


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Механіко-математичний факультет
Кафедра математичної статистики і диференціальних рівнянь

Затверджено

На засіданні кафедри математичної
статистики
і диференціальних рівнянь
механіко-математичного факультету
Львівського національного університету
імені Івана Франка
(протокол №1 від 28.08.2020 р.)

В.о. завідувача кафедри:

 Олег БУГРІЙ

Силабус з навчальної дисципліни
“Аналіз і візуалізація даних на мові R”,
що викладається в межах ОПШ
“ Статистичний аналіз даних ”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 112 - Статистика

Львів 2020 р.

Назва дисципліни	Аналіз і візуалізація даних на мові R
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Факультет та кафедра, за якою закріплена дисципліна	Механіко-математичний факультет Кафедра математичної статистики і диференціальних рівнянь
Галузь знань, шифр та назва спеціальності	11 - Математика та статистика 112 – Статистика
Викладачі дисципліни	Базилевич Ірина Богданівна, кандидат фізико-математичних наук, доцент, жоцент кафедри математичної статистики і диференціальних рівнянь
Контактна інформація викладачів	iryna.bazylevych@lnu.edu.ua , http://new.mmf.lnu.edu.ua/employee/bazylevych_i_b ; Головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю): ауд. 267, головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Інформація про дисципліну	Дисципліна “Аналіз і візуалізація даних на мові R” є вибірковою дисципліною з спеціальності 112-Статистика для освітньої програми “Статистичний аналіз даних”, яка викладається в 4-му семестрі в обсязі 3,5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Курс розроблено для ознайомлення студентів з мовою програмування R. Це мова, яку використовують статистики для аналізу даних. Вона виникла в 1993 році, є безкоштовною і має великі можливості. Її ще називають мовою статистичного програмування.
Мета та цілі дисципліни	<i>Мета:</i> надати студентам поняття та методи роботи в R <i>Цілі:</i> забезпечити знайомство студентів з властивостями мови програмування R.
Література для вивчення дисципліни	<ol style="list-style-type: none"> 1. В. Гнатюк Вступ до R на прикладах. В-во Харківського Національного Економічного Університету, Харків – 2010 2. Р. Майборода. Комп'ютерна статистика.- К.: 2016. 3. Р. Є. Майборода Комп'ютерна статистика : підручник – К. : 4. ВПЦ "Київський університет", 2019. – 589 с. 5. Майборода Р.Є. "Самостійна робота по курсу "Асимптотична статистика"" , - 2019 6. Р. Майборода, О. Сугакова. Аналіз даних за допомогою пакета R.- К:2015. 7. Braun W.J., Murdoch D.J. A First Course in Statistical Programming with R. Cambridge University Press, 2007 8. Bretz F., Hothorn T., Westfall P. Multiple Comparisons Using R. Chapman, Hall/CRC, 2011 -194. 9. Dalgaard P. Introductory statistics with R. 2 ed. — USA: Springer Science Business Media, 2008. — 363 p 10. An Introduction to Statistical Learning: with Applications in R / G. James .

