

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
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 № ____ від _____ 2020 __ р.)
Завідувач кафедри Савула Я.Г.

Силабус з навчальної дисципліни
“Прикладне статистичне моделювання”,
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 113 – прикладна математика

Львів 2020 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, доцент кафедри прикладної математики
Контактна інформація викладачів	mykhaylo.shcherbatyy@lnu.edu.ua ; https://ami.lnu.edu.ua/employee/shcherbatyy 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	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 засоби її вилучення; тестування та оцінка параметрів у випадку гетероскедантичності та автокореляції регресійних залишків; симулятивні моделі; навчання та тестування моделей. Для вивчення курсу використовується Statistics and Machine Learning Toolbox (пакет MatLab). Студенти також можуть використовувати інші програмні продукти для роботи із статистичними моделями.
Коротка анотація дисципліни	Дисципліна “Прикладне статистичне моделювання” є дисципліною вільного вибору студента циклу професійної та практичної підготовки з спеціальності 113 – прикладна математика для освітньої програми Прикладна Математика, яка викладається в 8-му семестрі (5 кредитів ECTS). Статистичне моделювання полягає в отриманні знань із даних спостережень за розглядуваним процесом. В курсі розглядаються різноманітні регресійні моделі, які будуються на основі вхідної та вихідної інформації про досліджуваний процес. Побудова регресійних моделей є одним із методів статистичного (та машинного) навчання із вчителем. Розглядаються алгоритмічна та програмна реалізація різноманітних регресійних моделей. Матеріали курсу ілюструються прикладами застосувань регресійних моделей в різних галузях (економіці, біології, медицині та інших).

<p>Мета та цілі дисципліни</p>	<p>Метою курсу є навчити студентів:</p> <ul style="list-style-type: none"> • будувати регресійні моделі різноманітних процесів на основі даних спостережень за досліджуваним процесом; • 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>	<ol style="list-style-type: none"> 1. <i>Айвазян С.А.</i> Методы эконометрики: учебник. // М.: Магистр: ИНФРА-М, 2010. – 512 с. 2. <i>Лавренюк С.П., Оліскевич М.О.</i> Основи економетрії. Тексти лекцій. // Львів: Вид. центр ЛНУ ім. І.Франка, 2003. – 367 с. 3. <i>Kroese D. P., Chan J. C.C.</i> Statistical Modeling and Computation. // Springer, 2014. – 400 p. 4. <i>Mendenhall W., Sincich T.</i> A second course in statistics: regression analysis./ 7 th ed. //Prentice Hall, 2012 – 797 p. 5. <i>James G., Witten D., Hastie T., Tibshirani R.</i> An Introduction to Statistical Learning with Applications in R. //Springer, 2015. – 426 p. 6. <i>Faul A. C.</i> A concise introduction to machine learning. // Taylor & Francis Group, 2020. – 314 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.
<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; • використовувати графічні можливості статистичних пакетів для аналізу даних.

<p>Ключові слова</p>	<p>Класична лінійна модель множинної регресії, ANOVA дисперсійний аналіз, Коваріаційна матриця, Кореляційна матриця, Узагальнена лінійна модель множинної регресії (УЛММР), 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.</p>
<p>Формат курсу</p>	<p>Дистанційний (Microsoft Teams), очний. Проведення лекцій, лабораторних робіт і консультацій.</p>
<p>Теми</p>	<p>Тема 1. Особливості побудови статистичних моделей. Постановка задачі. Етапи побудови статистичної моделі. Специфікація моделі. Формування вхідної інформації. Оцінка параметрів моделі. Аналіз якості моделі.</p> <p>Тема 2. Основні поняття регресійного аналізу. Функція регресії, коефіцієнт детермінації, коефіцієнт множинної кореляції. Коваріаційна та кореляційна матриці. Вихідні дані регресійного аналізу.</p> <p>Тема 3. Класична лінійна модель множинної регресії (КЛММР). Оцінка параметрів КЛММР. Скалярна та векторна форма подання КЛММР. Оцінка параметрів моделі методом найменших квадратів (МНК). Оцінка коефіцієнтів КЛММР методом максимальної правдоподібності (ММП).</p> <p>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 (критерій Дарбіна-Уотсона, коефіцієнт множинної кореляції, оцінений коефіцієнт множинної кореляції, коефіцієнт детермінації, критерій Фішера. Застосування критерію Фішера для порівняння двох моделей.</p> <p>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 коефіцієнтів моделі (критерій Стюдента). Критерій Фішера перевірки лінійного зв'язку між регресором та регресантами.</p> <p>Тема 6. Мультиколінеарність та методи її усунення. Ознаки і причини мультиколінеарності. Методи усунення мультиколінеарності (перехід до зміщених оцінок, метод головних компонент, відбір найбільш суттєвих пояснюючих змінних).</p> <p>Тема 7. Помилки специфікації моделі. Помилки специфікації моделі, викликані невірним вибором кількості пояснюючих змінних.</p> <p>Тема 8. Узагальнена лінійна модель множинної регресії (УЛММР). Узагальнений метод найменших квадратів (УМНК). Загальний вигляд УЛММР. Типи УЛММР. Приведення УЛММР до КЛММР. УМНК оцінки параметрів моделі.</p> <p>Тема 9. УЛММР з гетероскедантичними залишками. Деякі способи перевірки гомо-/ гетероскедантичності для УЛММР. Тест рангової кореляції Спірмана. Критерій Глейсера. Тест Гольдфельда-Куандта.</p> <p>Тема 10. УЛММР з автокорельованими залишками. Перевірка гіпотези про наявність/відсутність автокорельованості регресійних залишків (критерій Дарбіна-Уотсона). Оцінка авторегресійного параметра в моделі з автокорельованими залишками.</p>

