



Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»



Кафедра
математичного
аналізу та теорії
ймовірностей

СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ ЗАСОБАМИ МОВИ R

Робоча програма кредитного модуля навчальної
дисципліни (Силабус)

1. Реквізити навчальної дисципліни

Рівень вищої освіти	Другий(магістерський)
Галузь знань	11 Математика та статистика
Спеціальність	111 Математика
Освітня програма	Страхова та фінансова математика
Статус дисципліни	Вибіркова
Форма навчання	очна
Рік підготовки, семестр	1 курс, осінній семестр
Обсяг дисципліни	120 годин (36 години – Лекції, 18 години – Практичні, 66 години – СРС)
Семестровий контроль/ контрольні заходи	Залік/МКР, ДКР
Розклад занять	http://rozklad.kpi.ua
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: канд. ф.-м.н., доцент кафедри математичного аналізу та теорії ймовірностей, Круглова Наталія Володимирівна, natahak@ukr.net , моб. +38(095)2207958 Практичні: канд. ф.-м.н., доцент кафедри математичного аналізу та теорії ймовірностей, Круглова Наталія Володимирівна,

	natahak@ukr.net , моб. +38(095)2207958
Розміщення курсу	https://campus.kpi.ua ,

2. Програма навчальної дисципліни

3. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Цілі дисципліни	Ціллю навчальної дисципліни є: формування у здобувачів освіти умінь застосовувати сучасні комп'ютерні технології до постановки, аналізу та розв'язання основних задач статистичного аналізу даних
Предмет навчальної дисципліни	Предмет навчальної дисципліни «Аналіз даних засобами мови R» - вивчення основ роботи у статистичному програмному середовищі R та оволодіння методами розв'язання основних статистичних задач у цьому середовищі, візуалізація даних за допомогою відомих бібліотек мови R.
Компетентності	Метою навчальної дисципліни є формування у студентів здатностей: Здатність учитися, здобувати нові знання, умінь, у тому числі в галузях, відмінних від математики (ЗК1); Здатність до пошуку, оброблення й аналізу інформації, необхідної для розв'язування наукових і професійних завдань (ЗК4); Знання на рівні новітніх досягнень, необхідні для дослідницької та/або інноваційної діяльності у сфері математики та її практичних застосувань (ФК1); Здатність до розвитку нових та удосконалення існуючих математичних методів аналізу, моделювання, прогнозування, розв'язування нових проблем у нових галузях знань (ФК8); Здатність розв'язувати прикладні задачі аналізу даних математичними методами та методами комп'ютерної статистики і обирати для цього адекватні математичні засоби (ФК10); Здатність сформулювати у слухачів уявлення про класичні та сучасні математичні теорії, взаємозв'язок та різницю між ними і застосування їх у природничих, економічних та технічних науках, у фінансовій та страховій сферах (ФК13); Здатність використовувати методи математичної статистики у теоретичних і прикладних дослідженнях; Спроможність доводити розв'язок задачі до зрозумілого результату – адекватного математичного та алгоритмічного апаратів; Здатність самостійно вивчати та використовувати сучасну математичну статистичну літературу та літературу з програмування, здатності до розвитку гнучкості мислення, творчої самостійності та дій.

Програмні результати навчання

Головних результатів та сфер застосувань основних математичних теорій, що використовуються при математичному моделюванні: фінансової математики, методів математичної економіки та імітаційного моделювання, комп'ютерної статистики (PH4);
Математичних дисциплін, в яких вивчаються моделі природничих процесів; математичних методів аналізу та прогнозування; математичних способів інтерпретації числових даних; принципів функціонування природничих процесів (PH5);
Математичних моделей у природничих та суспільних науках, а також обмежень для їх застосування; математичних методів оцінки параметрів моделей та прогнозування на підставі моделей; знання в галузі інформатики й сучасних інформаційних технологій (PH6);
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.

УМІННЯ:

Роботи з базовою версією мови програмування R та інтегрованим середовищем розробки для R - RStudio;
Використання та обробки основних типів даних мови програмування R: векторів, матриць, масивів, факторів, списків та фреймів;
Експортувати/імпортувати дані; складати власні функції;
Графічно аналізувати за допомогою функцій пакету ggplot2 різні типи вхідних даних;
Моделювати випадкові величини, розраховувати щільності та функції розподілу.