	<p>Springer, 2013- 430 p.</p> <p>11. Højsgaard S., Edwards D., Lauritzen S. Graphical Models with R. Springer, 2012.- 182 p.</p> <p>12. J.D. Long, P. Teetor. R Cookbook. Proven Recipes for Data Analysis, Statistics and Graphics. – ISBN 978-1-492-04068-2. - 2019</p> <p>13. Matloff N. The Art of R Programming: A Tour of Statistical Software Desing, ISBN 978-1-59327-2, published by No Starch Ptes, 2011</p> <p>14. Zumel N., Mount J. Practical Data Science with R. Manning Publications, 2014. -416 p.</p>
Обсяг курсу	Загальний обсяг: 105 годин. Аудиторних занять: 48 год., з них 32 год. лекційних та 16 годин практичних занять. Самостійної роботи: 57 год.
Очікувані результати навчання	У результаті вивчення даного курсу студент повинен: знати: основні операції мови програмування R вміти: застосовувати мову програмування R до розв’язання математичних задач, аналізувати статистичні дані з допомогою мови R.
Ключові слова	Код, операція, умовні оператори, оператори типу «apply», R-studio
Формат курсу	Очний
Теми	Див. Схема курсу
Підсумковий контроль, форма	Залік у кінці семестру
Пререквізити	Для вивчення даного курсу студентам потрібні базові знання з: - математичного аналізу ; -алгебри.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація); дедуктивні методи на основі узагальнень; евристичні методи (проблемна лекція); інтерактивні методи (дискусія)
Необхідне обладнання	Комп’ютер із необхідним програмним забезпеченням, доступ до Internet мережі.
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням:</p> <ul style="list-style-type: none"> • Змістовий модуль 1: 35 балів – 14 балів за виконання лабораторних робіт, 11 балів – контрольна робота • Змістовий модуль 2: 35 балів – 14 балів за виконання лабораторних робіт, 11 балів – контрольна робота • Контрольне оцінювання: 30% семестрової оцінки, максимальна кількість балів 30. <p>Підсумкова максимальна кількість балів 100.</p> <p>Академічна доброчесність: Очікується, що роботи студентів будуть оригінальними дослідженнями чи міркуваннями. Списування та втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в написанні завдань є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі</p>

студенти відвідають усі лекції та практичні 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, передбачених курсом.

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.

Політика виставлення балів. Враховуються бали, набрані при поточному контролі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; 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.

Оцінювання практичних робіт. Бали оцінювання аудиторного виконання практичних робіт нараховуються за наступним співвідношенням:

100% – студент в повному обсязі володіє навчальним матеріалом, має повне розуміння розглянутої теми, надає правильні відповіді на запитання по темі, код програми функціонує відповідно до завдання;

75% – студент достатньо розуміє розглянутий матеріал та принципи написаного ним коду програми, присутні неточності та незначні помилки у відповідях на запитання по темі, код програми функціонує відповідно до завдання;

50% – студент не досить добре розуміє розглянутий матеріал та написаний ним код програми, вагається та надає неточні/не конкретні відповіді на запитання по темі, код програми функціонує з помірними недоліками;

25% – студент погано розуміє розглянутий матеріал та написаний ним код програми, студент в більшості надає помилкові відповіді на питання по темі, код програми не функціонує належним чином;

0 - студент зовсім не засвоїв розглянутий матеріал, написаний ним код програми не відповідає темі/не функціонує взагалі.

Бали оцінювання домашнього завершення виконання практичних робіт та наданого звіту нараховуються за наступним співвідношенням:

100% – звіт цілком і повністю відображає індивідуальне завдання студента, містить правильні висновки, ілюстрований (за потреби) відповідними графіками і таблицями які правильно відображають суть виконаного завдання, студент має повне розуміння розглянутої теми, надає правильні відповіді на запитання по темі, код програми функціонує відповідно до завдання;

75% – звіт в достатній мірі відображає індивідуальне завдання студента, містить допустимі висновки, ілюстрований (за потреби) відповідними графіками і таблицями які частково відображають суть виконаного завдання, студент достатньо розуміє принципи написаного ним коду програми, присутні неточності та незначні помилки у відповідях на запитання по темі, код програми функціонує відповідно до завдання;

50% – звіт містить загальні формулювання завдання, висновки нечіткі, необхідні ілюстрації чи таблиці відсутні, студент не досить добре розуміє

