархічні методи класифікації	Проблемна лекція з питання: Застосування кластерного аналізу. Ієрархічні методи класифікації
Тема 3. Кластерний аналіз. Ітеративні методи класифікації	Проблемна лекція з питання: Застосування кластерного аналізу. Ітеративні методи класифікації
Разом за розділом 1	
<b>Розділ 2. Основи дискримінантного аналізу та аналіз даних методами нечіткої кластеризації</b>	
Тема 4. Дискримінантний аналіз	Лекція, дискусія
Тема 5. Аналіз даних методами нечіткої кластеризації	Лекція, дискусія
Тема 6. Практика застосування методів класифікації з використання ППП	Кейс-метод – Практика застосування методів класифікації з використання ППП

## 8. Методи контролю

Оцінювання знань, вмінь та навичок студентів включає ті види занять, які згідно з програмою навчальної дисципліни “Методи класифікації даних в пакеті Statistica” передбачають лекційні, практичні заняття, самостійну роботу та виконання індивідуального науково-дослідного завдання.

Перевірка та оцінювання знань студентів проводиться в наступних формах:

- поточне оцінювання роботи і знань студентів під час практичних занять;
- складання проміжного контролю знань за розділами (тестування);
- оцінювання виконання та захист індивідуального науково-дослідного завдання;
- складання екзамену.

**Поточне оцінювання** знань студентів здійснюється під час проведення практичних і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Об'єктами поточного контролю є:

- активність та результативність роботи студента протягом семестру над вивченням програмного матеріалу дисципліни;
- відвідування занять;
- виконання контрольного завдання;
- складання проміжного контролю знань (тестування).

Контроль систематичного виконання самостійної роботи та активності на практичних заняттях проводиться за такими критеріями:

- розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються;
- ступінь засвоєння фактичного матеріалу навчальної дисципліни;
- ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;
- уміння поєднувати теорію з практикою при розгляді практичних ситуацій, розв'язанні задач, проведенні розрахунків при виконанні індивідуальних завдань, та завдань, винесених на розгляд в аудиторії;



– оволодіння методами економіко-статистичної обробки даних з використанням комп'ютерних технологій;

– логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки.

Оцінювання знань студента під час виконання завдань для самостійної роботи проводиться за чотирирівневою шкалою. Оцінка «відмінно» ставиться за умови відповідності виконаного завдання студента або його усної відповіді до всіх зазначених критеріїв. Відсутність тієї чи іншої складової знижує оцінку.

При оцінюванні практичних занять увага приділяється також їх якості та самостійності, своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу). Якщо якась із вимог не буде виконана, то оцінка буде знижена.

*Проміжний контроль (тестування)* рівня знань передбачає виявлення опанування студентом лекційного матеріалу та вміння застосування його для вирішення практичної ситуації і проводиться у вигляді тестування. При цьому тестове завдання може містити як запитання, що стосуються суто теоретичного матеріалу, так і запитання, спрямовані на вирішення невеличкого практичного завдання.

## 9. Схема нарахування балів та критерії оцінювання навчальних досягнень

Структура засобів контролю та розподіл балів із дисципліни “Методи класифікації даних в пакеті Statistica” наведена в табл.

Узагальнена схема нарахування балів (денна форма навчання)

Поточний контроль, самостійна робота, індивідуальні завдання									Експе- наційна робота	Сума
Поточне оцінювання					Проміжний контроль (тестування)	Контрольна робота, передбачена навчальним планом	Індиві- дуальне завдання	Разом		
Розділ 1			Розділ 2							
T1	T2	T3	T4	T5						
2	2	2	2	2	20	30	-	60	40	100

T1, T2 ... – теми розділів.

Узагальнена схема нарахування балів (заочна форма навчання)

Поточний контроль, самостійна робота, індивідуальні завдання										Експе- наційна робота	Сума
Поточне оцінювання					Проміжний контроль (тестування)	Контрольна робота, передбачена навчальним планом	Індиві- дуальне завдання	Разом			
Розділ 1			Розділ 2								
T1	T2	T3	T4	T5							
2	2	2	2	2	20	30	-	60	40	100	

T1, T2 ... – теми розділів.

*Проміжний тестовий* контроль проводиться один раз на семестр. Загальна тривалість тестів – 1,5 години. Поточне тестування складається з 10 тестів і 1 практичної задачі. Одна правильна відповідь на кожен з тестів дорівнює 1 бал, задача – 10 балів. Тестове завдання містить запитання одиничного і множинного вибору різного рівня складності.

Тести можуть бути застосовані як з метою контролю, так і для закріплення теоретичних знань і практичних навичок.

Тести для проміжного контролю обираються із загального переліку тестів за відповідними темами.

*Підсумковий контроль* за курсом - у формі екзамену.

Для допуску до складання підсумкового контролю (екзамену/заліку) здобувач вищої освіти повинен набрати **не менше 10 балів** з навчальної дисципліни під час поточного контролю, самостійної роботи, індивідуального завдання/контрольної роботи.

Екзамен проводиться у відповідності до екзаменаційних білетів, які містять чотири завдання (два теоретичних питання, одне практичне завдання - задача і тести). Вони дають можливість здійснити оцінювання знань студента за дисципліною.

*Критерії оцінювання результативності роботи студентів при виконанні самостійної роботи*

Якісними критеріями оцінювання виконання індивідуальних завдань студентами є:

*1. Повнота виконання завдання:*

- Елементарна;
- Фрагментарна;
- Повна;
- Неповна.

*2. Рівень самостійності студента*

- під керівництвом викладача;
- консультація викладача;
- самостійно.

*3. Сформованість навчально-інформаційних умінь* (роботи з підручником, володіння різними способами читання, складання плану, рецензій, конспекту, вміння користуватися бібліотекою, спостереження, експеримент тощо)

*4. Сформованість навчально-інтелектуальних умінь* (визначення понять, аналіз, синтез, порівняння, класифікація, систематизація, узагальнення, абстрагування, вміння відповідати на запитання, виконувати творчі завдання тощо);

*5. Рівень сформованості фахових методичних вмінь* (вміння застосовувати на практиці набуті знання):

- низький – володіння умінням здійснювати первинну обробку навчальної інформації без подальшого її аналізу;
- середній – уміння вибирати відомі способи дій для виконання фахових завдань;
- достатній – застосовує набуті знання у стандартних практичних ситуаціях;
- високий – володіння умінням творчо-пошукової діяльності.

*Критерії оцінювання здобувачів вищої освіти за відповіді на питання поточного контролю*

Завданням поточного контролю є перевірка розуміння та засвоєння певного матеріалу, вироблених навичок проведення розрахункових робіт, умінь самостійно опрацьовувати тексти, здатності осмислити зміст теми чи розділу, умінь публічно чи письмово представити певний матеріал (презентація).

Якісними критеріями оцінювання виконання завдань поточного контролю є:

*1. Повнота відповіді або виконання завдання:*

- елементарна;
- фрагментарна;
- повна;
- неповна.

*2. Рівень сформованості логічних умінь:*

- елементарні дії;
- операція, правило, алгоритм;
- правила визначення понять;
- формулювання законів і закономірностей;
- структурування суджень, доводів, описів.

### *Критерії оцінювання здобувачів вищої освіти за виконання контрольної роботи*

Контрольна робота студента складається з двох частин: перша – опрацювання теоретичного питання, друга – виконання практичного завдання. Максимальна оцінка за виконання завдань контрольної роботи – 30 балів.

Виконання завдань оцінюється за такими критеріями:

- 1) теоретичне питання:
  - повнота й ґрунтовність викладу;
  - аргументованість тверджень;
  - суб'єктне усвідомлення змісту;
  - термінологічна коректність;
- 2) практичне завдання:
  - технологічна грамотність;
  - методична грамотність;
  - обґрунтованість висновків;
  - правильність оформлення.

#### **Шкала оцінювання контрольної роботи**

Кількість балів	Теоретичне питання	Практичне завдання
21-30	Повне засвоєння та суб'єктне усвідомлення матеріалу. Твердження чітко аргументовані. Продemonстровано термінологічну грамотність	Продemonстровано методичну й технологічну грамотність. Методичні рішення обґрунтовано. Оформлення відповідає вимогам.
11-20	Часткове засвоєння матеріалу, суб'єктне його не усвідомлення. Нечітка аргументація тверджень. Часткова термінологічна некоректність.	Наявність незначних методичних і технологічних помилок, а також помилок в оформленні роботи. Методичні рішення обґрунтовані частково.
0-10	Теоретичний матеріал не засвоєно. Аргументація відсутня. Термінологічна неграмотність.	Методична й технологічна неграмотність. Неправильне оформлення роботи.

### ***Критерії оцінювання на екзамені***

Оцінювання знань студента проводиться за чотирирівневою шкалою (відмінно, добре, задовільно, незадовільно). За екзамен студент може отримати максимум 40 балів:

1. Для отримання оцінки «відмінно» (35-40 балів) студент повинен:

- укластися у встановлений строк підготовки відповіді;
- викласти теоретичний матеріал чітко, коротко, зв'язно й обґрунтовано;
- навести вірне рішення задачі та тестів.

2. Для одержання оцінки «добре» (25-34 бала) студент повинен:

- укластися у встановлений строк підготовки відповіді;
- викласти теоретичний матеріал зв'язно й обґрунтовано;
- навести вірне рішення задачі;
- можливі помилки у відповідях на тести.

3. Для отримання оцінки «задовільно» (15-24 бала) студент повинен:

- викладати теоретичний матеріал у доступній для розуміння формі;
- можливі помилки при розв'язанні задачі та в тестах.

4. Оцінку «незадовільно» (1-14 балів) отримують студенти, відповіді яких можуть бути оцінені нижче вимог, сформульованих у попередніх пунктах.

Кожне завдання заліку оцінюється окремо. Загальна оцінка дорівнює сумі оцінок за всі завдання (види контролю).

**Підсумкова оцінка** з навчальної дисципліни визначається як сума балів, набраних здобувачем вищої освіти протягом семестру при виконанні контрольних заходів, передбачених програмою навчальної дисципліни та балів, набраних ним при складанні семестрового екзамену (поточне оцінювання за роботу протягом семестру, поточний контроль знань (тестування), контрольна робота та оцінка за результатами підсумкового екзамену).

Максимальна сума балів, яку може набрати здобувач вищої освіти при підсумковому семестровому контролі, складає 100.

#### Зведена шкала оцінювання роботи студентів з дисципліни

Види робіт	Максимум балів
Поточне оцінювання	10
Поточний контроль знань (тестування)	20
Контрольна робота	30
<b>РАЗОМ</b>	<b>60</b>
<b>Екзамен</b>	<b>40</b>
<b>ВСЬОГО</b>	<b>100</b>

У відповідності до набраних студентом балів оцінка знання матеріалу проводиться за чотирирівневою системою згідно з Методикою переведення показників успішності знань студентів.

#### Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка за національною шкалою
	для <b>чотирирівневої</b> шкали оцінювання
90-100	відмінно
80-89	добре
70-79	
60-69	
50-59	задовільно
1-49	незадовільно

## 10. Рекомендована література

### Основна література

1. Корепанов О.С. Методи класифікації даних в пакеті Statistica: дистанційний курс на платформі LMS Moodle. URL : <https://moodle.karazin.ua/course/view.php?id=867>.
2. Єріна А. М., Єрін Д. Л. Статистичне моделювання та прогнозування : навч. посіб. К. : КНЕУ, 2014. 348 с.
3. Моделювання та оптимізація систем: підручник / Дубовой В. М., Кветний Р. Н., Михальов О. І., Усов А. В. Вінниця : ПП «ТД«Еднльвейс», 2017. 804 с.
4. Статистичне моделювання та прогнозування: Практикум із використанням MS Excel та ППП Statistica / упоряд. Г. С. Корепанов, О. С. Корепанов, Ю. О. Лазебник, Т. Г. Чала, Д. І. Черненко // Економ. факультет ХНУ ім. В. Н. Каразіна, Х. : Видавництво ФОП Тагасв П. О., 2017. 92 с.



