

УДК 004.43  
ББК 32.372.1  
Б98

**Бюиссон Ф.**

Б98 Анализ поведенческих данных на R и Python / пер. с англ. А. В. Логунова. – М.: ДМК Пресс, 2022. – 368 с.: ил.

**ISBN 978-5-97060-992-7**

Задействуйте всю мощь поведенческих данных в своей компании, используя инструменты, специально разработанные для их анализа. Автор, эксперт в области экономики и бихевиористики, показывает, как повысить ценность и результаты аналитических проектов за счет понимания того, что движет поведением людей. Практическая часть книги содержит полные примеры и упражнения на языках R и Python, которые помогут вам получать более глубокую информацию о данных.

Издание предназначено для бизнес-аналитиков и других специалистов, исследующих данные и владеющих программированием на R или Python. Для чтения требуется минимальное знакомство с линейной и логистической регрессией.

УДК 004.43  
ББК 32.372.1

Authorized Russian translation of the English edition of Behavioral Data Analysis with R and Python ISBN 9781492061373. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same. Russian language edition copyright © 2022 by DMK Press. All rights reserved.

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.

ISBN 978-1-492-06137-3 (англ.)  
ISBN 978-5-97060-992-7 (рус.)

© Florent Buisson, 2021  
© Перевод, оформление, издание,  
ДМК Пресс, 2022

# Содержание

<b>От издательства.....</b>	<b>11</b>
<b>Предисловие.....</b>	<b>12</b>
<b>Благодарности .....</b>	<b>21</b>
<b>Об авторе.....</b>	<b>22</b>
<b>Об иллюстрации на обложке (колофон) .....</b>	<b>23</b>
<b>Часть I. ПОНИМАНИЕ ПОВЕДЕНИЙ .....</b>	<b>24</b>
<b>Глава 1. Причинно-поведенческий каркас для анализа данных.....</b>	<b>25</b>
Почему для объяснения человеческого поведения нужна причинно-следственная аналитика.....	26
Различные типы аналитики.....	26
Люди – сложные существа.....	27
Чтоб ей пусто было! Скрытые опасности, когда разбирательства отданы на усмотрение регрессии.....	30
Данные .....	31
Почему корреляция не есть каузация: спутывающий фактор в действии ...	32
Слишком много переменных может испортить всю обедню .....	34
Выводы .....	40
<b>Глава 2. Понимание поведенческих данных.....</b>	<b>41</b>
Базовая модель человеческого поведения.....	42
Личностные характеристики .....	43
Познание и эмоции .....	45
Намерения .....	46
Действия.....	48
Поведения бизнеса .....	49
Как соединять поведения и данные.....	50
Развивать бихевиористски целостный менталитет.....	51
Не доверять и проверять .....	52
Выявлять категорию.....	53

Уточнять поведенческие переменные .....	55
Понимать контекст.....	56
Выводы .....	59

## **Часть II. ПРИЧИННО-СЛЕДСТВЕННЫЕ ДИАГРАММЫ И РАСПУТЫВАНИЕ .....60**

### **Глава 3. Введение в причинно-следственные диаграммы.....61**

Причинно-следственные диаграммы и причинно-поведенческий каркас.....	62
Причинно-следственные диаграммы представляют поведения .....	63
Причинно-следственные диаграммы представляют данные .....	65
Фундаментальные структуры причинно-следственных диаграмм.....	69
Цепочки.....	69
Развилки.....	73
Сталкиватели.....	75
Распространенные преобразования причинно-следственных диаграмм.....	77
Нарезка/дезагрегирование переменных .....	77
Агрегирование переменных .....	78
А что делать с циклами? .....	80
Пути .....	84
Выводы .....	85

### **Глава 4. Строительство причинно-следственных диаграмм с нуля .....87**

Деловая задача и настройка данных.....	88
Данные и пакеты.....	89
Понимание интересующей взаимосвязи.....	89
Выявление переменных-кандидатов на включение .....	91
Действия.....	93
Намерения .....	94
Познание и эмоции .....	95
Личностные характеристики .....	96
Поведения бизнеса.....	99
Временные тренды.....	100
Подтверждение наблюдаемых переменных для включения на основе данных .....	101
Взаимосвязи между числовыми переменными.....	102
Взаимосвязи между категориальными переменными.....	105
Взаимосвязи между числовыми и категориальными переменными .....	108
Итеративное расширение причинно-следственной диаграммы .....	110
Выявление косвенных индикаторов для ненаблюдаемых переменных.....	111
Выявление дальнейших причин .....	112
Итеративный повтор.....	113
Упрощения причинно-следственной диаграммы .....	113
Выводы .....	115

<b>Глава 5. Использование причинно-следственных диаграмм для распутывания аналитических расчетов .....</b>	<b>116</b>
Деловая задача: продажи мороженого и бутилированной воды.....	117
Критерий дизъюнктивной причины .....	120
Определение.....	120
Первый блок .....	120
Второй блок .....	122
Критерий боковой двери .....	123
Определения.....	123
Первый блок .....	126
Второй блок .....	127
Выводы .....	129

