

*Свиридова Л.Е.,
Аспирант 1 курс, кафедра
«Системного анализа и исследования операций»
Сибирский государственный университет науки и технологий имени
академика М.Ф.Решетнева
Россия, г. Красноярск*

ИСПОЛЬЗОВАНИЕ ЯЗЫКА R В ПРИКЛАДНОЙ СТАТИСТИКЕ

***Аннотация:** Возможности R безграничны, и могут использоваться там, где необходима обработка данных. Под обработкой можно понимать не только статистический анализ, но и анализ графики, а также продвинутое математическое моделирование.*

***Ключевые слова:** статистика, базы данных, язык R, данные.*

***Annotation:** The possibilities of R are limitless and can be used where data processing is necessary. By processing, we can understand not only statistical analysis, but also analysis of graphics, as well as advanced mathematical modeling.*

***Keywords:** statistics, databases, R language, data.*

Прикладная статистика – наука о способах обработки статистических данных. Как самостоятельная научно-практическая область она появилась сравнительно недавно, однако её развитие проходит стремительно. В ее составе обширные и весьма совершенствующиеся научные направления. Применение прикладной статистики и других статистических методов зачастую используют специалисты в своих областях исследования, таких как технические исследования, экономика, менеджмент, социология, история, геология, то есть непрофессионалы в статистической науке ¹.

¹ Орлов А. И. Прикладная статистика - состояние и перспективы // Научный журнал КубГАУ - Scientific Journal of KubSAU. 2016. №119

R, о котором пойдет речь, является не только языком программирования, но и свободной средой для статистического счета и графического анализа. Первоначально язык R разработали Росс Айхэка и Роберт Джентлмен, они работали в Оклендском университете на факультете статистики. Новая разработка была анонсирована в августе 1993 г. R задумывали как аналог среды S-Plus, распространявшейся платно. Этот фактор сдерживал распространение данной среды. Создатели хотели, чтобы R был новой реализацией языка S, с отличием в некоторых чертах, к примеру, подходом к использованию глобальных и локальных переменных и работой с памятью².

R представляет собой пакет для изучения данных со свободным исходным кодом, обладающим поддержкой большого и активного исследовательского сообщества. Самой популярной конференцией, где происходит общение и обмен опытом для R-сообщества является useR! (The R User Conference), которая проходит ежегодно с 2004 года³. При этом, все еще пользуются популярностью пакеты для изучения статистических данных и их обработки, например Microsoft Excel , SAS , IBM SPSS , Stata и Minitab⁴. Стоит отметить, что в 2010 году R оказался в числе победителей в конкурсе журнала Infoworld в номинации на лучшее открытое программное обеспечение для разработки приложений⁵.

Возможности R безграничны, и могут использоваться там, где необходима обработка данных. Под обработкой можно понимать не только статистический анализ, но и анализ графики, а также продвинутое математическое моделирование. Также R можно использовать в специализированных программах вроде MATLAB или Octave, но зачастую больше всего он применим в статистическом анализе, начиная от вычисления средних величин до преобразований временных рядов. В настоящее время, этот язык настолько

2 Шипунов А. Б. и др. Наглядная статистика. Используем R! //М.: ДМК Пресс. – 2012. – Т. 298. – С. 25.

3 R: The R Project for statistical Computing. [Электронный ресурс]. URL: <https://www.r-project.org/conferences.html>.

4 Кабаков Р. R в действии. Анализ и визуализация данных на языке R. // Litres, 2017. С 13-14

5 InfoWorld Bossie Awards 2010 [Электронный ресурс]. URL: <https://web.archive.org/web/20140911220615/http://www.infoworld.com/d/open-source/bossie-awards-2010-the-best-open-source-application-development-software-140%26current%3D10>

популярен, что очень многие серьезные компании его устанавливают для работы и составляет труд отыскать университет, где с ним не работали или даже о нем не слышали⁶.

Специалисты гуманитарной направленности, привыкшие работать с удобными и понятными графическими интерфейсами, сталкиваются с трудностями пакета R, так как в нем присутствует интерфейс командной строки. Первоначально это может оттолкнуть пользователя-гуманитария, однако с помощью знания синтаксиса языка программирования и умения специальной формулировки требований программе, можно очень эффективно использовать все функции R. Однако, преодолев все трудности, пользователь может получить опыт создания уникальных процедур статистической обработки данных.

