

1. Початок роботи з системою R

R - це середовище програмування для статистичного аналізу даних. Це середовище складається з базової програми R, що працює як інтерпретатор мови статистичного програмування S, та окремих пакетів, які реалізують спеціальні методи та технології статистичної обробки даних. Базова програма створена у рамках проекту GNU, як альтернативна програмна реалізація мови S (ця мова та комерційний пакет S+ для її реалізації були розроблені у Bell Laboratories під керівництвом Дж. Чемберса). На відміну від S+, програма R є некомерційною і вільно розповсюджується за умови дотримання вимог GNU General Public License. Офіційна сторінка проекту R: <http://www.r-project.org/>.

Отримати останню версію інсталятора базової програми R для операційної системи Windows можна за адресою: <http://cran.r-project.org/bin/windows/base/>. Інсталятор завантажується у вигляді exe-файлу. Для інсталяції програми досить запустити цей файл і відповідати на його запитання. При першій спробі роботи з R рекомендовано погоджуватись з усіма пропозиціями, які робить інсталятор. Проблеми можуть виникнути, якщо на вашому комп'ютері встановлені різні права доступу для різних користувачів. Справа в тому, що R наприкінці кожної сесії роботи зберігає на диску «робочий простір» (workspace) – сукупність даних та програм, які були завантажені під час сесії. На початку наступної сесії workspace завантажується з диску. Якщо під час інсталяції для зберігання workspace буде обрано директорію, недоступну певному користувачеві, то при роботі з R можуть виникати повідомлення про неможливість завантаження або зберігання workspace. Для усунення таких повідомлень потрібно або вибрати директорію вільного доступу при інсталяції, або змінити директорію, використовуючи пункт File->ChangeDir... у головному меню головного вікна програми R.


Після інсталяції R його можна запустити і отримати приблизно таке вікно:

Тут вгорі знаходиться головне меню, а нижче відкрито вікно «консолі R», в якій можна давати інструкції програмі та отримувати її відповіді. Синім кольором у цьому вікні виведено початкову інформацію про вашу версію базової програми R. Далі червоним кольором вказана інструкція, яку R виконав автоматично при завантаженні. Нарешті, червоний символ `>` є запрошенням користувачу вводити власні інструкції. Для перевірки роботи системи можна після `>` ввести

`2+2`

і натиснути Enter. Результат буде виведений на консоль.

Як бачимо, R виводить результати виконання безпосередньо після інструкції синім кольором, після чого переходить у режим очікування наступної інструкції, про що повідомляє червоним знаком `>`. При роботі з R можна виконувати одразу багато інструкцій, що записані в окремому файлі. Найпростіший спосіб зробити це – завантажити такий файл в якому-небудь текстовому редакторі, зробити там сору, а потім – paste на консолі. При цьому, якщо інструкції у файлі розміщені у окремих рядочках, розділових знаків між ними не потрібно. Інструкції, вміщені в одному рядочку, розділяють символом `;`. Якщо довга інструкція не вміщується у одному рядочку, її можна розбити на декілька рядочків, причому, при переході до наступного рядочку R автоматично виводить символ продовження `+`. R сам

здогадується, що інструкція не закінчена за її синтаксисом. Тому деякі синтаксичні помилки (як от – забуті дужки) можуть сприйматись як незакінчені інструкції. У цьому випадку R виставить + на початку наступного рядочка і перейде у режим очікування. Натисніть escape, щоб перейти у режим введення нової інструкції без продовження аналізу попередньої. Програми, що складаються з інструкцій R, називають скриптами (script). Вони мають стандартне розширення .r. У базовій програмі є можливість відкрити вікно редактора для створення нового скрипту, або завантажити файл зі скриптом, використовуючи пункти головного меню File->Newscript або File->Openscript. Виконати завантажений у вікні редактора скрипт повністю можна, використовуючи Edit->Runall. Можна також виконати виділену частину скрипту, використовуючи кнопку  (Run line or selection).

