

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ**
СИСТЕМА УПРАВЛІННЯ ЯКІСТЮ
Система забезпечення якості освітньої діяльності та якості вищої освіти
сертифікована на відповідність ДСТУ ISO 9001:2015 / ISO 9001:2015

Кафедра цифрової економіки та системного аналізу

ЗАТВЕРДЖЕНО

вченою радою

(пост. п. 9 від «23» жовтня 2021 р.)

Ректор

 А. Мазаракі



**ІНСТРУМЕНТАЛЬНІ ЗАСОБИ
БІЗНЕС-АНАЛІТИКИ /
BUSINESS INTELLIGENCE TOOLS**

**РОБОЧА ПРОГРАМА /
COURSE OUTLINE**

освітній ступінь	бакалавр / bachelor
галузь знань	12 Інформаційні технології / Information Technologies
спеціальність	124 Системний аналіз / System analysis
спеціалізація	Інформаційні технології та бізнес-аналітика (Data Science) / Information technologies and business intelligence (Data Science)

Київ 2021

**Розповсюдження і тиражування без офіційного дозволу КНТЕУ
заборонено**

Автор: А. А. Роскладка, доктор економічних наук, професор

Робочу програму розглянуто і затверджено на засіданні кафедри
цифрової економіки та системного аналізу 08.11.2021р., протокол № 3

Рецензенти: О. А. Харченко, кандидат технічних наук, доцент,
В. В. Кулаженко, кандидат економічних наук,
М. Г. Шарафутдінов, директор з розвитку компанії «Center
Research & Development», бізнес-аналітик

**ІНСТРУМЕНТАЛЬНІ ЗАСОБИ
БІЗНЕС-АНАЛІТИКИ /
BUSINESS INTELLIGENCE TOOLS**

**РОБОЧА ПРОГРАМА /
COURSE OUTLINE**

освітній ступінь	бакалавр / bachelor
галузь знань	12 Інформаційні технології / Information Technologies
спеціальність	124 Системний аналіз / System analysis
спеціалізація	Інформаційні технології та бізнес-аналітика (Data Science) / Information technologies and business analytics (Data Science)

**1. СТРУКТУРА ДИСЦИПЛІНИ ТА РОЗПОДІЛ ГОДИН ЗА ТЕМАМИ
(ТЕМАТИЧНИЙ ПЛАН)**

Освітня програма «Інформаційні технології та бізнес-аналітики»

Для вивчення дисципліни відводиться 180 год. / 6 кредитів ЄКТС (лекцій – 28 год., лабораторних занять – 56 год., самостійної роботи – 96 год.), підсумковий контроль – екзамен

Назва теми	Кількість годин				Форма контролю
	Усього годин / кредитів	з них:			
		лекції	лабораторні заняття / МК	самостійна робота студентів	
Тема 1. Моделювання даних у бізнес-аналітиці	38	8	14	16	Т, ІДЗ, П, Пр
Тема 2. Управління даними в R	34	6	14	14	Т, ІДЗ, П, Пр, МК
Тема 3. Описова та індуктивна бізнес-аналітика	34	6	8	20	Т, ІДЗ, П, Пр
Тема 4. Візуалізація даних та результатів бізнес-аналізу	28	4	8	16	Т, ІДЗ, П, Пр
Тема 5. Прогностична бізнес-аналітика	24	2	12	10	Т, ІДЗ, Пр, П, МК
Тема 6. Розширена бізнес-аналітика	22	2	0	20	Т, Пр
Разом	180/6	28	56	96	х
Підсумковий контроль – екзамен					

Примітка: Т – тестування; ІДЗ – виконання індивідуальних домашніх завдань; П – перевірка індивідуальних завдань; Пр. – презентація індивідуального завдання; МК – модульний контроль.

2. ТЕМАТИКА ТА ЗМІСТ ЛЕКЦІЙНИХ, ПРАКТИЧНИХ (СЕМІНАРСЬКИХ) ЗАНЯТЬ І САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ

Результати навчання	Навчальна діяльність ¹	Робочий час здобувача, год
1	2	3
<p>Знати: Засоби представлення даних у середовищі R.</p> <p>Вміти: працювати з інтерфейсом RStudio та основними пакетами R.</p>	<p style="text-align: center;">Тема 1. Моделювання даних у бізнес-аналітиці.</p> <p style="text-align: center;">Лекція 1. Основні поняття бізнес-аналітики. Мова R.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Аналітичні дані. Види аналітики. 2. Актуальні бізнес-задачі аналітики даних. 3. Знайомство з R. Доступ до інсталяції програмного середовища. 4. Редагування та виконання коду. 5. Принципи пакетної організації обчислень 	2
	<p style="text-align: center;">Лекція 2. Графічний інтерфейс GUI для R. RStudio.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Робота з консоллю. 2. Термінал RStudio. Редагування та виконання коду в терміналі. 3. Автоматичне створення функцій. 4. Коментування коду. Історія команд. 5. Робочий простір Workspace. 	2
	<p style="text-align: center;">Лекція 3. Об'єктно-орієнтоване програмування в R. Пакети gWidgets2 та Shiny.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Розробка графічного інтерфейсу для скриптів R. 2. Встановлення пакетів сімейства gWidgets2. 3. Візуальні об'єкти gWidgets2 та засоби керування ними. 4. Створення інтерактивних веб-додатків мовою R. 5. Інтерфейс користувача та віджети. 6. Використання реактивних виразів Shiny. 	2
	<p style="text-align: center;">Лекція 4. Імпорт, експорт, введення та виведення даних в R.</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Імпорт аналітичних даних з текстового файлу, файлу Excel, XML-файлу. 2. Добування бізнес-даних із веб-сторінок. 3. Імпорт даних із систем управління базами даних. 4. Введення даних з клавіатури. 5. Виведення даних та експорт результатів аналізу. 	2
	<p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-5. <i>Додатковий:</i> 8-10, 12-17, 19-22, 24, 26-29. <i>Інтернет-ресурси:</i> 30-38.</p>	

