

Чернівецький національний університет імені Юрія Федьковича

(повне найменування закладу вищої освіти)

Факультет математики та інформатики

(назва інституту/факультету)

Кафедра математичного моделювання

(назва кафедри)



“ЗАТВЕРДЖУЮ”

Декан

Мартинюк О.В.

2024 року

12 ” 08

РОБОЧА ПРОГРАМА

навчальної дисципліни

Аналіз і візуалізація даних в R

(вказіть назву навчальної дисципліни (іноземною, якщо дисципліна викладається іноземною мовою))

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/інституту, на якому здійснюється підготовка фахівців за вказаною освітньо-професійною програмою)

Мова навчання українська

(вказати: на яких мовах читається дисципліна)

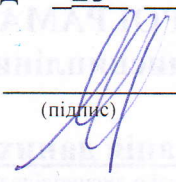
Чернівці 2024 рік

Робоча програма навчальної дисципліни «**Аналіз і візуалізація даних в R**» складена відповідно до освітньо-професійної програми «Системний аналіз», затвердженій Вченою радою Чернівецького національного університету імені Юрія Федьковича (протокол № 7 від 29 квітня 2024 р.)
(назва освітньо-професійної програми, дата останнього затвердження)

Розробники: Дорошенко І.В., доцент, кандидат фізико-математичних наук
(П.І.Б. авторів, посада, науковий ступінь, вчене звання)

Погоджено з гарантом ОП і затверджено на засіданні кафедри математичного моделювання

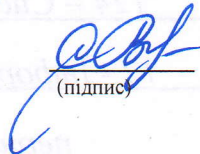
Протокол № 18 від “25” червня 2024 року

Завідувач кафедри  Черевко І.М.
(підпис) (прізвище та ініціали)

Схвалено методичною радою факультету математики та інформатики

Протокол № 11 від “25” червня 2024 року

Голова методичної ради факультету
математики та інформатики

 Сікора В.С.
(підпис) (прізвище та ініціали)

1. 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.

2. 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.

Розвинути у студентів наступні компетентності:

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК4. Здатність спілкуватися державною мовою як усно, так і письмово

ЗК6. Здатність вчитися й оволодівати сучасними знаннями.

ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК8. Здатність генерувати нові ідеї (креативність).

ЗК9. Здатність працювати в команді.

ЗК11. Здатність приймати обґрунтовані рішення.

ФК1. Здатність до математичного формулювання та досліджування неперервних та дискретних математичних моделей, обґрунтовування вибору методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук, аналізу та інтерпретування.

ФК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

Програмні результати навчання:

ПРН1. Застосовувати знання основних форм і законів абстрактнологічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

ПРН3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.

ПРН8. Використовувати методологію системного аналізу об'єктів, процесів і систем для задач аналізу, прогнозування, управління та проектування динамічних процесів в макроекономічних, технічних, технологічних і фінансових об'єктах.

3. Опис навчальної дисципліни

3.1. Загальна інформація

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

3.2. 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
Змістовний модуль 1. Основи R та Rstudio												
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						
Разом за ЗМ 1	42	8	-	14		20						
Змістовий модуль 2. Розвідувальний аналіз та складання звітів												
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						
Разом за ЗМ 2	48	7		16		25						
Усього годин	90	15	-	30		45						

3.3 Тематика лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Основи роботи в системі R	2
2	Складні структури даних в R	4
3	Функціональне програмування	4
4	Графічні можливості R	4
5	Основи статистики в R	2
6	Перевірка статистичних гіпотез	4
7	Основи кореляційного аналізу в R	4
8	Основи регресійного аналізу в R	4

3.4. Самостійна робота студента

№ з/п	Назва теми (форма контролю)	Кількість балів
1	Основи роботи з R та RStudio. Встановлення системи, RStudio. Робота зі скриптами. Встановлення пакетів. Використання R в якості калькулятора. Основні арифметичні операції над векторами.	5
2	Найпростіші об'єкти R. Матриці. Масиви. Фактори.	5
3	Складні структурні одиниці мови R. Списки. Фрейми. Функції групи apply.	5
4	Візуалізація даних. Функції qplot, ggplot. Різноманітні geoms.	5
5	Описова статистика. Гістограма. P-P і Q-Q діаграми. Обчислення числових характеристик вибірок.	10
6	Перевірка гіпотез про вид розподілу. Критерії перевірки	10

