



ПРИКЛАДНИЙ СТАТИСТИЧНИЙ АНАЛІЗ ДАНИХ

Робоча програма кредитного модуля навчальної дисципліни «Прикладний статистичний аналіз даних» (Силабус)

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	5 курс, весняний семестр
Обсяг дисципліни	120 годин (36 години – Лекції, 18 години – Практичні, 66 години – СРС)
Семестровий контроль/ контрольні заходи	Залік/модульна контрольна робота, розрахункова робота
Розклад занять	ozklad.kpi.ua/Schedules/ScheduleGroupSelection.aspx
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: канд. ф.-м.н., доцент кафедри математичного аналізу та теорії ймовірностей, Диховичний Олександр Олександрович, a.dyx@ukr.net , моб. +38(067)9005262 Практичні: канд. ф.-м.н., доцент кафедри математичного аналізу та теорії ймовірностей, Диховичний Олександр Олександрович, a.dyx@ukr.net , моб. +38(067)9005262
Розміщення курсу	https://campus.kpi.ua , https://do.ipk.kpi.ua/enrol/index.php?id=4258.1

Програма навчальної дисципліни

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

Цілі дисципліни	Ціллю навчальної дисципліни є: формування у здобувачів освіти уміння застосовувати сучасні комп'ютерні технології до постановки, аналізу та розв'язання основних задач прикладного статистичного аналізу даних
Предмет навчальної дисципліни	Предмет навчальної дисципліни « Прикладний статистичний аналіз даних » - вивчення основ прикладного статистичного аналізу даних та розв'язання базових задач засобами програмного середовища R.
Компетентності	<p>Метою навчальної дисципліни є формування у студентів здатностей:</p> <ul style="list-style-type: none"> ● Здатність учитися, здобувати нові знання, уміння, у тому числі в галузях, відмінних від математики (ЗК1); ● Здатність використовувати у професійній діяльності знання з галузей математичних, природничих, соціально-гуманітарних та економічних наук (ЗК2); ● Здатність вирішувати проблеми у професійній діяльності на основі абстрактного мислення, аналізу, синтезу та прогнозу (ЗК3); ● Здатність застосовувати міждисциплінарні підходи при критичному осмисленні математичних проблем (ФК2); ● Спроможність розробляти математичну модель ситуації з реального світу та переносити математичні знання у нематематичні контексти (ФК5); ● 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 використовувати методи математичної статистики у теоретичних і прикладних дослідженнях; ● Спроможність доводити розв'язок задачі до зрозумілого результату – адекватного математичного та алгоритмічного апаратів; ● Здатність аналізувати отримані результати, здатності до узагальнення, постановки задачі та вибору шляхів її розв'язання; ● 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	<p>Після засвоєння навчальної дисципліни студенти мають продемонструвати такі результати навчання:</p> <ul style="list-style-type: none"> ● Відтворювати знання фундаментальних розділів математики в обсязі, необхідному для володіння

	<p>математичним апаратом відповідної галузі знань і використання математичних методів у обраній професії (PH2);</p> <ul style="list-style-type: none"> ● Володіти математичними методами аналізу, прогнозування та оцінки параметрів моделей, математичними способами інтерпретації числових даних та принципами функціонування природничих процесів (PH4); ● Знати головні результати та сфери застосувань основних математичних теорій, що використовуються при математичному моделюванні, фінансовій та актуарній математиці, методів математичної економіки та імітаційного моделювання, комп'ютерної статистики (PH14); ● Основних принципів роботи у середовищі R та R-studio; ● Основних типів даних, операцій, над ними, елементів програмування, методів графічного відображення даних та результатів їх обробки; ● Методів роботи у середовищі R з основними ймовірнісними розподілами; ● Методів розв'язання засобами мови R задач описової статистики та графічного аналізу даних; ● Базових принципів перевірки засобами мови R основних статистичних гіпотез; ● Методів розв'язання засобами мови R задач дисперсійного аналізу даних; ● Методів побудови засобами мови R моделей лінійного регресійного аналізу даних; <ul style="list-style-type: none"> ● Методів логістичної регресії ● Методів факторного аналізу ● Методів кластерного аналізу ● Методів дискримінантного аналізу ● Методів аналізу моделей часових рядів та їх прогнозування. <p>УМІННЯ:</p> <ul style="list-style-type: none"> ● Працювати з базовою версією мови програмування R та R-studio; ● Використовувати основні типи даних мови програмування R: вектори, масиви, матриці, фактори, списки та фрейми, якими зображують статистичні дані; ● Працювати з функціями експорту та імпорту даних; складати власні функції, застосовувати основні оператори мови програмування R; ● Графічно відображати засобами мови програмування R різні типи статистичних даних та результати їхнього первинного аналізу; ● Використовувати реалізовані у середовищі R функціональні можливості роботи з базовими статистичними розподілами: та проводити розрахунки їх основних ймовірнісних характеристик;
--	--

