

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

«ЗАТВЕРДЖУЮ»
декан факультету інформаційних
технологій
Глазунова О.Г.
«16» травня 2022 р.

Протокол №10 від 19 травня 2022 р.

"СХВАЛЕНО"
на засіданні кафедри економічної кібернетики
Протокол №10 від 06 травня 2022 р.

Завідувач кафедри
Дмитро ЖЕРЛІЦІН

"РОЗГЛЯНУТО"
Гарант ОПП «Цифрова економіка»

Дмитро ЖЕРЛІЦІН

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Аналітика з R»

Галузь знань **05 «Соціальні та поведінкові науки»**
Спеціальність **051 «Економіка»**
Освітньо-професійна програма **«Цифрова економіка»**

Розробник:

доцент кафедри економічної кібернетики, к.е.н., доцент Негрей М.В.

Київ – 2022 р.

1. Опис навчальної дисципліни «Аналітика з R»

Галузь знань, напрям підготовки, спеціальність, ступень вищої освіти	
Освітній ступінь	Бакалавр
Галузь знань	05 Соціальні та поведінкові науки;
Спеціальність	051 Економіка;
Освітня програма	Цифрова економіка
Характеристика навчальної дисципліни	
Вид	Обов'язкова складова
Загальна кількість годин	150
Кількість кредитів ECTS	5
Кількість змістовних модулів	2
Курсовий проект (робота) (за наявності)	Не передбачено
Форма контролю	Екзамен
Показники навчальної дисципліни	
Форма навчання	Денна
Рік підготовки	3
Семестр	5
Лекційні заняття	30
Практичні, семінарські заняття	
Лабораторні заняття	30
Самостійна робота студента	90
Індивідуальні завдання	
Кількість тижневих аудиторних годин для денної форми навчання	4 год.

2. Мета, завдання та компетентності навчальної дисципліни

Мета: формування системи теоретичних знань і практичних навичок аналізу та статистичних розрахунків в середовищі R; опанування основних принципів роботи та базових методів і моделей у середовищі R.

Завдання вивчення курсу: вивчення основних принципів роботи в середовищі R, орієнтованих на застосування сучасних наукових методів, моделей та засобів інформаційних технологій; набуття вмінь створення, використання й адаптації моделей в R.

Засвоївши курс студент повинен:

знати:

- основні принципи організації середовища R,
- функціонал роботи R,
- основи статистичних розрахунків в R,
- підходи до візуалізації в R,
- принципи побудови моделей з R,
- алгоритм побудови економетричних моделей в R,
- використання моделей в управлінні економічними процесами;

вміти:

- працювати із даними в R;
- графічно представляти дані в R;
- застосувати функціонал роботи R для розв'язання практичних задач;
- самостійно здійснювати підбір інформації необхідної для вирішення поставлених задач, аналізувати отримані результати;

володіти

- методами аналізу статистичних даних,
- використовувати сучасне програмне забезпечення.

Набуття компетентностей:

загальні компетентності (ЗК):

ЗК7. Навички використання інформаційних і комунікаційних технологій.

фахові (спеціальні) компетентності (ФК):

СК8. Здатність аналізувати та розв'язувати завдання у сфері економічних та соціально-трудових відносин.

СК15. Уміння використовувати сучасні технології, інтелектуальні методи, цифрові та програмні інструменти з обробки й аналізу даних.

У результаті вивчення навчальної дисципліни студент повинен показати певні програмні результати, а саме

ПРН13. Ідентифікувати джерела та розуміти методологію визначення і методи отримання соціально-економічних даних, збирати та аналізувати необхідну інформацію, розраховувати економічні та соціальні показники.

3. Програма навчальної дисципліни

Змістовий модуль 1. Основи аналітичної роботи з R

Тема 1. Основи аналізу даних та сучасні інформаційні технології їх оброблення.

