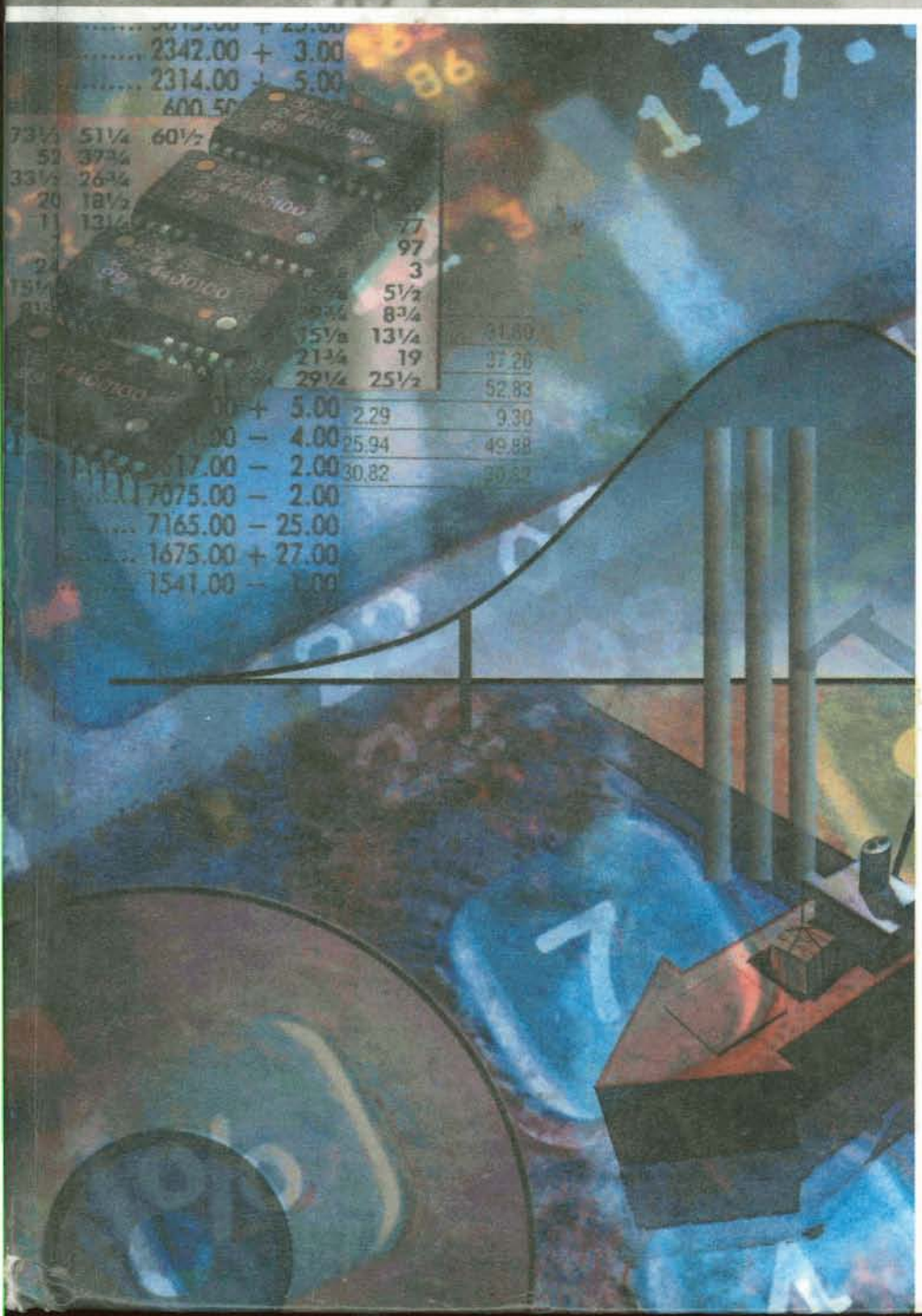
ПРАКТИЧЕСКАЯ БИЗНЕС-СТАТИСТИКА



ЧЕТВЕРТОЕ ИЗДАНИЕ

ЭНДРЮ Ф. СИГЕЛ

Практическая бизнес-статистика

Четвертое издание

Practical Business Statistics

FOURTH EDITION

ANDREW F. SIEGEL
Department of Management Science
Department of Finance
Department of Statistics
Department of Molecular Biotechnology
University of Washington

Irwin
McGraw-Hill

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogota Caracas Lisbon London Madrid
Mexico City Milan New Delhi Seoul Singapore Sydney Taipei Toronto

Практическая бизнес-статистика

Четвертое издание

ЭНДРЮ Ф. СИГЕЛ
Университет штата Вашингтон
Факультет научного менеджмента
Факультет финансов
Факультет статистики
Факультет молекулярной биотехнологии

ВИЛЬЯМС

Москва • Санкт-Петербург • Киев
2002

ББК 88.5я75
С34
УДК 681.3.07

Издательский дом "Вильямс"

Зав. редакцией *С.Н. Тригуб*

Перевод с английского *А.И. Мороза, О.Л. Пелявского, А.В. Редины* и *Е.Л. Усенко*

Под редакцией канд. экон. наук. *А.П. Горбачика*

По общим вопросам обращайтесь в Издательский дом "Вильямс" по адресу:
info@williamspublishing.com, http://www.williamspublishing.com

Сигел, Эндрю.

С34 Практическая бизнес-статистика. : Пер. с англ. — М. : Издательский дом "Вильямс", 2002. — 1056 с. : ил. — Парал. тит. англ.

ISBN 5-8459-0306-8 (рус.)

Эта книга представляет собой прекрасно организованный вводный курс статистических методов анализа данных. Дидактически грамотно представленный теоретический материал не перегружен математическими подробностями и дополняется большим количеством тщательно отобранных примеров. Здесь есть анализ финансового состояния предприятий и конъюнктуры фондового рынка, прогнозирование уровня продаж и результатов избирательных кампаний, анализ качества продукции и эффективности рекламы, изучение аудитории средств массовой информации и много других непростых и практически важных задач. Книга может быть полезна преподавателям, студентам, научным сотрудникам, аналитикам консалтинговых фирм и рекламных агентств, всем тем, кто занимается (или еще только собирается заняться) прикладным статистическим анализом эмпирических экономических и социальных данных.

ББК 88.5я75

Все названия программных продуктов являются зарегистрированными торговыми марками соответствующих фирм.

Никакая часть настоящего издания ни в каких целях не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами, будь то электронные или механические, включая фотокопирование и запись на магнитный носитель, если на это нет письменного разрешения издательства *Irvin, McGraw-Hill*.

