

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет прикладної математики та інформатики
Кафедра прикладної математики

Затверджено

На засіданні
кафедри прикладної математики
факультету прикладної математики та
інформатики
Львівського національного університету
імені Івана Франка
(протокол № 1 від 31 серпня 2023р.)

Завідувач кафедри



Юрій ЯЩУК

Силабус з навчальної дисципліни
“Прикладне статистичне моделювання”,
що викладається в межах першого (бакалаврського) рівня
вищої освіти для здобувачів
з спеціальності 113 Прикладна математика

Львів 2023 р.

Назва дисципліни	Прикладне статистичне моделювання
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка м. Львів, вул. Університетська 1
Факультет та кафедра, за якою закріплена дисципліна	Факультет прикладної математики та інформатики Кафедра прикладної математики
Галузь знань, шифр та назва спеціальності	11 – математика та статистика 113 – прикладна математика
Викладачі дисципліни	Щербатий Михайло Васильович, кандидат фізико-математичних наук, доцент кафедри прикладної математики Переймибіда Андрій Андрійович, кандидат фізико-математичних наук, доцент кафедри прикладної математики (лабораторні заняття)
Контактна інформація викладачів	mykhaylo.shcherbatyy@lnu.edu.ua ; https://ami.lnu.edu.ua/employee/shcherbatyy andrii.pereimybida@lnu.edu.ua https://ami.lnu.edu.ua/employee/pereimybida-andriy-andriyovych Головний корпус ЛНУ ім. І. Франка, каб. 278. м. Львів, вул. Університетська, 1
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/лабораторних занять (за попередньою домовленістю).
Сторінка курсу	https://ami.lnu.edu.ua/course/applied-statistical-modeling-applied-mathematics
Інформація про дисципліну	Дисципліна “Прикладне статистичне моделювання” є дисципліною на вибір зі спеціальності 113 – прикладна математика, яка викладається в 8-му семестрі в обсязі 5-ти кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Курс присвячений застосуванню методів статистичного аналізу для дослідження процесів, моделі яких будуються на основі даних спостережень. Основна увага приділяється регресійним моделям та їхньому застосуванню для аналізу та прогнозування поведінки різноманітних процесів. Поряд з класичною лінійною моделлю множинної регресії розглянуто окремі питання регресійних моделей: мультиколінеарність та способи вибору найбільш суттєвих пояснюючих змінних; тестування та оцінка параметрів у випадку гетероскедантичності та автокореляції регресійних залишків; симультивні моделі; навчання та тестування моделей. Розглядаються алгоритмічна та програмна реалізація різноманітних регресійних моделей. Матеріали курсу ілюструються прикладами застосувань регресійних моделей в різних галузях (економіці, біології, медицині та інших).
Мета та цілі дисципліни	Метою курсу є навчити студентів: <ul style="list-style-type: none"> будувати регресійні моделі різноманітних процесів на основі даних спостережень за досліджуваним процесом;

