

**Чернівецький національний університет імені Юрія Федьковича**

(повне найменування закладу вищої освіти)

**Факультет математики та інформатики**

(назва факультету/навчально-наукового інституту)

**Кафедра математичного моделювання**

(назва кафедри, що забезпечує викладання)

**“ЗАТВЕРДЖУЮ”**

**Декан факультету  
математики та інформатики**

**Ольга МАРТИНЮК**



2025 року

**РОБОЧА ПРОГРАМА  
навчальної дисципліни**

**Аналіз і візуалізація даних в R**

(назва навчальної дисципліни)

**вибіркова**

(вказати: обов'язкова)

**Освітньо-професійна програма «Системний аналіз»**

(назва програми)

**Спеціальність 124 Системний аналіз**

(вказати: код, назва)

**Галузь знань 12 Інформаційні технології**

(вказати: шифр, назва)

**Рівень вищої освіти перший (бакалаврський)**

(вказати: перший (бакалаврський) / другий (магістерський) / третій (освітньо-науковий))

**Факультет математики та інформатики**

(назва факультету/ навчально-наукового інституту, на якому здійснюється підготовка фахівців за вказаною освітньою програмою)

**Мова навчання українська**

(вказати: на якій мові читається дисципліна)

**Чернівці 2025 рік**

Робоча програма навчальної дисципліни «Аналіз і візуалізація даних в R» складена відповідно до освітньо-професійної програми «Системний аналіз»

**Розробник:**

Дорошенко Ірина Вікторівна, доцент кафедри математичного моделювання, кандидат фіз.-мат. наук, доцент

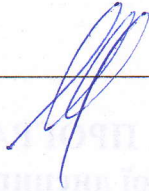
**Викладач, що забезпечує читання даної навчальної дисципліни:**

Дорошенко Ірина Вікторівна, доцент кафедри математичного моделювання, кандидат фіз.-мат. наук, доцент

**Затверджено** на засіданні кафедри математичного моделювання

Протокол № 15 від «24» червня 2025 року

Завідувач кафедри \_\_\_\_\_

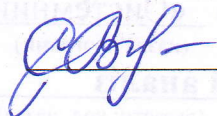


**Ігор ЧЕРЕВКО**

**Схвалено** методичною радою факультету математики та інформатики

Протокол № 12 від «25» червня 2025 року

Голова методичної ради \_\_\_\_\_



**Віра СІКОРА**

**Мета навчальної дисципліни:** вивчення основ роботи у статистичному програмному середовищі R та оволодіння методами розв'язання основних задач математичної статистики у цьому середовищі, візуалізація даних за допомогою відомих бібліотек мови R.

Призначення дисципліни – формування у здобувачів освіти уміння застосовувати сучасні комп'ютерні технології до постановки, аналізу та розв'язання основних задач статистичного аналізу даних із застосуванням середовища R.

**Результати навчання:** формування практичних умінь у сфері статистики; основних принципів роботи з R; основних графічних оболонок; основних структурних одиниць мови програмування R та операцій над ними; основних елементів програмування: циклів, розгалужень, функцій; основних методів графічного відображення даних та результатів їх обробки за допомогою спеціального пакету ggplot2; методів роботи засобами мови R з основними ймовірнісними розподілами; методів розв'язання засобами мови R задач описової статистики; методів перевірки засобами мови R основних статистичних гіпотез про вид розподілу вибірки; методів знаходження коефіцієнтів простої лінійної регресії; методів діагностики моделей регресії.

Дисципліна формує такі **компетенції** у відповідності до стандарту вищої освіти [1] та освітньої програми:

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК07. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК12. Здатність працювати в команді

ФК1. Здатність використовувати системний аналіз як сучасну міждисциплінарну методологію, що базується на прикладних математичних методах та сучасних інформаційних технологіях і орієнтована на вирішення задач аналізу і синтезу технічних, економічних, соціальних, екологічних та інших складних систем.

