



ПО 01. АНАЛІЗ ДАНИХ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Перший (бакалаврський)</i>
Галузь знань	<i>11 Математика і статистика</i>
Спеціальність	<i>113 Прикладна математика</i>
Освітня програма	<i>Наука про дані та математичне моделювання</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>3 курс, весняний семестр</i>
Обсяг дисципліни	<i>4,5 кредити ЕКТС / 135 год.: (36 год. – лекції, 18 год. – лабораторні роботи, 81 год. – СРС)</i>
Семестровий контроль/ контрольні заходи	<i>Екзамен, МКР, лабораторні роботи</i>
Розклад занять	<i>http://rozklad.kpi.ua лекція – 1 раз на тиждень, лабораторні – 1 раз на 2 тижні</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: канд. техн. наук, доцент Тавров Данило Юрійович e-mail: tavrov.danylo@i111.kpi.ua Лабораторні: канд. техн. наук, доцент Тавров Данило Юрійович</i>
Розміщення курсу	

Програма навчальної дисципліни

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

Метою навчальної дисципліни є засвоєння студентами основних принципів, ідей та методів аналізу даних, що потрібно для дальшого опанування сучасних методів науки про дані, набуття практичних навичок статистичного опрацювання даних в обраній предметній галузі.

Предмет навчальної дисципліни:

- методи розвідкового аналізу даних та візуалізації;
- методи статистичного виведення на великих вибірках;
- методи регресійного аналізу даних;
- методи оцінювання непараметричних статистичних моделей;
- аналіз головних компонент;
- методи беесівського виведення.

Після засвоєння навчальної дисципліни студенти мають продемонструвати такі компетентності та програмні результати навчання за освітньою програмою:

Загальні компетентності

- ЗК1. Здатність учитися і оволодівати сучасними знаннями;
- ЗК2. Здатність застосовувати знання у практичних ситуаціях;

- ЗК3. Здатність генерувати нові ідеї (креативність);
- ЗК4. Здатність бути критичним і самокритичним;
- ЗК5. Здатність проведення досліджень на відповідному рівні;
- ЗК6. Здатність до абстрактного мислення, аналізу та синтезу;
- ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- ЗК8. Знання та розуміння предметної області та розуміння професійної діяльності;
- ЗК9. Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань/видів економічної діяльності);
- ЗК10. Навички у використанні інформаційних і комунікаційних технологій.

Фахові компетентності

- ФК1. Здатність використовувати й адаптувати математичні теорії, методи та прийоми для доведення математичних тверджень і теорем;
- ФК2. Здатність виконувати завдання, сформульовані у математичній формі;
- ФК3. Здатність обирати та застосовувати математичні методи для розв'язання прикладних задач, моделювання, аналізу, проектування, керування, прогнозування, прийняття рішень;
- ФК9. Здатність до проведення математичного і комп'ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів;
- ФК12. Здатність до пошуку, систематичного вивчення та аналізу науково-технічної інформації, вітчизняного й закордонного досвіду, пов'язаного із застосуванням математичних методів для дослідження різноманітних процесів, явищ та систем;
- ФК13. Здатність зрозуміти постановку завдання, сформульовану мовою певної предметної галузі, здійснювати пошук та збір необхідних вихідних даних;
- ФК14. Здатність сформулювати математичну постановку задачі, спираючись на постановку мовою предметної галузі, та обирати метод її розв'язання, що забезпечує потрібні точність і надійність результату;
- ФК18. Навички розв'язування специфічних математичних та комп'ютерних задач машинного навчання;
- ФК19. Здатність розв'язувати професійні задачі за допомогою систем комп'ютерної математики.

Програмні результати навчання

- РН1. Демонструвати знання й розуміння основних концепцій, принципів, теорій прикладної математики і використовувати їх на практиці;
- РН3. Формалізувати задачі, сформульовані мовою певної предметної галузі; формувати їх математичну постановку та обирати раціональний метод вирішення; розв'язувати отримані задачі аналітичними та чисельними методами, оцінювати точність та достовірність отриманих результатів;
- РН6. Володіти основними методами розробки дискретних і неперервних математичних моделей об'єктів та процесів, аналітичного дослідження цих моделей на предмет існування та єдиності їх розв'язку;
- РН7. Вміти проводити практичні дослідження та знаходити розв'язок некоректних задач;
- РН10. Володіти методиками вибору раціональних методів та алгоритмів розв'язання математичних задач оптимізації, дослідження операцій, оптимального керування і прийняття рішень, аналізу даних;
- РН12. Розв'язувати окремі інженерні задачі та/або задачі, що виникають принаймні в одній предметній галузі: в соціології, економіці, екології та медицині;
- РН13. Використовувати в практичній роботі спеціалізовані програмні продукти та програмні системи комп'ютерної математики;
- РН14. Виявляти здатність до самонавчання та продовження професійного розвитку;
- РН15. Уміти організувати власну діяльність та одержувати результат у рамках обмеженого часу;

РН21. Вміти формулювати та розв'язувати задачі аналізу даних і базові задачі машинного навчання.