Сучасні особливості підготовки та первинного оброблення аналітичної інформації. Пакети аналізу даних. Підходи щодо аналізу великих даних та представлення результатів аналізу. Застосування сучасних методів аналізу в управлінні соціально-економічними процесами.

Тема 2. Підготовка та оброблення даних з R. Робота з Data.frame.

Основні принципи організації середовища R. Робота з командною консоллю інтерфейсу R. Робота з меню пакету R commander. Об'єкти, пакети, функції, пристрої. Типи даних мови R. Вектори і матриці. Фактори. Списки і таблиці. Імпортування даних в R. Представлення дати і часу. Часові ряди. Організація обчислень: функції, відгалуження, цикли. Робота з даними data.frame.

Тема 3. Описова статистика і нормальний закон розподілу в R.

Пакет psych. Гіпотеза про рівність середніх двох генеральних сукупностей. Використання рангових критеріїв. Гіпотеза про однорідність дисперсії. Дисперсійний аналіз. Оцінка кореляції двох випадкових величин. Критерій хі-квадрат. Тест Фішера.

Тема 4. Випадкові змінні та інструменти програмування в R.

Оцінка вибірових параметрів з використанням спеціальних функцій. Використання вбудованих та власних функцій, додаткових пакетів. Аналіз викидів. Заповнення пропущених значень у таблицях даних. Відтворення результатів при використанні генератора випадкових чисел. Закони розподілу, реалізовані в R. Підбір закону і параметрів розподілу в R. Перевірка на нормальність розподілу.

Змістовий модуль 2. Аналітичні інструменти з мовою програмування R

Тема 5. Графічний аналіз з R.

Діаграми розсіювання `plot()` і параметри графічних функцій. Гістограми і функція `cdplot()`. Діаграми розмахів. Кругові і стовпчикові діаграми. Діаграми Клівленда і одномірні діаграми розсіювання. Категоризовані графіки. Пакет `ggplot2` та побудова складних діаграм.

Тема 6. Регресійний аналіз з R.

Проста лінійна регресія. Поліноміальні і нелінійні моделі регресії. Множинна регресія. Моделі згладжування. Узагальнені моделі регресії. Логістична регресія. Коваріаційний аналіз. Індуктивні моделі. Моделювання структурних рівнянь.

Тема 7. Специфічні аналітичні інструменти, функції та пакети R.

Створення та робота з інструментами створення звітів R Markdown. Створення звітів у форматі `.html`. Робота з форматом даних `data.table`. Бібліотека `data.table`. Відбір, фільтрування та групування даних `data.table`. Систематизація інших аналітичних пакетів R.

Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин					
	усього	у тому числі				
		л	п	лаб.	інд.	с. р.
1	2	3	4	5	6	7
Змістовий модуль 1. Основи аналітичної роботи з R						
Тема 1. Основи аналізу даних та сучасні інформаційні технології їх оброблення.	18	4	4			10
Тема 2. Підготовка та оброблення даних з R. Робота з data.frame.	18	4	4			10
Тема 3. Описова статистика і нормальний закон розподілу в R.	18	4	4			10
Тема 4. Випадкові змінні та інструменти програмування в R.	21	3	3			15
Разом за змістовим модулем 1	75	15	15			45
Змістовий модуль 2. Аналітичні інструменти з мовою програмування R						
Тема 5. Графічний аналіз з R.	25	5	5			15
Тема 6. Регресійний аналіз з R	25	5	5			15
Тема 7. Специфічні аналітичні інструменти, функції та пакети R	25	5	5			15
Разом за змістовим модулем 2	75	15	15			45
Усього годин	150	30	30			90

6. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1.	Встановлення R. Ознайомлення з основними компонентами R	4
2.	Збирання та первинне оброблення даних	4
3.	Робота із таблицями. Аналіз статистичних даних	4
4.	Побудова та аналіз законів розподілу в R	3
5.	Побудова графіків в R	5
6.	Побудова регресійних моделей в R	5
7.	Робота з R Markdown	5
	Разом	30