### Допоміжна література

1. Богомаз О.О., Корепанов О.С. Порівняльний аналіз стану навколишнього середовища за регіонами України. *Fundamental and Applied Research in the Modern World*: матер. IV Міжнар. наук.-практ. конф. (Бостон, США, 18-20 листопада 2020 р.). Бостон, США. 2020. С. 358–365. URL: <https://sci-conf.com.ua/wp-content/uploads/2020/11/FUNDAMENTAL-AND-APPLIED-RESEARCH-IN-THE-MODERN-WORLD-18-20.11.20.pdf>.
2. Корепанов О.С., Дашутін В.В. Методичні засади аналізу регіональної диференціації розвитку тваринництва в Україні. *Бізнес Інформ*. 2020. № 10. С. 214–220. URL: [https://www.business-inform.net/export\\_pdf/business-inform-2020-10\\_0-pages-214\\_220.pdf](https://www.business-inform.net/export_pdf/business-inform-2020-10_0-pages-214_220.pdf).
3. Корепанов О.С., Лазебник Ю.О., Чала Т.Г., Корнієнко В.В. Статистичний аналіз та прогнозування складових національного валютного ринку в умовах вторгнення РФ в Україну. *Бізнес Інформ*. Харків. 2023. № 1. С. 31–39. URL: [https://www.business-inform.net/export\\_pdf/business-inform-2023-1\\_0-pages-31\\_39.pdf](https://www.business-inform.net/export_pdf/business-inform-2023-1_0-pages-31_39.pdf).
4. Лазебник Ю.О., Корепанов О.С., Хавалиць Д.Л. Оцінка диференціації розвитку підприємств сільського господарства за регіонами України. *Інвестиції: практика та досвід*. 2019. № 21. С. 15–20. URL: [http://www.investplan.com.ua/pdf/21\\_2019/5.pdf](http://www.investplan.com.ua/pdf/21_2019/5.pdf)
5. Box G. Time Series Analysis : Forecasting and Control / G. Box, G. Jenkins, G. Reinsel / John Wiley & Sons. 2018.
6. Chala, T., Korepanov, O., Lazebnyk, I., Chernenko, D., Korepanov, G. Statistical modeling and forecasting of wheat and meslin export from Ukraine using singular spectral analysis. *Statistics in Transition New Series*. ISSN 1234-7655. Poland. 2023. 24(1), pp. 169–197. DOI <https://doi.org/10.59170/stattrans-2023-010>. URL: [https://sit.stat.gov.pl/SiT/2023/1/gus\\_sit\\_2023\\_00\\_tetyana\\_chala\\_oleksiy\\_korepanov\\_iuliia\\_lazebnyk\\_daryna\\_chernenko\\_georgii\\_korepanov\\_statistical\\_modelling\\_and\\_forecasting.pdf](https://sit.stat.gov.pl/SiT/2023/1/gus_sit_2023_00_tetyana_chala_oleksiy_korepanov_iuliia_lazebnyk_daryna_chernenko_georgii_korepanov_statistical_modelling_and_forecasting.pdf)
7. Kleinbaum D. Applied Regression Analysis and Other Multivariable Methods / D. Kleinbaum, L. Kupper, A. Nizam et al. Cengage Learning, 2013.
8. Korepanov O., Goncharenko Y., Mekhovich S., Dyshkantiuk O., Salamatina S., Rusavska V. Economic-Mathematical Modelling of the Distribution of Investments in the Tourism and Hospitality. *International Journal of Engineering and Advanced Technology (IJEAT)*. ISSN: 2249–8958. Vol.-9. Issue-1. October 2019. PP. 138-142. URL: <https://www.ijeat.org/wp-content/uploads/papers/v9i1/A1093109119.pdf>
9. Sukumar Arun. Applied Business Analysis / Arun Sukumar. Lucian Tipi ; Jayne Revill, 2016.

### 11. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. Нормативно-правова база / Кабінет Міністрів України; Урядовий портал. Офіц. веб-сайт. URL : <http://www.kmu.gov.ua/control/npd/list>.
2. Нормативно-правові акти / Урядовий портал. Єдиний веб-портал виконавчої влади України. URL : [http://www.me.gov.ua/control/uk/publish/category/main?cat\\_id=32854](http://www.me.gov.ua/control/uk/publish/category/main?cat_id=32854).
3. Статистична інформація / Державна служба статистики України; за ред. О. Г. Осауленка / Офіційний сайт державної служби статистики України. URL : <http://www.ukrstat.gov.ua> .
4. Загальноакадемічний портал наукової періодики. Національна академія наук України / Л. Костенко, Є. Копанєва // Національна бібліотека України імені В.І. Вернадського. URL : <http://www.nbuv.gov.ua/portal/natural/NTI/search.html>.

## 12. Особливості навчання за денною формою в умовах дії обставин непереборної сили

В умовах дії форс-мажорних обмежень освітній процес в університеті здійснюється відповідно до наказів/розпоряджень ректора/проректора або за змішаною формою навчання, або повністю дистанційно в синхронному режимі.

У разі проведення екзамену/заліку в дистанційній формі використовується LMS платформа Moodle з автентифікацією здобувача у режимі відеоконференції з Zoom. Реєстрація (допуск до складання) учасників освітнього процесу, а також обмін контрольними (підсумковими) завданнями та відповідями на них здійснюється винятково з корпоративної електронної пошти Харківського національного університету імені В. Н. Каразіна (домен – karazin.ua) із забезпеченням академічної доброчесності. При проведенні екзамену/заліку в дистанційній формі використовуються технічні і програмні засоби, які дозволяють забезпечити аудіо- і відео- фіксацію.

Режим доступу до курсу на LMS платформі Moodle:  
<https://moodle.karazin.ua/course/view.php?id=867>.

Режим доступу до курсу на LMS платформі Classroom:  
<https://classroom.google.com/u/0/c/MjYzMjQ4NjUyMTk1> (код курсу: e76r4kt).

Міністерство освіти і науки України  
Харківський національний університет імені В. Н. Каразіна

Кафедра статистики, обліку та аудиту

“ЗАТВЕРДЖУЮ”

Декан економічного факультету

Віталій ДЯЧЕК



“*Віталій Дячек*” 2024 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**“Методи класифікації даних в пакеті Statistica”**

(шифр і назва навчальної дисципліни)

рівень вищої освіти \_\_\_\_\_ другий (магістерський)  
галузь знань \_\_\_\_\_ 05 «Соціальні та поведінкові науки»  
спеціальність \_\_\_\_\_ 051 «Економіка»  
освітня програма \_\_\_\_\_ «Економічна аналітика та статистика»  
вид дисципліни \_\_\_\_\_ за вибором  
факультет \_\_\_\_\_ економічний

2024/2025 навчальний рік



Програму рекомендовано до затвердження вченою радою факультету (інституту, центру)  
“27” серпня 2024 року, протокол № 11

**РОЗРОБНИКИ ПРОГРАМИ:**

Олексій КОРЕПАНОВ, професор закладу вищої освіти кафедри статистики, обліку та аудиту  
ХНУ імені В. Н. Каразіна, докт. екон. наук, професор.  
(вказати авторів, їхні посади, наукові ступені та вчені звання)

Програму схвалено на засіданні кафедри статистики, обліку та аудиту

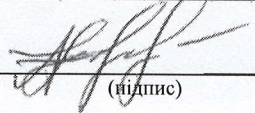
Протокол від “26” серпня 2024 року № 1

Завідувач кафедри статистики, обліку та аудиту

  
(підпис) Оксана НЕСТЕРЕНКО

Програму погоджено з гарантом освітньо-професійної програми (керівником проектної групи)  
«Економічна аналітика та статистика»  
назва освітньої програми

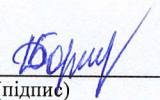
Гарант освітньо-професійної програми  
(керівник проектної групи) «Економічна аналітика та статистика»

  
(підпис) Олексій КОРЕПАНОВ

Програму погоджено науково-методичною комісією  
економічного факультету  
(назва факультету, для здобувачів вищої освіти якого викладається навчальна дисципліна)

Протокол від “27” серпня 2024 року № 1

Голова науково-методичної комісії

  
(підпис) Дар'я ЗАГОРСЬКА



## ВСТУП

Програма навчальної дисципліни “**Методи класифікації даних в пакеті Statistica**” складена відповідно до освітньо-професійної програми підготовки **магістра**

спеціальності (напряму) 051 «Економіка»

освітня програма «Економічна аналітика та статистика»

### 1. Опис навчальної дисципліни

**1.1. Метою** викладання навчальної дисципліни є формування системи знань і практичних навичок застосування багатовимірної класифікації об'єктів для аналізу складних соціально-економічних явищ та процесів.

**1.2. Основними завданнями** вивчення дисципліни є:

- вивчення методів рішення задач класифікації або дискримінації об'єктів спостереження згідно визначеним ознакам.

Після завершення курсу студент має набути такі **загальні, фахові та спеціалізовані компетентності**:

ЗК2. Здатність до абстрактного мислення, аналізу, синтезу та встановлення взаємозв'язків між явищами та процесами.

ЗК8. Здатність проводити дослідження та презентувати результати

СК1. Здатність застосовувати науковий, аналітичний, методичний інструментарій для управління економічною діяльністю.

СК3. Здатність збирати, аналізувати та обробляти статистичні дані, науково-аналітичні матеріали, які необхідні для розв'язання комплексних економічних завдань.

СК4. Здатність використовувати сучасні інформаційні технології та економіко-математичні методи і моделі для дослідження економічних та соціальних процесів.

ДСК3. Здатність до фундаментального аналізу, статистичного забезпечення управління для прийняття обґрунтованих рішень на загальнодержавному та регіональному рівнях.

**1.3.** Кількість кредитів – 5.

**1.4.** Загальна кількість годин – 150.

**1.5.** Характеристика навчальної дисципліни

За вибором	
Денна форма навчання	Заочна (дистанційна) форма навчання
Рік підготовки	
1-й	1-й
Семестр	
2-й	2-й
Лекції	
32 год.	6 год.
Практичні, семінарські заняття	
16 год.	2 год.
Лабораторні заняття	
-	-

Самостійна робота	
102 год.	142
у тому числі, індивідуальні завдання	
	-

### 1.6. Заплановані результати навчання.

Згідно з вимогами освітньо-професійної програми студенти повинні досягти таких **результатів навчання**:

РН.1. Формулювати, аналізувати та синтезувати рішення науково-практичних проблем.

РН.4. Розробляти соціально-економічні проекти та систему комплексних дій щодо їх реалізації з урахуванням їх цілей, очікуваних соціально-економічних наслідків, ризиків, законодавчих, ресурсних та інших обмежень.

РН.7. Обирати ефективні методи управління економічною діяльністю, обґрунтовувати пропонувані рішення на основі релевантних даних та наукових і прикладних досліджень.

РН.8. Збирати, обробляти та аналізувати статистичні дані, науково-аналітичні матеріали, необхідні для вирішення комплексних економічних завдань.

РН.9. Приймати ефективні рішення за невизначених умов і вимог, що потребують застосування нових підходів, методів та інструментарію соціально-економічних досліджень.

РН.10. Застосовувати сучасні інформаційні технології та спеціалізоване програмне забезпечення у соціально-економічних дослідженнях та в управлінні соціально-економічними системами.

РН.11. Визначати та критично оцінювати стан та тенденції соціально-економічного розвитку, формувати та аналізувати моделі економічних систем та процесів.

РН.12. Обґрунтовувати управлінські рішення щодо ефективного розвитку суб'єктів господарювання, враховуючи цілі, ресурси, обмеження та ризики.

РН.14. Розробляти сценарії і стратегії розвитку соціально-економічних систем.

РН.15. Організовувати розробку та реалізацію соціально-економічних проектів із врахуванням інформаційного, методичного, матеріального, фінансового та кадрового забезпечення.

РН.20. Професійно застосовувати статистичні прикладні пакети програм для розв'язання науково-дослідних та прикладних аналітичних завдань в економічній діяльності.

Після завершення курсу студент отримує наступні **знання**:

- ідентифікувати мету та задачі, які вирішують методи кластерного аналізу;
- визначати та описувати типи задач класифікації та варіанти кінцевих прикладних цілей дослідження для даного типу задач класифікації;
- демонструвати розуміння математичного апарату методів класифікації;
- усвідомлювати доцільність і ефективність застосування тих або інших методів класифікації;

**вміння**:

- застосовувати навички проведення систематизації задач класифікації відповідно до кінцевих прикладних цілей дослідження;
- визначати критерії якості, цільову функцію, значення якої дозволять зіставити різні схеми класифікації та здійснити вибір алгоритму класифікації;
- демонструвати навички використання різних методів багатомірної класифікації для вирішення практичних завдань.

## 2. Тематичний план навчальної дисципліни

### Розділ 1. Теоретичні засади багатомірної класифікації об'єктів та методика проведення ієрархічних та ітеративних методів класифікації

#### Тема 1. Основні поняття методів багатомірної класифікації.

Поняття класифікації. Формалізація загальної задачі класифікації. Комбінаційні угруповання та їх безперервне узагальнення. Проста типологізація. Зв'язна упорядкована типологізація. Структурна типологізація. Класифікація динамічних траєкторій розвитку систем. Типологізація математичних ставлень задач класифікації.

#### Тема 2. Кластерний аналіз. Ієрархічні методи класифікації.

Загальні поняття про кластерний аналіз. Характеристика методів кластерного аналізу. Відстань між об'єктами та міри близькості: евклідова відстань, зважена евклідова відстань, Хемінгова відстань, відстань Чебишева, ступінна відстань.