2. Пререквізити та постреквізити дисципліни

Пререквізитами для вивчення дисципліни є дисципліни ЗО 10 «Математичний аналіз», ЗО 11 «Алгебра та геометрія», ЗО 25 «Математична статистика». Особливо корисними для успішного опанування дисципліни буде детальне освоєння основ матричної алгебри, безумовної оптимізації функцій багатьох змінних, властивостей випадкових величин та оцінок їхніх параметрів. Корисним є володіння базовими навичками програмування для швидкого опанування мови R.

Вивчення дисципліни є передумовою вивчення дисципліни ПО 10 «Переддипломна практика».

3. Зміст навчальної дисципліни

РОЗДІЛ 1. ВСТУП В АНАЛІЗ ДАНИХ

Тема 1.1. Вступ в аналіз даних та мову R

Тема 1.2. Розвідковий аналіз даних в R

Тема 1.3. Статистичне виведення на великих вибірках

Тема 1.4. Бутстреп

РОЗДІЛ 2. МЕТОДИ РЕГРЕСІЙНОГО АНАЛІЗУ

Тема 2.1. Множинна лінійна регресія

Тема 2.2. Нелінійність у лінійних моделях

Тема 2.3. Аналіз панельних даних

Тема 2.4. Логістична регресія

Тема 2.5. Робота з пропущеними даними

РОЗДІЛ 3. НЕПАРАМЕТРИЧНІ МЕТОДИ

Тема 3.1. Аналіз головних компонент

Тема 3.2. Вибір моделей

Тема 3.3. Непараметричні методи регресійного аналізу

РОЗДІЛ 4. БЕЄСІВСЬКЕ ВИВЕДЕННЯ

Тема 4.1. Основні поняття про беєсівське виведення

Тема 4.2. Алгоритми MCMC

Тема 4.3. Ієрархічні беєсівські моделі

4. Навчальні матеріали та ресурси

Базова література

1. Майборода Р.Є. Комп'ютерна статистика : підручник / Р.Є. Майборода – Київ : ВПЦ «Київський університет», 2019. – 589 с. – Режим доступу: <https://probability.knu.ua/userfiles/mre/cscolor.pdf> (дата звернення: 11.04.2021).

2. Прикладна економетрика : навч. посіб. : у двох частинах. Частина 1 : [Електронне видання] / Л. С. Гур'янова, Т. С. Клебанова, С. В. Прокопович та ін. – Харків : ХНЕУ ім. С. Кузнеця, 2016. – 235 с. Режим доступу: http://repository.kpi.kharkov.ua/bitstream/KhPI-Press/38420/3/Book_2016_Prykladna_ekonometryka_Ch_1.PDF (дата звернення: 11.04.2021).

Допоміжна література

3. James G. An Introduction to Statistical Learning with Applications in R / G. James, D. Witten, T. Hastie, R. Tibshirani. – [2nd ed.]. – New York : Springer New York, 2021. – 615 p. – Mode of Access:

https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023.pdf.download.html (access date: 19.06.2024).

4. Golemund G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data / G. Golemund, H. Wickham. – O'Reilly Media, Inc., 2017. – 510 p. – Mode of Access: <https://digitallibrary.tsu.ge/book/2019/september/books/R-for-Data-Science.pdf> (access date: 19.06.2024).

5. Gelman A. Bayesian Data Analysis / A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. — [3rd ed.]. — Chapman and Hall/CRC, 2013. — 675 p. — Mode of Access: <http://www.stat.columbia.edu/~gelman/book/BDA3.pdf> (access date 19.06.2024).

6. Gelman A. Regression and Other Stories / A. Gelman, J. Hill, A. Vehtari. — Cambridge University Press, 2020. — 546 p. — Mode of Access: <https://users.aalto.fi/~ave/ROS.pdf> (access date: 19.06.2024).

