

Чарльз Уилан

Голая статистика

Самая
интересная
книга
о самой
скучной
науке



Charles Wheelan

Naked Statistics:

Stripping the Dread from the Data

W. W. Norton & Company

Чарльз Уилан

Голая статистика

**Самая интересная книга
о самой скучной науке**

Перевод с английского Ивана Веригина

Москва
Издательство «Манн, Иванов и Фербер»
2016

УДК 311.1

ББК 65.051

У36

Научный редактор Александр Минько

*Издано с разрешения Janklow & Nesbit Associates
и литературного агентства Prava I Perevodi*

Книга рекомендована к изданию Федором Царевым

Уилан, Чарльз

У36 Голая статистика. Самая интересная книга о самой скучной науке / Чарльз Уилан ; пер. с англ. И. Веригина ; [науч. ред. А. Минько]. — М. : Манн, Иванов и Фербер, 2016. — 352 с.

ISBN 978-5-00057-953-4

Статистика помогает принимать важные решения, находить скрытые взаимосвязи между явлениями, лучше понимать ситуацию в бизнесе и на рынке. Автор книги профессор Чарльз Уилан с юмором и блестящими наглядными примерами рассказывает о том, как это происходит.

Эта книга будет полезной для студентов, которые не любят и не понимают статистику, но хотят в ней разобраться; маркетологов, менеджеров и аналитиков, которые хотят понимать статистические показатели и анализировать данные; а также для всех, кому интересно, как устроена статистика

УДК 311.1

ББК 65.051

Все права защищены.

Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

Правовую поддержку издательства обеспечивает юридическая фирма «Вегас-Лекс»

VEGAS LEX

ISBN 978-5-00057-953-4

© Charles Wheelan, 2013

© Перевод на русский язык, издание на русском языке, оформление. ООО «Манн, Иванов и Фербер», 2016

Оглавление

ВВЕДЕНИЕ	9
<i>Почему я ненавижу вычисления, но обожаю статистику</i>	
ГЛАВА 1. В чем суть?	17
ГЛАВА 2. Описательная статистика.....	35
<i>Кто же все-таки лучший бейсболист всех времен и народов?</i>	
Приложение к главе 2.....	57
ГЛАВА 3. Дезориентирующее описание.....	61
<i>«Он — выдающаяся личность!» и другие истинные, но вводящие в заблуждение утверждения</i>	
ГЛАВА 4. Корреляция	89
<i>Откуда Netflix известно, какие фильмы мне нравятся?</i>	
Приложение к главе 4.....	97
ГЛАВА 5. Основы теории вероятностей.....	101
<i>Не покупайте расширенную гарантию для своего 99-долларового принтера</i>	
ГЛАВА 5½. Загадка Монти Холла	127
ГЛАВА 6. Проблемы с вероятностью	133
<i>Как самоуверенные знатоки математики едва не разрушили глобальную финансовую систему</i>	
ГЛАВА 7. Почему так важны данные	151
<i>«Мусор на входе — мусор на выходе»</i>	

ГЛАВА 8. Центральная предельная теорема.....	171
<i>Леброн Джеймс статистики</i>	
ГЛАВА 9. Статистические выводы	191
<i>Почему моему преподавателю статистики казалось, что я пытаюсь его обмануть</i>	
Приложение к главе 9.....	216
ГЛАВА 10. Опрос общественного мнения.....	221
<i>Откуда нам известно, что 64% американцев поддерживают смертную казнь (ошибка выборки $\pm 3\%$)</i>	
Приложение к главе 10.....	238
ГЛАВА 11. Регрессионный анализ.....	241
<i>Вошебный эликсир</i>	
Приложение к главе 11.....	268
ГЛАВА 12. Типичные регрессионные ошибки	273
<i>Важное предупреждение</i>	
ГЛАВА 13. Программы статистического оценивания	289
<i>Изменит ли вашу жизнь поступление в Гарвардский университет</i>	
ЗАКЛЮЧЕНИЕ	309
<i>Пять вопросов, на которые поможет ответить статистика</i>	
ПРИЛОЖЕНИЕ	327
<i>Статистическое программное обеспечение</i>	
ПРИМЕЧАНИЯ.....	331
ОТ АВТОРА	339
ПРЕДМЕТНО-ИМЕННОЙ УКАЗАТЕЛЬ	343

Посвящается Кэтрин

Введение

*Почему я ненавидел вычисления,
но обожал статистику*

Я всегда недолюбливал математику. Мне вообще не нравятся числа как таковые. На меня не производят впечатления заумные формулы, не имеющие реального практического применения. Но особенно, учась в средней школе, я не любил алгебру, по той простой причине, что никто так и не смог мне толком объяснить, почему я должен изучать ее. Как вычислить площадь под параболой? Кому это нужно?

Кстати, один из самых значимых моментов в моей жизни пришелся на время учебы в выпускном классе. Это было в конце первого семестра; я готовился к сдаче последнего экзамена, однако чувствовал, что шансов на высокий результат мало. (Должен сказать, что к тому времени меня уже приняли в колледж, в который я давно мечтал поступить, поэтому какая-либо мотивация особо усердствовать при подготовке к школьным экзаменам у меня отсутствовала.) Вытянув экзаменационный билет и взглянув на вопросы, я понял, что быть беде. Причем даже не потому, что я не знал правильных ответов, а потому, что я вообще не понимал, о чем идет речь. Я не впервые приходил на экзамены плохо подготовленным, но по крайней мере, как правило, знал, в каких вопросах «мелко плаваю». Однако на сей раз я, похоже, не знал почти ничего. Поломав какое-то время над вопросами экзаменационного билета голову и поняв, что катастрофа неизбежна, я подошел к столу, за которым сидела наша преподавательница (помню, ее звали Кэрол Смит). «Миссис Смит, — произнес я, — я вообще не понимаю, о чем говорится в моем экзаменационном билете».

Должен сказать, что я не нравился миссис Смит гораздо больше, чем она нравилась мне. Да, сейчас я могу сознаться, что иногда злоупотреблял своими правами председателя ученической ассоциации и планировал общешкольные собрания таким образом, чтобы время их проведения совпадало с уроками по началам анализа, которые вела миссис Смит (уроки

приходилось отменять). Да, мы с одноклассниками время от времени клали букет цветов на стол миссис Смит перед ее приходом в класс (предполагалось, что это были цветы от некоего «тайного обожателя») и буквально давились от смеха, наблюдая, как она, войдя в класс и заметив букет, ужасно смущалась и краснела. И еще: поступив в колледж, я сразу же перестал выполнять домашние задания по математике.

Поэтому, когда я подошел к миссис Смит и сообщил, что не понимаю вопросов в экзаменационном билете, она не посочувствовала мне. «Чарльз, — сказала она громко, обращаясь, по-видимому, не только ко мне, но и ко всем присутствующим в классе, — если бы вы работали в течение семестра и добросовестно готовились к экзамену, то вопросы не показались бы вам непонятными». Это был железный аргумент.

