



R - значит победа

КОРОСОВ

Андрей Викторович

Петрозаводский госуниверситет,

korosov@psu.karelia.ru

© 2012 Петрозаводский государственный университет

Опубликована: 28 декабря 2012 года

Excel всех расслабил. Формулы, ссылки, базы данных, базовая статистика, сводная таблица, поиск решения, макросы, диаграммы, векторные рисунки, Visual Basic, интеграция в MS-среду – эти и многие другие возможности, организованные самим дружелюбным образом, нашептывают: «Биологу для расчетов лучше Excel ничего не найти». И правда, – все здорово (хотя и здорово дорого), все устраивает, но только до тех пор, пока не столкнешься с серьезными или новыми задачами. Стоит захотеть найти собственные векторы, выполнить шкалирование, применить набирающую популярность технологию бутстреп-анализа, сделать статистику по данным о биоразнообразии – и функций универсального пакета не хватает, требуются специальные программы, причем несколько разных и дорогих.

Однако есть прекрасная бесплатная альтернатива, точнее дополнение к Excel или LibreOffice.Calc, в лице языка R.



Звучит странно, но это объектно-ориентированный интерпретатор, по словам авторов – калькулятор-переросток. В среде R с клавиатуры вводится строчка команды, например `3+4`, строчной ниже появляется результат. Команды могут быть разными и ну очень эффективными: `a%*%b` – дает произведение матриц `a` и `b`, `t(x)` транспонирует матрицу `x`; `rbinom(n,M,S)` – создает набор `n` случайных значений, распределенных по биномиальному закону со средней `M` и стандартным отклонением `S`; `f=splinefun(x,y)` – из векторов точек с координатами `x` и `y` создает сплайн-сглаживание; `curve(f,...)` – рисует линию с координатами в векторе `f`; `nlm(fun,p)` – минимизирует функцию `fun` с `p` значениями ее параметров (аналог «Поиска решения» в Excel), `princomp(x, ...)` – дает расчет главных компонент для матрицы `x`; `hclust(d,...)` – кластеризует объекты по матрице расстояний `d` методом Варда. Можно писать длинные коды (скрипты – программы), сохраняя в файлах; можно редактировать файлы с данными, строить разнообразные диаграммы и пр.

Разнообразие функций необычайно велико (несколько тысяч). Этот продукт свободно распространяется (The Comprehensive R Archive Network, 2012; The R Project for Statistical Computing, 2012), он дает возможность самим пользователям создавать новые команды (пакеты) и добавлять в общую копилку. Существенно то, что международная организация CRAN взялась объединять и стандартизировать этот язык. Для биологов разработано несколько специальных пакетов функций, которыми можно воспользоваться, усвоив достаточно простой синтаксис этого языка.

Те, кто изучал в школе Бейсик, легко освоит и R. К счастью, на русском языке имеется немало удачных учебных пособий, многие из которых представлены в группе «Язык программирования R для статистических вычислений...» на сайте vk.com. Начать можно с небольшого вики-учебника (Язык программирования R, 2011), а продолжить – учебником А. Шипунова и Е. Балдина (2008) и книгами с сайта twirpx.com (2012), например (Зарядов, 2010). Многие функции обработки биологической

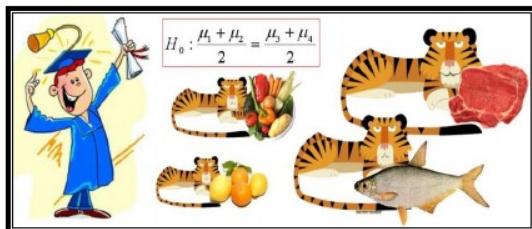
информации описаны только по-английски (Seefeld, Linder, 2007; Burns, 2010). Для тех пользователей, кто привык к развитым интерфейсам среды Windows, разработана серия свободно распространяемых оболочек программирования, например RStudio (Мастицкий, 2010); многие считают, что в ней очень удобно работать.

Мое знакомство с R началось с замечательной книги В. К. Шитикова и Г. С. Розенберга «Рандомизация и бутстреп: статистический анализ в биологии с использованием R» (2012). В ней на конкретных примерах разбираются сложные эколого-биологические задачи, решение которых иными методами крайне затруднительно или вовсе невозможно. Серьезно звучащий лейтмотив книги – «как из эмпирических данных извлечь всю доступную информацию, не впадая в ошибки любого рода?». Такая постановка вопроса вполне соответствует политике нашего журнала, ориентированной на поиск экологических закономерностей количественными методами. Остается освоить R.

Российская академия наук
Институт экологии Волжского бассейна

В.К. Шитиков, Г.С. Розенберг

Рандомизация и бутстреп:
статистический анализ в биологии и экологии
с использованием R.



Версия 5.4 от 12.11.2012

Тольятти 2012

Библиография

Зарядов И. С. Введение в статистический пакет R. [Introduction to the R statistical package] М.: Изд-во РУДНБ, 2010. 207 с. URL: <http://www.twirpx.com/file/556977/?rand=5181753> (дата обращения 21.12.2012).

Мастицкий С. R: Анализ и визуализация данных. [Data analysis and visualization] 2011. URL: <http://r-analytics.blogspot.de/p/rstudio.html> (дата обращения: 21.12.2012).

Шипунов А., Балдин Е. Введение в R. Ч. 1. [Introduction to R. Part 1.] // Linux Format. 2008. Vol. 100/101: Н. 98-101. URL: <http://www.inp.nsk.su/~baldin/DataAnalysis/R/> (дата обращения: 21.12.2012).

Шитиков В. К., Розенберг Г. С. Рандомизация и бутстреп: статистический анализ в биологии с использованием R. [Randomization and bootstrap: a statistical analysis in biology using R.] Тольятти, 2012. 151 с. URL: <http://www.ievbras.ru/ecostat/Kiril/Article/A32/Starb.pdf> (дата обращения: 21.12.2012).

Язык программирования R [Programming language R] // Викиучебник, 2011. URL:

Коросов А. В. R - значит победа // Принципы экологии. 2012. № 4. С. 65-67.

http://ru.wikibooks.org/wiki/Язык_программирования_R (дата обращения: 21.12.2012).

Burns P. The R Inferno // Burns statistics. 2011 URL: http://www.burns-stat.com/pages/Tutor/R_inferno.pdf (дата обращения: 21.12.2012).

<http://www.twirpx.com> (дата обращения: 21.12.2012).

Seefeld K., Linder E. Statistics Using R with Biological Examples. University of New Hampshire. 2007. URL: http://cran.r-project.org/doc/contrib/Seefeld_StatsRBio.pdf (дата обращения: 21.12.2012).

The Comprehensive R Archive Network. URL: <http://cran.gis-lab.info/> (дата обращения: 21.12.2012).

The R Project for Statistical Computing. URL: <http://www.r-project.org/> (дата обращения: 21.12.2012).

R means victory

**KOROSOV
Andrey**

Petrozavodsk state university, korosov@psu.karelia.ru