КЛАСИЧНИЙ ПРИВАТНИЙ УНІВЕРСИТЕТ

Кафедра економіки та статистики

СИЛАБУС

навчальної дисципліни

ПП 4. Пакети аналізу даних у Python

2026-2027 навчальний рік

Викладач	Порохня Василь Михайлович
Контактний телефон	+38(061)764-67-50 (внутр. 169)
Ідентифікатор та код доступу (Zoom / Google Meet)	963 860 1816, код доступу 1349
E-mail:	vprhnp76@gmail.com
Сторінка курсу на сайті підтримки навчальних програм КПУ	http://www.zhu.edu.ua/cpu_edu/course/view.php?id=4682
Консультації	Консультації проводяться в дистанційному режимі за попереднім погодженням із викладачем через месенджери або електронну пошту

ІНФОРМАЦІЯ ПРО НАВЧАЛЬНУ ДИСЦИПЛІНУ

Рівень вищої освіти	Другий (магістерський)	
Галузь знань	Е Природничі науки, математика та статистика	
Спеціальність	Е8 Статистика	
Освітня програма	Комп'ютерна статистика та аналіз даних	
Форма навчання	Очна (денна), заочна	
Мова викладання	Українська	
Курс (рік навчання), семестр	Курс 1, семестр 1	
Цикл підготовки	Професійної підготовки	
Статус дисципліни	Обов’язкова	
Обсяг дисципліни	Кількість кредитів ECTS	5
	Загальна кількість годин	150
	Аудиторних годин	60
	Лекції	24
	Практичні	36
	Самостійна робота	90
Форма контролю	Екзамен	
Попередні компоненти (Пререквізити)	-	
Паралельні та подальші компоненти (Постреквізити)	Паралельні: ПП 1. Математико-статистичні методи і моделі в прийнятті управлінських рішень Подальші: ПП 6. Статистичні алгоритми навчання, П 1 Переддипломна практика, А 1 Підготовка та захист кваліфікаційної роботи магістра.	

МЕТА, ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета навчальної дисципліни

Формування здатності застосовувати пакети Python для підготовки, статистичного аналізу та візуалізації даних, обґрунтовувати вибір аналітичних процедур, перевіряти коректність обчислень та інтерпретувати результати прикладних досліджень

Завдання навчальної дисципліни:

- опанування основ роботи з даними в середовищі Python;
- набуття навичок імпорту даних із файлів різних форматів та реляційних баз даних;
- засвоєння способів перевірки якості, очищення й перетворення даних;
- опанування засобів описової статистики, перевірки статистичних гіпотез та аналізу зв'язків;
- набуття навичок застосування пакетів pandas, SciPy і statsmodels;
- формування вмінь виконувати регресійний аналіз, аналіз часових рядів та аналіз виживаності;
- розвиток навичок візуалізації, документування обчислень і представлення результатів аналізу.

У результаті вивчення дисципліни здобувачі вищої освіти повинні знати:

- основні типи й структури даних Python, призначення масивів NumPy та структур даних pandas;
- способи імпорту й експорту даних, особливості роботи з файлами CSV, XLSX, JSON, HTML та реляційними базами даних;
- способи виявлення пропусків, дублікатів, помилок типізації та аномальних значень;
- статистичні засади оцінювання параметрів і перевірки гіпотез;
- призначення, умови застосування та обмеження методів кореляційного, регресійного аналізу й аналізу якісних ознак;
- можливості пакетів pandas, SciPy і statsmodels;
- основні процедури аналізу часових рядів та даних про тривалість до настання події;
- принципи візуалізації результатів і забезпечення відтворюваності обчислень.

Після вивчення дисципліни здобувачі вищої освіти повинні вміти:

- визначати вимоги до даних відповідно до мети статистичного дослідження;
- завантажувати дані з різних джерел і перевіряти правильність їх зчитування;
- очищувати, перетворювати, об'єднувати та групувати дані;
- обчислювати описові статистики й будувати графічні подання даних;
- обирати та виконувати статистичні тести з урахуванням їхніх припущень;
- досліджувати зв'язки між кількісними та якісними ознаками;
- будувати регресійні моделі, перевіряти їхні основні припущення та пояснювати результати оцінювання;
- застосовувати базові процедури аналізу часових рядів та аналізу виживаності;
- перевіряти коректність програмної реалізації статистичних процедур;
- оформлювати програмний код, аналітичний звіт і висновки за результатами дослідження.

КОМПЕТЕНТНОСТІ ТА ПРОГРАМНІ РЕЗУЛЬТАТИ НАВЧАННЯ

Інтегральна компетентність

Здатність розв’язувати складні, зокрема міждисциплінарні, задачі та проблеми у сфері статистики й комп’ютерного аналізу даних шляхом інтегрування сучасних знань, проведення досліджень та/або інноваційної діяльності й ухвалення обґрунтованих рішень за умов невизначеності та неповної інформації.

