

# Содержание

Вступительное слово.....	17
Отзывы.....	19
Обращение к студентам: почему я написал эту книгу.....	26
Благодарности.....	29
А теперь о шестом издании... ..	31
Предисловие от издательства.....	33
Об авторе.....	35

## ЧАСТЬ I

Ура! У меня статистика.....	37
<b>1 Статистика или садистика?.....</b>	<b>41</b>
Почему статистика?.....	41
И почему SPSS?.....	42
Пятиминутка истории статистики.....	42
Чем (не) является статистика.....	44
Что такое описательная статистика?.....	44
Что такое инференциальная статистика?.....	45
Другими словами.....	46
Что я вообще делаю на курсе по статистике?.....	47
10 способов работы с этой книгой (и одновременного изучения статистики).....	48
О пиктограммах.....	51
Ключ к маркерам сложности.....	52
Предметный указатель.....	52
Реальная статистика.....	52
Выводы по главе.....	53
Время практики.....	53
Сайт с материалами для обучения.....	54

## ЧАСТЬ II

<b>Σигма Фрейд и описательная статистика .....</b>	<b>55</b>
<b>2 Серединный мир .....</b>	<b>57</b>
Вычисление среднего значения .....	58
Вычисление взвешенной средней.....	60
Вычисление медианы.....	61
Вычисление моды.....	64
Пирожок с двумя начинками.....	65
Как выбрать нужную меру центральной тенденции (и все, что вам нужно сейчас знать о шкалах измерения) .....	66
Хоть розой назови ее, хоть нет: номинальная шкала измерения .....	67
Любой порядок подойдет: порядковая шкала измерения .....	67
1 + 1 = 2: интервальная шкала измерения .....	67
Можно ли не иметь ничего? Пропорциональная шкала измерения .....	67
Подводя итог.....	68
Вычисление описательных статистик при помощи компьютера.....	69
Выводы SPSS .....	70
Реальная статистика.....	72
Выводы по главе .....	73
Время практики .....	73
Сайт с материалами для обучения .....	77
<b>3 Vive la Différence! .....</b>	<b>78</b>
Почему важно понимать изменчивость?.....	78
Вычисление диапазона .....	79
Вычисление стандартного отклонения.....	80
Почему $n - 1$ ? Что не так с просто $n$ ?.....	83
И что такого?.....	84
Вычисление дисперсии .....	85
Стандартное отклонение против дисперсии.....	85
Вычисление мер изменчивости при помощи компьютера.....	86
Выводы SPSS .....	87
Второй вывод SPSS .....	88
Реальная статистика.....	88
Выводы по главе .....	89
Время практики .....	89
Сайт с материалами для обучения .....	91
<b>4 Лучше один раз увидеть.....</b>	<b>92</b>
Зачем нужно иллюстрировать данные?.....	92
Десять способов сделать красивую картинку (не заваливать горизонт?) .....	93

Обо всем по порядку: как создать распределение частот .....	94
Самый классный интервал .....	94
Тени сгущаются: создание гистограммы.....	96
Метод подсчета.....	98
Следующий шаг: полигон частот.....	98
Кумулятивная частота .....	99
Другие классные способы	
графического представления данных.....	101
Столбиковые диаграммы .....	102
Линейчатые диаграммы .....	103
Линейные графики.....	103
Круговые диаграммы .....	104
Иллюстрация данных при помощи компьютера	
(т. е. SPSS) .....	104
Создание гистограммы .....	105
Создание столбиковой диаграммы .....	107
Создание линейного графика.....	108
Создание круговой диаграммы .....	109
Реальная статистика.....	110
Выводы по главе .....	111
Время практики.....	111
Сайт с материалами для обучения .....	112
<b>5 Мороженое и преступность.....</b>	<b>113</b>
О чем говорит корреляция?.....	113
Типы коэффициентов корреляции: вкус 1 и вкус 2 .....	114
Расчет простого коэффициента корреляции.....	116
Визуальное представление корреляции: диаграмма	
рассеяния .....	118
Кучи корреляций: матрица корреляции .....	122
Как понять, что значит коэффициент корреляции.....	123
Эмпирический метод (оценим «на глазок»).....	123
Детерминируй это: квадрат коэффициента корреляции .....	124
Чем больше съедается мороженого, тем... выше	
преступность (или ассоциация против причинности) .....	125
Вычисление коэффициента корреляции при помощи	
SPSS.....	127
Создание диаграммы рассеяния .....	129
Другие классные корреляции .....	130
Делим на части: немного о частичной корреляции.....	131
Вычисление частичной корреляции при помощи SPSS .....	133
Реальная статистика.....	134
Выводы по главе .....	135
Время практики.....	135
Сайт с материалами для обучения .....	139

<b>6 Только правда .....</b>	<b>140</b>
Предисловие к надежности и достоверности.....	140
Так что там с этими измерениями? .....	141
Надежность: повторяйте, пока не получится правильно .....	142
Оценки за тест: правда или действие? .....	142
Наблюдаемая оценка = истинная оценка + оценка ошибки .....	143
Виды надежности .....	144
Ретестовая надежность .....	144
Надежность параллельных форм .....	145
Надежность внутренней согласованности.....	146
Взаимная надежность оценивающих.....	150
Много – это сколько? Наконец-то: интерпретация коэффициентов надежности .....	151
А если не получается установить надежность, что тогда? .....	152
И еще одно .....	152
Достоверность: в чем правда, брат?.....	153
Виды достоверности.....	153
Содержательная достоверность.....	153
Критериальная достоверность .....	154
Конструктивная достоверность .....	155
А если не получается установить достоверность – что тогда?.....	156
Дружественное напутствие.....	157
Достоверность и надежность – очень близкие родственники.....	158
Реальная статистика.....	158
Выводы по главе .....	159
Время практики .....	159
Сайт с материалами для обучения .....	160

### ЧАСТЬ III

<b>Карты, деньги, вероятности.....</b>	<b>161</b>
<b>7 Гипотезы и вы .....</b>	<b>163</b>
Итак, вы хотите быть ученым.....	163
Выборка и генеральная совокупность.....	164
Нулевая гипотеза .....	165
Назначение нулевой гипотезы .....	165
Альтернативная гипотеза .....	167
Ненаправленная альтернативная гипотеза.....	168
Направленная альтернативная гипотеза.....	168
Различия между нулевой и альтернативной гипотезами .....	170
Критерии качества хорошей гипотезы .....	171
Реальная статистика.....	174
Выводы по главе .....	175

Время практики .....	175
Сайт с материалами для обучения .....	176
<b>8 Нормальны ли ваши кривые? .....</b>	<b>177</b>
Почему вероятность? .....	177
Кривая нормального распределения (также известная как «колокольчик») .....	178
Эй, это ненормально! .....	179
Еще немного о нормальной кривой .....	181
Наш любимый стандартный показатель: z-оценка .....	184
Что показывает z-оценка? .....	187
Что на самом деле показывает z-оценка? .....	190
Проверка гипотез и z-оценки: первый шаг .....	192
Вычисление z-оценок при помощи SPSS .....	193
Толстые и тощие распределения частот .....	193
Среднее значение .....	194
Изменчивость .....	194
Асимметрия .....	194
Экссесс .....	196
Реальная статистика .....	198
Выводы по главе .....	198
Время практики .....	199
Сайт с материалами для обучения .....	201

## ЧАСТЬ IV

<b>Значимая разница: применение статистики вывода .....</b>	<b>203</b>
<b>9 Значимо значимый .....</b>	<b>205</b>
Концепция значимости .....	205
Если бы мы были идеальны .....	206
Самая важная таблица в мире (только для этого семестра) ...	209
Подробнее про таблицу 9.1 .....	209
Возвращаясь к ошибкам первого рода .....	210
Значимость против осмысленности .....	212
Предисловие к статистике вывода .....	214
Как работают выводы? .....	215
Как выбрать статистический критерий для проверки? .....	215
Как пользоваться схемой? .....	216
Предисловие к проверке значимости .....	216
Как работает проверка значимости: план .....	217
Лучше один раз увидеть .....	219
Повысьте свое доверие .....	220
Реальная статистика .....	221
Выводы по главе .....	222
Время практики .....	222
Сайт с материалами для обучения .....	223

<b>10 Строго по одному</b> .....	<b>224</b>
Предисловие к одновыборочному $Z$ -критерию .....	224
Путь к мудрости и знанию .....	225
Вычисление $Z$ -статистики .....	225
Так как мне интерпретировать $z = 2,38, p < 0,05$ ? .....	229
Расчет $Z$ -критерия при помощи SPSS .....	229
Спецэффекты: а эти различия настоящие? .....	231
Как понимать размер эффекта? .....	232
Реальная статистика .....	233
Выводы по главе .....	233
Время практики .....	234
Сайт с материалами для обучения .....	235
<b>11 (t)ет-а-тет</b> .....	<b>236</b>
Предисловие к $t$ -критерию для независимых выборок .....	236
Путь к мудрости и знанию .....	237
Вычисление $t$ -критерия .....	238
Время для примера .....	239
Так как мне интерпретировать $t_{(58)} = -0,14, p > 0,05$ ? .....	242
Размер эффекта и (t)ет-а-тет .....	242
Расчет и понимание размера эффекта .....	243
Два очень классных калькулятора размера эффекта .....	244
Расчет $t$ -критерия при помощи SPSS .....	245
Реальная статистика .....	247
Выводы по главе .....	247
Время практики .....	248
Сайт с материалами для обучения .....	250
<b>12 (t)ет-а-тет (снова)</b> .....	<b>251</b>
Предисловие к $t$ -критерию для зависимых средних .....	251
Путь к мудрости и знаниям .....	252
Вычисление значения $t$ -критерия .....	253
Так как мне интерпретировать $t_{(24)} = 2,45, p < 0,05$ ? .....	256
Расчет $t$ -критерия при помощи SPSS .....	256
Размер эффекта для (t)ет-а-тета (снова) .....	260
Реальная статистика .....	260
Выводы по главе .....	261
Время практики .....	261
Сайт с материалами для обучения .....	264
<b>13 Двух групп недостаточно?</b> .....	<b>265</b>
Предисловие к дисперсионному анализу .....	265
Путь к мудрости и знанию .....	266
Разновидности дисперсионного анализа .....	266
Вычисление статистического критерия Фишера .....	268
Так как мне интерпретировать $F_{(2,27)} = 8,80, p < 0,05$ ? .....	273