Міра схожості для об'єднання двох кластерів. Метод “найближчого сусіда”. Метод “дальнього сусіда”. Метод середнього зв'язку. Метод медіанного зв'язку.

Методи ієрархічного кластерного аналізу: метод одиночного зв'язку, метод повних зв'язків, метод середнього зв'язку, метод Уорда.

#### Тема 3. Кластерний аналіз. Ітеративні методи класифікації.

Загальні поняття про ітеративні методи кластерного аналізу. Математичне описання методу k-середніх. Модифікації методу k-середніх.

Алгоритм обчислювальних процедур ітеративних методів класифікації.

Модифікації методу розшуку згущень. Схема ітеративного алгоритму типу “форель”. Способи добору радіусу сфери. Оцінка сталості угруповань об'єктів за методом розшуку згущень.

Поняття про функціонали або критерії якості. Найбільш поширені функціонали якості: сума квадратів відстаней до центрів класів, сума внутрішньокласових відстаней між об'єктами, сумарна внутрішньокласова дисперсія.

### Розділ 2. Основи дискримінантного аналізу та аналіз даних методами нечіткої кластеризації

#### Тема 4. Дискримінантний аналіз.

Загальні положення дискримінантного аналізу. Дискримінантні змінні та шкала відношень. Геометрична інтерпретація дискримінантної функції та дискримінантних змінних. Центри множин, які розділяються, та константа дискримінації.

Розрахунок коефіцієнтів дискримінантної функції. Алгоритм класифікації при наявності двох та більше навчаючих вибірок. Взаємозв'язок між дискримінантними змінними та дискримінантними функціями.

#### Тема 5. Аналіз даних методами нечіткої кластеризації.

Постановка задачі нечіткої кластеризації. Концептуальний зв'язок між кластерним аналізом і теорією нечітких множин. Методи нечіткої кластеризації

Алгоритм розв'язування задачі нечіткої кластеризації.

Виконання алгоритму FCM в системі MATLAB.

Приклади реалізації алгоритму FCM.

#### Тема 6. Практика застосування методів класифікації з використання ППП.

Приклади реалізації різних алгоритмів класифікації у ППП Statistica.

### 3. Структура навчальної дисципліни\*

Назви розділів і тем	Кількість годин											
	денна форма						заочна форма					
	усь- го	у тому числі					усь- го	у тому числі				
		л	п	лаб.	інд.	с.р.		л	п	дист	інд.	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13
<b>Розділ 1. Теоретичні засади багатомірної класифікації об'єктів та методика проведення ієрархічних та ітеративних методів класифікації</b>												
Тема 1. Основні поняття методів багатомірної класифікації	16	4	-	-	-	12	22,00	1	-	-	-	21
Тема 2. Кластерний аналіз. Ієрархічні методи класифікації	25	6	4	-	-	15	22,75	0,5	0,25	-	-	22
Тема 3. Кластерний аналіз. Ітеративні методи класифікації	35	6	4	-	-	25	22,75	0,5	0,25	-	-	22
Разом за розділом 1	76	16	8	-	-	52	67,5	2	0,5	-	-	65
<b>Розділ 2. Основи дискримінантного аналізу та аналіз даних методами нечіткої кластеризації</b>												
Тема 4. Дискримінантний аналіз	25	6	4	-	-	15	26,50	1	0,5		-	25
Тема 5. Аналіз даних методами нечіткої кластеризації	25	6	4	-	-	15	28,00	1,5	0,5		-	26
Тема 6. Практика застосування методів класифікації з використання ППП	24	4	-	-	-	20	28,00	1,5	0,5		-	26
Разом за розділом 2	74	16	8	-	-	50	82,5	4	1,5		-	77
<b>Усього годин</b>	<b>150</b>	<b>32</b>	<b>16</b>	<b>-</b>	<b>-</b>	<b>102</b>	<b>150</b>	<b>6</b>	<b>2</b>		<b>-</b>	<b>142</b>

\* структура навчальної дисципліни (години на вивчення) може змінюватись у зв'язку із змінами законодавства, ступенем засвоєння матеріалу студентами, коригуванням структури навчального року (графіку навчального процесу) тощо. Можлива зміна послідовності розкриття навчального матеріалу в межах окремого розділу або теми, але так, щоб не порушувалась загальна логіка його викладу.

#### 4. Теми практичних занять

№ з/п	Назва теми	Зміст практичного заняття	денна	заочна
1	Основні поняття методів багатомірної класифікації	Визначити сутність класифікації об'єктів за допомогою методів багатомірної класифікації.	4	-
2	Кластерний аналіз. Ієрархічні методи класифікації	За вихідними даними здійснити класифікацію підприємств згідно ієрархічного агломеративного алгоритму з використанням звичайної та зваженої евклідової відстані, а також принципів: „найближчого” і „дальнього” сусіда, центру ваги і середнього зв'язку.	6	0,25
3	Кластерний аналіз. Ітеративні методи класифікації	За вихідними даними здійснити класифікацію об'єктів за допомогою методу k-середніх. За допомогою одного з функціоналів якості дати оцінку якості розподілу на угруповання.	6	0,25
4	Дискримінантний аналіз	За вихідними даними, що характеризують сільськогосподарське виробництво, для регіонів України, які підлягають дискримінації, обчислити значення дискримінантної функції та провести класифікацію регіонів.	6	0,5
5	Аналіз даних методами нечіткої кластеризації	Розв'язування задач нечіткої кластеризації за алгоритмом FCM.	6	0,5
6	Практика застосування методів класифікації з використання ППП	Застосування методів класифікації з використання ППП	4	0,5
	<b>Разом</b>		<b>32</b>	<b>2</b>

#### 5. Завдання для самостійної роботи

Самостійна робота студента включає: опрацювання навчального матеріалу, підготовку до практичних занять, виконання індивідуальних завдань, підготовку до усіх видів контролю.

№ з/п	Види, зміст самостійної роботи (назва теми)	Кількість годин	
		денне	заочне
1	Проробка конспекту лекції, вивчення рекомендованої літератури по темах, які викладаються на лекціях	20	30
2	Підготовка до практичних занять	20	30
3	Підготовка до поточного контролю	20	30
4	Підготовка до підсумкового контролю (екзамен)	22	22
5	Виконання індивідуального завдання (ІНДЗ) з курсу	20	30
	<b>Разом</b>	<b>102</b>	<b>142</b>



## 6. Індивідуальні завдання

Не передбачено.

## 7. Методи навчання

Основними відмінностями активних та інтерактивних методів навчання від традиційних вважаються не лише методика і техніка викладання, але і висока ефективність навчального процесу, яка проявляється через:

- високу мотивацію студентів;
- закріплення теоретичних знань на практиці;
- підвищення самосвідомості студентів;
- вироблення здатності ухвалювати самостійні рішення;
- вироблення здатності приймати колективні рішення;
- вироблення здатності до соціальної інтеграції;
- придбання навичок вирішення конфліктів;
- розвиток здатності до компромісів.

При викладанні дисципліни для активізації навчального процесу передбачено використання таких сучасних навчальних технологій як кейс-метод, проблемні лекції, міні-лекції, робота в малих групах.

**Кейс-метод** – метод аналізу конкретних ситуацій, який дає змогу наблизити процес навчання до реальної практичної діяльності спеціалістів і передбачає розгляд виробничих, управлінських та інших ситуацій, інцидентів у процесі вивчення навчального матеріалу.

**Проблемні лекції** спрямовані на розвиток логічного мислення студентів. Коло питань теми лекції обмежується кількома ключовими моментами і увага студентів має бути сконцентрована на матеріалі, що не знайшов відображення в підручниках. При проведенні проблемної лекції має бути розданий друкований матеріал і виділені головні висновки з питань, що розглядаються.

При викладанні лекційного матеріалу студентам пропонуються питання для самостійного розмірковування. При цьому лектор ставить питання, які спонукають студента шукати вирішення проблемної ситуації. Така ситуація змушує студентів сконцентруватися і почати активно мислити в пошуках правильної відповіді.

На початку проведення проблемної лекції необхідно чітко сформулювати проблему, яку необхідно вирішити студентам. При викладанні лекційного матеріалу слід уникати прямої відповіді на поставлені питання, а висвітлювати матеріал таким чином, щоб отриману інформацію студент міг використати при вирішенні проблеми.

**Міні-лекції** передбачають викладення навчального матеріалу за короткий проміжок часу й характеризуються значною ємністю, складністю логічних побудов, образів, доведень та узагальнень.

На початку проведення міні-лекції за обраними темами лектор акцентує увагу студентів на необхідності представити викладений лекційний матеріал у так званому структурно-логічному вигляді. На розгляд виносять питання, які зафіксовані у плані лекції, але викладають їх стисло. Лекційне заняття, проведене у такий спосіб, пробуджує у студента активність та увагу до сприйняття матеріалу, а також спрямовує його на використання системного підходу при відтворенні інформації, яку він отримав під час лекції.

Проблемні лекції та міні-лекції доцільно поєднувати з такою формою активізації навчального процесу, як робота в малих групах.

**Дискусії** передбачають обмін думками та поглядами учасників щодо певної теми, а також розвивають мислення, допомагають формувати погляди і переконання, виробляють вміння формулювати думки й висловлювати їх, вчать оцінювати пропозиції інших людей, критично підходити до власних поглядів.

Розподіл форм та методів активізації процесу навчання навчальної дисципліни представлено в табл.

### Використання навчальних технологій для активізації процесу навчання

Тема	Практичне застосування навчальних технологій
<b>Розділ 1. Теоретичні засади багатомірної класифікації об'єктів та методика проведення ієрархічних та ітеративних методів класифікації</b>	
Тема 1. Основні поняття методів багатомірної класифікації	Лекція, дискусія
Тема 2. Кластерний аналіз. Ієрархічні методи класифікації	Проблемна лекція з питання: Застосування кластерного аналізу. Ієрархічні методи класифікації
Тема 3. Кластерний аналіз. Ітеративні методи класифікації	Проблемна лекція з питання: Застосування кластерного аналізу. Ітеративні методи класифікації
Разом за розділом 1	
<b>Розділ 2. Основи дискримінантного аналізу та аналіз даних методами нечіткої кластеризації</b>	
Тема 4. Дискримінантний аналіз	Лекція, дискусія
Тема 5. Аналіз даних методами нечіткої кластеризації	Лекція, дискусія
Тема 6. Практика застосування методів класифікації з використання ППП	Кейс-метод – Практика застосування методів класифікації з використання ППП

## 8. Методи контролю

Оцінювання знань, вмінь та навичок студентів включає ті види занять, які згідно з програмою навчальної дисципліни “Методи класифікації даних в пакеті Statistica” передбачають лекційні, практичні заняття, самостійну роботу та виконання індивідуального науково-дослідного завдання.

Перевірка та оцінювання знань студентів проводиться в наступних формах:

- поточне оцінювання роботи і знань студентів під час практичних занять;
- складання проміжного контролю знань за розділами (тестування);
- оцінювання виконання та захист контрольного завдання;
- складання екзамену.

**Поточне оцінювання** знань студентів здійснюється під час проведення практичних і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Об'єктами поточного контролю є:

- активність та результативність роботи студента протягом семестру над вивченням програмного матеріалу дисципліни;
- відвідування занять;
- виконання індивідуального розрахункового завдання;
- складання проміжного контролю знань (тестування).

Контроль систематичного виконання самостійної роботи та активності на практичних заняттях проводиться за такими критеріями:

- розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються;
- ступінь засвоєння фактичного матеріалу навчальної дисципліни;
- ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;
- уміння поєднувати теорію з практикою при розгляді практичних ситуацій, розв'язанні задач, проведенні розрахунків при виконанні індивідуальних завдань, та завдань, винесених на розгляд в аудиторії;

– оволодіння методами економіко-статистичної обробки даних з використанням комп'ютерних технологій;

– логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки.

Оцінювання знань студента під час виконання завдань для самостійної роботи проводиться за чотирирівневою шкалою.

Оцінка «відмінно» ставиться за умови відповідності виконаного завдання студента або його усної відповіді до всіх зазначених критеріїв. Відсутність тієї чи іншої складової знижує оцінку.

При оцінюванні практичних занять увага приділяється також їх якості та самостійності, своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу). Якщо якась із вимог не буде виконана, то оцінка буде знижена.

*Проміжний контроль (тестування)* рівня знань передбачає виявлення опанування студентом лекційного матеріалу та вміння застосування його для вирішення практичної ситуації і проводиться у вигляді тестування. При цьому тестове завдання може містити як запитання, що стосуються суто теоретичного матеріалу, так і запитання, спрямовані на вирішення невеличкого практичного завдання.

## 9. Схема нарахування балів та критерії оцінювання навчальних досягнень

Структура засобів контролю та розподіл балів із дисципліни “Методи класифікації даних в пакеті Statistica” наведена в табл.