ФК4. 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.

ФК10. Здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них.

Наведені результати навчання за відповідною дисципліною співвідносяться із такими **програмними результатами навчання:**

ПР3. Вміти визначати ймовірнісні розподіли стохастичних показників та факторів, що впливають на характеристики досліджуваних процесів, досліджувати властивості та знаходити характеристики багатовимірних випадкових векторів та використовувати їх для розв'язання прикладних задач, формалізувати стохастичні показники та фактори у вигляді випадкових величин, векторів, процесів.

ПР12. Застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу.

## Опис навчальної дисципліни

### Загальна інформація

Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин						Вид підсумкового контролю
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота	індивідуальні завдання	
Денна	3	6	3	90	15	15	-	15	45	-	залік

### Структура змісту навчальної дисципліни

Назви змістових модулів і тем	Кількість годин												
	денна форма						Заочна форма						
	усього	у тому числі					усього	у тому числі					
		л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13	
<b>Змістовний модуль 1. Основи R та Rstudio</b>													
Тема 1. Основи роботи в системі R	6	2	-	2		2							
Тема 2. Складні структури даних в R	9	2	-	4		3							
Тема 3. Функціональне програмування в R	9	1		4		5							
Тема 4. Вступ до графічного аналізу в R	8	1		4		5							
<b>Разом за ЗМ 1</b>	<b>42</b>	<b>8</b>	<b>-</b>	<b>14</b>		<b>20</b>							
<b>Змістовний модуль 2. Розвідувальний аналіз та складання звітів</b>													
Тема 1. Описова статистика в R	10	1		2		5							
Тема 2. Перевірка статистичних гіпотез в R	11	2		4		5							
Тема 3. Елементи кореляційно аналізу в R	12	2		4		5							
Тема 4. Основи регресійного аналізу в R	16	2		6		10							
<b>Разом за ЗМ 2</b>	<b>48</b>	<b>7</b>		<b>16</b>		<b>25</b>							
<b>Усього годин</b>	<b>90</b>	<b>15</b>	<b>-</b>	<b>30</b>		<b>45</b>							

### Тематика лекційних занять з переліком питань

№	Назва теми
1	<i>Тема 1. Основи роботи в системі R.</i> Основні типи об'єктів в R: вектор, матриця, список, дата-фрейм. Огляд середовища розробки Rstudio. Базові арифметичні та логічні операції в R. Імпорт та експорт даних у R: робота з файлами .csv, .txt. Робота з консоллю та скриптами: основні команди та прийоми
2	<i>Тема 2. Складні структури даних в R</i> Створення та модифікація списків (lists). Особливості роботи з матрицями та масивами. Побудова дата-фреймів та їх редагування. Фактори (factors) та їхнє призначення для категоріальних даних. Індксація та фільтрація складних структур даних. Об'єднання, злиття та групування даних.
3	<i>Тема 3. Функціональне програмування в R</i> Створення власних функцій у R. Використання функцій <i>apply</i> , <i>lapply</i> , <i>sapply</i> , <i>tapply</i> .
4	<i>Тема 4. Вступ до графічного аналізу в R</i> Огляд базових графічних функцій R. Налаштування параметрів графіків. Побудова діаграм розсіювання та гістограм. Основи побудови графіків з <i>ggplot2</i>
5	<i>Тема 5. Описова статистика в R</i> Обчислення середнього, медіани, моди, стандартного відхилення. Показники асиметрії та ексцесу. Зведені таблиці та частоти. використання функцій <i>summary</i> , <i>describe</i> , <i>table</i> . Побудова діаграм для візуалізації розподілу даних. Інтерпретація отриманих статистичних показників.
6	<i>Тема 6. Перевірка статистичних гіпотез в R</i> Вибір тесту залежно від типу даних та досліджуваної задачі. Використання функцій <i>t.test</i> , <i>chisq.test</i> , <i>shapiro.test</i> .
7	<i>Тема 7. Елементи кореляційно аналізу в R</i> Візуалізація кореляційної матриці. Побудова діаграм розсіювання для оцінки зв'язків. Використання функцій <i>cor</i> , <i>cor.test</i> .
8	<i>Тема 8. Основи регресійного аналізу в R</i> Однофакторна лінійна регресія: побудова та інтерпретація. Багатофакторна лінійна регресія. Оцінка якості моделі: R-квадрат, р-значення, залишки. Побудова графіків для перевірки припущень регресії. Використання функції <i>lm</i> та аналіз результатів.

