

Чернівецький національний університет імені Юрія Федьковича
(повне найменування вищого навчального закладу)

Факультет математики та інформатики

(назва інституту / факультету)

Кафедра математичного моделювання

(назва кафедри)

СИЛАБУС навчальної дисципліни

Аналіз даних

(назва навчальної дисципліни)

вибіркова

(вказати: обов'язкова / вибіркова)

Освітньо-професійна програма Інформаційні технології та управління проектами,
Системний аналіз

(назва програми)

Спеціальність 122 – Комп'ютерні науки, 124 – Системний аналіз

(вказати: код, назва)

Галузь знань 12 – Інформаційні технології

(вказати: шифр, назва)

Рівень вищої освіти перший (бакалаврський)

(вказати: перший бакалаврський/другий магістерський)

Факультет математики та інформатики

(назва факультету/інституту, на якому здійснюється підготовка фахівців за вказаною освітньо-професійною програмою)

Мова навчання українська

(вказати: на якій мові читається дисципліна)

Розробники: Черевко Ігор Михайлович, професор каф. мат. мод., доктор. фіз.-мат. наук,
професор

(П.І.Б. авторів, посада, науковий ступінь, вчене звання)

Профайл викладача: <http://matmod.fmi.org.ua/pro-kafedru/spivrobitnyky/cherevko-igor-myhailovych/>

Контактний тел.: (0372) 58-48-25

E-mail: i.cherevko@chnu.edu.ua

Сторінка курсу в Moodle: <http://e-learning.fpm.chnu.edu.ua/course/view.php?id=25>

Консультації: Зазначте формат і розклад проведення консультацій
Онлайн-консультації: среда з 14-20-15-20.
Очні консультації: за попередньою домовленістю.

1. Анотація дисципліни (призначення навчальної дисципліни).

Це є базовий курс для майбутніх аналітиків. Завдяки отриманим знанням вони зможуть проводити первинний статистичний аналіз даних, висувати гіпотези та оцінювати їх на основі зібраних даних. Прикладний аналіз даних служить базою прийняття управлінських рішень. На сучасному етапі його значимість значно зросла у зв'язку із величезним обсягом інформації, який необхідно проаналізувати. Такі обсяги інформації важко обробляти без спеціалізованих пакетів, зокрема найпопулярніших Pandas та R, що передбачають володіння відповідними програмними засобами та навичками використання спеціалізованих модулів та бібліотек.

Перш ніж застосовувати методи аналізу, дані повинні бути підготовані та попередньо проаналізовані. На подальшому етапі з'являється можливість перевірки статистичних гіпотез щодо подібності розподілів та їх характеристик як за параметричними, так і за непараметричними критеріями. І нарешті на заключних етапах з'являється можливість побудови прогнозів поведінки характеристик досліджуваних об'єктів в залежності від прогнозованих оцінок входних параметрів. Аналіз та застосування згаданих підходів і складає основний зміст даного курсу.

2. Мета навчальної дисципліни. Задачі аналізу емпіричних даних займають центральне місце при проведенні експериментальних досліджень в будь-якій області знань. Аналіз даних складає алгоритмічну основу автоматизованих систем (САПР, АСУП, АСУТП та ін.). Метою та завданнями дисципліни «Аналіз даних» є систематичне вивчення задач, що пов'язані із етапами обробки даних, побудовою та оцінкою математичних моделей експериментальних даних, застосування сучасних ППП обробки статистичних даних на комп'ютері.

Дисципліна формує такі компетентності за ОП

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

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

ЗК11. Здатність приймати обґрунтовані рішення.

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.

СК6. Здатність до системного мислення, застосування методології системного аналізу для дослідження складних проблем різної природи, методів формалізації та розв'язування системних задач, що мають суперечливі цілі, невизначеності та ризику.

3. Пререквізити. Обчислювальні методи, Програмування, Системи та методи прийняття рішень, Теорія ймовірностей та математична статистика, Диференціальні рівняння, Аналіз даних та статистичне виведення на мові R (онлайн-курси платформи **Prometheus**).