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Навчання здійснюється на основі студентоцентрованого підходу та стратегії взаємодії викладача та студента для засвоєння студентами матеріалу та розвитку у них практичних навичок. Для лекційних занять використовуються пояснювально-ілюстративний метод та метод проблемного виконання, для проведення лабораторних занять використовується дослідницький метод навчання.

За дистанційної форми навчання заняття проводять за допомогою платформи для проведення онлайн-зустрічей Zoom.

Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань
1	Вступ в аналіз даних та мову R Основні питання: вступ в аналіз даних, основи програмування в R, робота з tidyverse.
2	Розвідковий аналіз даних в R Основні питання: розвідковий аналіз даних, дескриптивний аналіз даних, основи використання ggplot2.
3	Статистичне виведення на великих вибірках. Частина 1 Основні питання: основні поняття статистичного виведення, статистичні оцінки, середнє вибіркове та вибіркова дисперсія, асимптотичні розподіли.
4	Статистичне виведення на великих вибірках. Частина 2 Основні питання: довірчі інтервали, тестування гіпотез, тест Волда.
5	Бутстреп Основні питання: plug-in оцінки, алгоритм бутстрепу, застосування до оцінювання стандартних похибок та зміщень, застосування до побудови довірчих інтервалів.
6	Множинна лінійна регресія. Частина 1 Основні питання: функція умовного сподівання, лінійні моделі, оцінювання коефіцієнтів лінійної регресії за методом найменших квадратів, асимптотичні властивості цієї оцінки.
7	Множинна лінійна регресія. Частина 2 Основні питання: зміщення від неврахованих змінних, статистичне виведення для оцінок коефіцієнтів лінійної моделі.
8	Нелінійність у лінійних моделях Основні питання: нелінійні ефекти в лінійних моделях, практичні рекомендації.
9	Аналіз панельних даних. Логістична регресія Основні питання: аналіз панельних даних, регресійний аналіз із бінарною залежною змінною.

№ з/п	Назва теми лекції та перелік основних питань
10	Робота з пропущеними даними Основні питання: класифікація пропущених даних, метод MICE, регресійний аналіз із цензурованими даними.
11	Модульна контрольна робота
12	Аналіз головних компонент. Вибір моделей Основні питання: аналіз головних компонент, вибір моделей, крос-валідація.
13	Непараметричні методи регресійного аналізу Основні питання: лінійні згладжувачі, ядрова оцінка щільностей, ядрова регресія, локальна поліномна регресія, узагальнені лінійні моделі.
14	Основні поняття про беєсівське виведення Основні питання: основні поняття про беєсівське виведення
15	Алгоритми МСМС. Частина 1 Основні питання: методи Монте-Карло, алгоритм Метрополіса-Гейстингса.
16	Алгоритми МСМС. Частина 2 Основні питання: вибірка Гіббса, діагностика МСМС.
17	Ієрархічні беєсівські моделі. Частина 1 Основні питання: ієрархічні моделі, вибір моделей.
18	Ієрархічні беєсівські моделі. Частина 2 Основні питання: робастні регресійні моделі, беєсівські моделі для роботи з панельними даними.

Лабораторні роботи

№ з/п	Назва лабораторної роботи та перелік основних питань
1	Розвідковий аналіз даних в R Основні питання: дескриптивний аналіз обраного датасету, аналіз викидів та пропущених значень, візуалізація за допомогою ggplot2.
2	Статистичне виведення в R Основні питання: побудова довірчих інтервалів в R, тестування гіпотез в R, застосування бутстрепа в R для побудови довірчих інтервалів.
3	Регресійний аналіз в R Основні питання: побудова структурної лінійної моделі, оцінювання моделі в R, здійснення статистичного виведення та порівняння різних специфікацій моделей.
4	Непараметричне виведення в R Основні питання: ядрова оцінка функції умовного сподівання, оцінювання частково лінійної моделі в R, виконання аналізу головних компонент в R.
5	Беєсівське виведення в R Основні питання: оцінювання регресійних моделей в R в рамках беєсівської парадигми.

6. Самостійна робота студента

Студент повинен завчасно готуватись до лекцій. Перед лекціями потрібно повторити вивчений раніше теоретичний матеріал, підготувати питання щодо нової теми для обговорення на лекції. Для написання лабораторних робіт потрібно повторити відповідний теоретичний матеріал.

