

УДК 311.2

М.В. Богданова

Л.С. Паршинцева

## СОВРЕМЕННЫЕ СРЕДСТВА ЭЛЕКТРОННОЙ ОБРАБОТКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ИНФОРМАЦИИ

*Аннотация.* В статье проанализированы современные наиболее популярные пакеты прикладных программ для статистического анализа экономической информации, выявлены их достоинства и недостатки, разработаны рекомендации по выбору статистических пакетов.

*Ключевые слова:* пакеты прикладных программ, социально-экономическая информация, статистический анализ, открытый и закрытый исходный код, классификатор.

Marina Bogdanova

Lidiya Parshintseva

## MODERN MEANS OF ELECTRONIC PROCESSING OF SOCIO-ECONOMIC INFORMATION

*Annotation.* The article analyzes most popular software packages for statistical analysis of economic information, identified their merits and demerits, developed recommendations for choice of statistical packages.

*Keywords:* software packages, socio-economic information, statistical analysis, open and closed source code, user classifier.

Сегодня современное общество невозможно представить без информационных технологий. Они прочно вошли в каждую сферу человеческой деятельности. И это не случайно, ведь информация является одним из важнейших ресурсов. Статистическая информация – это основа при принятии управленческих решений в экономической, социальной, демографической, инновационной сферах деятельности. Именно поэтому применение информационных технологий в статистике дает конкурентное преимущество как отдельно взятому предприятию, так и стране в целом [2; 16].

Использование информационных технологий в статистике связано с применением компьютерных программ, которые условно можно разделить на бесплатные (free software) и коммерческие (commercial), с открытым (open source) и закрытым (closed source) исходным кодом. Коммерческие программы, в отличие от бесплатных, распространяются с целью извлечения прибыли и имеют достаточно высокую стоимость: годовая стоимость лицензии на зарубежные программы для одного компьютера варьирует в пределах от 100 до 6000 долл. США. Отечественные статистические пакеты, к которым относятся Stadia, Олимп, Мезозавр, Эвриста, САНИ и другие, стоят сравнительно дешевле. Так, например, стоимость лицензии Stadia для одного пользователя составляет 15000 руб. [3].

Бесплатные статистические программы делятся на программы с открытым и закрытым исходным кодом. Преимуществом программного обеспечения с открытым исходным кодом является возможность его усовершенствования путем доработки и изменения исходного кода.

По назначению все статистические программные продукты можно разделить на три основные группы: профессиональные, универсальные и специализированные. К профессиональным программам для статистического анализа, отличительной особенностью которых является возможность работы со сверхбольшими объемами данных, относятся SAS и BMDP. Универсальные программы предназначены для выполнения анализа данных по широкому кругу статистических методов в различных сферах человеческой деятельности. Среди большого разнообразия универсальных программ отдельно следует выделить такие бесплатные программы, как среда R, Macanova, DataPlot, BrightStat, InStat+, AtteStat, Microsir, PSPP, JASP, Epi Info, SOFA, StatBase, PAST и SAS University Edition. В специализированных пакетах реализовано ограниченное число статистических методов для де-

тального анализа по изучаемой проблеме. К специализированным бесплатным статистическим программам относятся Gretl, Jmulti, Matrixer, Caterpillar 1.00, SALTSTAT (BETA), G7, FACTOR и другие; среди коммерческих – наибольший интерес представляют такие пакеты, как Мезозавр, Эвриста и САНИ [4].

Одной из самых востребованных и мощных универсальных программ, предназначенных для анализа данных, особенно среди зарубежных компаний, таких как Google, InterContinental Hotels Group, Shell и других, является среда R. Так, по мнению научного сотрудника Google Дэре-ла Прегибона: «R позволяет статистикам выполнять сложные виды анализа без глубокого знания внутреннего устройства вычислительных систем». R представляет собой язык программирования для обработки статистических данных. В отличие от целого ряда других статистических программ среда R имеет ряд преимуществ, среди которых: широкое представление статистических методов, кросс-платформенность (среда R компилируется и работает на разных платформах, таких как UNIX, Windows и MacOS), бесплатность, качественная графика, возможность работы с разными форматами данных, открытый исходный код, большое множество проверенных пакетов (более 7 тыс.), возможность создания скриптов и новых пакетов и другие. Как сказал главный экономист Google Хэл Вариян: «Огромная красота R заключается в том, что вы можете модифицировать его, чтобы делать всевозможные вещи» [17].