Самостійна робота

№ з/П	Назва теми	Кількість годин
1	Основи аналізу даних та сучасні інформаційні технології їх оброблення.	10
2	Підготовка та оброблення даних з R. Робота з data.frame.	10
3	Описова статистика і нормальний закон розподілу в R.	10
4	Випадкові змінні та інструменти програмування в R.	15
5	Графічний аналіз з R.	15
6	Регресійний аналіз з R	15
7	Специфічні аналітичні інструменти, функції та пакети R	15
	Разом	90

7. Контрольні питання, комплекти тестів для визначення рівня засвоєння знань студентами.

Перелік питання для визначення рівня засвоєння знань студентами

1. Історія виникнення R.
2. Основні принципи організації середовища R.
3. Робота з командною консоллю інтерфейсу R.
4. Робота з меню пакету R commander.
5. Об'єкти, пакети, функції, пристрої.
6. Типи даних мови R.
7. Вектори і матриці.
8. Факторні змінні.
9. Списки і таблиці.
10. Data.Frame
11. Об'єднання та розділення таблиць Data.Frame.
12. Групування даних Data.Frame.
13. Бібліотека psych
14. Бібліотека
15. Імпортування даних в R.
16. Представлення дати і часу.
17. Часові ряди.
18. Організація обчислень: функції.
19. Організація обчислень: відгалуження.
20. Організація обчислень: цикли.
21. Векторизовані обчислення в R з використанням apply-функцій.
22. Оцінка вибірових параметрів з використанням спеціальних функцій.
23. Використання функцій summary () і додаткових пакетів.

24. Аналіз викидів.
25. Заповнення пропущених значень в таблицях даних.
26. Відтворення результатів при використанні генератора випадкових чисел.
27. Закони розподілу, реалізовані в R.
28. Підбір закону і параметрів розподілу в R.
29. Перевірка на нормальність розподілу.
30. Гіпотеза про рівність середніх двох генеральних сукупностей.
31. Використання рангових критеріїв.
32. Гіпотеза про однорідність дисперсій.
33. Дисперсійний аналіз.
34. Оцінка кореляції двох випадкових величин.
35. Критерій χ^2 -квадрат.
36. Тест Фішера.
37. Діаграми розсіювання `plot()` і параметри графічних функцій.
38. Гістограми і функція `qplot()`.
39. Бібліотека `ggplot2`.
40. Діаграми розмахів.
41. Кругові і стовпчикові діаграми.
42. Одномірні діаграми розсіювання.
43. Категоризовані графіки.
44. Проста лінійна регресія.
45. Поліноміальні моделі регресії.
46. Нелінійні моделі регресії.
47. Множинна регресія.
48. Моделі згладжування.
49. Узагальнені моделі регресії.
50. Моделі пробіт- і логіт- регресії.
51. Коваріаційний аналіз.
52. Звіти R Markdown
53. Робота з пакетом `data.table`

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

ОС Бакалавр Спеціальність 051 Економіка (Економічна кібернетики)	Кафедра Економічної кібернетики 2022-2023 н.р.	Екзаменаційний білет №7 з дисципліни «Аналітика з R»	Затверджую: Зав кафедри проф.Д. Жерліцин _____ 2022 р.
1. Екзаменаційне питання (максимальна оцінка 10 балів за відповідь)			
Розподіл Пуассона в R			
2. Задача (максимальна оцінка 10 балів за розв'язання задачі)			
Використовуючи дані із отриманого варіанту, виконати наступні завдання:			
1. Оцінити параметри лінійної регресії. 2. Обчислити стандартну похибку та побудувати 95% довірчі інтервали для коефіцієнтів регресії. 3. Перевірити нульову гіпотезу про відсутність зв'язку між X та Y. 4. Провести ANOVA-аналіз регресійної моделі.			

