

Міністерство освіти і науки України
Національний університет водного господарства та природокористування
Навчально-науковий інститут автоматики, кібернетики
та обчислювальної техніки
Кафедра прикладної математики

"ЗАТВЕРДЖУЮ"

Проректор з науково-педагогічної,
методичної та виховної роботи

_____ О.А. Лагоднюк

" ____ " _____ 2020 р.

04-01-79

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
Аналіз даних та математична статистика

PROGRAM OF THE DISCIPLINE
Data Analysis and Mathematical Statistics

спеціальність 113 «Прикладна математика»

specialty 113 «Applied mathematics»

Робоча програма навчальної дисципліни «Аналіз даних та математична статистика» для здобувачів вищої освіти першого (бакалаврського) рівня, які навчаються за освітньо-професійною програмою «Прикладна математика» спеціальності 113 «Прикладна математика». Рівне: НУВГП, 2019. 12 с.

Розробник: О.П. Остапчук, кандидат технічних наук, доцент кафедри прикладної математики.

Робочу програму схвалено на засіданні кафедри прикладної математики

Протокол від « 20 » грудня 2019 року № 6

Завідувач кафедри _____(П. М. Мартинюк)

Керівник групи забезпечення спеціальності
113 «Прикладна математика» _____(О.В. Прищепя)

Схвалено науково-методичною радою з якості навчально-наукового інституту автоматизації, кібернетики та обчислювальної техніки

Протокол від « 24 » грудня 2019 року № 4

Голова науково-методичної ради з якості _____(П.О.Тадєєв)

© Остапчук О.П., 2019
© НУВГП, 2019

ВСТУП

Робоча програма навчальної дисципліни «Аналіз даних та математична статистика» є складовою частиною нормативно-методичного забезпечення навчального процесу за галуззю знань 11 «Математика і статистика». Програма складена відповідно до стандарту освіти з підготовки здобувачів вищої освіти першого (бакалаврського) рівня за спеціальністю 113 «Прикладна математика».

Предметом вивчення навчальної дисципліни є основні методи аналізу статистичних даних, які найчастіше застосовуються у наукових та практичних дослідженнях.

Міждисциплінарні зв'язки: дисципліна «Аналіз даних та математична статистика» є складовою частиною циклу фундаментальних дисциплін. Вивчення курсу передбачає наявність систематичних та ґрунтових знань із дисциплін «Математичний аналіз», «Теорія ймовірностей», «Прикладне програмне забезпечення», «Програмування».

Вимоги до знань та умінь визначаються галузевими стандартами вищої освіти України.

Анотація

Математичні методи аналізу даних широко використовуються при дослідженні різноманітних систем і процесів – природних, технічних, екологічних, економічних, соціальних тощо. З огляду на це формування відповідних знань та навичок є необхідною складовою підготовки фахівців зі спеціальностей прикладна математика, комп'ютерні науки, економічна кібернетика, інформатика, та багатьох інших.

Під час вивчення даної дисципліни передбачається розглянути основні типи даних, що підлягають аналізу; методи побудови описової статистики й емпіричних функцій розподілу; критерії перевірки статистичних гіпотез щодо однорідності вибірок та порівняння емпіричних функцій розподілу з теоретичними моделями; критерії та методи перевірки наявності статистичного зв'язку між ознаками; теоретичні основи та основні методи кореляційного, регресійного, дисперсійного аналізу; методи класифікації даних.

Ключові слова: випадкова величина, гіпотеза, генеральна сукупність, вибірка, кореляція, регресія, дисперсійний аналіз, коваріаційний аналіз, кластерний аналіз.

Abstract

Mathematical methods of data analysis are widely used in the study of various systems and processes - natural, technical, environmental, economic, social and the like. Given this, the formation of relevant knowledge and skills is a necessary component of the training of specialists in the field of applied mathematics, computer science, economic cybernetics, computer science, and many others.