	<p>розглянутий матеріал та представлений код програми, надає неточні/не конкретні відповіді на запитання по темі, код програми функціонує з помірними недоліками;</p> <p>25% – звіт не містить формулювання завдання, висновки необґрунтовані чи неповні, необхідні ілюстрації чи таблиці відсутні, студент погано розуміє розглянутий матеріал та представлений код програми, студент в більшості надає помилкові відповіді на питання по темі, код програми не функціонує належним чином;</p> <p>0 – звіт відсутній/не відповідає темі, студент зовсім не засвоїв розглянутий матеріал, написаний ним код програми не відповідає темі/не функціонує взагалі.</p> <p>Оцінювання контрольного тестування (завдання з тематики кожного змістового модуля) відбувається шляхом оцінки письмових відповідей студента на поставлені запитання.</p> <p>Відсотки нарахування балів оцінювання відповіді на кожне запитання нараховуються за наступним співвідношенням:</p> <p>75-100% – тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно;</p> <p>50-75% – відтворюється значна частина розглянутої теми, проте присутні неточності та/або невідповідності;</p> <p>25-50% – виявлено множинні неточності та невідповідності, пояснення відсутні чи частково помилкові;</p> <p>0-25% – тему майже не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи.</p> <p>Критерії оцінювання результатів неформальної освіти:</p> <p>Нарахування балів відбувається за публікацію студентом тез доповідей на конференціях, наукових статей, за участь студента у діяльності наукових гуртків, семінарів, круглих столів, конкурсів, участь у заходах неформальної освіти, за отримання сертифікатів про проходження навчання на різних освітніх платформах (Coursera, Prometheus тощо), курсах провідних ІТ компаній за тематикою навчальної дисципліни. Кількість балів визначається відсотком покриття результатів відповідної активності до вимог результатів навчання з навчальної дисципліни</p>
<p>Питання для оцінювання</p>	<ol style="list-style-type: none"> 1. Арифметичні дії в \mathbb{R}. 2. Логічні операції в \mathbb{R}. 3. Присвоєння значення об'єктам. 4. Елементарні функції в \mathbb{R} 5. Статистичні операції в \mathbb{R}. 6. Вектори. Задання векторів. 7. З'єднання двох вектор в одну структуру. 8. Послідовності. Команда, яка створює послідовності. 9. За допомогою логічних операцій виділення за певними критеріями членів послідовності (вектора). 10. Списки. 11. Матриці 12. Визначення розмірності матриці 13. Знаходження власних значень і власних векторів матриці. 14. Додавання матриць. 15. Множення матриць.

16. Операція `diag()`.
17. Фрейми даних
18. Читання та редагування даних.(у фреймі)
19. Робота з розподілами.(генерування вибірки, знаходження квантилів розподілу, знаходження функції розподілу, знаходження щільності розподілу).
20. Основні розподіли і їх позначення в R.
21. Знаходження гістограми в R.
22. Побудова графіків в R.
23. Побудова вусатої коробочки в R.
24. Побудова матричної діаграми.
25. Побудова P-P діаграми.
26. Побудова Q-Q діаграми.
27. Зображення точок на графіку.
28. Команда **`text(x, y, labels), cex, pch-`**.
29. Лінійна регресія: випадок однієї пояснювальної змінної, випадок багатьох пояснювальних змінних.
30. Критерій χ^2 -квадрат перевірки простої гіпотези.
31. Критерій χ^2 -квадрат перевірки гіпотези про незалежність.
32. Критерій Колмогорова.
33. Значення нескінченності та невизначеності в R.
34. Функція `гер()`.
35. Індксація векторів.
36. Фактори.
37. Функція `sort`
38. Функції `sum, prod`.
39. Функція `outer`
40. Експорт, імпорт даних.
41. Створення власних функцій.
42. Структури управління `for, If, else`.
43. Вибір з кількох умов `switch`.
44. Цикли `while, repeat`.
45. Стовпцеві та кругові діаграми.
46. Точки та лінії на площині.
47. Елементи тривимірної графіки.
48. Географічні карти.
49. Вибіркова медіана.
50. Середньоквадратичне відхилення.
51. Функції для підрахунку описових статистик
52. Генерація вибірки.
53. Гістограми.
54. Вектори, фактори.
55. Матриці і масиви.
56. Списки.

