

ТЕХНОЛОГІЇ АНАЛІЗУ ДАНИХ

Викладач: к. ф.-м.н., доц. Світлана Петрівна Швидка

Кафедра: фундаментальної та прикладної математики, 1й корп. ЗНУ, ауд. 21-в (1^й поверх)

Email: svetlana.shvydka@gmail.com

Телефон: (061) 289-12-52 (кафедра), 289-41-11 (деканат)

Facebook Messenger:

Освітня програма, рівень вищої освіти		Прикладна математика. Магістр.				
Статус дисципліни		Вибіркова				
Кредити ECTS	5	Навч. рік	2021-2022 1 семестр	Рік навчання - 2	Тижні	14
Кількість годин	150	Кількість змістових модулів¹	6	Лекційні заняття – 20 год Лабораторні заняття – 22 год Самостійна робота – 108 год.		
Вид контролю	Залік					
Посилання на курс в Moodle			https://moodle.znu.edu.ua/course/view.php?id=11430			
Консультації:			щосереди, 11.25-12.45 або за домовленістю чи ел. поштою			

ОПИС КУРСУ

Метою вивчення навчальної дисципліни «Технології аналізу даних» є отримання нових знань стосовно застосування програмного середовища R до розв'язання задач з теорії ймовірностей та математичної статистики; удосконалення навичок написання програмного коду на мові програмування R, використання пакетів мови R, які є надзвичайно потужними інструментами для статистичного аналізу і створення статистичної графіки та набуття вмінь проведення статистичного дослідження і розробки методів розв'язання задач математичної статистики з використанням програми R.

Основними завданнями вивчення дисципліни «Додаткові розділи теорії ймовірностей та математичної статистики» є:

1. Ознайомлення з етапами аналізу даних у середовищі R.
2. Ознайомлення з основними методами статистичного аналізу.
3. Оволодіння прикладними (комп'ютерними) аспектами статистичного аналізу даних.
4. Розвиток навичок аналізу статистичних даних з використанням як стандартних методів R так і спеціальних пакетів.
5. Розвиток навичок візуалізації отриманих результатів.

¹ 1 змістовий модуль = 15 годин (0,5 кредита ECTS). Детальна формула розрахунку – в рекомендаціях.



Змістове наповнення курсу, що викладається на лекційних і практичних заняттях та засвоюється студентом під час самостійної роботи, забезпечує набуття компетентностей:

- КЗ 1. Здатність до абстрактного мислення, аналізу та синтезу;
- КЗ 2. Здатність застосовувати знання у практичних ситуаціях;
- КЗ 3. Здатність до розуміння предметної області та професійної діяльності;
- КЗ 5. Здатність вчитися і оволодівати сучасними знаннями.

ОЧІКУВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

У разі успішного завершення курсу студент зможе:

1. (ПР-1) знати теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації.
2. (ПР-3) використовувати базові знання інформатики й сучасних інформаційних систем та технологій, навички програмування, технології безпечної роботи в комп'ютерних мережах, методи створення баз даних та інтернет-ресурсів, технології розроблення алгоритмів і комп'ютерних програм мовами високого рівня із застосуванням об'єктно-орієнтованого програмування для розв'язання задач проектування і використання інформаційних систем та технологій.
3. Будувати ймовірнісні розподіли для дискретної та неперервної випадкової величини у середовищі R.
4. Проводити обробку та аналіз статистичних даних у системі R.
5. Будувати статистичний розподіл та його графічне зображення.
6. Проводити дослідження кореляційної залежності даних.
7. Будувати парну, множинну лінійну і нелінійну регресійні моделі та проводити їх оцінку.
8. Будувати узагальнені регресійні моделі (Generalized Additive models) у R.
9. Проводити прогнозування з використанням статистичних моделей.

ОСНОВНІ НАВЧАЛЬНІ РЕСУРСИ

Матеріали на платформі Moodle

<https://moodle.znu.edu.ua/course/view.php?id=11430>



КОНТРОЛЬНІ ЗАХОДИ

Поточні контрольні заходи (тах 60 балів):

Лабораторні роботи – 6 робіт, виконання і захист кожної оцінюється в 5 балів. Загалом **30 балів**.
Самостійні роботи – 6 робіт, виконуються самостійно, а складання кожної оцінюється в 3 бали. Загалом **18 балів**.

Поточні контрольні роботи – 2 тести по 6 балів кожен (проводяться на базі Moodle). Загалом **12 балів**.

Підсумкові контрольні заходи (тах 40 балів):

Залік – 40 балів.