Стоит заметить, что популярные пакеты для изучения статистических данных тоже имеют поддержку R, позволяя расширять задачи данных пакетов, открывая и запуская программы. Объединение данных средств можно наблюдать с пакетом STATISTICA. Можно включать скрипты R прямо в программе, увеличивая возможности STATISTICA благодаря библиотекам R. Немаловажно, что с десятой версии пакета STATISTICA интеграция с R происходит автоматически. Стоит сказать, что такая практика уже была реализована с программами, имевшими интерфейс командной строки (SAS, Stata, SYSTAT)⁷.

Преимущества среды R:

- бесплатность и кроссплатформенность;
- богатство арсенала методов статистики;
- качественность векторной графики;
- содержит больше 7000 надежных пакетов;

⁶ Шипунов А. Б. и др. Наглядная статистика. Используем R! // М.: ДМК Пресс. – 2012. – Т. 298. – С. 25.

⁷ Бородкин Л.И. Рец. на: Т. Arnold, L. Tilton. Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text. Springer, 2017. - 211 p. // Историческая информатика. — 2018. - № 2. - С.126-132. DOI: 10.7256/2585-7797.2018.2.26986. [Электронный ресурс]. URL: http://e-notabene.ru/istinf/article_26986.html.

- гибкость для использования в плане создания и редактирования скриптов и пакетов, а также взаимодействия с другими языками (C, Java и Python и др.) и работы с форматами данных для SAS, SPSS и STATA;

- поддержка большого живого сообщества разработчиков и юзеров;
- регулярное обновление ПО;
- обработка большого массива информации (до 1 Гб);
- возможность создания таблиц любых размеров;
- параллельные вычисления (пакеты `foreach` и `doParallel`)⁸ ;
- большое количество бесплатных курсов на русском языке даже для непрофессионалов⁹ .

Вполне очевидным недостатком можно назвать затруднительность для понимания пользователя незнакомого с программированием.

Подводя итог, можно сказать, что использование языка R многогранно и многофункционален и учитывая достоинства, его использование может решить много задач. Даже его существенный недостаток имеет двойное дно и может обернуться преимуществом, потому что возможности интерфейса командной строки могут достигать таких целей, каких пользователь программы с визуальным интерфейсом смог бы достичь часами ручного труда.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Орлов А.И. Прикладная статистика - состояние и перспективы // Научный журнал КубГАУ - Scientific Journal of KubSAU. 2016. №119.
2. Шипунов А.Б. и др. Наглядная статистика. Используем R! //М.: ДМК Пресс. – 2012. – Т. 298. – С. 25.
3. R: The R Project for statistical Computing. [Электронный ресурс]. URL: <https://www.r-project.org/conferences.html>.

8 Мамедов В.С. Анализ средств языка программирования R //Молодой исследователь Дона. – 2017. – №. 1 (4).

9 Статистические методы в гуманитарных исследованиях. Электронный курс ТГУ на портале Coursera. [Электронный ресурс]. URL: <https://www.coursera.org/learn/statistics-for-humanities>

4. Кабаков Р. R в действии. Анализ и визуализация данных на языке R. // Litres, 2017. С 13-14
5. InfoWorld Bossie Awards 2010 [Электронный ресурс]. URL: <https://web.archive.org/web/20140911220615/http://www.infoworld.com/d/open-source/bossie-awards-2010-the-best-open-source-application-development-software-140%26current%3D10>
6. Бородкин Л.И. Рец. на: Т. Arnold, L. Tilton. Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text. Springer, 2017. - 211 p. // Историческая информатика. — 2018. - № 2. - С.126-132. DOI: 10.7256/2585-7797.2018.2.26986. [Электронный ресурс]. URL: http://enotabene.ru/istinf/article_26986.html.
7. Мамедов В.С. Анализ средств языка программирования R // Молодой исследователь Дона. – 2017. – №. 1 (4).
8. Статистические методы в гуманитарных исследованиях. Электронный курс ТГУ на портале Coursera. [Электронный ресурс]. URL: <https://www.coursera.org/learn/statistics-for-humanities>.