Я молча вернулся на место. Через несколько минут Брайан Арбеттер, гораздо лучше меня разбирающийся в математическом анализе, подошел к миссис Смит и что-то прошептал ей на ухо. Она что-то тихо ответила ему, а затем произошло нечто неожиданное. «Попрошу минутку внимания, — обратилась миссис Смит к классу. — Оказалось, что по ошибке я принесла на экзамен билеты для второго семестра». С момента начала экзамена прошло уже достаточно много времени, поэтому было решено прервать его и перенести на другой день.

Не могу описать эйфорию, охватившую меня тогда. Одним словом, все закончилось как нельзя лучше. Со временем я женился на замечательной девушке. У нас родилось трое детей. Я опубликовал несколько книг и побывал в таких местах, как Тадж-Махал и храмовый комплекс Ангкор-Ват. Тем не менее день, когда моя преподавательница математики понесла заслуженное наказание, остается одним из самых памятных в моей жизни. (То обстоятельство, что в тот день я чуть не провалил экзамен, не оказало существенного влияния на мою дальнейшую счастливую жизнь.)

Инцидент, случившийся на экзамене по математике, весьма красноречиво (но не до конца) иллюстрирует мои отношения с этим предметом. Что любопытно, к школьному курсу физики я не испытывал такой неприязни. Более того, физика мне нравилась, несмотря на то что она тоже относится к точным наукам и широко использует математический аппарат. Как это объяснить? *Дело в том, что физика гораздо ближе к жизни и практике, чем математика.* Я прекрасно помню, как учитель физики показывал нам во время ежегодного чемпионата США по бейсболу, как использовать базовую формулу

ускорения, чтобы оценить дальность хоумрана*. Это здорово, притом что у той же формулы есть множество других сфер применения.

Во время учебы в колледже одним из моих любимых предметов была теория вероятностей — опять же потому, что она позволяет лучше понять ряд интересных реальных ситуаций. Теперь я знаю, что моя неприязнь к математическому анализу, который мы изучали в старших классах школы, объясняется тем, что никто нам так и не растолковал, какое отношение этот предмет имеет к реальной жизни. Если вас не приводит в восхищение элегантность самих математических формул, — а меня, безусловно, нет, — то ничего, кроме смертельной скуки, они у вас не вызывают. Не исключаю, что в этом во многом виноваты наши школьные учителя, которые не сумели привить нам любовь к математике.

Теперь настало время поговорить собственно о статистике (в рассказе о которой не обойтись без теории вероятностей). Я обожаю статистику: ее можно использовать для объяснения очень многих вещей, от тестирования ДНК до бессмысленности участия в разного рода лотереях. Статистика способна помочь в выявлении факторов, связанных с такими недугами, как рак и заболевания сердца, а также в обнаружении манипуляций с проведением стандартизованных тестов. Благодаря ей вы даже можете выиграть некоторые игровые шоу. В детстве я любил смотреть знаменитую телепрограмму под названием *Let's Make a Deal* («Совершим сделку») с ее не менее знаменитым ведущим Монти Холлом. В конце каждого выпуска передачи участник, добравшийся до финала, становился вместе с Монти Холлом перед тремя большими дверьми — Дверью № 1, Дверью № 2 и Дверью № 3, — и Монти Холл объяснял ему, что за одной из них скрывается очень ценный приз — скажем, новый автомобиль, а за двумя другими — козел. Финалист должен был выбрать одну из дверей и получить то, что находилось за нею.

Вероятность того, что финалист выберет дверь, за которой скрывался самый ценный приз, составляла 1 к 3. Однако в игре *Let's Make a Deal* был предусмотрен интересный трюк, приводивший в восхищение статистиков и ставивший в тупик остальных. После того как финалист указывал на какую-то из трех дверей, Монти Холл открывал одну из двух оставшихся дверей, за которой всегда оказывался козел. Допустим, к примеру, что финалист выбрал

* Хоумран — удар в бейсболе, при котором мяч перелетает через все игровое поле; дает право совершить перебежку по всем базам и принести своей команде очко.
Прим. перев.

Дверь № 1. После этого Монти Холл открывал Дверь № 3 — за ней находился козел. При этом две другие двери — Дверь № 1 и Дверь № 2 — оставались закрытыми. Если ценный приз скрывался за Дверью № 1, то финалист становился победителем игры, если же за Дверью № 2, то считался проигравшим. Но далее ситуация становилась еще более интригующей: Монти Холл спрашивал у финалиста, не передумал ли он и не считает ли, что ценный приз находится не за Дверью № 1, а за Дверью № 2. Напоминаю, что к этому времени Дверь № 1 и Дверь № 2 остаются закрытыми, и единственная новая информация, которой располагает финалист, состоит в том, что за одной из них скрывается козел.

Следует ли финалисту отказаться от своего прежнего выбора и указать на Дверь № 2?

Отвечаю: да, следует. Почему? Объяснение найдете в главе 5½.

Парадокс статистики в том, что она вездесуща — начиная с так называемых средних показателей и заканчивая голосованием на выборах президента, — но при этом пользуется репутацией неинтересной и малопонятной. Многие книги и курсы по статистике перегружены математическими формулами и специальным жаргоном. Поверьте, все эти технические подробности важны и по-своему привлекательны, но для человека, который не страдает избытком интуиции и воображения, выглядят как абракадабра, способная вызвать исключительно отторжение. Если вы не понимаете, *зачем* изучать статистику, то лучше не беритесь. Именно поэтому в каждой главе книги я пытаюсь ответить на основной вопрос, который безуспешно задавал в школе своему преподавателю математики: *зачем все это нужно лично мне?*

Эта книга об интуиции. Я старался по возможности избегать употребления математических формул, уравнений и графиков, в тех же случаях, когда без них нельзя было обойтись, я преследовал четкую конкретную цель. Множество приведенных мною примеров призваны убедить вас в целесообразности изучения этой дисциплины. *Статистика может быть действительно интересной и по большей части не так сложна, как кажется поначалу.*

Идея написать эту книгу родилась через несколько лет после моей неудавшейся попытки постичь сущность математического анализа под чутким руководством миссис Смит. В магистратуре мне предстояло изучать экономику и политологию. Но прежде чем читать нам курс экономики, меня (что неудивительно) и большинство моих сокурсников направили в так называемый

математический лагерь, чтобы мы ликвидировали там свои многочисленные пробелы в познании этого предмета. На протяжении трех недель мы чуть ли не круглосуточно изучали математику в плохо проветриваемом полуподвальном помещении.

В какой-то из таких дней я как никогда был близок к тому, что принято называть прозрением. Преподаватель пытался объяснить нам условия, при которых сумма бесконечного ряда сходится к конечному числу. Постарайтесь следить за ходом моих рассуждений, а я попробую описать суть данной концепции. (Возможно, сейчас вы испытываете те же ощущения, что и я, сидя в душном полуподвальном помещении.) Бесконечный ряд представляет собой последовательность чисел, уходящую куда-то в... бесконечность, например $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots$. Многоточие означает, что эта последовательность продолжается до бесконечности.

