

Содержание

Предисловие	17
Большая работа: пересмотр материала и проверка в студенческой аудитории	17
Стиль книги	18
Примеры	18
Статистические графики	18
Ситуации для анализа	19
Что нового в четвертом издании	19
Структура книги	19
Руководство к Excel	22
Не обязательное для изучения приложение StatPad	22
Руководство для преподавателя	22
Тесты	22
Благодарности	23
Обращение к студентам	24
Об авторе	25
ЧАСТЬ I. ВВЕДЕНИЕ И ОПИСАТЕЛЬНАЯ СТАТИСТИКА	27
Глава 1. Введение: роль статистики в бизнесе	28
1.1. Почему именно статистика?	28
Почему необходимо изучать статистику?	29
Сложна ли статистика?	29
Влияет ли знание статистики на гибкость принятия решений?	29
1.2. Что такое статистика?	29
Статистика рассматривает общую картину	30
Статистика не игнорирует отдельные объекты	30
Посмотрим на данные	31
Статистика в менеджменте	31
1.3. Четыре основных этапа статистического анализа	31
Оценка неизвестной величины	33
Проверка гипотез	34
1.4. Что такое вероятность	36
1.5. Общий совет	37
1.6. Дополнительный материал	37
Резюме	37
Основные термины	38
Контрольные вопросы	39
Задачи	39
Проект	41
Глава 2. Структуры данных: классификация различных типов наборов данных	42
2.1. Сколько переменных?	42
Одномерные данные	43

Двумерные данные	44
Многомерные данные	45
2.2. Количественные данные: числа	47
Дискретные количественные данные	47
Непрерывные количественные данные	47
Остерегайтесь чисел, не имеющих содержательную интерпретацию	48
2.3. Качественные данные: категории	49
Порядковые качественные данные	49
Номинальные качественные данные	50
2.4. Временные ряды и данные об одном временном срезе	50
2.5. Источники данных, включая Internet	52
2.6. Дополнительный материал	59
Резюме	59
Основные термины	61
Контрольные вопросы	61
Задачи	62
Упражнения с использованием базы данных	69
Проекты	69
Глава 3. Гистограммы: взгляд на распределение данных	70
3.1. Последовательность данных	71
Числовая ось	72
3.2. Использование гистограмм для отображения частот	73
Гистограммы и столбиковые диаграммы	77
3.3. Нормальное распределение	79
3.4. Несимметричные распределения и преобразование данных	81
Проблема с асимметрией	85
Выход с помощью преобразования	85
Интерпретация и вычисление логарифма	87
3.5. Бимодальные распределения	88
3.6. Выбросы (сильно отклоняющиеся значения)	91
Работа с выбросами (сильно отклоняющимися значениями)	92
3.7. Гистограммы, построенные вручную: метод “ствол и листья”	97
3.8. Дополнительный материал	99
Резюме	99
Основные термины	101
Контрольные вопросы	101
Задачи	102
Упражнения с использованием базы данных	114
Проекты	115
Ситуация для анализа	115
Необходимость контроля производственных потерь	115
Вопросы для обсуждения	116
Глава 4. Обобщающие показатели: интерпретация типических значений и перцентилей	117
4.1. Чему равно наиболее типическое значение?	118
Среднее: типическое значение для количественных данных	118
Взвешенное среднее: учет важности	121
Медиана: типическое значение для количественных и порядковых данных	126
Мода: типическое значение даже для номинальных данных	133
Какие показатели нужно использовать	135
4.2. Что такое перцентиль	137
Экстремумы, квартили и блочные диаграммы	137
Функция кумулятивного распределения показывает перцентили	142
4.3. Дополнительный материал	149