Вычисление $F$ -отношения Фишера с помощью SPSS .....	274
Размер эффекта для однонаправленного дисперсионного анализа .....	277
Реальная статистика .....	278
Выводы по главе .....	279
Время практики .....	279
<b>14 Два лишних фактора .....</b>	<b>281</b>
Предисловие к многофакторному дисперсионному анализу .....	281
Путь к мудрости и знанию .....	282
Новый сорт ANOVA .....	283
Основное событие: главные эффекты в многофакторном дисперсионном анализе .....	284
Еще интереснее: эффекты взаимодействия .....	286
Вычисление $F$ -отношения с помощью SPSS .....	288
Расчет размера эффекта для многофакторного ANOVA .....	292
Реальная статистика .....	292
Выводы по главе .....	293
Время практики .....	293
<b>15 Двоюродные или просто хорошие друзья? .....</b>	<b>295</b>
Предисловие к проверке коэффициента корреляции .....	295
Путь к мудрости и знанию .....	296
Расчет статистического критерия .....	296
Так как мне интерпретировать $r_{(28)} = 0,437, p < 0,05$ ? .....	300
Причины и связи (снова!) .....	300
Значимость против существенности (опять и снова!) .....	301
Вычисление коэффициента корреляции при помощи SPSS (снова) .....	301
Реальная статистика .....	303
Выводы по главе .....	304
Время практики .....	304
<b>16 Как предсказать, кто выиграет Суперкубок.....</b>	<b>307</b>
Предисловие к линейной регрессии .....	307
Вообще, что такое предсказание? .....	308
Логика предсказания.....	309
Рисуем лучшую прямую в мире (по вашим данным) .....	313
Насколько хороши ваши предсказания?.....	315
Расчет прямой регрессии при помощи SPSS .....	316
Чем больше предсказываемых переменных, тем лучше? Возможно .....	320
Важные правила применения множественных независимых переменных.....	322
Реальная статистика.....	323

Выводы по главе .....	324
Время практики .....	324

## ЧАСТЬ V

### Больше статистики! Больше инструментов!

<b>Больше веселья!</b> .....	<b>329</b>
------------------------------	------------

### **17 Что делать с ненормальными .....** **331**

Предисловие к непараметрическим критериям .....	331
Представляем критерий согласия хи-квадрат (одновыборочный) .....	332
Вычисление значения критерия согласия хи-квадрат .....	333
Так как мне интерпретировать $\chi^2 = 20,6, p < 0,05$ ? .....	336
Предисловие к критерию независимости хи-квадрат.....	336
Расчет значения критерия независимости хи-квадрат.....	337
Расчет критерия хи-квадрат при помощи SPSS .....	339
Критерий согласия и SPSS.....	339
Критерий независимости и SPSS.....	340
Другие непараметрические критерии, о которых вам нужно знать.....	343
Реальная статистика.....	344
Выводы по главе .....	345
Время практики .....	345

### **18 Еще несколько (важных) статистических процедур,**

<b>о которых вам стоит знать.....</b>	<b>348</b>
---------------------------------------	------------

Многомерный дисперсионный анализ.....	348
Дисперсионный анализ повторных измерений .....	349
Ковариационный анализ .....	350
Множественная регрессия .....	351
Метаанализ .....	351
Дискриминантный анализ.....	352
Факторный анализ.....	353
Анализ пути .....	353
Моделирование структурных уравнений .....	354
Выводы по главе .....	355

### **19 Интеллектуальный анализ данных .....** **356**

Наш подопытный набор данных – кто не любит младенцев? .....	359
Подсчет результатов.....	360
Подсчет по частотам.....	360
Сводные таблицы и таблицы сопряженности: находим скрытые паттерны.....	364
Создание сводной таблицы .....	364
Изменение сводной таблицы .....	366



Выводы по главе .....	369
Время практики .....	369
<b>20 Подборка программ для статистического анализа.....</b>	<b>371</b>
Как выбрать идеальную программу для статистики?.....	371
Что же у нас есть? .....	374
Сначала о бесплатном.....	374
Час расплаты.....	377
Выводы по главе .....	380

## ЧАСТЬ VI

<b>Десять вещей (умножить на два), которые вы хотите узнать и запомнить.....</b>	<b>381</b>
--	------------

<b>21 Десять (или более) лучших (и самых увлекательных) интернет-сайтов для статистики .....</b>	<b>382</b>
--	------------

Не думали поучить статистику в Стокгольме? .....	383
Кто есть кто и что случилось? .....	383
Все здесь .....	383
Hyperstat.....	384
Данные? Вам нужны данные?.....	385
Больше и больше ресурсов.....	386
И конечно, YouTube... ..	386
Наконец.....	386

<b>22 Десять заповедей сбора данных.....</b>	<b>387</b>
--	------------

<b>Приложение А. SPSS Statistics меньше, чем за 30 мин .....</b>	<b>390</b>
--	------------

Запуск SPSS .....	391
Приветственное окно SPSS .....	391
Панель инструментов и строка состояния SPSS.....	393
Использование справки SPSS .....	393
Краткий тур по SPSS .....	395
Открываем файл.....	395
Простая таблица и диаграмма .....	396
Простой анализ.....	397
Создание и редактирование файла данных .....	397
Определение переменных .....	397
Настраиваемое определение переменных:	
использование представления переменных .....	398
Определение значения меток данных.....	399
Изменение меток значений .....	401
Как открыть файл данных.....	402
Печать из SPSS .....	402
Печать файла данных SPSS .....	402
Печать выделения в файле данных SPSS .....	403

Создание диаграммы в SPSS.....	403
Создание простой диаграммы.....	403
Создание столбиковой диаграммы .....	403
Сохранение диаграммы .....	405
Улучшение диаграмм SPSS .....	405
Редактирование диаграммы.....	406
Работа с заголовками и подзаголовками.....	406
Работа со шрифтами.....	407
Работа с осями .....	407
Как изменить ось шкалы (y).....	408
Как изменить ось категорий (x).....	408
Описание данных .....	409
Частоты и сопряженные таблицы.....	409
Применение t-критерия для независимых выборок .....	411
Как рассчитать t-критерий для независимых выборок.....	412
Выход из SPSS .....	413
<b>Приложение В. Таблицы.....</b>	<b>414</b>
Таблица В.1. Площади под кривой нормального распределения .....	414
Таблица В.2. Значения T, необходимые для отклонения нулевой гипотезы .....	416
Таблица В.3. Критические значения для дисперсионного анализа, или F-критерия.....	417
Таблица В.4. Значения коэффициента корреляции, необходимые для отклонения нулевой гипотезы.....	421
Таблица В.5. Критические значения хи-квадрат.....	422
<b>Приложение С. Наборы данных.....</b>	<b>424</b>
Глава 2. Набор данных 1 .....	424
Глава 2. Набор данных 2 .....	425
Глава 2. Набор данных 3 .....	425
Глава 2. Набор данных 4 .....	425
Глава 3. Набор данных 1 .....	425
Глава 3. Набор данных 2 .....	426
Глава 3. Набор данных 3 .....	426
Глава 3. Набор данных 4 .....	426
Глава 4. Набор данных 1 .....	427
Глава 4. Набор данных 2 .....	427
Глава 4. Набор данных 3 .....	427
Глава 5. Набор данных 1 .....	428
Глава 5. Набор данных 2 .....	428
Глава 5. Набор данных 3 .....	428
Глава 5. Набор данных 4 .....	428
Глава 5. Набор данных 5 .....	429
Глава 5. Набор данных 6 .....	429

Глава 6. Набор данных 1 .....	429
Глава 6. Набор данных 2 .....	430
Глава 11. Набор данных 1 .....	431
Глава 11. Набор данных 2 .....	431
Глава 11. Набор данных 3 .....	432
Глава 11. Набор данных 4 .....	432
Глава 11. Набор данных 5 .....	433
Глава 11. Набор данных 6 .....	433
Глава 12. Набор данных 1 .....	434
Глава 12. Набор данных 2 .....	434
Глава 12. Набор данных 3 .....	434
Глава 12. Набор данных 4 .....	435
Глава 12. Набор данных 5 .....	435
Глава 13. Набор данных 1 .....	436
Глава 13. Набор данных 2 .....	436
Глава 13. Набор данных 3 .....	437
Глава 13. Набор данных 4 .....	437
Глава 14. Набор данных 1 .....	438
Глава 14. Набор данных 2 .....	438
Глава 14. Набор данных 3 .....	439
Глава 14. Набор данных 4 .....	440
Глава 15. Набор данных 1 .....	441
Глава 15. Набор данных 2 .....	441
Глава 15. Набор данных 3 .....	442
Глава 15. Набор данных 4 .....	442
Глава 15. Набор данных 5 .....	442
Глава 16. Набор данных 1 .....	443
Глава 16. Набор данных 2 .....	443
Глава 16. Набор данных 3 .....	443
Глава 17. Набор данных 1.....	444
Глава 17. Набор данных 2.....	444
Глава 17. Набор данных 3.....	445
Глава 17. Набор данных 4.....	446
Глава 17. Набор данных 5.....	447
Глава 19. Набор данных 1 и Глава 19. Набор данных 2 .....	448
Набор данных. Пример .....	448

**Приложение D. Ответы на задания .....** 449

**Приложение E. Математика: самые основы .....** 488

    Большие правила: приветствуем СВД У ПМ .....

    Маленькие правила .....

