

RSTUDIO ЯК ОДИН ІЗ ІНСТРУМЕНТІВ ДОСЛІДЖЕННЯ ЗАДАЧ З КУРСУ «КОМП'ЮТЕРНА МАТЕМАТИКА»

Грод Іван Миколайович

доктор фізико-математичних наук,
професор кафедри математики та методики її навчання,
Тернопільський національний педагогічний університет імені Володимира Гнатюка
igrod@ukr.net

Андрєва Юлія Юрїївна

магістрантка спеціальності «Середня освіта. Математика»,
Тернопільський національний педагогічний університет імені Володимира Гнатюка
andreeva3553@gmail.com

Основним завданням курсу «Комп'ютерна математика» є формування інформатично-математичних знань, вивчення можливостей використання сучасних інформаційно-комунікаційних технологій розв'язування прикладних задач, дослідження різноманітних процесів і явищ на основі математичного моделювання та застосування методів знаходження оптимальних розв'язків.

Найбільш вдалим пакетом для дослідження поставлених задач є зокрема такі: SAS, SPSS, Python, Statistica, MatLab. Однак, впродовж останніх років, у п'ятірку лідерів входить мова програмування **R**, наразі вона використовується такими компаніями-гігантами, як Google, Facebook, Boeing та New York Times, стаючи все більш розповсюдженою.

Застосування *RStudio* як інструменту, для засвоєння отриманих теоретичних знань в рамках курсу «Комп'ютерна математика», доповнює та ілюструє математичну теорію прикладами і реальними розрахунками, за допомогою моделювання і чисельного експериментування, формує основні поняття про сучасні методи дослідження різноманітних об'єктів, побудови відповідних математичних чи інформаційних моделей та їх дослідження. Такий підхід вписується в концепцію активного і інтерактивного навчання і сприяє появі навичок самостійного наукового дослідження.

До головних переваг **R**, безсумнівно, варто віднести, те що **R** – це безкоштовне програмне забезпечення з відкритим вихідним кодом та можливістю запусити свій код без будь-якого компілятора. Крім того, **R** – це векторна мова, на відміну від інших мов, може виконувати декілька операцій одночасно, розділяючи завдання між декількома вузлами обробки, при цьому скорочуючи час виконання і аналізуючи великі набори даних;

Існує багато «оболонок» для **R**, зовнішній вигляд і функціональність яких можуть сильно відрізнятися. Коротко розглянемо лише найбільш популярний варіант – *RStudio*, який теж має некомерційну версію продукту.

RStudio – інтегроване середовище розробки (IDE) (<https://www.rstudio.com/>). *RStudio* має більш зручний інтерфейс, розділені області та додаткові модулі (наприклад, історія команд, робоча область), що спрощує роботу з **R**.

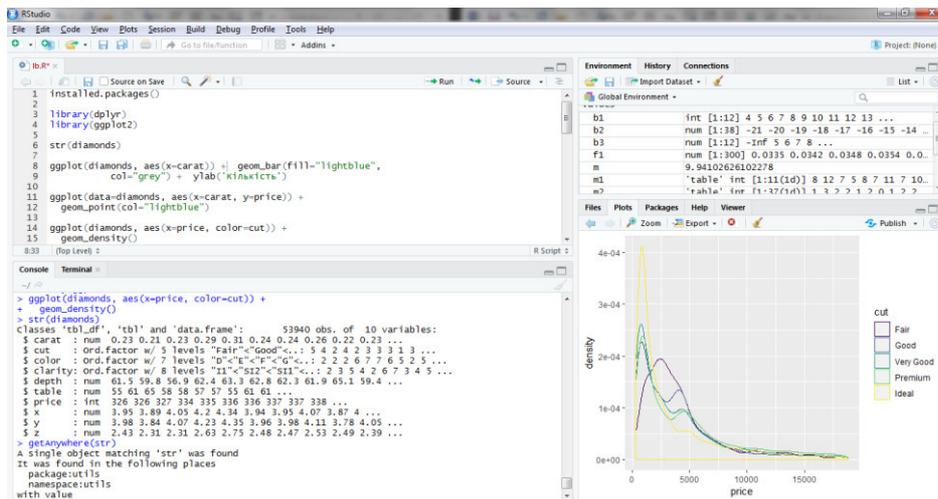


Рис. 1 Вікно середовища RStudio для програмування на мові R.