Authorized translation from the English language edition published by Irvin, McGraw-Hill Companies, Copyright © 2000 Andrew F. Siegel

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the Publisher.

Russian language edition published by Williams Publishing House according to the Agreement with R&I Enterprises International, Copyright © 2002

ISBN 5-8459-0306-8 (рус.)
ISBN 0-07-117788-4 (*англ.*)

© Издательский дом "Вильямс", 2002
© Andrew F. Siegel, 2000

Оглавление

Предисловие 17

ЧАСТЬ I. ВВЕДЕНИЕ И ОПИСАТЕЛЬНАЯ СТАТИСТИКА 27

Глава 1. Введение: роль статистики в бизнесе 28

Глава 2. Структуры данных: классификация различных типов наборов данных 42

Глава 3. Гистограммы: взгляд на распределение данных 70

Глава 4. Обобщающие показатели: интерпретация типических значений и перцентилей 117

Глава 5. Изменчивость: изучение разнообразия 169

ЧАСТЬ II. ВЕРОЯТНОСТЬ 219

Глава 6. Вероятность: разбираемся в случайных ситуациях 220

Глава 7. Случайные величины: работа с неопределенными значениями 278

ЧАСТЬ III. СТАТИСТИЧЕСКИЙ ВЫВОД 339

Глава 8. Построение случайной выборки: предварительное планирование для сбора данных 340

Глава 9. Доверительные интервалы: допущение о неточности оценок 395

Глава 10. Проверка статистических гипотез: выбор между реальностью и совпадением 445

ЧАСТЬ IV. РЕГРЕССИЯ И ВРЕМЕННЫЕ РЯДЫ 517

Глава 11. Корреляция и регрессия: измерение и прогнозиро-вание взаимосвязей 518

Глава 12. Множественная регрессия: прогнозирование одного фактора на основе нескольких других 611

Глава 13. Составление отчетов: представление результатов множественной регрессии 720

Глава 14. Временные ряды: анализ изменений во времени 743

ЧАСТЬ V. МЕТОДЫ И ПРИМЕНЕНИЯ 807

Глава 15. Дисперсионный анализ: проверка различий для нескольких выборок и многое другое 808

Глава 16. Непараметрические методы: проверка гипотез для порядковых данных или данных, не подчиня-ющихся нормальному распределению 847

Глава 17. Анализ "хи-квадрат": поиск закономерностей для качественных данных 878
Глава 18. Контроль качества: выявление вариации и управление ею 908
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

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 *относительной* 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: *точное* 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
Не событие 238
Правило дополнения (правило *не*) 238
Одно событие *и* другое 239
Если два события не могут наблюдаться вместе 240
Правило пересечения (*и*) для несовместимых событий 240
Одно событие *или* другое 240
Правило объединения (*или*) для несовместимых событий 241
Нахождение *или* из *и* и наоборот 242

Одно событие *при условии* 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 (*и*) 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 *оценка* 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 прогнозиро-вание 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 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 ценообра-зования 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 подчиня-ющихся 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 *всех* 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

Предисловие

Традиционный курс бизнес-статистики изменился и в основном к лучшему: больше внимания уделяют интерпретации данных, понятиям и идеям, лучше демонстрируют взаимосвязь статистики с различными видами экономической деятельности, все делается для более глубокого понимания фундаментальных статистических принципов. Книга *Практическая бизнес-статистика* (*Practical Business Statistics*) является лидером этих перемен. В предисловии к первым трем изданиям говорилось следующее.

"Традиционный курс бизнес-статистики меняется. В настоящее время в связи с доступностью и широким использованием компьютеров для обработки числовой информации нет необходимости в подробном освещении многих старомодных тем. Это дало огромную возможность для подачи в учебное время нового материала, который необходим бизнес-менеджеру, – нового материала, который подводит фундамент под понятия и приложения статистики применительно к бизнесу и экономике. Например, менеджерам не обязательно знать, как вывести формулу для вычисления коэффициентов регрессии в методе наименьших квадратов, но они должны уметь интерпретировать коэффициенты регрессии, чтобы, опираясь на эти важные показатели, принимать верные решения в сложных ситуациях.

Практическая бизнес-статистика была написана с учетом этих изменений. Студенты экономических специальностей могут выучить статистику на "отлично", если изучение статистики сопровождать реальными практическими примерами и простыми и доступными объяснениями, из которых видно, почему стоит осваивать статистический взгляд на мир. Студенты-экономисты отличаются от студентов других специальностей и заслуживают книги, созданной специально в соответствии с их потребностями и интересами."

Большая работа: пересмотр материала и проверка в студенческой аудитории

В начале работы над этой книгой был сборник материалов для самостоятельного чтения, который я раздал своим студентам в дополнение к рекомендуемому учебнику. Все доступные книги представляли статистику излишне сложно, а я хотел представить прямые и легкие способы понимания предмета. Я также хотел придать предмету аромат современной коммерческой деятельности. Вся полезная информация, которую в качестве обратной связи я получил от студентов за все эти годы, была переработана и учтена в книге.

Даже перед публикацией первого издания *Практическая бизнес-статистика* прошла несколько стадий рецензирования и проверки в студенческой аудитории. Сейчас, когда в университетах и колледжах всей страны да и всего мира используют три предыдущих издания, подготовка четвертого издания дала мне возможность улучшить книгу с учетом всех полученных мною дополнительных рецензий и полезных положительных комментариев.

Стиль книги

Мне нравилось писать эту книгу. Везде, где это было возможно, *я показывал* на примерах "область применения" и объяснял, как мы, статистики, *действительно* размышляем о предмете, что это означает и насколько это полезно. Такой подход помог мне вдохнуть жизнь в предмет, который, к сожалению, имеет непривлекательный образ в глазах многих людей. Конечно, в книге даны также и традиционные объяснения, что позволяет иметь два взгляда на вещи: вот, что мы говорим, и вот, что это означает.

Я был взволнован, услышав, когда некоторые из моих студентов, которые просто боялись математики, сказали, что книгу действительно *приятно читать*! И это *после* того, как они получили свои оценки на экзамене!