**Приложение F. Бонус: рецепт брауни .....** 492

**Глоссарий .....** 494

**Предметный указатель .....** 500

# Вступительное слово

Сегодня, в эпоху цифровых технологий и огромного потока различных данных, возрождается интерес широкого круга специалистов к статистике. Статистический анализ позволяет получать из собираемых повсеместно данных новые знания о жизни людей, поведении потребителей, психологии принятия решений. Методы прикладной статистики активно применяются в технических исследованиях, экономике, теории и практике управления, социологии, медицине, геологии, истории.

В то же время недостаточная образовательная подготовка, обилие различного рода манипуляций с данными, непонимание функций статистики в последние десятилетия во многом дискредитировали ее в глазах общества.

В нашей стране недоверие к статистике тесно связано с непониманием ее основ: неумением «читать» данные и делать объективные выводы на их основе, неумением выявлять лжестатистику и отделять правду от спекуляций. Отдельная проблема – непрофессионализм журналистов в работе со статистическими данными, порой неверно интерпретируемыми ими в поисках ярких фактов и заголовков.

Сегодня статистику изучают во многих учебных заведениях, она включена в программы подготовки различных специалистов, в том числе гуманитарного профиля. В то же время приходится констатировать, что подготовку сильных и востребованных на рынке социологов, маркетологов и аналитиков осуществляет небольшое число российских вузов. И дело не только в педагогических кадрах, но и в учебных материалах. Сегодня преподавание статистики осложняет серьезный дефицит качественных учебных пособий, книг, электронных образовательных ресурсов.

Российское исследовательское сообщество, представителем которого является Аналитический центр НАФИ, приветствует издание качественной и доступной для понимания широкого круга читате-

лей книги Нила Дж. Салкинда «Статистика для тех, кто (думает, что) ненавидит статистику».

За многие годы совместной работы с крупнейшими вузами мы хорошо изучили методические потребности студентов и можем смело сказать, что представляемая вашему вниманию книга отвечает им в полной мере.

В 2022 г. будет отмечаться юбилей – 220 лет официальной российской статистике. Полагаем, что данная книга успеет к этому времени завоевать популярность у читателей и внести вклад в подготовку российских специалистов, имеющих дело со статистическим анализом.

Данная книга относится к особому, не похожему ни на один из принятых в современной учебной литературе жанру. Это организованный курс с увлекательными примерами и доступными объяснениями. Книгу действительно легко и интересно читать.

Желаем вам приятного погружения в интереснейший мир анализа данных!

*Тимур Аймалетдинов,*  
кандидат социологических наук,  
заместитель генерального директора  
аналитического центра НАФИ, nafl.ru

## ОТЗЫВЫ

Книга «Статистика для тех, кто *(думает, что)* ненавидит статистику» действительно заставляет студентов изучать эту область науки и наслаждаться ею и исследованиями в целом. Студентам особенно нравятся 10 заповедей и ссылки на интернет-сайты.

– *Валери Джейнсик (Valarie Janesick),  
профессор управления обучением,  
Университет Южной Флориды*

Я просто хотела сказать, что, проходя онлайн-обучение сестринскому делу в Университете штата Нью-Йорк, столкнулась с «Введением в статистику», и ваша книга спасла мою карьеру! Я бросила обучение из-за статистики, пару раз даже восстанавливалась и опять бросала. Я прочитала первые две главы книги и уже поняла суть. Я знаю, что дальше будет сложнее, но я так благодарна вам за простую для понимания методику. Прошлым вечером я сказала мужу, что вполне могу полюбить статистику. Отдельное спасибо вам за обзор основ математики. Никто еще не объяснял ее так понятно, как вы: в старшей школе я была в коррекционном классе по математике и все равно не понимала ее. Теперь я больше не боюсь ни математики, ни статистики.

– *Меган Уилер (Meghan Wheeler),  
дипломированная медсестра*

Очень признательна вам за помощь и отзывчивость. Я учусь использовать SPSS, пока готовлюсь к докторской программе, которая начнется осенью. Прошло уже двадцать лет с тех пор, как я изучала статистику. Спасибо этой доходчивой книге за то, что помогла мне освоиться с текущими трендами.

– *Сильвия Миллер-Мартин  
(Sylvia Miller-Martin)*

Я полюбил статистику со времени получения второго высшего образования. Ваша книга «Статистика для тех, кто (думает, что) ненавидит статистику» прояснила запутанные и не до конца понятные вопросы, которые волновали меня много лет. Это обязательная для прочтения книга для всех начинающих или продолжающих свой путь в науке. Я обожаю ее и обязательно буду использовать в будущем.

– *Рональд А. Штраубе (Ronald A. Straube),  
Региональный медицинский центр, г. Мишен, Техас*

Д-р Салкинд, я чувствую себя просто обязанным поблагодарить Вас за такую отличную книгу. Два года назад я купил дом. Люди, жившие здесь раньше, оставили эту книгу. Я не выбросил ее, потому что я книжный маньяк. В общем, я начал учиться в магистратуре по курсу психологии и решил использовать Вашу книгу. Она оказалась для меня просто подарком. Это, определенно, лучшая книга по статистике, с которой я когда-либо имел дело, особенно учитывая простоту и доступность содержащихся в ней объяснений. Она точно стоила потраченных на дом ста тысяч! Благослови Вас бог!

– *Брайан Райт (Brian Wright)*

Проектная команда, состоявшая из Дениз, Рене, Шона и Триш, выбрала в качестве рабочей гипотезы утверждение, что брауни из обычной муки будет пользоваться большей популярностью, чем брауни без глютена. Они выбрали рецепт, приведенный в приложении F «Награда». Дениз испекла брауни без глютена, Рене сделала обычные, а выбор осуществляли наши сокурсники в колледже. Для оценки брауни мы использовали опрос с порядковой шкалой от 1 до 5, где 1 означало худший, а 5 – лучший брауни в вашей жизни. Брауни без глютена выиграли, тем самым опровергнув гипотезу исследования. Сравнение проводилось по среднему значению и по моде. У брауни без глютена среднее значение/мода составляло 4, а у обычных – 3. Широта распределения для брауни без глютена была больше, чем для обычных. Все участники опроса были без ума от брауни.

Все это произошло, потому что я спросила преподавателя, будем ли мы как-то использовать информацию из приложения F. Ни преподаватель, ни сокурсники даже и не заглядывали в него. Преподаватель сказал, что я могу испечь брауни и принести их на занятия. Именно тогда я сказала, что у меня целиакия и в доме есть только мука без глютена. Из-за особой текстуры выпечка без глютена обычно нравится меньше. Преподаватель всегда хотел попробовать что-нибудь безглютеновое. Вот так и появился на свет наш учебный эксперимент.

– *Дениз Проске (Denise Proske),  
колледж Тускула*

Я просто хотел на минуту привлечь Ваше внимание, чтобы сообщить, что выбрал Вашу книгу, «Статистика для тех, кто (думает, что) ненавидит статистику», в качестве учебника по своему предмету. Я полностью согласен с курсом, которого Вы придерживаетесь в книге, и знаю, что наши студенты точно так же одобрят его.

– Карл Р. Кравиц (*Karl R. Krawitz*),  
д-р педагогических наук,  
Университет Бейкера, Оверленд Парк, Канзас

Я «нетрадиционная» (так милые ребята из Университета Дейтона называют пожилых) студентка магистратуры и получаю большое удовольствие от Вашей «Статистики для тех, кто (думает, что) ненавидит статистику». Несмотря на то что моя работа заключается в том, чтобы рассказывать об исследованиях, участие в них и статистических расчетах является для меня совершенно новой задачей. Так что считайте меня одной из множества тех, кто ценит Ваш подход к статистике и чувство юмора (оно, определенно, помогает уменьшить страх перед этим предметом).

Еще раз спасибо за такое «человеческое» отношение к этой теме.

С наилучшими пожеланиями,

Памела Грегг (*Pamela Gregg*),  
администратор по коммуникациям,  
Исследовательский институт Университета Дейтона

Я подумал, что нужно отправить Вам небольшой позитивный отзыв! Друг заставил меня купить Вашу книгу, когда я учился психологии в колледже. Предназначалась она не для курса по статистике, как Вы могли бы подумать. Он был у нас 2 года назад, и я не вынес из него НИЧЕГО! А приступив к дипломной работе, я стал ощущать легкое (а честно говоря, огромное) беспокойство по поводу анализа данных, который мне нужно было сделать. Именно тогда приятель и посоветовал купить Вашу книгу. Моя первая реакция? «Я не собираюсь покупать еще один учебник по статистике просто из любви к искусству!» Что ж, после множества напоминаний я все-таки купил эту книгу (в то время это было второе издание). Сейчас я снова изучаю статистику, но в этот раз, на третьем году аспирантуры, держу Вашу книгу при себе в качестве успокоительного средства!

Спасибо за то, что все эти годы делали статистику сносной!

– Эшли Шир (*Ashley Shier*),  
магистр педагогики,  
Университет Цинциннати,  
аспирант школы психологии

Привет, д-р Салкинд! Просто хочу поблагодарить Вас за все вложенные в книгу усилия. Я использую ее в своем курсе «Количественные методы исследований в антропологии» в Университете



Северной Аризоны. Мы по-доброму называем вашу книгу «ненавистой».

Еще раз спасибо!

– *Бриттон Л. Шепардсон, PhD, преподаватель,  
ассистент кафедры антропологии,  
Университет Северной Аризоны*

Примеры Салкинда помогают понять, как правильно использовать ключевые концепции, и успешно сдать тесты. Книга легко читается благодаря разнообразной подаче информации, например блокам с техническими деталями и заметками на память, индексу сложности, различным спискам из «10 вещей», иконкам, рисункам и комиксам. Даже ее название вызывает у студентов смех, а юмор может быть отличным лекарством от стресса!

– *Мэри Бет Зени (Mary Beth Zeni),  
Школа сестринского дела, штат Флорида*

Привет, м-р Салкинд.