When studying this discipline, it is supposed to consider the main types of data to be analyzed; methods for constructing descriptive statistics and empirical distribution

functions; criteria for testing statistical hypotheses regarding sample uniformity and comparing empirical distribution functions with theoretical models; criteria and methods for checking the presence of a statistical relationship between features; theoretical foundations and basic methods of correlation, regression, analysis of variance; data classification methods.

Keywords: random variable, hypothesis, universal set, sampling, correlation, regression, analysis of variance, analysis of covariance, cluster analysis.

1. Опис предмета навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, рівень вищої освіти	Характеристика навчальної дисципліни
		денна форма навчання
Кількість кредитів відповідних ECTS – 5	Галузь знань 11 «Математика та статистика»	Цикл навчальних дисциплін фахової підготовки
Модулів – 2	Спеціальність 113 «Прикладна математика»	Рік підготовки:
Змістових модулів – 2		3-й
Загальна кількість годин – 150		Семестри
		5-й
Тижневих годин для денної форми навчання: аудиторних – 3,33 СРС – 5	Рівень вищої освіти: бакалавр	Лекції
		26 год.
		Практичні, семінарські
		-
		Лабораторні
		26 год.
		Самостійна робота
		98 год.
Індивідуальні завдання:		
-		
Вид контролю:		
іспит		

Примітка.

Співвідношення кількості аудиторних занять, самостійної та індивідуальної роботи студентів в процентах до загальної кількості годин складає відповідно:
для денної форми навчання – 35% до 65%.

2. Мета навчальної дисципліни

Мета курсу - вивчити основні методи аналізу статистичних даних, які найчастіше застосовуються у наукових дослідженнях та на практиці.

Завдання: сформувати у студентів теоретичні і практичні навички застосування методів аналізу даних при обробці даних експерименту за допомогою комп'ютерних програм аналізу даних.

У результаті вивчення даного курсу здобувачі вищої освіти повинні **знати:** статистичні методи та їх теоретичні основи, класичні і сучасні методи статистичного аналізу даних.

При вивченні дисципліни здобувачі вищої освіти повинні **вміти:**

- самостійно ставити задачу дослідження та забезпечувати її розв'язання;
- вибирати основні методи та алгоритми обробки даних;
- реалізувати дані методи, застосовуючи сучасні пакети прикладних програм та системи програмування для аналізу даних;
- аналізувати результати досліджень та забезпечувати їх застосування;
- адекватно інтерпретувати отримані результати;
- формулювати висновки та рекомендації.

Згідно з програмою курсу здобувач вищої освіти повинен виконувати лабораторні роботи, завдання із самостійної роботи.

3. Програма навчальної дисципліни

МОДУЛЬ 1

ЗМІСТОВИЙ МОДУЛЬ 1

Етапи та методи попередньої обробки даних

Тема 1. Вступ. Основні задачі аналізу даних.

Етапи аналізу даних. Класифікація типів змінних. Кількісні, ординарні та номінальні дані та робота з ними. Залежні та незалежні змінні, зв'язок між ними. Групування даних. Описові статистики для згрупованих даних.

Тема 2. Попередня обробка даних.

Етапи попередньої обробки даних. Числові характеристики вибірки. Перевірка статистичних гіпотез. Перевірка стохастичності вибірки. Критерій серій на базі медіани вибірки, критерій зростаючих та спадаючих серій., критерій Аббе. Видалення аномальних спостережень у скалярних та векторних вибірках. Розвідувальний аналіз. Методи візуалізації даних. Графічні експрес-методи аналізу даних. Перевірка статистичних гіпотез.

МОДУЛЬ 2

ЗМІСТОВИЙ МОДУЛЬ 2 Основні методи аналізу даних

Тема 3. Кореляційний аналіз.