Примеры

Примеры приближают статистику к жизни, показывают полезность и важность каждой темы. Книга *Практическая бизнес-статистика* содержит много реальных примеров, отобранных из широкого круга экономических источников. Фондовый рынок рассматривается в главе 5 для демонстрации рыночной изменчивости и риска, измеренных с помощью стандартного отклонения. Контроль качества рассматривается в разных главах для иллюстрации отдельных тем, а также в главе 18, которая посвящена именно этой теме. Опросы общественного мнения и предвыборные опросы также рассматриваются в разных главах (и особенно в главе 9), поскольку они представляют в чистом виде статистический вывод для реальной, хорошо знакомой всем нам, жизненной ситуации. Поиск данных в Internet рассмотрен в главе 2. На примере цен на рекламу в главе 12 показано, как множественная регрессия может выявить взаимосвязи в сложном наборе данных. Для демонстрации прогнозирования на основе временных рядов в главе 14 рассмотрены данные об уровне безработицы. Использование наглядных примеров значительно облегчает обучение студентов.

Статистические графики

Чтобы наглядно показать, что происходит с данными, *Практическая бизнес-статистика* включает более 200 рисунков, иллюстрирующих основные характеристики и зависимости. Эти графики точны, поскольку выполнены с помощью компьютера. Например, имеющая форму колокола кривая нормального распределения здесь представлена точно, в отличие от аналогичных кривых, изображенных в некоторых книгах, где эта форма непроизвольно улучшена художником. Но точность нельзя заменить ничем! Это помогает студентам лучше понять и запомнить статистические понятия.

Ситуации для анализа

В конце глав 3–12 помещены ситуационные задачи, которые демонстрируют полезность статистического мышления как неотъемлемой части более широкой коммерческой деятельности. Эти задачи часто не имеют единственно верного решения, но дают возможность для размышлений и открытых обсуждений.

Что нового в четвертом издании

Здесь подытожены основные изменения. Многие примеры и задачи взяты из предыдущих изданий, но теперь они основаны на более свежих данных. Книга проиллюстрирована копиями экрана Excel, которые показывают, как выполнять многие операции непосредственно с помощью электронной таблицы. В главе 2 рассматриваются новые источники данных, в частности Internet. Глобальная сеть Internet также часто предлагается в качестве средства поиска данных для проектов, представленных в конце каждой главы. Связанные с Internet задачи или источники данных помечены в книге специальной пиктограммой. Внося изменения, я старался сохранить те части книги из первых трех изданий, которые хорошо зарекомендовали себя среди студентов и преподавателей.

Структура книги

Студент всегда должен знать, *почему* предлагаемый материал является важным. Поэтому каждая часть книги начинается с краткого обсуждения рассматриваемого предмета и соответствующих глав. Каждая глава начинается с краткого обзора основной темы, который демонстрирует важность этой темы для бизнеса, прежде чем будут изложены детали и примеры.

Ключевые слова, основные термины и понятия выделены жирным шрифтом. Они собраны в списке основных понятий в конце каждой главы, а также включены в словарь терминов. Это облегчает их изучение и дает возможность сконцентрировать внимание на основных идеях. Предметный указатель поможет легко и быстро найти как главные темы, так и отдельные детали. Например, попытайтесь найти термины "корреляция", "временной ряд" или "доверительный интервал".

Каждая глава заканчивается разделом "Краткое содержание и задачи", который начинается с резюме содержания основного материала. Затем идет перечень основных терминов. Контрольные вопросы дают обзор основных тем, указывая также на их важность. Предложенные *задачи* дают студенту возможность применить статистику в новых ситуациях. "Упражнения с использованием базы данных" предлагают также ряд практических задач с использованием данных о служащих из приложения А. "Проекты" призваны приблизить статистику к потребностям и интересам студентов. Студенты могут самостоятельно ставить задачу и, руководствуясь своим опытом или интересами, подбирать данные из различных источников, включая Internet, различные публикации или отчеты компаний. И наконец, "Ситуации для анализа" (по одной в каждой из глав 3–12) дают возможность для размышлений и обсуждений, часто без единственно верного ответа.

Радсон, Висконсинский университет, штат Милуоки; Терренс Релли, Бэдсонский колледж; Питер Шухман, Ричмондский университет; Бала Шетти, Техасский университет, Л. Дуайт Снизен-мл., Аризонский университет; Тед Тзукахара, колледж Святой Марии; Эдвард А. Усил, Американский университет; Мишель Уэгманн, школа менеджмента Келлера; Мустафа Йилмаз, Северо-восточный университет; и Гейри Йошимото, государственный университет Святого Клода.

Мои слова признательности также всем тем, чей вклад в предыдущие издания помог принять этой книге завершенный вид: Сангиту Чаттерджи, Северо-восточный университет; Джей Девоур, Калифорнийский государственный политехнический университет; Берту Холланду, Темплийский университет; Уинстону Лину, государственный университет Нью-Йорка в Буффало; Герберту Спиреру, Коннектикутский университет; Дональду Уэстерфильду, Уэбстерский университет; Уэйну Уинстону, Индианский университет; Джеку Юркевичу, Пейсовский университет; Бетти Торн, Стетсоновский университет; Деннису Петруска, Янгстаунский государственный университет; Х. Кариму, университет Западного побережья; Мартину Янгу, Мичиганский университет; Ричарду Спинетто, университет Колорадо в Булстере; Полю Пашке, Орегонский государственный университет; Ларри Амманну, Техасский университет в Далласе; Дональду Марксу, университет Аляски; Кевину Нгу, университет Оттавы; Рахмату Таваллялли, Уэльский университет; Дэвиду Ауэру, Западно-Вашингтонский университет; Мюрею Коте, Техасский университет; Петеру Лакнеру, Нью-йоркский университет; Дональду Адольфсону, университет Бригхам Янга; А. Рахулджи Парса, Дрейковский университет.

Обращение к студентам

Когда вы начинаете этот курс, у вас может быть некоторое предвзятое мнение о том, что представляет собой статистика. Если ваше мнение положительное, сохраните его и поделитесь им с однокурсниками. Если же негативное — то отложите его в сторону и оставайтесь непредубежденным. Поскольку этот предмет часто не изучают в высших учебных заведениях многих стран, то относительно небольшое количество студентов развили в себе статистический взгляд на мир, взгляд, который помогает им справляться с горами данных при принятии решений в условиях неопределенности.

По нескольким причинам статистика легче для вашего поколения, чем для предыдущих. Сейчас, когда компьютеры могут делать за вас черновую работу по проведению расчетов данных, у вас есть время глубже изучить основные понятия статистики и осознать, как это поможет вам победить конкурентов в бизнесе.

Хорошо усвойте вводный материал, чтобы понять, почему статистику стоит изучать. Пользуйтесь расположенными в конце глав резюме, списками основных понятий и другими материалами. Не забывайте, что в конце книги приведены подробные решения задач, а также словарь терминов, который всегда поможет вам быстро вспомнить нужное слово. И не волнуйтесь. Если вы осознали необходимость статистики для вашего бизнеса, вещи, которые вы должны выучить, легко найдут свое место.