	гіпотези про розподіл вибірки: Колмогорова-Смірнова, критерій .	
7	Елементи кореляційно-регресійного аналізу	10
8	Знаходження коефіцієнтів кореляції Фехнера, Спірмена та Кендала	5

4. Освітні технології, методи навчання і викладання навчальної дисципліни

У викладання курсу використовуються такі освітні технології:

- Лекції та їх презентації.
- Онлайн-лекції.
- Практичні та лабораторні заняття.
- Групова робота, коли студенти розв'язують практичні завдання.
- Онлайн-тести та опитування: Використання системи MOODLE

Електронні підручники і ресурси репозитарію ЧНУ

Методи навчання:

МН 1 - лекція-візуалізація;

МН 8 – робота з тестами;

МН 9 – робота в групах;

МН 12 – дистанційне навчання з використанням відповідних онлайн-платформ

5. Критерії оцінювання результатів навчання з навчальної дисципліни

5.1. Критерієм підсумкового 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 та практичної роботи, студент повинен оформити і завантажити для подальшої перевірки на сайт електронного навчання звіт разом із працездатними файлами програмної реалізації завдань ЛР та ПР (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.

Розподіл балів, які отримують студенти

Поточне оцінювання (аудиторна та самостійна робота)									Кількість балів (залік)	Сумарна к-ть балів
Змістовий модуль №1					Змістовий модуль № 2					
T1	T2	T3	T4	T5	T1	T2	T3	T4	40	100
5	10	5	5	5	5	5	10	10		

5.2. Шкала оцінювання: національна та ЄКТС

Оцінка за національною шкалою	Оцінка за шкалою ECTS	
	Оцінка (бали)	Пояснення за розширеною шкалою
Зараховано	A (90-100)	відмінно
Зараховано	B (80-89)	дуже добре
	C (70-79)	добре
Зараховано	D (60-69)	задовільно
	E (50-59)	достатньо
Незараховано	FX (35-49)	(незадовільно) з можливістю повторного складання
	F (1-34)	(незадовільно) з обов'язковим повторним курсом

5.3. Засоби оцінювання

Засобами оцінювання та демонстрування результатів навчання можуть бути:

- перевірка викладачем та захист студентами письмових звітів про виконання кожної практичної та лабораторної роботи,
- експрес-опитування,
- тестові завдання.

6. Форми поточного та підсумкового контролю

1. Поточний (захист практичних та лабораторних робіт)
2. Модульний (тестування, виконання завдань)
3. Підсумковий (залік)

7. Рекомендована література

7.1. Фахова (основна)

1. Гнатюк В. Вступ до R на прикладах: навчальний посібник.- Навчальний посібник. ХНЕУ, 2010, 107с.
2. Jenine K. Harris, Statistics With R .Washington University in St.Louis, USA, 2020.
3. Demidenko E. Advanced Statistics with Applications in R, , 2020, Wiley Series in Probability and Statistics, John Wiley and Sons USA.- 880 p.
4. Майборода Р.Є Комп'ютерна статистика. Професійний старт. Навчальний посібник. Київський університет», 2018. – 482 с.
<http://probability.univ.kiev.ua/userfiles/mre/compsta1.pdf>
5. Короліук 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.
6. Турчин В.М. Теорія ймовірностей і математична статистика,- Дніпропетровськ, ІМА-пресс, 2014. - 566 с.
7. Keon-Woong Moon. Learn ggplot2 Using Shiny App, - Springer, 2017.-P. 356.
8. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning, - Springer: New , 2017.
9. Ugarte M.D., Militino A.F., Arnholt A.T. Probability and statistics with R. – Boca Raton, London, New York: CRC Press, Taylor&Francis Group, 2008. – 700 p.
10. Venables W.N., Ripley B.D. Modern applied statistics with S. – Springer, 2002. – 495 p.
11. Sarah Stowell. Using R for Statistics. Apress, 2014.

8.1. Додаткова

1. Thomas Rahrhf. Data Visualisation with R. Springer International Publishing, New York, 2017.

Методичне забезпечення

1. Дорошенко І.В., Антонюк С.В., Бодрик Н.П. Статистика: програма, методичні вказівки і контрольні завдання. - Чернівці: Видавничий дім «Родовід» , 2013 – 76с.

8. 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.ipk.kpi.ua/login/index.php>