Узагальнена схема нарахування балів (денна форма навчання)

Поточний контроль, самостійна робота, індивідуальні завдання									Експе- наційна робота	Сума
Поточне оцінювання					Проміжний контроль (тестування)	Контрольна робота, передбачена навчальним планом	Індиві- дуальне завдання	Разом		
Розділ 1			Розділ 2							
T1	T2	T3	T4	T5						
2	2	2	2	2	20	30	-	60	40	100

T1, T2 ... – теми розділів.

Узагальнена схема нарахування балів (заочна форма навчання)

Поточний контроль, самостійна робота, індивідуальні завдання									Експе- наційна робота	Сума
Поточне оцінювання					Проміжний контроль (тестування)	Контрольна робота, передбачена навчальним планом	Індиві- дуальне завдання	Разом		
Розділ 1			Розділ 2							
T1	T2	T3	T4	T5						
2	2	2	2	2	20	30	-	60	40	100

T1, T2 ... – теми розділів.

*Проміжний тестовий* контроль проводиться один раз на семестр. Загальна тривалість тестів – 1,5 години. Поточне тестування складається з 10 тестів. Одна правильна відповідь на кожен з тестів дорівнює 2 балам. Тестове завдання містить запитання одиничного і множинного вибору різного рівня складності.

Тести можуть бути застосовані як з метою контролю, так і для закріплення теоретичних знань і практичних навичок.

Тести для проміжного контролю обираються із загального переліку тестів за відповідними темами.

*Підсумковий контроль* за курсом - у формі **екзамену**.

Для допуску до складання підсумкового контролю (екзамену/заліку) здобувач вищої освіти повинен набрати **не менше 10 балів** з навчальної дисципліни під час поточного

контролю, самостійної роботи, індивідуального завдання/контрольної роботи.

Екзамен проводиться у відповідності до екзаменаційних білетів, які містять чотири завдання (два теоретичних питання, одне практичне завдання - задача і тести). Вони дають можливість здійснити оцінювання знань студента за дисципліною.

*Критерії оцінювання результативності роботи студентів при виконанні самостійної роботи*

Якісними критеріями оцінювання виконання індивідуальних завдань студентами є:

*1. Повнота виконання завдання:*

- Елементарна;
- Фрагментарна;
- Повна;
- Неповна.

*2. Рівень самостійності студента*

- під керівництвом викладача;
- консультація викладача;
- самостійно.

*3. Сформованість навчально-інформаційних умінь* (роботи з підручником, володіння різними способами читання, складання плану, рецензій, конспекту, вміння користуватися бібліотекою, спостереження, експеримент тощо)

*4. Сформованість навчально-інтелектуальних умінь* (визначення понять, аналіз, синтез, порівняння, класифікація, систематизація, узагальнення, абстрагування, вміння відповідати на запитання, виконувати творчі завдання тощо);

*5. Рівень сформованості фахових методичних вмінь* (вміння застосовувати на практиці набуті знання):

- низький – володіння умінням здійснювати первинну обробку навчальної інформації без подальшого її аналізу;
- середній – уміння вибирати відомі способи дій для виконання фахових завдань;
- достатній – застосовує набуті знання у стандартних практичних ситуаціях;
- високий – володіння умінням творчо-пошукової діяльності.

*Критерії оцінювання здобувачів вищої освіти за відповіді на питання поточного контролю*

Завданням поточного контролю є перевірка розуміння та засвоєння певного матеріалу, вироблених навичок проведення розрахункових робіт, умінь самостійно опрацьовувати тексти, здатності осмислити зміст теми чи розділу, умінь публічно чи письмово представити певний матеріал (презентація).

Якісними критеріями оцінювання виконання завдань поточного контролю є:

*1. Повнота відповіді або виконання завдання:*

- елементарна;
- фрагментарна;
- повна;
- неповна.

*2. Рівень сформованості логічних умінь:*

- елементарні дії;
- операція, правило, алгоритм;
- правила визначення понять;
- формулювання законів і закономірностей;
- структурування суджень, доводів, описів.

### *Критерії оцінювання здобувачів вищої освіти за виконання контрольної роботи*

Контрольна робота студента складається з двох частин: перша – опрацювання теоретичного питання, друга – виконання практичного завдання. Максимальна оцінка за виконання завдань контрольної роботи – 30 балів.

Виконання завдань оцінюється за такими критеріями:

- 1) теоретичне питання:
  - повнота й ґрунтовність викладу;
  - аргументованість тверджень;
  - суб'єктне усвідомлення змісту;
  - термінологічна коректність;
- 2) практичне завдання:
  - технологічна грамотність;
  - методична грамотність;
  - обґрунтованість висновків;
  - правильність оформлення.

#### **Шкала оцінювання контрольної роботи**

Кількість балів	Теоретичне питання	Практичне завдання
21-30	Повне засвоєння та суб'єктне усвідомлення матеріалу. Твердження чітко аргументовані. Продemonстровано термінологічну грамотність	Продemonстровано методичну й технологічну грамотність. Методичні рішення обґрунтовано. Оформлення відповідає вимогам.
11-20	Часткове засвоєння матеріалу, суб'єктне його не усвідомлення. Нечітка аргументація тверджень. Часткова термінологічна некоректність.	Наявність незначних методичних і технологічних помилок, а також помилок в оформленні роботи. Методичні рішення обґрунтовані частково.
0-10	Теоретичний матеріал не засвоєно. Аргументація відсутня. Термінологічна неграмотність.	Методична й технологічна неграмотність. Неправильне оформлення роботи.

### ***Критерії оцінювання на екзамені***

Оцінювання знань студента проводиться за чотирирівневою шкалою (відмінно, добре, задовільно, незадовільно). За екзамен студент може отримати максимум 40 балів:

1. Для отримання оцінки «відмінно» (35-40 балів) студент повинен:

- укластися у встановлений строк підготовки відповіді;
- викласти теоретичний матеріал чітко, коротко, зв'язно й обґрунтовано;
- навести вірне рішення задачі та тестів.

2. Для одержання оцінки «добре» (25-34 бала) студент повинен:

- укластися у встановлений строк підготовки відповіді;
- викласти теоретичний матеріал зв'язно й обґрунтовано;
- навести вірне рішення задачі;
- можливі помилки у відповідях на тести.

3. Для отримання оцінки «задовільно» (15-24 бала) студент повинен:

- викладати теоретичний матеріал у доступній для розуміння формі;
- можливі помилки при розв'язанні задачі та в тестах.

4. Оцінку «незадовільно» (1-14 балів) отримують студенти, відповіді яких можуть бути оцінені нижче вимог, сформульованих у попередніх пунктах.

Кожне завдання заліку оцінюється окремо. Загальна оцінка дорівнює сумі оцінок за всі завдання (види контролю).

**Підсумкова оцінка** з навчальної дисципліни визначається як сума балів, набраних здобувачем вищої освіти протягом семестру при виконанні контрольних заходів, передбачених програмою навчальної дисципліни та балів, набраних ним при складанні семестрового екзамену (поточне оцінювання за роботу протягом семестру, поточний контроль знань (тестування), контрольна робота та оцінка за результатами підсумкового екзамену).

Максимальна сума балів, яку може набрати здобувач вищої освіти при підсумковому семестровому контролі, складає 100.

#### Зведена шкала оцінювання роботи студентів з дисципліни

Види робіт	Максимум балів
Поточне оцінювання	10
Поточний контроль знань (тестування)	20
Контрольна робота	30
<b>РАЗОМ</b>	<b>60</b>
<b>Екзамен</b>	<b>40</b>
<b>ВСЬОГО</b>	<b>100</b>

У відповідності до набраних студентом балів оцінка знання матеріалу проводиться за чотирирівневою системою згідно з Методикою переведення показників успішності знань студентів.

#### Шкала оцінювання

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка за національною шкалою
	для <b>чотирирівневої</b> шкали оцінювання
90-100	відмінно
80-89	добре
70-79	
60-69	
50-59	задовільно
1-49	незадовільно

## 10. Рекомендована література

### Основна література

1. Корепанов О.С. Методи класифікації даних в пакеті Statistica: дистанційний курс на платформі LMS Moodle. URL : <https://moodle.karazin.ua/course/view.php?id=867>.
2. Єріна А. М., Єрін Д. Л. Статистичне моделювання та прогнозування : навч. посіб. К. : КНЕУ, 2014. 348 с.
3. Моделювання та оптимізація систем: підручник / Дубовой В. М., Кветний Р. Н., Михальов О. І., Усов А. В. Вінниця : ПП «ТД«Еднльвейс», 2017. 804 с.
4. Статистичне моделювання та прогнозування: Практикум із використанням MS Excel та ППП Statistica / упоряд. Г. С. Корепанов, О. С. Корепанов, Ю. О. Лазебник, Т. Г. Чала, Д. І. Черненко // Економ. факультет ХНУ ім. В. Н. Каразіна, Х. : Видавництво ФОП Тагаєв П. О., 2017. 92 с.



### Допоміжна література

1. Богомаз О.О., Корепанов О.С. Порівняльний аналіз стану навколишнього середовища за регіонами України. *Fundamental and Applied Research in the Modern World*: матер. IV Міжнар. наук.-практ. конф. (Бостон, США, 18-20 листопада 2020 р.). Бостон, США. 2020. С. 358–365. URL: <https://sci-conf.com.ua/wp-content/uploads/2020/11/FUNDAMENTAL-AND-APPLIED-RESEARCH-IN-THE-MODERN-WORLD-18-20.11.20.pdf>.
2. Корепанов О.С., Дашутін В.В. Методичні засади аналізу регіональної диференціації розвитку тваринництва в Україні. *Бізнес Інформ*. 2020. № 10. С. 214–220. URL: [https://www.business-inform.net/export\\_pdf/business-inform-2020-10\\_0-pages-214\\_220.pdf](https://www.business-inform.net/export_pdf/business-inform-2020-10_0-pages-214_220.pdf).
3. Корепанов О.С., Лазебник Ю.О., Чала Т.Г., Корнієнко В.В. Статистичний аналіз та прогнозування складових національного валютного ринку в умовах вторгнення РФ в Україну. *Бізнес Інформ*. Харків. 2023. № 1. С. 31–39. URL: [https://www.business-inform.net/export\\_pdf/business-inform-2023-1\\_0-pages-31\\_39.pdf](https://www.business-inform.net/export_pdf/business-inform-2023-1_0-pages-31_39.pdf).
4. Лазебник Ю.О., Корепанов О.С., Хавалиць Д.Л. Оцінка диференціації розвитку підприємств сільського господарства за регіонами України. *Інвестиції: практика та досвід*. 2019. № 21. С. 15–20. URL: [http://www.investplan.com.ua/pdf/21\\_2019/5.pdf](http://www.investplan.com.ua/pdf/21_2019/5.pdf)
5. Box G. Time Series Analysis : Forecasting and Control / G. Box, G. Jenkins, G. Reinsel / John Wiley & Sons. 2018.
6. Chala, T., Korepanov, O., Lazebnyk, I., Chernenko, D., Korepanov, G. Statistical modeling and forecasting of wheat and meslin export from Ukraine using singular spectral analysis. *Statistics in Transition New Series*. ISSN 1234-7655. Poland. 2023. 24(1), pp. 169–197. DOI <https://doi.org/10.59170/stattrans-2023-010>. URL: [https://sit.stat.gov.pl/SiT/2023/1/gus\\_sit\\_2023\\_00\\_tetyana\\_chala\\_oleksiy\\_korepanov\\_iuliia\\_lazebnyk\\_daryna\\_chernenko\\_georgii\\_korepanov\\_statistical\\_modelling\\_and\\_forecasting.pdf](https://sit.stat.gov.pl/SiT/2023/1/gus_sit_2023_00_tetyana_chala_oleksiy_korepanov_iuliia_lazebnyk_daryna_chernenko_georgii_korepanov_statistical_modelling_and_forecasting.pdf)
7. Kleinbaum D. Applied Regression Analysis and Other Multivariable Methods / D. Kleinbaum, L. Kupper, A. Nizam et al. Cengage Learning, 2013.
8. Korepanov O., Goncharenko Y., Mekhovich S., Dyshkantiuk O., Salamatina S., Rusavska V. Economic-Mathematical Modelling of the Distribution of Investments in the Tourism and Hospitality. *International Journal of Engineering and Advanced Technology (IJEAT)*. ISSN: 2249–8958. Vol.-9. Issue-1. October 2019. PP. 138-142. URL: <https://www.ijeat.org/wp-content/uploads/papers/v9i1/A1093109119.pdf>
9. Sukumar Arun. Applied Business Analysis / Arun Sukumar. Lucian Tipi ; Jayne Revill, 2016.