Загальні компетентності

ЗК 1. Здатність до безперервного навчання, професійної рефлексії, самостійного оволодіння сучасними знаннями та адаптації до змін у професійній сфері.

Спеціальні (фахові) компетентності

СК 1. Здатність застосовувати поглиблені сучасні знання зі статистики та комп’ютерного аналізу даних для розв’язання складних професійних і дослідницьких задач.

СК 2. Здатність інтегрувати статистичні, економетричні та обчислювальні методи для дослідження складних міждисциплінарних проблем.

СК 4. Здатність ідентифікувати, критично аналізувати й формалізувати складні задачі у сфері статистики та аналізу даних, визначати вимоги до даних і обмеження розв’язання, зокрема за неповної інформації.

СК 7. Здатність добирати, адаптувати та програмно реалізовувати статистичні методи й обчислювальні алгоритми для розв’язання складних прикладних задач.

Програмні результати навчання

РН 1. Застосовувати поглиблені сучасні знання зі статистики та комп’ютерного аналізу даних для розв’язання складних професійних і дослідницьких задач.

РН 2. Формалізувати складні задачі статистичного дослідження та обґрунтовувати вибір методів їх розв’язання відповідно до мети дослідження й обмежень даних.

РН 5. Застосовувати статистичні методи для опрацювання, аналізу та інтерпретації даних під час розв’язання прикладних задач.

РН 14. Застосовувати програмні засоби для підготовки, перевірки якості, опрацювання, аналізу та візуалізації даних, програмно реалізовувати статистичні методи й обчислювальні алгоритми для розв’язання прикладних задач.

ЗМІСТ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

РОЗДІЛ 1. ОСНОВИ СТАТИСТИЧНОГО АНАЛІЗУ ДАНИХ ЗАСОБАМИ PYTHON

Тема 1. Вступ до аналізу даних у Python

Тема 2. Типи даних у Python. Імпорт даних із файлів форматів XLSX та CSV.

Тема 3. Імпорт даних із джерел у форматах JSON і HTML та з реляційних баз даних засобами SQL.

Тема 4. Математичні основи аналізу даних: теорія ймовірностей та математична статистика.

Тема 5. Перевірка статистичних гіпотез.

Тема 6. Аналіз зв’язків: кореляція та асоціація.

Тема 7. Опис якісних ознак.

Тема 8. Регресійний аналіз.

РОЗДІЛ 2. ЗАСТОСУВАННЯ SCIPY.STATS, PANDAS ТА STATSMODELS В АНАЛІЗІ ДАНИХ

Тема 9. Особливості застосування модуля scipy.stats в аналізі даних.

Тема 10. Особливості застосування пакета pandas в аналізі даних.

Тема 11. Особливості застосування пакета statsmodels: регресійний аналіз та аналіз часових рядів.

Тема 12. Особливості застосування пакета statsmodels: аналіз виживаності.

ФОРМИ КОНТРОЛЮ ТА ОЦІНЮВАННЯ

Контроль результатів навчання та рівня сформованості компетентностей здобувачів здійснюється на засадах прозорості та академічної доброчесності відповідно до Положення про критерії оцінювання результатів навчання здобувачів вищої освіти у Класичному приватному університеті. http://virtuni.education.zp.ua/info_cpu/node/5478

У процесі вивчення навчальної дисципліни використовуються такі види контролю:

1. **Поточний контроль:** здійснюється під час виконання практичних робіт, тематичного тестування та перевірки самостійної роботи. Оцінюються правильність підготовки даних, обґрунтованість застосування статистичних методів, коректність програмного коду, візуалізація та інтерпретація результатів.

2. **Семестровий (підсумковий) контроль:** проводиться у формі екзамену за матеріалом усіх тем навчальної дисципліни.

Кожна практична робота, самостійна робота, тематичний тест та екзаменаційна відповідь оцінюються за 100-бальною шкалою. Підсумкова оцінка визначається як середньозважена величина з урахуванням установлених вагових коефіцієнтів. Для тематичного тестування використовується середня арифметична оцінок за тематичні тести.

Підсумковий бал обчислюється за формулою:

$$\textbf{Підсумковий бал} = (B_1 \times k_1) + (B_2 \times k_2) + \ldots + (B_n \times k_n) + (B_{екз.} \times k_{екз.})$$

де:

$B_1, B_2, B_3, \ldots, B_n$ – оцінки за практичні роботи, контрольну роботу та тематичне тестування, визначені за 100-бальною шкалою; $k_1, k_2, k_3, \ldots, k_n$ – відповідні вагові коефіцієнти;

$B_{екз.}$ – оцінка за екзамен, визначена за 100-бальною шкалою; $k_{екз.}$ – ваговий коефіцієнт екзамену.