Резюме	149
Основные термины	151
Контрольные вопросы	152
Задачи	153
Упражнения с использованием базы данных	164
Проекты	164
Ситуация для анализа	165
Управленческие прогнозы о производстве и маркетинге, или “Случай подозрительного потребителя”	165
Вопросы для обсуждения	168
Глава 5. Изменчивость: изучение разнообразия	169
5.1. Стандартное отклонение: традиционный выбор	171
Определение и формула для стандартного отклонения и дисперсии	172
Использование калькулятора или компьютера	174
Интерпретация стандартного отклонения	175
Интерпретация стандартного отклонения для нормального распределения	177
Стандартное отклонение выборки и генеральной совокупности	187
5.2. Размах: быстрая и поверхностная оценка	189
5.3. Коэффициент вариации: мера <i>относительной</i> изменчивости	192
5.4. Результаты прибавления константы или изменения шкалы	194
5.5. Дополнительный материал	196
Резюме	196
Основные термины	198
Контрольные вопросы	199
Задачи	200
Упражнения с использованием базы данных	215
Проекты	216
Ситуация для анализа	216
Следует ли продолжать работу с этим поставщиком?	216
ЧАСТЬ II. ВЕРОЯТНОСТЬ	219
Глава 6. Вероятность: разбираемся в случайных ситуациях	220
6.1. Пример: за какой из дверей спрятан приз?	222
6.2. Как исследовать неопределенность	223
Случайный эксперимент: <i>точное</i> определение случайной ситуации	223
Выборочное пространство: перечень возможных событий	224
Результат: что происходит в действительности	226
События: они либо происходят, либо нет	227
6.3. Насколько вероятно событие?	229
Каждое событие имеет свою вероятность	229
Откуда берутся значения вероятности	230
Относительная частота и закон больших чисел	230
Теоретическое значение вероятности	233
Правило равной вероятности	233
Субъективная оценка вероятности	234
Анализ методом Байеса и частотный анализ	236
6.4. Как совместить информацию о нескольких событиях	237
Диаграммы Венна позволяют увидеть все возможности	237
<i>Не</i> событие	238
Правило дополнения (правило <i>не</i>)	238
Одно событие <i>и</i> другое	239
Если два события не могут наблюдаться вместе	240
Правило пересечения (<i>и</i>) для несовместимых событий	240
Одно событие <i>или</i> другое	240
Правило объединения (<i>или</i>) для несовместимых событий	241
Нахождение <i>или</i> из <i>и</i> и наоборот	242

Одно событие <i>при условии</i> другого: учет имеющейся информации	243
Правило вычисления условной вероятности при наличии дополнительной информации	244
Условные вероятности для несовместимых событий	245
Независимые события	246
Правило пересечения (<i>и</i>) для независимых событий	248
Связь между независимыми и несовместимыми событиями	249
6.5. Как решать вероятностные задачи	249
Дерево вероятностей	249
Правила построения дерева вероятностей	250
Таблица совместных вероятностей	257
6.6. Дополнительный материал	259
Резюме	259
Основные термины	262
Контрольные вопросы	263
Задачи	264
Упражнения с использованием базы данных	274
Проект	275
Ситуация для анализа	276
Детективная история: кто же все-таки ответствен за увеличение количества дефектов в последнее время?	276
Глава 7. Случайные величины: работа с неопределенными значениями	278
7.1. Дискретные случайные величины	280
Вычисление среднего и стандартного отклонения	281
7.2. Биномиальное распределение	285
Определение биномиального распределения и биномиального соотношения	285
Вычисление среднего и стандартного отклонения: короткий путь	288
Вычисление вероятностей	290
7.3. Нормальное распределение	295
Представление вероятности как площади под кривой	297
Стандартное нормальное распределение Z и соответствующие вероятности	298
Решение задач на вычисление вероятности при нормальном распределении	302
Четыре способа вычисления вероятности	308
Внимание! Не все распределено нормально!	308
7.4. Аппроксимация биномиального распределения нормальным	309
7.5. Распределение Пуассона и экспоненциальное распределение	315
Распределение Пуассона	315
Экспоненциальное распределение	318
7.6. Дополнительный материал	320
Резюме	320
Основные термины	323
Контрольные вопросы	324
Задачи	325
Упражнения с использованием базы данных	335
Проекты	335
Ситуация для анализа	336
Стоимость опциона на аренду нефтяного месторождения	336
ЧАСТЬ III. СТАТИСТИЧЕСКИЙ ВЫВОД	339
Глава 8. Построение случайной выборки: предварительное планирование для сбора данных	340
8.1. Генеральные совокупности и выборки	341
Что такое репрезентативная выборка	343