¹+20% інтерактивних методів навчання виділено курсивом

	<p style="text-align: center;">Лабораторне заняття 1 Робота в Rstudio</p> <p>Мета: ознайомитися з основами роботи у візуальному середовищі <i>RStudio</i>.</p> <p>Завдання: <i>Встановити останні версії програм R і RStudio. Вивести список пакетів, встановлених за замовчуванням. Перевірити, які дані наявні у цих пакетах для тестових цілей. Вивести документацію за визначеними датасетами. Проаналізувати роботу з областю «Environment». Завантажити пакети у робоче середовище. Провести тестові розрахунки в терміналі: додавання, віднімання, множення і ділення. Створити R Script і R Notebook. Проаналізувати base пакет.</i></p>	2
	<p style="text-align: center;">Лабораторні заняття 2-3 Створення графічного інтерфейсу користувача в R</p> <p>Мета: ознайомитися із принципами побудови графічного інтерфейсу програм в R за допомогою сімейства пакетів <i>gWidgets2</i>.</p> <p>Завдання: <i>Створити візуальні елементи <code>gwindow</code> та <code>gbasicdialog</code>. Дослідити властивості візуальних елементів: кнопки <code>gbutton</code>, мітки <code>glabel</code>, текстового поля <code>gedit</code>, альтернативного списку <code>gradio</code>, випадючого списку <code>gcombobox</code>, множинного вибору <code>gcheckboxgroup</code>, календаря <code>gcalendar</code>, вікна вибору папки або файлу <code>gfile</code>. Побудувати функції-реакції на дію із візуальними елементами.</i></p>	4
	<p style="text-align: center;">Лабораторні заняття 4-5 Основні підходи до розробки веб-програм у середовищі R</p> <p>Мета: створення інтерактивних веб-додатків у середовищі R з використанням пакету <i>Shiny</i>.</p> <p>Завдання: <i>Створити новий проект в Shiny. Побудувати простий інтерфейс користувача. Додати основні віджети із бібліотеки Shiny. Визначити реакції на дії користувача. Використати скрипти R і файли даних для роботи у веб-середовищі. Створити та дослідити реактивні вирази в Shiny. Вивчити методи поширення створених веб-додатків.</i></p>	4
	<p style="text-align: center;">Лабораторне заняття 6 Імпорт даних та експорт результатів аналізу</p> <p>Мета: вивчення методів імпорту даних до системи R та експорту результатів аналізу із середовища R до інших систем обробки інформації.</p> <p>Завдання: <i>Здійснити імпорт даних із текстового файлу з роздільниками з використанням команди <code>read.table()</code>, <code>read.csv()</code>, <code>read.csv2()</code>, <code>read.delim()</code>, <code>read.delim2()</code>. Налаштувати параметри імпорту. Використати різні типи роздільників у файлі імпорту. Здійснити імпорт даних із середовища Excel з використанням пакетів <i>RODBC</i>, <i>xlsx</i>, ...</i></p> <p><i>Здійснити імпорт з XML-файлу. Проаналізувати властивості функцій <code>readLines()</code>, <code>grep()</code>, <code>gsub()</code> для веб-скрепінгу. Вивчити можливості імпорту файлів із <i>SPSS</i>, <i>SAS</i>, <i>Stata</i>, <i>netCDF</i>, <i>HDF5</i>. Дослідити можливості системи R при імпорті даних із СУБД <i>Microsoft SQL Server</i>, <i>Microsoft Access</i>, <i>MySQL</i>, <i>Oracle</i>, <i>PostgreSQL</i>,</i></p>	2