### 11. Посилання на інформаційні ресурси в Інтернеті, відео-лекції, інше методичне забезпечення

1. Нормативно-правова база / Кабінет Міністрів України; Урядовий портал. Офіц. веб-сайт. URL : <http://www.kmu.gov.ua/control/npd/list>.
2. Нормативно-правові акти / Урядовий портал. Єдиний веб-портал виконавчої влади України. URL : [http://www.me.gov.ua/control/uk/publish/category/main?cat\\_id=32854](http://www.me.gov.ua/control/uk/publish/category/main?cat_id=32854).
3. Статистична інформація / Державна служба статистики України; за ред. О. Г. Осауленка / Офіційний сайт державної служби статистики України. URL : <http://www.ukrstat.gov.ua>.
4. Загальноакадемічний портал наукової періодики. Національна академія наук України / Л. Костенко, Є. Копанєва // Національна бібліотека України імені В.І. Вернадського. URL : <http://www.nbuv.gov.ua/portal/natural/NTI/search.html>.

## 12. Особливості навчання за денною формою в умовах дії обставин непереборної сили

В умовах дії форс-мажорних обмежень освітній процес в університеті здійснюється відповідно до наказів/розпоряджень ректора/проректора або за змішаною формою навчання, або повністю дистанційно в синхронному режимі.

У разі проведення екзамену/заліку в дистанційній формі використовується LMS платформа Moodle з автентифікацією здобувача у режимі відеоконференції з Zoom. Реєстрація (допуск до складання) учасників освітнього процесу, а також обмін контрольними (підсумковими) завданнями та відповідями на них здійснюється винятково з корпоративної електронної пошти Харківського національного університету імені В. Н. Каразіна (домен – karazin.ua) із забезпеченням академічної доброчесності. При проведенні екзамену/заліку в дистанційній формі використовуються технічні і програмні засоби, які дозволяють забезпечити аудіо- і відео- фіксацію.

Режим доступу до курсу на LMS платформі Moodle:  
<https://moodle.karazin.ua/course/view.php?id=867>.

Режим доступу до курсу на LMS платформі Classroom:  
<https://classroom.google.com/u/0/c/MjYzMjQ4NjUyMTk1> (код курсу: e76r4kt).

## **2. МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДЛЯ ВИКОНАННЯ КУРСОВИХ, ЛАБОРАТОРНИХ, ПРАКТИЧНИХ, САМОСТІЙНИХ РОБІТ ТОЩО**

## Тематика та плани практичних занять

### Практичне заняття 1. «Основні поняття методів багатомірної класифікації»

#### Питання для дискусії:

1. У якій формі можуть представлятися вихідні дані в задачах класифікації об'єктів?
2. Що таке навчальна вибірка?
3. Які кінцеві прикладні цілі ставить перед собою дослідник при проведенні класифікації?
4. Які фактори називаються типоутворюючими?
5. Як будується комбінаційне угруповання?
6. Яка ідея покладена в основу методу відбору найбільш інформативних ознак-детермінантів?

### Практичні заняття 2-3. «Кластерний аналіз. Ієрархічні методи класифікації»

#### Питання для дискусії:

1. Чим обумовлена необхідність розвитку методів кластерного аналізу і їхнього використання?
2. Які задачі дозволяють вирішувати методи кластерного аналізу?
3. У чому суть агломеративних методів кластерного аналізу?
4. У чому суть дивізімних методів кластерного аналізу?
5. Чому вибір способу обчислення відстані між об'єктами є вузловим моментом дослідження?
6. Які відстані між об'єктами найбільш часто використовуються в задачах кластерного аналізу?
7. Які показники можуть бути використані в якості мір подібності?
8. У чому полягає сутність ієрархічних агломеративних методів?

#### Приклади розв'язання типових задач:

##### Задача 1

За вихідним даними, представленим у табл. 1.1, потрібно провести класифікацію шести промислових підприємств ( $n = 6$ ) по двом показникам:  $x^{(1)}$  - рентабельність (%),  $x^{(2)}$  - продуктивність праці (тис. грн/чіл.)

Характеристики аналізованих підприємств

№ підприємства	1	2	3	4	5	6
$x^{(1)}$	23,4	17,5	9,7	18,2	6,6	8,0
$x^{(2)}$	9,1	5,2	5,5	9,4	7,5	5,7

Класифікацію проведемо по ієрархічному агломеративному алгоритму з використанням звичайного й зваженого ( $w_1 = 0,75$  і  $w_2 = 0,25$ ) евклидова відстані, а також принципів: «найближчого» і «далекого» сусіда, центру ваги й середньому зв'язку.

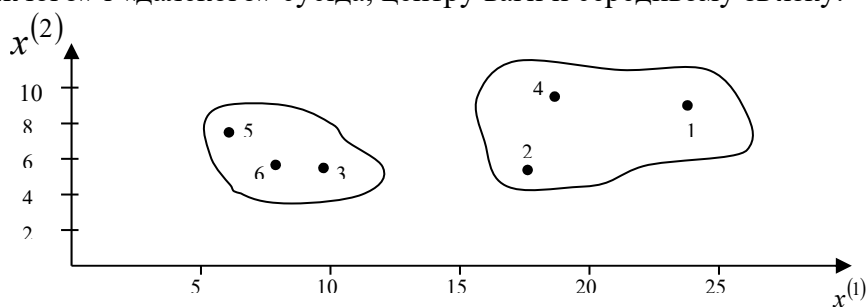


Рис. 1.1. Вихідні дані для класифікації.

На підставі попереднього якісного аналізу можна висунути припущення, що по відомих характеристиках підприємства 1, 2, 4 належать до однієї типологічної групи, а підприємства 3, 5, 6 – до іншої, що узгодиться з розташуванням шести спостережень на площині (мал. 1.1).

1) Проведемо класифікацію, вибравши при звичайній евклідовій відстані принцип «найближчого сусіда». Відповідно звичайній евклідовій метриці відстань між спостереженнями 1 і 2 рівно:

$$d_{1,2} = \sqrt{(23,4 - 17,5)^2 + (9,1 - 5,2)^2} = 7,07.$$

При цьому очевидно, що  $d_{1,1} = 0$ . Знаходимо відстані між усіма шістьма спостереженнями й будемо матрицю відстаней:

$$D_1 = \begin{pmatrix} 0 & 7,07 & 14,6 & 5,21 & 16,88 & 15,77 \\ 7,07 & 0 & 7,81 & 4,26 & 11,14 & 9,51 \\ 14,16 & 7,81 & 0 & 9,35 & 3,69 & 1,71 \\ 5,21 & 4,26 & 9,35 & 0 & 11,75 & 10,85 \\ 16,88 & 11,14 & 3,69 & 11,75 & 0 & 2,28 \\ 15,77 & 9,51 & 1,71 & 10,85 & 2,28 & 0 \end{pmatrix}.$$

З матриці відстаней  $D_1$  випливає, що об'єкти 3 і 6 найбільш близькі ( $d_{3,6} = d_{6,3} = 1,71$ ), тому об'єднаємо їх в один кластер. Після об'єднання об'єктів одержимо п'ять кластерів:  $S_1, S_2, S_{(3,6)}, S_4, S_5$ .

Відстань між кластерами будемо знаходити за принципом «найближчого сусіда», скориставшись формулою перерахування. Так, відстань між кластером  $S_1$  і кластером  $S_{(3,6)}$  рівно:

$$\begin{aligned} d_{1,(3,6)} &= d(S_1, S_{(3,6)}) = \frac{1}{2}d_{1,3} + \frac{1}{2}d_{1,6} - \frac{1}{2}|d_{1,3} - d_{1,6}| = \\ &= \frac{1}{2}(14,16 + 15,77) - \frac{1}{2}|14,16 - 15,77| = 14,16. \end{aligned}$$

Як видно з розрахунку, відстань  $d_{1,(3,6)}$  дорівнює відстані від об'єкта 1 до найближчого до нього об'єкта, що входить у кластер  $S_{(3,6)}$ , тобто  $d_{1,(3,6)} = d_{1,3} = 14,16$ . Провівши аналогічні розрахунки, одержимо матрицю відстаней:

$$D_2 = \begin{pmatrix} 0 & 7,07 & 14,16 & 5,21 & 16,88 \\ 7,07 & 0 & 7,81 & 4,26 & 11,14 \\ 14,16 & 7,81 & 0 & 9,35 & 3,69 \\ 5,21 & 4,26 & 9,35 & 0 & 11,75 \\ 16,88 & 11,14 & 3,69 & 11,75 & 0 \end{pmatrix}.$$

З матриці відстаней  $D_2$  випливає, що найбільш близькі кластери  $S_{(3,6)}$  й  $S_5$  ( $d_{5,(3,6)} = 3,69$ ). Після їхнього об'єднання маємо чотири кластери:  $S_1, S_2, S_{(3,5,6)}, S_4$ .

Будемо нову матрицю спостережень:

$$D_3 = \begin{pmatrix} 0 & 7,07 & 14,16 & 5,21 \\ 7,07 & 0 & 7,81 & 4,26 \\ 14,16 & 7,81 & 0 & 9,35 \\ 5,21 & 4,26 & 9,35 & 0 \end{pmatrix}.$$

Об'єднаємо спостереження 2 і 4, що мають найменшу відстань ( $d_{2,4} = 4,26$ ). Після об'єднання маємо три кластери:  $S_1$ ,  $S_{(2,4)}$  і  $S_{(3,5,6)}$ . Обчислимо нову матрицю спостережень:

$$D_4 = \begin{pmatrix} 0 & 5,21 & 14,16 \\ 5,21 & 0 & 7,81 \\ 14,16 & 7,81 & 0 \end{pmatrix}.$$

Об'єднаємо кластери  $S_1$  й  $S_{(2,4)}$ , відстань між якими, згідно з матрицею  $D_4$ , мінімально ( $d_{1,(2,4)} = 5,21$ ). У результаті цього одержимо два кластери:  $S_{(1,2,4)}$  і  $S_{(3,5,6)}$ . Матриця відстаней буде мати вигляд:

$$D_5 = \begin{pmatrix} 0 & 7,81 \\ 7,81 & 0 \end{pmatrix}.$$

З матриці  $D_5$  випливає, що на відстані  $d_{(1,2,4),(3,5,6)} = 7,81$  всі шість спостережень поєднуються в один кластер.

На підставі графічної вистави результатів кластерного аналізу можна зробити висновок, що найкращим є розбивка шести підприємств на два кластери:  $S_{(1,2,4)}$  і  $S_{(3,5,6)}$ , коли гранична відстань перебуває в інтервалі  $5,21 \leq d_{\text{нор}} \leq 7,81$ .

## Задача 2

Проведемо класифікацію, вибравши при звичайній евклідовій відстані принцип «далекого сусіда».

Як і в першому випадку, будемо використовувати звичайне евклідово відстань, тому матриця  $D_1$  залишиться без зміни. Згідно агломеративному алгоритму в один кластер поєднуються об'єкти 3 і 6, як найбільше близькі ( $d_{3,6} = 1,71$ ). Після об'єднання маємо п'ять кластерів:  $S_1$ ,  $S_2$ ,  $S_{(3,6)}$ ,  $S_4$  і  $S_5$ .

У вигляді того, що відстань між кластерами вимірюємо за принципом «далекого сусіда», у формулі перерахування ухвалюємо  $\delta = 1/2$ , а не  $-1/2$ , як у першому випадку. Тоді, наприклад, відстань між кластером  $S_1$  і кластером  $S_{(3,6)}$  визначається по формулі:

$$\begin{aligned} d_{1,(3,6)} &= d(S_1, S_{(3,6)}) = \frac{1}{2}d_{1,3} + \frac{1}{2}d_{1,6} + \frac{1}{2}|d_{1,3} - d_{1,6}| = \\ &= \frac{1}{2}(14,16 + 15,77) + \frac{1}{2}|14,16 - 15,77| = 15,77. \end{aligned}$$

Таким чином, відстань  $d_{1,(3,6)}$  дорівнює відстані від об'єкта 1 до найбільш віддаленого від нього об'єкта, що входить у кластер  $S_{(3,6)}$ , тобто  $d_{1,(3,6)} = d_{1,6} = 15,77$ .

