Вступ до програмування в R

Матеріали для студентів спеціальності економічна кібернетика, 1 курс

Юрій Клебан

12/05/2022

Зміст

# Про курс

Юрій Клебан  
30.10.2022

(n.d.)

Матеріали підготовлені для читання курсу ***“Вступ до програмування в R”*** [05.250] для студентів 1-го року навчання, спеціальності економічна кібернетика.

## Опис навчальної дисципліни

Навчальна дисципліна спрямована на вивчення основ практичного застосування популярної мови R для проведення статистичних досліджень в економіці.

У процесі вивчення курсу розглядаються теми, що стосуються теоретичних основ та практичної реалізації алгоритмів завантаження, підготовки та обробки економічних даних.

Місце навчальної дисципліни у підготовці здобувачів: програмні результати дисципліни використовуються під час вивчення таких навчальних дисциплін: “Алгоритми та структури даних”, “Аналіз даних в R”, “Прикладне математичне моделювання в R”. Закріплення на практиці здобутих програмних результатів відбувається під час проходження Навчальної практики з курсу «Економіко-математичне моделювання».

## Мета дисципліни

Мета навчальної дисципліни – формування у студентів теоретичних знань та практичних навичок використання мови програмування R для роботи з даними та базовими структурами мови (типи даних, розгалуження, цикли, функції).

## Дотримання принципів доброчесності

Викладач та слухач цього курсу, як очікується, повинні дотримуватися Кодексу академічної доброчесності університету:

* ☒ будь-яка робота, подана здобувачем протягом курсу, має бути його власною роботою здобувача; не вдаватися до кроків, що можуть нечесно покращити Ваші результати чи погіршити/покращити результати інших здобувачів;
* ☒ якщо буде виявлено ознаки плагіату або іншої недобросовісної академічної поведінки, то студент буде позбавлений можливості отримати передбачені бали за завдання;
* ☒ не публікувати у відкритому доступі відповіді на запитання, що використовуються в рамках курсу для оцінювання знань здобувачів;
* ☒ під час фінальних видів контролю необхідно працювати самостійно; не дозволяється говорити або обговорювати, а також не можна копіювати документи, використовувати електронні засоби отримання інформації.

Порушення академічної доброчесності під час виконання контрольних завдань призведе до втрати балів або вживання заходів, які передбачені Кодексу академічної доброчесності НаУОА.

## Зміст

## Зміст

[Про курс](#про-курс)

[Вступ](#вступ)

[Список використаних джерел](#список-використаних-джерел)

## Підтримка курсу

Курс створено у межах проекту “Підготовка, обробка та ефективне використання даних для наукових досліджень (на основі R)”, що підтримує Європейський союз за програмою [House of Europe](https://houseofeurope.org.ua/).

logo house of europe / logo moving forvard together

|  |  |
| --- | --- |
|  | Матеріали курсу створені з використанням ряду технологій:   * ☒ [R Language](https://www.r-project.org) - безкоштована мова програмування для виконання досліджень у сфері статистики, машинного навчання та візуалізацї результатів. * ☒ [Quarto Book](https://quarto.org) - система для публікації наукових та технічних текстів з відкритим кодом (R/Python/Julia/Observable). * ☒ [JupyterLab](https://github.com/jupyterlab/jupyterlab) - середовище розробки на основі [Jupyter Notebook](https://jupyter.org/). JupyterLab є розширеним веб-інтерфейсом для роботи з ноутбуками. * ☒ Git/Github - система контролю версій та, відповідно, сервіс для організації зберігання коду, а також публікації статичних сторінок. |

# Вступ

Юрій Клебан  
30.10.2022

Фахівці спеціальності економічна кібернетика у майбутньому працюватимуть з великими масивами даних, що накопичуються велизними темпами у даний момент і збиралися у попередні дисятиліття. Підготовка, обробка і трансформація даних у зручний формат прийняття рішень забирає все більше часу, а звичні рашіне інструменти аналізу даних, як наприклад, Microsoft Excel не мають достатньо вбудованих можливостей для виконнання задач бізнесу.

На даний час існує велика кількість мов програмування, що інтегруються у суспільні сфери діяльності людини та роботи технічних систем: біоінформатика, а також економіка та бізнес.

Однією з мов програмування, що отримали широке поишення серед економістів-науковців, аналітиків та практиків математичного моделювання (machine learning) є мова програмування (R)(n.d.). Свою популярність ця мова програмування здобула завдяки простоті у використанні, доступності (безкоштовні як базові компоненти для написання коду, так і середовища розробки), розширюваності (кожен розробник має можливість створювати власні пакети та публікувати їх у відкритому доступі).

Основними задачами курсу “Вступ до прграмування в R” є ознайомлення студентів з базовми конструкціями мови програмування R, вивчення способів роботи з найпоширенішими типами даних,

# 1. Що таке R?

Юрій Клебан  
30/11/2022

R є поширеною мовою програмування для роботи з даними (DataScience) та машинного навчання (Machine Learning). Але Ви можете скористатися засобами R і для простіших задач: обчислення, візуалізація даних.

Синтаксис мови програмування R є досить простим для вивчення та використання, а широкий набір готових пакетів дозволяє використати готові розробки для виіршення широкого спектру задач від статистичних обчислень до навчання нейронних мереж для розпізнавання/класифікації зображень.

Важливо відмітити, що мова програмування R є безкоштовною (free) і має відкритий код (open source).

R має ряд корисних властивостей, серед яких варто виділити:

* **Візуалізація даних**. Побудова різноманітих видів графіків, робота з мапами, широкий спектр бібліотек та налаштувань до них.
* **Повторне використання коду**. На відміну від електронних таблиць, що мають обмеження на кількість спостережень (наприклад, MS Excel), R дозволяє працювати з великими масивами даних та перезапускати обчислення у потрібний момент не створюючи додаткових копій даних.
* **Машинне навчання**. R дозволяє використати для побудови, навчання та тестування моделей, а також оптимізації гіперпараметрів та відбору факторів дуже велику кількість алгоритмів. Існують також спеціальні пакети, що об’єднують у собі усі описані функції та алгоритми, наприклад, caret та mlr.
* **Автоматизація**. Написаний код та проекти можна перетворити у готові до публікації та впровадження продукти (deployment) або використовувати напрацьовані алгоритми для швидкого вирішення схожих задач (pipeline).

Також можна виділити досить корисні фічі **Розробка веб-застосунків** та **Звітність**, адже, використовуючи спеціальні бібліотеки (shiny, shinydashboard, flexdashboard, rmarkdown, knitr тощо), результати виконаної роботи можна “оживити” або сформувати “на льоту” готові до презентації документи.

# 2. Коротка історія миови R

Юрій Клебан  
30/11/2022

# 3. Вcтановлення та налаштування R

Юрій Клебан  
30/11/2022

# 4. Документація та допомога

Юрій Клебан  
30/11/2022

# 5. Робота з пакетами

Юрій Клебан  
30/11/2022

# 6. Змінні

Юрій Клебан  
30/11/2022

# 7. Базові типи даних

Юрій Клебан  
30/11/2022

# 8. Оператори

Юрій Клебан  
30/11/2022

# 9. Корисні математичні функції

Юрій Клебан  
30/11/2022

# 10. Створення функції

Юрій Клебан  
30/11/2022

# 11. Вектори

Юрій Клебан  
30/11/2022

# 12. Дата-фрейми

Юрій Клебан  
30/11/2022

# 13. Списки

Юрій Клебан  
30/10/2022

# Список використаних джерел

n.d. *The Comprehensive R Archive Network*. <https://cran.r-project.org/>.