

Чернівецький національний університет імені Юрія Федьковича

Факультет математики та інформатики

Кафедра математичного моделювання

СИЛАБУС
навчальної дисципліни

Прикладний статистичний аналіз з використанням Python
вбіркова

Освітньо-професійна програма Системний аналіз

Спеціальність 124 – Системний аналіз

Галузь знань 12 – Інформаційні технології

Рівень вищої освіти перший (бакалаврський)

Факультет математики та інформатики

Мова навчання українська

Розробник: Юрченко Ігор Валерійович,
доцент кафедри математичного моделювання,
кандидат фіз.-мат. наук, доцент

Профайл викладача <http://matmod.fmi.org.ua/pro-kafedru/spivrobotnyky/yurchenko-igor-valeriyovich/>

Контактний тел. 0372-58-48-25

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

Сторінка курсу в Moodle

Консультації Онлайн-консультації: п'ятниця з 13:00 до 14:20.
Очні консультації: за попередньою домовленістю.

1. Анотація дисципліни (призначення навчальної дисципліни).

Навчальна дисципліна призначена для ознайомлення студентів з основами кореляційного, дисперсійного, регресійного, коваріаційного, факторного статистичного аналізу та їх застосування для розв'язання соціально-економічних задач з використанням комп'ютерної техніки.

2. Мета навчальної дисципліни: ознайомити студентів із теоретичними і практичними питаннями застосування методів факторного, канонічного аналізу, видобування даних в соціально-економічних дослідженнях з використанням комп'ютерних технологій (бібліотеки мови Python).

3. Пререквізити. Навчальні дисципліни: “Програмування”, “Теорія ймовірностей і математична статистика”, “Бібліотеки мови Python”.

4. Результати навчання. У результаті вивчення навчальної дисципліни студенти повинні

знати основні теоретичні і практичні питання кореляційного, дисперсійного, регресійного, коваріаційного, факторного статистичного аналізу,

вміти моделювати на ПК статистичний розподіл вибірки, оцінювати параметри розподілу, перевіряти статистичні гіпотези; будувати за допомогою ПК моделі різних видів статистичного аналізу; використовувати сучасне програмне забезпечення (Python) для проведення статистичних розрахунків та розв'язання прикладних соціально-економічних задач (згідно зі стандартом вищої освіти [8]):

- *спеціальні компетентності* К26 – здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них;
- *результати навчання* ПР12 – застосовувати методи і засоби роботи з даними і знаннями, методи математичного, логіко-семантичного, об'єктного та імітаційного моделювання, технології системного і статистичного аналізу; ПР13 – проектувати, реалізовувати, тестувати, впроваджувати, супроводжувати, експлуатувати програмні засоби роботи з даними і знаннями в комп'ютерних системах і мережах).

5. Опис навчальної дисципліни

5.1. Загальна інформація

Назва навчальної дисципліни: “Прикладний статистичний аналіз з використанням Python”											
Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин						Вид підсумкового контролю
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота	індивідуальні завдання	
Денна	4	7	3	90	30	–	–	30	30	–	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.													
НЕ 1.1 (лекція). Основні поняття вибіркового методу. Огляд бібліотек Python для статистичного аналізу.	3	2				1	–	–	–	–	–	–	
НЕ 1.2 (лекція). Прикладні задачі точкового та інтервального оцінювання.	5	4				1	–	–	–	–	–	–	
НЕ 1.3 (лекція). Прикладні задачі перевірки статистичних гіпотез.	5	4				1	–	–	–	–	–	–	
НЕ 1.4 (лекція). Застосування кореляційного та регресійного	6	4				2	–	–	–	–	–	–	