4. Результати навчання

Завдання дисципліни. Знання і досвід, набуті в цьому курсі, будуть корисними в майбутній практичній діяльності студентів при моделюванні на ЄОМ різноманітних прикладних процесів та аналізі експериментальних даних.

У результаті вивчення навчальної дисципліни студент повинен

знати:

- основні теоретичні і практичні принципи, форми, методи, прийоми статистичної обробки чисельних даних, зокрема результатів експерименту із застосуванням ЄОМ;

вміти:

- застосовувати статистичні методи для розв'язання задач практичного змісту при моделюванні на ЄОМ різноманітних прикладних процесів та аналізі експериментальних даних.

Наведені результати навчання за відповідною дисципліною співвідносяться із такими **програмними результатами навчання:**

ПР1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

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.

ПР8. Використовувати методологію системного аналізу об'єктів, процесів і систем для задач аналізу, прогнозування, управління та проектування динамічних процесів в макроекономічних, технічних, технологічних і фінансових об'єктах.

5. Опис навчальної дисципліни

5.1. Загальна інформація

Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин						Вид підсумкового контролю
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота	індивідуальні завдання	
Денна	4	8	3	90	22	–	–	22	46	–	екзамен
Заочна											

5.2. Дидактична карта навчальної дисципліни

Назви змістових модулів і тем	Кількість годин												
	денна форма						Заочна форма						
	усього	у тому числі					усього	у тому числі					
		л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13	
Змістовий модуль 1. Попередня обробка даних													
Тема 1. Задачі попередньої обробки даних	6	2				4							
Тема 2. Моделювання псевдо-випадкових чисел.	10	2		2		6							
Тема 3. Емпіричний розподіл даних.	8	2		2		4							
Тема 4. Дослідження законів розподілу експериментальних даних.	14	4		4		6							
Разом за ЗМ 1	38	10		8		20							
Змістовий модуль 2. Аналіз експериментальних даних та статистичні моделі													
Тема 1. Лінійні регресійні моделі.	10	2		4		4							
Тема 2. Нелінійні багатофакторні регресійні моделі.	10	2		2		6							

Тема 3. Планування експерименту в задачах ідентифікації об'єктів.	6	2			4						
Тема 4. Аналіз даних в системі "Statistica". Опис, структура та головні функціональні модулі системи Statistica.	14	4		4	6						
Тема 5. Статистичний аналіз часових рядів в системі Statistica	12	2		4	6						
Разом за ЗМ 2	52	12		14	26						
Усього годин	90	22		22	46						

5.3. Зміст завдань для самостійної роботи

№	Назва теми
1	Особливості побудови програмного забезпечення для статистичної обробки даних. Типи статистичних пакетів і загальні вимоги до них.
2	Методика обчислення числових характеристик експериментальних даних. Додаткові відомості із теорії ймовірностей та математичної статистики. Оцінка числових характеристик та параметрів випадкових величин та процесів.
3	Регресійний аналіз простіших поліноміальних моделей. Особливості регресійного аналізу при порушенні базових припущень. Визначення показників зв'язку при криволінійній залежності.
4	Методи перевірки ортогональності планів. Побудова неповних факторних планів, генератор планів. Схеми проведення імітаційного експерименту

6. Система контролю та оцінювання

Види та форми контролю

1. Поточний (усне опитування, розв'язування задач)
2. Модульний (контрольні роботи)
3. Підсумковий (екзамен)

Засоби оцінювання:

- контрольні роботи;
- командні проекти;
- аналітичні звіти про виконання індивідуальних завдань та самостійної роботи.

Критерії оцінювання результатів навчання з навчальної дисципліни

Білет містить три питання, з яких одне питання теоретичне і два практичних.

