

УДК 316:303.4.025(035.3)

ББК 60.5+87.256.64я44

Д99

*Печатается по решению Комитета по гуманитарному
и социально-экономическому направлению науки и образования
при ученом совете Южного федерального университета
(протокол № 7 от 4 июля 2022 г.)*

Рецензенты:

заведующий кафедрой социальных и гуманитарных наук
Южно-Российского государственного политехнического университета
(НПИ) им. М. И. Платова, доктор философских наук,
профессор *К. В. Воденко*;

заведующий кафедрой социологии, истории, политологии
Института управления в экономических, экологических и социальных
системах Южного федерального университета,
доктор социологических наук, профессор *А. В. Рачина*

Дятлов, А. В.

Д99

Проблемы оптимизации выборочных данных с неполнотой
в диапазоне : монография / А. В. Дятлов ; Южный федеральный
университет. – Ростов-на-Дону ; Таганрог : Издательство Южного
федерального университета, 2022. – 232 с.

ISBN 978-5-9275-4238-3

DOI 10.18522/801303706

В монографии рассматриваются основные принципы построения выборки в
социологических исследованиях и влияние конструкции выборки на точность полу-
чаемых результатов. Главное внимание уделяется проблемам при неполном диапазоне
данных в эмпирических исследованиях и методам поиска решений при неполном ди-
апазоне выборочных данных.

Адресована студентам, аспирантам, а также всем тем, кто интересуется прак-
тикой социологических эмпирических исследований выборочных совокупностей с
неполным диапазоном данных.

ISBN 978-5-9275-4238-3

УДК 316:303.4.025(035.3)

ББК 60.5+87.256.64я44

© Южный федеральный университет, 2022

© Дятлов А. В., 2022

© Оформление. Макет. Издательство

Южного федерального университета, 2022

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
Глава первая. ПРОБЛЕМЫ ПРИ НЕПОЛНОМ ДИАПАЗОНЕ ДАННЫХ В ЭМПИРИЧЕСКИХ ИССЛЕДОВАНИЯХ: ТЕОРЕТИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ	10
I. Неполный диапазон данных в эмпирических социологических исследованиях: определение, измерение, проблемы	10
1. Неполный диапазон данных.....	10
2. Проблемы с диапазоном данных: труднодоступные единицы. Потеря и (или) выпадение данных	13
3. Виды потерь данных	16
4. Влияние неполноты диапазона данных на точность оценок	30
II. Статистические аспекты пробелов в диапазоне данных.....	37
1. Случайные и системные ошибки	37
2. Источники ошибок.....	47
3. Механизм выпадающих единиц. Классификация отсутствующих данных ..	61
III. Влияние неполного диапазона данных на достоверность и точность оценок	66
1. Влияние на достоверность информация по совокупности	66
2. Возможности ограничения влияния выпадающих единиц в диапазоне данных на достоверность	75
Глава вторая. МЕТОДЫ ПОИСКА РЕШЕНИЙ ПРИ НЕПОЛНОМ ДИАПАЗОНЕ ДАННЫХ ВЫБОРКИ	85
I. Разработка методов оптимизации при отсутствующих данных.....	88
1. Прямая оценка отсутствующих значений.....	90
2. Оценка как функция сопротивления	92
3. Оценка с помощью вероятности потерь	93
4. Повторный отбор при выпадающих (отсутствующих) единицах	94
5. Выпадающие единицы как подвыборка.....	97
II. Контрольные методы снижения потерь данных	98

III. Методы сокращения доли выпадающих единиц.....	101
IV. Взвешивание данных	106
1. Необходимость взвешивания.....	107
2. Процедуры взвешивания.....	109
3. Взвешивание данных на практике.....	117
4. Модифицированные процедуры взвешивания данных.....	121
V. Методы атрибуции значений отсутствующих данных (Imputation)	129
1. Сущность метода атрибуции	130
2. Модификации метода атрибуции значений	131
VI. Многомерная атрибуция значений.....	151
1. Развитие идеи многомерной атрибуции значений.....	151
2. Сущность метода многомерной атрибуции значений	154
3. Основные гипотезы метода многомерной атрибуции значений	157
4. Правила обобщения при многомерной атрибуции.....	163
Глава третья. СРАВНИТЕЛЬНЫЕ ПРЕИМУЩЕСТВА И ОГРАНИЧЕНИЯ РАЗЛИЧНЫХ МЕТОДОВ ОПТИМИЗАЦИИ ВЫБОРОЧНЫХ ДАННЫХ	169
I. Оптимизация выборки при неполном диапазоне данных	169
1. Характеристика выборки	170
2. Сравнение характеристик выборки с параметрами генеральной совокупности.....	173
3. Влияния на выборку	174
II. Сравнительный анализ использования методов взвешивания и атрибуции значений	185
1. Базовые гипотезы сравнения	187
2. Модель исследования сравнения методов.....	189
3. Симуляционные информационные массивы. Анализ и оценки.....	193
III. Концепция решения проблемы недостающих данных в эмпирических исследованиях.....	208
1. Виды недостающих данных.....	208
2. Критерии при выборе оптимизационного подхода (метода).....	209
3. Возможные стратегии в зависимости от доли недостающих данных	213
ЗАКЛЮЧЕНИЕ.....	217
СПИСОК ЛИТЕРАТУРЫ	224