

Министерство образования и науки РФ  
Алтайский государственный университет

С.В. Дронов

МЕТОДЫ И ЗАДАЧИ  
МНОГОМЕРНОЙ СТАТИСТИКИ

*Монография*



Барнаул

---

Издательство  
Алтайского государственного  
университета

2015

УДК 517.0  
ББК 22.172  
Д 758

*Рецензенты:*

д .ф.-м.н. Родионов Е.Д. (АлтГУ),  
к.ф.-м.н. Кизбикенов К.О. (АлтГПА).

**Д 758 Дронов, С.В.**

Методы и задачи многомерной статистики [Текст] : монография. –  
Барнаул : Изд-во Алт. ун-та, 2015. – 275 с.  
ISBN 978-5-7904-1957-7

В книге систематически рассмотрены основные современных методы анализа многомерных статистических данных. Включены подробные алгоритмы и примеры. Она также содержит новые результаты, связанные с оцифровкой нечисловых категоризованных данных, которые получены автором и ранее публиковались только в виде статей в специализированных журналах. Изложение областей многомерной статистики, ставших уже традиционными, основано на многолетнем опыте преподавания автором курсов многомерного статистического анализа и эконометрики. Предназначена для широкого круга исследователей, применяющих в своей практике методы многомерной статистики.

УДК 517.0  
ББК 22.172

ISBN 978-5-7904-1957-7

© Дронов С.В., 2015.

© Оформление: Издательство Алтайского государственного  
университета, 2015.

# Оглавление

<b>1</b>	<b>Предварительные сведения</b>	<b>7</b>
1.1	Теория вероятностей и математическая статистика . . . . .	8
1.2	Нечисловые данные . . . . .	14
1.3	Нормальное распределение . . . . .	16
<b>2</b>	<b>Изучение связи числовых показателей</b>	<b>22</b>
2.1	Понятие статистической связи . . . . .	22
2.2	Проверка независимости . . . . .	24
2.3	Коэффициент парной корреляции Пирсона . . . . .	26
2.4	Корреляционная теория . . . . .	31
<b>3</b>	<b>Группировка и цензурирование</b>	<b>41</b>
3.1	Группировка в одномерном случае . . . . .	41
3.2	Одномерное цензурирование . . . . .	44
3.3	Многомерная группировка . . . . .	46
3.4	О цензурировании многомерных данных . . . . .	48
<b>4</b>	<b>Алгоритмы кластерного анализа</b>	<b>50</b>
4.1	Постановка задачи . . . . .	50
4.2	Иерархические алгоритмы . . . . .	51
4.3	Оцифровка иерархий . . . . .	60
4.4	Коэффициент кластерных различий . . . . .	63
4.5	Два классических алгоритма . . . . .	66
4.5.1	Алгоритм $k$ -средних . . . . .	67
4.5.2	Алгоритм ФОРЕЛЬ . . . . .	68

<b>5</b>	<b>Оценка силы связи нечисловых показателей</b>	<b>70</b>
5.1	Вводные замечания . . . . .	70
5.2	Ранговые шкалы . . . . .	72
5.2.1	Экспертные оценки . . . . .	76
5.2.2	Согласованные группы экспертов . . . . .	79
5.3	Два категорированных показателя . . . . .	82
5.3.1	Таблицы сопряженности общего вида . . . . .	82
5.3.2	Четырехпольные таблицы . . . . .	84
5.3.3	Коэффициент относительного риска . . . . .	86
5.4	Коэффициент $E$ . . . . .	89
<b>6</b>	<b>Регрессионный анализ</b>	<b>94</b>
6.1	Постановка задачи . . . . .	94
6.2	Нормальное уравнение регрессии . . . . .	98
6.3	Задачи с ограничениями . . . . .	104
6.4	Оптимальный выбор матрицы плана . . . . .	105
6.5	Задача статистического прогноза . . . . .	108
<b>7</b>	<b>Дисперсионный анализ</b>	<b>112</b>
7.1	Вводные замечания . . . . .	112
7.1.1	Проверка гипотезы нормальности . . . . .	114
7.1.2	Однородность дисперсий . . . . .	116
7.2	Однофакторный анализ . . . . .	117
7.3	Несколько факторов . . . . .	119
7.4	Некоторые примеры . . . . .	120
<b>8</b>	<b>Методы анализа соответствий</b>	<b>125</b>
8.1	Случай двух показателей . . . . .	125
8.1.1	Расстояние хи-квадрат . . . . .	127
8.1.2	Матрицы рассеивания . . . . .	129
8.1.3	Оцифровка для задач кластерного анализа . . . . .	133
8.2	Множественный анализ соответствий . . . . .	137
8.2.1	Бинарная матрица как таблица сопряженности . . . . .	138
8.2.2	Максимальные корреляции . . . . .	140
8.3	О выборе размерности меток . . . . .	141
8.4	Пример с профессиями . . . . .	143
8.5	Случай смешанных данных . . . . .	146