5. Визначити коефіцієнт детермінації, стандартну похибку моделі.	
6. Записати регресійну модель. Зобразити графічно.	
3. Тестові завдання різних типів (максимальна оцінка 10 балів за відповідь на всі тестові завдання) за посиланням: https://elearn.nubip.edu.ua/mod/quiz/view.php?id=129560	
Екзаменатор	доц. Негрей М.В.

8. Методи навчання

В процесі викладання навчальної дисципліни за характером пізнавальної діяльності застосовуються переважно методи гейміфікації та пояснювально-ілюстративний, евристичний методи, метод демонстраційних прикладів, методи парної та групової роботи, а також частково кожен із зазначених методів залежно від видів робіт на занятті. (Табл. 1).

Таблиця 1

Класифікація методів навчання

Засади	Найменування	Характеристики
1. Джерело знань: слово образ досвід	словесні, наочні, практичні	
2. Етапи навчання	підготовка до вивчення нового матеріалу, вивчення нового матеріалу, закріплення вправ, контроль і оцінка	
3. Спосіб педагогічного керівництва	пояснення педагога, самостійна робота	керівництво: безпосереднє; опосередковане
4. Стиль викладання (пояснення)	інформаційно-повідомлювальний, пояснювальний, інструктивно-практичний, пояснювально-спонукальний	
4. Логіка навчання	індуктивні, дедуктивні, аналітичні, синтетичні	
5. Дидактичні цілі	організація навчальної діяльності, стимулювання і релаксація, контроль і оцінка	
6. Дидактичні завдання	методи оволодіння знаннями, методи формування умінь і навичок, застосування здобутих знань, умінь і навичок	
7. Характер пізнавальної діяльності	пояснювально- ілюстративні, репродуктивні проблемного викладу, частково-пошукові (евристичні), дослідницькі методи	репродуктивні продуктивні

При викладанні навчальної дисципліни використовуються такі методи навчання:

- М1. Лекція (дискусія, проблемна)
- М2. Лабораторна робота
- М3. Проблемне навчання
- М4. Проектне навчання(індивідуальне, малі групи, групове)
- М5. Он-лайн навчання

Та методи контролю:

- МК1. Тестування
- МК2. Контрольне завдання
- МК3. Розрахункова робота
- МК4. Методи усного контролю
- МК5. Екзамен

9. Форми контролю

Кожна з форм контролю має особливості й залежить від мети, змісту та характеру навчання. У процесі навчання дисципліни використовуються наступні форми контролю:

- **Поточний контроль:** усне опитування (індивідуальне, фронтальне, групове), комп'ютерне тестування, виконання практичних завдань на комп'ютері згідно програми;
- **Підсумковий контроль:** тестування та співбесіда за результатами роботи.

10. Розподіл балів, які отримують студенти. Оцінювання знань студента відбувається за 100-бальною шкалою і переводиться в національні оцінки згідно з табл. 1 «Положення про екзамени та заліки у НУБіП України» (наказ про введення в дію від 27.12.2019 р. № 1371)

Поточний контроль		Рейтинг з навчальної роботи R_{HP}	Рейтинг з додаткової роботи $R_{ДР}$	Рейтинг штрафний $R_{ШТР}$	Підсумкова атестація (залік)	Загальна кількість балів
Змістовий модуль 1	Змістовий модуль 2					
0-100	0-100	0-70	0-20	0-5	0-30	0-100

Примітки.

Рейтинг з поточної роботи (навчальної, додаткової), штрафний рейтинг та показники підсумкової атестації визначаються відповідно п. 2 та п.4 Положення про екзамени та заліки у Національному університеті біоресурсів і природокористування України, що затверджене Вченою радою НУБіП України « 27 » грудня 2019 р. протокол №5.