	<ul style="list-style-type: none"> • Засобами мови R: <ul style="list-style-type: none"> ✓ проводити первинний статистичний аналіз одновимірних даних; ✓ перевіряти основні статистичні гіпотези; ✓ проводити дисперсійний аналіз; ✓ будувати та аналізувати моделі лінійного регресійного аналізу; ✓ будувати та аналізувати моделі логістичної регресії; ✓ проводити факторний аналіз; ✓ проводити кластерний аналіз; ✓ проводити дискримінантний аналіз; ✓ будувати та аналізувати моделі часових рядів, будувати для них прогноз.
--	--

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

Пререквізити: Дисципліна «Комп'ютерна статистика» 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 «Комп'ютерна статистика» передуватиме 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
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
Розділ 1. Основи програмування у системі R				
Тема 1.1. Основи роботи з системою R Основні об'єкти мови програмування R та операції над ними. Базові прийоми програмування та графіка мови програмування R	8	4	-	4
Разом за розділом 1	8	4	-	4

Розділ 2. Розв’язування статистичних задач мовою програмування R				
Тема 2.1. Описова статистика одновимірних числових даних	6	2	-	4
Тема 2.2. Статистичні критерії	8	2	2	4
Тема 2.3. Дисперсійний аналіз	10	4	2	4
Тема 2.4. Регресійний аналіз	12	6	2	4
Тема 2.5. Логістична регресія	10	4	2	4
Тема 2.6. Факторний аналіз	8	2	2	4
Тема 2.7. Кластерний аналіз	10	4	2	4
Тема 2.8. Дискримінантний аналіз	8	2	2	4
Тема 2.9. Аналіз часових рядів	10	4	2	4
МКР	12		2	10
ДКР	12			12
Разом за розділом 2	106	30	18	58
Залік	6	2		4
Всього годин	120	36	18	66

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. Майборода Р.Є, Сугакова О.В. Аналіз даних за допомогою R.- Навчальний посібник. «Київський університет», 2015. – 65 с.
4. Оленко А.Я. Комп’ютерна статистика. Навчальний посібник.— К., ВПЦ “Київський університет”, 2007. -174с.
5. Jenine K. Harris, Statistics With R .Washington University in St.Louis, USA, 2020.
6. Sarah Stowell. Using R for Statistics. Apress, 2014.
7. Турчин В.М. Теорія ймовірностей і математична статистика,- Дніпропетровськ, ІМА-прес, 2014. - 566 с.
8. Shumway R.H., Staffer D.S. Time series analysis and its applications. With R examples, 3d edition, Springer, 2011. – 202 p. . [Електронний ресурс] – pzs.dstu.dp.ua/DataMining/times/bibl/TimeSeries.pdf

Додаткова література

9. Thomas Rahlf. Data Visualisation with R. Springer International Publishing, New York, 2017.
10. Shao J. Mathematical statistics.- Springer-Verlag: New York, 1998. -530 p.
11. Shumway Robert H., Stoer David S. Time Series Analysis and Its Applications With R Examples.- Springer, 2010.
12. James G., Witten D., Hastie T., Tibshiani R An introduction to statistical learning with applications in R.- Springer, 2013.
13. Лупан І.В., Авраменко О.В. Комп'ютерні статистичні пакети. Навчально-методичний посібник, Кировоград, 2010.

Інформаційні ресурси

14. <http://cran.r-project.org/bin/windows/base/>
15. <https://buildmedia.readthedocs.org/media/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf>
16. Дистанційний курс «Аналіз даних мовою R». Круглова Н.В., Диховичний О.О. <https://do.ipk.kpi.ua/login/index.php>
17. <https://loginom.ru/blog/logistic-regression-roc-auc>

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 та завдання на СРС)
1.	<u>Основи роботи з середовищем R. R-studio. Основні об'єкти та операції над ними у середовищі R.</u> <i>Рекомендована література:</i> [2], глава 1,2; [3], §1,2
2.	<u>Програмування у середовищі R: Імпорт/експорт даних; функції; основні оператори; базова графіка.</u> <i>Рекомендована література:</i> [2], глава3,4, 6,[3], §2, §4
3.	<u>Описова статистика одновимірних числових даних. Первинний графічний аналіз даних. Інтервальні оцінки.</u> <i>Рекомендована література:</i> [2], глава 3 [1], глава 4,8.
4.	<u>Статистичні критерії.</u> <i>Рекомендована література:</i> [3], §8 , [1], глава 9.
5.	<u>Однофакторний дисперсійний аналіз . Обґрунтування та розрахунок основних статистик. Побудова та перевірка критеріїв. Статистичні висновки.</u> <i>Рекомендована література:</i> [3], §10 , [1], глава 9
6.	<u>Двофакторний дисперсійний аналіз . Обґрунтування та розрахунок основних статистик.</u>