	<p>3. Числові і текстові функції. 4. Функції для роботи із символьними даними 5. Функції для роботи з об'єктами</p> <p style="text-align: center;">Лекція 7. Програмування процедур обробки даних в R. План лекції</p> <p>1. Умовні оператори. Функції if і switch 2. Управління потоками в R. Цикли з використанням for, while і repeat 3. Функції користувача. Аргументи та результати функцій. 4. Сімейство apply функцій</p> <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-5. <i>Додатковий:</i> 8-10, 12-17, 19-22, 24, 26-29. <i>Інтернет-ресурси:</i> 30-38.</p>	2
	<p style="text-align: center;">Лабораторні заняття 8-9 Типи даних в R</p> <p>Мета: набути практичних навичок використання різних типів даних в R. Завдання: <i>Визначити функції оголошення різних типів даних: текстові – is.character(); числові – is.numeric(); цілочислові – is.integer(); логічні – is.logical(); комплексні числа – is.complex(). Застосувати функцію typeof() для визначення типу даних. Виконати практичні завдання із застосуванням різноманітних структур даних: скалярів, векторів, матриць, масивів даних, таблиць даних, списків.</i></p> <p style="text-align: center;">Лабораторне заняття 10 Управління даними в R</p> <p>Мета: засвоїти основні процедури управління даними в R. Завдання: <i>Провести аналіз структури датасету. Перевірити наявність спеціальних значень даних: NA (Not Available), NaN (Not arithmetic Number), Inf (Infinity), NULL. Виконати перекодування даних. Здійснити сортування за визначеними полями набору даних. Відфільтрувати дані за визначеною умовою. Провести об'єднання векторів за допомогою функції cbind(). Здійснити розбиття структури даних на частини з використанням функції split().</i></p> <p style="text-align: center;">Лабораторне заняття 11 Вбудовані функції в R</p> <p>Мета: набути практичних навичок використання вбудованих функцій в системі R. Завдання: <i>Розрахувати значення виразу з використанням арифметичних функцій та операторів. Використати логічні оператори для створення складних структур в R. Перетворення символів та тексту за допомогою функцій для роботи із символьними даними. Провести аналіз структур в R, використовуючи функції для роботи з об'єктами.</i></p>	4 2 2

	<p style="text-align: center;">Лабораторні заняття 12-13 Основи програмування в R</p> <p>Мета: ознайомитися з принципами програмування обчислювальних процедур для роботи з даними в R. Завдання: Вивчити алгоритм створення власних функцій в R, правила записів аргументів та значень функцій. Створити власні функції користувача. Набути практичних навичок використання різних структур управління даними: умовні конструкції – функції <i>if()</i>, <i>if-else()</i>, <i>switch()</i>; цикли – <i>for</i>, <i>while</i>, <i>repeat</i>; сімейство <i>apply</i> функцій – <i>apply()</i>, <i>lapply()</i>, <i>sapply()</i>, <i>rapply()</i>, <i>tapply()</i>, <i>replicate()</i>, <i>by()</i>, <i>outer()</i>.</p> <p style="text-align: center;">Лабораторне заняття 14 Модульний контроль 1 «Управління даними в бізнес-аналітиці»</p> <p>Мета: перевірити рівень знань, отриманих студентами при вивчення теми «Управління даними в R». Завдання: Встановити потрібні пакети для обробки даних. Завантажити набір даних. Провести аналіз датасету. Вивести весь датасет та вказані його частини на екран. Сформувати допоміжний датасет, який містить певні поля початкового набору даних. Упорядкувати датасет за визначеним критерієм і вивести на екран упорядкований набір. Провести фільтрацію даних датасету за певною умовою. Експортувати датасет у текстовий файл та файл Excel.</p>	4
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Управління даними в R», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульного контролю.</p> <p><i>Індивідуальна робота:</i> Розробити шаблон веб-додатку за допомогою пакету Shiny. В розробленому додатку передбачити процедури введення даних та засобів виведення результатів аналізу у вигляді текстової та графічної інформації, засоби керування діями в проєкті за допомогою командних кнопок Action Button. Використати розроблений шаблон для виконання лабораторних робіт за темою «Управління даними в R»</p> <p style="text-align: center;">Список рекомендованих джерел: <i>Додатковий:</i> 12, 29. <i>Інтернет-ресурси:</i> 30-32, 37-38.</p>	14
<p>Знати: основні методи описової та індуктивної аналітики</p> <p>Вміти: використовувати функції R для опису структури набору даних;</p>	<p style="text-align: center;">Тема 3. Описова та індуктивна бізнес-аналітика</p> <p style="text-align: center;">Лекція 8. Описова аналітика даних в R</p> <p style="text-align: center;">План лекції</p> <ol style="list-style-type: none"> 1. Описова статистика даних бізнес-процесів. 2. Оцінка вибіркового параметрів. 3. Найбільш поширені помилки при виконанні статистичного аналізу. 4. Таблиці частот та спряженості. Кореляції. 5. Перевірка статистичної значущості кореляцій. Візуалізація кореляцій. 	2