Суть кореляційного аналізу. Характеристики лінійного парного та множинного статистичного зв'язку кількісних даних: індекс кореляції, коефіцієнт детермінації, кореляційне відношення, множинний коефіцієнт кореляції і т.д. Частинні коефіцієнти кореляції. Аналіз рангових кореляцій. Ранг, об'єднаний ранг. Метод середнього рангу. Таблиці рангів. Парні рангові коефіцієнти кореляції: Спірмена, Кендела. Індекси Фехнера. Кореляційне відношення Пірсона. Коефіцієнт конкордації. Кореляційний аналіз номінальних даних. Оцінювання та перевірка на значимість характеристик статистичного зв'язку.

Тема 4. Регресійний аналіз.

Поняття регресії. Основні задачі регресійного аналізу. Класична постановка задачі. Послідовність проведення регресійного аналізу. Лінійна регресія. Діаграма розсіювання. Метод частинних середніх. Проста лінійна регресія. Метод найменших квадратів. Визначення параметрів при парному лінійному регресійному аналізі. Алгоритм ручного рахунку. Лінеаризація у випадку криволінійних залежностей. Множинна лінійна регресія. Перевірка на значимість параметрів моделі. Перевірка гіпотез у лінійній регресії. Розв'язок задач у випадку порушення класичних припущень (мультиколінеарність, неоднорідність, та корельованість похибок і т.д.). Вибір оптимальної регресійної моделі.

Тема 5. Дисперсійний аналіз.

Моделі дисперсійного аналізу. Однофакторний дисперсійний аналіз. Контрасти та довірчі інтервали для них. Багатофакторний дисперсійний аналіз. Визначення впливів головних ефектів та взаємодій. Перехресна класифікація з повтореннями. Ієрархічна класифікація.

Тема 6. Коваріаційний аналіз.

Поняття коваріації. Коваріаційний аналіз. Основна модель та її зв'язок з моделями регресійного та дисперсійного аналізу. Двокроковий метод найменших квадратів.

Тема 7. Класифікація і кластерний аналіз.

Поняття класифікації. Основні постановки задач класифікації. Основні поняття кластерного аналізу. Застосування кластерного аналізу. Типи відстаней і міри подібностей. Дискримінантний аналіз. Навчаюча вибірка. Дискримінантні функції.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин			
	денна форма			
	Всього	лекції	лаб.	с.р.
1	2	3	4	5
Модуль 1				
Змістовий модуль 1				
Етапи та методи попередньої обробки даних				
Тема 1. Вступ. Основні задачі аналізу даних.	12	4	2	6
Тема 2. Попередня обробка даних.	22	4	4	14
Разом за змістовим модулем 1	34	8	6	20
Модуль 2				
Змістовий модуль 2				
Основні методи аналізу даних				
Тема 3. Кореляційний аналіз.	22	4	4	14
Тема 4. Регресійний аналіз.	22	4	4	14
Тема 5. Дисперсійний аналіз.	22	4	4	14
Тема 6. Коваріаційний аналіз.	24	2	4	18
Тема 7. Класифікація і кластерний аналіз.	26	4	4	18
Разом за змістовим модулем 2	116	18	20	78
Всього годин	150	26	26	98

5. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1.	Попередня обробка даних. Визначення числових характеристик варіаційного ряду.	2
2.	Перевірка статистичних гіпотез.	4
3.	Кореляційний аналіз. Визначення показників зв'язку при парній лінійній та криволінійній залежності, коефіцієнтів множинної лінійної та частинної кореляції.	4
4.	Лінійний та нелінійний регресійний аналіз.	4
5.	Однофакторний та двофакторний дисперсійний аналіз.	4
6.	Коваріаційний аналіз.	4
7.	Кластерний аналіз.	4
	Всього	26

6. Завдання до самостійної роботи

Розподіл годин самостійної роботи для студентів включає наступні пункти:

1. Підготовка до аудиторних занять (0,5 год. на 1 год. аудиторних занять) – 26 год.
2. Підготовка до контрольних заходів (6 год. на 5 кредит) – 30 год.
3. Опрацювання окремих тем програми або їх частин, які не викладаються на лекціях – 42 год.