	Побудова та перевірка критеріїв. Статистичні висновки. <i>Рекомендована література:</i> [3], §11 .
7.	<u>Кореляційний аналіз.</u> Діаграми розіювання. Коефіцієнти кореляції Пірсона, Спірмена, Кендалла. <i>Рекомендована література:</i> [1], глава 5
8.	<u>Регресійний аналіз.</u> Розрахунок параметрів простої лінійної регресії. Розрахунок похибок оцінок, Побудова довірчих інтервалів. Перевірка узгодженості моделі. Прогноз. Графічне зображення результатів. <i>Рекомендована література:</i> [1], глава 10, [3], §6
9.	<u>Множинна регресія.</u> Модель. Оцінки. Властивості оцінок. Аналіз рівняння регресії. Перевірка значущості рівняння множинної регресії. Довірчі інтервали для коефіцієнтів регресії. Прогноз. <i>Рекомендована література:</i> [1], глава 10, [3], §6
10.	<u>Логістична регресія.</u> Оцінки. Аналіз рівняння регресії. Перевірка значущості рівняння регресії. Прогноз. <i>Рекомендована література:</i> [17], [12], §4
11.	<u>ROC- аналіз.</u> <i>Рекомендована література:</i> [17]
12	<u>Факторний аналіз.</u> <i>Рекомендована література:</i> [3] , §11
13- 14	<u>Кластерний аналіз</u> <i>Рекомендована література:</i> [12] , §4
15	<u>Дискримінантний аналіз</u> <i>Рекомендована література:</i> [12] , §4, [13] , §7
16	<u>Часові ряди.</u> Моделі. Розклад часового ряду на компоненти. Тренд. Періодична компонента. Згладжування. Прогноз. <i>Рекомендована література:</i> [8], глава 2, [15]
17	<u>Моделі ARIMA.</u> Модель. Побудова моделі. Прогноз на підставі моделі . <i>Рекомендована література:</i> [8], глава 3, [15]
18	<u>Залік.</u>

Практичні заняття

№ з/п	Теми практичних занять (перелік дидактичних матеріалів, посилання на літературу та завдання на СРС)
1.	<u>Описова статистика</u> одновимірних числових даних. Статистики середнього положення. Статистики розкиду. Статистики форми розподілу. Гістограма. Box-whisker plots. P-P і Q-Q діаграми. <u>Довірчі інтервали.</u> <i>Рекомендована література:</i> [1], глава 6
2.	<u>Перевірка гіпотез.</u> Критерії перевірки простих гіпотез про параметри вибірки. Критерії перевірки гіпотези про розподіл вибірки. <i>Рекомендована література:</i> [3], §8 , [1], глава 9
3.	<u>Дисперсійний аналіз</u> . Розрахунок основних статистик. Побудова та перевірка критеріїв. <i>Рекомендована література:</i> [3], §10 , [1], глава 9

4.	<u>Регресійний аналіз</u> . Розрахунок параметрів простої лінійної регресії. Розрахунок похибок оцінок, побудова довірчих інтервалів. Перевірка узгодженості моделі. Прогноз. Графічне зображення результатів. <i>Рекомендована література:</i> [1], глава 10
5.	<u>Логістична регресія</u> . Оцінки. Аналіз рівняння регресії. Перевірка значущості рівняння регресії. Прогноз. <i>Рекомендована література:</i> [17]
6.	<u>Факторний аналіз</u> . <i>Рекомендована література:</i> [3] , §11
7.	<u>Кластерний аналіз</u> <i>Рекомендована література:</i> [12] , §4
8.	<u>Часові ряди</u> . Моделі. Розклад часового ряду на компоненти. Знаходження тренду, періодичної компоненти. Згладжування. Прогноз. <u>Моделі ARIMA</u> . Модель. Побудова моделі. Прогноз на підставі моделі . <i>Рекомендована література:</i> [8,15]
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	ДКР	МКР	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;
- 2) МКР;
- 3) ДКР;
- 4) відповідь на заліку.

Система рейтингових (вагових) балів та критерії оцінювання

1. Робота на практичних заняттях

Ваговий бал – 30. Максимальна кількість балів на всіх практичних заняттях дорівнює 30 балам.

2. Модульний контроль

Ваговий бал - 40. Максимальна кількість балів за контрольну роботу дорівнює 40 балів.

Критерій оцінювання МКР:

відсутність на контрольній роботі – 0 балів,

оцінка МКР (в балах) дорівнює величині відсотка (від максимальної кількості балів 40) її виконання.

При виконання < 60% контрольна робота не зараховується.

3. Домашня контрольна робота

Ваговий бал –30.

Критерій оцінювання ДКР:

Невиконання ДКР – 0 балів

ДКР виконується і захищається частинами, які за змістом відповідають модульній контрольній роботі. Ця частина ДКР здається до написання МКР, а сама МКР є її захистом.

Оцінка ДКР (в балах) дорівнює величині відсотка (від максимальної кількості балів 30) її виконання з урахуванням результату написання відповідної МКР.

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

За несвоєчасне (пізніше ніж на тиждень) подання розрахункової 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.

Поточний контроль може проводитись у вигляді тестових контрольних робіт в Moodle.

Робочу програму навчальної дисципліни (силабус):

Складено:

Доцент кафедри МАтаТЙ, канд. фіз.-мат. наук, доцент Диховичний О.О.

Ухвалено кафедрою МАтаТЙ (протокол № 16 від 8.07.22)

Погоджено Методичною комісією ФМФ (протокол № 8 від 11.07.2022)