<p>здійснювати перевірку статистичних гіпотез в середовищі R.</p>	<p>Лекція 9. Моделювання законів розподілу випадкових даних План лекції</p> <ol style="list-style-type: none"> 1. Закони розподілу випадкових даних: біноміальний, рівномірний, експонентний, нормальний розподіл, розподіл Пуасона. 2. Підгонка розподілів. 3. Аналіз викидів бізнес-даних. 4. Заповнення пропущених значень у таблицях даних. 5. Відновлюваність результатів спостережень з використанням генератора випадкових чисел. <p>Лекції 10. Індуктивна аналітика в R План лекції</p> <ol style="list-style-type: none"> 1. Основні кроки побудови та верифікації моделей. 2. Статистична перевірка гіпотез. 3. Класичні методи і критерії статистики. Тест Шапіро-Уїлка. Рангові критерії Уїлкінсона-Манна-Уїтні. 4. Тести Стьюдента для залежних і незалежних вибірок даних. 5. Непараметричні тести міжгрупових різниць. 6. Рандомізація, бустреп та оцінка статистичної потужності вибірки. 7. Оцінка кореляції. Тест хі-квадрат (Пірсона). 8. Тест Колмогорова-Смірнова. 9. Критерій Фішера. Точний тест Фішера. 10. Критерії Мак-Немара та Кохрана-Мантеля-Хензеля. <p>Список рекомендованих джерел: <i>Основний:</i> 1-3, 5. <i>Додатковий:</i> 10, 12, 19, 21, 27-29. <i>Інтернет-ресурси:</i> 30-32, 36-38.</p>	<p>2</p> <p>2</p>
	<p>Лабораторне заняття 15 Описова бізнес-аналітика даних у R Мета: отримання кількісних та якісних характеристик даних для проведення описової бізнес-аналітики. Завдання: Створити нову змінну з числовими та символічними елементами. Обчислити суму елементів, середнє значення, медіану. Визначити довжину змінної, стандартне відхилення, округлити стандартне відхилення до певної кількості знаків після коми. Визначити максимальне, мінімальне та середнє значення набору даних. Обчислити квантилі, квартилі, IQR та стандартну похибку на основі наведених даних. Розрахувати скошеність даних візуально та за допомогою метрик описової аналітики.</p> <p>Лабораторне заняття 16 Моделювання законів розподілу випадкових даних у R Мета: вивчення основних законів розподілу випадкових чисел, які найчастіше використовуються для моделювання даних. Завдання: Визначити закон розподілу на основі даних. Побудувати біноміальний закон розподілу із заданими параметрами. Згенерувати набір випадкових чисел, розподілених за рівномірним, нормальним, експоненціальним законом, законом Пуассона. Визначити імовірність потрапляння випадкової величини у визначений інтервал. Створити та дослідити графіки функції та щільності розподілу за вказаним законом.</p>	<p>2</p> <p>2</p>

	<p style="text-align: center;">Лабораторні заняття 17-18 Перевірка статистичних гіпотез у R</p> <p>Мета: засвоєння основних методів індуктивної аналітики.</p> <p>Завдання: Побудувати графік щільності розподілу випадкової величини і визначити візуально схожість із нормальним законом розподілу. Перевірити гіпотезу про нормальний закон розподілу з використанням критерію Пірсона (хі-квадрат). Перевірити гіпотезу про нормальний закон розподілу з використанням критерію Шапіро-Уїлка. Визначити є значущим відхилення від нормативу, застосувавши гіпотезу Стьюдента про рівність математичного сподівання нормально розподіленої випадкової величини заданому значенню. Визначити співставність параметрів двох вибірок за допомогою перевірки гіпотези про рівність математичних сподівань двох вибірок за критерієм Стьюдента. Перевірити для двох наборів даних гіпотезу про рівність дисперсій з використанням критерію Фішера. Провести кореляційний аналіз за вказаним полем для двох наборів даних. Зробити висновок про взаємозв'язок між ними на основі візуального аналізу.</p>	4
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Описова та індуктивна бізнес-аналітика», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера.</p> <p><i>Індивідуальна робота:</i> Розробити шаблон проєкту для визначення основних кількісних характеристик набору даних, введених з клавіатури або отриманих із зовнішніх джерел. Для розробки шаблону використати пакети Shiny та gWidgets2. Передбачити обчислення мінімального, максимального та середнього значення набору даних, визначення найбільш імовірного закону розподілу значень вибірки, перевірку основних гіпотез щодо математичних сподівань та дисперсій.</p> <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Додатковий:</i> 12, 29. <i>Інтернет-ресурси:</i> 30-32, 37, 38.</p>	20
<p>Знати: методи візуалізації даних, засоби роботи з графічними пакетами R</p> <p>Вміти: використовувати пакети : Base, Grid, Lattice, ggplot2 для проведення процедур розвідувальної бізнес-аналітики.</p>	<p style="text-align: center;">Тема 4. Візуалізація даних та результатів бізнес-аналізу</p> <p style="text-align: center;">Лекції 11. Розвідувальна аналітика даних у R План лекції</p> <ol style="list-style-type: none"> 1. Створення графіки в R. 2. Робота з діаграмами показників бізнес-процесу. 3. Прості методи візуалізації бізнес-даних: точкові діаграми, стовпчикові діаграми, кругові та секторні діаграми, спінограми, гістограми, діаграми розмахів, діаграми ядерної оцінки функції щільності, скрипкові діаграми. 4. Графічні параметри візуалізації даних: розташування діаграм на сторінці, системи координат, колір та прозорість, налаштування відображення графіка, легенда, заголовки графіка та осей, розмітка осей, підписи та анотації, фасети, графічні теми. 	2