4. 2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Пререквізити: Дисципліна «СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ ЗАСОБАМИ МОВИ R» викладається в другому семестрі першого курсу підготовки магістрів і базується на знаннях, отриманих при вивченні дисциплін «Теорія ймовірностей», «Основи математичної статистики», «Основи теорії випадкових процесів», «Лінійний регресійний аналіз», які вивчаються на бакалаврському рівні вищої освіти за освітньою програмою «Страхова та фінансова математика», а також «Прикладні моделі нелінійного регресійного аналізу» (ПО3), «Аналіз часових рядів» (ПО4), які вивчаються на магістерському рівні вищої освіти за освітньою програмою «Страхова та фінансова математика».

Постреквізити: Дисципліна «СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ ЗАСОБАМИ МОВИ R» передує освітнім компонентам «Обробка даних методами дисперсійного аналізу» (ЗО7), «Практика» (ПО8) та «Наукова робота за темою магістерської дисертації» (ПО9)

3.Зміст навчальної дисципліни

Назва розділів і тем	Кількість годин			
	Всього	у тому числі		
		Лекції	Практичні	СРС
1	2	3	4	5
Розділ 1.				
Тема 1.1. Основи роботи з графічними оболонками середовища R	8	2	2	4
Тема 1.2. Основні структурні одиниці мови програмування R та операції над ними	20	8	4	8
Тема 1.3. Створення власних функцій	16	4	2	10
Тема 1.4. Візуалізація даних	16	4	2	10
Разом за розділом 1	60	18	10	32
Розділ 2.				
Тема 2.1. Описова статистика за допомогою спеціальних бібліотек	8	4	2	2
Тема 2.2. Перевірка гіпотез про вид розподілу	6	2	2	2
Тема 2.3. Лінійна регресія	8	4	2	2
Тема 2.4. Логістична регресія	8	4		4
МКР	12		2	10
ДКР	10			10
Разом за розділом 2	52	14	8	30
Залік	8	4		4
Всього годин	120	36	18	66

5. 4. Навчальні матеріали та ресурси

Базова література

1. Майборода Р.Є Комп'ютерна статистика. Професійний старт. Навчальний посібник. Київський університет», 2018. – 482 с.
<http://probability.univ.kiev.ua/userfiles/mre/compsta1.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. Sarah Stowell. Using R for Statistics. Apress, 2014.
5. Турчин В.М. Теорія ймовірностей і математична статистика,- Дніпропетровськ, ІМА-прес, 2014. - 566 с.
6. Keon-Woong Moon. Learn ggplot2 Using Shiny App, - Springer, 2017.-P. 356.
7. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning, - Springer: New , 2017.

Додаткова література

8. Thomas Rahlf. Data Visualisation with R. Springer International Publishing, New York, 2017.
9. Чернова Н. И. Математическая статистика.- Учеб. пособие / Новосиб. гос. ун-т. Новосибирск, 2007. 148 с.

Інформаційні ресурси

10. <http://cran.r-project.org/bin/windows/base/>
11. Дистанційний курс «Аналіз даних мовою R». Круглова Н.В., Диховичний О.О.
<https://do.ipk.kpi.ua/login/index.php>
12. Дистанційний курс «Обчислювальна ймовірність та статистика». Круглова Н.В., Диховичний О.О.
<https://do.ipk.kpi.ua/course/view.php?id=2357>

6. Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Очна/дистанційна форма

Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних матеріалів, посилання на літературу)
1.	<u>Основи роботи з графічними оболонками мови R.</u> RStudio. RCommander. Встановлення пакетів. Робота з Help. <i>Рекомендована література:</i> [2], глава 1; [3], §1
2.	<u>Основні структурні одиниці середовища програмування R та операції з ними.</u> Основні арифметичні операції. Вектори. <i>Рекомендована література:</i> [2], глава 2, [3], §2
3	<u>Основні структурні одиниці середовища програмування R та операції з ними.</u> Матриці, масиви, фактори. Індексція. <i>Рекомендована література:</i> [2], глава 2, [3], §2
4	<u>Об'єкти складної структури.</u> Фрейми. Списки. <i>Рекомендована література:</i> [7], глава 2
5	<u>Об'єкти складної структури.</u> Функції сімейства apply для роботи з фреймами та списками. <i>Рекомендована література:</i> [4], глава 3, 4

