

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

ЗАТВЕРДЖЕНО

на засіданні кафедри економічної
кібернетики і системного аналізу
Протокол № 1 від 22.08.2023 р.

ПОГОДЖЕНО

Проректор з навчально-методичної роботи
Каріна НЕМАШКАЛО



ІНСТРУМЕНТАЛЬНІ ЗАСОБИ АНАЛІЗУ ДАНИХ (МОВА R)

робоча програма навчальної дисципліни (РПНД)

Галузь знань 12 "Інформаційні технології"
Спеціальність 124 "Системний аналіз"
Освітній рівень перший (бакалаврський)
Освітня програма Управління складними системами

Статус дисципліни вибіркова
Мова викладання, навчання та оцінювання українська

Розробники:
к.е.н., доцент

Любов ЧАГОВЕЦЬ

викладач

Владислав ПОЛЯНСЬКИЙ

Завідувач кафедри
економічної кібернетики
і системного аналізу

Лідія ГУР'ЯНОВА

Гарант програми

Оксана ПАНАСЕНКО

Харків
2024

ВСТУП

Актуальність навчальної дисципліни та її необхідність та роль у підготовці фахівців.

У сучасному корпоративному оточенні процес прийняття рішень ускладнений великою обсягом слабо структурованої інформації, що часто не має очевидних цінних властивостей. Оцінка цієї інформації часто проводиться без передварних етапів обробки даних, що призводить до втрати можливостей якісного використання кількісної інформації. Використання сучасних методів та моделей аналізу даних, а також застосування технологій машинного навчання можуть забезпечити ефективну обробку таких даних.

Метою навчальної дисципліни «Інструментальні засоби аналізу даних (мова R)» є засвоєння основних ідей сучасного аналізу даних, вивчення сучасних методів та моделей аналізу даних і машинного навчання.

Завданнями навчальної дисципліни є:

засвоєння основних аналізу даних та машинного навчання;

засвоєння основних принципів та підходів до обробки кількісних та якісних даних;

засвоєння прикладних аспектів роботи з великими даними, що описують стан та розвиток складних систем;

засвоєння основних методичних підходів та сучасного теоретичного та практичного базису з побудови моделей та алгоритмів обробки, візуалізації та аналізу даних, побудови рекомендаційних систем та машинного навчання.

Предметом дослідження є методи та моделі аналізу даних, зокрема засоби побудови дескриптивної, предиктивної аналітики, візуалізації даних та машинного навчання.

Об'єктом вивчення даної дисципліни є соціально-економічні системи.

Результати навчання та компетентності, які формує навчальна дисципліна визначено в табл. 1.

Таблиця 1

Результати навчання та компетентності, які формує навчальна дисципліна

Результати навчання	Компетентності, якими повинен оволодіти здобувач вищої освіти
РН 6	КЗ 1, КЗ 2, КЗ 4, КФ 4, КФ 5
РН12	КЗ 1, КЗ 2, КФ 7
РН14	КЗ 1, КЗ 2, КЗ 4, КФ 3, КФ 4, КФ 10

де, РН6. Знати та вміти застосовувати основні методи постановки та вирішення задач системного аналізу в умовах невизначеності цілей, зовнішніх умов та конфліктів.

РН12. Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу.

PH14. Розуміти і застосовувати на практиці методи статистичного моделювання і прогнозування, оцінювати вихідні дані.

КЗ 1. Здатність до абстрактного мислення, аналізу та синтезу.

КЗ 2. Здатність застосовувати знання у практичних ситуаціях.

КЗ 4. Знання та розуміння предметної області та розуміння професійної діяльності.

КФ 3. Здатність будувати математично коректні моделі статичних та динамічних процесів і систем із зосередженими та розподіленими параметрами із врахуванням невизначеності зовнішніх та внутрішніх факторів.

КФ 4. Здатність визначати основні чинники, які впливають на розвиток фізичних, економічних, соціальних процесів, виокремлювати в них стохастичні та невизначені показники, формулювати їх у вигляді випадкових або нечітких величин, векторів, процесів та досліджувати залежності між ними

КФ 5. Здатність формулювати задачі оптимізації при проектуванні систем управління та прийняття рішень, а саме: математичні моделі, критерії оптимальності, обмеження, цілі управління; обирати раціональні методи та алгоритми розв'язання задач оптимізації та оптимального керування.

