

I СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

1. Загальна інформація про навчальну дисципліну	
Повна назва навчальної дисципліни	Прикладна економетрія
Повна офіційна назва закладу вищої освіти	Сумський державний університет
Повна назва структурного підрозділу	Кафедра прикладної математики та моделювання складних систем факультету електроніки та інформаційних технологій
Розробник	Маринич Тетяна Олександрівна., к.е.н.
Рівень вищої освіти	Другий рівень вищої освіти, НРК – 7 рівень, QF-LLL – 7 рівень, FQ-EHEA – другий цикл
Семестр вивчення навчальної дисципліни	1 семестр
Обсяг навчальної дисципліни	Обсяг навчальної дисципліни становить 10 кредитів ЄКТС, 300 годин, з яких 80 годин становить контактна робота з викладачем (40 годин лекцій, 40 годин практичних занять), 220 годин становить самостійна робота
Мова(и) викладання	англійська
2. Місце навчальної дисципліни в освітній програмі	
Статус дисципліни	Обов'язкова
Передумови для вивчення дисципліни	інформатика; лінійна алгебра; теорія ймовірностей та математична статистика; математичні методи в економіці (за бажанням)
Додаткові умови	відсутні
Обмеження	відсутні
3. Мета навчальної дисципліни	
Курс призначений для покращення навичок студентів в галузях Business Intelligence та кількісного аналізу даних, підбору та верифікації алгоритмів, економетричних методів та моделей, знайомить студентів з можливостями використання статистичних інструментів для прийняття і покращення управлінських рішень, критичного оцінювання наукових публікацій.	
4. Зміст навчальної дисципліни	

Тема 1. Вступ до економетрики та аналізу даних.

Складові процесу “Data mining”. Прийоми й технології статистичного аналізу даних. Методи контрольованого і безконтрольного навчання. Типи даних та змінних, їх візуалізація. Введення до статистичної мови програмування R, IDE RStudio та редактора RMarkdown. Знайомство зі спеціальними пакетами R та Python для візуалізації, аналізу різних типів даних та моделювання.

Тема 2. Описова статистика, статистичний аналіз та візуалізація даних.

Застосування на практиці чисельних, табличних та графічних представлень описової статистики. Статистичний аналіз взаємозв'язків між змінними. Табличний формат даних. Прикладні кейси в R та Python.

Тема 3. Тестування гіпотез

Огляд основних понять та законів теорії ймовірності. Випадкова величина та типи розподілу даних. Закон нормального розподілу. Параметри, оцінки, довірчі інтервали. Види статистичних тестів та етапи тестування гіпотез. Реалізація T-test, ANOVA-test, Chi-Squared test в R та Python.

Тема 4. Збір та очищення даних.

Збір та підготовка даних у форматах csv, xlsx, з баз даних MySQL та через API. Робота з форматами даних XML, JSON. Очищення та підготовка даних до табличного формату. Робота з версіями файлів та у командах засобами платформи-репозитарію Github.

Тема 5. Одномірна та множинна лінійна регресія.

Метод найменших квадратів (МНК). Припущення класичної регресійної моделі та можливі наслідки їх порушення. Властивості МНК-оцінок коефіцієнтів регресії. Перевірка адекватності регресійної моделі та значущості коефіцієнтів; діагностичні графіки та тести. Покроковий регресійний аналіз. Використання категорійних незалежних змінних та ефектів взаємодії. Прогнозування та інтерпретація регресійної моделі. Потенційні проблеми класичної лінійної регресії (нелінійний характер даних; викиди; кореляція залишків; гетероскедастичність; колінеарність) та методи їх усунення. Квантильна регресія.

Тема 6. Діагностика та проблемні аспекти класичної регресії.

Моделі з порушенням передумов МНК. Узагальнений метод найменших квадратів. Трансформація змінних. Поліноміальна регресія. Робастна регресія. Зведення нелінійних моделей до лінійного вигляду. Вибір оптимальної моделі. Крос-валідація та регуляризація (Lasso, Ridge регресія).