	<p align="center">Лекція 12. Розширені ресурси візуалізації даних у R</p> <p align="center">План лекції</p> <ol style="list-style-type: none"> 1. Розширена графіка для просунутої бізнес-аналітики. Чотири графічні системи R: Base, Grid, Lattice, ggplot2. 2. Пакет lattice. Умовні змінні. Функції для зміни формату даних. 3. Пакет ggplot2. Інтерактивна графіка. Взаємодія з діаграмами: ідентифікація точок. 4. Створення інтерактивної графіки за допомогою пакета iplots. Пакет rggobi. 5. Діаграми розсіювання, бульбашкові діаграми, лінійні графіки, корелограми, мозаїчні діаграми, діаграми Клівленда. Категоризовані графіки. <p align="center">Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-3. <i>Додатковий:</i> 6, 9, 10, 12, 14, 17, 19, 21, 23, 26-29. <i>Інтернет-ресурси:</i> 30-32, 36-38.</p>	2
	<p align="center">Лабораторні заняття 19-20</p> <p align="center">Базові графічні процедури в R</p> <p>Мета: отримання практичних навиків роботи з візуалізаторами стандартного пакету <i>base</i>.</p> <p>Завдання: Відтворити в системі R графік, зображений на рисунку. До графіка додати заголовок та зазначені підписи за осями координат. Додати до графіка горизонтальну лінію, що відображує середнє значення параметру. Задати заголовок графіку. Використовуючи R команди для базової візуалізації, відтворіть графік <i>boxplot</i>, який доступний в базовій R інсталяції. Використовуючи команду <i>par</i>, розділити графік на два вікна. Створити вектор із заданих числових значень. Побудувати гістограму і стовпчикову діаграму та проаналізувати їх відмінність. За допомогою команди <i>rplot</i> згенерувати дві послідовності випадкових чисел, розподілених за нормальним законом і порівняти отримані гістограми. Для заданого датасету зобразити дані на 2D точковому графіку і провести аналіз отриманих результатів.</p>	4
	<p align="center">Лабораторні заняття №21-22</p> <p align="center">Розвинута графіка в R</p> <p>Мета: набуття практичних навичок роботи з розвинутими засобами розвідувальної аналітики в R.</p> <p>Завдання: інсталювати заданий пакет та завантажити датасет, що є його частиною. Побудувати прості стовпчикові вертикальну та горизонтальну діаграми частоти зміни параметру датасету з використанням функцій <i>plot</i> або <i>barplot</i>. Додати заголовки графіків та осей координат. Побудувати стовпчикові діаграми з вертикальним та горизонтальним розподілом на підгрупи. Для побудови діаграми з горизонтальним розбиттям на підгрупи використати додатковий аргумент функції <i>barplot</i>: <i>beside=TRUE</i>. Побудувати за допомогою команди <i>spine()</i> із пакету <i>vcd()</i> спінограми – особливий варіант стовпчикових діаграм, у яких висота кожного результуючого стовпця дорівнює 1, а висоти його складових відображають</p>	4

	<p>пропорції. Реалізувати в R приклад, який задано скриптом. Для заданого датасету побудувати вдосконалені діаграми розмахів (boxplot), порівнявши за ними різні параметри одного набору даних. При побудові використати додаткові параметри notch=TRUE, varwidth=TRUE, які дозволяють зобразити ширину фігур, пропорційну обсягу вибірок. Зафарбуйте фігури на графіку довільним кольором.</p> <p>Для датасету побудувати ще одну модифікацію діаграм – скрипкові діаграми за допомогою функції vioplot() із пакету з такою ж назвою vioplot.</p>	
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Візуалізація даних та результатів бізнес-аналізу», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера.</p> <p><i>Індивідуальна робота:</i> Розробити шаблон проєкту для виведення графічних результатів бізнес-аналізу. Передбачити керування параметрами візуалізацій (заголовки графіка, заголовки осей, тип лінії, точкових символів, колір, товщина тощо) за допомогою візуальних елементів керування з пакету Shiny та сімейства пакетів gWidgets2.</p> <p>Список рекомендованих джерел: <i>Додатковий:</i> 12, 29. <i>Інтернет-ресурси:</i> 30-32, 37-38.</p>	16
<p>Знати: принципи регресійного аналізу, методи створення прогнозів у середовищі R.</p> <p>Вміти: застосовувати лінійний, нелінійний та множинний кореляційно-регресійний аналіз для побудови якісних прогнозів в R.</p>	<p>Тема 5. Прогностична бізнес-аналітика</p> <p>Лекція 13. Лінійний регресійний аналіз у середовищі R.</p> <p>План лекції</p> <ol style="list-style-type: none"> 1. Регресійний бізнес-аналіз. 2. Моделі регресії при різних видах функції втрат. 3. МНК-регресія: проста лінійна, поліноміальна, множинна лінійна регресія. 4. Діагностика та відбір регресійних моделей. 5. Мультиколінеарність. <p>Список рекомендованих джерел: <i>Основний:</i> 1-3, 5. <i>Додатковий:</i> 8, 12, 14, 15, 19, 21, 26-29 <i>Інтернет-ресурси:</i> 30-32, 36-38.</p>	2
	<p>Лабораторне заняття 23</p> <p>Лінійний регресійний аналіз у середовищі R</p> <p>Мета: набуття практичних навичок моделювання на основі лійного регресійного аналізу в середовищі R.</p> <p>Завдання: Побудувати лінійну та поліноміальну регресійні моделі та вивести повну інформацію про ці моделі. Побудувати в одній системі координат множинну точкових даних для двох параметрів із заданого датасету; пряму, що відповідає лінійній моделі, розрахованій за допомогою функції lm(); криву, що відповідає поліноміальній моделі, розрахованій за допомогою функції lm(). Побудувати лінійну та квадратичну модель для наступних даних. Вивести інформацію про модель з використанням функцій регресійного аналізу: summary(), coefficients(), confint(), fitted(), residuals(), anova(), vcov(), AIC(), plot(), predict()</p>	2