	<p>Тема 11. Дослідження точності регресійної моделі в реалістичній ситуації. Перехресна перевірка. Вибір даних для навчальної та екзаменаційної вибірок.</p> <p>Тема 12. Нелінійні моделі регресії і їх лінеаризація. Деякі властивості нелінійних залежностей, які підлягають безпосередній лінеаризації (залежності гіперболічного, експоненціального, степеневого та логарифмічного типів). Підбір лінеаризуючого перетворення (підхід Бокса-Кокса).</p> <p>Тема 13. 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. Проблеми ототожнення окремого рівняння моделі та моделі в цілому. Необхідні і достатні умови ототожнення окремого рівняння моделі та моделі в цілому.</p> <p>Тема 14. Статистичне оцінювання невідомих параметрів СЛВР. Непрямий метод найменших квадратів для СЛВР. Ідентифікація (статистичне оцінювання значень невідомих параметрів) рекурсивних систем. Оцінювання структурних параметрів окремого рівняння двокроковим методом найменших квадратів.</p> <p>Тема 15. 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.</p>
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.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Лекції, лабораторні заняття. 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. Бали нараховуються за наступним співвідношенням: <ul style="list-style-type: none"> • 3 індивідуальні завдання: максимальна кількість балів за кожне завдання 25; • Виконання домашніх завдань, робота на лабораторних заняттях: максимальна кількість балів 25. Підсумкова максимальна кількість балів 100. Студент отримує залік, якщо підсумкова кількість балів більше 50 (51-100).

	<p>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 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 при виконанні індивідуальних завдань, виконанні домашніх завдань та поточній роботі протягом семестру. При цьому обов'язково враховуються присутність на заняттях та активність студента під час лабораторних занять. 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.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p>
<p>Питання до заліку чи екзамену.</p>	<p>Класична лінійна модель множинної регресії (КЛММР).</p> <p>Оцінка параметрів моделі методом найменших квадратів (МНК) та методом максимальної правдоподібності (ММП).</p> <p>Аналіз варіації результуючого показника. Оцінка якості статистичної моделі.</p> <p>Статистичні властивості оцінок КЛММР.</p> <p>Перевірка гіпотези про значимість коефіцієнтів регресійної моделі та побудова інтервалів довіри для коефіцієнтів моделі (критерій Стьюдента).</p> <p>Критерій Фішера перевірки лінійного зв'язку між регресором та регресантами.</p> <p>Ознаки і причини мультиколінеарності.</p> <p>Методи усунення мультиколінеарності (перехід до зміщених оцінок, метод головних компонент, відбір найбільш суттєвих пояснюючих змінних).</p> <p>Узагальнена лінійна модель множинної регресії (УЛММР). Узагальнений метод найменших квадратів (УМНК).</p> <p>УЛММР з гетероскедантичними залишками. Критерій Глейсера. Тест Гольдфельда-Куандта.</p> <p>УЛММР з автокорельованими залишками. Критерій Дарбіна-Уотсона. Оцінка авторегресійного параметра в моделі з автокорельованими залишками.</p> <p>Дослідження точності регресійної моделі в реалістичній ситуації. Перехресна перевірка. Вибір даних для навчальної та екзаменаційної вибірок.</p> <p>Нелінійні моделі регресії і їх лінеаризація. Підбір лінеаризуючого перетворення (підхід Бокса-Кокса).</p>

	<p>Статистичні моделі у вигляді систем лінійних взаємозалежних рівнянь (СЛВР). Умови ототожнення (ідентифікованості) рівнянь моделі у вигляді СЛВР.</p> <p>Статистичне оцінювання невідомих параметрів СЛВР. Оцінювання структурних параметрів окремого рівняння двокроковим методом найменших квадратів.</p> <p>Прогнозування поведінки модельованих систем з допомогою статистичних моделей.</p>
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.