Основною складовою Data Science, тобто науки про дані, є самі дані, та методи їх збирання, зберігання, обробки та виокремлення із загального масиву даних корисної інформації. Вихідні дані для статистичного аналізу є як правило результатом вимірювання кількох змінних або характеристик кількох об'єктів. Змінні діляться на числові і нечислові (якісні), принциповою відмінністю якісних змінних від кількісних є неможливість упорядкувати природним чином значення якісної змінної. Саме тому дані легше сприймати у вигляді таблиці, пакет *readr* Package – «*readr*» допомагає читати різні форми даних в R, а *data.table* Package – дозволяє швидше маніпулювати набором даних з мінімальним кодуванням.

У термінології R така таблиця називається кадром даних (*data frame*), рядки і стовпи в цій таблиці нерівноправні. Кожен стовпець представляє одну змінну (*variables*), кожен рядок – одне спостереження або випадок (*observation*), для створення набору даних використовується функція *data.frame*.

Основу математичного апарату Data Science складають основи теорії імовірностей та математичної статистики, статистичні оцінки, закони розподілу, регресія, кореляція. У пакеті *dplyr* (відомий своїми можливостями дослідження та трансформації даних) присутній набір функцій для роботи з деякими типовими дискретними і неперервними випадковими величинами та одновимірними розподілами їх імовірностей. Наприклад, вибірку з $n=100$ значень з нормального розподілу із середнім $mean=1$ і середньоквадратичними відхиленням $sd=0,3$ можна отримати командою *rnorm(100,1,3)*.

За час масового поширення технологій людина генерує величезну кількість даних, яку вона не здатна обробити і візуалізувати. Тому загальну форму і глобальні описові характеристики розподілу двох (і більше) змінних легше досліджувати на графіку (Рис. 1), пакети *ggplot2* і *ggedit for* є найбільш зручними для побудови графіків.

Програмування на мові R є об'єктно-орієнтованим, що робить скрипт написаний на ній більш чітким та компактним, відкритий код в RStudio, дозволяє за допомогою функції *getAnywhere(im'я)* переглянути код створення заданого в дужках об'єкта або методу.

Переваги об'єктно-орієнтованого програмування в **R** можна продемонструвати на прикладі. Нехай задана статистична регресія, виконаємо регресійний аналіз в **R**, використовуючи функцію `lm()`, дана функція в **R** повертає об'єкт, що містить усі результати – розрахункові коефіцієнти, їх стандартні похибки, залишки тощо, після чого аби використовувати лише, наприклад, розрахункові коефіцієнти, достатньо звернутись до них як до елемента матриці.

R дозволяє застосовувати одну функцію до об'єктів різних класів, в «середині» ж цих класів обирається внутрішній метод або процедура для виконання цієї функції. Наприклад, розглянемо функцію `plot()`, застосувавши її до списку чисел та отримаємо простий графік. Але якщо застосувати її до об'єкта, що є результатом того ж регресійного аналізу, отримаємо набір ділянок, що характеризуватимуть різні аспекти цього аналізу.

```
> numbers <-c(23,11,18,7,39,-2,-5)
> plot(numbers, type = "b")
```

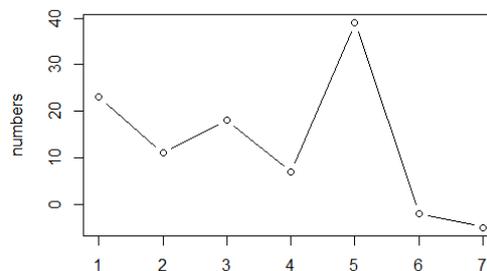
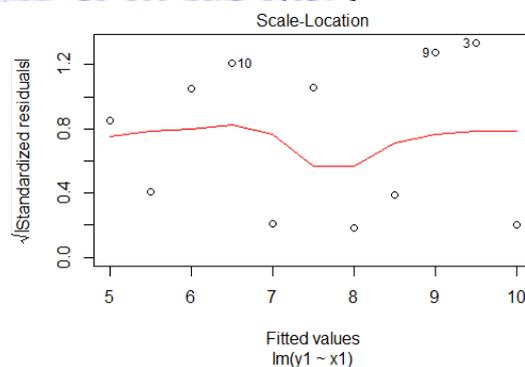


Рис. 2 Побудова простого графіка на основі числового вектору.