Аналогічно розглядаються всі інші елементи матриці відстаней:



$$D_2 = \begin{pmatrix} 0 & 7,07 & 15,77 & 5,21 & 16,88 \\ 7,07 & 0 & 9,51 & 4,26 & 11,14 \\ 15,77 & 9,51 & 0 & 10,85 & 3,69 \\ 5,21 & 4,26 & 10,85 & 0 & 11,75 \\ 16,88 & 11,14 & 3,69 & 11,75 & 0 \end{pmatrix}.$$

Згідно з матрицею  $D_2$  найбільш близькими кластерами будуть  $S_{(3,6)}$  і  $S_5$ ,  $d_{(3,6),5} = 3,69$ . Після їхнього об'єднання маємо чотири кластери:  $S_1$ ,  $S_2$ ,  $S_{(3,5,6)}$  і  $S_4$ . Будуємо матрицю відстаней  $D_3$ , скориставшись принципом далекого сусіда:

$$D_3 = \begin{pmatrix} 0 & 7,07 & 16,88 & 5,21 \\ 7,07 & 0 & 11,14 & 4,26 \\ 16,88 & 11,14 & 0 & 11,75 \\ 5,21 & 4,26 & 11,75 & 0 \end{pmatrix}.$$

Об'єднаємо об'єкти 2 і 4 в один кластер, як найбільше близькі (згідно з матрицею  $D_3$ ),  $d_{2,4} = 4,26$ . Після об'єднання маємо три кластери:  $S_1$ ,  $S_{(2,4)}$  і  $S_{(3,5,6)}$ . Будуємо нову матрицю  $D_4$  за принципом «далекого сусіда»:

$$D_4 = \begin{pmatrix} 0 & 7,07 & 16,88 \\ 7,07 & 0 & 11,75 \\ 16,88 & 11,75 & 0 \end{pmatrix}.$$

Об'єднаємо кластери  $S_1$  й  $S_{(2,4)}$ , відстань між якими  $d_{1,(2,4)} = 7,07$  мінімально, і одержимо два кластери:  $S_{(1,2,4)}$  і  $S_{(3,5,6)}$ , відстань між якими визначається по матриці:

$$D_5 = \begin{pmatrix} 0 & 16,88 \\ 16,88 & 0 \end{pmatrix}$$

і рівно  $d_{(1,2,4),(3,5,6)} = 16,88$ .

Графічні результати класифікації представлені на мал. 1.2.

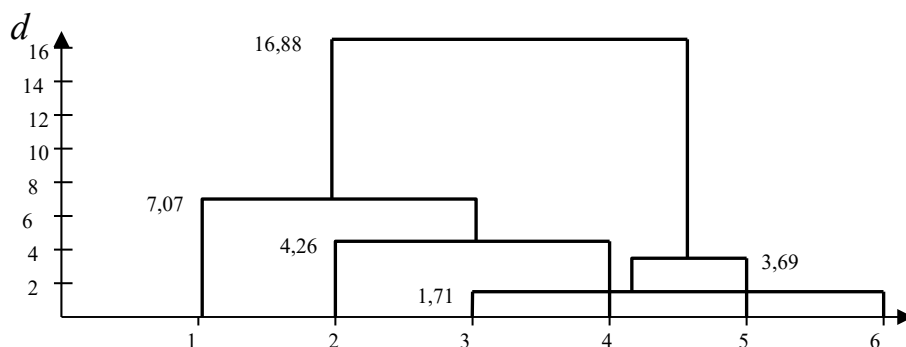


Рис. 1.2. Дендрограма (звичайне евклідово відстань, далекий сусід).

Як і в попередньому випадку найкращим є розбивка підприємств на два кластери (мал. 1.2):  $S_{(1, 2, 4)}$  і  $S_{(3, 5, 6)}$ , після передостаннього кроку класифікації, коли інтервал виміру відстані об'єднання найбільша  $7,07 \leq d_{\text{пор}} \leq 16,88$ .

#### Практичні заняття 4-5. «Кластерний аналіз. Ітеративні методи класифікації»

##### Питання для дискусії:

1. Яким чином проводиться класифікація об'єктів методом  $k$ -середніх?
2. До виконання яких кроків зводяться обчислювальні процедури більшості ітеративних методів класифікації?
3. У чому складається істотна відмінність методу пошуку згущень від інших ітеративних методів класифікації?
4. У чому полягає суть ітеративного алгоритму типу «форель»?
5. Якими способами вибирається радіус сфери для пошуку локальних згущень точок у методі пошуку згущень?
6. Для якої мети використовуються критерії якості класифікації?

##### Приклади розв'язання типових задач:

За вихідним даними, представленим у табл. 2.1, потрібно провести класифікацію шести промислових підприємств ( $n = 6$ ) по двом показникам:  $x^{(1)}$  - рентабельність (%),  $x^{(2)}$  - продуктивність праці (тис. грн/чіл.)

Таблиця 2.1

Характеристики аналізованих підприємств						
№ підприємства	1	2	3	4	5	6
$x^{(1)}$	23,4	17,5	9,7	18,2	6,6	8,0
$x^{(2)}$	9,1	5,2	5,5	9,4	7,5	5,7

На попередньому занятті класифікація аналізованих підприємств була проведена за допомогою ієрархічних методів класифікації. Тепер проведемо класифікацію підприємств за допомогою методу  $k$ -середніх і зрівняємо отримані результати.

Для початку з  $n$ -об'єктів досліджуваної сукупності відбираються випадковим образом або задаються дослідником виходячи з якихось апріорних міркувань  $k$ -об'єктів еталонів. Оскільки згідно результатів розв'язку цього завдання на практичному занятті №1 одержали розбивку підприємств на два кластери, то виберемо два об'єкти - еталона з найбільш високими й низькими показниками, наприклад, підприємства 1 і 5.

Запишемо вихідні значення еталонів і ваг:

$$\left. \begin{aligned} E_1^0 &= X_1 = (23,4; 9,1); \quad w_1^0 = 1 \\ E_2^0 &= X_5 = (6,6; 7,5); \quad w_2^0 = 1 \end{aligned} \right\} \text{ - нулева ітерація.}$$

На *першому кроці* беремо другий об'єкт і визначаємо його відстань до кожного з еталонів по евклідовій метриці:

$$\begin{aligned} d_{21} &= \sqrt{(17,5 - 23,4)^2 + (5,2 - 9,1)^2} = 7,072; \\ d_{22} &= \sqrt{(17,5 - 6,6)^2 + (5,2 - 7,5)^2} = 11,14. \end{aligned}$$

Отже, розглянутий об'єкт повинен бути приєднаний до першого еталона й перший еталон буде перелічений, а другий не міняється:

$$E_1^1 = \frac{w_1^0 \cdot E_1^0 + X_2}{w_1^0 + 1},$$

$$w_1^1 = w_1^0 + 1 = 2, \quad E_2^1 = E_2^0, \quad w_2^1 = w_2^0,$$

де  $X_2$  - вектор значень змінних для другого об'єкта,

$E_1^1$  - перелічене значення еталона;

$$E_1^1 = \left( \frac{23,4 + 17,5}{2}; \frac{9,1 + 5,2}{2} \right) = (20,45; 7,15).$$

На *другому кроці* перевіряємо, до якого еталона ближче всього перебуває третій об'єкт:

$$d_{31} = \sqrt{(9,7 - 20,45)^2 + (5,5 - 7,15)^2} = 10,876;$$

$$d_{32} = \sqrt{(9,7 - 6,6)^2 + (5,5 - 7,5)^2} = 3,689.$$

Тому що  $d_{32} = \min\{d_{31}; d_{32}\}$ , отже, третій об'єкт приєднується до другого еталона, цей еталон перераховується й вага його збільшується:

$$E_2^2 = \left( \frac{9,7 + 6,6}{2}; \frac{5,5 + 7,5}{2} \right) = (8,15; 6,5),$$

$$w_2^2 = w_2^1 + 1 = 2; \quad E_1^2 = E_1^1; \quad E_2^2 = E_2^1; \quad w_1^2 = w_1^1.$$

На *третьому кроці* перевіряємо, до якого еталона ближче всього перебуває четвертий об'єкт:

$$d_{41} = \sqrt{(18,2 - 20,45)^2 + (9,4 - 7,15)^2} = 3,182;$$

$$d_{42} = \sqrt{(18,2 - 8,15)^2 + (9,4 - 6,5)^2} = 10,46.$$

Отже, четвертий об'єкт приєднується до першого еталона, вага якого стає рівним:

$$E_1^3 = \left( \frac{20,45 \cdot 2 + 18,2}{3}; \frac{7,15 \cdot 2 + 9,4}{3} \right) = (19,7; 7,9);$$

$$w_1^3 = w_1^2 + 1 = 3; \quad E_2^3 = E_2^2; \quad w_2^3 = w_2^2.$$

На *четвертому кроці* перевіряємо, до якого еталона ближче всього перебуває шостий об'єкт:

$$d_{61} = \sqrt{(8,0 - 19,7)^2 + (5,7 - 7,9)^2} = 11,905;$$

$$d_{62} = \sqrt{(8,0 - 8,15)^2 + (5,7 - 6,5)^2} = 0,814.$$

Шостий об'єкт приєднується до другого еталона, вага якого стає рівним:

$$E_2^4 = \left( \frac{8,15 \cdot 2 + 8}{3}; \frac{6,5 \cdot 2 + 5,7}{3} \right) = (8,1; 6,233);$$

$$w_1^4 = w_1^3; \quad w_2^4 = w_2^3 + 1 = 3; \quad E_1^4 = E_1^3.$$

Таким чином, отримана розбивка аналізованих підприємств на два кластери: (1, 2, 4) і (3, 5, 6).

Після того як переглянуті всі об'єкти, крім першого й п'ятого, процес «зациклюється», тобто по тому ж правилу здійснюються перегляд і приєднання до відповідного до еталона кожного із шести об'єктів. При цьому відбувається перерахування еталонів і триває нарощування їх ваг. Результати розрахунків представлені в табл. 2.

Таблиця 2

Параметричні дані кластеризації об'єктів методом  $k$ -середніх

Номер ітерації	Еталони і їх ваги	
	1	2
0	$E_1^0 = (23,4; 9,1); w_1^0 = 1$	$E_2^0 = (6,6; 7,5); w_2^0 = 1$
1	$E_1^1 = (20,45; 7,15); w_1^1 = 2$	$E_2^1 = (6,6; 7,5); w_2^1 = 1$
2	$E_1^2 = (20,45; 7,15); w_1^2 = 2$	$E_2^2 = (8,15; 6,5); w_2^2 = 2$
3	$E_1^3 = (19,7; 7,9); w_1^3 = 3$	$E_2^3 = (8,15; 6,5); w_2^3 = 2$
4	$E_1^4 = (19,7; 7,9); w_1^4 = 3$	$E_2^4 = (8,1; 6,233); w_2^4 = 3$
5	$E_1^5 = (20,625; 8,2); w_1^5 = 4$	$E_2^5 = (8,1; 6,233); w_2^5 = 3$
6	$E_1^6 = (20,0; 7,6); w_1^6 = 5$	$E_2^6 = (8,1; 6,233); w_2^6 = 3$
7	$E_1^7 = (20,0; 7,6); w_1^7 = 5$	$E_2^7 = (8,5; 6,05); w_2^7 = 4$
8	$E_1^8 = (19,7; 7,9); w_1^8 = 6$	$E_2^8 = (8,5; 6,05); w_2^8 = 4$
9	$E_1^9 = (19,7; 7,9); w_1^9 = 6$	$E_2^9 = (8,12; 6,34); w_2^9 = 5$
10	$E_1^{10} = (19,7; 7,9); w_1^{10} = 6$	$E_2^{10} = (8,1; 6,23); w_2^{10} = 6$

Отже, на цьому процес завершується, тому що наступна розбивка (ітерації 5 - 10) дали такий же результат розділення, як і попереднє (ітерації 1 - 4).

Утворено два кластери:  $S_1 \{1, 2, 4\}$ ,  $S_2 \{3, 5, 6\}$ . Обчислюємо центри ваги отриманих кластерів (у загальному випадку ці центри можуть не збігатися із центрами еталонів):

$$C_1 = \left( \frac{23,4 + 17,5 + 18,2}{3}; \frac{9,1 + 5,2 + 9,4}{3} \right) = (19,7; 7,9);$$

$$C_2 = \left( \frac{9,7 + 6,6 + 8,0}{3}; \frac{5,5 + 7,5 + 5,7}{3} \right) = (8,1; 6,23).$$

Після цього будується остаточна розбивка: кожна багатомірна крапка ставиться до того кластера, центр якого ближче всіх до цієї крапки.

Для нашого прикладу визначаємо по черзі відстані всіх об'єктів ( $X_1, X_2, X_3, X_4, X_5, X_6$ ) до центрів двох кластерів (табл. 3).

Таблиця 3

Відстані до центрів класів

Центри кластерів	Об'єкти					
	1	2	3	4	5	6
$C_1$	3,89	3,48	10,28	2,12	13,11	11,91
$C_2$	15,57	9,46	1,76	10,59	1,97	0,54

Як видно з табл. 3, підтверджується отримана розбивка на два кластери:  $S_1 \{1, 2, 4\}$ ,  $S_2 \{3, 5, 6\}$ . На цьому алгоритм завершується.

Таким чином, класифікація аналізованих підприємств із використанням методу  $k$ -середніх дала такий же результат, як і класифікація ієрархічними методами кластерного аналізу.

