

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій

Затверджую

В. о. декана факультету комп'ютерних
інформаційних технологій



Ігор Якименко

" " 2023р.

Затверджую

В.о. проректора з науково-педагогічної
роботи



Віктор ОСТРОВЕРХОВ

" " 2023 р..

Затверджую

Директор ІНЦІОТ



Святослав Питель

" " 2023р.

РОБОЧА ПРОГРАМА

з дисципліни

«Системний аналіз економічних процесів»

Ступінь вищої освіти – другий (магістерський)

Галузь знань: 12 Інформаційні технології

Спеціальність: 124 Системний аналіз

Освітньо-професійна програма «Системний аналіз»

Кафедра економічної кібернетики та інформатики

| Форма навчання | Курс | Семестр | Лекції | Практ. | ІРС | Тре- нінг, КПЗ | СРС | Разом | Залік |
|----------------|------|---------|--------|--------|-----|----------------------|-----|-------|-------|
| Денна | I | II | 30 | 15 | 5 | 4 | 96 | 150 | II |
| Заочна | I | II | 8 | 4 | - | - | 138 | 150 | II |

Тернопіль 2023

Робочу програму склав: професор кафедри економічної кібернетики та інформатики
ПАСІЧНИК Роман Мирославович

Робоча програма затверджена на засіданні кафедри економічної кібернетики та інформатики, протокол № 1 від 28.08.2023 р.

Завідувач кафедри



проф. БУЯК Леся Михайлівна

Розглянуто та схвалено групою забезпечення спеціальності системний аналіз, протокол №1 від 30.08.2023 р.

Голова ГЗС



проф. ПАСІЧНИК Роман Мирославович

Гарант ОПШ



доц. БАБАЛА Людмила Василівна

СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Системний аналіз економічних процесів»

1. Опис дисципліни «Системний аналіз економічних процесів»

| Дисципліна – Системний аналіз економічних процесів | Галузь знань, напрям підготовки, освітньо- кваліфікаційний рівень | Характеристика навчальної дисципліни |
|--|---|---|
| Кількість кредитів ECTS 4 | Галузь знань – 12 «Інформаційні технології» | Вибіркова дисципліна, мова навчання - <i>українська</i> |
| Кількість залікових модулів - 3 | Спеціальність – 124 «Системний аналіз», | <i>Денна:</i> Рік підготовки:1 Семестр – 2 <i>Заочна:</i> Рік підготовки:1 Семестр – 2,3 |
| Кількість змістових модулів - 2 | Ступінь вищої освіти – магістр | <i>Денна:</i> лекції – 30 год.; лабор.- 15 год <i>Заочна:</i> лекції – 8 год.; лабор.- 4 год |
| Загальна кількість годин - - 150 | | Самостійна робота: 96 год., (тренінг(КПЗ) – 4 год.) Індивідуальна робота : 5 год. |
| Тижневих годин: 10 год., з них аудиторних – 3 год | | Вид підсумкового контролю – <i>залік</i> |

2. Мета й завдання вивчення дисципліни "Системний аналіз економічних процесів"

2.1. Мета вивчення дисципліни

Метою викладання дисципліни "Системний аналіз економічних процесів" є ознайомлення студентів з методологією підтримки прийняття рішень на основі статистичних методів із застосуванням їх для розв'язання прикладних задач. .

2.2. Завдання вивчення дисципліни

В результаті вивчення курсу "Системний аналіз економічних процесів" студенти повинні:

- **знати** основні поняття попереднього аналізу вибірок, базові структури даних середовища Python, пакетів Pandas та R, метод ієрархічного кластерного аналізу, метод к-середніх, основні статистичні гіпотези, методи відсіву вибірок, методи перевірки однорідності вибірок, основи лінійної та нелінійної регресій, основи прогнозування часових рядів;

- **вміти** здійснювати попередній аналіз вибірок, реалізовувати кластеризацію вибірок, перевіряти гіпотези однорідності, корельованості, нормальності вибірок, будувати регресійні залежності прогнозувати часові ряди.

3. Програма дисципліни " Системний аналіз економічних процесів "

Змістовий модуль 1 – Попередній аналіз та класифікація підвбірок

Тема 1. Моделювання соціально-економічних процесів. Збалансована система показників. Загальна імітаційна модель економічних об'єктів та її складові. Найпростіші моделі нерегулярних процесів. Часові ряди. Параметризовані моделі. Регресійні моделі та диференціальні рівняння.