С точки зрения статистического анализа интересными представляются пакеты fBasics, forecast, fpp, FactoClass и другие.

Главными недостатками R являются сложность в использовании для рядового пользователя и недостаток русскоязычной информации о R.

Так, например, ввод данных в программу может осуществляться несколькими способами.

1. **Вручную.** Для этого необходимо создать переменную и задать необходимые значения с использованием оператора присваивания `<-` и функции `c()`, преобразующей ряд чисел в один вектор. Например:

```
Преступления<-c(347,499,577,654,434,487,344,331,356,1209,329,265,335,367,563,396,588,847)
```

В этом примере используются данные о числе преступлений, совершенных несовершеннолетними и при их участии, в субъектах Центрального федерального округа в 2015 г., соответственно в R с помощью представленной выше команды будет создана переменная Преступления. Этот способ удобен, если имеется небольшое количество значений. В противном случае, целесообразно выполнять ввод данных посредством импорта.

2. **Импорт из файлов различных форматов.** При импорте данных из полностью закрытых форматов, таких как MS Excel, в отличие от импорта данных из текстовых файлов с расширениями типа \*.csv, \*.txt и другие, могут возникнуть некоторые сложности с загрузкой. В связи с этим, одним из способов решения этой проблемы является использование вместо MS Excel программы Электронная таблица бесплатного пакета офисных программ OpenOffice. Также при импорте данных из файлов различных форматов следует знать следующие особенности:

- функция `read.table()` переносит данные из выбранного файла в заданную переменную;
- первый аргумент представляет собой имя файла, который должен лежать в рабочей директории. Проверить путь рабочей директории можно с помощью команды `>getwd()`. Поменять рабочую директорию можно командой `setwd()`;
- аргумент `sep` используется, если разделителем полей является любой другой символ, за исключением пробелов;
- параметр `dec` используют, чтобы уточнить разделитель в десятичных дробях;
- аргумент `header`, равный значению TRUE указывается, если первая строка таблицы представляет собой заголовки;

– описание остальных параметров функции `read.table()` можно получить с помощью выполнения команды `help(read.table)` [13].

Например, в рабочей директории сохранен файл под именем `Data.csv` с данными, представленными по столбцам, с заголовками в первой строке, при этом точка с запятой выступает в качестве разделителей полей. Посредством команды, представленной ниже, данные из файла `Data.csv` будут импортированы в переменную `data`, созданную в программе R:

```
> data<-read.table("Data.csv",sep=";",header=TRUE)
```

Для расчета описательных статистик в R предназначена функция `>summary`.

Таким образом, как видно из описания различных способов ввода данных, пользователь R должен не только владеть различными статистическими методами, но и знать основы программирования и на начальном этапе базовые команды языка R.

Классификация статистических пакетов прикладных программ, включая R, и их основные характеристики для статистического анализа социально-экономической информации приведены в таблице 1. Анализ таблицы 1 позволяет сделать следующие основные выводы:

- 1) современные бесплатные статистические программы не уступают коммерческим по своим возможностям и интерфейсу;
- 2) для анализа временных рядов и прогнозирования целесообразно остановиться на одном из специализированных пакетов, таких как `Gretl`, `Jmulti`, `Matrixer`, `Caterpillar 1.00`, `G7` и `Эвриста`, которые включают в себя исчерпывающий спектр статистических методов в этой области;
- 3) анализ статистических данных посредством многомерных статистических методов, таких как кластерный, факторный и другие, удобно проводить посредством программы `FACTOR` [1].