### Тематика лабораторних занять з переліком питань

№	Назва теми
1	Основи роботи в системі R (Ознайомлення з середовищем R та Rstudio; базові операції, робота зі змінними та типами даних; завантаження та збереження даних)
2	Складні структури даних в R (Вектори, матриці, списки, data frame; маніпуляції з таблицями даних; імпорт та експорт даних з різних форматів)
3	Функціональне програмування (Визначення та використання функцій; аргументи функцій, передача змінних; анонімні функції та застосування <i>apply</i> , <i>lapply</i> , <i>sapply</i> , <i>map</i> )
4	Графічні можливості R (Базові графіки в R; створення багатопанельних графіків; ознайомлення з пакетом <i>ggplot2</i> .)
5	Основи статистики в R (Описова статистика: середнє, медіана, дисперсія, стандартне відхилення; побудова таблиць частот і діаграм розподілу; використання базових статистичних функцій.)

6	Перевірка статистичних гіпотез (t-тести, $\chi^2$ -тест; перевірка нормальності розподілу; використання функцій <code>t.test()</code> , <code>chisq.test()</code> , <code>shapiro.test()</code> )
7	Основи кореляційного аналізу в R (Коефіцієнт кореляції Пірсона та Спірмена; матриця кореляцій; візуалізація кореляційних зв'язків.)
8	Основи регресійного аналізу в R (Побудова простої лінійної регресії; множинна регресія; аналіз залишків та якість моделі.)

### Індивідуальні науково-дослідні завдання (ІНДЗ)

№	Завдання до тем
1	Аналіз відкритого набору даних у середовищі R: імпорт, попередній огляд та базові розрахунки
2	Створення та обробка складних структур даних у R: від списків до дата-фреймів
3	Розробка користувацьких функцій та застосування функцій сімейства <code>apply</code> для аналізу даних
4	Візуалізація даних у R: побудова гістограм, діаграм розсіювання та графіків за допомогою <code>ggplot2</code>
5	Обчислення описових статистичних характеристик та інтерпретація результатів
6	Практичне застосування t-тестів та перевірка нормальності розподілу
7	Дослідження кореляційних зв'язків між числовими змінними
8	Нелінійна та логістична регресія

\* ІНДЗ – для змістового модуля, або в цілому для навчальної дисципліни визначається викладачем, з урахуванням специфіки дисципліни.

### Завдання для самостійної роботи студентів

Самостійна робота складається з повторення матеріалу, засвоєного на лекціях, самостійного опанування частини теоретичного матеріалу, роботи з контрольними запитаннями та завданнями.

Студенти можуть отримувати до 1 бала в рахунок виконання завдань СРС під час кожного лекційного заняття за правильні відповіді на запитання лектора, активне обговорення багатоваріантних підходів до рішення представленої лектором проблеми (для активізації пошукової та дослідної діяльності студентів).

