

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра Комп'ютерних наук та інформаційних технологій (№ 302)

ЗАТВЕРДЖУЮ

Гарант освітньої програми/


(підпис) Мирослав МОМОТ
(ініціали та прізвище)

«_____» _____ 2024 р.

РОБОЧА ПРОГРАМА ОBOB'ЯЗКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Статистичні та імовірнісні методи дата-аналізу

(назва навчальної дисципліни)

Галузь знань: 12 «Інформаційні технології»
(шифр і найменування галузі знань)

Спеціальність: 122 – Комп'ютерні науки,
(код та найменування спеціальності)

Освітня програма: Комп'ютеризація обробки інформації та управління
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2024 рік

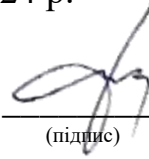
Розробник: Малєєва О.В., професор, д.т.н.
(прізвище та ініціали, посада, науковий ступінь та вчене звання)


(підпис)

Робочу програму розглянуто на засіданні кафедри комп'ютерних наук та інформаційних технологій

Протокол № 671/07 від « 27 » _серпня_ 2024 р.

Завідувач кафедри д.т.н., проф.
(науковий ступінь і вчене звання)


(підпис)

О.Є. Федорович
(ініціали та прізвище)

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, освітня програма, рівень вищої освіти	Характеристика навчальної дисципліни (денна форма навчання) Цикл професійної підготовки 4- й семестр – обов’язкова 5- й семестр – обов’язкова
Кількість кредитів – семестр 4 – 7,5 семестр 5 – 2	Галузь знань <u>12 «Інформаційні технології»</u> <small>(шифр та найменування)</small> Спеціальність: 6.122 – Комп’ютерні науки, <small>(код та найменування спеціальності)</small> Освітня програма: <u>Комп’ютеризація обробки інформації та управління</u> <small>(найменування освітньої програми)</small> Рівень вищої освіти: перший (бакалаврський)	Навчальний рік
Кількість модулів – 3		2024/2025
Кількість змістових модулів – 2		Семестр
Індивідуальне завдання РР “Статистичний аналіз варіаційних рядів” <small>(назва)</small>		4-й, 5-й
Загальна кількість годин: семестр 4 – 225, семестр 5 – 76		Лекції*
Кількість тижневих годин для денної форми навчання: - аудиторних семестр 4 – 5 години, семестр 5 – 1 година; - самостійної роботи студента: семестр 4 – 9 годин, семестр 5 – 3 години.		4- й семестр – 32 години, 5- й семестр – 0 годин
		Практичні, семінарські*
		4- й семестр – 16 годин, 5- й семестр – 16 годин
	Лабораторні*	
	4- й семестр – 32 години, 5- й семестр – 0 годин	
Самостійна робота	4- й семестр – 145 годин, 5- й семестр – 44 години	
Вид контролю	4- й семестр – модульний контроль, іспит 5- й семестр – діф. залік	

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: 80/145

* Аудиторне навантаження може бути зменшене або збільшене на одну годину залежно від розкладу занять.

2. Мета та завдання навчальної дисципліни

Мета вивчення: надання студентам основних положень теорії ймовірності та математичної статистики для проведення дата-аналізу стохастичних процесів та їх оцінювання.

Завдання: вивчення сучасних підходів теорії ймовірності для стохастичного дата-аналізу складних систем.

Компетентності, які набуваються:

Загальні:

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

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

ЗК5. Здатність спілкуватися іноземною мовою.

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

ЗК11. Здатність приймати обґрунтовані рішення.

Спеціальні (фахові):

СК2. 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.

СК3. Здатність до логічного мислення, побудови логічних висновків, використання формальних мов і моделей алгоритмічних обчислень, проектування, розроблення й аналізу алгоритмів, оцінювання їх ефективності та складності, розв'язності та нерозв'язності алгоритмічних проблем для адекватного моделювання предметних областей і створення програмних та інформаційних систем.

СК9. Здатність реалізувати багаторівневу обчислювальну модель на основі архітектури клієнт-сервер, включаючи бази даних, знань і сховища даних, виконувати розподілену обробку великих наборів даних на кластерах стандартних серверів для забезпечення обчислювальних потреб користувачів, у тому числі на хмарних сервісах.

СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

СК16. Здатність реалізовувати високопродуктивні обчислення на основі хмарних сервісів і технологій, паралельних і розподілених обчислень при розробці й експлуатації розподілених систем паралельної обробки інформації.