На этом месте мы впали в ступор. Используя какое-то доказательство (какое именно, уже не помню), преподаватель пытался убедить нас, что хоть такая последовательность чисел и может продолжаться до бесконечности, тем не менее она все равно сойдется (приблизительно) к какому-то конечному числу. Один из моих одноклассников, Уилл Уоршоер, сильно в этом сомневался (собственно, как и я). Разве так бывает?

Затем меня осенило: мне показалось, я понял, что именно пытается втолковать нам преподаватель. Я повернулся к Уиллу и изложил ему версию, которая только что возникла у меня в голове.

Допустим, вы стали ровно в двух футах от стены. Теперь придвиньтесь к стене на половину этого расстояния (1 фут). В результате вы окажетесь в одном футе от стены.

Еще раз придвиньтесь к стене на половину оставшегося расстояния (6 дюймов, или $\frac{1}{2}$ фута). Находясь в 6 дюймах от стены, повторите описанные выше действия (придвиньтесь к стене на 3 дюйма, или $\frac{1}{4}$ фута). Выполните их еще раз (придвиньтесь к стене на $1\frac{1}{2}$ дюйма, или $\frac{1}{8}$ фута). И так далее.

Постепенно вы почти упретесь в стену. (Например, окажетесь на расстоянии $1/1024$ дюйма от нее, а затем придвинетесь еще на половину этого пути, или на $1/2048$ дюйма.) Но ключевым здесь является слово *почти*: сколько бы раз вы ни повторяли это действие, расстояние между вами и стеной никогда не станет в точности равно нулю, поскольку, по определению, каждое такое продвижение приближает вас к стене лишь на половину оставшегося расстояния. Иными словами, вы все время будете оказываться бесконечно

близко к стене, но никогда не упретесь в нее. Если измерять ваши продвижения в футах, то соответствующую последовательность можно описать как $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \dots$

Именно в этом и заключалось мое прозрение. Сколько бы вы ни продвигались таким способом к стене (а вы будете делать это до бесконечности), совокупное расстояние, пройденное вами, не может превышать 2 футов, то есть вашего исходного расстояния от стены. С математической точки зрения, совокупное расстояние, пройденное вами, можно приравнять к 2 футам, что весьма удобно в плане вычислений. Математик сказал бы, что сумма бесконечного ряда $1 \text{ фут} + \frac{1}{2} \text{ фута} + \frac{1}{4} \text{ фута} + \frac{1}{8} \text{ фута} \dots$ сходится к 2 футам, то есть именно то, что пытался объяснить нам преподаватель.

Что показательно, в процессе объяснения мне удалось убедить в правильности моей версии не только Уилла, но и самого себя. Я уже не помню дословно математического доказательства того, что сумма бесконечного ряда при определенных условиях может сходиться к конечному числу (хотя могу найти его в соответствующем учебнике по математике), но исходя из собственного опыта готов утверждать, что благодаря интуиции математика и другие технические детали становятся гораздо понятнее (но необязательно наоборот).

Задача этой книги — доходчиво объяснить самые важные статистические концепции не только тем, кому приходится осваивать их в плохо проветриваемых, душных помещениях, но и тем, кого влечет магия чисел.

Хотя выше я был вынужден признать, что базовые инструменты статистики, к сожалению, менее интуитивно понятны и доступны, чем следовало бы, сейчас я намерен сделать несколько на первый взгляд противоречащее этому заявление, а именно: статистика может быть *более чем доступной для понимания* в том смысле, что каждый из нас, вооружившись исходными данными и компьютером, способен выполнить сложные статистические выкладки, нажав буквально несколько клавиш. Однако в случае, если исходных данных недостаточно или статистические методы используются некорректно, появляется риск, что наши выводы не только могут ввести нас в заблуждение, но и оказаться потенциально опасными. Рассмотрим следующую гипотетическую новость из интернета: *«Люди, которые делают короткие перерывы в работе в течение дня, имеют гораздо большие шансы умереть от рака»*. Представьте появление на экране такого сообщения,

когда вы занимаетесь веб-серфингом. Согласно весьма впечатляющим результатам обследования 36 000 работников (огромный массив данных, не правда ли?!), у тех, кто выходил из офиса на регулярные десятиминутные перерывы в течение каждого рабочего дня, вероятность заболевания раком в последующие пять лет оказалась на 41% выше, чем у тех, кто офисы не покидал. Понятно, что узнав такую новость, мы обязаны как-то на нее реагировать: возможно, провести общенациональную кампанию за запрет коротких перерывов в течение рабочего дня.

А может, следует подойти к проблеме с другой стороны и задуматься над тем, чем именно обычно занимаются работники во время таких десятиминут? Не мне вам рассказывать, что многие кучкуются неподалеку от входа в офисное помещение, покуривая сигареты (и создавая при этом облако дыма, через которое вынуждены проходить те, кто входит или выходит из здания). Смеею предположить, что именно сигареты, а не кратковременные перерывы в работе, являются основной причиной раковых заболеваний. Большинству читателей этот пример покажется абсурдным, но могу вас заверить, что многие статистические умозаключения, встречающиеся в реальной жизни, оказываются не менее абсурдными после их тщательного анализа.

Статистика подобна мощному оружию, полезному в случае его правильного применения и потенциально разрушительному в неумелых руках. Прочитав эту книгу, вы, конечно, не станете профессиональным статистиком, но по крайней мере она научит вас осторожному обращению со статистическими данными и убережет от их неверной интерпретации, которая может иметь непредсказуемые последствия.

Книга, которую вы держите в руках, — не учебник, и это обеспечило мне достаточно высокую степень свободы в выборе тем и способов изложения материала. *Цель этой книги — ознакомить читателей со статистическими концепциями в их непосредственной связи с повседневной жизнью.* Как ученые приходят к выводу о том, что некий фактор служит причиной раковых заболеваний? Каков механизм опросов общественного мнения (и что может исказить их результаты)? Кто «лжет, манипулируя статистическими данными», и как им это удастся? Как компания, выпустившая вашу кредитную карточку, использует информацию о совершаемых вами покупках, чтобы прогнозировать вероятность пропуска вами платежа? (Да-да, они и такое умеют!)

Если вы хотите правильно интерпретировать числа, озвученные в новостях, и использовать необычайную (и все более возрастающую) силу данных, то материал этой книги — именно то, что вам нужно. В конечном счете я надеюсь убедить вас в справедливости мысли, высказанной шведским математиком и писателем Андерсом Дункельсом: «Опираясь на статистику, легко лгать, но без статистики очень трудно выяснить истину».