Храните книгу как справочник. Она вам пригодится, когда ваш босс попросит немедленно написать докладную записку, которая требует быстрого просмотра данных, или подготовить ответ на аналитический материал конкурентов. С помощью *Практической бизнес-статистики, которая будет стоять у вас в шкафу*, вы быстро закончите работу и пойдете на обед. *Приятного аппетита*!

Эндрю Ф. Сигел (Andrew F. Siegel)

Об авторе

Эндрю Ф. Сигел (Andrew F. Siegel) — профессор факультета менеджмента и финансов школы бизнеса университета штата Вашингтон, Сиэтл. Он также является адъюнкт-профессором факультета статистики и факультета молекулярной биотехнологии, имеет звание доктора философии по статистике Станфордского университета (1977 г.), магистра наук по математике Станфордского университета (1975 г.) и бакалавра по математике и физике (с отличием) Бостонского университета (1973 г.). До работы в Сиэтле он преподавал и занимался исследованиями в Гарвардском университете, университете штата Висконсин, в корпорации RAND, Смитсоновском институте и в Принстонском университете. Он также периодически читал лекции (как приглашенный профессор) в Бургундском университете в Дижоне и в Сорбонне, Франция. Впервые преподавая статистику в школе бизнеса (университет штата Вашингтон, 1983 г.), он был удостоен звания лучшего профессора семестра на основании опроса студентов MBA. В 1993 году он получил должность профессора, которая финансируется грантом Батербау; эта профессорская должность была учреждена в честь выдающегося профессора И. Батербау (I. Butterbaugh), преподавателя бизнес-статистики. Другие его награды и премии: премии Бурлингтонского Северного Фонда за выдающиеся достижения (1986 и 1992 гг.); член-корреспондент Центра по изучению фьючерских рынков, Колумбийский университет, 1988 г.; награды за успехи в преподавании (исполнительная программа MBA), университет штата Вашингтон, 1986 и 1988 гг.; награда Фонда Пита Марвика (Peat Marwick Foundation) за исследование возможностей аудита, 1987 г., и Phi Beta Kappa, 1973 г.

Эндрю Ф. Сигел является членом Американской статистической ассоциации (American Statistical Association), где занимал должность секретаря-казначея секции экономической и бизнес-статистики. Им написаны также книги *Статистика и анализ данных. Введение* (*Statistics and Data Analysis: An Introduction*, Wiley, 1996, совместно с Charles J. Morgan), *Контрпримеры в вероятности и статистике* (*Counterexamples in Probability and Statistics*, Wadsworth, 1986, совместно с Joseph P. Romano) и *Современный анализ данных* (*Modern Data Analysis*, Academic Press, 1982, совместное редактирование с Robert L. Launer). Его статьи опубликованы в следующих изданиях: *Journal of the American Statistical Association, Journal of Business, Management Science, Journal of Finance, Encyclopedia of Statistical Sciences, American Statistician, Journal Financial and Quantitative Analysis, American Mathematical Monthly, Journal of the Royal Statistical Society, Annals of Statistic, Annals of Probability, Society for Industrial and Applied Mathematics Journal on Scientific and Statistical Computing; Journal of*

Computational Biology, Genome Research, Biometrika, Auditing: A Journal of Practice and Theory, Contemporary Accounting Research; Journal of Futures Markets и Journal of Applied Probability. Он работал консультантом в различных прикладных областях, таких как прогнозирование результатов выборов для крупной телевизионной сети, статистические алгоритмы распознавания речи для известной исследовательской лаборатории, тестирование телевизионной рекламы для маркетинговой фирмы, методы контроля качества продукции поставщиков для крупной промышленной компании, эффективность биотехнологических процессов в крупной лаборатории, автоматизация проектирования и запуск в производство электронного оборудования в Силиконовой Долине, анализ диверсификации портфеля активов финансовой компании.

ЧАСТЬ I

Введение и описательная статистика

В этой части...

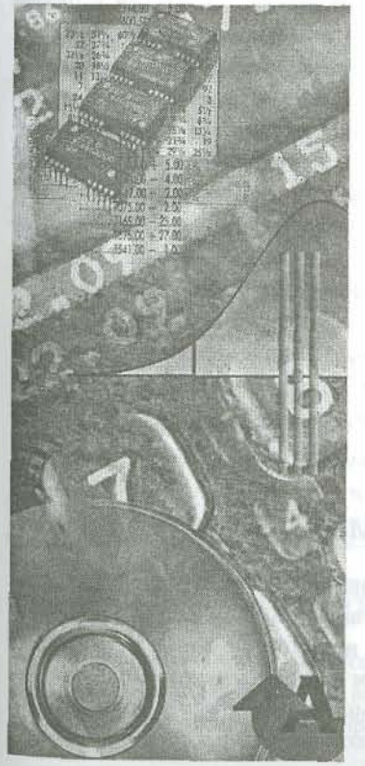
Глава 1. "Введение: роль статистики в бизнесе"

Глава 2. "Структуры данных: классификация различных типов наборов данных"

Глава 3. "Гистограммы: взгляд на распределение данных"

Глава 4. "Обобщающие показатели: интерпретация типических значений и перцентилей"

Глава 5. "Изменчивость: изучение разнообразия"

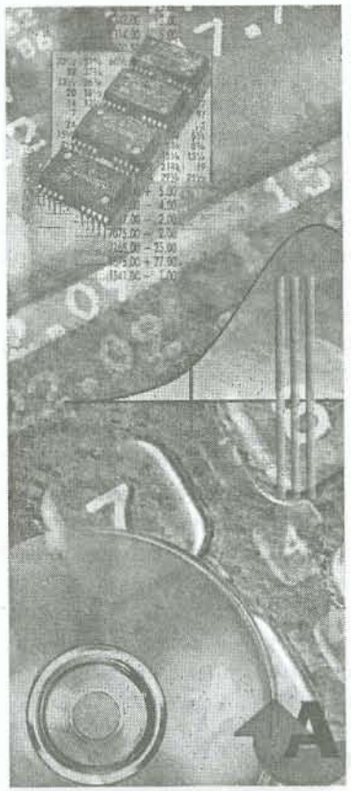


Добро пожаловать в мир статистики! Этот мир необходим вам, поскольку, когда вы располагаете информацией и знаете, как собрать необходимые дополнительные факты, вы сможете принять наилучшее управленческое решение. Как иначе можно руководить работой 12 отделов, 5809 служащими и производством 683 изделий? И даже для малого бизнеса необходимо понимать, что происходит в более широком бизнес-окружении, состоящем из потенциальных потребителей и конкурентов. Первые пять глав знакомят с ролью статистики в управлении экономической деятельностью (глава 1) и различными типами наборов данных (глава 2). Статистические показатели помогают увидеть "общую картину", которая иначе не проявилась бы в собранных данных. Глава 3 демонстрирует, как можно увидеть основные факты, касающиеся наборов чисел, с помощью *гистограмм*. Основные числовые показатели (среднее, медиана, перцентили и др.) представлены в главе 4. Одна из причин, по которой статистические методы имеют такое важное значение, заключается в том, что ситуации характеризуются *изменчивостью*, которая находит свое отражение в данных. В главе 5 рассказывается, как измерять степень такого разнообразия в данных.