## Практичні заняття 6-7. «Дискримінантний аналіз»

### Питання для дискусії:

1. Які ознаки називаються дискримінантними змінними?
2. Які допущення приймаються в дискримінантному аналізі?
3. Яким повинне бути співвідношення числа об'єктів спостереження й числа дискримінантних змінних?
4. Як визначається канонічна дискримінантна функція?
5. Яким чином визначають коефіцієнти дискримінантної функції?
6. Як визначається границя, що розділяє розглянуті групи?
7. Яким буде алгоритм використання дискримінантного аналізу для проведення багатомірної класифікації об'єктів?
8. Як зміна числа змінних впливає на результат дискримінантного аналізу?

### Типові задачі:

Аналізується діяльність 8 регіонів України по трьом економічним показникам, що характеризують сільськогосподарське виробництво в цих регіонах в 2000 г (табл. 3.1.): урожайність зернових культур (у господарствах усіх категорій; з 1 га; ц), середній річний удій молока від однієї корови (у сільськогосподарських підприємствах; кг.) і виробництво м'яса на душу населення (у господарствах усіх категорій; у забійній вазі; кг/чол).

Як видно з табл. 3.1, з регіонів виділено 2 групи: передова ( $X$ ) і відстаюча ( $Y$ ). Для регіонів, які підлягають дискримінації ( $Z$ ), потрібно обчислити значення дискримінантної функції й провести їхню класифікацію.

Метою дискримінантного аналізу є віднесення нового спостереження (рядка матриці  $Z$ ) або до  $X$ , або до  $Y$ .

Таблиця 3.1

Вихідні дані				
Регіони \ Показники		Урожайність зернових культур	Середній річний удій молока від однієї корови	Виробництво м'яса на душу населення
Передовий, $X$	Вінницька обл.	23,6	3015	52
	Київська обл.	24,7	2830	64
	Черкаська обл.	27,2	2920	52
Відстаючий, $Y$	Запорізька обл.	15,5	2432	37
	Луганська обл.	11,8	2305	20
	Миколаївська обл.	14,7	2487	38
$Z$	Волинська обл.	19,4	2558	41
	Чернівецька обл.	24,2	3188	31

Запишемо вихідні дані у вигляді матриць  $X$  і  $Y$ :

$$X = \begin{pmatrix} 23,6 & 3015 & 52 \\ 24,7 & 2830 & 64 \\ 27,2 & 2920 & 52 \end{pmatrix}; \quad Y = \begin{pmatrix} 15,5 & 2432 & 37 \\ 11,8 & 2305 & 20 \\ 14,7 & 2487 & 38 \end{pmatrix}.$$

При розв'язку завдань по навчальних вибірках спочатку визначають їхні вектори середніх:

$$\bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} = \begin{pmatrix} 25,2 \\ 2922 \\ 56 \end{pmatrix}; \quad \bar{Y} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \end{pmatrix} = \begin{pmatrix} 14 \\ 2408 \\ 31,7 \end{pmatrix}. \quad (3.1)$$

Використовуючи дані (3.1) можна знайти коваріаційні матриці:

$$S_x = (s_{ki})_x \text{ и } S_y = (s_{ki})_y$$

Елементи матриці  $S_x$  визначаються вираженням:

$$s_{ki}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \overline{x_j x_k} - \bar{x}_j \bar{x}_k, \quad (3.2)$$

де  $j, k = 1, 2, 3$ ;

$$\bar{x}_j = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{ij}.$$

Елементи матриці  $S_y$  визначаються по аналогічному вираженню. Визначимо елементи коваріаційних матриць  $S_x$  і  $S_y$  відповідно для передових і відстаючих регіонів згідно формули (3.2):

$$s_{11}(x) = \frac{23,6 \cdot 23,6 + 24,7 \cdot 24,7 + 27,2 \cdot 27,2}{3} - 25,2 \cdot 25,2 = 0,59;$$

$$s_{12}(x) = \frac{23,6 \cdot 3015 + 24,7 \cdot 2830 + 27,2 \cdot 2920}{3} - 25,2 \cdot 2922 = 141,4;$$

$$s_{13}(x) = \frac{23,6 \cdot 52 + 24,7 \cdot 64 + 27,2 \cdot 52}{3} - 25,2 \cdot 56 = 3,73;$$

$$s_{22}(x) = \frac{3015 \cdot 3015 + 2830 \cdot 2830 + 2920 \cdot 2920}{3} - 2922 \cdot 2922 = 3758;$$

$$s_{23}(x) = \frac{3015 \cdot 52 + 2830 \cdot 64 + 2920 \cdot 52}{3} - 2922 \cdot 56 = -385,3;$$

$$s_{33}(x) = \frac{52 \cdot 52 + 64 \cdot 64 + 52 \cdot 52}{3} - 56 \cdot 56 = 32;$$

$$s_{11}(y) = \frac{15,5 \cdot 15,5 + 11,8 \cdot 11,8 + 14,7 \cdot 14,7}{3} - 14 \cdot 14 = 2,53;$$

$$s_{12}(y) = \frac{15,5 \cdot 2432 + 11,8 \cdot 2305 + 14,7 \cdot 2487}{3} - 14 \cdot 2408 = 106;$$

$$s_{13}(y) = \frac{15,5 \cdot 37 + 11,8 \cdot 20 + 14,7 \cdot 38}{3} - 14 \cdot 31,7 = 12,2;$$

$$s_{22}(y) = \frac{2432 \cdot 2432 + 2305 \cdot 2305 + 2487 \cdot 2487}{3} - 2408 \cdot 2408 = 5808;$$

$$s_{23}(y) = \frac{2432 \cdot 37 + 2305 \cdot 20 + 2487 \cdot 38}{3} - 2408 \cdot 31,7 = 529,7;$$

$$s_{33}(y) = \frac{37 \cdot 37 + 20 \cdot 20 + 38 \cdot 38}{3} - 31,7 \cdot 31,7 = 66,1.$$

У результаті розрахунків отримані наступні коваріаційні матриці:

$$S_x = \begin{pmatrix} 0,59 & 141,4 & 3,73 \\ & 3758 & -385,3 \\ & & 32 \end{pmatrix}; \quad S_y = \begin{pmatrix} 2,53 & 106 & 12,2 \\ & 5808 & 529,7 \\ & & 66,1 \end{pmatrix}.$$

Далі знайдемо сумарну коваріаційну матрицю:

$$S_* = \frac{1}{n_1 + n_2 - 2} (n_1 S_x + n_2 S_y) = \frac{1}{3 + 3 - 2} (3S_x + 3S_y) = \begin{pmatrix} 2,34 & 185,5 & 11,9 \\ & 7174 & 108,3 \\ & & 73,6 \end{pmatrix},$$

а потім зворотну їй матрицю

$$S_*^{-1} = \begin{pmatrix} -0,27723 & 0,00664 & 0,03505 \\ & -0,00002 & -0,00105 \\ & & 0,00946 \end{pmatrix}.$$

Обчислимо вектор коефіцієнтів дискримінантної функції:

$$A = S_*^{-1}(\bar{X} - \bar{Y}) = S_*^{-1} \begin{pmatrix} 11,2 \\ 514 \\ 24,3 \end{pmatrix} = \begin{pmatrix} 1,16 \\ 0,039 \\ 0,083 \end{pmatrix}.$$

Після цього розрахуємо вектори дискримінантної функції для матриць вихідних даних:

$$\hat{U}_x = X \cdot A = \begin{pmatrix} 23,6 & 3015 & 52 \\ 24,7 & 2830 & 64 \\ 27,2 & 2920 & 52 \end{pmatrix} \cdot \begin{pmatrix} 1,16 \\ 0,039 \\ 0,083 \end{pmatrix} = \begin{pmatrix} 149,3 \\ 144,3 \\ 149,7 \end{pmatrix};$$

$$\hat{U}_y = Y \cdot A = \begin{pmatrix} 15,5 & 2432 & 37 \\ 11,8 & 2305 & 20 \\ 14,7 & 2487 & 38 \end{pmatrix} \cdot \begin{pmatrix} 1,16 \\ 0,039 \\ 0,083 \end{pmatrix} = \begin{pmatrix} 115,9 \\ 105,2 \\ 117,2 \end{pmatrix}.$$

Визначимо середні значення векторів дискримінантної функції:

$$\overline{\hat{u}}_x = \frac{1}{n_1} \sum_{i=1}^{n_1} \hat{U}_{x_i} = 147,8;$$

$$\overline{\hat{u}}_y = \frac{1}{n_2} \sum_{i=1}^{n_2} \hat{U}_{y_i} = 112,8,$$

а потім константу дискримінації:

$$\hat{C} = \frac{1}{2} (147,8 + 112,8) = 130,3.$$

Щоб визначити, до якої групи ставляться регіони, що підлягають дискримінації, розрахуємо для них дискримінантні функції

$$\hat{U}_1(Z) = A_1 Z_1 + A_2 Z_2 + A_3 Z_3 = 1,16 \cdot 19,4 + 0,039 \cdot 2558 + 0,083 \cdot 41 = 125,7;$$

$$\hat{U}_2(Z) = 1,16 \cdot 24,2 + 0,039 \cdot 3188 + 0,083 \cdot 31 = 155.$$

Зрівняємо отримані дискримінантні функції з константою дискримінації:

1)  $\hat{U}_1(Z) = 125,7 < \hat{C} = 130,3$ , тому Волинську область не можна віднести до передових регіонів;

2)  $\hat{U}_2(Z) = 155 > \hat{C} = 130,3$ , тому Чернівецьку область слід віднести до передових регіонів.

## **КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТИВНОСТІ РОБОТИ СТУДЕНТІВ НА ПРАКТИЧНОМУ ЗАНЯТТІ**

- міра повноти застосування практичних прийомів і методів аналізу теоретичних положень і концепцій учбової дисципліни;
- міра повноти застосування придбаних студентами умінь і навичок використання сучасних теоретичних методів в рішенні конкретних практичних завдань;
- міра повноти застосування творчого професійного мислення;
- міра повноти використання професійних знань в учбових умовах - оволодіння термінологією відповідної дисципліни;
- міра повноти оволодіння вміннями і навичками постановки і рішення проблем і завдань.



# ПЛАНІ САМОСТІЙНОЇ РОБОТИ

## Завдання для самостійної роботи

Самостійна робота студента включає: опрацювання навчального матеріалу, підготовку до практичних занять, виконання індивідуальних завдань, підготовку до усіх видів контролю.

№ з/п	Види, зміст самостійної роботи (назва теми)	Кількість годин	
		денне	заочне
1	Проробка конспекту лекції, вивчення рекомендованої літератури по темах, які викладаються на лекціях	15	25
2	Підготовка до практичних занять	25	35
3	Підготовка до поточного контролю	20	25
4	Підготовка до підсумкового контролю (екзамену)	38	41
5	Виконання контрольної роботи з курсу	20	20
	<b>Разом</b>	<b>118</b>	<b>146</b>

## Питання для самостійного опрацювання

### Тема 1. Основні поняття методів багатомірної класифікації

1. Що розуміється під класифікацією об'єктів?
2. У чому полягає проблема класифікації об'єктів за багатомірними даними?
3. У якій формі можуть представлятися вихідні дані в задачах класифікації об'єктів?
4. Що таке навчальна вибірка?
5. Які кінцеві прикладні цілі ставить перед собою дослідник при проведенні класифікації?
6. Які фактори називаються типоутворюючими?
7. Як будується комбінаційне угруповання?
8. Яка ідея покладена в основу методу відбору найбільш інформативних ознак-детермінантів?
9. Дати характеристику класифікації як необхідного попереднього етапу статистичної обробки багатомірних даних.
10. Як використовується класифікація в задачах планування вибірових обстежень?
11. Як класифікуються задачі розбивки об'єктів на однорідні групи залежно від наявності апіорної й попередньої вибіркової інформації?

### Тема 2. Кластерний аналіз. Ієрархічні методи класифікації

1. Що таке кластерний аналіз?
2. Яка мета кластерного аналізу?
3. Чим обумовлена необхідність розвитку методів кластерного аналізу і їхнього використання?
4. Які задачі дозволяють вирішувати методи кластерного аналізу?
5. У чому суть агломеративних методів кластерного аналізу?
6. У чому суть дивізімних методів кластерного аналізу?
7. Чому вибір способу обчислення відстані між об'єктами є вузловим моментом дослідження?
8. Які відстані між об'єктами найбільш часто використовуються в задачах кластерного аналізу?
9. Які показники можуть бути використані в якості мір подібності?

10. У чому полягає сутність ієрархічних агломеративних методів?
11. Як будується дендограма?
12. Які відстані між групами об'єктів є найбільш уживаними?
13. Чому якість проведення кластеризації залежить від алгоритму об'єднання в ієрархічних агломеративних методах?