Но я мечтаю о большем. Мне хочется, чтобы вы начали получать наслаждение от статистики. Идеи, положенные в ее основу, чрезвычайно интересны и актуальны. Главное — уметь отделять по-настоящему важные идеи от технических подробностей, которые способны стать для вас непреодолимым препятствием. Этому я и стараюсь вас научить на страницах данной книги.

Я заметил один любопытный феномен. Хотя студенты часто жалуются, что статистика — неинтересная и малопонятная наука, тем не менее, выйдя из аудитории, они охотно обсуждают свои спортивные достижения и средние результаты, которых добились летом, или коэффициент изменчивости погоды (в холодное время года), или свои баллы в колледже (этот вопрос не волнует их только во время каникул). Они признают, что «рейтинг распасовщика» — статистический показатель, выражающий в одном числе эффективность действий куортербека*, — весьма некорректно отражает качество его игры. Те же самые исходные данные (коэффициент удачного завершения, среднее число ярдов на каждую попытку паса, процент тачдаун-пасов** на каждую попытку паса и коэффициент перехватов мяча) можно было бы скомбинировать как-то по-другому, например присвоить каждой составляющей определенный весовой коэффициент и в результате создать другой, не менее надежный показатель эффективности действий куортербека. Однако все, кто интересуется американским футболом, должны признать, что наличие рейтинга распасовщика весьма удобно.

Является ли данный рейтинг идеальным? Разумеется нет. Статистика крайне редко предлагает единственно верный вариант оценивания чего бы то ни было. Предоставляет ли данный показатель возможность получить важную информацию? Разумеется да. Это превосходный инструмент, позволяющий быстро сравнивать эффективность действий двух куортербеков в один и тот же день. Я болею за команду Chicago Bears. Во время серии плей-офф 2011 года Chicago Bears играли с Packers (Packers одержали

* Куортербек — распасовщик, играющий помощник тренера в американском футболе. *Прим. перев.*

** Тачдаун — в американском футболе: пересечение мячом или игроком с мячом линии зачетного поля соперника. *Прим. перев.*

победу). Я мог бы описать этот матч множеством способов, потратив не одну страницу на его анализ. Но вот более сжатый вариант: рейтинг распасовщика куортербека Chicago Bears Джея Катлера составил в тот день 31,8, а куортербека Green Bay Аарона Роджерса — 55,4. Аналогично мы можем сравнить эффективность действий Джея Катлера с эффективностью его же действий в одной из предыдущих игр того же сезона против команды Green Bay, когда его рейтинг распасовщика равнялся 85,6. Эти показатели способны многое сказать тому, кто хочет понять, почему ранее в том сезоне Chicago Bears выигрывали у Packers, а затем потерпели поражение в серии плей-офф.

Это может служить весьма поучительным — и достаточно лаконичным — объяснением итогов футбольного сезона 2011 года. Однако нет ли здесь чрезмерного упрощения? Да, именно в этом и заключается сила и слабость любой описательной статистики. Один-единственный показатель говорит вам, что Джей Катлер продемонстрировал в играх плей-офф с участием Chicago Bears худшую эффективность, чем Аарон Роджерс. С другой стороны, тот же показатель ничего не скажет вам о том, потерпел ли тот или иной куортербек в ходе игры досадную неудачу (например, его идеальная передача не была поймана принимающим, а затем перехвачена), удавалось ли ему действовать с максимальной отдачей в определяющих с точки зрения конечного результата ключевых розыгрышах (поскольку весовые коэффициенты всех розыгрышей одинаковы и не зависят от их важности для конечного результата), насколько успешно действовала защита и т. д.

Парадоксально, что те же люди, которые свободно рассуждают о статистике в контексте спорта, погоды или академической успеваемости, начинают теряться, когда исследователь переходит к объяснению чего-нибудь наподобие коэффициента Джини — стандартного инструмента в экономике, демонстрирующего степень неравенства доходов. Ниже я объясню суть данного коэффициента, сейчас же для нас главное — *признать, что между коэффициентом Джини и рейтингом распасовщика нет принципиальных отличий*. Оба позволяют представить сложную информацию в виде единственного числового показателя. Как таковой коэффициент Джини обладает достоинствами большинства описательных статистик, а именно: обеспечивает удобный способ сравнения распределения дохода в двух странах или в одной стране в разные моменты времени.

Коэффициент Джини помогает оценить по шкале от 0 до 1, насколько равномерно распределяется в стране совокупный доход. Этот статисти-

ческий показатель можно вычислить для материального благосостояния или годового дохода, причем он может быть рассчитан на индивидуальном или семейном уровне. (Все эти значения будут сильно коррелированы, но не идентичны.) У коэффициента Джини, подобно рейтингу распасовщика, нет какого-либо собственного, внутренне присущего ему смысла — это всего лишь инструмент для сравнения. У страны, в которой все семьи имеют одинаковый уровень благосостояния, был бы нулевой коэффициент Джини. А в той стране, где все богатство сосредоточено в руках одной семьи, он равнялся бы единице. Как вы, наверное, догадались, чем ближе значение к единице, тем выше степень расслоения общества. Согласно данным Центрального разведывательного управления (между прочим, ЦРУ активно занимается сбором статистических данных)¹, коэффициент Джини для Соединенных Штатов равен 0,45. И что?

Если этот показатель поместить в определенный контекст, он может многое нам рассказать. Например, коэффициент Джини для Швеции составляет 0,23; для Канады — 0,32; для Китая — 0,42; для Южной Африки 0,65*. Анализ этих значений позволяет получить представление о том, какое место в мире занимают Соединенные Штаты с точки зрения неравенства распределения доходов. Можно также проанализировать, как коэффициент Джини изменяется со временем в одной и той же стране. Например, в 1997 году для Соединенных Штатов он равнялся 0,41, а в следующем десятилетии достиг 0,45 (самые последние данные ЦРУ относятся к 2007 году). Это дает возможность составить объективную картину нарастания неравенства в распределении богатства по мере процветания Соединенных Штатов (во всяком случае на рассматриваемом отрезке времени). Кроме того, мы можем сравнить изменения коэффициента Джини в разных странах примерно за один и тот же период времени. Скажем, в Канаде за указанный период он практически остался прежним. Швеция на протяжении двух последних десятилетий переживала фазу значительного экономического роста, однако коэффициент Джини в ней фактически снизился с 0,25 в 1992 году до 0,23 в 2005-м; это означает, что за указанный период Швеция не только стала богаче, но и доходы в ней начали распределяться более равномерно.

Можно ли считать коэффициент Джини идеальным показателем неравенства? Отнюдь нет — точно так же как рейтинг распасовщика нельзя считать

* Коэффициент Джини иногда умножают на 100, чтобы он выражался целым числом. В таком случае для Соединенных Штатов он равнялся бы 45.

идеальным показателем эффективности действий куортербека. Но несомненно одно: он позволяет нам получить весьма ценную информацию о социально значимом явлении — неравенстве в распределении богатства — в достаточно удобном формате.