Очікувані результати навчання:

ПР3. 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.

ПР4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

ПР10. Використовувати інструментальні засоби розробки клієнт-серверних застосувань, проектувати концептуальні, логічні та фізичні моделі баз даних, розробляти та оптимізувати запити до них, створювати розподілені бази даних, сховища та вітрини даних, бази знань, у тому числі на хмарних сервісах, із застосуванням мов веб-програмування.

ПР12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій Data Mining, Text Mining, Web Mining.

Пререквізити – Вища математика, Моделі та методи дискретної математики, Створення візуальних інтерфейсів, Структури даних, Мобільні та хмарні технології, Веб-технології та веб-дизайн.

Кореквізити – Оптимізація рішень в комп'ютерних системах управління, Технології системного аналізу, Статистичні та імовірнісні методи дата-аналізу (КР).

3. Програма навчальної дисципліни

Модуль 1.

Змістовний модуль 1. Основи теорії ймовірностей. Ряди розподілу випадкових величин.

Тема 1. Вступ до навчальної дисципліни «Статистичні та імовірнісні методи дата-аналізу»

Історія розвитку теорії ймовірностей та математичної статистики. Предмет та методи дисципліни. Зв'язок з іншими дисциплінами.

Тема 2. Основи теорії ймовірностей

Вступ до навчальної дисципліни. Випадкові події. Аксиоми теорії ймовірностей. Визначення імовірності. Геометрична ймовірність. Задачі з обчислення геометричної ймовірності. Умовна ймовірність. Незалежність подій. Формула повної імовірності. Формула Байєса. Схема Бернуллі, гранична теорема Пуассона. Локальна та інтегральна теореми Муавра – Лапласа.

Тема 3. Статистична сукупність. Ряди розподілу.

Предмет та метод дата-аналізу. Загальна, математична та спеціальні види статистики Сукупність та її елементи. Варіаційні ряди. Візуалізація результатів аналізу в процесі розв'язування прикладних задач (графічне зображення варіаційних рядів). Функція і щільність розподілу безперервних величин. Види кривих розподілу. Статистичні показники. Абсолютні величини. Відносні величини. Основні можливості програмного пакету STATISTICA. Інтерфейс

користувача. Робота з файлами даних в пакеті STATISTICA. Високопродуктивні обчислення на основі хмарних сервісів і технологій, паралельних і розподілених обчислень.

Тема 4. Середні величини.

Середня арифметична. Багаторівнева обчислювальна модель (загальна і часткові середні). Властивості середньої арифметичної. Середня гармонійна і геометрична. Загальний вид середніх. Медіана і мода. Квартилі і децилі. Середні в рядах розподілу випадкових величин.

Модульний контроль

Модуль 2.

Змістовний модуль 2. Статистичні гіпотези та критерії.

Тема 1. Вимір варіації

Показники варіації. Правило додавання дисперсій. Дисперсія в рядах розподілу випадкових величин. Моменти розподілу. Характеристики асиметрії й ексцесу.

Тема 2. Статистичні гіпотези та критерії.

Дискретні закони розподілу. Безперервні закони розподілу. Види розподілу: біноміальний, рівномірний, бета-розподіл. Задачі перевірки гіпотез. Види статичних критеріїв, потужність критерію. Типи задач з перевірки гіпотез. t-критерій. Критерії згоди. Число ступенів свободи. Критерій Пірсона, Романовського та Колмогорова-Смірнова.

Тема 3. Вимір зв'язку. Кореляційний, регресійний і дисперсійний аналіз.

Виявлення статистичних закономірностей недетермінованих явищ. Рівняння регресії. Метод найменших квадратів. Лінійний коефіцієнт кореляції. Коефіцієнт детермінації. Нелінійні рівняння регресії. Множинна регресія. Формальні моделі алгоритмічних обчислень (побудова рівнянь залежностей методом відхилень). Розрахунок коефіцієнта й індексу кореляції, коефіцієнта стійкості зв'язку. Розв'язання задач розпізнавання, класифікації, ідентифікації об'єктів керування (дисперсійний аналіз, Ф-критерій).

Модульний контроль

Змістовний модуль 3. Ряди динаміки та теорія вибірки.

Тема 1. Ряди динаміки.

Характеристики рядів динаміки. Методи вирівнювання динамічних рядів. Вирівнювання динамічних рядів за прямою, за показовою функцією, за параболою, за рядом Фур'є. Автокореляція, кореляція в рядах динаміки.