КФ 7. Здатність використовувати сучасні інформаційні технології для комп'ютерної реалізації математичних моделей та прогнозування поведінки конкретних систем а саме: об'єктно-орієнтований підхід при проектуванні складних систем різної природи, прикладні математичні пакети, застосування баз даних і знань.

КФ 10. Здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Зміст навчальної дисципліни

Змістовий модуль 1. Вступ до аналітики даних та інструментальні засоби обробки даних

Тема 1. Базові поняття аналізу даних

Поняття даних. Історичний поступ методів аналітики. Особливості обробки та аналізу даних. Інструменти та методи аналізу даних баговимірної структури. Архітектура та екосистема обробки великих даних.

Етапи моделювання та дослідження даних. Компоненти рішення даних. Можливості аналітики даних на R

Тема 2. Розвідувальний аналіз даних та передобробка.

Розподіл даних та вибірок. Сутність і причини виникнення помилок в даних.

Методи дослідження похибок в даних.

Методи бутстрепа.

Тема 3. Візуалізація аналіз даних.

Основи візуалізації даних.

Сучасний інструментарій для візуалізації даних.

Закони розподілу та використання основних функцій для роботи з розподілами.

Змістовий модуль 2. Алгоритми машинного навчання

Тема 4. Методи статистичного навчання

Особливості задач статистичного машинного навчання.

Проблеми побудови регресійних моделей на великих обсягах даних.

Підходи до статистичного навчання. Методи побудови та дослідження регресійних моделей. Оцінка параметрів моделі. Критерії якості моделі. Дослідження нетипових помилок моделі.

Тема 5. Ансупервайзорні методи машинного навчання

Задача навчання без учителя. Реалізація методів кластерного аналізу.

Найпоширеніші алгоритми кластеризації у R (K-Means та GMM). Основи кластеризації та візуалізації результатів.

Тема 6. Супервайзорні методи та алгоритми класифікації

Навчання з учителем. Особливості різних видів алгоритмів класифікації. Байєсовський класифікатор. KNN.

Оцінка якості класифікації. Види помилок. Проблема вибору моделі.

Дерева класифікації та ансамблі дерев. Деревоподібні моделі пошуку логічних закономірностей. Алгоритм CART.

Ансамблеве навчання та випадкові ліси.

Тема 7. Методи просторового скорочення

Методи зниження розмірності. Інтерпретація даних в задачах великої розмірності.

Побудова класифікаторів на опорних векторах. Машини опорних векторів (SVM). SVM з декількома класами. Взаємозв'язок з логістичною регресією.

Перелік лабораторних занять та завдань за навчальною дисципліною наведено в табл. 2.

Таблиця 2

Перелік лабораторних занять / завдань

Назва теми та / або завдання	Зміст
Лабораторна робота 1	Знайомство з R та RStudio. Робота з таблицями даних у R
Лабораторна робота 2	Передобробка даних та візуалізація в R
Лабораторна робота 3	Кореляційно-регресійний аналіз в R. Відбір та регуляризація лінійних моделей в R
Лабораторна робота 4	Кластеризація в R. Супервайзорні та ансупервайзорні методи класифікації
Лабораторна робота 5	Методи просторового скорочення

Перелік самостійної роботи за навчальною дисципліною наведено в табл. 3.

Перелік самостійної роботи

Назва теми	Зміст
Тема 1.	Огляд літературних джерел, опрацювання презентаційних матеріалів за вивченою темою, підготовка до практичних занять, виконання індивідуальних завдань теми «Знайомство з R та RStudio. Робота з таблицями даних у R»
Тема 2.	Огляд літературних джерел, опрацювання презентаційних матеріалів за вивченою темою, підготовка до практичних занять, виконання індивідуальних завдань теми «Передобробка даних»
Тема 3.	Огляд літературних джерел, опрацювання презентаційних матеріалів за вивченою темою, підготовка до практичних занять, виконання індивідуальних завдань теми «Візуалізація в R»
Тема 4.	Огляд літературних джерел, опрацювання презентаційних матеріалів за вивченою темою, підготовка до практичних занять, виконання індивідуальних завдань теми «Відбір та регуляризація лінійних моделей в R»
Тема 5.	Огляд літературних джерел, опрацювання презентаційних матеріалів за вивченою темою, підготовка до практичних занять, виконання індивідуальних завдань теми «Ансупервайзорні методи класифікації»
Тема 6.	Огляд літературних джерел, опрацювання презентаційних матеріалів за вивченою темою, підготовка до практичних занять, виконання індивідуальних завдань теми «Супервайзорні методи класифікації»
Тема 7.	Огляд літературних джерел, опрацювання презентаційних матеріалів за вивченою темою, підготовка до практичних занять, виконання індивідуальних завдань теми «Методи просторового скорочення»