Сума вагових коефіцієнтів поточного та підсумкового контролю дорівнює 1

$$k_1 + k_2 + k_3 + \ldots + k_n + k_{екз.} = 1$$

Структура компонентів оцінювання та їх питома вага

Види навчальної діяльності здобувача	Шкала оцінювання завдання	Ваговий коеф.	Питома вага у загальній оцінці
Практична робота 1. Імпорт даних із файлів різних форматів та реляційної бази даних	100	0,10	10
Практична робота 2. Перевірка якості, очищення та перетворення даних засобами pandas	100	0,10	10
Практична робота 3. Описова статистика та перевірка статистичних гіпотез засобами scipy.stats	100	0,10	10
Практична робота 4. Кореляційний аналіз та дослідження асоціації між ознаками	100	0,05	5
Практична робота 5. Опис і візуалізація якісних ознак	100	0,05	5
Практична робота 6. Побудова та діагностика регресійної моделі засобами statsmodels	100	0,10	10
Практична робота 7. Аналіз часового ряду засобами pandas і statsmodels	100	0,05	5
Практична робота 8. Аналіз даних про тривалість до настання події засобами statsmodels	100	0,05	5
Контрольна робота	100	0,15	15
Тестування	100	0,05	5
Екзамен	100	0,20	20
Разом	**100**	1	**100**
Додаткові досягнення з навчальної дисципліни (за наявністю) Результати науково-дослідної діяльності здобувача, зокрема за такі активності: – участь у предметних олімпіадах, конкурсах студентських наукових робіт з навчальної дисципліни; – участь у засіданнях наукових гуртків, зокрема в обговореннях у форматі круглих столів, присвячених актуальним темам з навчальної дисципліни; – публікації результатів досліджень у вигляді наукових статей, тез доповідей, а також виступи на наукових конференціях.	**До 10 балів**		

КРИТЕРІЇ ОЦІНЮВАННЯ

100-бальна система	шкала ЄКТС	Національна шкала	Характеристика вимог та критеріїв оцінювання
90–100	A	Відмінно	Здобувач самостійно визначає послідовність аналізу, перевіряє якість даних, обґрунтовано застосовує пакети Python. Програмний код працює коректно, статистичні припущення перевірено. Результати належно візуалізовано, висновки аргументовано, обчислення можна відтворити.
82–89	B	Добре	Правильно готує дані та застосовує основні статистичні процедури. Код і висновки загалом коректні. Окремі неточності в оформленні, візуалізації або поясненні результатів не впливають на основні висновки.
75–81	C	Добре	Виконує типові завдання з використанням відповідних пакетів Python. Допускає недоліки в перевірці даних, обґрунтуванні методу або інтерпретації результатів, які може виправити після зауважень.
67–74	D	Задовільно	Виконує аналіз за наданим зразком. Код потребує окремих виправлень, вибір процедур обґрунтований неповно. Має труднощі з перевіркою припущень та поясненням результатів.
60–66	E	Задовільно	Демонструє мінімально достатні знання основних пакетів. За інструкцією завантажує дані та виконує базові обчислення, проте потребує допомоги під час виправлення помилок і формулювання висновків.
35–59	FX	Незадовільно	Допускає суттєві помилки в підготовці даних, програмному коді або застосуванні статистичних методів. Отримані результати не забезпечують обґрунтованих висновків. Потрібне повторне складання.
0–34	F	Незадовільно	Не володіє основними прийомами роботи з даними в Python, не може виконати базові обчислення та пояснити їхній зміст. Потрібне повторне вивчення дисципліни.

ШКАЛА ОЦІНЮВАННЯ: 100-БАЛЬНА, НАЦІОНАЛЬНА ТА ECTS

Сума балів за всі види навчальної діяльності	Оцінка за шкалою ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проєкту (роботи), практики	для заліку
90-100	A	відмінно	зараховано
82-89	B	добре	
75-81	C		
67-74	D	задовільно	
60-66	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	F	незадовільно з обов’язковим повторним вивченням дисципліни	не зараховано з обов’язковим повторним вивченням дисципліни