№ з/п	Вид самостійної роботи	Кількість годин СРС
1.	Виконання лабораторних робіт	40
2.	Підготовка до лекцій	10
3.	Опрацювання лекційного матеріалу	28
4.	Підготовка до екзамену	3
	Загалом	81

7. Політика навчальної дисципліни (освітнього компонента)

Відвідування занять

Відповідно до Наказу 1-273 від 14.09.2020 р. заборонено оцінювати присутність або відсутність здобувача на аудиторному занятті. Матеріал занять, які були з тих чи тих причин пропущено, потрібно опановувати самостійно. У будь-якому випадку студентам рекомендується відвідувати всі види занять, оскільки на них викладається теоретичний матеріал та розвиваються навички, потрібні для виконання завдань на практичних заняттях. Система оцінювання орієнтована на отримання балів за активність студента, а також виконання завдань, що розвивають практичні уміння та навички.

Особливості виконання лабораторних робіт

Лабораторні роботи студенти виконують мовою R. Студенти можуть виконувати лабораторні роботи командами до 5 осіб включно. Звіт і презентацію у форматі PDF та код висилають викладачеві для попередньої перевірки та зберігання в електронній формі. Захист відбувається в усній формі, очно або в режимі конференції (за дистанційної форми навчання). На захисті студенти мають розповісти хід виконання роботи, сформулювати висновки, відповісти на запитання викладача. Роботи потрібно здавати та захищати вчасно, за запізнення оцінка знижується.

Правила складання модульної контрольної роботи

Модульну контрольну роботу студенти виконують під час відповідного лекційного заняття протягом фіксованого часу і здають у паперовому вигляді, якщо заняття відбуваються в очному режимі, або надсилають викладачеві в електронній формі, якщо заняття відбуваються в дистанційному режимі. Роботу студент виконує самостійно. Допустиме використання допоміжних матеріалів, але суворо заборонено обговорювати задачі з колегами чи питати порад у третіх осіб.

Правила складання екзамену

Екзамен студенти виконують письмово у встановлений час і здають у паперовому вигляді, якщо екзамен відбувається в очному режимі, або надсилають викладачеві в електронній формі, якщо екзамен відбувається в дистанційному режимі. Екзамен студент пише самостійно. Допустиме використання допоміжних матеріалів, але суворо заборонено обговорювати задачі з колегами чи питати порад у третіх осіб.

Пропущені контрольні заходи

Студент може спробувати переписати пропущену МКР тільки в тому випадку, якщо її було пропущено з поважної причини, яку підтверджено документально. Переписування МКР відбувається за домовленістю з викладачем. Пропущений екзамен не зараховується; у такому випадку студент отримує запис у відомості «не з'явився» та повинен скласти екзамен на додатковій сесії.

Академічна доброчесність

Політика та принципи академічної доброчесності визначено в 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 визначено в розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Процедура оскарження результатів контрольних заходів

Студенти мають можливість поставити будь-яке питання, яке стосується процедури проведення та/або оцінювання контрольних заходів, та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами. Студенти мають право оскаржити результати контрольних заходів, але обов'язково аргументовано, пояснивши, із яким критерієм не погоджуються відповідно до оціночного листа та/або зауважень.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Поточний контроль: МКР, лабораторні роботи.

Календарний контроль: проводиться двічі на семестр на 8-ому та на 14-ому тижнях, як моніторинг поточного стану виконання вимог силабуса; студент отримує «зараховано», якщо його поточний рейтинг не менший від 50% від максимальної кількості балів, яку може отримати студент до даного календарного контролю.

Семестровий контроль: екзамен.

Умови допуску до семестрового контролю: відсутність заборгованостей із лабораторних робіт та семестровий рейтинг не менший від 25 балів.

Рейтинг студента з кредитного модуля складається з:

- 1) балів за виконання та захист лабораторних робіт;
- 2) балів за виконання модульної контрольної роботи (МКР);
- 3) балів за відповідь на екзамені

СИСТЕМА РЕЙТИНГОВИХ БАЛІВ

1. Бали за виконання та захист лабораторних робіт

Протягом семестру студенти виконують 5 лабораторних робіт. Максимальна кількість балів за лабораторну роботу – 8. Бали нараховуються за:

- якість реалізації роботи: 5 балів;
- відповідь під час захисту лабораторної роботи: 3 бали.

Критерії оцінювання якості реалізації роботи:

5 балів — роботу виконано якісно, в повному обсязі;

4 бали — роботу виконано якісно, в повному обсязі, але вона має вади;

3 бали — роботу виконано не повністю або вона має суттєві помилки;

0 балів — роботу виконано неправильно або вона відсутня.

