

Код та назва дисципліни українською мовою/ Назва дисципліни англійською мовою	1-E8-112-4 Практична статистика для Data Science/ Practical Statistics for Data Science
Рекомендується для галузі знань	А – Освіта, Е – Природничі науки, математика та статистика, F – Інформаційні технології, G – Інженерія, виробництво та будівництво, 09 Біологія, 10 Природничі науки, 11 Математика та статистика, 12 Інформаційні технології, 13 Механічна інженерія, 14 Електрична інженерія, 16 Хімічна інженерія та біоінженерія, 17 Електроніка, автоматизація та електронні комунікації, 18 Виробництво та технології, 014.04 Середня освіта (Математика), 014.08 Середня освіта (Фізика та астрономія)
Кафедра	Статистики й теорії ймовірностей
П.І.П. НПП	Турчин Євген Валерійович
Рівень ВО	Перший (бакалаврський)
КУРС, семестр (в якому буде викладатись)	2-4 курси, парний
Мова викладання	Українська
Пререквізити (передумови вивчення дисципліни)	Базові знання зі шкільної математики та інформатики
Чому це цікаво/треба вивчати	Для успішного працевлаштуватися в ІТ-компанію на позицію аналітика даних, необхідно володіти інструментами, які актуальні для сфери діяльності ІТ-компанії. Водночас існує загальна тенденція: чим більше популярних бібліотек Python ви знаєте, тим ширші ваші можливості на ринку праці та тим вища ймовірність отримати конкурентну заробітну плату на посаді спеціаліста з аналізу даних.
Перелік тем з дисципліни	Розділ 1. Вступ до статистичного аналізу в Python Тема 1. Основи статистичного аналізу та його роль у дослідженнях даних Тема 2. Огляд можливостей Python для статистичного аналізу Тема 3. Робота у хмарному середовищі Google Colab: налаштування та основні функції Тема 4. Завантаження, збереження та обробка даних Тема 5. Підготовка даних до аналізу: очищення та нормалізація Розділ 2. Робота з бібліотекою NumPy Тема 1. Створення та управління масивами NumPy Тема 2. Векторні операції та ефективність обчислень Тема 3. Лінійна алгебра та статистичні функції в NumPy Тема 4. Генерація випадкових величин та їх застосування в аналізі даних Тема 5. Візуалізація основних характеристик масивів даних Розділ 3. Статистичний аналіз із SciPy.Stats Тема 1. Основні статистичні розподіли та їх застосування Тема 2. Оцінювання параметрів імовірнісних розподілів Тема 3. Тестування статистичних гіпотез (t-тест, χ^2-тест) Тема 4. Кореляційний та регресійний аналіз Тема 5. Аналіз часових рядів та прогнозування тенденцій Розділ 4. Візуалізація даних у Matplotlib Тема 1. Основи побудови графіків у Matplotlib Тема 2. Лінійні, точкові, гістограмні та коробчасті діаграми

	Тема 3. Налаштування стилю та оформлення графіків Тема 4. Взаємодія Matplotlib з NumPy, Pandas Тема 5. Аналіз тенденцій та зв'язків між змінними за допомогою графіків Розділ 5. Робота з табличними даними у Pandas Тема 1. Структура DataFrame та основні операції з таблицями Тема 2. Фільтрація, агрегування та трансформація даних Тема 3. Робота з категоріальними змінними та відсутніми значеннями Тема 4. Злиття, групування та об'єднання даних Тема 5. Обчислення описових статистичних характеристик Розділ 6. Основи машинного навчання у Scikit-Learn Тема 1. Огляд методів машинного навчання та їх застосування в аналізі даних Тема 2. Регресійний аналіз та класифікація Тема 3. Кластеризація даних та визначення схожих груп Тема 4. Оцінка якості моделей та вибір оптимальних параметрів Тема 5. Автоматизація аналізу великих наборів даних Розділ 7. Практичне застосування статистичного аналізу Тема 1. Аналіз фінансових даних та прогнозування трендів Тема 2. Використання статистичних методів у маркетингових дослідженнях Тема 3. Обробка соціальних та демографічних даних Тема 4. Аналіз медичних та біологічних даних Тема 5. Практичні задачі та виконання дослідницьких завдань
Як можна користуватися набутими знаннями і уміннями	Здатність виконувати усі стадії процесу статистичного аналізу даних, включаючи імпорт, попередню обробку даних (перетворення наборів даних до виду, зручного до аналізу), програмування (освоєння потужних інструментів мови Python, які спрощують статистичний аналіз даних та роблять його більш зрозумілим), розвідувальний аналіз даних (дослідження даних), моделювання даних (надання стислих зведених даних, що відображають виявлені тенденції у наборі даних), обмін інформацією (використання хмарного сервісу Google Colab, що забезпечує інтеграцію тексту, коду та результатів статистичного аналізу).
Очікувані результати навчання	Після опанування дисципліни ви зможете самостійно проводити повноцінний аналіз даних, використовувати інструменти Python для статистичних досліджень та застосовувати їх у практичних завданнях.
Інформаційне забезпечення	Робоча програма навчальної дисципліни, інтернет-ресурси
Види навчальних занять	Лекції, практичні заняття
Вид семестрового контролю	Диференційований залік
Максимальна кількість здобувачів на семестр	без обмежень

Декан факультету _____ Олександр ХАМІНІЧ