Исходя из вышесказанного, можно сделать вывод, что каждый пакет обладает определенным набором статистических методов, выбор которого должен в первую очередь отвечать целям и требованиям конкретного исследования и иметь удобный для работы интерфейс.

Таблица 1

**Классификация пакетов прикладных программ для статистического анализа социально-экономической информации и их основные характеристики**

Назначение пакета	Наименование пакета	Бесплатный пакет/ коммерческий пакет	Операционная система	Открытый/ закрытый исходный код	Специализация	Особенности
Профессиональные пакеты	SAS	Коммерческий	Кросс-платформенный	Закрытый	-	Возможность работы со сверхбольшими объемами данных. Широкий спектр методов [14]
	BMDP	Коммерческий	Кросс-платформенный	Закрытый	-	
Универсальные пакеты	среда R	Бесплатный	Кросс-платформенный	Открытый	-	Широкие возможности использования. Трудоемкость в освоении
	MacANOVA	Бесплатный	Кросс-платформенный	Открытый	-	Отсутствие литературы на русском языке. Сложности с установкой. Язык команд [9]

Назначение пакета	Наименование пакета	Бесплатный пакет/ коммерческий пакет	Операционная система	Открытый/ закрытый исходный код	Специализация	Особенности
	DataPlot	Бесплатный	Кросс-платформенный	Закрытый	-	Отсутствие версии программы и литературы по освоению программы на русском языке [6]
	BrightStat	Бесплатный	Кросс-платформенный	Закрытый	-	On-line сервер для статистического анализа данных [5]
	InStat+	Бесплатный	Для Windows	Закрытый	-	Пакет предназначен для выполнения простого статистического анализа [8]
	Microsiris	Бесплатный	Для Windows	Закрытый	-	Сложный интерфейс. Отсутствие литературы на русском языке [10]
	PSPP	Бесплатный	Кросс-платформенный	Открытый	-	Есть русскоязычная версия. Аналог SPSS [12]
	JASP	Бесплатный	Кросс-платформенный	Открытый	-	Аналог SPSS
	Epi Info	Бесплатный	Для Windows	Закрытый	-	Существует возможность создавать опросники и формы для ввода данных. Можно работать в on-line режиме [7]
Универсальные пакеты	SOFA	Бесплатный	Кросс-платформенный	закрытый	-	В программе не представлен регрессионный метод анализа. Преимуществом является графический метод анализа и возможность выполнять скрипты на Python
	StatBase	Бесплатный	Для Windows	закрытый	-	Программа не требует установки, может работать с USB-флеш-накопителя [11]
	PAST	Бесплатный	Кросс-платформенный	закрытый	-	

Назначение пакета	Наименование пакета	Бесплатный пакет/ коммерческий пакет	Операционная система	Открытый/ закрытый исходный код	Специализация	Особенности
	SAS University Edition	Коммерческий (бесплатный)	Кросс-платформенный	закрытый	-	Отсутствие русскоязычной версии программы. Интуитивно понятный интерфейс. Мощный язык программирования [14]
	MS Excel + AtteStat	Коммерческий	Для Windows	закрытый	-	-
	SPSS	Коммерческий	Кросс-платформенный	закрытый	-	Высокая стоимость пакета. Большой арсенал статистических методов
	Stata	Коммерческий	Кросс-платформенный	закрытый	-	
	Statistica	Коммерческий	Для Windows	закрытый	-	
Специализированные пакеты	Gretl	Бесплатный	Кросс-платформенный	открытый	анализ временных рядов и прогнозирование	Есть русскоязычная версия. Неудобный ввод данных
	Matrixer	Бесплатный	Для Windows	закрытый		-
	Caterpillar 1.00	Бесплатный	Для Windows	закрытый		Пакет предназначен только для анализа одномерных временных рядов [15]
	Jmulti	Бесплатный	Кросс-платформенный	открытый		Отсутствие литературы на русском языке
Специализированные пакеты	G7	Бесплатный	Для Windows	Закрытый	Анализ временных рядов и прогнозирование	-
	ЭВРИСТА 3.511 (учебная)	Коммерческий (бесплатный)	Для Windows	Закрытый		Отключены возможности импорта-экспорта данных (ввод – только вручную) и создания проектов
	Эвреста	Коммерческий	Для Windows	Закрытый		Специализируется на анализе финансовых рынков
	SALSTAT (BETA)	Бесплатный	Кросс-платформенный	Открытый		Дисперсионный анализ, проверка гипотез, графический метод анализа