6.	<u>Базові прийоми програмування.</u> Імпорт/експорт даних. Цикли. <i>Рекомендована література:</i> [2], глава3,4, [3], §2
7.	<u>Базові прийоми програмування.</u> Створення власних функцій <i>Рекомендована література:</i> [2], глава3,4, [3], §2
8.	<u>Візуалізація даних.</u> Функції першого пріоритету. <i>Рекомендована література:</i> [2], глава 6, [3], §4
9	<u>Візуалізація даних.</u> Можливості пакету ggplot2. <i>Рекомендована література:</i> [6]
10.	<u>Описова статистика.</u> Моделювання випадкових величин. Побудова гістограм, P-P і Q-Q діаграм. Обчислення числових характеристик вибірок. <i>Рекомендована література:</i> [2], глава 5, [3], §3
11.	<u>Описова статистика за допомогою спеціальних пакетів.</u> Функції пакетів psych, dplyr. <i>Рекомендована література:</i> [3], [10]
12.	<u>Перевірка гіпотез про вид розподілу.</u> Критерії перевірки гіпотези про розподіл вибірки. <i>Рекомендована література:</i> [3], §8 , [1], глава 9.
13.	<u>Лінійна регресія.</u> Кореляція. Проста лінійна регресія. <i>Рекомендована література:</i> [7], глава 3.
14.	<u>Лінійна регресія.</u> Множинна лінійна регресія. Діагностика моделі. <i>Рекомендована література:</i> [7], глава 3
15.	<u>Логістична регресія.</u> Модель з одним предиктором. Модель з двома і більше предикторами. <i>Рекомендована література:</i> [7], глава 4
16.	<u>Логістична регресія.</u> Непараметричні методи. <i>Рекомендована література:</i> [7], глава 4
17.	Захист ДКР.
18.	Залік.

Практичні заняття

№ з/п	Теми практичних занять (перелік дидактичних матеріалів, посилання на літературу)
1.	<u>Основи роботи з R та RStudio.</u> Встановлення системи, RStudio. Робота зі скриптами. Встановлення пакетів. Використання R в якості калькулятора. Основні арифметичні операції над векторами. <i>Рекомендована література:</i> [2], глава 1; [3], §1
2.	<u>Найпростіші об'єкти R.</u> Матриці. Масиви. Фактори. <i>Рекомендована література:</i> [2], глава 2, [3], §2
3.	<u>Складні структурні одиниці мови R.</u> Списки. Фрейми. Функції групи apply. <i>Рекомендована література:</i> [2], глава 2, [3], §2
4.	<u>Базові прийоми програмування.</u> Цикли. Розгалуження. Створення власних функцій. Експорт/імпорт даних. <i>Рекомендована література:</i> [2], глава 3, [3], §2-4
5.	<u>Візуалізація даних.</u> Функції qplot, ggplot. Різноманітні geoms. <i>Рекомендована література:</i> [6]
6.	<u>Описова статистика.</u> Гістограма. P-P і Q-Q діаграми. Обчислення числових характеристик вибірок. <i>Рекомендована література:</i> [7], глава 2
7.	<u>Перевірка гіпотез про вид розподілу.</u> Критерії перевірки гіпотези про розподіл вибірки:

	Колмогорова-Смірнова, критерій χ^2 . Рекомендована література: [3], §8, [1], глава 9
8.	<u>Лінійна регресія</u> . Кореляція. Найпростіша модель лінійної регресії. Рекомендована література: [7], глава 3
9	<u>МКР</u>

6. Самостійна робота студента/аспіранта

Вивчення дисципліни включає наступні види самостійної роботи:

- підготовка до лекційних та практичних занять, виконання домашніх завдань;
- виконання домашньої контрольної роботи (тестові завдання в дистанційних курсах на платформі Moodle);
- підготовка та виконання модульної контрольної роботи;
- підготовка до заліку.

Контрольні роботи

Запланована одна модульна контрольна робота, мета якої – виявити рівень засвоєння відповідних модулів, підрахування балів за кредитно-модульною системою модулів.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Рекомендовані методи навчання: вивчення основної та допоміжної літератури за тематикою лекцій, розв'язування задач на практичних заняттях та при виконанні домашніх робіт.