№ з/п	Назва теми	Завдання для самостійної роботи	Кількість год
1	Основи роботи в системі R	Імпортування даних у R з файлу формату CSV. Базові операції: підрахунок кількості рядків і стовпців, знаходження середнє та медіану для одного з числових стовпців.	2
2	Складні структури даних в R	Моделювання багатовимірних даних, конструювання складних об'єктів та робота з категоріальними змінними.	3
3	Функціональне програмування	Автоматизація обчислень за допомогою функцій; створення анонімних та рекурсивних функцій; функції вищого порядку.	5
4	Графічні	Візуалізація даних як інструмент аналітика;	5

	можливості R	робота з різними типами графіків; оформлення графічних результатів.	
5	Основи статистики в R	Гістограма P-P і Q-Q діаграми Статистичні показники центру і розсіювання; підготовка даних до подальшого аналізу.	5
6	Перевірка статистичних гіпотез	Практичне використання методів перевірки статистичних гіпотез; інтерпретація результатів.	5
7	Основи кореляційного аналізу в R	Знаходження коефіцієнтів кореляції Фехнера, Спірмена та Кендала. Виявлення зв'язків між змінними; оцінка сили та напрямку зв'язку	5
8	Основи регресійного аналізу в R	Моделювання залежностей; оцінка впливу однієї змінної на іншу; перевірка адекватності моделі.	10

### **Методи та освітні технології навчання**

Лекції, лабораторні роботи, тестування, аудиторне та дистанційне онлайн-навчання з використанням систем Moodle та Google Meet, виконання завдань ІНДЗ.

Методи навчання:

- вербальні методи (лекція, бесіда, диспут, пояснення, розповідь тощо);
- практичні методи (лабораторні роботи);
- наочні методи (демонстрація, ілюстрація);
- робота з інформаційними ресурсами: з навчально-методичною, науковою, нормативною літературою та інтернет-ресурсами;
- самостійна робота над індивідуальним завданням або за програмою навчальної дисципліни;
- виконання завдань ІНДЗ
- дистанційне навчання з використанням відповідних онлайн-платформ

### **Система контролю та оцінювання**

Система оцінювання рівня навчальних досягнень ґрунтується на принципах ECTS та є накопичувальною. Протягом семестру студенти виконують 7 лабораторних робіт та 3 практичних робіт. Кожна лабораторна та практична роботи оцінюється кількістю балів за таблицею, наведеною нижче.

Виконуючи завдання лабораторної, студент повинен оформити і завантажити для подальшої перевірки на сайт електронного навчання звіт разом із працездатними файлами програмної реалізації завдань ЛР (правила оформлення наведені на сторінці навчальної дисципліни на сайті).

50% балів, відведених на оцінювання ЛР, студент отримує за працюючий програмний продукт, в якому реалізовано всі завдання ЛР та оформлений звіт. Решта 50% балів виставляється після захисту студентом виконаного звіту. На захисті звіту з ЛР студент має відповісти на питання щодо постановки задачі та розробленого ним алгоритму реалізації кожного із завдань ЛР. При відповіді на теоретичні питання та питання щодо програмної реалізації алгоритму у випадку неістотної помилки знімається 10-20% балів, а у випадку істотної 20-40% балів,

якщо ж студент не опанував теоретичний матеріал, плутається в означеннях, наводить логічно невірні твердження, то знімається до 50% балів від усієї суми балів за ЛР.

Максимальна кількість, яку можна набрати на підсумковому модулі (тестування) – 40 балів.

Підсумкова оцінка виставляється за результатами суми балів, набраних на змістових модулях під час семестру та підсумковому модулі згідно з нижче наведеною таблицею.

**Методи** формування професійної компетентності: розповідь, пояснення, бесіда, демонстрація, візуалізація, дискусія тощо. Методи формування практичних умінь та навичок: розв'язування задач практичних робіт, виконання завдань, розробка та аналіз алгоритмів і програмного коду, захист звітів з практичних робіт.

**Засобами** оцінювання та демонстрування результатів навчання є: стандартизовані тести; аналітичні звіти з лабораторних робіт; презентації результатів виконаних завдань та досліджень ІНДЗ, усний контроль у вигляді індивідуального та фронтального опитування на лекціях та лабораторних заняттях.