	<ul style="list-style-type: none"> • аналізувати адекватність побудованої моделі та робити прогнозування за її допомогою; • розробляти відповідні алгоритми та програмне забезпечення для реалізації побудованих регресійних моделей; • використовувати для розв'язування різноманітних прикладних задач наявне в системах комп'ютерної математики програмне забезпечення з статистичного моделювання (зокрема Statistics and Machine Learning Toolbox, пакет MatLab)Matlab).
<p>Література для вивчення дисципліни</p>	<p>Основна література</p> <ol style="list-style-type: none"> 1. <i>Kroese D. P., Chan J. C.C.</i> Statistical Modeling and Computation. // Springer, 2014. – 400 p. 2. <i>Mendenhall W., Sincich T.</i> A second course in statistics: regression analysis./ 7 th ed. //Prentice Hall, 2012 – 797 p. 3. <i>James G., Witten D., Hastie T., Tibshirani R.</i> An Introduction to Statistical Learning with Applications in R. //Springer, 2015. – 426 p. 4. <i>Лавренюк С.П., Оліскевич М.О.</i> Основи економетрії. Тексти лекцій. // Львів: Вид. центр ЛНУ ім. І.Франка, 2003. – 367 с. 5. <i>Faul A. C.</i> A concise introduction to machine learning. // Taylor & Francis Group, 2020. – 314 p. <p>Додаткова література</p> <ol style="list-style-type: none"> 6. <i>Hill C.R., Griffiths W.E., Lim C.G.</i> Principles of econometrics. / 5th ed. //Wiley, 2017 – 878 p. 7. <i>Brunton S.L., Kutz J.N.</i> Data-Driven. Science and Engineering. Machine Learning, Dynamical Systems, and Control. // Cambridge University Press, 2019. – 472 p. 8. MATLAB Homepage: http://www.mathworks.com/products/matlab/ 9. GNU Octave Homepage: http://www.gnu.org/software/octave/
<p>Обсяг курсу</p>	<p>Загальний обсяг: 150 годин. Аудиторних занять: 70 год., з них 42 год. лекцій і 28 год. лабораторних занять. Самостійної роботи: 80 год.</p>
<p>Очікувані результати навчання</p>	<p>Після завершення цього курсу студент буде :</p> <p>знати: методи статистичного (регресійного) аналізу і їх можливості для побудови моделей різноманітних процесів на основі даних спостережень;</p> <p>вміти:</p> <ul style="list-style-type: none"> • будувати статистичні (регресійні) моделі на основі даних спостережень різних прикладних проблем; • оцінювати та інтерпретувати параметри моделей, будувати для отриманих параметрів інтервали довіри; • аналізувати особливі випадки регресійних моделей (мультиколінеарність, автокорельованість, гетероскедантичність) та будувати відповідні алгоритми оцінки параметрів; • оцінювати якість побудованих моделей; порівнювати різні моделі; • здійснювати прогноз (точковий та інтервальний) поведінки процесу, на основі побудованої моделі; • здійснювати програмну реалізацію розроблених алгоритмів; • використовувати вбудовані функції статистичних пакетів (зокрема Statistics and Machine Learning Toolbox, пакет MatLab) та на їх основі розробляти програмне забезпечення для реалізації інших алгоритмів статистичного моделювання;

	<ul style="list-style-type: none"> використовувати графічні можливості статистичних пакетів для аналізу даних. 					
Ключові слова	Класична лінійна модель множинної регресії, ANOVA дисперсійний аналіз, Коваріаційна матриця, Кореляційна матриця, Узагальнена лінійна модель множинної регресії (УЛІММР), Метод найменших квадратів, Метод максимальної правдоподібності, Узагальнений метод найменших квадратів, Мультиколінеарність, Метод головних компонент, Гетоскедантичність, Автокорельованість, Перехресна перевірка, Навчальна вибірка, Екзаменаційна вибірка, Нелінійні моделі регресії, Статистичні моделі у вигляді систем лінійних взаємозалежних рівнянь.					
Формат курсу	Очний. Проведення лекцій, лабораторних робіт і консультацій.					
Теми	Подано нижче у таблиці Схема курсу “Прикладне статистичне моделювання”					
Підсумковий контроль, форма	Залік у кінці семестру.					
Пререквізити	Для вивчення курсу студенти потребують базових знань з <ul style="list-style-type: none"> - Математичного аналізу; - Алгебри; - Теорії ймовірностей та математичної статистики; - Методів оптимізації; - Програмування; Систем комп’ютерної математики (напр., Matlab, Octave) або бібліотек чисельних методів інших програмних продуктів (напр. NumPy для Python). 					
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Презентації, лекції, лабораторні заняття. Домашні та індивідуальні завдання. Самостійна робота.					
Необхідне обладнання	Комп’ютер із одним із пакетів - MATLAB/Octave, R, Python, Internet доступ.					
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	Оцінювання проводиться за 100-бальною шкалою.					
	Оцінка за шкалою ECTS		Оцінка в балах	Оцінка за національною шкалою		залік зараховано
				Екзамен, диференційований залік		
	A	Відмінно	100 - 90	Відмінно	5	
	B	Дуже добре	81- 89	Добре	4	
	C	Добре	71 -80			
	D	Задовільно	61 - 70	Задовільно	3	
E	Достатньо	51- 60				
FX (F)	Незадовільно	0 - 50	Незадовільно	2	не зараховано	

Бали нараховуються за наступним співвідношенням:

- індивідуальні завдання : 69% семестрової оцінки; максимальна кількість балів 69 (4 завдання по 14, 17, 21 та 17 балів);
- контрольна робота: 21% семестрової оцінки; максимальна кількість балів 21;
- активність на заняттях: 10% семестрової оцінки; максимальна кількість балів 10.