Мне 19 лет, я работаю медсестрой на полную ставку и недавно приступила к обучению в бакалавриате по специальности «сестринское дело». Завтра у нас первое занятие по статистике. Я только что прочитала ваше обращение к студентам и хочу сказать, что вы описали все до единого мои симптомы. Мы с сокурсниками страшно переживаем по поводу изучения этого предмета и того, что нам предстоит в следующие три месяца. Прочитав эти две страницы, мне захотелось сказать, что вы уменьшили мои опасения. Я могу наконец перестать трястись перед неизвестностью и начать просто читать. Спасибо Вам за это! Сегодня я работаю в ночную смену. Надеюсь, у меня найдется время, чтобы прочитать заданные главы с меньшим волнением и на самом деле усвоить часть изучаемого материала. Спасибо еще раз. Я буду стараться и уже предвкушаю новые знания, которые получу из Вашей книги, от нашего отличного преподавателя и моих сокурсников.

С уважением,

*Лори Вайда (Lori Vajda), медсестра*

Огромное количество статистической информации было поглощено.

– *Бельдар с планеты Ремулак*

Уважаемый проф. Салкинд,

хочу поблагодарить Вас за удивительную книгу «Статистика для тех, кто (думает, что) ненавидит статистику». Я точно был одним из тех, кто ненавидит статистику, и до недавних пор игнорировал ее. Однако поскольку я почти заканчиваю аспирантуру, то подумал, что будет стыдно, если у меня не будет хотя бы минимальных знаний в статистике. Книга не только помогла понять предмет, но и вдох-

новила меня на дальнейшее чтение. Я даже успел за два дня порекомендовать Вашу книгу нескольким людям! Большое Вам спасибо за такую замечательную работу!

– Э. Дж. Пэдман (*A. J. Padman*)

Хочу сказать Вам «спасибо» за то, что написали чрезвычайно дружелюбную по отношению к читателю книгу «Статистика для тех, кто (думает, что) ненавидит статистику». Я учусь на психфаке и на каникулах занимаюсь самостоятельно (у нас в Алверно, для того чтобы начать изучать экспериментальную психологию – предмет, который я хочу взять этой весной, нужно сдать статистику). Другими словами, я довольно много изучаю ее сама (с небольшой помощью от ментора), поэтому очень рада, что у меня есть книга, где весь материал изложен в простой, а порой даже юмористической манере. Могу лишь предложить написать еще один учебник по статистике более высокого уровня, чтобы я могла прочесть и его тоже!

С искренним уважением,

*Дженни Сосерман (Jenny Saucerman)*

Мне понравился юмористический подход, который точно помогает снизить статистическую тревожность. Еще один плюс книги – это привлекательное и успокаивающее оформление. Стиль автора великолепен, а подача информации подходит моим студентам. Веселая и хорошо написанная книга, которую легко читать и использовать и которая доходчиво подает информацию о статистике. Несомненно, рекомендую ее.

– Миньцзюань Ван (*Minjuan Wang*),  
*Государственный университет Сан-Диего*

Позвольте поблагодарить Вас за удивительный учебник. Среди всех пособий, которые я использовала в эти годы, Вашему я бы присудила первое место за понятную и легко усваиваемую подачу материала.

– Кэролин Летше (*Carolyn Letsche*),  
*студентка магистратуры по социальной педагогике*

Книга Салкинды – единственная в своем роде. Ее с легкостью можно назвать лучшей книгой такого типа среди тех, которые мне встречались. С большим энтузиазмом рекомендую ее всем, кто интересуется предметом, и даже (в особенности) тем, кому он не интересен!

– Расс Шафер-Ландау (*Russ Shafer-Landau*),  
*Университет Висконсина*

«Статистика для тех, кто (думает, что) ненавидит статистику» – это, определенно, нужная книга для тех, кому приходится преодолеть

вать знакомое чувство тревоги, открывая обычную книгу по статистике, и для тех, кто, преодолев его, все равно не может увидеть в ней никакого смысла. Книгу Салкинда легко и приятно читать, и вряд ли потребуется какое-то предварительное знание предмета, чтобы следовать за ходом мыслей автора. Салкинду удалось разъяснить статистику людям, которые ее ненавидят или думали так раньше.

Из обзора в журнале «Статистические методы в медицинских исследованиях» (Statistical Methods in Medical Research).

– Д-р Андреа Винклер (*Dr. Andrea Winkler*),  
Бетлемская королевская больница, Лондон, Великобритания



Археологи раскапывают первое издание  
«Статистики для тех, кто (думает, что)  
ненавидит статистику»

# Обращение к студентам: почему я написал эту книгу

С помощью этого нового издания (теперь уже шестого по счету) я надеюсь обогатить ваш опыт обучения. Меня переполняет восторг от возможности продолжать дорабатывать эту книгу, и я надеюсь, что она принесет вам такое же удовольствие, какое приносит мне.

Что объединяет множество людей, изучающих вводный курс в статистику (будь то новички в этой области или те, кто просто освежает в памяти материал)? Это, по крайней мере в начале обучения, довольно высокий уровень беспокойства, которое чаще всего основывается на услышанном от старших товарищей. Некоторая часть того, что они слышали, правдива. Ведь изучение статистики требует затрат времени и усилий (и иногда преподаватель оказывается сущим монстром).

Но большая часть того, что они слышали (и что вызывает тревогу), не соответствует действительности. Так, нельзя согласиться с тем, что статистика невыносимо сложна и запутана. Тысячи напуганных студентов достигли успеха в том, в чем, по их мнению, должны были потерпеть поражение. Им это удалось, потому что они не спешили, последовательно изучали предмет, видели, как его основные принципы применяются в реальной жизни, и даже получали некоторое удовольствие в процессе. Это то, чего я хотел добиться первыми пятью изданиями «Статистики, для тех, кто...» и к чему я стремился еще сильнее, работая над текущим изданием.

В результате большого количества проб и ошибок, нескольких удачных и множества неудачных попыток и многочисленных отзывов от студентов и преподавателей всех уровней образования я решил попробовать учить статистике так, чтобы она перестала, по моему (и множества моих студентов) мнению, пугать и стала полезной. Я сделал все, от меня зависящее, чтобы вложить весь этот опыт в данную книгу.

Из нее вы узнаете все необходимое, чтобы понять, чем занимаетесь и что изучает базовая статистика. Вы познакомитесь с основ-

ными понятиями и наиболее часто применяемыми техниками организации и осмысления данных. В ней совсем немного теории (но кое-какая имеется) и доказательств или выводов, обосновывающих математические процедуры.

Почему же в «Статистике для тех, кто...» нет теории и прочего? Очень просто. Прямо сейчас она вам не нужна. Дело даже не в том, что она неважна. Скорее, в данный момент обучения хочу предложить вам материал на том уровне, который, как я думаю, будет вам понятен и который вы сможете освоить, приложив разумные усилия. При этом он не отвратит вас от выбора дополнительных курсов по статистике в будущем. Мы с вашим преподавателем хотим, чтобы у вас все получилось.

Так что, если вы ищете подробный разбор того, как выводится анализ отклонений по критерию Фишера, лучше найдите другую хорошую книгу от издательства SAGE (я с радостью посоветую вам, какую именно). Но если вы хотите узнать, почему и как статистика может быть полезной для вас, то вы попали, куда нужно. Эта книга поможет вам понять материал в научных статьях, объяснит, что означают результаты статистического анализа, и научит вас выполнять основные статистические операции.

Если вы хотите поговорить о любом аспекте изучения статистики, не стесняйтесь обращаться ко мне. Это можно сделать по электронной почте: [njs@ku.edu](mailto:njs@ku.edu). Желаю удачи! И дайте мне знать, как еще можно улучшить эту книгу, чтобы она отвечала потребностям начинающего студента. А если вам нужны файлы с данными, которые помогут в изучении статистики, или сходите на сайт издательства SAGE [edge.sagepub.com/salkind6e](http://edge.sagepub.com/salkind6e), или свяжитесь со мной по электронной почте и скажите, какое издание вы используете.

## И (НЕБОЛЬШАЯ) ЗАМЕТКА ДЛЯ ПРЕПОДАВАТЕЛЯ

Мне хотелось бы сказать вам две вещи.

Во-первых, я аплодирую вашим усилиям по обучению основам статистики. Несмотря на то что для некоторых студентов предмет может быть очень простым, у большинства он вызывает существенные трудности. Ваше терпение и упорный труд заслуживают уважения. Если я могу что-то сделать, чтобы помочь вам, пожалуйста, дайте мне знать об этом.

Во-вторых, «Статистика для тех, кто...» задумана не как упрощенная донельзя книга, каких вы, наверное, уже много видели. В ее названии нет ничего, кроме отображения того факта, что многие начинающие изучать предмет студенты очень беспокоятся, как все получится. Это не академическая или учебная версия пособий для начинающих. Я приложил все усилия, чтобы обращаться к студентам с уважением, которого они заслуживают, не покровительствовать им и сделать материал доступным для понимания. Насколько

я в этом преуспел, судить вам, но хочу еще раз подчеркнуть, что в этой книге содержится вся информация, необходимая для вводного курса. И даже если мой подход содержит в себе толику юмора, то в моих намерениях нет ничего несерьезного. Спасибо!

## ЧАСТЬ II

# Σигма Фрейд и описательная статистика



Да, все проясняется



Одна из вещей, которые основатель психоанализа Зигмунд Фрейд довольно хорошо умел делать, – это наблюдение за природой состояния своих пациентов и описание ее. Будучи проницательным наблюдателем, он разработал первую систематизированную и хорошо обоснованную теорию личности. Несмотря на то что вы можете сомневаться в правильности его идей, он был хорошим ученым.

В начале XX в. курсы по статистике (подобные тому, какой вы сейчас изучаете) не входили в учебные программы высшего образования. Эта область была относительно новой, а характер научных исследований не требовал той точности, которую набор статистических инструментов вывел на научную арену.

Но все изменилось. Теперь цифры имеют значение почти в любой сфере деятельности (о чем также говорил и Фрэнсис Гальтон, изобретатель корреляции и троюродный брат Чарльза Дарвина). Этот раздел «Статистики для тех, кто *(думает, что)* ненавидит статистику» посвящен способам применения статистики для описания и лучшего понимания результатов, после того как информация о них будет сгруппирована.