Параметры выборки и параметры генеральной совокупности	345
8.2. Случайная выборка	346
Извлечение случайной выборки	347
Извлечение выборки методом перемешивания генеральной совокупности	350
8.3. Выборочное распределение и центральная предельная теорема	352
8.4. Стандартная ошибка как <i>оценка</i> стандартного отклонения	359
Насколько отличается среднее выборки от среднего генеральной совокупности	360
Поправка для малой генеральной совокупности	363
Стандартная ошибка биномиальной доли	366
8.5. Другие методы построения выборки	367
Стратифицированная случайная выборка	368
Систематическая выборка	372
8.6. Дополнительный материал	374
Резюме	374
Основные термины	378
Контрольные вопросы	378
Задачи	380
Упражнения с использованием базы данных	390
Проекты	392
Ситуация для анализа	393
Можно ли извлечь пользу из этого исследования?	393
Глава 9. Доверительные интервалы: допущение о неточности оценок	395
9.1. Доверительный интервал для среднего значения и для доли признака в генеральной совокупности	398
t-таблица и t-распределение	401
Часто используемый 95% доверительный интервал	404
Другие доверительные уровни	409
9.2. Предположения, необходимые для корректного использования	412
Случайная выборка	413
Нормальное распределение	416
9.3. Интерпретация доверительного интервала	417
Какое событие имеет вероятность 95%?	417
Ваши жизненные достижения	418
9.4. Односторонние доверительные интервалы	419
Внимание! Не всегда можно использовать односторонний доверительный интервал	419
Вычисление одностороннего доверительного интервала	420
9.5. Интервалы предсказания	423
9.6. Дополнительный материал	426
Резюме	426
Основные термины	428
Контрольные вопросы	428
Задачи	430
Упражнения с использованием базы данных	440
Проекты	442
Ситуация для анализа	443
Многообещающие результаты опроса относительно заказов фирменных товаров по каталогу	443
Глава 10. Проверка статистических гипотез: выбор между реальностью и совпадением	445
10.1. Не все гипотезы одинаковы!	446
Нулевая гипотеза	446
Исследовательская гипотеза	447

О чем свидетельствует результат	447
Примеры гипотез	448
10.2. Проверка гипотезы о равенстве среднего генеральной совокупности некоторому заданному значению	450
Использование доверительных интервалов: простой способ	451
t-статистика: способ другой, результат тот же	459
10.3. Интерпретация проверки гипотезы	462
Ошибки I и II рода	462
Условия применимости	464
Гипотезы не могут быть вероятно истинными или вероятно ложными	464
Статистическая значимость и уровни проверки	465
Доверительная вероятность (p-значение)	466
10.4. Односторонняя проверка	469
Как выполнять проверку	471
10.5.4. Проверка того, принадлежит ли новое наблюдение той же генеральной совокупности	478
10.6. Сравнение двух выборок	479
t-тест для зависимых выборок	480
t-тест для независимых выборок	482
10.7. Дополнительный материал	487
Резюме	487
Основные термины	492
Контрольные вопросы	493
Задачи	495
Упражнения с использованием базы данных	512
Проекты	513
Ситуация для анализа	513
Так много рекламы, так мало времени	513
ЧАСТЬ IV. РЕГРЕССИЯ И ВРЕМЕННЫЕ РЯДЫ	517
Глава 11. Корреляция и регрессия: измерение и прогнозирование взаимосвязей	518
11.1. Исследование взаимосвязей с помощью диаграмм рассеяния и корреляций	519
Диаграмма рассеяния демонстрирует взаимосвязь	520
Корреляция характеризует силу взаимосвязи	521
Формула для вычисления коэффициента корреляции	522
Различные типы взаимосвязей	525
Отсутствие взаимосвязи	531
Нелинейная взаимосвязь	533
Неодинаковая вариация	536
Разделение совокупности на группы	540
Выбросы (резко отклоняющиеся значения)	542
Корреляция — это не причинная обусловленность	544
11.2. Регрессия: предсказание одного фактора на основании другого	546
Прямая линия характеризует линейную взаимосвязь	547
Прямые линии	549
Построение линии на основе данных	549
Насколько полезна построенная линия	555
Стандартная ошибка оценки: насколько велики ошибки предсказания	555
R^2 : как много объяснено	558
Доверительные интервалы и проверка гипотез для регрессии	558
Предположение о линейности определяет генеральную совокупность	559
Стандартные ошибки для наклона и сдвига	560
Доверительные интервалы для коэффициентов регрессии	561
Проверка того, является ли связь реальной или случайной	562
Другие методы проверки значимости взаимосвязи	563