Тема 7. Моделі класифікації.

Узагальнені лінійні регресійні моделі (GLM). Логістична регресія: однокласова та багатокласова. Метод максимальної правдоподібності. Оцінка якості прогнозу. AUC-ROC криві. Порівняння методів класифікації. Моделі дерева рішень, випадкового лісу та градієнтного бустінгу.

Тема 8. Регресії з панельними даними.

Пулова регресійна модель. Регресії з фіксованими ефектами, випадковими ефектами та фіксованими часовими ефектами. Діагностика моделей, інтерпретація та прогнозування.

Тема 9. Регресія з інструментальними змінними.

Визначення інструментальних змінних та перевірка їх валідності. Узагальнена регресійна модель з інструментальними змінними. Пошук інструментальних змінних. Розгляд прикладних кейсів.

Тема 10. Аналіз та моделювання часових рядів.

Декомпозиція та стаціонарність часових рядів. Метод експоненційного згладжування. Авторегресійні моделі. Прогнозування часових рядів з використанням алгоритмів машинного навчання (Neural Networks, Gradient Boosting, XGBoost, Prophet).

Тема 11. Додаткові теми аналізу часових рядів.

Векторні авторегресії (VAR). Коінтеграція та моделі корекції помилки (VEC). Структурні векторні моделі авторегресії (SVAR).

5. Очікувані результати навчання навчальної дисципліни

Після успішного вивчення навчальної дисципліни здобувач вищої освіти зможе:

PH1.	збирати, очищувати і аналізувати дані різного типу й походження
PH2.	правильно застосовувати та інтерпретувати основні статистичні інструменти, методи і моделі
PH3.	будувати та оптимізувати моделі для прогнозування та виявлення причинно-наслідкових зв'язків
PH4.	перевіряти адекватність, точність та економічний ефект економетричних моделей
PH5.	використовувати прикладні програмні середовища та пакети (R, Python) для аналізу, моделювання та прогнозування
PH6.	застосовувати сучасні техніки машинного навчання; підбирати оптимальні алгоритми аналізу даних
PH7.	презентувати та візуалізувати дані та результати досліджень

6. Роль навчальної дисципліни у досягненні програмних результатів

Програмні результати, досягнення яких забезпечує навчальна дисципліна:

ПРН2. Уміти формалізувати задачі певної предметної галузі, формулювати їх математичну постановку та обирати раціональний метод вирішення; розв'язувати отримані задачі аналітичними та чисельними методами, оцінювати точність та достовірність отриманих результатів.

ПРН5. Бути здатним проводити аналітичне дослідження математичних моделей об'єктів і процесів на предмет існування та єдиності їх розв'язку.

ПРН6. Бути здатним проводити дослідження та знаходити розв'язок некоректних задач з використанням методів регуляризації.

ПРН7. Уміти поєднувати методи математичного та комп'ютерного моделювання з неформальними процедурами експертного аналізу для пошуку оптимальних рішень.

ПРН12. Виявляти здатність до самонавчання та продовження професійного розвитку.

ПРН17. Володіти математичними методами обробки великих наборів даних. Вміти обирати до застосування оптимальні методи для конкретної задачі побудови моделі поведінки складної системи за існуючим набором даних та будувати графові ймовірнісні моделі для розв'язання технічних задач.

ПРН18. Уміти розробляти та використовувати на практиці алгоритми, пов'язані з спрощенням даних, що описують поведінку системи, класифікацією даних за певними ознаками без навчання та за попередньої наявності класів даних.

ПРН19. Уміти оцінити на адекватність результат обробки великих масивів даних.

ПРН20. Уміти застосовувати набуті знання про дані на практиці у сферах професійної діяльності.

7. Види навчальних занять та навчальної діяльності

7.1 Види навчальних занять

Видами навчальних занять при вивченні дисципліни є лекції (Л), семінарські (СЗ) та практичні заняття (ПЗ).

Тема 1.

Л 1. Вступ до економетрики та аналізу даних.