Глава 2 рассказывает о метриках средних значений и о том, как вычисление нескольких различных типов средних дает вам отличную точку данных, описывающую набор величин. Также благодаря этой главе можно узнать, когда какую среднюю применять. Глава 3 завершает рассказ об инструментах для описания наборов данных обсуждением изменчивости, включая стандартное отклонение и дисперсию. Когда вы перейдете к главе 4, то уже будете готовы узнать о том, чем распределения (или наборы данных) отличаются друг от друга и что эта разница означает. Глава 5 разбирается с природой отношений между переменными, а именно с корреляциями. И наконец, глава 6 расскажет вам о том, почему при описании некоторых характеристик эффективных методов измерения важны параметры надежности и обоснованности.

Когда вы прочитаете раздел II, то будете отлично подготовлены к пониманию того, какую роль вероятность и предположения играют в социальных, поведенческих и других науках.

# 2

## Серединный мир

### Вычисление и значение средних

Уровень сложности: 😊 😊 😊 😊 (довольно легко)

#### В ЭТОЙ ГЛАВЕ ВЫ УЗНАЕТЕ:

- ◆ что такое меры центральной тенденции;
- ◆ как вычислить среднюю для набора данных;
- ◆ как вычислить медиану для набора данных;
- ◆ как вычислить моду для набора данных;
- ◆ что такое шкалы измерений и как они применяются;
- ◆ как выбрать меру центральной тенденции.

**В**ы терпеливо ждали, и вот теперь пришло время начать наконец работу с настоящими живыми данными. Именно этим вы будете заниматься в данной главе. Когда данные собраны, обычно первым шагом становится их организация при помощи простых показателей, описывающих эти данные. Самый простой способ сделать это – посчитать среднее значение, которое может быть нескольких видов.

**Среднее** – это одна величина, которая лучше всего отражает целую группу величин. И не важно, описывает ли эта группа величин количество верных ответов на контрольной по правописанию у 30 пятиклашек, или процент подач у каждого игрока «Нью-Йорк Янки» (который, кстати, был не очень высок в сезоне 2015 г.), или количество людей, которые зарегистрировались как демократы или республиканцы для грядущих выборов. Во всех этих примерах можно обобщить группы данных при помощи среднего. Вы можете рассматривать среднюю как центральную точку или как ось вращения циркулярной пилы. Это точка, в которой все значения в наборе величин приходят в равновесие.

Средние, которые также называются **мерами центральной тенденции**, представлены тремя видами: среднее значение, медиана и мода.



Введение  
в главу 2

## ВЫЧИСЛЕНИЕ СРЕДНЕГО ЗНАЧЕНИЯ



Вычисление  
среднего  
значения

**Среднее значение** – самый распространенный тип вычисляемого среднего. Это просто сумма всех величин в группе, разделенная на количество этих величин. Да, если у вас есть оценки за контрольную у 30 пятиклашек, вы просто сложите все показатели, а затем разделите сумму на количество учеников, т. е. на 30.

Формула для вычисления среднего значения представлена ниже (см. формулу 2.1).

$$\bar{X} = \frac{\sum x}{n}, \quad (2.1)$$

где буква  $X$  с черточкой наверху (произносится как «икс с чертой») – это среднее значение группы величин, или просто среднее;  $\sum$ , или греческая буква «сигма» – это знак суммы, который указывает на необходимость сложить все, что следует за ним;  $x$  – это каждая отдельная величина в группе величин;  $n$  – это размер выборки, для которой вы вычисляете среднее.



Для вычисления среднего значения сделайте следующие шаги:

1. Выпишите весь набор величин в один или несколько столбиков. Каждая величина – это  $x$ .
2. Вычислите сумму всех величин.
3. Разделите сумму на количество величин.

Например, если вам нужно рассчитать среднее количество покупателей в трех магазинах, вы можете вычислить среднее значение.

**Таблица 2.1** Количество покупателей в магазинах

Магазин	Годовое количество покупателей
Лэнхэм Парк	2150
Вильямсбург	1534
Центральный	3564



«Самые  
плохие»  
средние

Среднее количество покупателей в каждом магазине равно 2416 чел. Формула 2.2 показывает, как эта величина была рассчитана при помощи формулы 2.1.

$$\bar{X} = \frac{\sum x}{n} = \frac{2150 + 1534 + 3564}{3} = \frac{7248}{3} = 2416. \quad (2.2)$$

Если бы вам нужно было вычислить среднее количество учеников в классах с подготовительного по 6-й, вы бы делали то же самое.

Класс	Количество детей
Подготовительный	18
1	21
2	24
3	23
4	22
5	24
6	25

Среднее количество учеников в классе равно 22,43. Формула 2.3 показывает, как эта величина была рассчитана при помощи формулы 2.1.

$$\bar{X} = \frac{\sum x}{n} = \frac{18 + 21 + 24 + 23 + 22 + 24 + 25}{7} = \frac{157}{7} = 22,43. \quad (2.3)$$

Видите, мы же говорили, что это легко. Ничего сложного.

- Среднее значение иногда обозначается буквой  $M$  и называется также типичным (или центральным) значением. Если, читая книгу по статистике или отчет об исследовании, вы увидите что-то вроде  $M = 45,87$ , то, вероятно, это будет означать, что среднеарифметическое равно 45,87.
- В приведенной формуле маленькая  $n$  обозначает размер выборки, для которой рассчитывается среднее. Большая  $N$  будет означать размер генеральной совокупности. В некоторых книгах и научных статьях между ними не делают различия.
- Среднее значение выборки – это мера центральной тенденции, которая точнее прочих отражает среднюю всей генеральной совокупности.
- Среднее значение похоже на ось вращения циркулярной пилы. Это самая центральная точка, где сумма всех значений с одной стороны от нее равна сумме всех значений с другой.
- Наконец, плохо это или хорошо, но среднее значение очень чувствительно к экстремальным значениям величин. Экстремальное значение может сместить среднее в ту или иную сторону и сделать его менее репрезентативным для набора величин и менее ценным показателем центральной тенденции. Все это, конечно, зависит от величин, для которых рассчитывается среднее значение. Если у вас есть экстремальные величины, а среднее значение не работает так, как вам хотелось бы, у нас есть решение! Расскажем о нем позже.





Среднюю также называют арифметической средней. Кроме нее, есть еще и другие виды средней, о которых вы можете прочитать в различных источниках, например гармоническая средняя. Они применяются в особых случаях и здесь вам не нужны. Если вы хотите быть технически подкованным, то знайте, что арифметическая средняя (та самая, которую мы все это время обсуждали) также определяется как точка, около которой сумма всех отклонений равна 0. Так, если у вас есть значения 3, 4 и 5, среднее для которых равно 4, сумма отклонений вокруг среднего ( $-1$ , 0 и  $+1$ ) равна 0.



Помните, что слово «средняя» означает только один показатель, который наилучшим образом представляет набор данных, и что есть еще много других типов средних значений. Какой тип среднего вам использовать, зависит от того, на какой вопрос вы отвечаете и какой тип данных пытаетесь обобщить. Это имеет отношение к шкалам измерения, о которых мы поговорим позже в данной главе.

## Вычисление взвешенной средней

Вы только что увидели примеры расчетов обычной средней. Однако вы можете столкнуться с ситуацией, когда одно и то же значение встречается несколько раз, и тогда вам понадобится рассчитать взвешенную среднюю. **Взвешенную среднюю** легко посчитать, умножив значения на частоту их появления, сложив все произведения и затем поделив сумму на общее количество случаев. Это удобнее, чем складывать каждую величину по отдельности.



Для вычисления взвешенной средней выполните следующие шаги:

1. Запишите все значения выборки, для которой рассчитывается средняя, в приведенную ниже таблицу.
2. Запишите частоту, с которой встречается каждое значение.
3. Перемножьте величины на частоту, как показано в третьем столбике.
4. Просуммируйте все значения в столбце «Значение × частота».
5. Разделите на сумму частот.

Например, вот таблица, в которой сгруппированы результаты профессионального тестирования 100 пилотов гражданской авиации.

Значение оценки	Частота	Значение × частота
97	4	388
94	11	1034
92	12	1104

Продолжение ⇒

Значение оценки	Частота	Значение × частота
91	21	1911
90	30	2700
89	12	1068
78	9	702
60 (не летайте с этим парнем!)	1	60
ИТОГО	100	8967

Взвешенная средняя равна  $8967/100$ , или  $89,67$ . Вычислять среднюю таким способом гораздо легче, чем вводить 100 различных значений в калькулятор или специальную программу.



В основах статистики проводится важное различие между величинами, относящимися к выборке (части генеральной совокупности) и ко всей генеральной совокупности (или популяции) в целом. Для этого статистики прибегают к следующим договоренностям. Для показателей выборки (таких как средняя по выборке) используются латинские буквы. Для параметров генеральной совокупности (таких как средняя по генеральной совокупности) берут греческие буквы. Так, например, среднее значение оценок за контрольную у 100 пятиклассников будет обозначаться  $\bar{X}_5$ , тогда как средняя оценка для генеральной совокупности всех пятиклассников будет обозначаться греческой буквой «мю» –  $\mu_5$ .

## ВЫЧИСЛЕНИЕ МЕДИАНЫ

Медиана – это тоже средняя, только совсем другого рода. **Медиана** определяется как средняя точка в наборе величин. Это точка, в которой одна половина, или 50 %, значений лежит выше нее, а вторая половина, или 50 %, значений – ниже ее. У медианы есть особые свойства, о которых мы поговорим в этом разделе позже, а пока давайте сосредоточимся на том, как ее вычислить. Для вычисления медианы нет стандартной формулы.



Вычисление медианы

Для нахождения медианы выполните следующие шаги:

1. Запишите все величины в порядке от большего к меньшему или от меньшего к большему.
2. Найдите значение в самом центре списка. Это и есть медиана.



