

Б А К А Л А В Р И А Т

А.А. Григорьев, Е.А. Исаев

МЕТОДЫ И АЛГОРИТМЫ ОБРАБОТКИ ДАННЫХ

У Ч Е Б Н О Е П О С О Б И Е



Электронно-
Библиотечная
Система
znanium.com



Уважаемый читатель!

Вы держите в руках книгу,
дополнительные материалы которой
доступны Вам **БЕСПЛАТНО**
в интернете на www.znanium.com
Специального программного
обеспечения не требуется

ВЫСШЕЕ ОБРАЗОВАНИЕ – БАКАЛАВРИАТ

серия основана в 1996 г.



А.А. ГРИГОРЬЕВ
Е.А. ИСАЕВ

МЕТОДЫ И АЛГОРИТМЫ ОБРАБОТКИ ДАННЫХ

УЧЕБНОЕ ПОСОБИЕ

2-е издание, переработанное и дополненное

*Рекомендовано Межрегиональным учебно-методическим советом
профессионального образования в качестве учебного пособия
для студентов высших учебных заведений, обучающихся
по экономическим направлениям подготовки и специальностям
(протокол № 8 от 22.06.2020)*

Электронно-
Библиотечная
znanium.com

Москва
ИНФРА-М
2021

УДК 330.47(075.8)

ББК 65.050я73

Г83

Рецензенты:

Сиденко А.В., доктор экономических наук, профессор, президент Российской академии диалектико-системных исследований и разработок, директор Института статистики Российской академии диалектико-системных исследований и разработок.

Корнилов В.В., кандидат физико-математических наук, доцент кафедры инноваций и бизнеса в сфере информационных технологий, заместитель руководителя Школы бизнес-информатики Национального исследовательского университета «Высшая школа экономики»

Григорьев А.А.

Г83

Методы и алгоритмы обработки данных : учебное пособие / А.А. Григорьев, Е.А. Исаев. — 2-е изд., перераб. и доп. — Москва : ИНФРА-М, 2021. — 383 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/1032305.

ISBN 978-5-16-015581-4 (print)

ISBN 978-5-16-108009-2 (online)

В учебном пособии рассматриваются некоторые методы и алгоритмы обработки данных, последовательность решения задач обработки и анализа данных для создания модели поведения объекта с учетом всех компонент его математической модели. Описываются виды технологических методов использования программно-аппаратных средств для решения задач в этой области. Рассматриваются алгоритмы распределений, регрессий временных рядов, их преобразование с целью получения математических моделей и прогноза поведения информационно-экономических систем (объектов).


Второе издание дополнено материалами, которые востребованы исследователями в части корректного использования алгоритмов кластеризации. Приводятся элементы классификации алгоритмов с определением их возможностей, достоинств и недостатков. Представляются процедуры обоснования и проверки адекватности результатов кластерного анализа, проводятся сравнение и оценка различных методов кластеризации, даются сведения о визуализации многомерных данных и примеры практического применения алгоритмов кластеризации.

Соответствует требованиям федеральных государственных образовательных стандартов высшего образования последнего поколения.

Для студентов экономических специальностей, специалистов, аспирантов.

УДК 330.47(075.8)

ББК 65.050я73

Материалы, отмеченные знаком , доступны в электронно-библиотечной системе Znanium.com

ISBN 978-5-16-015581-4 (print)
ISBN 978-5-16-108009-2 (online)