Підсумкова максимальна кількість балів 100. Студент отримує залік, якщо підсумкова кількість балів більше 50 (51-100).

Для кожного завдання встановлено терміни здачі. Завдання, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку.

Критерії оцінювання індивідуальних завдань. За кожне завдання студент отримує SZ балів, які обчислюється за формулою

$$SZ = S * k,$$

де k – коефіцієнт виконання завдання, $k \in [0, 1]$, S – максимальне кількість балів за дане завдання.

k – коефіцієнт виконаного завдання	Критерії оцінювання
$k \in [0.9, 1]$	Студент повністю виконав умови завдання; алгоритм реалізовано правильно; відповідає на практично на всі запитання, пов'язані з тематикою завдання; проводить чіткий аналіз та порівняння отриманих результатів.
$k \in [0.7, 0.9)$	Студент повністю виконав умови завдання; алгоритм реалізовано правильно; на деякі питання, пов'язані з тематикою завдання, відповідає з незначними неточностями; проводить аналіз отриманих результатів з незначними неточностями.
$k \in [0.5, 0.7)$	Студент виконав завдання з незначними помилками, але самостійно їх виправляє, якщо на них вкаже викладач; на деякі запитання, пов'язані з тематикою завдання, відповідає з неточностями; проводить аналіз отриманих результатів з неточностями.
$k \in [0.3, 0.5)$	Студент виконав завдання частково; алгоритм реалізовано з помилками, які частково може виправити, якщо на них вкаже викладач; на запитання відповідає з помилками; проводить аналіз отриманих результатів з помилками.
$k \in [0.1, 0.3)$	Студент виконав завдання частково, алгоритм реалізовано з помилками, які самостійно не може виправити; переважно не відповідає на запитання.
$k \in (0, 0.1)$	Студент виконав завдання частково або з грубими помилками, які самостійно не може виправити; демонструє незнання матеріалу.
$k = 0$	Студент не виконав завдання.

Академічна доброчесність: Очікується, що роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в написанні програм та поданні результатів є підставою

	<p>для зменшення балів при оцінці завдання чи його незарахування викладачем.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та лабораторні заняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання всіх видів робіт та індивідуальних завдань, передбачених курсом.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали набрані за виконання індивідуальних завдань, контрольну роботу та активність на заняттях. Не допускається пропуски та запізнення на заняття без поважних причин; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін. Жодні форми порушення академічної доброчесності не толеруються.</p>
<p>Питання до контрольної роботи.</p>	<ol style="list-style-type: none"> 1. Класична лінійна модель множинної регресії (КЛММР). 2. Оцінка параметрів моделі методом найменших квадратів (МНК) та методом максимальної правдоподібності (ММП). 3. Аналіз варіації результуючого показника. Оцінка якості статистичної моделі. 4. Статистичні властивості оцінок КЛММР. 5. Перевірка гіпотези про значимість коефіцієнтів регресійної моделі та побудова інтервалів довіри для коефіцієнтів моделі (критерій Стьюдента). 6. Критерій Фішера перевірки лінійного зв'язку між регресором та регресантами. 7. Ознаки і причини мультиколінеарності. 8. Методи усунення мультиколінеарності (перехід до зміщених оцінок, метод головних компонент, відбір найбільш суттєвих пояснюючих змінних). 9. Узагальнена лінійна модель множинної регресії (УЛММР). Узагальнений метод найменших квадратів (УМНК). 10. УЛММР з гетероскедантичними залишками. Критерій Глейсера. Тест Гольдфельда-Куандта. 11. УЛММР з автокорельованими залишками. Критерій Дарбіна-Уотсона. Оцінка авторегресійного параметра в моделі з автокорельованими залишками. 12. Дослідження точності регресійної моделі в реалістичній ситуації. Перехресна перевірка. Вибір даних для навчальної та екзаменаційної вибірок. 13. Нелінійні моделі регресії і їх лінеаризація. Підбір лінеаризуючого перетворення (підхід Бокса-Кокса). 14. Статистичні моделі у вигляді систем лінійних взаємозалежних рівнянь (СЛВР). Умови ототожнення (ідентифікованості) рівнянь моделі у вигляді СЛВР. 15. Статистичне оцінювання невідомих параметрів СЛВР. Оцінювання структурних параметрів окремого рівняння двокроковим методом найменших квадратів.