Контрольний захід		Термін виконання	% від загальної оцінки
Поточний контроль (тах 60%)			
Змістовий модуль 1	Лабораторна робота №1	1 тиждень	2,5%
	Лабораторна робота №1	2 тиждень	2,5%
	Самостійна робота №1	2 тиждень	3%
Змістовий модуль 2	Лабораторна робота №2	3 тиждень	2,5%
	Лабораторна робота №2	4 тиждень	2,5%
	Самостійна робота №2	4 тиждень	3%
Змістовий модуль 3	Лабораторна робота №3	5 тиждень	2,5%
	Лабораторна робота №3	6 тиждень	2,5%
	Самостійна робота №3	6 тиждень	3%
	Тестове завдання №1	6 тиждень	6%
Змістовий модуль 4	Лабораторна робота №4	7 тиждень	2,5%
	Лабораторна робота №4	8 тиждень	2,5%
	Самостійна робота №4	8 тиждень	3%
Змістовий модуль 5	Лабораторна робота №5	9 тиждень	2,5%
	Лабораторна робота №5	10 тиждень	2,5%
	Самостійна робота №5	10 тиждень	3%
Змістовий модуль 6	Лабораторна робота №6	11 тиждень	2,5%
	Лабораторна робота №6	12 тиждень	2,5%
	Самостійна робота №6	12 тиждень	3%
	Тестове завдання №2	12 тиждень	6%
Підсумковий контроль (тах 40%)			
Завдання заліку (тест)		За розкладом	40%
Разом			100%

Шкала оцінювання: національна та ECTS

За шкалою ECTS	За шкалою університету	За національною шкалою	
		Екзамен	Залік
A	90 – 100 (відмінно)	5 (відмінно)	Зараховано
B	85 – 89 (дуже добре)	4 (добре)	
C	75 – 84 (добре)		
D	70 – 74 (задовільно)	3 (задовільно)	
E	60 – 69 (достатньо)		

ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ
Силабус навчальної дисципліни



FX	35 – 59 (незадовільно – з можливістю повторного складання)	2 (незадовільно)	Не зараховано
F	1 – 34 (незадовільно – з обов’язковим повторним курсом)		



РОЗКЛАД КУРСУ ЗА ТЕМАМИ І КОНТРОЛЬНІ ЗАВДАННЯ

Тиждень і вид заняття	Тема заняття	Контрольний захід	Кількість балів
Змістовий модуль 1. Ймовірнісні розподіли для дискретної випадкової величини			
Тиждень 1 Лекція	Історія мови програмування R . Інтерфейс R . Арифметичні операції, змінні, масиви. Типи змінних. Базові ймовірнісні розподіли для дискретної випадкової величини.		
Тиждень 1 Лабораторне заняття	Установка R . Робота з консоллю. Пакет Stats .	Лабораторна робота №1	
Тиждень 2 Лекція	Функції в R , що реалізовані для розподілів для дискретної випадкової величини.		
Тиждень 2 Лабораторне заняття	Ймовірнісні розподіли для дискретної випадкової величини	Захист лабораторної роботи №1 (разом з самостійною роботою №1)	5
	Самостійна робота №1. Від’ємний біноміальний розподіл в R .		3
Змістовий модуль 2. Ймовірнісні розподіли для неперервної випадкової величини			
Тиждень 3 Лекція	Базові ймовірнісні розподіли для неперервної випадкової величини.		
Тиждень 3 Лабораторна заняття	Функції в R , що реалізовані для розподілів для неперервної випадкової величини.	Лабораторна робота №2	
Тиждень 4 Лекція	Приклади застосування базових ймовірнісних розподілів для неперервної випадкової величини.		
Тиждень 4 Лабораторне заняття	Ймовірнісні розподіли для неперервної випадкової величини	Захист лабораторної роботи №2	5
	Самостійна робота №2. Розподіл Вейбулла.	Захист самостійної роботи №2	3
Змістовий модуль 3. Організація даних: статистичний розподіл вибірки			
Тиждень 5 Лекція	Функції в R , що дозволяють проводити базові статистичні розрахунки.		
Тиждень 5 Лабораторне заняття	Побудова дискретного статистичного розподілу. Побудова інтервального	Лабораторна робота №3	

ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ
Силабус навчальної дисципліни