СЗ 1. Завдання та обмеження статистичного аналізу. Огляд методів контрольованого і неконтрольованого статистичного навчання.

ПР 1. Знайомство з R, RStudio, RMarkdown. Робота зі спеціальними пакетами R та Python для візуалізації та статистичного аналізу.

Тема 2.

Л 2. Описова статистика, статистичний аналіз та візуалізація даних.

СЗ 2. Застосування чисельних, табличних та графічних засобів описової статистики на практиці.

ПР 2-3. Графічний та статистичний аналіз даних в R та Python.

Тема 3. Тестування гіпотез

Л 3. Висновувальна статистика: принципи, етапи та методи.

Л 4. Тестування гіпотез та види статистичних тестів.

СЗ 3. Формулювання висновків щодо генеральної сукупності даних на підставі

аналізу даних вибірки.

ПР 4-5. Прикладні кейси із застосуванням тестів T-test, ANOVA-test, Chi-Squared test в R та Python.

Тема 4.

Л 5. Збір та очищення даних.

ПР 6-8. Збір даних: API, JSON, XML, SQL. Очищення та підготовка таблиці даних.

Тема 5.

Л 6. Одномірна та множинна лінійна регресія.

Л 7. Фіктивні змінні та ефекти взаємодії.

ПР 9. Побудова та оцінювання множинної лінійної регресії в R/Python.

Тема 6.

Л 8. Узагальнені лінійні моделі. Трансформація змінних та робастна регресія.

Зведення нелінійних моделей до лінійного вигляду.

Л 9-10. Вибір оптимальної моделі. Методи кросвалідації та регуляризації.

ПР 10-11. Регресійний аналіз в R/Python.

ПР 12. Кросвалідація, лассо та рідж-регресія в R/Python.

Тема 7.

Л 11. Задачі класифікації: логістична регресія, моделі дерева рішень та випадкового лісу.

Л 12. Логістична регресія: оцінювання та прогнозування.

ПР 13. Моделі дерева рішень та випадкового лісу в R/Python.

ПР 14. Оцінювання та прогнозування логістичної регресії в R/Python.

Тема 8.

Л 13. Лінійні регресії з панельними даними.

Л 14. Вибір оптимальної моделі та інтерпретація панельних регресій.

ПР 15. Панельні лінійні регресії з фіксованими та випадковими ефектами.

Тема 9.

Л 15. Побудова та оцінювання узагальненої регресійної моделі з інструментальними змінними. Пошук інструментальних змінних та перевірка їх валідності.

ПР 16. Побудова та оцінювання узагальненої регресійної моделі з інструментальними змінними в R/Python.

Тема 10.

Л 16. Аналіз часових рядів. Стаціонарність, декомпозиція, коінтеграція та причинність за Грейнджером.

Л 17. Моделі експоненційного згладжування; ARIMA/SARIMAX.

Л 18. Використання нейронних мереж та інших методів машинного навчання для прогнозування часових рядів.

ПР 17. Аналіз стаціонарності, декомпозиції, коінтеграції та причинності за Грейнджером в R/Python.

ПР 18. Моделювання часових рядів в R/Python.

Тема 11.

Л 19. Векторні авторегресії та моделі корекції помилки.

Л 20. Структурні векторні моделі авторегресії (SVAR).

ПР 19-20. Моделювання часових рядів в R/Python.

7.2 Види навчальної діяльності

- НД 1. Підготовка до практичних занять
- НД 2. Виконання та презентація результатів практичної роботи
- НД 3. Виконання ситуативних вправ
- НД 4. Підготовка мультимедійних презентацій
- НД 5. Електронне навчання у системах МІХ
- НД 6. Підготовка до поточних контролів
- НД 7. Участь у лекціях-дискусіях
- НД 8. Виконання курсової роботи.