Тема 2. Прикладні пакети аналізу даних.

Основні завдання та інструменти аналізу даних. R та Python переваги, недоліки, рекомендовані сфери застосування. Середовище IDLE розроблення програм на Python. Пакет статистичного аналізу Pandas. Базові структури даних в Python списки, кортежі словники. Формування структур та вибір елементів із них. Методи сортування словників по ключах та значеннях. Пакет R та середовище аналітика RStudio

Тема 3. Попередній аналіз даних із пакетом Pandas

Генеральна сукупність та вибірка. Варіаційний ряд. Гістограма. Ядрова оцінка щільності розподілу. Коробкова діаграма. Її елементи та принцип побудови. Порівняння розподілів та обсягів кількох вибірок. Проведення попереднього аналізу за допомогою пакету Pandas.

Тема 4. Ієрархічний кластерний аналіз

Віддаль між кластерами. Центроїдний метод та його недоліки. Методи далекого та найближчого сусіда. Віддаль Соренсона. Рекомендовані методи попереднього кластерного аналізу. Метод покрокового агрегування кластерів. Дендрограми та алгоритми їх побудови. Критерій зупинки. Насипи (лікті) для масштабних кластеризацій.

Відбір змінних кластеризації. Сурогатні змінні. Роль стандартизації змінних. Методи стандартизації. Стандартизація за розмахом та за допомогою z-міток. Інтерпретація результатів кластерного аналізу. Обмеженість на розмірності в методі ієрархічного кластерного аналізу. Приклади використання кластерного аналізу. Процедура NBClust.

Тема 5. Метод k-середніх.

Етапи побудови моделі. Експертне задання центрів кластерів. Ініціалізація датчиків випадкових чисел. Обмеження кількості змін центрів кластерів. Кількість проведення процедур кластеризації. Вибір найкращої кластеризації за критерієм якості. Допустимість мінімальної зміни центрів кластерів. Метод голосування кількості кластерів. Середнє значення показників по кластеру. Інтерпретація типових представників кластеру.

Змістовий модуль 2 – Перевірка статистичних гіпотез та регресійний аналіз

Тема 6. Основні статистичні гіпотези.

Гіпотеза погодження. Гіпотеза нормальності для економії кількості вимірювань. Гіпотеза експоненціальності для часу безвідмовної роботи. Гіпотези однорідності про рівність розподілів після виконання певних дій на покращення. Гіпотези незалежності. Гіпотеза про параметр розподілу. (л5)
Помилки першого та другого роду. Рівень значимості. Критичні значення. Статистичний критерій. Перевірка статистичних гіпотез про вигляд розподілу. Критерій Шапіро

Тема 7. Порівняння характеристик вибірок

Логарифмування вибірок. Зменшення зернистості гістограм. Відсів викидів. Візуальний критерій нормальності розподілу. Гіпотези рівності дисперсій. Критерій Флінгена. Критерій рівності математичних сподівань Стюдента. Випадок парних вибірок. Перевірка однорідності вибірок. А-В тестування. Перевірка корельованості вибірок. Критерій χ^2 квадрат. Тести порівняння відносних частот у вибірок. Р-тестування. Критерій Манна-Уїтні порівняння медіан..

Тема 8. Лінійна регресія

Перевірка гіпотези про лінійну залежність вибірок. Види залежностей, параметри. Критерій якості. Підбір невідомих параметрів. Коефіцієнт детермінації як міра якості прогнозу. Аналіз залишків. Проблема колінеарності. Пошук колінеарних залежностей. Методи заповнення пропусків в даних. Усунення змінних, що породжують колінеарність. Перевірка гіпотези про рівність нулю коефіцієнта при неінформативних змінних. Покроковий метод зменшення моделі.

Найпростіша нелінійна модель регресії. Бібліотека PolynomialFeatures. Побудова множини моделей. Аналіз коефіцієнтів детермінації. Аналіз значимостей та знаків коефіцієнтів моделі. Відбір адекватних моделей.

Тема 9. Аналіз часових рядів.