Например, вот доходы пяти разных домохозяйств:

\$135 456  
 \$25 500  
 \$32 456  
 \$54 365  
 \$37 668

Вот упорядоченный список от большего к меньшему:

\$135 456  
 \$54 365  
 \$37 668  
 \$32 456  
 \$25 500

В списке пять величин, в самом центре стоит \$37 668. Это и есть медиана.

А что делать, если количество значений четное? Давайте добавим к нашему списку величину \$34 500, чтобы теперь было 6 значений дохода. Вот они отсортированы по убыванию:

\$135 456  
 \$54 365  
 \$37 668  
 \$34 500  
 \$32 456  
 \$25 500

Когда количество величин четное, медиана – это просто среднее между двумя центральными значениями. В данном случае два центральных значения – это \$34 500 и \$37 668. Их среднее значение равно \$36 084, и это медиана для шести величин.

Что, если два центральных значения одинаковы, как в следующем наборе данных?

\$45 678  
 \$25 567  
 \$25 567  
 \$13 234

В этом случае медиана будет такая же, как обе эти центральные величины. В данном примере это \$25 567.

Если бы у нас был ряд значений, показывающий количество дней, затраченных на реабилитацию после спортивной травмы семью различными пациентами, то числа могли бы выглядеть так:

43  
 34  
 32  
 12  
 51  
 6  
 27

Как и раньше, мы можем отсортировать значения (51, 43, 34, 32, 27, 12, 6), а потом выбрать центральное значение в качестве медианы, которая в данном случае равна 32. Итак, медианное количество дней, затраченных на реабилитацию, равно 32.



Если вы знаете о медианах, то должны знать и о **процентильных точках**. Процентильные точки используются, чтобы определить долю случаев, находящихся до определенной точки в распределении или наборе величин. Например, если величина находится на 75-м процентиле, это означает, что она равна или превышает 75 % других величин в распределении. Медиана также называется 50-м процентилем, потому что это точка, ниже которой находятся 50 % случаев в распределении. Другие процентиля также весьма полезны, например 25-й процентиль, часто называемый  $Q_1$ , и 75-й процентиль, называемый  $Q_3$ . Что же такое  $Q_2$ ? Конечно, медиана.

Вот и ответ на вопрос, который, вероятно, интересовал вас с тех пор, как мы начали говорить про медиану. Зачем использовать медиану вместо среднего? Это нужно делать по одной очень простой причине. Медиана вовсе не чувствительна к крайним значениям, на которые реагирует среднее значение.

Когда у вас есть набор величин, в котором одно или несколько значений являются **крайними** (или **экстремальными**), медиана лучше обозначает самое центральное значение этого набора, чем любая другая мера центральной тенденции. Да, даже лучше, чем среднее значение.

Что мы имеем в виду под «крайним»? Наверное, проще всего предположить крайнюю величину как ту, которая очень отличается от группы, в которую она входит. Например, посмотрите на список пяти доходов, с которыми мы уже работали раньше (вот он еще раз):

\$135 456  
\$54 365  
\$37 668  
\$32 456  
\$25 500

Значение \$135 456 сильнее отличается от остальных четырех, чем любая другая величина в наборе. Мы будем считать ее крайней величиной.

Лучшим способом проиллюстрировать полезность медианы в качестве меры центральной тенденции будет расчет и средней, и медианы для набора данных, содержащего одно или несколько крайних значений, а затем сравнение, какая из них лучше всего описывает группу. Приступим.

Среднее значение пяти величин, приведенных выше, – это сумма всех величин, разделенная на 5, что дает в результате \$57 089. С другой стороны, медиана для этого набора равна \$37 668. Какое из этих значений лучше описывает группу? Это \$37 668, потому что оно явно лежит ближе к центру группы, а мы предполагаем, что «среднее» (в данном случае мы используем медиану в качестве показателя среднего значения) показывает типичное значение или занимает центральное положение в наборе данных. В реальности среднее



значение \$57 089 оказывается выше второго самого большого числа и не особо типично для этого распределения.

Именно по этой причине определенные социальные и экономические показатели (зачастую имеющие отношение к доходу) публикуются с использованием медианы в качестве меры центральной тенденции (например, «Медианный доход средней американской семьи равен...»), вместо того чтобы обобщать данные через среднее значение. В этих данных слишком много крайних величин, которые **скосят** или существенно исказят то, что является настоящей центральной точкой в наборе или распределении данных.



Вы уже узнали, что иногда среднее обозначается заглавной буквой  $M$  вместо  $\bar{X}$ . Для медианы тоже используются другие обозначения. Нам нравится буква  $M$ , но некоторые путаются, что она обозначает, поэтому используют для медианы сокращение *Med* или *Mdn*. Пусть это вас не смущает. Просто помните о том, что такое медиана и что она показывает, и вам будет несложно приспособиться к разным символам.



Вот еще несколько интересных и важных фактов о медиане:

- в то время как среднее значение – это средняя точка совокупности значений, медиана – это центральная точка набора наблюдений;
- из-за того что медиана принимает во внимание только количество наблюдений, а не их значения, экстремальные значения (иногда называемые **выбросами**) на нее не влияют.

## ВЫЧИСЛЕНИЕ МОДЫ

Третья, и последняя, мера центральной тенденции, о которой мы поговорим, мода, является самой обобщенной и наименее точной. Однако она играет очень важную роль в понимании характеристик набора данных. **Мода** – это значение, которое встречается чаще всего. Для его вычисления нет никакой формулы.



Для вычисления моды выполните следующие шаги:

1. Сделайте список всех значений, которые встречаются в распределении.
2. Укажите, сколько раз встречается каждое значение.
3. Определите самое частое значение (оно будет модой).



Применение  
SPSS для  
описательной  
статистики

Например, опрос 300 человек об их политических предпочтениях может показать следующие результаты (см. табл. 2.4).

Мода – это наиболее часто встречаемое значение. В данном примере это «Независимые кандидаты». Это и есть мода этого распределения.

Предпочитаемая партия	Количество раз
Демократы	90
Республиканцы	70
Независимые кандидаты	140

Если бы посмотрели на ответы теста, состоящего из 100 вопросов, то мы могли бы обнаружить, что экзаменуемые чаще всего выбирают ответ «А». Данные могли бы выглядеть следующим образом (см. табл. 2.5):

Выбранный вариант ответа	A	B	C	D
Количество раз	57	20	12	11

В этом тесте из 100 вопросов с множественным выбором из четырех вариантов (A, B, C и D) ответ A был выбран 57 раз. Этот ответ и есть мода.

Хотите знать самую простую и часто совершаемую ошибку при вычислении моды? Она состоит в выборе *частоты* повторения категории вместо *названия* категории. Вместо определения того, что модой является категория «Независимые кандидаты» (в нашем первом примере), кто-нибудь может легко прийти к заключению, что модой является число 140. Почему? Потому что человек смотрит на то, сколько раз встречалось значение, а не на само часто встречающееся значение! Эту ошибку очень просто сделать, поэтому будьте начеку, когда вам зададут такой вопрос.

## Пирожок с двумя начинками

Если каждое значение в распределении встречается одинаковое количество раз, то тогда в нем фактически нет ни одной моды. Но если с одинаковой частотой встречается более одного значения, то распределение будет **мультимодальным**. Например, набор величин может быть бимодальным (с двумя модами), как демонстрирует следующая совокупность данных (см. табл. 2.6):

Цвет волос	Количество случаев (или частота)
Рыжий	7
Блонд	12
Черный	45
Каштановый	45

Распределение в примере выше является бимодальным, поскольку частота значений «черный» и «каштановый» одинакова. У вас даже может быть бимодальное распределение, когда моды относительно близки друг к другу, но не совпадают полностью (например, 45 человек с черными волосами и 44 – с каштановыми). Вопрос в том, насколько одно количество случаев отличается от другого?

Может ли распределение быть тримодальным? Конечно, это возможно в том случае, когда три значения имеют одну и ту же частоту. Это маловероятно, особенно когда вы имеете дело с большой совокупностью **точек данных** (или наблюдений), но, определенно, возможно. Настоящий ответ на выделенный выше вопрос состоит в том, что категории должны быть взаимоисключающими, т. е. у вас не может быть одновременно рыжих и черных волос (хотя если осмотреться сейчас в аудитории, то вы можете со мной не согласиться). Конечно, вы можете одновременно иметь волосы, выкрашенные в два цвета, но цвет волос каждого человека должен быть учтен только в одной категории.

## КАК ВЫБРАТЬ НУЖНУЮ МЕРУ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ (И ВСЕ, ЧТО ВАМ НУЖНО СЕЙЧАС ЗНАТЬ О ШКАЛАХ ИЗМЕРЕНИЯ)

То, какую меру центральной тенденции использовать, зависит от определенных характеристик данных, с которыми вы работаете (особенно от **шкалы измерения**, в которой эти данные находятся). И эта шкала измерения будет диктовать, какую именно меру центральной тенденции нужно выбрать.

Но давайте на минутку вернемся назад, чтобы убедиться, что мы говорим на одном языке, начиная с представления о том, что же такое измерение.

Измерение – это присвоение значений результатам наблюдений в соответствии с набором правил, т. е. все очень просто. В итоге мы получаем разные шкалы, о которых пойдет речь далее, а результат наблюдений – это все то, что нам интересно было измерить, например цвет волос, пол, результаты тестирования или рост.

Эти шкалы измерений, или правила, представляют собой определенные уровни, на которых мы наблюдаем результаты. У каждого уровня есть свой набор параметров, и шкалы измерения представлены в четырех видах: номинальная, порядковая, интервальная и пропорциональная.

Давайте перейдем к краткому обсуждению всех четырех шкал измерения и приведем примеры, а затем поговорим, как эти шкалы соотносятся с разными мерами центральной тенденции.

## **Хоть розой назови ее, хоть нет: номинальная шкала измерения**

**Номинальная шкала измерения** определяется характеристиками результатов наблюдения, которые попадают в один и только один класс или категорию. Например, номинальной переменной может быть пол (женский или мужской), этническая принадлежность (европеец или афроамериканец), а также политические предпочтения (республиканец, демократ или независимый). Переменные номинального уровня – это «имена» (от лат. *nominal*), и номинальный уровень может быть наименее полным уровнем измерения. Номинальные шкалы измерения имеют взаимоисключающие категории. Например, политическая принадлежность не может быть одновременно республиканской и демократической.