**Формами** поточного контролю є усна чи письмова (тестування, лабораторні роботи, ІНДЗ) відповідь студента.

Формою підсумкового контролю є залік.

### Критерії оцінювання поточного та підсумкового контролю

#### Розподіл балів, які отримують студенти

Поточне оцінювання ( <i>аудиторна та самостійна робота</i> )								Кількість балів (залік)	Сумарна к-ть балів
Змістовий модуль №1				Змістовий модуль №2					
T1	T2	T3	T4	T1	T2	T3	T4	40	100
5	10	5	5	5	10	10	10		

#### Шкала оцінювання: національна та ЄКТС

100-бальна шкала	Оцінка за національною шкалою		Оцінка за шкалою ЄКТС	
			Оцінка	Пояснення за розширеною шкалою
90-100	Зараховано	Відмінно	A	відмінно
80-89		Добре	B	дуже добре
70-79			C	добре
60-69		Задовільно	D	задовільно
50-59			E	достатньо
35-49	Незараховано	Незадовільно	FX	(незадовільно) з можливістю повторного складання

1-34			F	(незадовільно) з обов'язковим самостійним повторним опрацюванням освітнього компонента до перескладання
------	--	--	---	--

### Перелік питань для самоконтролю та підсумкового контролю навчальних досягнень студентів

1. Призначення та можливості середовища R для аналізу даних.
2. Основні типи даних та об'єктів у R: вектори, матриці, списки, дата-фрейми, фактори.
3. Імпорт і експорт даних у R, робота з робочою директорією.
4. Виконання базових арифметичних та логічних операцій у R.
5. Створення та редагування складних структур даних (списки, матриці, дата-фрейми).
6. Перетворення типів даних та робота з категоріальними змінними (факторами).
7. Фільтрація, сортування та об'єднання таблиць даних.
8. Написання власних функцій у R та використання аргументів за замовчуванням.
9. Анонімні функції, рекурсія та функції вищого порядку.
10. Використання функцій *apply*, *lapply*, *sapply* для обробки даних.
11. Побудова базових графіків у R: гістограми, діаграми розсіювання, *boxplot*.
12. Основи роботи з пакетом *ggplot2* та концепція Grammar of Graphics.
13. Налаштування вигляду графіків: кольори, легенди, підписи.
14. Збереження графіків у різних форматах.
15. Обчислення описових статистичних характеристик: середнє, медіана, мода, стандартне відхилення.
16. Побудова таблиць частот і зведених таблиць.
17. Інтерпретація статистичних показників та підготовка висновків.
18. Формулювання та перевірка нульової та альтернативної статистичних гіпотез.
19. Проведення t-тестів для однієї та двох вибірок.
20. Перевірка нормальності розподілу даних (тест Шапіро-Уїлка).
21. Інтерпретація результатів перевірки статистичних гіпотез та р-значення.
22. Розрахунок коефіцієнтів кореляції (Пірсона, Спірмена).
23. Побудова діаграм розсіювання та кореляційних матриць.
24. Інтерпретація сили та напрямку зв'язку між змінними.
25. Побудова моделі простої лінійної регресії у R.
26. Оцінка якості регресійної моделі ( $R^2$ , р-значення, залишки).
27. Інтерпретація коефіцієнтів регресії та побудова регресійної прямої.

## **Зарахування результатів неформальної/інформальної освіти**

Здобувачі вищої освіти має право на участь у неформальній/інформальній освіті.

У межах поточного контролю можуть визнаватися результати неформальної/інформальної освіти за умови наявності сертифікату або освітньої декларації про результати неформальної/інформальної освіти з питань, що відповідає тематиці курсу («Порядок визнання у Чернівецькому національному університеті імені Юрія Федьковича результатів навчання, здобутих шляхом неформальної та/або інформальної освіти»), <https://www.chnu.edu.ua/media/4g5fzssb/poriadok-vyznannia-rezultativ-navchannia-zdobutykh-shliakhom-neformalnoi-ta-abo-informalnoi-osvity.pdf> ).

