

УДК 311:004.9R
ББК 60.6с515
Ш63

Ш63 А.Б. Шипунов, Е.М. Балдин, П.А. Волкова, А.И. Коробейников,
С.А.Назарова, С.В. Петров, В.Г. Суфиянов
Наглядная статистика. Используем R! – М.: ДМК Пресс – 298 с.: ил.

ISBN 978-5-97060-473-1

Если вам необходима статистическая обработка данных для курсовой, диплома, статьи или диссертации; вы хотите лучше понимать результаты тех статистических методов, которые применяете; вы устали от того, что программы анализа данных не способны выполнить нестандартные задачи; вам необходимо перегруппировать ваши данные, но жаль тратить на это часы ручного труда; вам нужно освоить самые современные методы, еще не нашедшие отражения в большинстве статистических пакетов, то эта книга – для вас!

Изложение построено на базе самого современного программного обеспечения – статистической среды R, которая принадлежит к числу наиболее динамически развивающихся программ в своем классе.

Освоив R, вы сможете: полностью автоматизировать свою работу; запускать статистическую обработку прямо из текста документа; получать графики высокого качества и сохранить их в переносимых графических форматах; в любой момент повторить ваш анализ (например, если поменялись требования к иллюстрациям или исходные данные); использовать сотни «библиотек»-плагинов, разработанных для R; применять самые современные методы; разрабатывать собственные программы анализа данных: от коротких «макросов» до полноценных пакетов, реализующих новейшие алгоритмы; и, естественно, проводить любой стандартный анализ данных, получая при этом графики любой степени сложности.

УДК311:004.9R
ББК 60.6с515

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-5-97060-473-1

© А.Б. Шипунов и др.
© Оформление, издание, ДМК Пресс

Оглавление

Предисловие	7
Глава 1. Что такое данные и зачем их обрабатывать?	10
1.1. Откуда берутся данные	10
1.2. Генеральная совокупность и выборка	12
1.3. Как получать данные	13
1.4. Что ищут в данных	17
Глава 2. Как обрабатывать данные	21
2.1. Неспециализированные программы	21
2.2. Специализированные статистические программы	22
2.2.1. Оконно-кнопочные системы	22
2.2.2. Статистические среды	24
2.3. Из истории S и R	24
2.4. Применение, преимущества и недостатки R	25
2.5. Как скачать и установить R	27
2.6. Как начать работать в R	28
2.6.1. Запуск	28
2.6.2. Первые шаги	29
2.7. R и работа с данными: вид снаружи	30
2.7.1. Как загружать данные	30
2.7.2. Как сохранять результаты	36
2.7.3. R как калькулятор	37
2.7.4. Графики	38
2.7.5. Графические устройства	40
2.7.6. Графические опции	42
2.7.7. Интерактивная графика	43
Глава 3. Типы данных	46
3.1. Градусы, часы и километры: интервальные данные	46
3.2. «Садись, двойка»: шкальные данные	49
3.3. Красный, желтый, зеленый: номинальные данные	50
3.4. Доли, счет и ранги: вторичные данные	55
3.5. Пропущенные данные	59
3.6. Выбросы и как их найти	61