ГЛАВА 1

Введение: роль статистики в бизнесе

Люди, занимающиеся бизнесом, вынуждены принимать решения в условиях постоянного давления обстоятельств, зачастую не имея полной и точной информации. Конечно же, любую доступную информацию следует использовать максимально полно. *Статистический анализ* помогает извлекать информацию из данных и оценивать качество этой информации. *Вероятность* помогает понять риски и случайности и обеспечивает оценки правдоподобности получения различных потенциальных результатов.



Даже те, кто считает, что принятие решения в бизнесе должно быть основано на интуиции и опыте специалистов (и, следовательно, не нуждается в излишней количественной информации), должны согласиться, что следует принимать во внимание всю имеющуюся ценную информацию. Таким образом, статистические методы следует рассматривать как важную часть процесса принятия решений, позволяющую выработать обоснованные стратегические решения, сочетающие интуицию специалиста с тщательным анализом имеющейся информации.

Мы начнем с обзора тех преимуществ, которые обеспечивает знание статистических методов, затем рассмотрим некоторые основные положения статистики и теории вероятности и их роль в экономической деятельности.

1.1. Почему именно статистика?

Действительно ли знание статистики так необходимо для успешного ведения бизнеса? Или достаточно полагаться на интуицию, опыт и догадку? Давайте сформулируем вопрос иначе: согласны ли вы игнорировать большую часть потенциально полезной информации, которая содержится в данных?

Почему необходимо изучать статистику?

Изучив статистику, вы станете более компетентно работать с данными и будете чувствовать себя гораздо увереннее в неопределенных ситуациях. Часто данные содержат много информации, которая не является очевидной, — статистика поможет извлечь и понять эту информацию. Нужна очень высокая квалификация, чтобы разрабатывать стратегию на основе знаний, опыта и интуиции. Если знания представлены в виде наборов чисел, статистика поможет ответить на такие вопросы. Насколько можно доверять этим числам и выводам из них? Как можно обобщить все эти данные? Используя статистику для получения знаний, вы приумножите свой опыт, что, безусловно, поможет принять правильное стратегическое решение.

Не следует пренебрегать статистикой. Методы статистики постоянно используются во всем мире, а снижение стоимости вычислительной техники увеличивает возможность принятия решений на основе количественной информации.

Сложна ли статистика?

Статистика не сложнее любой другой науки. Конечно, чтобы понять основные идеи и концепции, нужно хорошо поработать. Хотя некоторое внимание следует уделить отдельным деталям и вычислениям, все же намного легче стать опытным *пользователем* статистики, чем опытным статистиком, разбирающимся во всех мелочах. Сейчас стало легче использовать статистику, поскольку компьютеры теперь легко выполняют большое количество однотипных вычислений, что позволяет сконцентрироваться на содержательной интерпретации результатов. Возможно, некоторые непреклонные чистые статистики недовольны отсутствием изложения технических деталей при обучении статистическим методам, но сейчас уже хорошо видно, что эти детали заняли подобающее им место. Жизнь слишком коротка, чтобы человек посвящал ее разбору таких сложных технических приемов, как деление в столбик или обращение матрицы.

Влияет ли знание статистики на гибкость принятия решений?

Знание статистики помогает вам принимать хорошие решения. Статистика — не жесткая наука, она не отвергает опыт и интуицию. Получив знания о работе с данными и об основных свойствах неопределенных событий, вы улучшите информационную основу принятия решений и разовьете свою интуицию. Рассматривайте статистику как один из компонентов процесса принятия решений, но не как весь процесс. Статистика дополняет, но не заменяет деловой опыт, здравый смысл и интуицию.

1.2. Что такое статистика?

Статистика — это искусство и наука сбора и анализа данных. Поскольку *данными* называют любой вид зарегистрированной информации, статистика играет важную роль во всех сферах деятельности человека.

Статистика рассматривает общую картину

Если вы имеете большой и сложный набор данных, состоящий из многих небольших порций информации, статистика поможет классифицировать и проанализировать ситуацию, предоставив полезный обзор и резюме основных характеристик этих данных. Если данные пока отсутствуют, статистика поможет собрать их, обеспечивая получение ответов на ваши вопросы, а также то, что вы затратите не слишком большие усилия на выполнение этой работы.

Статистика не игнорирует отдельные объекты

При правильном подходе статистика уделяет необходимое внимание каждому из изучаемых объектов. Полный и тщательный статистический анализ обобщит основные свойства, которые характерны для каждого из объектов, а также *подготовит вас к интерпретации любых исключений*. Если данные содержат особые случаи, которые неадекватно отражены в "общей картине", то работа специалиста-статистика еще не завершена. Например, вы можете прочитать, что в 1996 году средний размер американской семьи составлял 2,65 человека[1]. Хотя это полезная статистическая информация, она не дает полную картину размера всех семей в США. Далее вы узнаете, что с помощью статистических методов можно легко описывать полное распределение размеров всех семей в США.

Пример. Данные в менеджменте

Данные очень часто используют в менеджменте. Ниже приведен краткий перечень видов ежедневно используемой в менеджменте информации (по сути, перечень используемых данных).

1. Финансовые отчеты (и другие виды бухгалтерской отчетности).
2. Курсы и объемы ценных бумаг, процентные ставки (и другая информация, относящаяся к инвестициям).
3. Состояние бюджета (и другие сообщения правительства).
4. Отчеты о продажах (и другие внутренние отчеты).
5. Результаты обзора состояния рынка (и другие маркетинговые отчеты).
6. Данные о качестве продукции (и другие производственные отчеты).
7. Отчеты о производительности рабочих (и другие внутренние данные фирмы).
8. Цена и объем проданной продукции (и другие данные о продажах).
9. Расходы на рекламу и результаты рекламной компании (и другая рекламная информация).