Назначение пакета	Наименование пакета	Бесплатный пакет/ коммерческий пакет	Операционная система	Открытый/ закрытый исходный код	Специализация	Особенности
	FACTOR	Бесплатный	Для Windows	Закрытый	Многомерные статистические методы	Количество переменных для анализа не ограничено

Библиографический список

1. Богданова, М. В. Кластерная классификация факторных показателей нормы управляемости и компонентный анализ / М. В. Богданова, Л. С. Пантелина // Вестник Университета (Государственный университет управления). – 2013. – № 7. – С. 13–19.
2. Дуброва, А. М. Многомерные статистические методы и основы эконометрики : учеб. пособ. / А. М. Дуброва, В. С. Мхитарян, Л. И. Трошин. – М. : МЭСИ, 2002. – С. 79.
3. Научно-производственное общество «Информатика и компьютеры» [Электронный ресурс]. – Режим доступа : <http://statsoft.msu.ru/products.htm> (дата обращения : 10.04.2017).
4. Прикладная математическая статистика : учеб. пособ. / О. Г. Берестнева, О. В. Марухина, Г. Е. Шевелев. – Томск : Томский политехнический университет, 2012. – 188 с.
5. Программа BrightStat [Электронный ресурс]. – Режим доступа : <https://secure.brightstat.com/> (дата обращения : 18.04.2017).
6. Программа DataPlot [Электронный ресурс]. – Режим доступа : <http://www.itl.nist.gov/div898/software/dataplot/> (дата обращения : 18.04.2017).
7. Программа Epi Info [Электронный ресурс]. – Режим доступа : <https://www.cdc.gov/epiinfo/index.html> (дата обращения : 18.04.2017).
8. Программа InStat+ [Электронный ресурс]. – Режим доступа : <https://www.reading.ac.uk/ssc/resourcepage/instat.php> (дата обращения : 11.04.2017).
9. Программа MacANOVA [Электронный ресурс]. – Режим доступа : <http://www.stat.umn.edu/macanova/macanova.home.html> (дата обращения : 11.04.2017).
10. Программа Microsirir [Электронный ресурс]. – Режим доступа : <http://www.microsirir.com/> (дата обращения : 11.04.2017).
11. Программа PAST [Электронный ресурс]. – Режим доступа : <http://www.nhm.ac.uk/about-us/page-not-found.html> (дата обращения : 11.04.2017).
12. Программа PSPP [Электронный ресурс]. – Режим доступа : <https://www.gnu.org/software/pspp/> (дата обращения : 11.04.2017).
13. Программа R [Электронный ресурс]. – Режим доступа : <https://cran.r-project.org/mirrors.html> (дата обращения : 11.04.2017).
14. Программа SAS [Электронный ресурс]. – Режим доступа : [https://www.sas.com/ru\\_ru/home.html](https://www.sas.com/ru_ru/home.html) (дата обращения : 18.04.2017).
15. Программная реализация метода Гусеница [Электронный ресурс]. – Режим доступа : <http://www.gistatgroup.com/gus/programs.html> (дата обращения : 18.04.2017).
16. Симчера, В. М. Методы многомерного анализа статистических данных / В. М. Симчера. – М. : Финансы и статистика, 2008. – 400 с.
17. Vance, A. Data Analysts Captivated by R's Power [Electronic resource] / A. Vance // The New York Times. – 2009. – Mode of access : <http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html> (accessed date : 11.04.2017).