<b>9</b>	<b>Снижение размерности</b>	<b>150</b>
9.1	Сущность задачи снижения размерности . . . . .	150
9.2	Метод главных компонент . . . . .	153
9.2.1	Формальная постановка задачи . . . . .	154
9.2.2	Вычисление главных компонент . . . . .	156
9.2.3	Численная иллюстрация метода . . . . .	159
9.3	Метод экстремальной группировки признаков . . . . .	161
9.3.1	Критерий квадратов . . . . .	162
9.3.2	Критерий модулей . . . . .	165
9.4	Метод корреляционных плеяд . . . . .	167
9.4.1	Построение плеяд . . . . .	167
9.4.2	Дерево зависимостей показателей . . . . .	168
<b>10</b>	<b>Факторный анализ</b>	<b>170</b>
10.1	Постановка задачи . . . . .	170
10.1.1	Чем этот метод отличается от остальных? . . . . .	170
10.1.2	Интерпретация решения . . . . .	172
10.2	Математическая модель . . . . .	175
10.2.1	Подходы к решению основного уравнения ФА . . . . .	177
10.2.2	Центроидный метод . . . . .	179
10.3	Алгоритм факторного анализа по шагам . . . . .	182
10.4	Вращение решений . . . . .	185
10.5	Оценивание значений латентных факторов . . . . .	188
10.5.1	Метод Бартлетта . . . . .	188
10.5.2	Метод Томсона . . . . .	189
10.6	Практика выявления латентных факторов . . . . .	191
<b>11</b>	<b>Дискриминантный анализ</b>	<b>196</b>
11.1	Постановка задачи для двух классов . . . . .	197
11.2	Линейное прогностическое правило . . . . .	201
11.3	Рекомендации для нелинейного случая . . . . .	204
11.4	Дискриминация на практике . . . . .	206
11.5	Более двух классов . . . . .	209
11.6	Проверка качества дискриминации . . . . .	211
<b>12</b>	<b>Многомерное шкалирование</b>	<b>214</b>
12.1	Вводные замечания . . . . .	215
12.2	Модель Торгерсона . . . . .	217
12.2.1	Стресс-критерий . . . . .	220
12.3	Алгоритм Торгерсона . . . . .	221

12.4	Пример с предпочтениями профессий . . . . .	224
12.5	Методы шкалирования индивидуальных различий . . . . .	228
<b>13</b>	<b>Простейшие задачи временных рядов</b>	<b>231</b>
13.1	Общие положения . . . . .	231
13.2	Критерии случайности . . . . .	233
13.2.1	Подсчет экстремальных точек . . . . .	234
13.2.2	Распределение длины фазы . . . . .	237
13.2.3	Критерии, основанные на ранговой корреляции	240
13.3	Тренд и сезонность . . . . .	244
13.3.1	Полиномиальные тренды . . . . .	245
13.3.2	Выбор степени тренда . . . . .	248
13.3.3	Сглаживание . . . . .	250
13.3.4	О спектральном анализе временных рядов . . . . .	253
13.3.5	Оценка сезонных колебаний . . . . .	254
<b>14</b>	<b>Post-hoc задача в кластерном анализе</b>	<b>257</b>
14.1	Вводные замечания . . . . .	257
14.2	Применение дисперсионного анализа . . . . .	258
14.3	Учет искажений кластерной структуры . . . . .	260
14.4	Кластерная переменная . . . . .	263
14.4.1	Ранжирование кластеров . . . . .	264
14.4.2	Применение анализа соответствий . . . . .	266
14.4.3	Специализированные методы: регрессия . . . . .	267
	<b>Приложение: некоторые важные распределения</b>	<b>269</b>
	<b>Библиографический список</b>	<b>272</b>

*Научное издание*

**Дронов Сергей Вадимович**

**МЕТОДЫ И ЗАДАЧИ  
МНОГОМЕРНОЙ СТАТИСТИКИ**

*Монография*

Публикуется в авторской редакции

Издательская лицензия ЛР № 020261 от 14.01.1997.

Подписано в печать 21.09.2015. Формат  $60 \times 84 \frac{1}{16}$ .

Усл.печ.л. 15,9. Тираж 100. Заказ № 276.

Издательство Алтайского государственного университета;  
типография Алтайского государственного университета;  
656049 Барнаул, ул. Димитрова, 66.