Кількість годин лекційних та лабораторних занять та годин самостійної роботи наведено в робочому плані (технологічній карті) з навчальної дисципліни.

МЕТОДИ НАВЧАННЯ

У процесі викладання навчальної дисципліни для набуття визначених результатів навчання, активізації освітнього процесу передбачено застосування таких методів навчання, як: словесні: лекція (усі теми); наочні: презентація (усі теми); практичні: лабораторні роботи (усі теми), робота в малих групах (Тема 1, 3, 5-7).

Під час викладання дисципліни "Інструментальні засоби аналізу даних (мова R)" для активізації навчального процесу передбачено застосування сучасних навчальних технологій, таких, як робота в малих групах. Розділ форм та методів активізації процесу навчання за темами навчальної дисципліни наведено у таблиці 4.

**Розподіл форм та методів активізації процесу навчання
за темами навчальної дисципліни**

Тема	Практичне застосування навчальних технологій
Тема 1.	Робота в малих групах
Тема 2.	Презентація результатів роботи в малих групах за темою заняття
Тема 3.	Робота в малих групах
Тема 4.	Презентація результатів роботи в малих групах за темою заняття
Тема 5.	Робота в малих групах
Тема 6.	Робота в малих групах
Тема 7.	Робота в малих групах за темою заняття

Робота в малих групах дає змогу структурувати лекційні або практично-семінарські заняття за формою і змістом, створює можливості для участі кожного здобувача в роботі за темою заняття, забезпечує формування особистісних якостей та досвіду соціального спілкування. Після висвітлення проблеми або стислого викладання матеріалу здобувачам пропонується об'єднуватися у групи по 2 – 3 особи та презентувати своє бачення та сприйняття матеріалу.

ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

Університет використовує 100 бальну накопичувальну систему оцінювання результатів навчання здобувачів вищої освіти.

Поточний контроль здійснюється під час проведення лекційних, лабораторних занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретної роботи і оцінюється сумою набраних балів:

– для дисциплін з формою семестрового контролю екзамен (іспит): максимальна сума – 60 балів; мінімальна сума, що дозволяє здобувачу вищої освіти скласти екзамен (іспит) – 35 балів.

Підсумковий контроль включає семестровий контроль та атестацію здобувача вищої освіти.

Семестровий контроль проводиться у формах семестрового екзамену (іспиту). Складання семестрового екзамену (іспиту) здійснюється під час екзаменаційної сесії.

Максимальна сума балів, яку може отримати здобувач вищої освіти під час екзамену (іспиту) – 40 балів. Мінімальна сума, за якою екзамен (іспит) вважається складеним – 25 балів.

Підсумкова оцінка за навчальною дисципліною визначається сумуванням балів за поточний та підсумковий контроль.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: індивідуальні завдання за лабораторними роботами (33 бали), письмова контрольна робота (10 балів), експрес-опитування (12 балів), презентація (5 балів).

Семестровий контроль: екзамен (40 балів)

Більш детальну інформацію щодо системи оцінювання наведено в робочому плані (технологічній карті) з навчальної дисципліни.

Приклад екзаменаційного білета.

Харківський національний економічний університет імені Семена Кузнеця	
Кафедра	Економічної кібернетики і системного аналізу
Навчальна дисципліна	«Інструментальні засоби аналізу даних (мова R)»
Спеціальність 124	«Системний аналіз»
Освітня програма	«Управління складними системами»

Екзаменаційний білет

Завдання 1.