	<p>57. Логічні типи даних і оператори.</p> <p>58. Експорт та імпорт даних.</p> <p>59. Перенаправлення даних з екрану у файл.</p> <p>60. Функція scan.</p> <p>61. Імпорт даних з ECSEL, SPSS.</p> <p>62. Введення даних з клавіатури.</p> <p>63. Отримання інформації про об'єкти.</p> <p>64. NA, NaN.</p> <p>65. Inf.</p> <p>66. NULL.</p> <p>67. Кодування значень.</p> <p>68. Виключення відсутніх значень з аналізу.</p> <p>69. Функції apply, sapply, replicate, tapply, rapply, by, outer.</p> <p>70. Регресійна модель.</p> <p>71. Функція plot().</p> <p>72. Лінійні графіки.</p> <p>73. Гістограми (Гнатюк)</p> <p>74. Злиття двох наборів</p> <p>75. Експорт, імпорт.</p> <p>76. Переформатовування</p> <p>77. Програмування в R.</p> <p>78. Векторизація функцій.</p> <p>79. Структури управління.</p> <p>80. Кореляційна матриця.</p> <p>81. Візуалізація кореляцій.</p> <p>82. Ранги та рангові кореляції.</p> <p>83. Множини в мові R.</p>
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

Схема курсу
“ Аналіз і візуалізація даних на мові R ”
для студентів спеціальності 112 – Статистика

Тижні	Лекційний курс		Практичні заняття		К-сть год СР	Література
	Назва теми	К-сть год	Назва теми	К-сть год		
1	2	3	4	5	6	7
1	Ознайомлення з мовою програмування R. Основні арифметичні операції, основні функції в R.	2	Встановлення на комп'ютер мови R, а також R-STUDIO. Ознайомлення з консоллю. Виконання основних дій. Основні математичні функції в R	1	4	[1],[2], [3], [4], [7],[8],[9],[12] Сайт курсу
2	Вектори, послідовності, списки в R	2	Вектори, послідовності списки	1	4	[1],[2], [3],[4],[12],[13] Сайт курсу
3	Матриці, масиви, фрейми в R	2	Матриці, масиви, фрейми	1	4	[1],[2], [7],[13] Сайт курсу
4	Ймовірність, розподіли в R, основні статистичні операції	2	Розподіли в R, розв'язування ймовірнісних задач за допомогою R, основні статистичні операції в мові R	1	3	[2],[3],[4], [5], [12] Сайт курсу
5	Умовні оператори	2	Умовні оператори	1	4	[1][2],[4] Сайт курсу
6	Фактори. Поділ відрізка на частини	2	Фактори. Поділ відрізка на частини. Створення статистичного ряду	1	3	[3], [13] Сайт курсу
7	Програмування в R. Написання власних функцій	2	Програмування в мові R. Створення власних функцій	1	4	[1],[2],[4], [5], [12] Сайт курсу
8	Графіка в R. Зображення точок, ліній	2	Контрольна робота 1	1	3	[1],[3],[4],[11], [12],[13] Сайт курсу
9	Діаграми в R.	2	Зображення точок, ліній, діаграм. Візуалізація даних	1	4	[3],[12] Сайт курсу
10	Контурні графіки	2	Контурні графіки. Розбиття на вікна.	1	3	[3] Сайт курсу
11	Експорт, імпорт, зберігання даних	2	Зберігання даних в R. Імпорт з EXCEL, SPSS	1	4	[1], [12] Сайт курсу
12	Тривимірна графіка. Географічні карти	2	Тривимірна графіка. Географічні карти.	1	3	[3],[10],[12], [14] Сайт курсу
13	Статистичні гіпотези в мові R	2	Статистичні гіпотези	1	4	[2],[3],[4] Сайт курсу
14	Регресійний аналіз в мові R	2	Регресійний аналіз	1	3	[3], [6], [12] Сайт курсу

15	Часові ряди в мові R	2	Часові ряди	1	4	[12] Сайт курсу
16	Множини в мові R	2	Контрольна робота 2	1	3	[13] Сайт курсу
	Разом	32		16	57	
	Викладач: Базилевич І.Б.		Викладач: Базилевич І.Б.			