Застосування *RStudio* є доцільним для статистичної обробки даних, їх інтерполяції і екстраполяції, апроксимації поліномами із застосуванням методу найменших квадратів, реалізації методу статистичних випробувань (методу Монте-Карло), формулювання та перевірка гіпотез тощо.

В *RStudio* є бібліотеки, які дозволяють реалізувати із генетичні та еволюційні алгоритми, і більш прості завдання, пов'язані, наприклад, з кластерним аналізом

```
> plot(lm1)
hit <Return> to see next plot: return()
hit <Return> to see next plot: return(lm1)
hit <Return> to see next plot: return(lm1$df.residual)
```



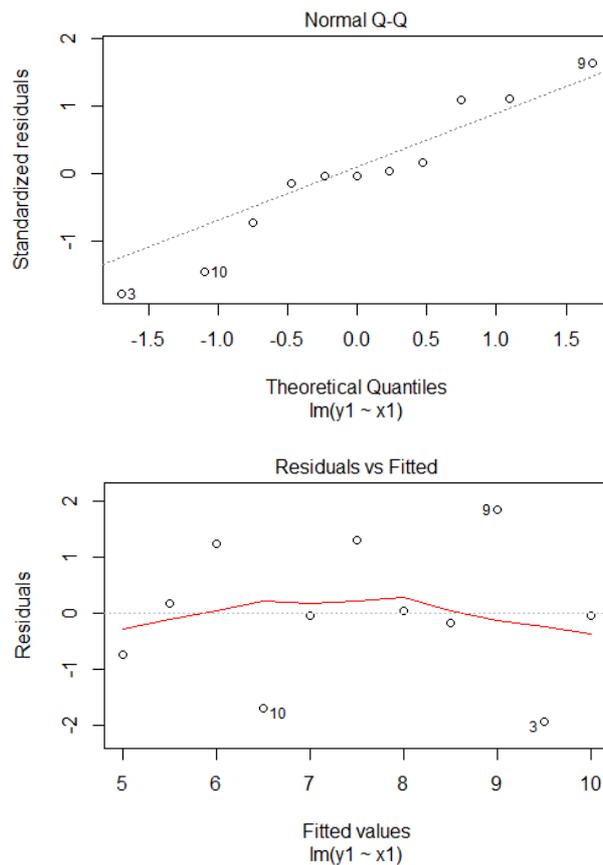


Рис. 3 Побудова графіків на основі результатів регресійного аналізу.

Отже, використання *RStudio* як інструменту для розв'язання задач в курсі «Комп'ютерної математики» дозволяє студенту уникнути проведення громіздких, рутинних операцій, однотипних обчислень, та зосередитися безпосередньо на аналізі досліджуваного явища. А значна кількість алгоритмів та методів реалізованих в програмному середовищі дає можливість їх використовувати при проведенні наукових досліджень, для знаходження оптимального розв'язку досить широкого класу математичних задач з візуалізацією основних етапів розв'язування.

Список використаних джерел:

1. Evans R. R Programming/ Robin Evans – Michaelmas, – 2014p., – С.4-6.
2. Matloff N. S. The art of R programming : tour of statistical design./Matloff, Norman S. – San Francisco, – 2011p., – С.207-216.
3. Why R is important for data science professionals![Електронний ресурс] – Режим доступу: <https://blog.eduonix.com/bigdata-and-hadoop/why-r-is-important-for-data-science-professionals/>.
4. Єфименко В. В. Методика навчання комп'ютерної математики майбутніх учителів інформатики(автореферат)/ В.Єфименко – Київ, – 2015р., – С.9-12.
5. Зорин А. Введение в прикладной статистический анализ в пакете R(навчально-методичний посібник)/ А.Зорин, В., М.Федоткин –ННГУ, – 2010р., – С.3-21.