## **Любой порядок подойдет: порядковая шкала измерения**

Как следует из названия, **порядковая шкала измерения** имеет дело с порядком, и измеряемой характеристикой является то, в каком порядке находятся объекты. Отличным примером является ранжирование кандидатов на вакансию. Если мы знаем, что Расс – это кандидат № 1, Шелдон – № 2, а Ханна – № 3, то мы имеем дело с порядковой классификацией. Мы не знаем, намного ли выше на этой шкале стоит Расс по отношению к Шелдону, чем Шелдон по отношению к Ханне. Мы знаем только, что лучше быть № 1, чем № 2, а № 2 лучше, чем № 3, но не знаем, насколько.

## **1 + 1 = 2: интервальная шкала измерения**

Теперь мы кое к чему подходим. Когда мы говорим об **интервальной шкале измерения**, оценивание основывается на некотором непрерывном континууме, таком, что мы можем говорить о том, насколько именно высокий результат больше, чем низкий. Например, если в орфографическом тесте вы верно написали 10 слов, то это в 2 раза больше, чем пять правильно написанных слов. Отличительной характеристикой интервальной шкалы является то, что интервалы (или промежутки) между точками на шкале равны между собой. Десять правильных слов на два больше, чем 8, что на три больше, чем 5.

## **Можно ли не иметь ничего? Пропорциональная шкала измерения**

Что ж, вот небольшая головоломка для вас. Критерии оценки в **пропорциональной шкале измерения** характеризуются наличием на шкале абсолютного нуля. Он означает отсутствие любого прояв-

ления измеряемого признака. В чем состоит головоломка? Есть ли в наших наблюдениях возможность получить отсутствие того, что измеряется? В некоторых областях такое может быть. Например, в физике или биологии у вас может наблюдаться отсутствие характеристики, такое как абсолютный ноль (отсутствие движения молекул) или нулевая освещенность. В социальных и поведенческих науках немного сложнее. Даже если вы получите ноль в том орфографическом тестировании или не ответите ни на один вопрос IQ-теста (на китайском языке), означает ли это, что у вас полностью отсутствуют орфографические или интеллектуальные способности?

### Подводя итог...

Шкалы (или правила) измерения представляют определенные уровни, на которых замеряются наблюдаемые результаты. Подводя итог, мы можем сказать следующее:

- любой результат наблюдения можно задать в одной из четырех шкал измерения;
- шкалы измерения можно расставить в порядке от наименее (номинальной) к наиболее тонко настроенной (пропорциональной);
- чем «выше» находится шкала измерения, тем точнее, детальнее и информативнее собираемые данные. Может быть достаточно того, чтобы знать, что некоторые люди богатые, а некоторые – бедные (и это номинальное, или категориальное, разграничение), но гораздо лучше знать точно, сколько денег они зарабатывают (пропорциональная шкала). Имея всю информацию, мы всегда сможем отделить богатых от бедных, если захотим;
- наконец, более тонко настроенные шкалы содержат все качества шкал, находящихся под ними. Например, интервальная шкала обладает характеристиками порядковой и номинальной. Если вы знаете, что средний уровень подачи у «Кабс» составляет 0,350, то вам известно, что это на 100 очков лучше, чем у «Тайгерс» (которые выбивают 0,250), но вы также знаете и то, что «Кабс» лучше, чем «Тайгерс» (но не знаете, насколько), и то, что «Кабс» отличаются от «Тайгерс» (но неизвестно, чем именно).

Мы определили шкалы измерения и обсудили три меры центральной тенденции, приведя довольно понятные примеры каждой из них. Однако самый важный вопрос все еще остается без ответа: когда же и какую меру использовать?

Вообще, выбор меры центральной тенденции зависит от типа описываемых вами данных, что, в свою очередь, означает уровень, на котором данные замеряются. Не вызывает вопросов, что центральную тенденцию для качественных, категориальных или номинальных данных (таких как расовая принадлежность, цвет глаз,

группа доходов, политические предпочтения и район проживания) можно описать только при помощи моды.

Например, если вас интересует мера центральной тенденции, описывающая преобладающие политические предпочтения в группе, то вы не можете использовать среднее. Что вообще вы узнаете с его помощью? То, что каждый человек наполовину республиканец? А вот утверждение, что из 300 человек почти половина (140) предпочитает независимую партию, лучше всего опишет значение этой переменной. В общем, медиану и среднюю лучше всего использовать с количественными данными, такими как рост, размер дохода в долларах (не по категориям), возраст, результаты контрольной, время реакции и количество часов, прослушанных для получения диплома.

Также будет справедливо сказать, что средняя точнее измеряет центральную тенденцию, чем медиана, а медиана – точнее, чем мода. Это означает, что при всех прочих равных лучше использовать среднюю. И она действительно применяется чаще всего. Однако бывают случаи, когда средняя не подходит в качестве меры центральной тенденции. Это возможно тогда, например, когда мы имеем дело с категориальными или номинальными данными, такими как пациенты на стационарном и амбулаторном лечении. Тогда мы используем моду.

Итак, вот три инструкции, которые могут вам помочь. Помните, что всегда могут быть исключения.

1. Используйте моду, когда имеете дело с категориальными данными, чьи значения могут попадать только в один класс (такими как цвет волос, политические предпочтения, место жительства и религия). В подобных случаях эти категории называются взаимоисключающими.
2. Используйте медиану, когда имеете дело с крайними значениями и не хотите исказить средние показатели (как в обычной средней), например как в случае, когда интересующая вас переменная – это доход, выраженный в долларах.
3. Наконец, используйте среднюю, когда имеете дело с данными без крайних значений, которые при этом не являются категориальными (например, с числовой оценкой за тест или количеством секунд, за которое можно проплыть 50 ярдов).

## ВЫЧИСЛЕНИЕ ОПИСАТЕЛЬНЫХ СТАТИСТИК ПРИ ПОМОЩИ КОМПЬЮТЕРА



Если вы этого еще не сделали, то сейчас самое время обратиться к приложению А и ознакомиться с основами работы в SPSS. Потом возвращайтесь сюда.

Давайте посчитаем в SPSS некоторые описательные статистики. Набор данных, который мы будем использовать, называется «Глава 2. Набор данных 1» (Chapter 2 Data set 1) и состоит из 20 результатов тестирования на предубеждения. Все наборы данных доступны в приложении С и на сайте издательства SAGE по адресу [edge.sagepub.com/salkind6e](http://edge.sagepub.com/salkind6e). Их также можно получить по запросу на почту автора: [njs@ku.edu](mailto:njs@ku.edu). В данном наборе используется только одна переменная:

Таблица 2.7	
Переменная набора данных «Глава 2. Набор данных 1» (Chapter 2. Data set 1)	
Переменная	Определение
Prejudice (Предубеждения)	Результат тестирования на предубеждения в шкале от 1 до 100



Ниже приведены этапы вычисления мер центральной тенденции, которые обсуждались в этой главе. Попробуйте самостоятельно сделать это, строго соблюдая все этапы. В этом и других упражнениях предполагается, что набор данных уже открыт в SPSS (не важно, вводили вы их или скачали).

1. Выберите в меню **Анализ** (Analyze) ⇒ **Описательные статистики** (Descriptive Statistics) ⇒ **Частоты** (Frequencies).
2. Щелкните 2 раза по переменной «Prejudice», чтобы переместить ее в блок **Переменные** (Variable(s)).
3. Нажмите на **Статистики** (Statistics), **Таблица**, и вы увидите диалоговое окно **Частоты: Статистики** (Frequencies: Statistics), показанное на рис. 2.1.
4. В блоке **Положение центра распределения** (Central Tendency) поставьте галочки напротив **Среднее значение** (Mean), **Медиана** (Median) и **Мода** (Mode).
5. Нажмите **Продолжить** (Continue).
6. Нажмите **ОК**.

## Выводы SPSS

На рис. 2.2 показаны выводы процедур SPSS для переменной Prejudice.

В таблице **Статистики** (Statistics) вы можете увидеть рассчитанные значения среднего, медианы и моды вместе с указанием размера выборки и подтверждением, что ни одни данные не были пропущены. SPSS не использует в выводах такие символы, как  $X$ . В выводах также представлены частоты для каждого значения в количественном и процентном выражении, что тоже является полезной описательной информацией.

Рис. 2.1 Диалоговое окно SPSS Частоты: Статистики (Frequencies: Statistics)

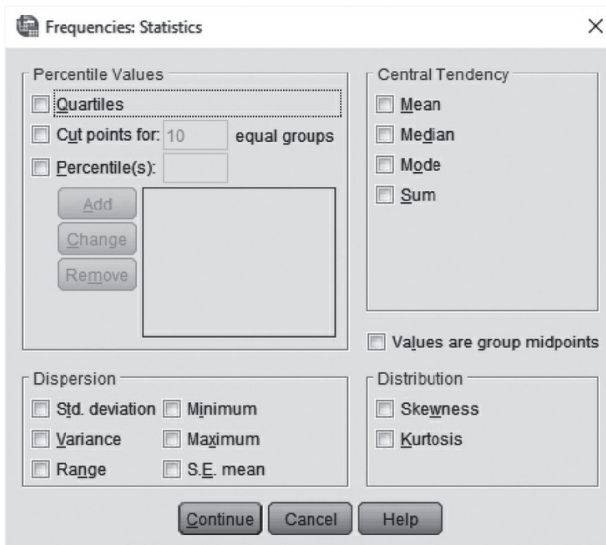


Рис. 2.2 Описательные статистики из SPSS

→ **Frequencies**

[DataSet1] C:\Textbook Stuff\Stat for People 6e\Chapte

**Statistics**

Prejudice

N	Valid	20
	Missing	0
Mean		84.70
Median		87.00
Mode		87

**Prejudice**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	55	1	5.0	5.0
	64	1	5.0	10.0
	67	1	5.0	15.0
	76	1	5.0	20.0
	77	1	5.0	25.0
	81	2	10.0	35.0
	82	1	5.0	40.0
	87	4	20.0	60.0
	89	1	5.0	65.0
	93	1	5.0	70.0
	94	2	10.0	80.0
	96	1	5.0	85.0
	99	3	15.0	100.0
Total	20	100.0	100.0	





Это немного странно, но если вы выберете в SPSS **Анализ** (Analyze) ⇒ **Описательные статистики** (Descriptive Statistics) ⇒ **Описательные** (Descriptives), вместо того чтобы пойти в **Частоты** (Frequencies), а затем нажмете **Параметры** (Options), то там не будет параметра медианы или моды, как можно было бы ожидать, поскольку они тоже являются базовыми описательными статистиками. Какой из этого вывод? Программы статистического анализа обычно отличаются друг от друга, используют разные названия для одних и тех же вещей и разные предположения о том, что где должно находиться. Если вы не можете найти то, что вам нужно, вероятно, оно там есть. Просто продолжайте искать. Не забывайте также пользоваться **Помощью** (Help) для навигации среди всей этой новой информации, пока не найдете то, что требуется.