Побудувати в одній системі координат множинну точкових даних із таблиці початкових даних; 2) пряму, що відповідає лінійній моделі, розрахованій за допомогою функції $lm()$; параболу, що відповідає квадратичній моделі, розрахованій за допомогою функції $lm()$.

Лабораторне заняття 24 Логістична регресія в R

2

Мета: аналіз регресійних моделей з категоріальними значеннями показника.

Завдання: Виміряти ймовірність двійкового відгуку як значення показника на основі математичного рівняння, що зв'язує категоріальні значення (True/False або 0/1) зі змінними предиктора. Використати функцію $glm(formula, data, family)$ для визначення параметрів логістичної регресії. Дослідити вбудований набір даних «mtcars», що описує різні моделі автомобілів з різними характеристиками двигуна. У наборі даних «mtcars» інтерпретувати режим передачі (автоматичний або ручний), що описується стовпцем *am*, двійковим значенням (0 або 1). Створити модель логістичної регресії між стовпцями «*am*» і трьома іншими стовпцями - *hp*, *wt* і *cyl*.

Здійснити графічне відображення результатів логістичної регресії.

Лабораторне заняття 25-26 Множинна лінійна регресія в R.

4

Мета: побудова та дослідження параметрів множинної лінійної регресії в R.

Завдання: Створити регресійну модель для декількох факторів з використанням функції $lm()$ в R. Використайте синтаксис функцій $lm()$ для випадку множинної регресії: $lm(y \sim x_1 + x_2 + x_3 \dots, data)$. Для датасету *car* порівняти різні моделі автомобілей з точки зору пробігу на галон (міль на галон, *mpg*), робочого об'єму циліндра (*disp*), кінських сил (*hp*), ваги автомобіля (*wt*) та інших параметрів. Встановити зв'язок між *mpg* в якості змінної відгуку, *disp*, *hp* та *wt* в якості змінних-предикторів. Створити модель відношень між змінними і отримати коефіцієнти множинної регресійної моделі. Застосувати отримане рівняння множинної регресії для прогнозування нових значень пробігу автомобіля.

Лабораторне заняття 27 Прогностична аналітика в R.

2

Мета: набуття базових навичок побудови прогнозів в системі R на основі лінійного та нелінійного регресійного аналізу.

Завдання: Обрати один із датасетів, представлених на веб-ресурсі <https://vincentarelbundock.github.io/Rdatasets/datasets.html>. Датасет повинен містити не менше 4 чисельних характеристик ($n_numeric \geq 4$). Обрати два довільні параметри датасету.

Для обраних параметрів побудувати: лінійну модель; одну з нелінійних моделей; побудувати в одній системі координат точки початкових даних, лінійну та нелінійну моделі. Всі три графіка зобразити різними кольорами. Зобразити взаємозв'язок між двома довільно обраними параметрами датасету з використанням функції $scatterplot()$. Дослідити всі можливі двовимірні кореляції

між чисельними параметрами датасету з використанням функції `cor()`. Зробити інтерпретацію отриманих результатів кореляційного аналізу. Зобразити діаграми розсіювання для всіх чисельних параметрів за допомогою функції `scatterplotMatrix()`. Побудувати множинну регресійну модель за допомогою функції `lm()`, що зв'язує одну залежну з трьома незалежними чисельними змінними. Записати отримане рівняння. Отримати статистичні характеристики моделі з використанням функції `summary()`. Дати інтерпретацію отриманим результатам. Використати функцію `confint()` для визначення довірчих інтервалів моделі множинної регресії. Провести графічний аналіз моделей лінійної, нелінійної та множинної регресії з використанням діаграм, побудованих за допомогою функції `par(mfrow=c(2,2))`. Дати інтерпретацію результатам аналізу щодо адекватності моделей.

Лабораторне заняття 28

Модульний контроль 2 «Описова, індуктивна, розвідувальна та прогностична аналітика даних у середовищі R».

Мета: перевірити рівень знань, отриманих студентами при вивчення тем «Описова та індуктивна бізнес-аналітика», «Візуалізація даних та результатів бізнес-аналізу», «Прогностична бізнес-аналітика».