	16. Прогнозування поведінки модельованих систем з допомогою статистичних моделей.
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

Схема курсу “Прикладне статистичне моделювання”

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література, Ресурси в інтернеті	Завдан ня, год.	Термін виконан ня
1	Тема 1. Особливості побудови статистичних моделей. Етапи побудови статистичної моделі. Постановка задачі. Специфікація моделі. Формування вхідної інформації. Оцінка параметрів моделі. Аналіз якості моделі. Прогнозування на основі статистичних моделей. Статистичне моделювання та машинне навчання.	Лекція, Самостійна робота	[1,3,5,7]	4 2	1 тиждень
	Тема 1. Побудова моделей на основі статистичних даних. Програмні засоби для дослідження статистичних моделей.	Лабораторна Самостійна робота	[1,3,5,7,8,9]	2 3	1 тиждень
2	Тема 2. Основні поняття регресійного аналізу. Функція регресії, коефіцієнт детермінації, коефіцієнт множинної кореляції. Коваріаційна та кореляційна матриці. Вихідні дані регресійного аналізу.	Лекція, Самостійна робота	[1,2,4,6,7]	2 2	1 тиждень
	Тема 2. Підготовка та аналіз вихідних даних. Побудова діаграм. Обчислення коваріаційної та кореляційної матриць. <i>Індивідуальне завдання 1.</i>	Лабораторна Самостійна робота	[1,3,7,8,9]	2 3	2 тижні
3	Тема 3. Класична лінійна модель множинної регресії (КЛММР). Оцінка параметрів КЛММР. Скалярна та векторна форма подання КЛММР. Оцінка	Лекція, Самостійна робота	[2,4,6,7]	4 3	1 тиждень

	параметрів моделі методом найменших квадратів (МНК). Оцінка параметрів КЛММР методом максимальної правдоподібності (ММП).				
	Тема 3. Оцінка параметрів моделі методом найменших квадратів (МНК).	Лабораторна Самостійна робота	[2,4,5,6,8,9]	2 3	1 тиждень
4	Тема 4. Аналіз варіації результуючого показника. Оцінка якості статистичної моделі. Вибіркова варіація результуючого показника. ANOVA дисперсійний аналіз. ANOVA таблиця. Критерії степені відповідності моделі даним спостережень: коефіцієнт множинної кореляції, оцінений коефіцієнт множинної кореляції, коефіцієнт детермінації, критерій Фішера. Застосування критерію Фішера для порівняння двох моделей.	Лекція, Самостійна робота	[2,4,6]	2 3	1 тиждень
	Тема 4. ANOVA дисперсійний аналіз. Застосування критерію Фішера для порівняння двох моделей. <i>Індивідуальне завдання 2.</i>	Лабораторна Самостійна робота	[2,6,8,9]	2 3	2 тижні
5	Тема 5. Статистичні властивості оцінок КЛММР. Обґрунтованість, незміщеність та оптимальність МНК оцінок параметрів моделі. Коваріаційна матриця оцінок параметрів моделі. Властивості оцінок, які справедливі при нормальності регресійних залишків. Перевірка гіпотези про значимість коефіцієнтів регресійної моделі та побудова інтервалів довіри для	Лекція, Самостійна робота	[2,3,4,6]	4 3	1 тиждень

	коєфіцієнтів моделі (критерій Ст'юдента). Критерій Фішера перевірки лінійного зв'язку між регресором та регресантами.				
	Тема 5. Перевірка гіпотези про значимість коєфіцієнтів регресійної моделі. Критерій Фішера Перевірка лінійного зв'язку між регресором та регресантами на основі критерію Фішера.	Лабораторна Самостійна робота	[2,3,6,8,9]	2 4	1 тиждень
6	Тема 6. Мультиколінеарність та методи її усунення. Помилки специфікації моделі. Ознаки і причини мультиколінеарності. Методи усунення мультиколінеарності (перехід до зміщених оцінок, метод головних компонент, відбір найбільш суттєвих пояснюючих змінних). Помилки специфікації моделі, викликані невірним вибором кількості пояснюючих змінних.	Лекція, Самостійна робота	[2,3,6,7]	2 3	1 тиждень
	Тема 6. Відбір найбільш суттєвих пояснюючих змінних в регресійних моделях. <i>Індивідуальне завдання 3.</i>	Лабораторна Самостійна робота	[2,6,7,8]	2 3	5 тижнів
7	Тема 7. Узагальнена лінійна модель множинної регресії (УЛММР). Узагальнений метод найменших квадратів (УМНК). Загальний вигляд УЛММР. Типи УЛММР. Приведення УЛММР до КЛММР. УМНК оцінки параметрів моделі.	Лекція, Самостійна робота	[2,4,6]	4 3	1 тиждень
	Тема 7. Приведення	Лабораторна	[2,4,6]	2	1