### Читаем выводы SPSS

Данный вывод довольно простой и понятный.

Среднее значение для 20 переменных равно 84,70 (помните, что шкала возможных значений – от 0 до 100). Медиана, или точка, в которой 50 % всех переменных оказывается выше нее, а 50 % – ниже, равна 87 (что довольно близко к среднему значению), а наиболее часто встречающийся результат, или мода, равен 87.



Вывод SPSS может быть забит информацией или будет показывать только основное. Все зависит от типа выполняемого анализа. В примере выше мы получили только основное, и, честно говоря, именно это нам и требовалось. По мере чтения этой книги вы увидите разные выводы и узнаете, что они означают, но в некоторых случаях обсуждение полного набора информации в выводе совсем не входит в задачи книги. Мы сфокусируемся на выводах, которые имеют прямое отношение к тому, что вы изучаете в главе.

## РЕАЛЬНАЯ СТАТИСТИКА



Большой смысл применение описательных статистик имеет при использовании их в опросах и голосованиях. Таких опросов уже буквально миллионы (как во время любых президентских выборов в США). В своей статье Роджер Моррелл (Roger Morrell) и его коллеги описывают особенности пользования интернетом 550 взрослых нескольких возрастных категорий, включая людей среднего (40–59 лет), раннего пожилого (60–74 лет) и глубоко пожилого возраста (75–92 лет). Получив 71 % откликов, что довольно хорошо, они обнаружили несколько очень интересных (хотя и не сказать, что неожиданных) результатов:

- пользователи интернета имеют явные возрастные и демографические различия;

- пользователи среднего и раннего пожилого возраста схожи в своих интернет-привычках;
- два основных препятствия для пользования интернетом заключаются в отсутствии доступа к компьютеру и недостаточных знаниях об интернете;
- глубоко пожилые взрослые заинтересованы в использовании интернета меньше всех;
- основные вопросы при обучении использованию интернета – это работа с электронной почтой, доступ к медицинской информации и информации о путешествиях для удовольствия.

Этот опрос, для обработки которого в основном использована описательная статистика, был проведен почти 20 лет назад, и с тех пор очень многое изменилось. Но причина, по которой мы включаем его в пример реальных данных, заключается в том, что он является хорошей иллюстрацией исторической узловой точки, которую можно использовать для сравнения в будущих исследованиях. Описательные статистики и подобные исследования часто применяются именно в таких целях.

*Хотите узнать больше?* Найдите в интернете или библиотеке статью...

*Morrell R. W., Mayhorn C. B. & Bennett J. A survey of World Wide Web Use in middle-aged and older adults. Human Factors, 2010. P. 175–182.*



Реальная статистика

## ВЫВОДЫ ПО ГЛАВЕ

Независимо от того, насколько вы продвинуты в статистике, почти всегда начнете с простого описания того, что есть. Отсюда происходит важность понимания простой сущности центральной тенденции. Далее мы отправимся к другой важной описательной составляющей – изменчивости, т. е. тому, насколько данные отличаются друг от друга. Это мы изучим в главе 3!



Выводы по главе

## ВРЕМЯ ПРАКТИКИ

1. Вручную посчитайте среднее значение, медиану и моду для следующего набора из 40 результатов контрольной по химии (табл. 2.8).

93	85	99	77
94	99	86	76
95	99	97	84

Продолжение ⇒

91	89	77	87
97	83	80	98
75	94	81	85
78	92	89	94
76	94	96	94
90	79	80	92
77	86	83	81



Проблема 2

2. Рассчитайте среднее значение, медиану и моду для следующих трех наборов данных (табл. 2.9), приведенных также в файле «Глава 2. Набор данных 2» (Chapter 2. Data set 2). Сделайте это вручную или при помощи программы вроде SPSS. Покажите свои расчеты, а если вы используете SPSS, то распечатайте страницу выводов.

Таблица 2.9 Три набора данных из файла «Глава 2. Набор данных 2» (Chapter 2. Data set 2)		
Набор 1	Набор 2	Набор 3
3	34	154
7	54	167
5	17	132
4	26	145
5	34	154
6	25	145
7	14	113
8	24	156
6	25	154
5	23	123

3. Рассчитайте средние для следующих наборов данных (табл. 2.10), сохраненных так же, как «Глава 2. Набор данных 3» (Chapter 2. Data set 3), при помощи SPSS. Распечатайте страницу выводов.

Таблица 2.10 Данные по больнице	
Количество коек	Заболеваемость (на 1000 обратившихся)
234	1,7
214	2,4
165	3,1
436	5,6
432	4,9
342	5,3
276	5,6
187	1,2
512	3,3
553	4,1

4. Вы менеджер ресторана быстрого питания. Одна из ваших задач – в конце каждого дня сообщать начальнику, какое блюдо продается лучше всего. Используйте свои глубокие знания описательных статистик и напишите один параграф, чтобы рассказать начальнику, что происходило сегодня. Вот данные. Не используйте SPSS для расчета важных величин, лучше сделайте это вручную. Не забудьте приложить к отчету копию ваших расчетов.

**Таблица 2.11** Продажи ресторана быстрого питания за один день

Блюдо	Количество продаж	Цена
Огромный бургер	20	\$2,95
Бэби-бургер	18	\$1,49
Куриные малышки	25	\$3,50
Свиной бургер	19	\$2,95
Бургер «Вкусняшка»	17	\$1,99
Хот-дог	20	\$1,99
Продажи ИТОГО	119	

5. Представьте, что вы глава огромной корпорации и планируете расширение. Вам хотелось бы, чтобы новый магазин давал те же результаты, что и остальные три в вашей империи (см. табл. 2.12). Набросайте вручную, каких финансовых результатов вы хотите добиться от нового магазина. И помните: нужно решить, что использовать в качестве среднего: среднее значение, медиану или моду. Удачи, юный джедай!

**Таблица 2.12** Показатели розничных магазинов

Среднее	Магазин 1	Магазин 2	Магазин 3	Новый магазин
Продажи (тыс. долл.)	323,6	234,6	308,3	
Кол-во проданных штук	3454	5645	4565	
Кол-во посетителей	4534	6765	6654	

6. Ниже в табл. 2.13 приведены рейтинги (по шкале от 1 до 5) для угощений на вечеринке Суперкубка. Вам нужно решить, какая еда имеет самый высокий рейтинг (5 – победитель, 1 – проигравший). Определите, какой тип среднего использовать и почему. Сделайте это вручную или в SPSS.



Проблема 6

**Таблица 2.13** Рейтинг предпочтений снеков

Еда	Фанаты Севера	Фанаты Востока	Фанаты Юга	Фанаты Запада
Начос с добавками	4	4	5	4
Фруктовая кружка	2	1	2	1

Продолжение ⇒

Еда	Фанаты Севера	Фанаты Востока	Фанаты Юга	Фанаты Запада
Острые крылышки	4	3	3	3
Гигантская пицца с доп. начинкой	3	4	4	5
Цыпленок в пиве	5	5	5	4

- При каких условиях вы скорее будете использовать медиану, чем среднее значение в качестве меры центральной тенденции? Почему? Приведите пример двух ситуаций, когда медиана может оказаться полезнее, чем среднее значение.
- Предположим, вы работаете с набором данных, в котором встречаются сильно отличающиеся (очень большие или очень маленькие по сравнению с остальными) значения. Какой мерой центральной тенденции вы воспользуетесь и почему?
- Для этого упражнения используйте набор из 16 отсортированных значений, представляющих уровни дохода от \$50 000 до \$200 000. Какая мера центральной тенденции подойдет лучше всего и почему?

\$199 999

\$98 789

\$90 878

\$87 678

\$87 245

\$83 675

\$77 876

\$77 743

\$76 564

\$76 465

\$75 643

\$66 768

\$65 654

\$58 768

\$54 678

\$51 354

- Используйте данные файла «Глава 2. Набор данных 4» (Chapter 2. Data set 4) и вручную рассчитайте среднее отношение трех групп людей к опыту использования общественного городского транспорта (где 10 – очень положительное, а 1 – отрицательное отношение).
- Взгляните на следующие цифры, характеризующие продажи пирогов в закусочной «Леди Берд» (табл. 2.14), и определите среднее количество заказов для каждой недели.



**Таблица 2.14** Данные о продажах пирогов

Неделя	Шоколадный шелк	Яблочный	Дуглас Каунти
1	12	21	7
2	14	15	12
3	18	14	21
4	27	12	15

## САЙТ С МАТЕРИАЛАМИ ДЛЯ ОБУЧЕНИЯ

 **SAGE** edge™

Получите ресурсы для прокачки своих навыков! Заходите на сайт [edge.sagepub.com/salkind6e](http://edge.sagepub.com/salkind6e) и получайте доступ к практическим викторинам, электронным флеш-картам, оригинальным и специально отобраным видео, наборам данных, научным статьям и многому другому!