Тема 2. Основи теорії вибірки.

Помилки спостереження. Генеральна сукупність і вибірка. Помилки вибірки при повторному і неповторному доборі. Три задачі теорії вибірки: визначення довірчих інтервалів, довірчій імовірності і чисельності вибірки. Способи добору: випадковий добір, районований добір, серійний добір. Мала вибірка.

Модульний контроль

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Денна форма				
	Усього	У тому числі			
		л	п	лаб.	с. р.
Модуль 1					
Змістовий модуль 1. Основи теорії ймовірностей. Ряди розподілу випадкових величин.					
Тема 1. Вступ до навчальної дисципліни «Статистичні та імовірнісні методи дата-аналізу»	1	1	-	-	-
Тема 2. Основи теорії ймовірностей.	33	6	6	-	21
Тема 3. Статистична сукупність. Ряди розподілу.	26	3	1	4	18
Тема 4. Середні величини.	17	2	2	4	9
Модульний контроль	2	2	-	-	-
Разом за змістовним модулем 1	79	14	9	8	48
Модуль 2					
Змістовий модуль 2. Статистичні гіпотези та критерії.					
Тема 1. Вимір варіації.	13	2	2	4	5
Тема 2. Статистичні гіпотези та критерії.	25	3	2	8	12
Тема 3. Вимір зв'язку. Кореляційний, регресійний і дисперсійний аналіз.	30	3	1	8	18
Модульний контроль	2	2	-	-	-
Разом за змістовним модулем 2	70	10	5	20	35
Змістовий модуль 3. Ряди динаміки та теорія вибірки					
Тема 1. Ряди динаміки	29	3	1	4	21
Тема 2. Основи теорії вибірки	13	3	1	-	9
Модульний контроль	2	2			
Разом за змістовним модулем 3	44	8	2	4	30
Модуль 2					
Індивідуальне завдання		-	-	-	32
Усього за 4-й семестр	225	32	16	32	145
Курсова робота з дисципліни	60	-	16	-	44
Усього за 5-й семестр	60	-	16	-	44
Усього годин з дисципліни	285	32	32	32	189

5. Теми семінарських занять

№ з/п	Назва теми	Кількість годин
1	Не передбачено навчальним планом	

6. Теми практичних занять

№ з/п	Назва теми	Кількість годин
1	Визначення ймовірності. Залежні і незалежні події. Додавання і множення ймовірностей.	2
2	Умовна ймовірність. Формула повної імовірності.	2
3	Формули Бернуллі і Пуассона. Локальна і інтегральна формули Муавра – Лапласа.	2
4	Побудова дискретних і інтервальних варіаційних рядів. Графічне зображення варіаційних рядів.	1
5	Обчислення середньої арифметичної зваженої в інтервальних рядах. Розрахунок загальної і часткової середніх. Розрахунок середньої гармонійної і геометричної. Визначення медіани і моди, кватилей і децилей у дискретних та інтервальних рядах.	2
6	Розрахунок показників варіації. Розрахунок загальної дисперсії на основі міжгрупової і внутрігрупової. Характеристики асиметрії та ексцесу.	2
7	Визначення закономірностей випадкових явищ, їх властивостей та операцій над ними. Оцінка відповідності варіаційних рядів дискретним та безперервним розподілам за критеріями згоди (Пірсона, Колмогорова-Смірнова, Романовського). Дисперсійний аналіз.	2
8	Розподілена обробка великих наборів даних Розрахунок характеристик рядів динаміки. Аналітичне вирівнювання динамічних рядів. Визначення автокореляції, кореляції в рядах динаміки.	2
9	Рішення задач теорії вибірки (визначення довірчих інтервалів, довірчій імовірності і чисельності вибірки).	1
	Разом	16

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Основні прийоми роботи в пакеті Statistica	4
2	Розвідницький аналіз даних: обчислення основних	4

	статистик.	
3	Оцінка виду розподілу за статистичними критеріями.	4
4	Дослідження статистичної значущості розходжень в групах(ч.1).Т-критерій	4
5	Дослідження статистичної значущості розходжень в групах(ч. 2). Дисперсійний аналіз.	4
6	Багатовимірний аналіз даних Дослідження статистичної залежності даних: аналіз таблиця спряженості, кореляційний аналіз.	4
7	Проведення регресійного аналізу: множинна регресія, нелінійна регресія.	4
8	Часові ряди.	4
	Разом	32