Критерії оцінювання відповідей під час захисту лабораторної роботи:

3 бали — відповіді під час захисту повні, добре аргументовані;

2 бали — у цілому відповіді правильні, але мають вади чи незначні помилки;

1 бал — відповіді неповні або мають істотні вади;

0 балів — немає відповідей або відповіді неправильні.

Максимальна кількість балів за виконання та захист лабораторних робіт:

8 балів x 5 робіт = 40 балів.

У випадку, якщо лабораторну роботу здано з запізненням від графіка, викладач має право нарахувати штрафні бали. Загальний обсяг штрафних балів за всі лабораторні роботи не може перевищувати **5 балів** за семестр.

2. Модульний контроль

Студенти пишуть одну модульну контрольну роботу. Робота містить чотири завдання практичного характеру. Максимальна кількість балів за відповідь на кожне завдання — 2.5.

Критерії оцінювання відповідей на кожне завдання:

2.5 бали – відповідь повна і правильна;

2 бали – відповідь правильна, але неповна чи містить незначні неточності;

1.5 бали – відповідь неповна і неточна;

0 балів – відповідь неправильна або відсутня.

Максимальна кількість балів за модульний контроль:

2.5 бали x 4 задачі = **10 балів**.

3. Бали за відповідь на екзамені

Екзаменаційний білет складається з 1 теоретичного питання (доведення певного твердження) вагою в 20 балів та 3 задач по 10 балів кожна.

Критерії оцінювання теоретичного питання екзаменаційної роботи:

20 балів – відповідь правильна, повна;

18-19 балів – відповідь майже правильна, можливі незначні вади або термінологічні помилки;

15-17 балів – у відповіді є істотні помилки;

12-14 балів – у відповіді є істотні помилки;

0 балів – відповіді немає.

Критерії оцінювання розв'язань кожної окремої задачі:

10 балів – задачу розв'язано правильно;

8-9 балів – задачу розв'язано майже правильно, можливі незначні вади чи арифметичні помилки;

6-7 балів – задачу розв'язано не повністю, існують вади;

0 балів – розв'язок відсутній.

Максимальна кількість балів за відповідь на екзамені:

20 балів x 1 теоретичне питання + 10 балів x 3 задачі = **50 балів**

Розрахунок шкали (R) рейтингу:

Сума вагових балів контрольних заходів протягом семестру складає:

$$R_c = 40 + 10 = 50 \text{ балів.}$$

Із цієї суми за семестр студент може набрати певну кількість балів r_c .

Екзаменаційна складова (R_E) шкали дорівнює 50 балам, тобто складає 50% від R .

За відповідь на екзамені студент може набрати певну кількість балів r_E .

Рейтингова шкала з дисципліни складає

$$R = R_c + R_E = 100 \text{ балів.}$$

Неодмінною умовою допуску до екзамену є відсутність заборгованостей із лабораторних робіт, а також набрані $r_c = 25$ балів.

Якщо студент не виконав умов допуску до екзамену, то він у першу відомість отримує оцінку «не допущено», а потім повинен виконати ці умови (здавати лабораторні роботи та набирати бали для допуску). Після отримання допуску студент складає екзамен (на першому та/або другому перескладанні) за наведеними раніше правилами, а його остаточну рейтингову оцінку R_D розраховують за формулою: $R_D = r_c + r_E$.

За умови дистанційного семестрового контролю оцінка розраховується за формулою за умови відсутності заборгованостей із лабораторних робіт та набраних щонайменше $r_c = 25$ балів. Перерахунок стартового рейтингового балу з дисципліни у 100-у шкалу здійснюється за формулою

$$R_D = 60 + \frac{40(r_c - 25)}{25}$$

Перескладання отриманої за формулою 100-бальної оцінки не допускається. Якщо умови допуску до екзамену до кінця семестру не виконано, то студент у першу відомість отримує оцінку

«не допущено», а потім виконує умови допуску (на перше перескладання) або спілкування за темами дисципліни у режимі конференції (на комісії) під час додаткової сесії.

Оцінку за дисципліну виставляють відповідно до значення ***R_D*** згідно з таблицею.

<i>Кількість балів</i>	<i>Оцінка</i>
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконано умов допуску	Не допущено

Робочу програму навчальної дисципліни (силабус):

Склав доцент кафедри ПМА, канд. техн. наук, доцент Данило ТАВРОВ

Ухвалено кафедрою ПМА (протокол № 18 від 10.06.2024).

Погоджено Методичною комісією факультету (протокол № 12 від 21.06.2024)