Закінчивши роботу зі скриптом, його можна зберегти, використовуючи File->Save. Однак, редактор скриптів, реалізований у R, має дуже обмежену функціональність. Тому доцільно для редагування більш-менш складних скриптів використовувати зручніші текстові редактори. Ви можете користуватись тим редактором, до якого звикли, але спеціально для редагування R-скриптів зручно застосовувати редактор Notepad++. Це некомерційна, вільно розповсюджувана програма, яку можна отримати на офіційному сайті <http://notepad-plus-plus.org/> . Працюючи з файлами, що мають розширення .r, Notepad++ трактує їх як скрипти системи R і виділяє логічну структуру скрипту.

Базова програма R є лише інтерпретатором одного з діалектів мови програмування S. Конкретні статистичні алгоритми, як правило, виконуються у вигляді скриптів на цій мові і зібрані у пакети (packages) R. При інсталяції разом з базовою програмою інстальюються і основні пакети, які реалізують найбільш популярні технології статистичної обробки. Частина цих пакетів автоматично завантажується при запуску R. Ті пакети, які не завантажились автоматично, але інстальовані на комп'ютері, можна завантажити функцією library:

```
>library(splines)
```

завантажує пакет splines, призначений для роботи зі сплайнами. Ця ж функція, викликана без параметрів, дає перелік усіх завантажених пакетів з коротким поясненням їх функцій (у мінімальному варіанті цього переліку кілька десятків пакетів). Якщо потрібний пакет не було інстальовано на комп'ютері, його можна завантажити з інтернет-архіву, використовуючи пункти головного меню

Packages->Installpackage(s).

Спочатку програма пропонує вибрати інтернет-архів, з якого робиться інсталяція. Варіант 0-cloud, що пропонується за умовчанням, як правило, працює цілком задовільно. Після цього треба у списку вибрати потрібний для вас пакет. Якщо цей пакет використовує які-небудь інші, котрих немає на вашому комп'ютері, вони будуть інстальовані автоматично. Після того, як пакети інстальовані, вони зберігаються на вашому комп'ютері, але для роботи з ними під час сеансу їх треба підключати, використовуючи функцію library. Усі об'єкти, створені або завантажені під час сеансу роботи з R і не видалені спеціальною інструкцією, зберігаються у робочому просторі. Наприкінці сеансу R запитує, чи зберігати робочий простір на диску:

Save workspace image?

Якщо вибрати збереження, то цей робочий простір буде відновлено на початку наступного сеансу. Такою можливістю варто користуватись дуже обережно, оскільки «старі» об'єкти можуть спотворювати роботу R у новому сеансі. Зберігати робочий простір доцільно лише в тому випадку, коли ви збираєтесь наступного разу продовжити свою роботу з того самого місця, на якому зупинились зараз. Існує великий обсяг англійської літератури по R та мові програмування S. Найбільш корисною для початківця є книжка W.N. Venables, B.D. Ripley «Modern Applied Statistics with S», котру можна знайти, наприклад, тут:
http://www.planta.cn/forum/files_planta/modern_applied_statistics_with_s_192.pdf

Крім того, дуже зручно користуватись офіційними керівництвами з мови S, що входили у стандартну поставку пакету S+ (На даний час комерційний проект S+ припинено, версій цієї програми під сучасну Windows, наскільки нам відомо, не існує). При роботі з такою літературою слід мати на увазі, що R як діалект дещо відрізняється від стандартного S з пакету S+. Крім того, кожен додатковий пакет може мати свої особливості, які пояснюються у документації до нього. Це може створити певні незручності, але досвід показує, що такі проблеми легко розв'язуються пошуком у інтернеті по назві пакету або функції із додаванням ключових слів "rsoftware".

Лабораторна робота № 1. Робота з даними (найпростіші операції і типи даних).

Для виконання певної операції необхідно натиснути клавішу Enter. Якщо необхідно скопіювати команду, яку ви вже виконували раніше в цьому файлі, можна скористатись клавішами ↑ та ↓.

Для виклику help, що стосується певної команди, слід набрати ?<<назва команди>> і натиснути клавішу Enter; якщо вас цікавить опція, а не команда, можна спробувати набрати ??<<назва опції>>. Коментарі відмічаються символом #.

Як було сказано вище, зберігати робочий файл рекомендується, скопіювавши його в Блокноті, в текстовому файлі. R пропонує збереження лише значень змінних.

Завдання

1. Створити особисту папку.
2. Створити особистий скрипт № групи, ПІБ.
3. Ознайомитися з найпростішими операціями і типами даних.