8. Самостійна робота

№ з/п	Назва теми	Кількість годин
4-й семестр		
Змістовий модуль 1		
1	Тема 2. Задачі з обчислення геометричної ймовірності. Відхилення відносної частоти від постійної імовірності в незалежних випробуваннях. Закон великих чисел.	21
2	Тема 3. Робота з файлами даних в пакеті STATISTICA. Графічні можливості програмного пакету STATISTICA.	18
3	Тема 4. Метод моментів.	9
Змістовий модуль 2		
4	Тема 1. Метод доданків обчислення середньої та дисперсії. Метод найбільшої правдоподібності.	5
5	Тема 2. Види розподілу: біноміальний, рівномірний, бета-розподіл.	12
6	Тема 3. Визначення виду залежності методом ломаної. Рангова кореляція.	18
Змістовий модуль 3		
7	Тема 1. Корелювання послідовних різниць. Методи прогнозування. Метод експоненційного вирівнювання. Помилки прогнозування.	21
8	Тема 2. Помилки вибірки при районуваному та серійному доборі. Індексний метод.	9
9	Індивідуальне завдання	32
Разом за 4-й семестр		145

№ з/п	Назва теми	Кількість годин
4-й семестр		
Змістовий модуль 1		
5-й семестр		
1	Аналіз існуючих систем, публікацій за предметною областю	10
2	Математичні методи дослідження	10
3	Вибір інструментальних засобів	5
4	Реалізація дослідження	15
5	Оформлення розрахунково-пояснювальної записки за зазначеними вимогами.	4
Разом за 5-й семестр		44
Усього годин з дисципліни		189

9. Індивідуальні завдання

РР на тему «Статистичний аналіз варіаційних рядів»

10. Методи навчання

Проведення лекцій, практичних занять, лабораторних робіт, індивідуальні консультації з питань нового матеріалу, самостійна робота студентів.

11. Методи контролю

Виконання самостійних письмових робіт, захист лабораторних робіт, написання тестів з теоретичного матеріалу, іспит (4 семестр), діф. залік (5 семестр).

12. Критерії оцінювання та розподіл балів, які отримують здобувачі

12.1. Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Складові навчальної роботи	Бали за одне заняття	Кількість занять	Сумарна кількість балів
Змістовний модуль 1			
Робота на практичних заняттях	0...2	4	0...8
Виконання і захист лабораторних робіт	0...5	2	0...10
Модульний контроль (самостійні письмові роботи)	0...6	2	0...12

Модульний контроль (тест)	0...8	1	0...8
Змістовний модуль 2			
Робота на практичних заняттях	0...2	3	0...6
Виконання і захист лабораторних робіт	0...5	5	0...23
Модульний контроль (самостійні письмові роботи)	0...6	1	0...6
Модульний контроль (тест)	0...7	1	0...7
Змістовний модуль 3			
Робота на практичних заняттях	0...2	1	0...2
Виконання і захист лабораторних робіт	0...5	1	0...5
Модульний контроль (самостійні письмові роботи)	0...6	1	0...6
Модульний контроль (тест)	0...7	1	0...7
Усього за семестр			0...100

Семестровий контроль (іспит) проводиться у разі відмови студента від балів поточного тестування й за наявності допуску до іспиту. Під час складання семестрового іспиту студент має можливість отримати максимум 100 балів.

Білет для іспиту складається з теоретичного тесту та задачі. За тест (20 запитань) студент отримає максимально 70 балів. За повне та правильне вирішення задачі – 30 балів.

12.2 Критерії оцінювання роботи студента протягом семестру

Задовільно (60-74). Мати мінімум знань та умінь. Відпрацювати та захистити всі лабораторні роботи. Знати основи теорії ймовірностей. Знати способи графічного зображення варіаційних рядів. Знати способи обчислення середніх величин, дисперсії; критерій Пірсона, метод найменших квадратів. Знати деякі методи вирівнювання динамічних рядів. Знати основні задачі теорії вибірки.

Добре (75-89). Твердо знати мінімум знань, виконати усі завдання. Показати вміння виконувати та захищати всі лабораторні роботи в обумовлений викладачем строк з обґрунтуванням рішень та заходів, які запропоновано у роботах. Знати формули повної імовірності, Байєса, схему Бернуллі. Вміти будувати функцію та щільність розподілу. Знати різні види

середніх величин. Вміти обчислювати показники варіації. Знати основні закони розподілу та критерії згоди. Вміти будувати рівняння регресії та оцінювати тісноту зв'язку. Вміти визначати автокореляцію у часових рядах. Знати основи теорії вибірки.