Результаты компьютерных вычислений для данных о производственных затратах	564
Проверки других гипотез о коэффициенте регрессии	566
Новое наблюдение: неопределенность и доверительный интервал	568
Среднее значение Y : неопределенность и доверительный интервал	570
Регрессия может вводить в заблуждение	572
Линейная модель может оказаться неверной	573
Трудно предсказать вмешательство исходя из наблюдаемого опыта	576
Сдвиг может быть лишен смысла	576
Представление Y на основании X и представление X на основании Y	577
Скрытый “третий фактор” может быть полезен	578
11.3. Дополнительный материал	578
Резюме	578
Основные термины	583
Контрольные вопросы	583
Задачи	586
Упражнения с использованием базы данных	605
Проекты	607
Ситуация для анализа	608
Еще один этап производства: нужен ли он?	608
Глава 12. Множественная регрессия: прогнозирование одного фактора на основе нескольких других	611
12.1. Интерпретация результатов множественной регрессии	613
Коэффициенты регрессии и уравнение регрессии	618
Интерпретация коэффициентов регрессии	620
Прогнозы и ошибки прогнозирования	621
Насколько хороши наши прогнозы	625
Типичная ошибка прогнозирования: стандартная ошибка оценки	625
Объясненный процент вариации: R^2	625
Статистический вывод в случае множественной регрессии	626
Предположения	626
Значима ли модель? F -тест или тест R^2	628
Таблицы критических значений для тестирования R^2	632
Какие переменные являются значимыми: t -тест для каждого коэффициента	641
Другие проверки, касающиеся коэффициента регрессии	644
Какие переменные оказывают большее влияние	645
Сравнение стандартизованных коэффициентов регрессии	645
Сравнение коэффициентов корреляции	647
12.2. Сложности и проблемы, связанные с множественной регрессией	649
Мультиколлинеарность: не слишком ли схожи между собой объясняющие переменные?	650
Выбор переменной: может быть, мы пользуемся “не теми” переменными?	656
Классификация перечня X -переменных по приоритетам	657
Автоматизация процесса выбора переменных	658
Неправильный выбор модели: возможно, уравнение регрессии имеет неправильную форму?	660
Анализ данных с целью выявления нелинейности или неравной изменчивости	661
Использование диагностической диаграммы для выяснения наличия проблем	662
Использование процентных изменений для моделирования экономических временных рядов	669
12.3. Нелинейные взаимосвязи и неравная изменчивость	672
Преобразование взаимосвязи в линейную форму: интерпретация результатов	673
Подгонка кривой с помощью полиномиальной регрессии	679
Моделирование взаимодействия между двумя X -переменными	682

12.4. Индикаторные переменные: прогнозирование на основе категорий	684
Интерпретация и проверка значимости коэффициентов регрессии для индикаторных переменных	686
Раздельные регрессии	691
12.5. Дополнительный материал	692
Резюме	692
Основные термины	695
Контрольные вопросы	696
Задачи	697
Упражнения с использованием базы данных	714
Проект	716
Ситуация для анализа	717
Контроль качества продукции	717
Глава 13. Составление отчетов: представление результатов множественной регрессии	720
13.1. Как организовать свой отчет	722
Абзац, содержащий реферат	723
Вводная часть	724
Раздел “Анализ и методы”	724
Раздел “Выводы и резюме”	726
Включение ссылок	726
Раздел приложений	728
13.2. Рекомендации и советы	728
Помните о своей аудитории	728
О чем писать в первую очередь, во вторую, в последнюю?	729
Другие источники	729
13.3. Пример: формула оперативного ценообразования для ответа на запросы потребителей	730
13.4. Дополнительный материал	736
Резюме	736
Основные термины	738
Контрольные вопросы	738
Задачи	739
Упражнение с использованием базы данных	741
Проект	741
Глава 14. Временные ряды: анализ изменений во времени	743
14.1. Обзор анализа временных рядов	744
14.2. Анализ трендов и сезонности	755
Тренд и циклический компонент: скользящее среднее	757
Сезонный индекс: среднее значение отношения к скользящему среднему отражает сезонное поведение	759
Поправка на сезон: деление ряда на сезонный индекс	763
Долгосрочный тренд и прогноз с поправкой на сезонные колебания: линия регрессии	766
Прогноз: тренд с учетом сезонности	770
14.3. Моделирование циклического поведения с помощью ARIMA-процессов Бокса–Дженкинса	771
Процесс случайного шума не обладает памятью: отправная точка	774
Процесс авторегрессии (AR) обладает памятью о своем прошлом	775
Процесс скользящего среднего (MA) имеет ограниченную память	776
Процесс авторегрессии и скользящего среднего (ARMA) сочетает в себе AR и MA	778
Чистый интегрированный (I) процесс помнит, где он находился, и затем движется случайно	782
Процесс авторегрессионного интегрированного скользящего среднего (ARIMA) помнит свои изменения	785