Студентам можуть бути зараховані додаткові бали, отримані через неформальну освіту, до загальної суми балів, набраної з освітньої компоненти, за умови, що результати з проблеми, за якою відбувалося навчання, відповідають тематиці курсу.

### **Рекомендована література**

#### **Основна**

1. Стандарт вищої освіти України перший (бакалаврський) рівень, галузь знань 12 – Інформаційні технології, спеціальність 124 – Системний аналіз // Затверджено і введено в дію наказом Міністерства освіти і науки України від 13.11.2018 р. № 1245.– 23 с.  
URL: <https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/12/21/124-sistemn.analiz-bakalavr-1.pdf>
2. Гнатюк В. Вступ до R на прикладах: навчальний посібник.- Навчальний посібник. ХНЕУ, 2010, 107с.
3. Jenine K. Harris, Statistics With R. Washington University in St. Louis, USA, 2020.
4. Demidenko E. Advanced Statistics with Applications in R, , 2020, Wiley Series in Probability and Statistics, John Wiley and Sons USA.- 880 p.
5. Майборода Р.Є Комп'ютерна статистика. Професійний старт. Навчальний посібник. Київський університет», 2018. – 482 с.  
<http://probability.univ.kiev.ua/userfiles/mre/compsta1.pdf>
6. Королюк В.С., Царков Є.Ф., Ясинський В.К. Ймовірність, статистика та випадкові процеси. Теорія та комп'ютерна практика. В 3-х томах. Т.2: Статистика. Комп'ютерне статистичне моделювання. – Чернівці: Видавництво "Золоті литаври", 2008. – 580 с.
7. Турчин В.М. Теорія ймовірностей і математична статистика,- Дніпропетровськ, ІМА-пресс, 2014. - 566 с.
8. Дорошенко І.В., Антонюк С.В., Бодрик Н.П. Статистика: програма, методичні вказівки і контрольні завдання. - Чернівці: Видавничий дім «Родовід» , 2013 – 76с.
9. Keon-Woong Moon. Learn ggplot2 Using Shiny App, - Springer, 2017.-P. 356.
10. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning, - Springer: New , 2017.

11. Ugarde M.D., Militino A.F., Arnholt A.T. Probability and statistics with R. – Boca Raton, London, New York: CRC Press, Taylor&Francis Group, 2008. 700p.
12. Venables W.N., Ripley B.D. Modern applied statistics with S. – Springer, 2002. – 495 p.
13. Sarah Stowell. Using R for Statistics. Apress, 2014.

### **Додаткова**

1. Thomas Rahlf. Data Visualisation with R. Springer International Publishing, New York, 2017.
2. Дорошенко І.В., Антонюк С.В., Бодрик Н.П. Статистика: програма, методичні вказівки і контрольні завдання. - Чернівці: Видавничий дім «Родовід», 2013 – 76с.

### **Інформаційні ресурси**

1. <http://cran.r-project.org/bin/windows/base/>
2. Дистанційний курс «Аналіз даних мовою R». Круглова Н.В., Диховичний О.О. <https://do.ipokpi.ua/login/index.php>

### **Політика академічної доброчесності**

Дотримання політики щодо академічної доброчесності учасниками освітнього процесу при вивченні навчальної дисципліни регламентовано такими документами:

1. «Етичний кодекс Чернівецького національного університету імені Юрія Федьковича» <https://www.chnu.edu.ua/universytet/normatyvni-dokumenty/etychnyi-kodeks-chernivetskoho-natsionalnoho-universytetu-imeni-yuriiia-fedkovycha/>
2. «Положенням про виявлення та запобігання академічного плагіату у Чернівецькому національному університету імені Юрія Федьковича» <https://www.chnu.edu.ua/universytet/normatyvni-dokumenty/polozhennia-pro-vyivlennia-ta-zapobihannia-akademichnomu-plahiatu/>