1. Значення коефіцієнта ексцесу більше нуля характерно для
Виберіть одну відповідь:

- a. низьковершинного розподілу
- b. середньовершинного розподілу
- c. високовершинного розподілу

2. Значення коефіцієнта ексцесу більше нуля характерно для
Виберіть одну відповідь:

- a. низьковершинного розподілу
- b. середньовершинного розподілу
- c. високовершинного розподілу

3. Дисперсія дозволяє оцінити:
Виберіть одну відповідь:

- a. міру розсіювання
- b. частоту розподілу
- c. розмах варіації
- d. групування

4. Розрахунок критерію Пірсона засновано на порівнянні
Виберіть одну відповідь:

- a. критичних і теоретичних частот
- b. накопичених частот
- c. розрахункових і теоретичних частот
- d. емпіричних і теоретичних частот

5. У регресійній моделі факторними ознаками називаються:
Виберіть одну відповідь:

- a. залежні змінні
- b. незалежні змінні
- c. немає правильної відповіді
- d. ендогенні змінні

6. Для перевірки статистичної значимості регресійної моделі в цілому використовується критерій:

- Виберіть одну відповідь:
- a. Стьюдента

- b. Фішера
- c. Стандартної помилки
- d. Хі-квадрат

7. Рівняння регресії має вигляд $Y=2,02+0,78X$. На скільки одиниць свого виміру в середньому зміниться Y при збільшенні X на одну одиницю свого виміру:

Виберіть одну відповідь:

- a. збільшиться на 0,78;
- b. збільшиться на 2,02;
- c. не зміниться
- d. збільшиться на 2,80;

8. З урахуванням співвідношення між доходом Y (тис. грн) та числом робітників X (люд): $Y= 8,65 + 268*X$, магазин, прийнявши на роботу додатково одну людину, може очікувати такий додатковий дохід:

Виберіть одну відповідь:

- a. $8,65 + 268$
- b. 8,65
- c. немає правильної відповіді
- d. 268

9. Параметр a_0 рівняння регресії визначає:

Виберіть одну відповідь:

- a. тангенс кута нахилу прямої до осі абсцис
- b. тангенс кута нахилу прямої до осі ординат
- c. точку перетину прямої з віссю ординат
- d. точку на перетині прямої з віссю абсцис

10. Який тип даних дозволяє поєднувати цифровий та текстовий формат даних:

Виберіть одну відповідь:

- a. масив
- b. матриця
- c. список
- d. вектор

11. Це є зібранням функцій R, даних та скомпільованого програмного коду у певному форматі

Виберіть одну відповідь:

- a. Пакет
- b. Робочий простір
- c. Директорія

12. Чи вірно, що деякі R-пакети мають власні функції, аналогічні стандартній `summary()`, для виведення компактних описових зведень за таблицями даних.

Виберіть одну відповідь:

- Правильно
- Неправильно

13. Директорію, в якій пакети R зберігаються на комп'ютер можна вважати

Виберіть одну відповідь:

- a. Робочим простором
- b. Бібліотекою
- c. Датафреймом

14. У якості міри подібності в кластерному аналізі використовуються

Виберіть одну відповідь:

- a. міра відстані, коефіцієнти кореляції, коефіцієнти асоціативності
- b. тільки коефіцієнти кореляції
- c. тільки міра відстані

15. Для оцінки якості класифікації використовуються такі функціонали:

Виберіть одну або декілька відповідей:

- a. найближчого сусіда
- b. сумарна внутрішньокласова дисперсія
- c. сума внутрішньокласових відстаней
- d. сума квадратів відстаней до центрів класів
- e. функціонал Бартлета

16. Під час використання агломеративного методу групування на першому кроці всі об'єкти

Виберіть одну відповідь:

- a. розглядаються як самостійні кластери
- b. належать двом кластерам

17. Об'єднана інформація про атрибути розщеплення і предиката розщеплення в вузлі називається

Виберіть одну відповідь:

- a. критерієм якості
- b. атрибутом прогнозування
- c. атрибутом розподілу
- d. критерієм розщеплення

18. Якість побудованого дерева вище, якщо

Виберіть одну відповідь:

- a. Ціна крос-перевірки вище, ніж ціна помилки навчання
- b. Ціна крос-перевірки мінімальна, ціна помилки навчання перевищує помилку крос-перевірки
- c. Ціна крос-перевірки мінімальна, ціна помилки навчання мінімальна

19. Відповідно до критерію відсіювання Р. Кеттела кількість факторів визначається

Виберіть одну відповідь:

- a. відповідно до величини навантаження на фактори
- b. за точкою перегину на графіку власних значень
- c. аналітично на графіку розсіювання факторів власних значень

20. Репрезентативні ознаки мають такі властивості

Виберіть одну відповідь:

- a. всі перераховані вище відповіді вірні
- b. некорельовані з ознаками, що належать до діагностичного набору
- c. ознаки некорельовані між собою

Завдання 2.