Итак, мы медленно продвигаемся к получению ответа на вопрос, поставленный в названии этой главы: в чем суть? А в том, что статистика помогает нам обрабатывать *данные*, хотя на самом деле это всего лишь еще одно название информации. Подчас эти данные тривиальны, как в случае спортивной статистики, а подчас проливают свет на природу человеческого общества, как в случае коэффициента Джини.

Но, как любят повторять в телевизионных рекламных роликах, *это еще не все!* Хол Вариян, главный экономист компании Google, в интервью The New York Times сказал, что в следующем десятилетии работа со статистическими данными станет «модной профессией», а точнее «сексуальной» (дословное выражение Хола Вариана: *the sexy job*)². Я, наверное, окажусь первым, кто пришел к выводу о весьма превратном представлении некоторых экономистов о том, *что* следует считать «сексуальным». Тем не менее предлагаю рассмотреть несколько никак не связанных между собой вопросов.

- Как учесть учебные заведения в подтасовке результатов стандартизированных тестов?
- Откуда Netflix* известно о том, какого рода фильмы вам нравятся?
- Как определить, какие вещества и образ жизни вызывают раковые заболевания, учитывая, что мы не можем проводить над людьми экспериментов, приводящих к заболеванию раком?
- Можно ли рассчитывать на более успешный исход хирургической операции, если молиться за пациента?
- Существует ли реальная экономическая выгода в получении диплома какого-либо из престижных колледжей или университетов?
- Что является причиной роста заболеваемости аутизмом?

Статистика способна помочь нам (или, как мы рассчитываем, поможет в ближайшем будущем) получить ответы на эти вопросы.

* Netflix — американская компания, поставщик фильмов и сериалов на основе потокового мультимедиа. *Прим. перев.*

Наш мир все быстрее и быстрее генерирует все бóльшие и бóльшие объемы данных. Тем не менее, как справедливо отметила The New York Times, «данные — всего лишь исходный материал знаний»^{3*}. Статистика — самый мощный из имеющихся в нашем распоряжении инструментов для практического использования информации, например для оценивания эффективности действий бейсболистов или более справедливой оплаты труда преподавателей. Ниже приведен краткий обзор того, как статистика способна придать смысл исходным данным.

Описание и сравнение

Счет партии в боулинг является описательной (дескриптивной) статистикой. То же можно сказать и о каком-либо среднем показателе (например, в спорте). Большинство американских спортивных болельщиков в возрасте старше пяти лет неплохо разбираются в описательной статистике. Мы используем численные показатели в спорте и других сферах жизни для подытоживания информации. Насколько Микки Мэнтл был хорош как бейсболист? Его итоговый рейтинг как хиттера составил 0,298. Для бейсбольных болельщиков это весьма красноречивое число. Итоговый рейтинг 0,298 — выдающийся показатель, если принять во внимание, что в нем учитываются результаты Микки Мэнтла за восемнадцать лет карьеры профессионального бейсболиста⁴. (Хотя, согласитесь, если итог жизни человека можно выразить единственным числом, это несколько разочаровывает и настраивает на мысли о бренности человеческого бытия.) Разумеется, фанаты бейсбола должны помнить о существовании другой описательной статистики, которая, возможно, отражает ценность того или иного бейсболиста гораздо лучше, чем пресловутый средний показатель.

* Исторически так сложилось, что слово «данные» (data) используется во множественном числе (например, «эти данные являются весьма обнадеживающими»). Это слово можно употреблять и в единственном числе: «данное» (datum); в этом случае речь идет о каком-то отдельно взятом элементе данных (например, ответ одного человека на какой-то один вопрос анкеты, используемой при опросе общественного мнения). Употребление слова «данные» во множественном числе сигнализирует каждому, кто занимается серьезными исследованиями, о том, что вы знаете толк в статистике. С учетом сказанного многие специалисты по грамматике, а также многие издания, такие как The New York Times, в настоящее время согласны с тем, что слово «данные» может означать как единственное, так и множественное число, как свидетельствует приведенная мной цитата из The New York Times.

Академическая успеваемость учащихся школ и колледжей в США оценивается с помощью среднего балла. В стране используется шкала с буквенными обозначениями, где каждой букве соответствует определенный балл: как правило, A = 4 балла, B = 3 балла, C = 2 балла и т. д. По окончании учебного заведения, когда абитуриенты поступают в колледжи, а выпускники колледжей подыскивают себе работу, средний балл становится удобным инструментом для оценивания их академического потенциала. Тот, у кого средний балл 3,7, явно сильнее выпускника со средним баллом 2,5. Таким образом, средний балл является весьма полезной описательной статистикой. Его легко вычислить, понять и сравнивать с баллами других учащихся.

Тем не менее данный показатель не идеален. В нем не учитывается сложность учебных программ, которые проходят разные ученики. Как можно сравнивать знания учащегося со средним баллом 3,4, обучавшегося по относительно легкой программе, и его сверстника со средним баллом 2,5, изучавшего математику, физику, химию и другие сложные предметы? В свое время я посещал школу, которая пыталась решить эту проблему, присваивая таким дисциплинам дополнительные весовые коэффициенты, в результате чего оценка A по предмету повышенной трудности соответствовала пяти баллам, а по обычному предмету приравнивалась к четырем. Однако у данного подхода были существенные минусы. Моя мать довольно быстро уяснила, как эта «поправка» влияет на средний балл. Дело в том, что для таких учеников, как я (изучавших много сложных предметов), максимальная оценка A по любому из обычных предметов (например, по физкультуре или основам безопасности жизнедеятельности) не могла превышать 4 баллов, что снижало средний балл, как бы хорошо мы ни учились. В результате родители запретили мне посещать в школе курсы вождения автомобиля, поскольку даже самые высокие оценки по этому курсу уменьшали мои шансы на поступление в какой-либо престижный колледж и последующие занятия писательским трудом. Поэтому они отправили меня в частную (платную) школу вождения, которую мне пришлось посещать летом.

Глупость? Конечно! Но одной из тем, которые я затрону в этой книге, будет опасность чрезмерного увлечения любой из описательных статистик, поскольку это может привести к ошибочным умозаключениям и подтолкнуть к нежелательным действиям. В первоначальном варианте книги я использовал выражение «упрощенная описательная статистика», однако в конечном счете выбросил слово «упрощенная», поскольку оно показалось

мне заведомо избыточным. Описательная статистика для того и существует, чтобы упрощать, что всегда подразумевает некоторую потерю нюансов и деталей. Каждый, кто работает с числами, должен воспринимать это как данность.

Умозаключения

Сколько бездомных живет на улицах Чикаго? Как часто женатые пары занимаются сексом? На первый взгляд у этих вопросов нет ничего общего. На самом же деле на каждый из них можно ответить (правда, не с абсолютной точностью) с помощью базовых статистических инструментов. Одна из ключевых функций статистики — использование имеющихся данных для выдвижения аргументированных предположений, касающихся вопросов, исчерпывающий ответ на которые невозможно дать из-за отсутствия полной информации. Короче говоря, мы можем использовать данные из «известного мира» для построения обоснованных гипотез относительно «неизвестного мира».