аналізу в економіці												
НЕ 1.5 (лабораторне заняття). Огляд модулів SciPy, Matplotlib, NumPy мови Python	3			2		1	-	-	-	-	-	-
НЕ 1.6 (лабораторне заняття). Побудова гістограми та щільності розподілу засобами Matplotlib	6			4		2	-	-	-	-	-	-
НЕ 1.7 (лабораторне заняття). Точкове та інтервальне оцінювання, його візуалізація в Python	7			5		2	-	-	-	-	-	-
Разом за змістовим модулем 1	35	14		11		10	-	-	-	-	-	-
Змістовий модуль 2.												
НЕ 2.1 (лекція). Однофакторний дисперсійний аналіз в економіці.	3	2				1	-	-	-	-	-	-
НЕ 2.2 (лекція). Двофакторний дисперсійний аналіз в економіці.	3	2				1						
НЕ 2.3 (лекція). Поняття про коваріаційний аналіз.	3	2				1						
НЕ 2.4 (лекція). Загальні моделі багатовимірного аналізу (Big Data). Модель і властивості головних компонент.	4	3				1						

HE 2.5 (лекція). Факторний аналіз задач великої розмірності.	4	3			1						
HE 2.6 (лекція). Часові ряди в соціально- економічних задачах.	5	4			1						
HE 2.7 (лабораторне заняття). Кореляційний аналіз. Використання модуля SciPy.	6			4	2	–	–	–	–	–	–
HE 2.8 (лабораторне заняття). Регресійний аналіз, його візуалізація в Python.	9			5	4	–	–	–	–	–	–
HE 2.9 (лабораторне заняття). Перевірка статистичних гіпотез засобами модуля SciPy. Однофакторний дисперсійний аналіз.	9			5	4	–	–	–	–	–	–
HE 2.10 (лабораторне заняття). Двофакторний дисперсійний аналіз засобами Python.	9			5	4	–	–	–	–	–	–
Разом за змістовим модулем 2	55	16		19	20	–	–	–	–	–	–
Всього годин	90	30		30	30	–	–	–	–	–	–

5.3. Зміст лабораторних занять

№	Навчальний елемент	Зміст	Год.
1	Огляд модулів SciPy, Matplotlib, NumPy мови Python	Методи для роботи з розподілами випадкових величин, які зосереджені в модулі SciPy. Візуалізація даних засобами Matplotlib. Наукові обчислення в NumPy.	2

2	Побудова гістограми та щільності розподілу засобами Matplotlib	Побудова гістограми за набором даних (вибіркою) та порівняння її з графіком щільності нормального розподілу	4
3	Точкове та інтервальне оцінювання, його візуалізація в Python	Для трьох наборів випадковим чином згенерованих даних побудувати графіки з даними про середнє, медіану, квартилі, діапазони та викиди у формі boxplot()	5
4	Кореляційний аналіз. Використання модуля SciPy.	Обчислити коефіцієнт кореляції Пірсона для двох наборів даних на Python з використанням власної функції. Перевірити отриманий результат з використанням функції stats.pearsonr() модуля SciPy	4
5	Регресійний аналіз, його візуалізація в Python.	За згенерованими випадковим чином вихідними даними побудувати лінію регресії, обчислити коефіцієнт детермінації та довірчий інтервал для коефіцієнтів рівняння прямої регресії.	5
6	Перевірка статистичних гіпотез засобами модуля SciPy. Однофакторний дисперсійний аналіз.	Методами дисперсійного аналізу (ANOVA) при рівні значущості 0.05 перевірити нульову гіпотезу про рівність групових середніх. Також перевірити гіпотези про початкові припущення (вибірки зроблені з нормальних сукупностей з однаковими дисперсіями).	5
7	Двофакторний дисперсійний аналіз засобами Python.	Провести двофакторний дисперсійний аналіз. Перевірити гіпотезу про рівність дисперсій у групах. Для рівня значущості 0.05 перевірити гіпотезу про вплив факторів А та В та їхньої комбінації на результат синтезу	5

Лабораторні завдання викладені в повному обсязі в [3].

5.4. Зміст завдань для самостійної роботи

Самостійна робота складається з повторення матеріалу, засвоєного на лекціях, самостійного опанування частини теоретичного матеріалу, роботи з контрольними запитаннями та завданнями.