© Григорьев А.А., 2017
© Григорьев А.А., Исаев Е.А., 2020,
с изменениями

Оглавление

Введение	3
ГЛАВА 1. ОСНОВНЫЕ КОМПОНЕНТЫ В АЛГОРИТМИЗАЦИИ И ПОСЛЕДОВАТЕЛЬНОСТЬ РЕШЕНИЯ ЗАДАЧИ ОБРАБОТКИ ДАННЫХ	6
1.1. Цели и задачи обработки данных.....	6
1.2. Технологии вычислений.....	9
1.2.1. Параллельные вычисления.....	9
1.2.2. Виды параллельного взаимодействия.....	9
1.2.3. Параллельные вычислительные системы.....	10
1.2.4. Типы параллелизма.....	10
1.3. Вычислительный конвейер.....	12
1.3.1. Общий конвейер.....	12
1.3.2. Последовательные и векторные конвейеры.....	14
1.4. Введение в параллельную обработку.....	15
ГЛАВА 2. ПАРАЛЛЕЛЬНАЯ SQL-ОБРАБОТКА	19
2.1. Архитектура параллельной SQL-обработки. Разделение данных.....	19
2.2. Параллельная SQL-обработка.....	23
2.3. Конвейерная обработка данных.....	24
ГЛАВА 3. АНАЛИЗ ДАННЫХ.....	28
3.1. Этапы анализа.....	28
3.2. Подготовка исходного набора данных.....	29
3.3. Анализ данных (<i>data mining</i>).....	31
ГЛАВА 4. МЕТОДЫ И МОДЕЛИ АНАЛИЗА ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ	34
4.1. Этапы анализа экономических процессов.....	34
4.2. Прогнозирование экономических процессов.....	37
4.3. Многофакторные (эконометрические) модели анализа и прогнозирования экономических систем.....	47
4.4. Модели сетевого планирования и управления.....	51
4.5. Экономико-математические модели в маркетинге.....	61
4.6. Экономико-математические модели в области экономики и социологии труда.....	62
ГЛАВА 5. ЗАДАЧИ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ	64
5.1. Задачи анализа временных рядов.....	64
5.2. Формы тренда.....	66
5.3. Сезонность.....	67
ГЛАВА 6. О РАСПРЕДЕЛЕНИЯХ	72
6.1. Случайная переменная.....	72
6.2. Совместное распределение случайных величин.....	72
6.3. Оценивание.....	73
6.4. Характеристики оценок.....	73
6.5. Распределения.....	75
6.5.1. Нормальное распределение.....	75
6.5.2. χ^2 -распределение.....	75

6.5.3. t -распределение	76
6.5.4. F -распределение.....	78
6.5.5. Квантиль	79
ГЛАВА 7. МНК — ОДИН ИЗ МЕТОДОВ ОЦЕНКИ ПАРАМЕТРОВ ЛИНЕЙНОЙ ЗАВИСИМОСТИ	80
ГЛАВА 8. ПАРНАЯ РЕГРЕССИЯ. ВЕРОЯТНОСТНАЯ ПРИРОДА РЕГРЕССИОННЫХ МОДЕЛЕЙ	83
8.1. Теорема Гаусса — Маркова.....	85
8.2. Проверка гипотез и доверительные интервалы	86
ГЛАВА 9. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ	90
9.1. Оценка параметров	90
9.2. Анализ модели множественной регрессии	93
9.3. Использование регрессионных моделей	95
9.4. Пропущенные наблюдения.....	103
9.5. Прогнозирование на основе регрессионной модели из одного уравнения	105
9.6. Условный прогноз	111
ГЛАВА 10. АНАЛИЗ ОСТАТКОВ.....	113
10.1. Автокорреляция и автокорреляционная функция	113
10.2. Стационарные ряды. Свойства стационарных процессов	116
ГЛАВА 11. МОДЕЛЬ АВТОРЕГРЕССИИ И СКОЛЬЗЯЩЕГО СРЕДНЕГО	120
11.1. Модель авторегрессии	120
11.2. Модель скользящего среднего.....	122
11.3. Идентификация (выбор структуры) модели.....	125
ГЛАВА 12. РАСПРЕДЕЛЕННЫЕ ЛАГИ	131
12.1. Геометрический лаг	131
12.2. Полиномиальный распределенный лаг	133
ГЛАВА 13. МОДЕЛИ КАЧЕСТВЕННОГО ВЫБОРА	135
13.1. Модели бинарного выбора	135
13.2. Линейная вероятностная модель.....	136
13.3. Пробит-модель	138
13.4. Логит-модель	140
ГЛАВА 14. ПРОГНОЗИРОВАНИЕ НА ОСНОВЕ СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ.....	144
14.1. Модели уравнений	144
14.2. Рекурсивные системы уравнений	145
ГЛАВА 15. ВРЕМЕННЫЕ РЯДЫ.....	150
15.1. Классификация рядов динамики, правила их построения.....	150
15.2. Показатели изменения уровней рядов динамики	155
15.3. Компоненты временных рядов	163
15.4. Стационарные временные ряды и их основные характеристики.....	165
15.5. Сглаживание временных рядов с помощью скользящих средних	168
15.6. Использование взвешенных скользящих средних	172
15.7. Восстановление краевых значений.....	178
15.8. Использование скользящих средних в техническом анализе	181