Подумайте об этом. Вероятно, большая часть вашей деятельности зависит, по крайней мере, косвенно от данных. Вероятно, кто-то работает для вас и консультирует вас по этим вопросам, но вы редко видите фактические данные. Время от времени вы могли бы попросить дать вам "сырые данные", чтобы иметь представление о перспективе. Просмотрев данные и задав вопросы, можно получить неожиданные результаты: можно обнаружить, что качество данных не так высоко, как вы думали (и у вас появляется мысль: на чем же мы строим наши прогнозы?), или, наоборот, вы получите обнадеживающие результаты и обретете уверенность. Другими словами, данные заслуживают внимания.

[1] *Statistical Abstracts of the United States*: 1997 (117 edition) Washington, DC, 1977, p. 59.

Посмотрим на данные

О чем вы думаете, глядя на таблицы с данными (например, на последние страницы *The Wall Street Journal*)? И что видит в них профессиональный статистик? Удивительно, но в обоих случаях ответ, как правило, один: не так уж и много. Нужно поработать с числами: нарисовать на бумаге графики, вычислить ряд характеристик и сделать еще много работы, прежде чем прояснится содержащаяся в этих числах информация. Именно этим занимаются профессиональные статистики. Они считают, что это проще и полезнее, чем подолгу подробно рассматривать длинные столбики чисел. Так что пусть вас не обескураживает, если столбик чисел выглядит для вас не более чем столбик чисел.

Статистика в менеджменте

Что должен знать о статистике менеджер? Вы должны иметь представление об основных понятиях статистики, включая некоторые (но не обязательно все) детали. Следует осознавать случайность и неопределенность многих аспектов окружающего нас мира. Более того, вы должны:

1. понимать и использовать в качестве базовой для вашей деятельности информации результаты статистического анализа;
2. принимать непосредственное участие в статистическом исследовании, если вы отвечаете за сбор и/или анализ данных.

Для этого не обязательно самому уметь выполнять сложный статистический анализ, но чтобы правильно интерпретировать полученные результаты, необходим некоторый опыт статистического анализа данных, который поможет вам также научить других глубже вникать в результаты. Более того, иногда удобнее выполнить определенный статистический анализ самостоятельно.

В последующих разделах мы уделим основное внимание идеям и понятиям статистики, сопровождая их для лучшего понимания практическими примерами.

1.3. Четыре основных этапа статистического анализа

На начальной стадии статистического анализа данные или не собраны, или еще даже не принято решение о том, какие данные следует изучать. Эти вопросы решают на этапе *планирования* таким образом, чтобы получить действительно полезную информацию. Когда данные есть в наличии, на этапе *исследования* проводится их первичный (предварительный) анализ. Следующий этап — *оценки* — позволяет получить на основе данных числовое значение неизвестной величины. Наконец, на последнем этапе — *проверке гипотез* — данные используются для принятия решения о соответствии выдвинутого предположения действительности. Рассмотрим все эти этапы по очереди.

Планирование сбора данных

Планирование сбора данных в маркетинговых исследованиях называют *планированием выборочного исследования*, а в изучении оптимизации химического производственного процесса — *планированием эксперимента*. Эта стадия пла-

нирования исследования включает составление подробного плана сбора данных. Тщательное составление плана поможет избежать лишних расходов и разочарования, если окажется (и будет уже слишком поздно), что собранные данные неадекватны основным поставленным вопросам. Разумный план также включает определение необходимого объема данных, достаточного для анализа, но не настолько большого, чтобы быть излишне расточительным. Таким образом, заранее составленный план удерживает стоимость проекта в разумных рамках и гарантирует, что стадия анализа будет протекать достаточно гладко.

Статистика особенно полезна тогда, когда есть большая группа людей, фирм или других объектов (*генеральная совокупность*), которая вас интересует, но вы не можете себе позволить провести полное исследование. Вместо этого, чтобы получить полезное, но неидеальное понимание ситуации в этой генеральной совокупности, можно отобрать небольшую группу (*выборку*), состоящую из некоторых (но не всех) объектов генеральной совокупности. Процесс обобщения результатов исследования выборки на всю совокупность называется *статистическим выводом*. *Случайная выборка* является одним из наилучших способов извлечь для подробного изучения выборку из генеральной совокупности, которая слишком велика, чтобы ее можно было изучать полностью[2]. Случайное извлечение выборки преследует две цели.

1. Гарантировать, что процесс извлечения выборки является беспристрастным, т.е. все объекты генеральной совокупности имеют равные шансы быть отобранными. Поэтому в среднем выборки являются представительными (репрезентативными) для данной генеральной совокупности (хотя каждая отдельная случайная выборка обычно является репрезентативной только приближенно, но не идеально).
2. Случайность, введенная контролируемым способом на стадии планирования проекта, гарантирует валидность (корректность) последующих статистических выводов.

Предварительное исследование данных

Как только вы получаете набор данных, вам хочется проверить его. В ходе **предварительного исследования** данные анализируются с разных точек зрения, описываются и обобщаются. Это позволяет убедиться, что данные представляют собой именно то, что необходимо, и нет никаких очевидных проблем[3]. Но хорошо выполненное предварительное исследование двояко готовит вас к проведению формального анализа.

1. Путем проверки, что ожидаемые связи действительно существуют в данных, и, таким образом, запланированные методы анализа адекватны данным.
2. Путем обнаружения в данных неожиданной структуры, которую необходимо принять во внимание, что предполагает внесение изменений в план анализа.

[2] Подробности построения случайной выборки изложены в главе 8.

[3] Предварительное исследование данных используется там, где это необходимо (и особенно в главах 3, 4, 11, 12 и 14).

Предварительное исследование является первой стадией. Часто недостаточно полагаться на формальный, автоматизированный анализ, который предполагает, что набор данных, вводимых в компьютер, "ведет себя хорошо". Всякий раз при возможности самостоятельно изучайте данные, чтобы убедиться, что все в порядке, т.е. нет больших ошибок и наблюдаемые в данных зависимости между параметрами соответствуют типу запланированного анализа. Эта стадия поможет внести в данные коррективы, выбрать соответствующий метод анализа и обосновать использование необходимых в дальнейшем статистических методов.

Оценка неизвестной величины

Оценка неизвестной величины представляет собой наиболее обоснованное, основанное на имеющихся данных, предположение о возможном значении. Поэтому желательно (а часто необходимо) оценивать те параметры, которые невозможно определить точно. Ниже приведено несколько примеров неизвестных величин для оценивания.