ПОЛІТИКА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Політика щодо термінів виконання завдань та перескладання	З огляду на безпекову ситуацію, повітряні тривоги та перебої з електропостачанням, освітній процес базується на принципах гнучкості та академічної підтримки. Усі завдання виконуються здобувачами в асинхронному режимі без застосування штрафних санкцій чи зниження балів за порушення первинних термінів, якщо затримка зумовлена об'єктивними обставинами непереборної сили. Здавання та перескладання пропущених завдань чи тем здійснюється у зручному для студента темпі протягом семестру за погодженням із викладачем без будь-яких обмежень.
Політика академічної доброчесності	Здобувачі освіти зобов'язані неухильно дотримуватись принципів академічної доброчесності під час виконання усіх видів навчальної, наукової та творчої діяльності. Порушеннями академічної доброчесності вважаються: академічний плагіат, самоплагіат, фабрикація, фальсифікація, списування, обман, хабарництво, необ'єктивне оцінювання та академічне шахрайство. Здобувачам освіти забороняється вдаватися до будь-яких форм академічного шахрайства, зокрема: використовувати під час контрольних заходів заборонені матеріали або технічні засоби; посилатися на джерела, які не використовувались у роботі; здавати роботи з однаковим змістом, підготовлені різними особами. У разі встановлення порушення академічної доброчесності до здобувача освіти можуть бути застосовані такі заходи академічної відповідальності, як повторне проходження оцінювання (контрольна робота, екзамен, залік тощо), повторне вивчення відповідного освітнього компонента або відрахування з університету.
Політика щодо відвідування занять в умовах воєнного стану	Відвідування навчальних занять здобувачами вищої освіти є обов'язковим відповідно до затвердженого розкладу. *Організація освітнього процесу в умовах загрози та повітряних тривог* У разі включення сигналу «Повітряна тривога» під час проведення занять у режимі реального часу (Zoom / Google Meet) або перебування в приміщеннях Університету, проведення синхронного заняття негайно припиняється, а всі учасники освітнього процесу зобов'язані перейти до укриттів. Навчальне заняття, перерване через сигнал «Повітряна тривога» або яке не відбулося через її тривалість, автоматично переводиться в асинхронний режим. Відсутність здобувача освіти на синхронній конференції під час повітряної тривоги, відсутності електроенергії або зв'язку не є підставою для виставлення заборгованості чи зниження підсумкової оцінки, за умови опрацювання навчального матеріалу та виконання завдань в асинхронному режимі. *Застосування дистанційних технологій та форс-мажорних обставин* У разі переходу університету або його окремих структурних підрозділів на навчання із використанням дистанційних технологій через безпекову ситуацію чи інші форс-мажорні обставини, освітній процес регулюється *«Положенням про організацію освітнього процесу здобувачів вищої освіти з використанням дистанційних технологій при обмежені або відсутності фізичного відвідування закладів освіти здобувачами з причин непереборної сили (природні катаклізми, заходи карантинного порядку та інші форс-мажорні обставини) в «Класичному приватному університеті».* Формат участі у дистанційних заняттях адаптується до поточних технічних можливостей та умов безпеки здобувачів і викладачів.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

Основна література

1. Юрченко І. В. Прикладний статистичний аналіз з використанням Python : навчальний посібник. Вид. 3-тє, допов. Чернівці : Чернівецький національний університет, 2021. 107 с.
2. Мірошниченко І., Хлевна Ю. Прикладний статистичний аналіз в Python. Частина І : електронний навчальний посібник. 2025. URL: https://aranaur.github.io/applied-statistics/index.html.
3. McKinney W. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter. 3rd ed. O'Reilly Media, 2022. URL: https://wesmckinney.com/book/.
4. Bruce P., Bruce A., Gedeck P. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. 2nd ed. O'Reilly Media, 2020. URL: https://www.oreilly.com/library/view/practical-statistics-for/9781492072935/.

Додаткова література

5. Куссуль Н. М., Шелестов А. Ю., Тарасенко С. А., Яйлимова Г. О. Аналіз даних. Лабораторний практикум : навчальний посібник. Київ : КПІ ім. Ігоря Сікорського, 2022. 28 с. URL: https://ela.kpi.ua/handle/123456789/50425.
6. Іванчук Я. В., Месюра В. І., Яровий А. А., Манжілевський О. Д. Інтелектуальний аналіз даних та машинне навчання. Частина 1. Базові методи та засоби аналізу даних : навчальний посібник. Вінниця : ВНТУ, 2021. 69 с. URL: https://ir.lib.vntu.edu.ua/handle/123456789/36454.
7. VanderPlas J. Python Data Science Handbook : електронне видання. URL: https://jakevdp.github.io/PythonDataScienceHandbook/.

ІНФОРМАЦІЙНІ РЕСУРСИ В ІНТЕРНЕТІ

1. Python. URL: https://docs.python.org/3/.
2. NumPy. URL: https://numpy.org/doc/stable/.
3. pandas. URL: https://pandas.pydata.org/docs/.
4. SciPy. Статистичні функції scipy.stats. URL: https://docs.scipy.org/doc/scipy/reference/stats.html.
5. statsmodels. URL: https://www.statsmodels.org/stable/.
6. Matplotlib. URL: https://matplotlib.org/stable/.
7. Seaborn. URL: https://seaborn.pydata.org/.
8. UCI Machine Learning Repository. URL: https://archive.ics.uci.edu/.
9. Kaggle Datasets. URL: https://www.kaggle.com/datasets.