Начнем с вопроса о бездомных. Точно подсчитать их количество в крупном мегаполисе и дорого, и затруднительно. Тем не менее располагать численной оценкой этой группы населения необходимо с целью предоставления социальных услуг, обоснования права на получение части доходов штата и федеральных доходов и соответствующего представительства в Конгрессе. Одним из важных статистических методов является выборочное исследование — процесс сбора данных по какой-то небольшой области, например нескольких районов, где проводилась перепись населения, чтобы на их основе сделать умозаключение о количестве бездомных в городе в целом. Такой подход требует значительно меньших ресурсов, чем попытка сосчитать всех бездомных; к тому же при правильном проведении выборочного исследования можно получить очень близкий к точному результат.

Опрос общественного мнения — еще одна форма статистической выборки. Скажем, исследовательская организация опрашивает членов среднестатистических семей, чтобы выяснить их точку зрения на ту или иную проблему или их мнение о том или ином политическом деятеле. Сделать это, естественно, гораздо проще, дешевле и быстрее, чем обойти все домохозяйства в соответствующем штате или стране в целом. По расчетам Американского института общественного мнения (Институт Гэллага), методологически

правильный опрос 1000 семей дает практически такие же результаты, как и опрос *всех* семей в Соединенных Штатах.

Именно таким способом нам удалось выяснить, как часто, с кем и как американцы занимаются сексом. В середине 1990-х годов Национальный центр изучения общественного мнения при Чикагском университете провел масштабное исследование сексуального поведения населения страны. Результаты основывались на детальных опросах крупной репрезентативной выборки взрослых американцев. Если вы продолжите чтение этой книги, то в главе 10 узнаете подробности. *В каких еще книгах, посвященных статистике, вы могли бы почерпнуть подобные сведения?*

Оценивание риска и событий, имеющих вероятностный характер

Казино *никогда* не бывают внакладе в долгосрочной перспективе. Это не означает, что они зарабатывают деньги в *любой момент*, но в конечном итоге остаются прибыльными, как бы ни складывалась каждая отдельно взятая игра. Весь игорный бизнес построен на азартных играх, поэтому исход каждой из них непредсказуем. В то же время базовые вероятности наступления соответствующих событий — выпадения двадцати одного очка в блек-джек или zero при игре в рулетку — известны. И когда эти базовые вероятности выступают в пользу казино (а это происходит *всегда*), можно не сомневаться, что по мере увеличения количества ставок вероятность того, что истинным победителем окажется игорное заведение, повышается, несмотря на мелкие «досадные недоразумения», случающиеся по ходу дела.

Данный феномен характерен не только для казино, но и для многих других сфер нашей жизни. Компаниям постоянно приходится оценивать риски, связанные со всевозможными неблагоприятными факторами. Полностью исключить такие риски невозможно — точно так же как казино не может гарантировать, что, сделав ставку, вы не сорвете крупный куш, доставив тем самым владельцам заведения немалое огорчение. Однако любой бизнес, сталкивающийся с неопределенностью, может управлять рисками, организовав соответствующие процессы таким образом, чтобы снизить вероятность того или иного неблагоприятного исхода (начиная со стихийного бедствия и заканчивая выпуском бракованного изделия) до приемлемого уровня. Компании на Уолл-стрит зачастую пытаются оценивать риски, связанные с их

портфелями при разных сценариях, причем каждому из этих сценариев в зависимости от вероятности его реализации присваивается определенный вес. Финансовый кризис 2008 года отчасти спровоцировали события на рынке, наступление которых считалось крайне маловероятным (например, как если бы все игроки в казино за один вечер оказались в крупном выигрыше). Далее в этой книге я попытаюсь доказать, что модели, которыми руководствовались компании на Уолл-стрит, были изначально ущербными, а данные, использовавшиеся для оценивания ключевых рисков, — слишком ограниченными, однако сейчас я лишь хочу сказать, что в основу любой модели, имеющей дело с рисками, должны быть положены вероятности.

Когда отдельные люди и фирмы не в состоянии полностью устранить неприемлемые для них риски, они пытаются обезопасить себя другими способами. Вся страховая индустрия построена на требовании клиентов защитить их от того или иного негативного события, такого как автомобильная авария, пожар и т. п. Страховая отрасль зарабатывает деньги отнюдь не на устранении подобных случаев: ДТП происходят каждый день, собственно, как и пожары. (Бывает даже так, что автомобиль, врезавшись в дом, становится причиной пожара.) Она процветает за счет взносов владельцев страховых полисов, которых оказывается более чем достаточно, чтобы покрыть ожидаемые страховые выплаты в случае автомобильной аварии или пожара в доме. (Страховая компания может также попытаться снизить ожидаемые страховые выплаты путем поощрения методов безопасного вождения, установки детекторов дыма в каждой спальне, ограждений вокруг водоемов и т. п.)

В определенных случаях концепцию вероятности можно даже использовать для поимки мошенников. Фирма Caveon Test Security специализируется на так называемой экспертизе данных, позволяющей выявить некие закономерности, которые предполагают обман⁵. Например, эта компания (между прочим, основанная бывшим разработчиком тестов SAT*) обратит внимание общественности на результаты экзаменов в том или ином учебном заведении или каком-либо другом месте их проведения, если обнаруженное количество идентичных *неправильных ответов* окажется крайне маловероятным (обычно речь идет о картине, которая складывается реже чем один раз на миллион). При этом она руководствуется следующей математической

* Scholastic Aptitude Test — стандартизированный тест для поступающих в американские высшие учебные заведения. *Прим. ред.*

логикой: когда большая группа учащихся правильно отвечает на какой-то вопрос, из этого нельзя сделать однозначный вывод. Здесь возможны два варианта: либо они дружно списали правильный ответ у кого-то из своих товарищей, либо все как один очень умные ребята. Но когда большая группа учащихся отвечает на какой-то вопрос неправильно, это настораживает: все не могут ответить *одинаково* неправильно — по крайней мере вероятность такого сценария чрезвычайно мала. Это говорит о том, что они списали неправильный ответ у кого-то из одноклассников. Кроме того, Caveon Test Security выявляет экзамены, в ходе которых экзаменуемые отвечают на сложные вопросы значительно лучше, чем на простые (в таком случае предполагается, что ответы им были известны заранее), или количество исправлений неправильного ответа на правильный существенно превышает количество исправлений правильного ответа на неправильный (в таком случае предполагается, что после экзамена преподаватель или экзаменатор подменил листы с ответами).

