

УДК 004.42(476)

ИСПОЛЬЗОВАНИЕ ПРОГРАММЫ R ДЛЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ НАУЧНЫХ ДАННЫХ

Дубинич В.Н., Дубинич М.В.

УО «Гродненский государственный аграрный университет»

г. Гродно, Республика Беларусь

В среде Linux для статистической обработки результатов исследований возможно применение различного программного обеспечения как коммерческого, так и свободнораспространяемого. Достаточно удобными считаются электронные таблицы Calc и Gnumeric. Данные программы схожи с MS Office Excel по принципу работы пользователя. Однако, следует помнить, что применение электронных таблиц позволяет реализовать только первичный анализ данных, так как данный софт предназначен для офисного применения и не ориентирован на полноценную статистическую обработку научных данных.

Пакетами для статистического анализа, применяемыми как в Linux, так и в других операционных системах (MS Windows и MacOS), являются Scilab, Maxima, Minitab, MatLab, Octave, GenStat, JMP, Analyse-it, R и др.

Во многих американских и западноевропейских университетах для обработки данных используется среда R, а в последние годы она получила распространение и на постсоветском пространстве. Данная программа применяется для анализа данных в психологии, медицине, физике, химии, экологии, экономике и в других областях знаний [1, 3, 5].

Что же из себя представляет R? Это свободно распространяемая программная среда с открытым кодом, язык программирования для статистической обработки данных, среда математического моделирования и работы с графикой. Среда R создавалась как аналог коммерческой программы S-PLUS, разработанной на базе языка программирования S. Наиболее бурно R начал развиваться после появления возможности лёгкого написания пакетов под нужды конкретных научных центров, лабораторий и исследователей. Следствием данного факта явилось создание информационного ресурса CRAN (Comprehensive R Archive Network — <http://cran.r-project.org>), на котором хранятся новые версии R, дополнения необходимые для полноценной работы в

среде, издаётся журнал. На сегодняшний день количество пакетов составляет более 4300 [2, 4].

Широчайшую известность R приобрёл в биоинформатике, когда Robert Gentleman запустил в работу проект открытого программного обеспечения Bioconductor. В настоящее время данный проект является безусловным лидером биоинформатического программного обеспечения для различных направлений (genomics, proteomics, metabolomics и т. д.) [5].

Отличительными сторонами среды R в массе программ для статистической обработки является её высокая гибкость, бесплатность, кросс-платформенность, огромное количество дополнительных пакетов, возможность работать с массивами данных содержащие сотни тысяч наблюдений, вывод полученных результатов в виде двумерных и трёхмерных графиков высокого качества, возможность написания собственных пакетов, огромное сообщество пользователей и разработчиков[1].

Положительной стороной в работе с R является наличие пакета Sweave позволяющего обрабатывать данные и помещать их непосредственно в документ LaTeX в процессе написания научного отчёта. В этом случае выполняется принцип «грамматного программирования», т. е. команды на выполнение формул, тестов, построение графиков и т. п. вносятся в тексте отчёта, а чистовой вариант(после компиляции) содержит только результат включённый непосредственно в отчёт.

Безусловно в работе с программой имеются и некоторые сложности требующие от пользователя определённых знаний и навыков работы как с компьютером, так и со средой R. Например, необходимость использования командной строки, отсутствие коммерческого обслуживания, незначительное количество русскоязычной литературы.

Таким образом, в среде R реализовано множество функций призванных упростить работу по статистическому анализу данных, формированию научных отчётов, созданию графиков и диаграмм различной сложности.

ЛИТЕРАТУРА

1. Зарядов, И.С. Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации и графика: уч.-мет. пособие / И.С. Зарядов. - Москва: Издательство Российского университета дружбы народов, 2010. - 207с.
2. Шитиков, В.К. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R / В.К. Шитиков, Г.С.Розенберг. - Тольятти: Институт экологии Волжского бассейна, 2013. - 227 с.
3. Введение в R / А. Шипунов, Е. Балдин // Linux Format. - 2008. - №1. - С. 98-101.
4. Comprehensive R Archive Network // CRAN [Электронный ресурс]. - 2013. - Режим доступа: <http://cran.r-project.org>. Дата доступа 26.03.2013.
5. Мой опыт введения в R или «I Love R»// Habrahabr [Электронный ресурс]. - 2006. - Режим доступа: <http://habrahabr.ru/post/168817/>. Дата доступа 27.03.2013.