Тиждень і вид заняття	Тема заняття	Контрольний захід	Кількість балів
	статистичного розподілу. Емпірична функція розподілу.		
Тиждень 6 Лекція	Графічне зображення статистичних розподілів. Розвідувальний аналіз даних.		
Тиждень 6 Лабораторне заняття	Графічне зображення статистичних розподілів. Розвідувальний аналіз даних.	Захист лабораторної роботи №3	5
	Самостійна робота №3. Мозаїчні діаграми	Захист самостійної роботи №3	3
	Тестове завдання №1	Тестування в Moodle	6
Змістовий модуль 4. Парна лінійна регресія. Нелінійні моделі парної регресії.			
Тиждень 7 Лекція	Постановка задачі однофакторної лінійної регресії. Метод найменших квадратів. Перевірка адекватності моделі за допомогою критерію Фішера. Аналіз моделі лінійної регресії. Прогнозування на підставі моделі лінійної регресії. Довірчий інтервал для залежної змінної. Інтервал передбачення для залежної змінної.		
Тиждень 7 Лабораторне заняття	Реалізація парної лінійної регресії у середовищі R . Аналіз моделі лінійної регресії. Прогнозування на підставі моделі лінійної регресії.	Лабораторна робота №4	
Тиждень 8 Лекція	Нелінійні моделі парної регресії. Перехід від нелінійної моделі до лінійної. Аналіз та оцінка регресійної моделі. Оцінка регресійної моделі з використанням спеціальних функцій пакету car .		
Тиждень 8 Лабораторне заняття	Перехід від нелінійної моделі до лінійної. Аналіз та оцінка регресійної моделі. Оцінка регресійної моделі з використанням спеціальних функцій пакету car .	Захист лабораторної роботи №4	5
	Самостійна робота №4. Гетероскедастичність та гомоскедастичність. Тест Дарбіна-Уотсона.	Захист самостійної роботи №4	3
Змістовий модуль 5. Множинна лінійна регресія			
Тиждень 9 Лекція	Множинна лінійна регресія. Діаграми розсіювання. Кореляційна матриця для усіх змінних з набору даних.		
Тиждень 9 Лабораторне заняття	Побудова моделі множинної лінійної регресії.	Лабораторної роботи №5.	
Тиждень 10 Лекція	Аналіз моделі множинної лінійної регресії. Прогнозування на основі моделі.		



Тиждень і вид заняття	Тема заняття	Контрольний захід	Кількість балів
Тиждень 10 Лабораторне заняття	Аналіз моделі множинної лінійної регресії. Прогнозування на основі моделі.	Захист лабораторної роботи №5	5
	Самостійна робота №5. Кореляційна матриця, діаграми розсіювання та згладжуванні розподіли для усіх змінних набору даних.	Захист самостійної роботи №5	3
Змістовий модуль 6. Узагальнені регресійні моделі (Generalized Additive models) у \mathbf{R}			
Тиждень 11 Лекція	Узагальнені регресійні моделі (Generalized Additive models) у \mathbf{R} з використанням пакету mgcv . Щільності розподілу для дискретних та неперервних величин, які є доступними у пакеті mgcv .		
Тиждень 11 Лабораторне заняття	Протокол розвідувального аналізу даних. Вибір закону розподілу емпіричних даних. Аналіз описової статистики та візуалізація даних.	Лабораторна робота №6	
Тиждень 12 Лекція	Етапи побудови узагальненої регресійної моделі: поетапний підхід вперед, поетапний підхід назад. Визначення вхідних змінних, які суттєво впливають на вихідну змінну. Вибір моделі з використанням Інформаційного критерію Акаїке. Критерії перевірки адекватності моделі. Основні діагностичні графіки залишків моделі. Прогнозування.		
Тиждень 12 Лабораторне заняття	Побудова узагальненої регресійної моделі, перевірка її адекватності та прогнозування.	Захист лабораторної роботи №6	5
	Самостійна робота №6. Рандомізовані квантильні залишки.	Захист самостійної роботи №6	3
	Тестове завдання №2	Тестування в Moodle	6

ОСНОВНІ ДЖЕРЕЛА

Книги:

1. Барковський, В.В., Барковська Н.В., Лопатін О.К. Теорія імовірностей та математична статистика : навч. посібник. К. : Центр учбової літератури, 2010. 424 с.
2. Валєєв, К.Г., Джалладова І.А. Збірник задач з теорії ймовірностей та математичної статистики : навч. посіб. К. : КНЕУ, 2006. 352 с.