	УЛММР до КЛММР.	Самостійна робота		2	тиждень
8	Тема 8. УЛММР з гетероскедантичними залишками. Деякі способи перевірки гомо-/гетероскедантичності для УЛММР. Критерій Глейсера. Тест Гольдфельда-Куандта.	Лекція, Самостійна робота	[2,3,4,6]	2 3	1 тиждень
	Тема 8. Перевірка моделі на наявність гетероскедантичності.	Лабораторна Самостійна робота	[2,4,6,8]	2 3	1 тиждень
9	Тема 9. УЛММР з автокорельованими залишками. Перевірка гіпотези про наявність/відсутність автокорельованості регресійних залишків (критерій Дарбіна-Уотсона). Оцінка авторегресійного параметра в моделі з автокорельованими залишками.	Лекція, Самостійна робота	[2,3,4,6]	4 3	1 тиждень
	Тема 9. Перевірка моделі на наявність автокорельованості..	Лабораторна Самостійна робота	[2,4,6,8]	2 3	1 тиждень
10	Тема 10. Дослідження точності регресійної моделі в реалістичній ситуації. Перехресна перевірка. Вибір даних для навчальної та екзаменаційної вибірок.	Лекція, Самостійна робота	[1,2,3,5,6]	2 3	1 тиждень
	Тема 10. Вибір даних для навчальної та екзаменаційної вибірок.	Лабораторна Самостійна робота	[2,6,7,8,9]	2 2	1 тиждень
11	Тема 11. Нелінійні моделі регресії і їх лінеаризація. Деякі властивості нелінійних залежностей, які підлягають безпосередній лінеаризації (залежності гіперболічного, експоненціального, степеневого та логарифмічного типів).	Лекція, Самостійна робота	[1,2,6]	4 3	1 тиждень

	Підбір лінеаризуючого перетворення (підхід Бокса-Кокса				
	Тема 11. Застосування нелінійних моделей для аналізу даних. <i>Індивідуальне завдання 4.</i>	Лабораторна Самостійна робота	[1,2,6,8]	2 4	2 тижні
12	Тема 12. Статистичні моделі у вигляді систем лінійних взаємозалежних рівнянь (СЛВР). Умови ототожнення (ідентифікованості) рівнянь моделі у вигляді СЛВР. Загальний вигляд статистичної моделі у вигляді СЛВР. Структурна та приведена форма моделі.	Лекція, Самостійна робота	[2,6]	2 3	1 тиждень
	Тема 12. Побудова регресійних моделей у вигляді систем лінійних взаємозалежних рівнянь. Структурна та приведена форма моделі.	Лабораторна Самостійна робота	[2,6,8]	2 3	1 тиждень
13	Тема 13. Умови ототожнення (ідентифікованості) рівнянь моделі у вигляді СЛВР. Статистичне оцінювання невідомих параметрів СЛВР. Проблеми ототожнення окремого рівняння моделі та моделі в цілому. Необхідні і достатні умови ототожнення окремого рівняння моделі та моделі в цілому. Непрямий метод найменших квадратів для СЛВР. Ідентифікація (статистичне оцінювання значень невідомих параметрів) рекурсивних систем. Оцінювання структурних параметрів окремого рівняння двокроковим методом найменших квадратів.	Лекція, Самостійна робота	[2,6]	4 4	1 тиждень

	Тема 13. Контрольна робота.	Лабораторна		2	1 тиждень
14	Тема 14. Прогнозування поведінки модельованих систем з допомогою статистичних моделей. Точковий та інтервальний прогноз значень ендогенних змінних. Деякі загальні підходи до аналізу точності оцінювання та порівняння методів та моделей.	Лекція, Самостійна робота	[1,2,3,5,6]	2 6	1 тиждень
	Тема 14. Обговорення результатів. Підведення підсумків.	Лабораторна	[1-7]	2	1 тиждень