1. Объем продаж в следующем квартале.
2. Намерения правительства по изменению налоговых ставок.
3. Реакция населения Сиэтла на новый продукт.
4. Стоимость портфеля инвестиций в следующем году.
5. Изменение производительности при изменении стратегии.
6. Уровень брака в производственном процессе.
7. Победители следующих выборов.
8. Влияние продолжительного воздействия излучения экрана компьютера на здоровье.

Статистика может пролить свет на некоторые из этих ситуаций, предоставив хорошо обоснованное предположение исходя из надежных данных. Помните, что все статистические оценки являются только предположениями, а следовательно, часто бывают неточны. Однако они служат поставленным целям, если достаточно близки к соответствующим неизвестным величинам. Если известно, насколько (приближенно) точны эти оценки, то можно решить, в какой мере их стоит принимать во внимание.

Статистическая оценка также показывает величину неопределенности или ошибки в некотором предположении, рассчитанном для выборки, случайно взятой из большей по размеру генеральной совокупности. *Доверительный интервал* дает вероятное значение верхней и нижней границ оцениваемой неизвестной величины, что позволяет заявить: "Я не знаю точное значение неизвестной величины, но я достаточно уверен в том, что оно лежит между этими двумя числами".

Обычно вычисляют доверительные интервалы, поскольку они показывают, насколько надежной в действительности является оценка. Например, утверждение, что в следующем квартале объем продаж составит $11 300 000, содержит некоторую определенную информацию. Однако утверждение, что вы на 95% уверены в том, что в следующем квартале объем продаж будет находиться в пределах от $5 900 000 до $16 700 000, позволяет делать дополнительные и гораздо более глубокие выводы.

Доверительные интервалы представляют оценку в некоторой перспективе и позволяют избежать необходимости указывать одно число как точное значение, в то время как фактически это число точным не является[4].

Проверка гипотез

Проверка статистических гипотез заключается в использовании данных для осуществления выбора одной из двух (или более) различных возможностей при решении вопроса в неоднозначной ситуации. Проверка гипотезы исходя из данных дает определенное решение о том, какая из возможностей является верной. Процедура проверки гипотезы включает сбор данных, которые помогают осуществить выбор одной из возможностей, и использование статистического анализа для подтверждения принятого решения, если это решение не вытекает из простого беглого анализа данных[5].

Ниже приведено несколько примеров гипотез, которые можно было бы проверить с использованием данных.

1. Средний житель Нью-Йорка в следующем месяце планирует тратить на покупку вашего продукта по крайней мере $10.
2. Завтра на выборах вы победите.
3. Новое медицинское средство безопасно и эффективно.
4. Средство марки "X" эффективно стирает и отбеливает.
5. Ошибка в финансовом отчете меньше некоторой величины.
6. Исходя из прошлого опыта можно предсказать ситуацию на фондовом рынке.
7. Уровень производственного брака ниже, чем ожидают потребители продукции.

Обратите внимание, что каждая гипотеза сформулирована как определенное утверждение, которое может быть либо верным, либо неверным. Результатом проверки гипотезы является заключение о том, что данные либо подтверждают гипотезу, либо нет.

Часто статистические методы используют, чтобы решить, можно ли в качестве допустимой возможности рассматривать просто "чистую случайность". Например, если опрос 300 человек свидетельствует, что 53% планируют завтра голосовать за вас, можно ли сделать вывод, что выборы закончатся в вашу пользу? Несмотря на то что здесь можно поставить много вопросов, давайте сейчас отвлечемся от деталей, как, например, от реальной возможности того, что некоторые люди до завтра изменят свое мнение, и сконцентрируемся только на элементе случайности (обусловленной тем фактом, что невозможно опросить всех избирателей и узнать предпочтение каждого из них). В этом примере тщательный анализ покажет, что существует реальная возможность того, что меньше 50% избирателей предпочтут вас, а наблюдение значения 53% находится в пределах диапазона ожидаемого случайного изменения.

---

[4] Подробно доверительные интервалы изложены в главе 9 и используются в главах 9–15.

[5] Подробно процедура проверки статистических гипотез изложена в главе 10 и используется в главах 10–18.

Пример. *Статистический контроль качества*

Ваши производственные процессы несовершенны (как и у других фирм), и время от времени некоторое изделие необходимо или вторично переработать, или просто выбросить. Нужно сказать спасибо вашей группе контроля, которая делает все, чтобы недоброкачественное изделие не попало к потребителю. Однако контроль, обнаружение и решение этих проблем — все это стоит немалых денег. Вот почему многие фирмы начали использовать статистический контроль качества.

Упрощая ситуацию, будем считать, что ваша сборочная линия *контролируема*, если изготовленные изделия имеют стабильные показатели, которые удовлетворяют техническим требованиям. В противном случае, сборочная линия считается *неконтролируемой*. Статистические методы помогают наблюдать за производственным процессом таким образом, что можно сэкономить финансовые средства тремя способами: (1) снизить затраты на сбор информации; (2) быстро выявлять проблемы и, следовательно, минимизировать ущерб; (3) по возможности не вмешиваться в процесс тогда, когда в этом нет необходимости. Ниже в общих чертах описано, как в данной ситуации реализуются четыре этапа статистического анализа.

На стадии планирования следует решить, что и как часто следует измерять. Например, можно принять решение извлекать случайную выборку объемом 5 изделий из каждой партии объемом 500 изделий. Каждое изделие в выборке оценивают по внешнему виду, выявляя очевидный брак, а также измеряют длину и ширину изделия. Результат стадии планирования исследования представляет собой план раннего выявления проблем. План должен работать в *реальном времени*, чтобы проблемы можно было выявлять немедленно, а не на следующей неделе.

В ходе предварительного исследования данные наносят на *карты контроля качества* и изучают те конфигурации, которые вызывают тревогу. Правильно определив направление изменения данных, можно даже предсказать и установить проблему раньше, чем она приведет к производственным потерям!

Статистическая оценка обеспечивает менеджмент информацией о ходе производственного процесса. Если производственный процесс хорошо управляется в установленных границах, то можно даже поднять сортность продукции, а значит, и цену. Оценки качества текущей продукции необходимы для удовлетворения текущих заказов, а прогноз качества на будущее полезен для стратегического планирования и выработки ценовой политики.