Інформаційні ресурси:



1. Майборода Р.Є., Сугакова О.В. Аналіз даних за допомогою пакета R: Навчальний посібник. К., 2015. URL: http://matphys.rpd.univ.kiev.ua/downloads/courses/mmatstat/Statistics_with_R.pdf
2. Шипунов А. Б., Балдин Е. М. Анализ данных с R. URL: <https://sociology.knu.ua/sites/default/files/course/materials/r1.pdf>
3. Шипунов А. Б., Балдин Е. М., Волкова П. А., Коробейников А. И., Назарова С. А., Петров С. В., Суфиянов В. Г. Наглядная статистика. Используем R! – 2014. – 296 с. URL: <https://cran.r-project.org/doc/contrib/Shipunov-rbook.pdf>
4. Зарядов И. С. Статистический пакет **R**: теория вероятностей и математическая статистика. М.: Изд-во РУДН, 2010. 141 с. URL: <https://sociology.knu.ua/sites/default/files/library/elopen/zaryadov.pdf>
5. Кабаков Р.И. **R** в действии. Анализ и визуализация данных а программе **R**. М. : ДМК Пресс, 2014. 588 с. URL: https://batrachos.com/sites/default/files/pictures/Books/Kabakov_2014_R_in_action.pdf

РЕГУЛЯЦІЇ І ПОЛІТИКИ КУРСУ²

Відвідування занять. Регуляція пропусків.

Відвідування усіх занять є обов'язковим. У разі поважної причини відсутності студента на занятті, його потрібно відпрацювати під час поточних контрольних заходів і при виконання індивідуального завдання. Контрольні заходи, які пропущено з поважних причин відпрацьовуються на консультаціях відповідно до часу, зазначеного на початку даного Силабусу.

Політика академічної доброчесності

Кожний студент зобов'язаний дотримуватися принципів академічної доброчесності.

Використання комп'ютерів/телефонів на занятті

Використання технічних засобів (мобільних телефонів, ноутбуків, планшетів та інших персональних гаджетів) під час лекційних і практичних занять дозволено в навчальних цілях. Мобільні телефони під час занять повинні бути переведені в режим «без звуку».

Під час проведення заходів поточного і підсумкового контролю використання власних технічних засобів заборонено. У разі їх виявлення результат оцінюється в 0 балів.

Комунікація

У разі очного навчання комунікація студентів з викладачем здійснюється під час аудиторних занять і на консультаціях. При дистанційному навчанні та при очному за потреби – через Viber (група з дисципліни, приватні повідомлення відповідно до зазначеного на початку номеру телефону), Moodle (форум курсу, приватні повідомлення).

² Тут зазначається все, що важливо для курсу: наприклад, умови допуску до лабораторій, реактивів тощо. Викладач сам вирішує, що треба знати студенту для успішного проходження курсу!



Повідомлення про терміни тестування, про дистанційні групові заняття, консультації з кодами доступу для конференцій Zoom надсилаються в групу з дисципліни Viber та/або на приватну пошту всім студентам (розсилка).

Виконані індивідуальні завдання, викладені студентом на платформу Moodle **вчасно** – у зазначений термін та перевіряються викладачем протягом 3 робочих днів. Якщо завдання надсилається невчасно, то його терміни перевірки не дотримуються. На інші запити викладач відповідає протягом 3 робочих днів.



ДОДАТОК ДО СИЛАБУСУ ЗНУ – 2020-2021 рр.

ГРАФІК НАВЧАЛЬНОГО ПРОЦЕСУ 2020-2021 н. р. (посилання на сторінку сайту ЗНУ)

АКАДЕМІЧНА ДОБРОЧЕСНІСТЬ. Студенти і викладачі Запорізького національного університету несуть персональну відповідальність за дотримання принципів академічної доброчесності, затверджених **Кодексом академічної доброчесності ЗНУ**: <https://tinyurl.com/ya6yk4ad>. Декларація академічної доброчесності здобувача вищої освіти (додається в обов'язковому порядку до письмових кваліфікаційних робіт, виконаних здобувачем, та засвідчується особистим підписом): <https://tinyurl.com/y6wzzlu3>.

НАВЧАЛЬНИЙ ПРОЦЕС ТА ЗАБЕЗПЕЧЕННЯ ЯКОСТІ ОСВІТИ. Перевірка набутих студентами знань, навичок та вмінь (атестації, заліки, іспити та інші форми контролю) є невід'ємною складовою системи забезпечення якості освіти і проводиться відповідно до *Положення про організацію та методик проведення поточного та підсумкового семестрового контролю навчання студентів ЗНУ*: <https://tinyurl.com/y9tve4lk>.

