

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 Інформаційні технології та управління проектами

Спеціальність 122 – Комп'ютерні науки

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.

Дисципліна формує такі **компетенції** у відповідності до освітньої програми:

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.

ЗК6. Здатність вчитися й оволодівати сучасними знаннями.

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

ЗК12. Здатність оцінювати та забезпечувати якість виконуваних робіт.

ФК3. 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.

ФК7. Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів.

ФК11. 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.

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

ПРН1. 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.

ПРН3. 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.

ПРН5. Проектувати, розробляти та аналізувати алгоритми розв'язання обчислювальних та логічних задач, оцінювати ефективність та складність алгоритмів на основі застосування формальних моделей алгоритмів та обчислюваних функцій.

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.													
HE 1.1 (лекція). Основні поняття вибіркового методу. Огляд бібліотек Python для статистичного аналізу.	3	2				1	–	–	–	–	–	–	–

HE 1.2 (лекція). Прикладні задачі точкового та інтервального оцінювання.	5	4				1	-	-	-	-	-	-
HE 1.3 (лекція). Прикладні задачі перевірки статистичних гіпотез.	5	4				1	-	-	-	-	-	-
HE 1.4 (лекція). Застосування кореляційного та регресійного аналізу в економіці	6	4				2	-	-	-	-	-	-
HE 1.5 (лабораторне заняття). Огляд модулів SciPy, Matplotlib, NumPy мови Python	3			2		1	-	-	-	-	-	-
HE 1.6 (лабораторне заняття). Побудова гістограми та щільності розподілу засобами Matplotlib	6			4		2	-	-	-	-	-	-
HE 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.												
HE 2.1 (лекція). Однофакторний дисперсійний аналіз в економіці.	3	2				1	-	-	-	-	-	-
HE 2.2 (лекція).	3	2				1						

Двофакторний дисперсійний аналіз в економіці.												
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						
НЕ 2.5 (лекція). Факторний аналіз задач великої розмірності.	4	3				1						
НЕ 2.6 (лекція). Часові ряди в соціально-економічних задачах.	5	4				1						
НЕ 2.7 (лабораторне заняття). Кореляційний аналіз. Використання модуля SciPy.	6			4		2	-	-	-	-	-	-
НЕ 2.8 (лабораторне заняття). Регресійний аналіз, його візуалізація в Python.	9			5		4	-	-	-	-	-	-
НЕ 2.9 (лабораторне заняття). Перевірка статистичних гіпотез засобами модуля SciPy. Однофакторний дисперсійний аналіз.	9			5		4	-	-	-	-	-	-
НЕ 2.10	9			5		4	-	-	-	-	-	-

(лабораторне заняття). Двофакторний дисперсійний аналіз засобами Python.											
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

Система оцінювання рівня навчальних досягнень ґрунтується на принципах ECTS та є накопичувальною. Протягом семестру студенти виконують 7 лабораторних робіт. Кожна лабораторна робота оцінюється кількістю балів, наведеною в таблиці нижче.

Виконуючи завдання лабораторної роботи, студент повинен оформити і завантажити для подальшої перевірки на сайт електронного навчання звіт разом із працездатними файлами програмної реалізації завдань ЛР (правила оформлення наведені на сторінці навчальної дисципліни на сайті).

50% балів, відведених на оцінювання ЛР, студент отримує за працюючий програмний продукт, в якому реалізовано всі завдання ЛР та оформлений звіт. Решта 50% балів виставляється після захисту студентом виконаного звіту. На захисті звіту з ЛР студент має відповісти на питання щодо постановки задачі та розробленого ним алгоритму реалізації кожного із завдань ЛР. При відповіді на

теоретичні питання та питання щодо програмної реалізації алгоритму у випадку неістотної помилки знімається 10-20% балів, а у випадку істотної 20-40% балів, якщо ж студент не опанував теоретичний матеріал, плутається в означеннях, наводить логічно невірні твердження, то знімається до 50% балів від усієї суми балів за ЛР.

Максимальна кількість, яку можна набрати на підсумковому модулі (тестування) – 40 балів.

Підсумкова оцінка виставляється за результатами суми балів, набраних на змістових модулях під час семестру та підсумковому модулі згідно з нижченаведеною таблицею.

Розподіл балів, які отримують студенти

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

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

1. Юрченко І.В. Прикладний статистичний аналіз з використанням Python. – Чернівці: Технодрук, 2021. – 102 с.
2. Jake VanderPlas. Python Data Science Handbook. Essential Tools for Working with Data. – Beijing, Boston, Farnham, Tokyo: O'Reilly Media, Inc, 2016. – 576 p. – ISBN: 9-781-491-912-058.
3. Доля П.Г. Вступ до наукового Python. – Харків: Харківський національний університет ім. В.Н. Каразіна, 2016. – 265 с.

4. Ясинський В.К., Юрченко І.В. Прикладний статистичний аналіз. Методичні рекомендації до лабораторних робіт.– Чернівці: Рута, 2008.– 84 с.
5. Мамчич Т.І., Оленко А.Я., Осипчук М.М., Шпортюк В.Г. Статистичний аналіз даних з пакетом Statistica. Навчально-методичний посібник.– Дрогобич: Видавнича фірма "Відродження", 2006.– 208 с.
6. Королюк В.С., Царков Є.Ф., Ясинський В.К. Ймовірність, статистика та випадкові процеси. Теорія та комп'ютерна практика. В 3-х томах. Т.1.: Ймовірність. Теорія та комп'ютерна практика.– Чернівці: Золоті литаври, 2007.– 444 с.
7. Королюк В.С., Царков Є.Ф., Ясинський В.К. Ймовірність, статистика та випадкові процеси. Теорія та комп'ютерна практика. В 3-х томах. Т.2.: Математична статистика. Комп'ютерне статистичне моделювання. – Чернівці: Золоті литаври, 2008. – 580 с.
8. Королюк В.С., Царков Є.Ф., Ясинський В.К. Ймовірність, статистика та випадкові процеси. Теорія та комп'ютерна практика. В 3-х томах. Т.3: Випадкові процеси. Теорія та комп'ютерна практика – Чернівці: Золоті литаври, 2009. – 782 с.
9. Юрченко І.В., Ясинський В.К., Ясинська Л.І. Курс комп'ютерного статистичного моделювання.– Чернівці: Вид-во "Прут", 2003.– 416 с.

8. Інформаційні ресурси

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

<http://www.python.org>

<http://www.matplotlib.org>

<http://www.scipy.org>

9. Політика освітнього процесу

Здобувач зобов'язаний своєчасно та якісно виконувати всі отримані завдання; за необхідністю з метою з'ясування всіх незрозумілих під час самостійної та індивідуальної роботи питань, відвідувати консультації викладача.

Студенти мають дотримуватись правил академічної доброчесності відповідно до "Кодексу академічної доброчесності ЧНУ". Політика дотримання академічної доброчесності (відповідно до Закону України "Про вищу освіту") – викладання навчальної дисципліни ґрунтується на засадах академічної доброчесності – сукупності етичних принципів та визначених законом правил, якими мають керуватися учасники освітнього процесу під час навчання, викладання та провадження наукової (творчої) діяльності з метою забезпечення

довіри до результатів навчання та/або наукових (творчих) досягнень. Наявність академічного плагіату в студентських роботах є підставою для виставлення негативної оцінки. Списування студентів під час проведення модульної контрольної роботи є підставою для дострокового припинення її складання та виставлення негативної оцінки.

Складання / перескладання екзаменів та заліків відбувається за встановленим деканатом розкладом.