### **Часть III. УСТОЙЧИВЫЙ АНАЛИЗ ДАННЫХ .....**

130

<b>Глава 6. Работа с пропущенными данными.....</b>	<b>131</b>
Данные и пакеты.....	133
Визуализация пропущенных данных .....	134
Объем пропущенных данных .....	137
Корреляция пропущенности.....	139
Диагностика пропущенных данных .....	144
Причины пропущенности: классификация Рубина .....	147
Диагностика переменных MCAR.....	149
Диагностика переменных MAR .....	151
Диагностика переменных MNAR .....	153
Пропущенность как спектр .....	155
Работа с пропущенными данными .....	159
Введение во множественное вменение (MI) .....	160
Метод вменения по умолчанию: соотнесение с предсказательным средним значением.....	162
От PMM к нормальному вменению (только для R).....	164
Добавление вспомогательных переменных.....	166
Вертикальное масштабирование числа наборов вмененных данных .....	168
Выводы .....	169

<b>Глава 7. Измерение неопределенности с помощью бутстрапа .....</b>	<b>171</b>
Введение в бутстрап: «опрашивание» самого себя.....	172
Пакеты .....	172
Деловая задача: малые данные с выбросом .....	172
Бутстраповский интервал уверенности для выборочного среднего .....	174
Бутстраповские интервалы уверенности для нерегламентированной статистики .....	180
Бутстрап для регрессионного анализа.....	182
Когда следует использовать бутстрап .....	185

Условия достаточности традиционной центральной оценки .....	186
Условия достаточности традиционного интервала уверенности .....	187
Определение числа бутстраповских выборок .....	189
Оптимизирование бутстрапа на R и Python .....	191
R: пакет boot .....	191
Оптимизация на Python .....	194
Выводы .....	195

## **Часть IV. ДИЗАЙН И АНАЛИЗ ЭКСПЕРИМЕНТОВ .....** 196

### **Глава 8. Экспериментальный дизайн: основы.....** 198

Планирование эксперимента: теория изменения .....	199
Деловая цель и целевая метрика .....	200
Вмешательство .....	203
Поведенческая логика .....	205
Данные и пакеты .....	207
Определение случайного размещения и размера/мощности выборки .....	208
Случайное размещение .....	208
Размер выборки и анализ мощности .....	211
Анализирование и интерпретирование экспериментальных результатов.....	226
Выводы .....	229

### **Глава 9. Стратифицированная рандомизация .....** 230

Планирование эксперимента .....	232
Деловая цель и целевая метрика .....	232
Определение вмешательства .....	234
Поведенческая логика .....	235
Данные и пакеты .....	235
Определение случайного размещения и размера/мощности выборки .....	236
Случайное размещение .....	237
Анализ мощности с помощью бутстраповских симуляций .....	245
Анализ и интерпретация экспериментальных результатов .....	252
Оценка намерения относительно экспериментальной процедуры для стимулирования вмешательства .....	253
Оценка причинно-следственного эффекта среднего по соблюдающим требования испытуемым в целях обязательного вмешательства .....	254
Выводы .....	260

### **Глава 10. Кластерная рандомизация и иерархическое моделирование .....** 262

Планирование эксперимента .....	263
Деловая цель и целевая метрика .....	263
Определение вмешательства .....	263
Поведенческая логика .....	265
Данные и пакеты .....	265

Введение в иерархическое моделирование .....	266
Исходный код на R .....	267
Исходный код на Python .....	270
Определение случайного размещения и размера/мощности выборки .....	272
Случайное размещение .....	272
Анализ мощности .....	274
Анализ эксперимента .....	282
Выводы .....	282

## **Часть V. ПРОДВИНУТЫЕ ИНСТРУМЕНТЫ АНАЛИЗА ПОВЕДЕНЧЕСКИХ ДАННЫХ ..... 284**

### **Глава 11. Введение в модерацию ..... 285**

Данные и пакеты .....	286
Поведенческие разновидности модерации .....	286
Сегментация .....	286
Взаимодействия .....	293
Нелинейности .....	294
Как применять модерацию .....	297
Когда следует искать модерацию? .....	298
Несколько модераторов .....	309
Подтверждение модерации с помощью бутстрапа .....	315
Интерпретирование отдельных коэффициентов .....	317
Выводы .....	323

### **Глава 12. Опосредование и инструментальные переменные .... 325**

Опосредование .....	326
Понимание причинно-следственных механизмов .....	326
Причинно-следственные систематические смещения .....	328
Выявление опосредования .....	329
Измерение опосредования .....	331
Инструментальные переменные .....	336
Данные .....	336
Пакеты .....	337
Понимание и применение инструментальных переменных .....	337
Измерение .....	340
Применение инструментальных переменных: часто задаваемые вопросы .....	343
Выводы .....	344

### **Библиография ..... 346**

### **Предметный указатель ..... 350**