Статистическую проверку гипотез можно использовать для ответа на важный вопрос: контролируется данный процесс или нет? Поскольку производственный процесс может быть большим, длительным и сложным, не всегда можно оценить его, посмотрев на работу части оборудования. Максимально используя статистическую информацию, содержащуюся в имеющихся данных, вы надеетесь достичь двух целей. Во-первых, вы хотите определить момент выхода системы из-под контроля, прежде чем уровень качества станет недопустимым. Во-вторых, вам хочется минимизировать "ложную тревогу", чтобы не тратить напрасно время и деньги на вмешательство в процесс тогда, когда он фактически является управляемым.

Пример. *Запуск нового продукта*

Решение вопроса об освоении производства нового продукта является одним из самых важных, принимаемых компанией, и для этого компании необходимо располагать большим количеством информации. Большую часть этой информации получают в результате статистических исследований. Например, маркетинговое исследование целевой группы потребителей можно было бы использовать, чтобы оценить, какое количество людей хотели бы приобрести новое изделие по каждой из предложенных цен. Прошлые данные о себестоимости аналогичных типов изделий можно было бы использовать для оценки себестоимости нового продукта. Анализ прошлого опыта освоения выпуска изделий, как удачный, так и неудачный, можно использовать как руководство при запуске нового продукта. Анализ статистических данных о национальных и международных фирмах, выпускающих аналогичный продукт, поможет вам оценить масштаб будущей конкуренции. Прежде чем вкладывать финансовые средства в несколько отобранных рекламных роликов, неплохо было бы проверить реакцию выборки возможных потребителей на отдельные виды рекламы.

Четыре основных этапа статистического анализа могут в такой ситуации быть реализованы различными способами. Поскольку генеральная совокупность потребителей слишком велика, чтобы можно было изучить ее полностью, можно *запланировать исследование* репрезентативной выборки из генеральной совокупности (например, чтобы определить количество потребителей, готовых приобрести новый продукт, или чтобы посмотреть реакцию потребителей на конкретный рекламный ролик). Если есть данные, их всегда можно подвергнуть *предварительному исследованию*, что позволит получить более ясное представление о ситуации (например, можно ли в предположении о сегментации рынка выделить определенные группы потребителей?), а кроме того, выполнить обычную проверку данных перед использованием других статистических процедур. Можно вычислять различные оценки, которые, например, показывают потенциальный размер рынка, возможный первоначальный уровень продаж и себестоимость продукции. И, наконец, можно *проверять различные гипотезы*, чтобы, например, подтвердить гипотезу о достаточно высоком интересе потребителя и таким образом оправдать продолжение проекта или чтобы проверить эффективность рекламных роликов и выбрать тот из них, который (не случайно) лучше других с точки зрения реакции потребителей.

1.4. Что такое вероятность

Вероятность — это средство для работы с риском и неопределенностью. **Вероятность** показывает возможность (или шанс) наступления в будущем каждого из различных потенциальных событий, рассчитанную на основании информации о некоторой ситуации. Например, можно допустить, что вы располагаете почти всей информацией об интересующей вас ситуации (т.е. известны все подробности осуществления процесса, который приводит к успеху, неудаче или некоторым потерям). Тогда можно вычислить вероятность получения различных результатов для разных стратегий, чтобы увидеть преимущество каждой из них.

Например, вы знаете, что шансы на успех международного проекта равны только 8% (т.е. вероятность равна 0,08), но вы предполагаете, что правительство сможет удержать инфляцию на низком уровне, и тогда шанс на успех возрастет до 35% (риск все еще высок, но все же лучше, чем 8%). Вероятность не подскажет вам, стоит ли инвестировать проект, но зато вы сможете трезво оценить ситуацию.

Ниже приведены примеры различных ситуаций, где для принятия решения необходимо вычислить или оценить значение вероятности.

1. При заданной структуре инвестиционного портфеля и определенных предположениях о работе финансовых рынков, чему равна вероятность получения прибыли в течение одного года? В течение 10 лет?
2. Насколько велика вероятность того, что завтра пойдет дождь? Чему равна вероятность того, что наступающая зима будет достаточно холодной, чтобы ваш бизнес, связанный с масляными радиаторами, был прибыльным?
3. Чему равна вероятность того, что иностранная держава (в которой находится ваш завод) будет втянута в гражданскую войну в течение двух следующих лет?
4. Чему равна вероятность того, что студент колледжа, прошедший анкетирование при устройстве на работу, в течение ближайших месяцев станет ценным работником?

Вероятность — это понятие, в некотором смысле обратное статистике. В то время как статистика помогает переходить от наблюдений к обобщениям относительно рассматриваемой ситуации, вероятность имеет обратную направленность — исходя из характеристики ситуации можно выяснить, какие данные вы скорее всего получите, и возможность получения каждого из вариантов таких данных. Эта обратная связь может быть графически изображена следующим образом.

$$\begin{pmatrix}\text{Характеристика}\\ \text{ситуации}\end{pmatrix} \xrightarrow{\text{Вероятность}} \begin{pmatrix}\text{Какое событие вероятнее}\\ \text{всего произойдет}\end{pmatrix}.$$

$$(\text{Какое событие произошло}) \xrightarrow{\text{Статистический вывод}} (\text{Характеристика ситуации}).$$

Вероятность всегда идет рядом со статистикой, обеспечивая прочный фундамент для статистического вывода. В условиях неопределенности нельзя знать точно, какое событие произойдет, всегда есть некоторая вероятность ошибки. Используя понятие вероятности, вы узнаете, как контролировать ошибку так, чтобы она наблюдалась не более чем в 5% или 1% случаев.

1.5. Общий совет

Статистические результаты должны допускать простое непосредственное объяснение (даже если соответствующая теория намного сложнее). Ниже приведено несколько общих советов.

1. Доверяйте своим суждениям, учитывайте здравый смысл.
2. Сохраняйте здоровый скептицизм.
3. Не дайте себя ввести в заблуждение с помощью на первый взгляд оригинального статистического анализа. Он может опираться на нереальные или неподходящие предположения.

Из-за большой гибкости, доступной аналитику на каждой стадии статистического анализа, один из самых важных факторов, который надо принять во внимание при оценке результатов статистического анализа, звучит так: *Кто это финансировал*? Помните, что аналитик много раз делает выбор — при определении проблемы, при планировании сбора данных, при выборе структуры или модели для анализа, при интерпретации результатов.

1.6. Дополнительный материал

Резюме

Статистика — это искусство и наука сбора и анализа данных. Статистические методы следует рассматривать как важную часть процесса принятия решений, позволяющую вырабатывать обоснованные стратегические решения, сочетающие интуицию специалиста с тщательным анализом имеющейся информации. Использование статистики становится все более важным преимуществом в конкуренции.