Студенти можуть отримати до 10 балів в рахунок ІНДЗ, якщо самостійно зареєструються на безкоштовних курсах платформи Prometheus або на аналогічних курсах платформи Coursera (за узгодженням з викладачем), пройдуть навчання, отримають відповідний сертифікат і надішлють його на сайт дистанційного навчання викладачу разом зі скріншотом успішності на курсі. Кількість балів буде виставлена пропорційно до навчальних результатів студента (згідно зі статистикою сайту Prometheus чи Coursera).

Студенти можуть також в рахунок ІНДЗ:

- опрацювати модельні приклади з візуалізації даних засобами пакету Matplotlib мови Python з навчального посібника [5, с.262-341] та оформити звіт за результатами роботи;

- опрацювати модельні приклади роботи з даними засобами пакету Pandas мови Python з навчального посібника [5, с.129-219]; оформити звіт за результатами роботи.

Залежно від кількості опрацьованих прикладів ІНДЗ оцінюється від 1 до 15 балів.

6. Система контролю та оцінювання

Види та форми контролю

Формами поточного контролю є усна чи письмова (тестування, лабораторна робота, ІНДЗ) відповідь студента.

Формою підсумкового контролю є тестування.

Засоби оцінювання

Усний контроль у вигляді індивідуального та фронтального опитування на лекціях та лабораторних заняттях, захист лабораторних робіт та індивідуального навчально-дослідницького завдання; письмовий контроль у вигляді контрольних робіт, тестів, підсумкове тестове опитування.

Критерії оцінювання результатів навчання з навчальної дисципліни

Критерієм успішного проходження здобувачем освіти підсумкового оцінювання є досягнення ним мінімальних порогових рівнів оцінок за кожним запланованим результатом навчання навчальної дисципліни.

Мінімальний пороговий рівень оцінки визначається за допомогою якісних критеріїв і трансформується в мінімальну позитивну оцінку використовуваної числової (рейтингової) шкали.

Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота							Модуль- контроль	Сума
Змістовий модуль №1			Змістовий модуль № 2					
HE	HE	HE	HE	HE	HE	HE		
1.5	1.6	1.7	2.7	2.8	2.9	2.10		
5	7	8	10	10	10	10	40	100

Шкала оцінювання: національна та ЄКТС

Оцінка за національною шкалою	Оцінка за шкалою ЄКТС	
	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

1. Прикладна статистика: Методичні вказівки до лабораторних робіт / Укл.: І.В.Юрченко.– Чернівці: Рута, 2000.– 55 с.
2. Ясинський В.К., Юрченко І.В. Прикладний статистичний аналіз. Методичні рекомендації до лабораторних робіт.– Чернівці: Рута, 2008.– 84 с.
3. Юрченко І.В. Прикладний статистичний аналіз з використанням Python.– Чернівці: Технодрук, 2021.– 102 с.
4. Доля П.Г. Введение в научный Python.– Харків: Харківський національний університет ім. В.Н. Каразіна, 2016.– 265 с.
5. Плас Дж.Вандер. Python для сложных задач: наука о данных и машинное обучение.– СПб.: Питер, 2018.– 576 с.
6. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика.– М.: ЮНИТИ-ДАНА, 2001.– 656 с.
7. Королюк В.С. Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике.– К.: Наук. думка, 1978.– 582 с.
8. Стандарт вищої освіти України: перший (бакалаврський) рівень, галузь знань 12 – Інформаційні технології, спеціальність – 124 Системний аналіз. Затверджено і введено в дію наказом Міністерства освіти і науки України від 13.11.2018 № 1245. <https://mon.gov.ua/storage/app/media/vishcha-osvita/zatverdzeni%20standarty/12/21/124-sistemniy-analiz-bakalavr.pdf>

8. Інформаційні ресурси

<http://moodle.chnu.edu.ua>

<http://www.python.org>

<http://www.matplotlib.org>

<http://www.scipy.org>