Завдання: У файлі *COVID-19 (Ukraine).xlsx* міститься інформація про перебіг коронавірусної хвороби в Україні. На основі початкових даних створити три датасети:

- загальний датасет *COVID19* з усіма даними таблиці;
- датасет *COVID19_April* з даними за квітень 2020 року;
- датасет *COVID19_May* з даними за травень 2020 року.

Побудувати стовпчикові діаграми даних полів *New cases*, *New deaths*, *New recovered* для кожного створеного датасету.

Перевірити гіпотезу про рівність дисперсій кількості щоденних одужань від хвороби, які спостерігалися у квітні та травні (рівень значущості 0,01).

На основі датасету *COVID19*:

Визначити максимальні значення кількості нових випадків зараження, кількості нових смертей та нової кількості тих, що одужали від хвороби, а також дати, в яких досягалися ці максимальні значення.

Перевірити гіпотезу про нормальний розподіл кількості нових випадків зараження за критерієм Шапіро-Уїлка з рівнем значущості 0.05. Зробити висновок.

Перевірити гіпотезу про нормальний розподіл кількості нових смертей від хвороби за критерієм Пірсона з рівнем значущості 0.05. Зробити висновок.

Побудувати в одній системі координат точковий графік кількості нових випадків інфікування та лінію, що відповідає нормальному розподілу з тими ж параметрами математичного сподівання та середньоквадратичного відхилення.

Дослідити асиметрію розподілу нових випадків зараження. Написати короткий висновок щодо результатів дослідження.

З рівнем значущості 0.05 перевірити гіпотезу про те, що

2

	<p>математичне сподівання кількості нових випадків зараження становить 3000 осіб на добу.</p> <p>Побудувати модель лінійної регресії залежності кількості нових смертей від кількості нових уражених. Записати рівняння лінійної регресії та побудувати в одній системі координат точковий графік початкових даних та знайдену лінію регресії.</p> <p>Побудувати матрицю діаграм розсіювання значень залежної змінної Total deaths від інших змінних датасету. Визначити змінні, які на Вашу думку, мають найбільш тісний кореляційний зв'язок.</p> <p>Провести діагностику лінійної регресійної моделі. Побудувати систему діагностичних діаграм залежності кількості нових смертей від кількості нових уражених. Зробити висновок про дотримання умов МНК-регресії.</p>	
	<p>Самостійна робота: опрацювання матеріалу лекцій, підготовка до лабораторних занять з теми «Прогностична бізнес-аналітика», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера, підготовка до модульного контролю.</p> <p><i>Індивідуальна робота:</i></p> <p>На основі пакету Shiny та сімейства пакетів gWidgets2 розробити шаблон проекту для створення апроксимації лінійними та нелінійними моделями і отримання прогностичних значень показника за заданими значеннями факторів з набору даних. Передбачити виведення коефіцієнтів прогностичної моделі та створення графічної демонстрації процесу апроксимації у вигляді початкової множини точок із заданого набору даних та знайденої лінії регресії.</p> <p>Список рекомендованих джерел:</p> <p><i>Додатковий:</i> 12, 29.</p> <p><i>Інтернет-джерела:</i> 30-32, 37-38.</p>	10
<p>Знати: засоби розв'язування основних задач інтелектуально-го аналізу даних у середовищі R.</p>	<p>Тема 6. Розширена бізнес-аналітика</p> <p>Лекція 14. Основні задачі Data Mining в R</p> <p>План лекції</p> <ol style="list-style-type: none"> 1. Задачі Data Mining в R. Бінарні матриці та асоціативні правила. Пошук логічних закономірностей у даних. 2. Кластерний аналіз. Бібліотека NbClust. 3. Способи сканування бізнес-даних. CSS-селектори. 4. Web scraping з пакетом rvest. PhantomJS і збереження вмісту динамічних веб-сторінок. 5. Збір інформації за допомогою API. <p>Список рекомендованих джерел:</p> <p><i>Основний:</i> 1-3.</p> <p><i>Додатковий:</i> 6, 26, 25-8, 11, 13, 15, 16, 18-22.</p> <p><i>Інтернет-ресурси:</i> 25, 26.</p>	2
	<p>Самостійна робота: опрацювання матеріалу лекції, підготовка до екзамену за питаннями з теми «Розширена бізнес-аналітика», ознайомлення з основною та додатковою літературою, інтернет-ресурсами, виконання індивідуальних завдань з використанням комп'ютера.</p> <p><i>Індивідуальна робота:</i></p>	20

	<p>Підготувати презентацію доповіді обсягом не менше 10 слайдів і тривалістю не більше 10 хвилин за однією із тем розширеної аналітики даних:</p> <ul style="list-style-type: none"> • Розширена аналітична теорія та методи: кластеризація • Розширена аналітична теорія та методи: асоціативні правила • Розширена аналітична теорія та методи: регресія • Розширена аналітична теорія та методи: класифікація • Розширена аналітична теорія та методи: аналіз часових рядів • Розширена аналітична теорія та методи: аналіз тексту • Розширені аналітичні технології та інструменти: MapReduce та Hadoop • Розширені аналітичні технології та інструменти: аналітика баз даних <p style="text-align: center;">Список рекомендованих джерел:</p> <p><i>Основний:</i> 5. <i>Додатковий:</i> 7, 10, 16 ,18, 25.</p>	
Разом		180

5. СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

Основний

1. Horton N., Kleinman K. Using R for Data Management, Statistical Analysis and Graphics. - CRC Press, 2011. - 296 p.
2. Kabacoff R. R in Action. Data analysis and graphics with R. – Manning: Shelter island, 2015. – 608 p.
3. Maindonald & J. Brown J. Data analysis and graphics using R. - Cambridge University Press, 2010. - 565 p.
4. Roskladka A., Ivanova O., Kulazhenko V. *Data Scientist: a glance into the future // Зовнішня торгівля: економіка, фінанси, право*. 2019. № 3. С. 109-120
5. Zumel N., Mount J. Practical Data Science with R. - Manning Publications, 2014. - 417 p.
* Курсивом виділені джерела, що є у бібліотеці КНТЕУ

Додатковий

6. BBC Visual and Data Journalism cookbook for R graphics [Електронний ресурс]. – Режим доступу: <https://bbc.github.io/rcookbook/>
7. Burger S. Introduction to Machine Learning with R: Rigorous Mathematical Analysis. - Sebastopol: O'Reilly Media, 2018. - 226 p.
8. Casella G., Flienberg S., Olkin I. An Introduction to statistical learning with application in R. - Springer, 2017. - 440 p.
9. Cbang W. R Graphics Cookbook - O'Reilly Media, 2013. - 413 p.
10. Dietrich D., HelleR B., Yang B. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. – John Wiley & Sons, Inc., 2015. – 420 p.
11. Etaati L. Advanced Analytics with Power BI and R. - Auckland: RADACAD Systems Limited, 2017. - 179 p.
12. Gago M. Creation of Graphical User Interfaces for R. - Universitat Autònoma de Barcelona, 2013. - 89 p.
13. Golemund G. HandsOn Programming with R. - O'Reilly Media, 2014. - 225 p.
14. Hafner S. An Introduction to R for Beginners. - Hafner Consulting LLC, 2019. - 360 p.
15. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning with Applications in R. - Springer, 2013. - 441 p.
16. Ledolter J. Data Mining and Business Analytics with R. - New Jersey: Wiley, 2013. - 368 p.
17. Matloff N. Probability and Statistics for Data Science: Math + R + Data. – London: Chapman & Hall, 2019. – 376 p.
18. Nolan D., Lang D.T. Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving. London: Chapman and Hall, 2015. - 539 p.
19. Shmueli G., Bruce P., Yahav I., Patel N., Lichtendahl K. Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. – New Jersey: Wiley, 2018. – 544 p.
20. Torgo L. Data Mining with R: Learning with Case Studies. – London: Chapman and Hall, 2017. – 446 p.
21. Viswanathan V. Data Analytics with R: A hands-on approach. - New Jersey: Infivista Inc, 2015. - 420 p.
22. Walkowiak S. Big Data Analytics with R: Leverage R Programming to uncover hidden patterns in your Big Data. - Birmingham: Packt Publishing, 2016. - 506 p.
23. Wickham H. ggplot2: Elegant graphics for data analysis. - Springer, 2017. - 217 p.
24. Wickham H., Golemund G. R for Data Science. – Sebastopol: O'Reilly Media, 2017. – 522 p.
25. Zhao Y., Cen Y. Data Mining Applications with R. - Elsevier, 2014. - 470 p.
26. Гнатюк В. Вступ до R на прикладах. - Харків: ХНЕУ, 2010. - 107 с.
27. Майборода Р. Є., Сугакова О. В. Аналіз даних за допомогою пакета R: навчальний посібник. – К.: ВПЦ «Київський університет», 2015. – 65 с.
28. Негрей М., Гнот Т. Аналітика з R. – Київ: Компрінт, 2020. – 236 с.

29. Храмов Д. Пакет Shiny [Электронный ресурс]. – Режим доступа:
http://dkhramov.dp.ua/Comp.ShinyLesson01.html#X_c5IdgzaM8

Internet-ресурсы

30. Displaying a graphic with gWidgets2 // URL: <https://stackoverflow.com/questions/20478142/displaying-a-graphic-with-gwidgets2>
31. How to start Shiny // URL: <https://vimeo.com/rstudioinc/review/131218530/212d8a5a7a/>
32. Package ‘gWidgets2’ // URL: <https://cran.r-project.org/web/packages/gWidgets2/gWidgets2.pdf>
33. RStudio // URL: <https://www.rstudio.com/products/rstudio/download/#download>
34. RStudio-education // URL: <https://github.com/rstudio-education>
35. The Comprehensive R Archive Network // URL: <https://cran.r-project.org/>
36. The R Project for Statistical Computing // URL: <https://www.r-project.org/>
37. Verzani J. Examples for gWidgets // URL: <https://edyaaleh.files.wordpress.com/2015/03/gwidgets.pdf>
38. Verzani J. The gWidgets2 package // URL: <http://search.r-project.org/library/gWidgets2/doc/gWidgets2.pdf>