Відмінно (90-100). Повно знати основний та додатковий матеріал. Знати усі теми. Орієнтуватися у підручниках та посібниках. Досконально знати методи теорії ймовірностей. Знати загальний вид середніх та вміти обчислювати структурні середні. Знати правило додавання дисперсій та моменти розподілу. Вміти застосовувати декілька критеріїв для перевірки гіпотез. Знати основи кореляційного, регресійного та дисперсійного аналізу. Знати аналітичні способи вирівнювання часових рядів. Вміти визначати помилки вибірки при повторному і неповторному відборі.

Розподіл балів, які отримують студенти за виконання курсової роботи

Пояснювальна записка	Ілюстративна частина	Захист роботи	Сума
до 50	до 30	до 20	100

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

13. Методичне забезпечення

1. Імовірісно-статистичні методи в інформаційно-керуючих системах / О.Є.Федорович, О.В. Малєєва, Н.В. Нечіпорук. - Навч. посібник. - Харків: Нац. аерокосм. ун-т «Харків. авіац. ін-т», 2005. - 202 с.
2. Математична статистика в інформаційних системах/О.В. Малєєва, О.В. Максименко. - Навч. посібник з практичних занять. - Харків: Нац. аерокосм. ун-т «Харків. авіац. ін-т», 2006. – 62 с.
3. Малєєва, О. В. Статистичний аналіз даних (у програмному пакеті STATISTICA 10.0): навч. посіб. до лаб. практикуму / О. В. Малєєва, А. Ю. Юркевич. – Харків : Нац. аерокосм. ун-т ім. М. Є. Жуковського «Харків. авіац. ін-т», 2018. – 48 с.
4. Комплекс навчально-методичного забезпечення дисципліни; презентаційні матеріали; інтерактивні лабораторні роботи (<https://mentor.khai.edu/course/view.php?id=1302>)

14. Рекомендована література

Базова

1. Руденко В. М. Математична статистика. Навч. посіб. – К.: Центр учбової літератури, 2012. – 304 с.
2. Кармелюк Г. Теорія ймовірності та математична статистика Посібник з розв'язування задач. – К.: Центр навчальної літератури, 2017, 576 с.
3. Алілуйко А. П. Практикум з теорії ймовірностей та математичної статистики / Алілуйко А. М., Дзюбановська Н. В., Єрмоменко В. О., Мартинюк О. М., Шинкарик М. І. — Тернопіль, 2018. — 352 с.

Допоміжна

1. Бобик, О. І., Гладунський В.Н. Довідник з теорії ймовірностей і математичної статистики : Навч. посібник для студентів ВНЗ– Львів : ЛБІ НБУ, 2005.– 143 с.
2. Теорія ймовірностей та математична статистика : [наук. зб.] / Київ. нац. ун-т ім. Тараса Шевченка. – - Виходить щоквартально. - ISSN 0868-6904. - ISSN 1547-7363
3. Статистика: навч. посіб. / Р. В. Фещур, В. П. Кічор, А. Ф. Барвінський, М. Р. Тимошук . – 4-те вид., оновл. і доповн. – Л. : Бух. центр "Ажур", 2010. – 256 с.
4. Чернобай О. Б. Про деякі методичні особливості викладання курсу теорії ймовірностей : дис. – Ірпінь: Університет ДФС України., 2020.

15. Інформаційні ресурси

1. Теорія ймовірностей та математична статистика у прикладах і задачах : навч. посіб. для студ. вищ. навч. закл. / О.Б. Жильцов ; за ред. Г.О. Михаліна. — К. : Київ. ун-т ім. Б. Грінченка, 2015. — 336 с. http://elibrary.kubg.edu.ua/id/eprint/13578/1/O_Zhylytsov_KUBG_TY_UN.pdf
2. Теорія ймовірностей і математична статистика : навчальний посібник / В. М. Астахов, Г. С. Буланов, В. О. Паламарчук. – Краматорськ : ДДМА, 2009. – 64 с. <http://www.dgma.donetsk.ua/metod/vm/tims.pdf>
3. Сайт науково-технічної бібліотеки університету [Електронний ресурс]. – Режим доступу: <http://library.khai.edu>.