Разумеется, нетрудно заметить ограничения, присущие использованию вероятностей. Достаточно большая группа экзаменуемых может абсолютно случайно дать одинаково неправильные ответы на какой-то вопрос; к тому же чем больше учебных заведений будет проверяться, тем выше вероятность натолкнуться на подобную картину. Однако никакая статистическая аномалия не опровергает принципиальную правильность предлагаемого подхода. В 2008 году Делма Кинни, пятидесятилетний житель города Атланта, выиграл в мгновенную лотерею миллион долларов, а затем, в 2011-м, еще миллион⁶. Вероятность такого совпадения равна примерно один к 25 триллионам. Естественно, оснований арестовывать г-на Кинни за мошенничество, опираясь исключительно на аналогичные математические выкладки, нет (правда, не мешало бы проверить, не работает ли кто-то из его родственников в лотерейной комиссии штата). Вероятность — лишь один из инструментов в арсенале статистики, и этот инструмент требует умелого обращения.

Выявление важных зависимостей (работа статистика-детектива)

Действительно ли курение вызывает рак? У нас есть ответ на этот вопрос, однако процесс его получения был не так прост, как может показаться на первый взгляд. Научный метод диктует, что при проверке той или иной гипотезы

необходимо провести управляемый эксперимент, в ходе которого именно интересующая нас переменная (например, курение) должна определять разницу между экспериментальной и контрольной группами. Если между двумя этими группами в чем-то (в нашем случае — в частоте возникновения рака легких) прослеживается заметная разница, то можно с уверенностью заключить, что к такому результату привела именно искомая переменная. Однако мы не имеем права ставить над людьми подобные эксперименты. Если, согласно нашей рабочей гипотезе, курение является причиной раковых заболеваний, то было бы неэтично, скажем, разделить недавних выпускников колледжа на две группы, курящих и некурящих, и спустя двадцать лет со дня окончания колледжа, когда они соберутся отметить эту круглую дату, выяснять, кто из них заболел раком легких, а кто — нет. (Управляемые эксперименты над людьми оправданны, если нужно проверить, поможет ли новое лекарство или метод лечения улучшить состояние их здоровья. Но когда речь идет о вероятности летального исхода и нам это хорошо известно, мы не имеем права подвергать людей опасности лишь ради того, чтобы подтвердить или опровергнуть свое предположение.)*

Итак, нам не стоит проводить весьма сомнительный в этическом плане эксперимент, чтобы изучить последствия курения. А не проще ли вместо всей этой заумной методологии взять и сравнить во время встречи по случаю двадцатилетнего юбилея со дня окончания колледжа процент заболевания раком у бывших выпускников — курильщиков и некурильщиков?

Не проще! Курильщики и некурильщики, скорее всего, будут отличаться не только своим отношением к курению. Например, не исключено, что у курильщиков выработался ряд специфических привычек, таких как тяга к алкоголю или склонность к перееданию, что тоже негативно сказывается на их здоровье. Поэтому мы не можем быть твердо убеждены, что их нездоровый вид — следствие именно курения, а не каких-либо других пагубных пристрастий. Кроме того, у нас возникла бы серьезная проблема с данными, на которых основывается наш анализ. Курильщики, действительно заболевшие раком (не говоря уже о тех, кто к тому времени от него умер), вряд ли придут на празднование юбилея. В результате на точности любого анализа состояния здоровья тех, кто пришел (касается ли этот анализ вреда курения или чего-либо другого), существенно скажется то обстоятельство, что

* Разумеется, я заведомо упрощаю здесь многогранные и чрезвычайно сложные проблемы, которые ставит перед нами медицинская этика.

в этом праздновании, скорее всего, примут участие лишь те, кто не испытывает особых проблем со здоровьем. Чем больше лет пройдет с момента окончания учебы в колледже (скажем, будет отмечаться сорокалетний или пятидесятилетний юбилей), тем меньшей будет точность анализа.

Мы не можем относиться к людям как к подопытным кроликам. В итоге статистика оказывается сродни профессии детектива. Исходные данные могут подсказать нам модели, которые в конечном счете способны привести к правильным выводам. Вы наверняка смотрели увлекательные полицейские сериалы наподобие CSI: New York, где очень симпатичные детективы и эксперты-криминалисты скрупулезно исследуют всевозможные «мелочи»: ДНК из остатков слюны на сигаретном окурке, отпечатки зубов на яблоке, кусочек волокна из автомобильного коврика, — а затем используют полученные улики для поимки преступника. «Изюминка» сериала заключается в том, что поначалу эксперты не располагают традиционными вещественными доказательствами (например видеозаписью камер наружного наблюдения или живым свидетелем преступления), позволяющими им изобличить «плохого парня», поэтому им приходится прибегать к научным методам и логическим умозаключениям. Статистика, по сути, идет тем же путем. Исходные данные дают нам некое хаотическое нагромождение подсказок и намеков — так сказать, сцену преступления. А статистический анализ их упорядочивает и систематизирует таким образом, чтобы на их основе можно было сделать логический вывод.

После прочтения главы 11 вы сможете по достоинству оценить телевизионное шоу, которое я планирую предложить какому-либо из телеканалов: CSI: Regression Analysis («CSI: регрессионный анализ»). Это шоу лишь немного отличалось бы от множества других остросюжетных полицейских сериалов. Регрессионный анализ — инструмент, позволяющий исследователям вычленить взаимосвязь между двумя переменными, такими как курение и раковые заболевания, удерживая при этом постоянным (или «учитывая») влияние других важных переменных, таких как режим питания, физические упражнения, вес и т. п. Когда вы читаете в газете о том, что ежедневное употребление в пищу хлеба из отрубей снижает риск заболевания раком толстой кишки, вы не должны думать, что группу несчастных испытуемых насильно кормили хлебом из отрубей в подвале какой-то федеральной лаборатории, в то время как контрольная группа, находившаяся в соседнем здании, с удовольствием уплетала яичницу с беконом. Вовсе нет! Исследователи

собирают подробные сведения о тысячах людей (в том числе как часто они едят хлеб из отрубей), а затем используют регрессионный анализ, чтобы сделать две важные вещи: во-первых, выразить в количественной форме связь между употреблением в пищу хлеба из отрубей и снижением вероятности заболевания раком толстой кишки (например, гипотетический вывод о том, что у тех, кто ежедневно ест хлеб из отрубей, рак толстой кишки встречается на 9% реже, с учетом других факторов, которые могут вызывать это заболевание); и во-вторых, вычислить вероятность того, что связь между ежедневным поеданием хлеба из отрубей и снижением заболеваемости раком толстой кишки, наблюдаемая в этом исследовании, является простым совпадением — случайностью в данных именно для этой выборки людей, — а не устойчивой закономерностью: связью между режимом питания и состоянием здоровья человека.