1. Повна відповідь на кожне питання оцінюється 10 балами.
2. За кожну помилку, яка допущена у відповіді, знімається певна кількість балів, а саме:
 - а) при відповіді на теоретичне питання у випадку неістотної помилки знімається 1-2 бали, а у випадку істотної 3-5 балів, якщо ж студент не опанував теоретичний матеріал дисципліни, плутається в означеннях, наводить логічно невірні твердження, то знімається до 10 балів;

б) при оцінці практичного завдання за помилку, допущену при перетвореннях, знімається 1-2 бали; за істотну помилку, яка привела до неправильної відповіді, знімається 3-5 балів; якщо ж розв'язання задачі логічно неправильне, то знімається до 10 балів.

3. Максимальна кількість, яку можна набрати на підсумковому модулі (екзамені) 30 балів.

4. Підсумкова оцінка виставляється за результатами суми балів набраних на змістовних модулях під час семестру та підсумковому модулі (екзамені) згідно таблиці

Шкала оцінювання: національна та ЄКТС

Оцінка за національною шкалою	Оцінка за шкалою ECTS	
	Оцінка (бали)	Пояснення за розширеною шкалою
Відмінно	A (90-100)	відмінно
Добре	B (80-89)	дуже добре
	C (70-79)	добре
Задовільно	D (60-69)	задовільно
	E (50-59)	достатньо
Незадовільно	FX (35-49)	(незадовільно) з можливістю повторного складання
	F (1-34)	(незадовільно) з обов'язковим повторним курсом

7. Рекомендована література Основна

1. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных.– М.: Финансы и статистика, 1983.– 471 с.
2. Андерсен Т. Статистический анализ временных рядов.– М.: Мир, 1976.– 756 с.
3. Бендат Дж., Пирсол А. Прикладной анализ случайных данных.– М.: Мир, 1989.– 540 с.
4. Вучков И., Бояджијева Л., Солаков Е. Прикладной линейный регрессионный анализ.– М.: Финансы и статистика, 1987.– 239 с.
5. Каханер Д., Моулер К., Неш С. Численные методы и программное обеспечение.– М.: Мир, 1998.– 575 с.
6. Костылев А. А. и др. Статистическая обработка результатов эксперимента на Микро-ЭВМ.– М.: Финансы и статистика, 1991.– 305 с.
7. Мармоза А. Т. Практикум по математической статистике.– К.: 1990.– 191 с.
8. Полард Дж. Справочник по вычислительным методам статистики.– М.: Финансы и статистика, 1982.– 344 с.
9. Тюрин Ю. Н., Макаров А. А. Статистический анализ данных на компьютере.– М.: ИНФРА-М, 1998.– 528 с.
10. Тюрин Ю. Н., Макаров А. А., Фигурнов Н. Р. Анализ данных на компьютере: Учебник.– М.: Инфра -М, 2003.– 544 с.
11. Лесна Н.С. Багатовимірний аналіз статистичних даних: Навч. посібник.– Харків: ХНУРЕ, 2004.– 120 с.
12. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів.– Запоріжжя : КПУ, 2011.– 268 с.

Додаткова

1. Боровиков В.П. Популярное введение в программу STATISTICA.– М.: КомпьютерПресс, 1998.– 267 с.
2. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде WINDOWS.– М.: Финансы и статистика, 1999.– 384 с.
3. Глинский В.В., Ионин В.Г. Статистический анализ.– М.: ИИД «Филинь», 1998. 264 с.

4. Джонстон Дж. Эконометрические методы. Пер. с англ.– М.: Статистика, 1980, с. 322–341.
5. Кенделл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. Пер. с англ.– М.: Наука, 1976.– 474 с.
6. Афифи А., Эйзен С. Статистический анализ.- М.: Мир, 1982.– 488 с.
7. Аналіз даних – Prometheus.– <https://prometheus.org.ua> › dataanalysis

8. Інформаційні ресурси

<http://e-learning.fpm.chnu.edu.ua/course/view.php?id=25>