Курсова робота з дисципліни передбачає проведення усіх етапів аналізу даних на практичних кейсах і включає: статистичний та графічний аналіз; тестування гіпотез; побудову різних моделей та вибір оптимальної моделі за критеріями статистичної значущості та прогнозуальної якості або оцінювання усередненого прогнозу за декількома моделями; інтерпретацію результатів та їх впровадження для прийняття відповідних управлінських рішень. Дані та предмет дослідження для курсової роботи обираються студентами самостійно та узгоджуються з викладачем до кінця 1 модулю навчання. Приклади джерел даних представлені у навчально-методичних матеріалах до курсу на електронній платформі змішаного навчання.

8. Методи викладання, навчання

Дисципліна передбачає навчання через:

- МН1. інтерактивні лекції (або он-лайн лекції);
- МН2. аналіз конкретних ситуацій
- МН3. проблемно-пошуковий метод
- МН4. евристичне навчання
- МН5. дослідницька робота
- МН6. творчий метод
- МН7. проєктна робота
- МН8. практико-орієнтоване навчання.

Лекції надають студентам матеріали зі статистичного аналізу та економетричного моделювання з різних точок зору, що є основою для самостійного навчання здобувачів вищої освіти (РН 2, РН 5, РН 8). Електронний конспект лекцій та завдання до практичних робіт доступні студентам на платформі змішаного навчання <https://mix.sumdu.edu.ua/>.

Лекції доповнюються практичними заняттями, що надають студентам можливість застосовувати теоретичні знання на практичних прикладах (РН 1, РН 7, РН 9, РН 10). Практико-орієнтоване навчання передбачає розуміння студентами доменної галузі, визначення проблеми дослідження, аналізу економічного ефекту від збору, підготовки та аналізу даних, вміння критично оцінювати результати аналізу та впроваджувати їх для покращення управлінських рішень (результати навчання РН 5, РН 5, РН 7). Самостійному навчанню сприятиме підготовка до лекцій, практичних занять, а також робота в невеликих групах для підготовки презентацій, що будуть представлені іншим групам, а потім проаналізовані, обговорені та продемонстровані у звіті про виконання завдань практико-орієнтованого навчання. Під час підготовки до презентацій за результатами практико-орієнтованого навчання студенти розвиватимуть навички самостійного навчання, швидкого критичного читання, синтезу та аналітичного мислення.

9. Методи та критерії оцінювання

9.1. Критерії оцінювання

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

R = 100 балів.

Призначення рейтингових балів:

Шкала оцінювання ECTS	Визначення	Чотирибальна національна шкала оцінювання	Рейтингова відсоткова шкала оцінювання	Рейтингова бальна шкала оцінювання
A	ВІДМІННО – відмінне виконання лише з незначною кількістю помилок	5,0 (відмінно)	$0,90R \leq RD \leq 1,00R$	$90 \leq RD \leq 100$
B	ДОБРЕ – в загальному правильна робота з певною кількістю помилок	4,0 (добре)	$0,82 R \leq RD < 0,89R$	$82 \leq RD < 89$
C			$0,74 R \leq RD < 0,81R$	$74 \leq RD < 81$
D	ЗАДОВІЛЬНО – непогано, але зі значною кількістю недоліків	3,0 (задовільно)	$0,64 R \leq RD < 0,73R$	$64 \leq RD < 73$
E			$0,60R \leq RD < 0,63R$	$60 \leq RD < 63$
FX	НЕЗАДОВІЛЬНО – можливе повторне складання	2 (незадовільно)	$0,35R \leq RD < 0,59R$	$35 \leq RD < 59$
F			$RD < 0,35R$	$1 \leq RD < 34$

9.2 Методи поточного формативного оцінювання

За дисципліною передбачені такі методи поточного формативного оцінювання: опитування, тестування, перевірка та оцінювання письмових завдань (робіт, звітів про виконання практичних робіт), захист презентацій (виконаних завдань, кейсів), самооцінювання поточного тестування, обговорення та взаємооцінювання студентами виконаних практичних завдань.

9.3 Методи підсумкового сумативного оцінювання

Оцінювання протягом семестру проводиться у формі усних та письмових опитувань, індивідуальних презентацій та колективних дискусій. Всі роботи повинні бути виконані самостійно. Індивідуальні завдання, схожі між собою, будуть відхилені.