ГЛАВА 16. АНАЛИТИЧЕСКОЕ ВЫРАВНИВАНИЕ ВРЕМЕННЫХ РЯДОВ И МОДЕЛИРОВАНИЕ ТЕНДЕНЦИИ РАЗВИТИЯ 184

16.1. Проверка гипотезы о существовании тренда.....	184
16.2. Применение моделей кривых роста для анализа и прогнозирования основной тенденции развития.....	186
16.3. Методы выбора кривых роста.....	199
16.4. Оценка адекватности и точности выбранных моделей.....	203

ГЛАВА 17. СТАТИСТИЧЕСКИЙ АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ПЕРИОДИЧЕСКИХ КОЛЕБАНИЙ 212

17.1. Статистические методы оценки уровня сезонности. Прогнозирование с помощью тренд-сезонных моделей.....	212
17.2. Сезонная декомпозиция и корректировка временных рядов.....	218
17.3. Моделирование сезонных колебаний с помощью фиктивных переменных.....	221
17.4. Исследование периодических колебаний методами спектрального анализа.....	224
17.5. Фундаментальный анализ ценных бумаг в условиях эффективного рынка капитала. Прогнозирование в направлениях сверху вниз и снизу вверх.....	230

ГЛАВА 18. КЛАСТЕРНЫЙ АНАЛИЗ КАК ИНСТРУМЕНТ ОБРАБОТКИ ДАННЫХ 236

18.1. Особенности кластерного анализа.....	236
18.2. Задача кластеризации.....	241
18.3. Описание объектов кластеризации и выбор переменных.....	243
18.4. Сравнительный анализ иерархических и неиерархических методов кластеризации.....	247
18.5. Меры сходства в кластеризации.....	251
18.6. Проблемы кластерного анализа.....	258
18.6.1. Алгоритм последовательной кластеризации.....	258
18.6.2. Число кластеров.....	261
18.7. Методы кластерного анализа.....	264
18.7.1. Классификация методов кластерного анализа.....	264
18.7.2. Классификация алгоритмов кластеризации по методам.....	267
18.8. Алгоритмы кластеризации.....	267
18.8.1. Иерархические агломеративные алгоритмы.....	268
18.8.2. Иерархические дивизимные алгоритмы.....	274
18.8.3. Итеративные алгоритмы.....	287
18.8.4. Неиерархическая кластеризация.....	296
18.8.5. Нечеткая кластеризация.....	305
18.9. Самоорганизующиеся карты Кохонена.....	330
18.10. Достоинства, недостатки методов кластеризации.....	338
18.11. Представление результатов кластеризации.....	340
18.11.1. Визуализация многомерных данных.....	341
18.11.2. Реализации OLAP.....	360
18.11.3. Визуализаторы для оценки качества моделей.....	361
18.12. Примеры практических применений алгоритмов кластеризации.....	362
18.12.1. Кластеризация продукции автомобильного рынка.....	363
18.12.2. Кластеризация возрастных распределений населения.....	363
18.12.3. Нечеткий кластерный анализ показателей крупных городов России.....	364

Темы для самостоятельного изучения 369

Библиографический список 373

Рекомендованная литература 376



ПРИЛОЖЕНИЯ..... 378

**Приложение 1. Аддитивные и мультипликативные методы
прогнозирования 378-1**

**Приложение 2. Основные статистические формулы
с инструментами в EXCEL.....378-11**

Приложение 3. Алгоритм прогнозирования объема продаж.....378-23

Приложение 4. Состав методов прогнозирования378-39