

Орлов Є. В.

*доцент кафедри статистики та ММЕ
Одеський національний економічний університет
м. Одеса, Україна*

Шинкаренко В. М.

*доцент кафедри статистики та ММЕ
Одеський національний економічний університет
м. Одеса, Україна*

ВИКОРИСТАННЯ СЕРЕДИ ПРОГРАМУВАННЯ R В ПРОЦЕСІ ВИКЛАДАННЯ МАТЕМАТИЧНИХ ТА СТАТИСТИЧНИХ ДИСЦИПЛІН

При побудові математичних і статистичних моделей економічних явищ і процесів часто виникає необхідність вирішувати математичні завдання, які або мають досить складний аналітичний розв'язок, або вимагають використання чисельних методів. Раніше це було досить складною проблемою, яка вимагала або самостійного опанування спеціальних розділів математики і статистики або звернення до відповідних фахівців. Обидві дороги істотно ускладнювали доступність використання наявного апарату математичних і статистичних методів. Проте, за останні декілька десятиліть з появою різних комп'ютерних програм ситуація повністю змінилася.

При наявності у розпорядженні користувачів десятки пакетів прикладних програм виникає своєрідна проблема вибору: яке програмне забезпечення вибрати для розв'язку певної проблеми. Так, якщо мова йдеться про типові математичні задачі, які розглядаються в курсі „Вища математика та теорія ймовірностей - (а саме: дії над матрицями, обчислення визначників, побудова графіків, пошук похідних та інтегралів а також розв'язок різноманітних рівнянь), то слід порадити для початкового використання програму Mathcad. Вона має дуже прозорий інтерфейс (записуємо вираз і одразу отримуємо результат). Хоча, при цьому, містить інструменти для програмування та симуляцій різноманітних процесів. Біль професійною програмою для розв'язку математичних задач слід вважати Wolfram Mathematica, яка все потребує певного початкового рівня входження. Можливості цієї програми щодо розв'язку математичних задач повністю задовольняють всі потреби вченого економіста. Слід також відзначити можливість зберігати векторні зображення, що є необхідністю для відправлення в редакції провідних наукових видань. В деяких ситуаціях (якщо потрібно побудувати модель певного процесу) можна порадити програми Matlab та Maple.

Що до задач, які мають відношення до статистики та аналізу даних, ситуація більш цікава та не така однозначна. Ми не будемо торкатися електронних таблиць Excel, тому що, по-перше, їх достоїнства і так добре відомі, по-друге, на наш погляд, при існуючих спеціалізованих програмних пакетах використання Excel для розв'язку статистичних і тим більше математичних задач себе не оправдовує. Розглянемо з яких кроків складається повний цикл інформаційно-технологічного процесу:

- отримання доступу до оброблюваних даних (їх загрузка з різноманітних джерел та комплектація сукупності таблиць);
- редагування завантажених показників (різні програми для відокремлення одного значення від іншого використовують різні підходи, або точка з комою, або пропуск або щось інше, тому потрібно звести таблицю початкових даних саме до того виду, який потребує використовувана програма);
- анутовання даних (розбиття на різні фрагменти);
- отримання загальної структури даних (обчислення описових статистик);
- графічне представлення даних і результатів обчислень в зрозумілій інформативній формі (вдале візуальне представлення якогось явища, процесу дуже спрощує розуміння, та пошук подальший шляхів застосування отриманої інформації);

- моделювання даних (математичний опис залежностей, тестування статистичних гіпотез, машинне навчання та ін.);
- оформлення результатів (підготовка таблиць та діаграм достатньої публікаційної якості).

То, яке вибрати програмне забезпечення аналізу даних буде вибране для опанування залежить від специфіки поставленої задачі, ефективності налаштування алгоритмів обробки, зручності інтерфейсу а також від ціни програмного забезпечення. Широко використовувана Statistica має фіксований набір кнопок меню, які далеко не завжди можуть задовільнити творчого дослідника, який вважає за краще самостійно використовувати шлях обчислюваного процесу. Існують альтернативні програми, такі як Matlab, SPSS та інші, які в змозі комбінувати різні типи даних, надають доступ до проміжних результатів, дозволяють різноманітні стиль відображення та додавати свої розширення програмних модулів. Однак ціни та подібні програми дуже високі. Альтернативою їм можна вважати безкоштовну середу програмування R, яка постійно розвивається і є сучасним інструментом статистичних досліджень [1-3].

Середу програмування R безкоштовно встановлюється з офіційного сайту [4]. Для зручності використання можна встановити графічну оболонку RStudio, яку також можна безкоштовно скачати з офіційного сайту [5,6]. Студенти першого курсу дуже легко справляються з цією задачею. R є дуже потужним інструментом, тому потребує певних зусиль для оволодіння, але це себе повністю оправдовує тому що дозволяє розв'язувати більшість задач з Вищої математики і всі задачі Статистики (для якої він і був створений). На відміну від іншого світу, де на протязі останніх двадцяти років R перетворився в майже найважливіший інструмент статистика (про це свідчать різноманітні нагороди, широта використання в різних організаціях а також біля чотирнадцяти тисяч розроблених пакетів для різноманітних задач), в нашій країні поки що не знайшов такого поширеного використання. Але безумовно, наведення скриптів R стає загальновизнаним стандартом як в журнальних публікаціях, так і при неформальному спілкуванні вчених всього світу.

Отже, ми пропонуємо ввести елементи програмування на R в програму навчання математичних та статистичних дисциплін, можливо навіть замість електронних таблиць Excel, принаймні для студентів спеціальності Статистика. Однак сучасна велика зацікавленість в такій області як Data Science, де основними інструментами є R та Python, нам здається сприяє тому щоб студенти усіх економічних спеціальностей мали поняття про цей потужний та сучасний інструмент дослідження даних.

Список використаних джерел

1. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. Data Mining, Inference, and Prediction. Springer. 2009. 745 p.
2. Chollet F., Kalinowski T., Allaire J.J. Deep learning with R. Manning publication. 2022. 646p.
3. James G., Witten D., Hastie T., Tibshirani R., An introduction to statistical learning with applications in R. Springer. 2013. 450 p.
4. <https://www.r-project.org> центральний сайт R, де можна його завантажити.
5. <https://posit.co/download/rstudio-desktop/#download> завантаження RStudio.
6. <http://www.cookbook-r.com/Basics/> посібник по R.