Встановлення пакетів та ініціалізація бібліотек. Візуалізація та інтеграція даних. Аналіз діаграми розподілу спостережень. Тест Харки — Бера нормальності розподілу результуючого показника. Узагальнений тест Діккі-Фуллера нестационарності часового ряду. Пошук стаціонарних різницевого ряду. Перевірка стаціонарності. Основні особливості моделей Бокса-Дженкінса та ARIMA.

Тема 10. Прогнозування часових рядів

Дослідження автокореляційної та частково автокореляційної функцій. Підбір параметрів моделі ARIMA. Аналіз залишків моделі. Q-тест Льюнга — Бокса. Підбір параметрів моделі на основі критерію Акаїки. Створення та візуалізація прогнозів. Оцінки похибок прогнозів.

1. Структура залікового кредиту дисципліни "Системний аналіз економічних процесів"

денна

| | Кількість годин | | | | |
|---|-----------------|-------------------|-------------------|----------------|----------------|
| | Лекції | Практичні заняття | Самостійна робота | Індивід робота | Контр. заходи |
| Змістовий модуль 1 – Попередній аналіз та класифікація підвибірок | | | | | |
| Тема 1. Моделювання соціально-економічних процесів | 3 | 2 | 10 | | поточне опит. |
| Тема 2. Прикладні пакети аналізу даних | 3 | 1 | 10 | | поточне опит. |
| Тема 3. Попередній аналіз даних із пакетом Pandas | 3 | 2 | 10 | 1 | поточне опит. |
| Тема 4. Ієрархічний кластерний аналіз | 3 | 1 | 10 | 1 | поточне опит. |
| Тема 5. Метод к-середніх | 3 | 2 | 10 | | модульн контр |
| Змістовий модуль 2 – Перевірка статистичних гіпотез та регресійний аналіз | | | | | |
| Тема 6. Основні статистичні гіпотези. | 3 | 1 | 10 | 1 | поточне опит. |
| Тема 7. Порівняння характеристик вибірок | 3 | 2 | 9 | | поточне опит. |
| Тема 8. Лінійна регресія | 3 | 1 | 9 | 1 | поточне опит. |
| Тема 9. Аналіз часових рядів | 3 | 2 | 9 | | поточне опит. |
| Тема 10. Прогнозування часових рядів | 3 | 1 | 9 | 1 | ректорс. контр |
| Разом | 30 | 15 | 96 | 5 | |

Заочна форма

| | Кількість годин | | |
|---|-----------------|-------------------|-------------------|
| | Лекції | Практичні заняття | Самостійна робота |
| Тема 1. Моделювання економічних процесів | 1 | | 13 |
| Тема 2. Прикладні пакети аналізу даних | | 1 | 13 |
| Тема 3. Попередній аналіз даних із пакетом Pandas | 1 | | 14 |
| Тема 4. Ієрархічний кластерний аналіз | 1 | 1 | 14 |
| Тема 5. Метод к-середніх | 1 | | 14 |
| Тема 6. Основні статистичні гіпотези. | 1 | | 14 |
| Тема 7. Порівняння характеристик вибірок | | | 14 |
| Тема 8. Лінійна регресія | 1 | 1 | 14 |
| Тема 9. Аналіз часових рядів | 1 | 1 | 14 |
| Тема 10. Прогнозування | 1 | | 14 |

| | | | |
|---------------|---|---|-----|
| часових рядів | | | |
| Разом | 8 | 4 | 138 |

5. Тематика практичних занять

Практичне заняття 1. Моделювання економічних процесів

1. Збалансована система показників.
2. Загальна імітаційна модель економічних об'єктів та її складові.
3. Найпростіші моделі нерегулярних процесів. Часові ряди.
4. Параметризовані моделі. Регресійні моделі та диференціальні рівняння.

Практичне заняття 2. Прикладні пакети аналізу даних.

- 1 R та Python переваги, недоліки, рекомендовані сфери застосування.
- 2 Середовище IDLE розроблення програм на Python.
2. Пакет статистичного аналізу Pandas. Методи.
3. Пакет R та середовище аналітика RStudio

Практичне заняття 3. Попередній аналіз даних із пакетом Pandas

- 1 Варіаційний ряд. Гістограма.
- 2 Ядрова оцінка щільності розподілу.
- 3 Порівняння розподілів та обсягів кількох вибірок.
4. Проведення попереднього аналізу за допомогою пакету Pandas.