Оцінка студента формується таким чином:

1. Ситуативні завдання (обговорення на лекціях, семінарах)	20 балів
2. Практичні завдання дослідницького характеру (виконання, презентація, захист)	40 балів
3. Два підсумкові модульні контролі, що проводяться у письмовій формі за тестовими технологіями	20*2=40 40 балів

Шкала оцінювання курсової роботи з навчальної дисципліни: R=100 балів (при позитивному оцінюванні від 60 до 100 балів). Оцінюванню підлягають такі етапи курсової роботи:

- розуміння даних, доменної галузі, формулювання проблеми, актуальність – 10 б.
- чіткий та деталізований аналіз даних, що відповідає поставленій меті – 20 б.;
- правильність дизайну та стилю коду, коректність моделей – 20 б.;
- ефективність використаних підходів та методів – 20 б.;
- наявність чітких та обґрунтованих рекомендацій та висновків – 10 б.;

захист та презентація результатів роботи – 20 б.

10. Ресурсне забезпечення навчальної дисципліни	
10.1 Засоби навчання	<p>Навчальний процес потребує використання мультимедійного (МО) комп'ютерного обладнання (КО) з доступом до інтернет (І); встановлення програмного забезпечення відкритого коду (ПЗ): R, RStudio, Python, Jupyter або інші Python Notebook.</p> <p>Доступ до програмного забезпечення може здійснюватися також з використанням хмарних технологій, наприклад, https://labs.cognitiveclass.ai/.</p>
10.2 Інформаційне та навчально- методичне забезпечення	<p>1. Основна навчальна література:</p> <p>1.1. Комплект електронних навчально- методичних матеріалів до дисципліни / Укладач Т.О. Маринич: https://mix.sumdu.edu.ua/.</p> <p>1.2. Anderson D. R., Sweeney D. J., Williams T. A. Statistics for Business and Economics, 13th Edition, Cengage, 2017 [ел. видання].</p> <p>1.3. James G., Witten D., Hastie T., Tibshirani R. (2015). An Introduction to Statistical Learning With Applications in R. – NY: Springer. – 426 p. [ел. видання].</p> <p>1.4. Duchesnay E., Löfstedt T. Statistics and Machine Learning in Python [ел. видання].</p> <p>1.5. Hyndman R. J., Athanasopoulos G. (2017). Forecasting: Principles and Practice [ел. ресурс].</p> <p>1.6. Methodological instructions for the Home Assignment Project on the discipline «Informatics» for students of economic, medical and ITspecialists / Т. О. Marynych. — Sumy: SSU, 2018. – 45 p.</p> <p>1.7. Coursera: Getting and Cleaning Data (John Hopkins University) [ел. ресурс].</p> <p>2. Додаткова рекомендована література:</p> <p>2.1. Козак Ю. Г. Математичні методи та моделі для магістрів з економіки. Практичні застосування. Навч. посіб. / Ю. Г. Козак, В. М. Мацкул. – К.: Центр учбової літератури, 2017. – 254 с. [ел. видання]</p> <p>2.2. Назаренко А. М. Основи економетрики: Підручник . - К.: Центр навчальної літератури, 2004. – 392 с.</p> <p>2.3. Hanck Ch., Arnold M., Gerber A., Schmelzer M. Introduction to Econometrics with R / – 2018. – 400 p. [ел. видання].</p> <p>2.4. Downey A. B. Think Stats Exploratory Data Analysis in Python [ел. видання].</p> <p>2.5. An In3troduction to R (pdf) : the most up-to- date official R Intro [ел. видання].</p> <p>2.6. Grolemund G., Wickham H. (2016). R for Data Science. – O'Reilly [ел. видання].</p> <p>2.7. Кабаков Р. И. R в действии. Анализ и визуализация данных в программе R.– М.: ДМК Пресс, 2014.–588с. [ел. видання]</p>