3.7.	Меняем данные: основные принципы преобразования . . .	61
3.8.	Матрицы, списки и таблицы данных	63
3.8.1.	Матрицы	63
3.8.2.	Списки	65
3.8.3.	Таблицы данных	68
Глава 4.	Великое в малом: одномерные данные	72
4.1.	Как оценивать общую тенденцию	72
4.2.	Ошибочные данные	82
4.3.	Одномерные статистические тесты	83
4.4.	Как создавать свои функции	87
4.5.	Всегда ли точны проценты	90
Глава 5.	Анализ связей: двумерные данные	94
5.1.	Что такое статистический тест	94
5.1.1.	Статистические гипотезы	94
5.1.2.	Статистические ошибки	95
5.2.	Есть ли различие, или Тестирование двух выборок . . .	96
5.3.	Есть ли соответствие, или Анализ таблиц	102
5.4.	Есть ли взаимосвязь, или Анализ корреляций	109
5.5.	Какая связь, или Регрессионный анализ	114
5.6.	Вероятность успеха, или Логистическая регрессия . . .	123
5.7.	Если выборка больше двух	127
Глава 6.	Анализ структуры: data mining	142
6.1.	Рисуем многомерные данные	142
6.1.1.	Диаграммы рассеяния	143
6.1.2.	Пиктограммы	146
6.2.	Тени многомерных облаков: анализ главных компонент	149
6.3.	Классификация без обучения, или Кластерный анализ	155
6.4.	Классификация с обучением, или Дискриминантный анализ	164
Глава 7.	Узнаем будущее: анализ временных рядов	173
7.1.	Что такое временные ряды	173
7.2.	Тренд и период колебаний	173
7.3.	Построение временного ряда	174
7.4.	Прогноз	181
Глава 8.	Статистическая разведка	190
8.1.	Первичная обработка данных	190
8.2.	Окончательная обработка данных	190

8.3. Отчет	191
Приложение А. Пример работы в R	196
Приложение Б. Графический интерфейс (GUI) для R	207
Б.1. R Commander	207
Б.2. RStudio	209
Б.3. RKWard	211
Б.4. Revolution-R	211
Б.5. JGR	214
Б.6. Rattle	215
Б.7. rpanel	216
Б.8. ESS и другие IDE	218
Приложение В. Основы программирования в R	220
В.1. Базовые объекты языка R	220
В.1.1. Вектор	220
В.1.2. Список	221
В.1.3. Матрица и многомерная матрица	222
В.1.4. Факторы	223
В.1.5. Таблица данных	224
В.1.6. Выражение	224
В.2. Операторы доступа к данным	225
В.2.1. Оператор [с положительным аргументом	225
В.2.2. Оператор [с отрицательным аргументом	226
В.2.3. Оператор [со строковым аргументом	226
В.2.4. Оператор [с логическим аргументом	227
В.2.5. Оператор \$	227
В.2.6. Оператор [[.	228
В.2.7. Доступ к табличным данным	229
В.2.8. Пустые индексы	231
В.3. Функции и аргументы	231
В.4. Циклы и условные операторы	234
В.5. R как СУБД	235
В.6. Правила переписывания. Векторизация	238
В.7. Отладка	243
В.8. Элементы объектно-ориентированного программирования в R	246
Приложение Г. Выдержки из документации R	249
Г.1. Среда R	249
Г.2. R и S	250
Г.3. R и статистика	250

Г.4.	Получение помощи	250
Г.5.	Команды R	251
Г.6.	Повтор и коррекция предыдущих команд	252
Г.7.	Сохранение данных и удаление объектов	252
Г.8.	Внешнее произведение двух матриц	253
Г.9.	<code>c()</code>	254
Г.10.	Присоединение	254
Г.11.	<code>scan()</code>	255
Г.12.	R как набор статистических таблиц	256
Г.13.	Область действия	256
Г.14.	Настройка окружения	260
Г.15.	Графические функции	261
Г.15.1.	<code>plot()</code>	262
Г.15.2.	Отображение многомерных данных	263
Г.15.3.	Другие графические функции высокого уровня	264
Г.15.4.	Параметры функций высокого уровня	265
Г.15.5.	Низкоуровневые графические команды	266
Г.15.6.	Математические формулы	269
Г.15.7.	Интерактивная графика	269
Г.15.8.	<code>par()</code>	270
Г.15.9.	Список графических параметров	272
Г.15.10.	Края рисунка	275
Г.15.11.	Составные изображения	276
Г.15.12.	Устройства вывода	277
Г.15.13.	Несколько устройств вывода одновременно	278
Г.16.	Пакеты	279
Г.16.1.	Стандартные и сторонние пакеты	280
Г.16.2.	Пространство имен пакета	280
Приложение Д. Краткий словарь языка R		282
Приложение Е. Краткий словарь терминов		285
Литература		291
Об авторах		293