Практичне заняття 4. Ієрархічний кластерний аналіз

1. Центроїдний метод та його недоліки.
2. Віддаль Соренсона.
3. Метод покрокового агрегування кластерів.
4. Насипи (ліктя) для масштабних кластеризацій.

Практичне заняття 5. Метод k-середніх.

1. Експертне задання центрів кластерів.
2. Кількість проведення процедур кластеризації.
3. Вибір найкращої кластеризації за критерієм якості.
4. Метод голосування кількості кластерів.

Практичне заняття 6. Основні статистичні гіпотези.

1. Гіпотеза нормальності для економії кількості вимірювань.
2. Гіпотеза експоненціальності для часу безвідмовної роботи.
3. Гіпотези однорідності про рівність розподілів після виконання певних дій на покращення.
4. Гіпотеза про параметр розподілу.

Практичне заняття 7. Порівняння характеристик вибірок

1. Зменшення зернистості гістограм.
2. Візуальний критерій нормальності розподілу.
3. Гіпотези рівності дисперсій.
4. Критерій рівності математичних сподівань Стьюдента.
5. Критерій Манна-Уїтні порівняння медіан..

Практичне заняття 8. Лінійна регресія

- 1.Перевірка гіпотези про лінійну залежність вибірок.
- 2.Коефіцієнт детермінації як міра якості прогнозу.
- 3.Пошук колінеарних залежностей.
4. Усунення змінних, що породжують колінеарність.
- 5.Перевірка гіпотези про рівність нулю коефіцієнта при неінформативних змінних.

Практичне заняття 9. Аналіз часових рядів.

- 1, Візуалізація та інтеграція даних.
2. Аналіз діаграми розподілу спостережень. Тест Харки — Бера нормальності розподілу результуючого показника.
3. Узагальнений тест Діккі-Фуллера нестационарності часового ряду.
- 4.Основні особливості моделей Бокса-Дженкінса та ARIMA.

Практичне заняття 10. Прогнозування часових рядів

- 1.Дослідження автокореляційної та частково автокореляційної функцій.
- 2.Підбір параметрів моделі ARIMA.
3. Аналіз залишків моделі.
4. Оцінки похибок прогнозів.

6. Комплексне практичне індивідуальне завдання.

1. Попередній аналіз вибірок за допомогою статистичних пакетів.
2. Ієрархічний кластерний аналіз
3. Метод к-середніх.
4. Основні статистичні гіпотези
5. Порівняння характеристик вибірок
6. Лінійна регресія
7. Аналіз часових рядів
8. Прогнозування часових рядів

7. Самостійна робота

| № п/п | Тематика |
|-------|--|
| 1. | Основні завдання та інструменти аналізу даних. R та Python переваги, недоліки, рекомендовані сфери застосування. |
| 2. | Середовище IDLE розроблення програм на Python. |
| 3. | Пакет статистичного аналізу Pandas. |
| 4 | Базові структури даних в Python списки, кортежі словники. |
| 5 | Методи сортування словників по ключах та значеннях. Пакет R та середовище аналітика RStudio. |
| 6 | Методи роботи із типом даних Series, його індексами та значеннями. |
| 7 | Групові вибірки та присвоювання. Переіндексування та фільтрування. Тип даних DataFrame. |