6.1. Завдання для самостійної роботи

№ з/п	Назва теми	Кількість годин
1.	Непараметричні тести при перевірці гіпотез.	10
2.	Поліноміальні моделі в регресійному аналізі.	10
3.	Двокроковий метод найменших квадратів.	12
4.	Класифікація з навчанням.	10
Разом		42

6.2. Оформлення звіту про самостійну роботу

Підсумком самостійної роботи над вивченням дисципліни „Аналіз даних та математична статистика” є складання конспекту за темами, вказаними у п.6.1. Загальний обсяг конспекту визначається з умови повноти та якості викладеного матеріалу.

Конспект оформлюється на стандартному папері формату А4 або в зошиті. Конспект може бути рукописним або друкованим і виконується українською мовою. Перевірка конспекту з самостійної роботи відбувається у терміни, спільно обумовлені студентом і викладачем.

7. Методи навчання

1. Лекції проводяться з використанням технічних засобів навчання і супроводжуються демонстрацією презентацій за допомогою відеопроєктора або мультимедійної дошки .
2. Лабораторні роботи проводяться в комп'ютерному класі з використанням роздаткового матеріалу та методичних вказівок. Використовуються для розв'язання оригінальні задачі з метою набуття здобувачами вищої освіти навичок аналізувати та узагальнювати отриману інформацію.
3. Проведення контрольних тестувань.

8. Методи контролю

Для визначення рівня засвоєння здобувачами вищої освіти навчального матеріалу використовуються такі методи оцінювання знань:

- поточне тестування після вивчення кожного змістового модуля;
- розробка і захист звіту про виконання самостійної роботи;
- оцінка поточного модульного контролю;
- оцінка підсумкового контролю (екзамен).

Основними критеріями, що характеризують рівень **компетентності** здобувачів вищої освіти при оцінюванні результатів поточного та підсумкового контролів є:

- виконання всіх видів навчальної роботи, що передбачені робочою програмою навчальної дисципліни;
- вміння аналізувати процеси, що вивчаються;
- характер відповіді на поставлені питання (чіткість, лаконічність, логічність, послідовність тощо);
- вміння застосовувати теоретичні положення під час розв'язання практичних завдань;

Оцінювання результатів поточної роботи (завдань, що виконуються на практичних заняттях та консультаціях, результати самостійної роботи здобувачів вищої освіти) проводиться за такими **критеріями** (у % від кількості балів, виділених на завдання із заокругленням до цілого числа):

0% - завдання не виконано;

40% - завдання виконано частково, висновки не аргументовані і не конкретні, звіт підготовлено недбало;

60% - завдання виконано повністю, висновки містять окремі недоліки, судження здобувача вищої освіти не достатньо аргументовані, звіт підготовлено з незначним відхиленням від вимог;

80% - завдання виконано повністю і вчасно, проте містить окремі несуттєві недоліки не системного характеру;

100% - завдання виконано правильно, вчасно і без зауважень.

Контроль самостійної роботи проводиться за підготовленими звітами згідно зазначених тем.

9.1. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота							Іспит	Сума
Змістовий модуль 1			Змістовий модуль 2				40	100
18			42					
T1	T2	T3	T4	T5	T6	T7		
6	10	10	8	10	8	8		

Всього поточна складова оцінювання		60
Підсумкова складова оцінювання	Модульний контроль №1	20
	Модульний контроль №2	20
Всього підсумкова складова оцінювання		40
Разом		100

T1, T2,...,T7 – теми змістових модулів

9.2. Шкала оцінювання

У випадку підсумкового контролю у формі заліку сума набраних балів шкалою оприлюднюються до початку екзаменаційної сесії у електронному журналі академічної групи. У екзаменаційній відомості результати навчання проставляються за двома шкалами – 100-бальною та національною. Позитивні оцінки виставляються тільки тим студентам, які виконали всі види навчальної роботи, передбачені робочою програмою навчальної дисципліни, і набрали за результатами поточного та підсумкового контролів не менше 60 балів.