### **Тема 3. Кластерний аналіз. Ітеративні методи класифікації**

1. У чому полягає суть ітеративних методів кластерного аналізу?
2. Яким чином в ітеративних методах задаються початкові умови?
3. Яким чином проводиться класифікація об'єктів методом  $k$ -середніх?
4. До виконання яких кроків зводяться обчислювальні процедури більшості ітеративних методів класифікації?
5. У чому складається істотна відмінність методу пошуку згущень від інших ітеративних методів класифікації?
6. У чому полягає суть ітеративного алгоритму типу «форель»?
7. Якими способами вибирається радіус сфери для пошуку локальних згущень точок у методі пошуку згущень?
8. Для якої мети використовуються критерії якості класифікації?
9. Які найпоширеніші функціонали якості використовуються в кластерному аналізі?
10. Які найпростіші прийоми дозволяють судити про якість розбивки об'єктів на кластери?

### **Тема 4. Дискримінантний аналіз**

1. Що таке дискримінантний аналіз?
2. На які дві групи можна розбити всі процедури дискримінантного аналізу?
3. Які ознаки називаються дискримінантними змінними?
4. Які допущення приймаються в дискримінантному аналізі?
5. Яким повинне бути співвідношення числа об'єктів спостереження й числа дискримінантних змінних?
6. Як визначається канонічна дискримінантна функція?
7. Яким чином визначають коефіцієнти дискримінантної функції?
8. Як визначається границя, що розділяє розглянуті групи?
9. Яким буде алгоритм використання дискримінантного аналізу для проведення багатомірної класифікації об'єктів?
10. Як зміна числа змінних впливає на результат дискримінантного аналізу?
11. На підставі чого судять про доцільність включення (видалення) дискримінантної змінної?
12. Як розраховуються стандартизовані коефіцієнти дискримінантної функції?
13. У яких випадках застосовують стандартизовані коефіцієнти дискримінантної функції?

### **Тема 5. Аналіз даних методами нечіткої кластеризації**

1. Чим обумовлена необхідність розвитку методів кластерного аналізу і їхнього використання?
2. На чому заснований концептуальний зв'язок між кластерним аналізом і теорією нечітких множин?
3. У чому полягає задача нечіткої кластеризації?
4. Ким і коли були запропоновані основні ідеї алгоритму для розв'язування задачі нечіткої кластеризації?
5. Які задачі дозволяють вирішувати методи нечіткої кластеризації?
6. У вигляді яких послідовних кроків можна представити алгоритм FCM?

## МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДО ВИКОНАННЯ САМОСТІЙНОЇ КОНТРОЛЬНОЇ РОБОТИ, ПЕРЕДБАЧЕНОЇ НАВЧАЛЬНИМ ПЛАНОМ

Розрахункова контрольна робота є обов'язковою частиною самостійної роботи студента.

Написання розрахункового завдання передбачає збір, узагальнення та аналіз статистичних даних за обраним напрямом дослідження.

Студенти обирають проблемну ситуацію із запропонованих у переліку або за власним бажанням, збирають необхідні дані, здійснюють аналіз даних із використання методів ті інструментів, що були розглянуті продовж лекційних занять та роблять відповідні висновки.

Розрахункова контрольна робота оцінюється за критеріями:

- самостійності виконання;
- логічності та послідовності викладення матеріалу;
- деталізації плану;
- повноти та глибини розкриття теми, проблемної ситуації, аналітичної частини;
- наявності ілюстрацій (таблиці, рисунки, схеми і т. д.);
- кількості використаних джерел;
- використання статистичної інформації, додаткових літературних джерел та ресурсів мережі Internet;
- відображення практичного досвіду;
- обґрунтованості висновків;
- якості оформлення, презентації та захисту індивідуального розрахункового завдання.

Контрольна робота студента складається з двох частин: перша – опрацювання теоретичного питання, друга – виконання практичного завдання. Максимальна оцінка за виконання завдань контрольної роботи – 30 балів.

Виконання завдань оцінюється за такими критеріями:

- 1) теоретичне питання:
  - повнота й ґрунтовність викладу;
  - аргументованість тверджень;
  - суб'єктне усвідомлення змісту;
  - термінологічна коректність;
- 2) практичне завдання:
  - технологічна грамотність;
  - методична грамотність;
  - обґрунтованість висновків;
  - правильність оформлення.

Шкала оцінювання контрольної роботи

Кількість балів	Теоретичне питання	Практичне завдання
21-30	Повне засвоєння та суб'єктне усвідомлення матеріалу. Твердження чітко аргументовані. Продemonстровано термінологічну грамотність	Продemonстровано методичну й технологічну грамотність. Методичні рішення обґрунтовано. Оформлення відповідає вимогам.
11-20	Часткове засвоєння матеріалу, суб'єктне його не усвідомлення. Нечітка аргументація тверджень. Часткова термінологічна некоректність.	Наявність незначних методичних і технологічних помилок, а також помилок в оформленні роботи. Методичні рішення обґрунтовані частково.

0-10	Теоретичний матеріал не засвоєно. Аргументація відсутня. Термінологічна неграмотність.	Методична й технологічна неграмотність. Неправильне оформлення роботи.
------	--	--

### **3. ПРИКЛАДИ ЗАВДАНЬ СЕМЕСТРОВИХ ЕКЗАМЕНІВ (ПИСЬМОВИХ ЗАЛІКОВИХ РОБІТ)**

## Контрольні питання на екзамен

1. Поняття класифікації. Формалізація загальної задачі класифікації.
2. Комбінаційні угруповання та їх безперервне узагальнення.
3. Проста типологізація.
4. Зв'язна упорядкована типологізація.
5. Структурна типологізація.
6. Класифікація динамічних траєкторій розвитку систем.
7. Типологізація математичних задач класифікації.
8. Загальні поняття про кластерний аналіз.
9. Характеристика методів кластерного аналізу.
10. Відстань між об'єктами та міри близькості: евклідова відстань, зважена евклідова відстань.
11. Відстань між об'єктами та міри близькості: Хемінгова відстань, відстань Чебишева, ступінна відстань.
12. Міри схожості для об'єднання двох кластерів.
13. Метод “найближчого сусіда”.
14. Метод “дальнього сусіда”.
15. Метод середнього зв'язку.
16. Метод медіанного зв'язку.
17. Методи ієрархічного кластерного аналізу: метод одиночного зв'язку.
18. Методи ієрархічного кластерного аналізу: метод повних зв'язків.
19. Методи ієрархічного кластерного аналізу: метод середнього зв'язку.
20. Методи ієрархічного кластерного аналізу: метод Уорда.
21. Загальні поняття про ітеративні методи кластерного аналізу.
22. Математичне описання методу k-середніх.
23. Модифікації методу k-середніх.
24. Алгоритм обчислювальних процедур ітеративних методів класифікації.
25. Модифікації методу пошуку згущень.
26. Оцінка сталості угруповань об'єктів за методом пошуку згущень.
27. Основні поняття про функціонали або критерії якості.
28. Найбільш поширені функціонали якості: сума квадратів відстаней до центрів класів, сума внутрішньокласових відстаней між об'єктами, сумарна внутрішньокласова дисперсія.
29. Загальні положення дискримінантного аналізу.
30. Дискримінантні змінні та шкала відношень.
31. Геометрична інтерпретація дискримінантної функції та дискримінантних змінних.
32. Розрахунок коефіцієнтів дискримінантної функції.
33. Алгоритм класифікації при наявності двох та більше навчаючих вибірок.
34. Взаємозв'язок між дискримінантними змінними та дискримінантними функціями.
35. Постановка задачі нечіткої кластеризації.
36. Концептуальний зв'язок між кластерним аналізом і теорією нечітких множин.
37. Методи нечіткої кластеризації.

**Підсумковий контроль за курсом - у формі екзамену.**

До складання екзамену допускають студентів, що мають задовільну кількість балів із складених тестів з основних навчальних елементів змістовних модулів, написання та захисту індивідуального науково-дослідного завдання та інших завдань передбачених програмою дисципліни.

Екзамен проводиться у відповідності до екзаменаційних білетів, які містять чотири завдання (два теоретичних питання, одне практичне завдання - задача і тести) (рис. 1). Вони дають можливість здійснити оцінювання знань студента за дисципліною.

Харківський національний університет імені В. Н.Каразіна										
Факультет	<u>економічний</u>									
Спеціальність	<u>051 «Економіка»</u>									
Освітня програма	<u>«Бізнес-аналітика та міжнародна статистика»</u> <u>«Економічна аналітика та статистика»</u>									
Семестр	<u>2</u>									
Форма навчання	<u>денна / заочна</u>									
Рівень вищої освіти (освітньо-кваліфікаційний рівень):	<u>магістр</u>									
Навчальна дисципліна: <b>«Методи класифікації даних в пакеті Statistica»</b>										
<b>ЕКЗАМЕНАЦІЙНИЙ БІЛЕТ № <u>1*</u></b>										
1. Дати загальну характеристику методів кластерного аналізу.										
2. Проведіть нормування ознак №№ 4,2.										
	ВАТ "Металургійний ком- бінат "Азовсталь"	ВАТ "Кримський содовий завод"	ВАТ "Азовкабель"	ВАТ "Об'єднання "Склоп- ластик"	ВАТ "Джанкойський ма- шинобудівний завод"	ВАТ "Нікопольський пів- деннотрубний завод"	ВАТ "Нафтопереробний комплекс "Галичина"	ВАТ "Компанія "Дніпро"	ВАТ "АК "Свема"	ВАТ "Нікопольський завод феросплавів"
1	36,8	33,2	34,1	37,7	35,4	36,8	33,0	36,6	32,2	36,2
2	136	122	133	138	131	136	126	134	125	133
3	9,4	6,6	7,4	10	9,3	8,6	4,0	3,5	6,6	0,9
4	0,15	0,48	0,62	0,32	0,41	0,35	0,51	0,45	0,5	0,9
3. За наведеними ненормованими даними визначте Евклідові відстані між підприємствами №№ 1,4: а) 7,2; б) 9,6; в) 10,8; г) 14,4. Відповідь: 1) а; 2) б; 3) в; 4) г.										
4. Метод к-середніх це: а) агломеративний метод; б) дивізімний метод; в) ітеративний метод. Відповідь: 1) а; 2) б; 3) в.										
Затверджено на засіданні кафедри статистики, обліку та аудиту Протокол № _____ від „_____” _____ 20____ року										
Завідувач кафедри	_____ (підпис)				Оксана НЕСТЕРЕНКО (прізвище та ініціали)					
Екзаменатор	_____ (підпис)				Олексій КОРЕПАНОВ (прізвище та ініціали)					
* За завдання 1 – 10 балів, за завдання 2 – 20 балів, за завдання 3, 4 – по 5 балів. Всього – <b>40 балів</b> .										

Рис. 1. Зразок екзаменаційного білету.



### **Критерії оцінювання на екзамені**

Оцінювання знань студента проводиться за чотирирівневою шкалою (відмінно, добре, задовільно, незадовільно). За екзамен студент може отримати максимум 40 балів:

1. Для отримання оцінки «відмінно» (35-40 балів) студент повинен:

- укластися у встановлений строк підготовки відповіді;
- викласти теоретичний матеріал чітко, коротко, зв'язно й обґрунтовано;
- навести вірне рішення задачі та тестів.

2. Для одержання оцінки «добре» (25-34 бала) студент повинен:

- укластися у встановлений строк підготовки відповіді;
- викласти теоретичний матеріал зв'язно й обґрунтовано;
- навести вірне рішення задачі;
- можливі помилки у відповідях на тести.

3. Для отримання оцінки «задовільно» (15-24 бала) студент повинен:

- викладати теоретичний матеріал у доступній для розуміння формі;
- можливі помилки при розв'язанні задачі та в тестах.

4. Оцінку «незадовільно» (1-14 балів) отримують студенти, відповіді яких можуть бути оцінені нижче вимог, сформульованих у попередніх пунктах.

Кожне завдання заліку оцінюється окремо. Загальна оцінка дорівнює сумі оцінок за всі завдання (види контролю).

Підсумкова оцінка з навчальної дисципліни визначається як загальна оцінка, яка враховує оцінки з кожного виду контролю (оцінки поточного та проміжного контролю за роботу протягом семестру, індивідуального завдання та оцінка за результатами підсумкового контролю).

У відповідності до набраних студентом балів оцінка знання матеріалу проводиться за чотирирівневою системою згідно з Методикою переведення показників успішності знань студентів.

#### **Шкала оцінювання**

Сума балів за всі види навчальної діяльності протягом семестру	Оцінка за національною шкалою
	для чотирирівневої шкали оцінювання
90-100	відмінно
80-89	добре
70-79	
60-69	задовільно
50-59	
1-49	незадовільно