| | |
|----|---|
| 8 | Закон та функція розподілу випадкової величини. |
| 9 | Щільність розподілу випадкової величини. |
| 10 | Рівномірно та нормально розподілені випадкові величини |
| 11 | Математичне сподівання, дисперсія, стандартне відхилення. Логарифмічно нормальні розподіли. |
| 12 | Генеральна сукупність та вибірка. Варіаційний ряд. Гістограма. Ядрова оцінка щільності розподілу. Проведення попереднього аналізу за допомогою пакету Pandas. |
| 13 | Задача ієрархічного кластерного аналізу. Метод к-середніх. Методи агрегації даних шляхом об'єднання спостережень в групи. |
| 14 | Віддаль між кластерами. Центроїдний метод та його недоліки. Методи далекого та найближчого сусіда. Віддаль Соренсона |
| 15 | Рекомендовані методи попереднього кластерного аналізу. Насипи (лікті) для масштабних кластеризацій. |
| 16 | Етапи побудови моделі. Експертне задання центрів кластерів. Метод голосування кількості кластерів. |
| 17 | Гіпотеза про параметри розподілу. Перевірка статистичних гіпотез про вигляд розподілу. Критерій Шапіро |
| 18 | Критерій рівності математичних сподівань Стьюдента. Випадок парних вибірок. Критерій Манна-Уїтні порівняння медіан. |
| 19 | Покроковий метод зменшення моделі. Найпростіша нелінійна модель регресії. Бібліотека PolynomialFeatures. Побудова множини моделей. Аналіз коефіцієнтів детермінації. Аналіз значимостей та знаків коефіцієнтів моделі. Відбір адекватних моделей. |
| 20 | Сфери застосування лінійної регресії. Випадок множинної сезонності. Аналіз коротких сезонних часових рядів. |
| 21 | Основні особливості моделей Бокса-Дженкінса та ARIMA. |
| 22 | Підбір параметрів моделі ARIMA. Аналіз залишків моделі. Q-тест Льюнга — Бокса. Підбір параметрів моделі на основі критерію Акаїки. Створення та візуалізація прогнозів. Оцінки похибок прогнозів. |

8. Засоби оцінювання та методи демонстрування результатів навчання

- У процесі вивчення дисципліни «Системний аналіз економічних процесів» використовуються наступні засоби оцінювання та методи демонстрування результатів навчання:

- поточне опитування;
- залікове модульне тестування та опитування;
- аналітичні звіти, реферати, есе;
- оцінювання результатів КПІЗ;
- розрахункові роботи;
- ректорська контрольна робота;
- залік;

9. Критерії, форми поточного та підсумкового контролю

Підсумковий бал (за 100-бальною шкалою) з дисципліни "Системний аналіз економічних процесів" визначається як середньозважена величина, в залежності від питомої ваги кожної складової залікового кредиту:

| Заліковий модуль 1 | Заліковий модуль 2 | Заліковий модуль 3 | Разом |
|---|--|--|-------|
| 30% | 40% | 30% | 100% |
| 1. Усне опитування під час заняття (5 теми по 10 балів = 50 балів) 2. Письмова робота = 50 балів | 1. Усне опитування під час заняття (5 тем по 10 балів = 50 балів) 2. Письмова робота = 50 балів | 1. Написання та захист КПЗ = 60 балів. 3. Виконання завдань під час тренінгу = 40 балів | |

10. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна

| № | Найменування | Номер теми |
|----|-----------------------------|------------|
| 1. | Дошка, крейда | 1-10 |
| 2. | Персональний комп'ютер | 1-10 |
| 3. | Програмне середовище Python | 1-10 |
| 4. | Прикладні пакети R, Pandas | 1-10 |

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

1. Wes McKinney. Python for Data Analysis. O'Reilly Media, 2013. <https://bedford-computing.co.uk/learning/wp-content/uploads/2015/10/Python-for-Data-Analysis.pdf>
2. Йорн Гіз. Підручник із ієрархічної кластеризації та дендрограм SciPy. <https://joernhees.de/blog/2015/08/26/scipy-hierarchical-clustering-and-dendrogram-tutorial/>
3. Marco Peixeiro. Time Series Forecasting in Python. https://www.methsoft.ac.cn/scipaper_files/document_files/Manning.Time.Series.Forecasting.in.Python.pdf
4. 8host.com. Прогнозування часових рядів за допомогою ARIMA в PYTHON 3. <https://www.8host.com/blog/prognozirovanie-vremennyx-ryadov-s-pomoshhyu-arima-v-python-3/>
5. Joel Grus. Data Science from Scratch. O'Reilly Media. 2019. <https://covid19.uthm.edu.my/wp-content/uploads/2020/04/Data-Science-from->

Scratch-First-Principles-with-Python-by-Joel-Grus-z-lib.org_.epub_.pdf

6. Dirk P. Kroese, Zdravko I. Botev, Thomas Taimre, Radislav Vaisman. Data Science and Machine Learning Mathematical and Statistical Methods. 2022. <https://people.smp.uq.edu.au/DirkKroese/DSML/DSML.pdf>.

7. Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda. Applied Text Analysis with Python. O'Reilly Media. 2018. <https://www.oreilly.com/library/view/applied-text-analysis/9781491963036/>