14.4. Дополнительный материал	787
Резюме	787
Основные термины	791
Контрольные вопросы	792
Задачи	794
Проекты	804
ЧАСТЬ V. МЕТОДЫ И ПРИМЕНЕНИЯ	807
Глава 15. Дисперсионный анализ: проверка различий для нескольких выборок и многое другое	808
15.1. Использование блочных диаграмм для одновременного представления нескольких выборок	810
15.2. F-тест определяет, значимо ли различаются средние	812
Данные и источники вариации	812
Допущения	813
Гипотезы	814
F-статистика	815
F-таблица	817
Результат F-теста	822
Результат вычислений с помощью компьютера: однофакторная ANOVA-таблица	823
15.3. Тест наименьшего значимого различия: какие пары различаются?	824
15.4. Более сложные планы дисперсионного анализа	827
Разнообразие — вот, что придает вкус жизни	828
Двухфакторный дисперсионный анализ	828
Три фактора и более	830
Ковариационный анализ, ANCOVA	830
Многомерный дисперсионный анализ (MANOVA)	830
Как читать таблицу ANOVA	830
15.5. Дополнительный материал	834
Резюме	834
Основные термины	836
Контрольные вопросы	837
Задачи	837
Упражнения с использованием базы данных	845
Проекты	846
Глава 16. Непараметрические методы: проверка гипотез для порядковых данных или данных, не подчиняющихся нормальному распределению	847
16.1. Проверка гипотезы о равенстве медианы некоторому заданному значению	849
Критерий знаков	850
Гипотезы	850
Допущение	851
16.2. Тестирование различий в двух связанных выборках	855
Использование критерия знаков для разностей	855
Гипотезы	856
Условие	857
16.3. Проверка значимости различия двух независимых выборок	858
Процедура, основанная на ранжировании <i>всех</i> данных	858
Гипотезы	859
Допущения	860
16.4. Дополнительный материал	864
Резюме	864
Основные термины	867

Контрольные вопросы	868
Задачи	869
Упражнения с использованием базы данных	876
Проекты	876
Глава 17. Анализ “хи-квадрат”: поиск закономерностей для качественных данных	878
17.1. Обобщение качественных данных с помощью частот и процентов	879
17.2. Проверка того, что значения процентов в генеральной совокупности равны некоторым заданным значениям	881
Критерий “хи-квадрат” в отношении равенства процентов	881
17.3. Проверка взаимосвязи между двумя качественными переменными	888
Понятие независимости переменных	888
Критерий “хи-квадрат” независимости	889
17.4. Дополнительный материал	895
Резюме	895
Основные термины	897
Контрольные вопросы	897
Задачи	898
Упражнения с использованием базы данных	905
Проекты	906
Глава 18. Контроль качества: выявление вариации и управление ею	908
18.1. Процессы и причины вариации	911
Диаграмма Парето показывает, на что обратить внимание	913
18.2. Что такое карты контроля и как их читать	915
Контрольные границы показывают выход из-под контроля одного наблюдения	916
Как выявить проблему даже в пределах контрольных границ	917
18.3. Отображение количественных измерений в $\bar{X}$ - и R -картах	919
18.4. Построение карт контроля для процента брака	926
18.5. Дополнительный материал	930
Резюме	930
Основные термины	932
Контрольные вопросы	932
Задачи	933
Проекты	944
Приложение А. База данных служащих	945
Приложение Б. Самопроверка: решение некоторых задач, а также упражнений, использующих базу данных	948
Приложение В. Статистические таблицы	976
Приложение Г. Краткое руководство по применению StatPad	1017
Словарь терминов	1025
Предметный указатель	1047