ПОВТОРНЕ ВИВЧЕННЯ ДИСЦИПЛІН, ВІДРАХУВАННЯ. Наявність академічної заборгованості до 6 навчальних дисциплін (в тому числі проходження практики чи виконання курсової роботи) за результатами однієї екзаменаційної сесії є підставою для надання студенту права на повторне вивчення зазначених навчальних дисциплін. Порядок повторного вивчення визначається *Положенням про порядок повторного вивчення навчальних дисциплін та повторного навчання у ЗНУ*: <https://tinyurl.com/y9pkmmp5>. Підстави та процедури відрахування студентів, у тому числі за невиконання навчального плану, регламентуються *Положенням про порядок переведення, відрахування та поновлення студентів у ЗНУ*: <https://tinyurl.com/ycds57la>.

НЕФОРМАЛЬНА ОСВІТА. Порядок зарахування результатів навчання, підтверджених сертифікатами, свідоцтвами, іншими документами, здобутими поза основним місцем навчання, регулюється *Положенням про порядок визнання результатів навчання, отриманих у неформальній освіті*: <https://tinyurl.com/y8gbt4xs>.

ВИРІШЕННЯ КОНФЛІКТІВ. Порядок і процедури врегулювання конфліктів, пов'язаних із корупційними діями, зіткненням інтересів, різними формами дискримінації, сексуальними домаганнями, міжособистісними стосунками та іншими ситуаціями, що можуть виникнути під час навчання, регламентуються *Положенням про порядок і процедури вирішення конфліктних ситуацій у ЗНУ*: <https://tinyurl.com/y9yfw9v>. Конфліктні ситуації, що виникають у сфері стипендіального забезпечення здобувачів вищої освіти, вирішуються стипендіальними комісіями факультетів, коледжів та університету в межах їх повноважень, відповідно до: *Положення про порядок призначення і виплати академічних стипендій у ЗНУ*: <https://tinyurl.com/yd6bq6p9>; *Положення про призначення та виплату соціальних стипендій у ЗНУ*: <https://tinyurl.com/y9r5dpwh>.

ПСИХОЛОГІЧНА ДОПОМОГА. Телефон довіри практичного психолога (061)228-15-84 (щоденно з 9 до 21).

ЗАПОБІГАННЯ КОРУПЦІЇ. Уповноважена особа з питань запобігання та виявлення корупції (Воронков В. В., 1 корп., 29 каб., тел. +38 (061) 289-14-18).

РІВНІ МОЖЛИВОСТІ ТА ІНКЛЮЗИВНЕ ОСВІТНЄ СЕРЕДОВИЩЕ. Центральні входи усіх навчальних корпусів ЗНУ обладнані пандусами для забезпечення доступу осіб з інвалідністю та інших маломобільних груп населення. Допомога для здійснення входу у разі потреби надається черговими охоронцями навчальних корпусів. Якщо вам потрібна спеціалізована допомога, будь-ласка, зателефонуйте (061) 228-75-11 (начальник охорони). Порядок супроводу (надання допомоги) осіб з інвалідністю та інших маломобільних груп населення у ЗНУ: <https://tinyurl.com/ydhcsagx>.

РЕСУРСИ ДЛЯ НАВЧАННЯ. Наукова бібліотека: <http://library.znu.edu.ua>. Графік роботи абонементів: понеділок – п'ятниця з 08.00 до 17.00; субота з 09.00 до 15.00.

ЕЛЕКТРОННЕ ЗАБЕЗПЕЧЕННЯ НАВЧАННЯ (MOODLE): <https://moodle.znu.edu.ua>

Якщо забули пароль/логін, направте листа з темою «Забув пароль/логін» за адресами:

- для студентів ЗНУ - moodle.znu@gmail.com, Савченко Тетяна Володимирівна
- для студентів Інженерного інституту ЗНУ - alexvasik54@gmail.com, Василенко Олексій Володимирович

У листі вкажіть: прізвище, ім'я, по-батькові українською мовою; шифр групи; електронну адресу.

Якщо ви вказували електронну адресу в профілі системи Moodle ЗНУ, то використовуйте посилання для відновлення паролю <https://moodle.znu.edu.ua/mod/page/view.php?id=133015>.

Центр інтенсивного вивчення іноземних мов: <http://sites.znu.edu.ua/child-advance/>

Центр німецької мови, партнер Гете-інституту: <https://www.znu.edu.ua/ukr/edu/ocnu/nim>

Школа Конфуція (вивчення китайської мови): <http://sites.znu.edu.ua/confucius>