Конвертація 100-бальної шкали у 4-бальну здійснюється за наступною таблицею.

Сума балів за всі види навчальної діяльності	Оцінка за національною шкалою
	для екзамену
90 – 100	відмінно
82-89	добре
74-81	
64-73	задовільно
60-63	
35-59	незадовільно з можливістю повторного складання
0-34	незадовільно з обов'язковим повторним вивченням дисципліни

10. Методичне забезпечення

Методичне забезпечення навчальної дисципліни „Аналіз даних та математична статистика” включає:

1. Грицюк П.М., Остапчук О.П. Аналіз даних: навчальний посібник. Рівне: НУВГП, 2008. – 218 с. [Електронний ресурс]. – Режим доступу: <http://ep3.nuwm.edu.ua/14267/>
2. Остапчук О.П. Методичні вказівки для виконання лабораторних робіт з дисципліни „ Аналіз даних та математична статистика ” **для здобувачів** вищої освіти першого (бакалаврського) рівня за спеціальністю 113 "Прикладна математика" денної форми навчання (електронний варіант).

3. Опорний конспект лекцій (електронний і паперовий варіанти) за всіма темами курсу.
4. Пакети тестових завдань для всього курсу дисципліни (навчальна платформа Moodle). [Електронний ресурс]. – Режим доступу: <https://exam.nuwm.edu.ua/course/view.php?id=2983>

11. Рекомендована література

Базова

1. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: основы моделирования и первичная обработка данных. Москва: Финансы и статистика, 1983. 471 с.
2. Себер Дж. Линейный регрессионный анализ. Москва: Мир, 1980. 456 с.
3. Шеффе Г. Дисперсионный анализ. Изд. 2-е. Москва: Наука, 1980. 512 с.

Допоміжна

1. Бахрушин В.С. Аналіз даних: навчальний посібник. Запоріжжя: ГУ “ЗІДМУ”, 2006. 128 с.
2. Боровиков В.П. Популярное введение в программу STATISTIKA. Москва: Комп’ютер-Пресс, 1998. 268 с.
3. Гутер Р.С., Овчинский Б. В. Элементы численного анализа и математической обработки данных эксперимента. Москва. 1971. 192 с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика Москва: Высшая школа, 2003. 479 с.
5. Жлуктенко В.І., Наконечний С.І., Савіна С.С. Теорія ймовірностей і математична статистика. Київ: КНЕУ, 2001. 336 с.
6. Кендалл М. Дж., Стюарт А. Многомерный анализ. Москва: Наука, 1975. 511 с.
7. Классификация и кластер /Дж. Вен Райзман. Москва: Мир, 1980. 389 с.
8. Лук’янова В.В. Комп’ютерний аналіз даних. Київ: Академія, 2003. 344 с.
9. Протасов К.В. Статистический анализ экспериментальных данных Москва: Мир, 2005. – 142 с.

12. Інформаційні ресурси

1. Обласна наукова бібліотека (м. Рівне, майдан Короленка, / [Електронний ресурс]. – Режим доступу: <http://www.lib.rv.ua>
2. Рівненська централізована бібліотечна система (м. Рівне, вул. Київська, 44) / [Електронний ресурс]. – Режим доступу: <http://www.cbs.rv.ua>
3. Наукова бібліотека НУВГП (м. Рівне, вул. О.Новака, 75) (інформаційні ресурси в цифровому репозиторії) / [Електронний ресурс]. – Режим доступу: <http://nuwm.edu.ua/naukova-biblioteka>
4. Сайт циклу дисциплін “Аналіз даних”, “Математичні методи аналізу даних”, “Комп’ютеризовані технології аналізу даних”. [Електронний ресурс]. – Режим доступу: <http://datan.ucoz.ru>
5. <http://www.google.com.ua>