У таблиці по восьми підприємствам легкої промисловості наведено значення показників рентабельності (x_1) та продуктивності праці (x_2).

Потрібно провести кластерний аналіз та визначити кластерну структуру даних. Зокрема:

1. Провести стандартизацію даних;
2. Побудувати графік "кам'янистого осипу" (або дендрограму) за методом Уорда і зробити припущення про кількість кластерів;
3. Побудувати кластерне розбиття методом K-means (можна іншим) (згідно до визначеної

кількості кластерів у пункті 2);

4. Візуалізувати (наприклад, візуальним розміщенням кластерів в R);

5. Визначити конкретні елементи кожного кластеру та дати умовну характеристику кожної групи (тобто через які фактори конкретний кластер виділяється з-поміж інших);

6. Сформулювати висновки по результатам.

Підприємство	X1	X2
П 1	2,95	5,6
П 2	3,1	7,1
П 3	1,65	4,5
П 4	2,7	9,8
П 5	1,5	9,2
П 6	1,15	5,7
П 7	1,75	4,7
П 8	1,25	4,2

Критерії оцінювання

Відповіді студентів оцінюються за системою згідно з кваліфікаційними вимогами до бакалаврів спеціальності 124 «Системний аналіз». Виконання кожного екзаменаційного завдання оцінюється таким чином (табл. 5):

Таблиця 5

Оцінювання екзаменаційних завдань

Завдання	Кількість балів
Завдання 1	20
Завдання 2	20
Разом	40

Перше завдання спрямоване на оцінку когнітивних знань студента за дисципліною, що дозволяє визначити рівень володіння навчальним матеріалом, для цього використовуються наступні критерії (табл. 6):

Таблиця 6

Критерії оцінювання завдання 1

Кількість вірних відповідей	Кількість балів	Кількість вірних відповідей	Кількість балів
1	1	11	11
2	2	12	12
3	3	13	13
4	4	14	14
5	5	15	15
6	6	16	16
7	7	17	17
8	8	18	18
9	9	19	19
10	10	20	20

Друге завдання спрямоване на оцінку компетентностей, пов'язаних з умінням застосовувати отримані знання для вирішення широкого класу задач із застосуванням інструментальних засобів, зокрема мови R. Оцінювання завдання здійснюється у відповідності з критеріями (табл. 7)

Таблиця 7

Бал	Вимоги до відповіді на завдання
20	На поставлені завдання дано правильні відповіді. Немає логічних та арифметичних помилок, розв'язання виконано в логічній послідовності. Хід розв'язання задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання. Наведено логічні висновки про результати розрахунків.
19	На поставлені завдання дано відповіді з помилками арифметичного характеру. Розв'язання виконано в логічній послідовності. Хід розв'язання задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання. Наведено логічні висновки про результати розрахунків.
18	На всі поставлені завдання дано відповіді з помилками арифметичного та логічного характеру, що не впливають на загальний відповідь завдань. Хід рішення задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання. Логічні висновки результатів розрахунків відсутні.
17	На всі поставлені завдання дано відповіді із незначними помилками арифметичного та логічного характеру, що не впливають на загальний відповідь завдань. Хід рішення задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання. Логічні висновки результатів розрахунків відсутні.
16	На всі поставлені завдання дано відповіді із незначними помилками арифметичного та незначними помилками логічного характеру, що не впливають на загальний відповідь завдань. Хід рішення задачі не супроводжується поясненнями, які використовують понятійний апарат методу розв'язання
15	На всі поставлені завдання дано відповіді із незначними помилками арифметичного та логічного характеру, що впливають на загальний відповідь завдань. Хід рішення задачі супроводжується поясненнями, які використовують понятійний апарат методу розв'язання
14	Наведено рішення завдання формальна постановка і загальний алгоритм рішення, однак, рішення не містить істотні логічні помилки, що впливають на підсумковий результат, остаточні висновки вірні.
13	Наведено рішення завдання формальна постановка і загальний алгоритм рішення, однак, рішення містить логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні. Хід рішення задачі супроводжується поясненнями
12	Наведено рішення завдання формальна постановка і загальний алгоритм рішення, однак, рішення містить істотні арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні. Хід рішення задачі не супроводжується поясненнями
11	Наведена формальна постановка і загальний алгоритм рішення, однак, рішення містить істотні арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні
10	Наведена формальна постановка рішення, рішення містить істотні логічні та арифметичні помилки, що впливають на підсумковий результат, остаточні висновки не вірні.