Студенту рекомендується вести докладний конспект лекцій. Важливим аспектом якісного засвоєння матеріалу, відпрацювання методів та алгоритмів вирішення основних завдань дисципліни є самостійна робота. Вона містить читання літератури, огляд літератури за темою, підготовку до занять, виконання домашньої контрольної роботи, підготовку до МКР та іспиту.

Академічна доброчесність

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>

Норми етичної поведінки

Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>

1. Види контролю та рейтингова система оцінювання результатів навчання (РСО) (очна\дистанційна форма)

Розподіл навчального часу за видами занять і завдань з дисципліни згідно з робочим навчальним планом.

Форма навчання	Кредитні модулі	Всього	Розподіл навчального часу за видами занять	Семестрова атестація
----------------	-----------------	--------	--	----------------------

		кредитів	годин	Лекції	Практичні заняття	Д К Р	М К Р	СРС	
Денна	1	4	120	36	18	1	1	66	залік

Рейтинг студента з дисципліни складається з балів, що він отримує за

- 1) відповіді на практичних заняттях, виконання домашніх робіт;
- 2) МКР;
- 3) ДКР;
- 4) Захист ДКР, відповідь на заліку.

Система рейтингових (вагових) балів та критерії оцінювання

1. Робота на практичних заняттях

Виконання домашніх завдань дозволяє закріпити знання, набуті під час роботи на практичному занятті. За виконання кожної домашньої роботи студент отримує до 5 балів:

- якщо задачі повністю розв'язані, то студент отримує 5 балів;
- якщо відповідь правильна, але у розв'язку є неточності, то студент отримує 1-4 бали;
- якщо незадовільна відповідь, метод розв'язування задачі неправильний – 0 балів.

Ваговий бал -40 балів.

2. Модульний контроль

Модульна контрольна: ваговий бал – 30 балів.

Складається з 10 тестових задач. Ваговий бал кожної задачі – 3 бали. Розв'язок задачі оцінюється в 0-3 бали наступним чином:

- якщо задача повністю розв'язана, то студент отримує 3 бали;
- якщо відповідь правильна, але у розв'язку є неточності, то студент отримує 1-2 бали;
- якщо незадовільна відповідь, метод розв'язування задачі неправильний – 0 балів.

При виконання < 60% контрольна робота не зараховується.

3. Домашня контрольна робота

Ваговий бал –30.

Критерій оцінювання ДКР:

Невиконання ДКР – 0 балів

Оцінка ДКР (в балах) дорівнює величині відсотка (від максимальної кількості балів 30) її виконання з урахуванням захисту ДКР.

При виконанні менше 60% ДКР вона не зараховується і повинна бути доопрацьована.

За несвоєчасне (пізніше ніж на тиждень) подання роботи – зараховується не більше 60% ДКР.

Для отримання “зараховано” з першої проміжної атестації (8 тиждень) студент повинен мати не менше ніж 50% від запланованої кількості балів. Для отримання “зараховано” з другої проміжної атестації (14 тиждень) студент повинен мати не менше ніж 50% від запланованої кількості балів.

В залежності від рейтингу R можливі наступні дії:

1. якщо $R < 60$, то студент вважається **недопущеним** і перескладає залік у додаткову сесію;
2. якщо $60 \leq R$ і студент має хоча б одну позитивну атестацію, зараховану модульну контрольну роботу та ДКР (виконану не менше, ніж на 60%) студент допускається до заліку;

Таблиця переведення рейтингової оцінки з навчальної дисципліни R: (згідно з Табл. 1)

R	Оцінка ECTS	Традиційна оцінка
95...100	A	відмінно
85...94	B	дуже добре
75...84	C	добре
65...74	D	задовільно
60...64	E	достатньо
$R \leq 60$	FX	незадовільно
$R < 60$ або не виконані інші умови допуску до заліку	F	не допущений

2. Додаткова інформація з дисципліни (освітнього компоненту)

У випадку дистанційної форми навчання організація освітнього процесу здійснюється з застосуванням електронної пошти, Telegram, відео-конференцій в Zoom та освітньої платформи Moodle.

Робочу програму навчальної дисципліни (силабус):

Складено доцентом кафедри МА та ТЙ, канд. фіз.-мат. наук Кругловою Н.В.

Ухвалено кафедрою МА та ТЙ (протокол № 12 від 19.06.2023 р.)

Погоджено Методичною комісією ФМФ (протокол № 10 від 27.06.2023 р.)