Для визначення рейтингу студента (слухача) із засвоєння дисципліни $R_{\text{дис}}$ (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи $R_{\text{НР}}$ (до 70 балів): $R_{\text{дис}} = R_{\text{НР}} + R_{\text{АТ}}$.

Загальний рейтинг здобувача вищої освіти із засвоєння дисципліни у балах переводиться у національної оцінки наступним чином:

Шкала оцінювання: національна та ECTS

Рейтинг здобувача вищої освіти, бали	Оцінка національна за результати складання екзаменів заліків	
	екзаменів	заліків
90-100	відмінно	зараховано
74-89	добре	
60-73	задовільно	
0-59	незадовільно	не зараховано

11. Методичне забезпечення

Електронний навчальний курс, розроблений на базі платформи LMS Moodle, розміщений на навчальному порталі НУБіП України за адресою: <https://elearn.nubip.edu.ua/course/view.php?id=2455>

12. Рекомендована друкована література

Основна

1. Негрей М., Гнот Т. Аналітика з R: навчальний посібник / Негрей М., Гнот Т. – Київ: ФОП Ямчинський О. В., 2022. – 236 с.
2. *Економетрика з R* : навчальний посібник / А.В. Скрипник, Д.М. Жерліцин, Ю.О. Нам'ясенко. – Київ: ФОП Ямчинський О.В., 2020. – 248 с.
3. Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.: ISBN 978-5-947060-077-1.
4. Шитиков В.К., Мاستицкий С.Э. (2017) Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 351 с. – Электронная книга, адрес доступа: <https://github.com/ranalytics/data-mining>
5. Роберт И. Кабаков. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.: ISBN 978-5-947060-077-1

Допоміжна

1. Grant V. Farnsworth (2008) Econometrics in R. October 26, 2008. 50 p. <https://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf>

2. Machine Learning with R - Second Edition. by Brett Lantz. Publisher: Packt Publishing. Release Date: July 2015. ISBN: 9781784393908. <https://www.oreilly.com/library/view/machine-learning-with/9781784393908/>
3. Deborah Nolan, Duncan Temple Lang (2015) Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving. 1st Edition. Chapman and Hall/CRC. Published April 21, 2015. 539 p. ISBN 9781482234817
4. XML and web technologies for data sciences with R / Nolan, Deborah; Lang, Duncan Temple. New York [u.a.] : Springer, 2014. 663 p. - ISBN 978-1-4614-7899-7
5. [Joseph Adler](#) (2012) R in a Nutshell, 2nd Edition. A Desktop Quick Reference. [O'Reilly Media](#). October 2012. 724 p
6. Winston Chang (2013) R Graphics Cookbook
7. Мэтлофф Норман. Искусство программирования на R. Погружение в большие данные. – СПб., Питер, 2019. – 416 с. (Серия «Библиотека программиста»). ISBN 978-5-4461-1101-5.

Допоміжна (Internet джерела)

8. <http://r-analytics.blogspot.com/> блог «Анализ и визуализация данных» С. Мастицкого;
9. <http://www.algorithmist.ru/search/label/R> серія статей С. Едунова по реалізації в R різних алгоритмів ;
10. <http://rrus.wordpress.com/> блог «R»
11. <http://www.inp.nsk.su/~baldin/DataAnalysis/index.html> - матеріали Е. Балдіна;
12. <https://m7876.wiki.zoho.com/Introduction-to-R.html> «Введение в R»

13. Інформаційні ресурси

1. World Bank Open Data. Режим доступу: <https://data.worldbank.org>
2. World Economic Forum Режим доступу: <https://www.weforum.org/>
3. Державна служба статистики України. Режим доступу: <http://www.ukrstat.gov.ua>
4. Міністерство Фінансів України. Режим доступу: <https://www.minfin.gov.ua/>
5. Національний банк України. Режим доступу: <https://bank.gov.ua/control/uk/index>
6. Організація економічного співробітництва та розвитку. Режим доступу: <https://data.oecd.org>