9	Подано повну математичну постановку задачі та описано вірно алгоритм оцінювання, виконано розрахунки із незначними логічними помилками, технічні розрахунки виконано невірно, висновків немає
8	Подано повну математичну постановку задачі та описано вірно алгоритм оцінювання, виконано розрахунки із деякими логічними помилками, технічні розрахунки виконано невірно, висновків немає
7	Подано повну математичну постановку задачі та описано вірно алгоритм оцінювання, виконано розрахунки із багатьма значними логічними помилками, висновків немає
6	Наведено повну математичну постановку задачі та повну етапність алгоритму оцінювання
5	Подано повну математичну постановку задачі та повна етапність алгоритму оцінювання
4	Наведено повну математичну постановку та основні етапи алгоритму оцінювання
3	Подано повну математичну постановку та початок алгоритму оцінювання
2	Наведено неповну математичну постановку та початок алгоритму оцінювання
1	Здобувач зміг записати неповну математичну постановку задачі
0	Завдання не розв'язано

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Актуальні проблеми системного аналізу та моделювання процесів управління / За ред. В. Пономаренка, Л. Гур'янової, Я. Пеліової, Е. Ніжинського – Братислава-Харків, ВШЕМ – ХНЕУ ім. С. Кузнеця, 2023. – 409 с. Режим доступу: <http://repository.hneu.edu.ua/handle/123456789/29952>.

2. Полянський В. О. Моделі класифікації країн по типу економічної динаміки. Сучасні проблеми моделювання соціально-економічних систем : Матеріали XIII міжнар. науково-практ. Інтернет-конф., м. Братислава – Харків, 8-9 квіт. 2021 р. Харків, 2021. URL: <https://mpsesm.org/book/2021/thesis04-979.html>

Додаткова

3. Ch V Raghavendran. Hand Book on R ProgrammingScholars' Press, 2021. -- 124 p.

4. Chagovets L., Chahovets V., Chernova N. Machine Learning Methods Applications for Estimating Unevenness Level of Regional Development. DataCentric Business and Applications. Evolvments in Business Information Processing and Management. (Volume 3), Springer, Cham, 2020. – Pp. 115-139.

5. Dayal V. Wrangling Data. Quantitative Economics with R. Singapore, 2021. URL: https://doi.org/10.1007/978-981-16-3434-5_3

6. Einspruch, Eric L. An Introductory Guide to R: Guilford Publications, 2022. – 196 p.

7. Чаговець Л. О., Чаговець В. В., Діденко А. С. Технологія Data Mining у моделюванні нерівномірності соціально-економічного розвитку регіонів. Бізнес Інформ. 2020. № 3. С. 82-91.

8. Jorge Munoz-Gama, Xixi Lu. Process Mining Workshops ICPM 2021 International Workshops, Eindhoven, The Netherlands, October 31 – November 4, 2021, Revised Selected Papers in Lecture Notes in Business Information Processing (2022).

9. Sergio Consoli, Diego Reforgiato Recupero, Michaela Saisana. Data Science for Economics and Finance. Methodologies and Applications: Springer; 1st ed., 2021. – 369 p.

10. Silvia Bozza, Franco Taroni. Bayes Factors for Forensic Decision Analyses with R. Springer Texts in Statistics (2022).

11. Прокопович С. В., Чаговец Л.О., Холод В.А. Застосування методів Data Science у комплексному оцінюванні економічного розвитку регіонів. Управління розвитком. 2020. № 3. С. 43-56

Інформаційні ресурси

12. Електронний курс «Інструментальні засоби аналізу даних (мова R)». – [Електронний ресурс]. – Режим доступу: <https://pns.hneu.edu.ua/course/view.php?id=9758>.