Разумеется, в телешоу CSI: Regression Analysis будут участвовать профессиональные актеры, которые выглядят на экране гораздо лучше реальных ученых, исследующих такие данные. Этим актерам и актрисам (многие из которых, между прочим, несмотря на молодой возраст, будут иметь ученые степени) предстоит изучить огромные массивы данных и использовать новейшие статистические инструменты для ответа на важные социальные вопросы (например, каковы самые эффективные методы борьбы с преступностью и насилием и какие социальные типы чаще всего становятся террористами). Далее в этой книге мы обсудим концепцию «статистически значимого» вывода, то есть когда в результате анализа выявляется связь между двумя переменными, которая не является случайной. Ученые рассматривают такой статистический вывод как «явную улику». Я предполагаю, что в телешоу CSI: Regression Analysis героиней будет девушка-исследователь, работающая поздно вечером в компьютерной лаборатории, поскольку днем она интенсивно тренируется в составе олимпийской сборной США по пляжному волейболу. Получив распечатку со статистическим анализом, девушка видит именно то, на что и рассчитывала: ярко выраженную статистически значимую связь между некой, по ее мнению, важной переменной и развитием аутизма. Естественно, она тут же спешит поделиться своим открытием с коллегами!

Девушка берет распечатку и бежит по коридору; скорость ее передвижения замедляют лишь высокие каблуки и очень узкая короткая черная юбка. Моя героиня вбегает в комнату к коллеге, симпатичному загорелому парню (и когда он только успел так загореть, ежедневно просиживая

по четырнадцать часов за компьютером?), и демонстрирует ему распечатку. Он задумчиво теребит пальцами свою аккуратно подстриженную эспаньолку, вынимает из ящика письменного стола пистолет калибра 9 мм марки Glock и сует его в боковой карман своего костюма от Hugo Boss за 5000 долларов (и откуда, интересно, взялся у него такой костюмчик, учитывая, что размер его годовой заработной платы составляет примерно 38 000 долларов?). Затем они быстрым шагом направляются в кабинет к боссу, прожженному ветерану сыска, которому уже удалось наладить отношения со своей женой и вылечиться от алкоголизма...

Ладно, вам вовсе не обязательно смотреть телевизор, чтобы оценить важность подобных статистических исследований, практически все важнейшие социальные проблемы решаются с помощью систематического анализа огромных массивов данных. (Во многих случаях их сбор — весьма дорогостоящий и трудоемкий — играет решающую роль в этом процессе, что я постараюсь продемонстрировать в главе 7.) Возможно, я несколько приукрасил своих героев в CSI: Regression Analysis, но это отнюдь не снижает актуальности решаемых ими вопросов. Существует научная литература о террористах и террористах-смертниках — теме, которую было бы очень трудно изучать на живых примерах, используя добровольцев в качестве подопытных кроликов. Одну из таких книг, *What Makes a Terrorist* («Как человек становится террористом»), написал мой преподаватель статистики в магистратуре. Материал книги основан на данных, собранных по результатам террористических актов в разных странах. Вот один из важных выводов, сделанных ее автором, экономистом Принстонского университета Аланом Крюгером: «Террористы отнюдь не всегда оказываются выходцами из беднейших слоев населения или малообразованными людьми, наоборот, обычно они принадлежат к среднему классу; уровень их образования также достаточно высок»⁷.

В чем тут дело? В этой ситуации проявляется одно из ограничений регрессионного анализа. С помощью статистического анализа мы можем изолировать сильную связь между двумя переменными, но далеко не всегда можем объяснить причину ее существования, а в некоторых случаях даже не знаем наверняка, носит ли она причинно-следственный характер (то есть что изменение одной переменной действительно влечет за собой изменение другой переменной). Что касается терроризма, то профессор Крюгер считает, что, поскольку террористы мотивированы определенными политичес-

кими целями, те, кто наиболее образован и богат, движимы сильным желанием изменить общество. Особенно таких людей возмущает подавление свободы — еще один фактор, связанный с терроризмом. Согласно исследованию, выполненному Крюгером, странам с высоким уровнем политических репрессий присущ более высокий уровень террористической деятельности (при условии и неизменности прочих факторов).

Это обсуждение возвращает меня к вопросу, поставленному в названии главы: в чем суть? Точно не в том, чтобы заниматься сложными математическими выкладками или поражать друзей и коллег мудреными статистическими методами. Суть в том, чтобы узнать вещи, которые позволяют нам лучше понимать свою жизнь.

Ложь, наглая ложь и статистика

Даже в идеальных условиях статистический анализ лишь в редких случаях позволяет выявить «истину». Мы обычно выстраиваем некую версию, основанную на косвенных доказательствах, базирующихся на несовершенных данных. В результате появляются многочисленные причины, по которым интеллектуально честные люди не соглашаются со статистическими результатами или выводами. На самом фундаментальном уровне мы можем не соглашаться с самой постановкой рассматриваемого вопроса. Любители спорта будут до бесконечности спорить по поводу «лучшего бейсболиста всех времен и народов» ввиду отсутствия четкого определения того, что именно следует считать «самым лучшим». Изогранные описательные статистики могут в той или иной степени проливать свет на этот вопрос, но они никогда не дадут на него исчерпывающего ответа. Как указывается в следующей главе, гораздо более значимые социальные вопросы пали жертвой той же фундаментальной проблемы. Что происходит с экономическим благополучием американского среднего класса? Ответ на этот вопрос зависит от того, как мы трактуем понятия «средний класс» и «экономическое благополучие».

Существуют определенные ограничения на данные, которые мы в состоянии собрать, и на виды эксперимента, который можем провести. Исследование корней терроризма, выполненное Аланом Крюгером, не могло охватить жизни тысяч молодых людей на протяжении нескольких десятилетий, чтобы проследить, кто из них стал террористом. Это физически невозможно.

Не можем мы и создать две идентичные страны, отличающиеся лишь наличием в *одной* из них мощного репрессивного аппарата, а затем сравнить количество террористов-смертников, появившихся в каждой из них. Даже когда крупномасштабные контролируемые эксперименты на людях проводятся, они оказываются чрезвычайно трудоемкими, сложными и дорогостоящими. Ученые выполнили одно такое исследование, чтобы выяснить, помогают ли молитвы снизить количество и тяжесть послехирургических осложнений (вы, наверное, помните, что это был один из вопросов, поднимавшихся ранее в настоящей главе), и *оно обошлось в 2,4 миллиона долларов* (его результаты обсуждаются в главе 13).

Министр обороны США Дональд Рамсфелд однажды сделал заявление, ставшее знаменитым: «Вы начинаете войну с армией, которая у вас на данный момент есть, а не которую вы хотели бы или можете иметь в будущем». Каким бы ни было ваше мнение о Дональде Рамсфелде (и о войне в Ираке, результаты которой он пытался объяснить), этот афоризм относится не только к армии, но и к исследованиям. Мы выполняем статистический анализ, используя доступные нам данные, методологии и ресурсы. Такой подход не похож на операции сложения или деления в столбик, когда применение правильного метода дает правильный ответ, а компьютер всегда обеспечивает более высокую точность и намного реже ошибается, чем человек. Статистический анализ гораздо больше напоминает работу следователя (что может служить гарантией высокого коммерческого потенциала телешоу CSI: Regression Analysis). А умные и честные люди всегда будут спорить относительно того, о чем именно говорят нам те или иные данные.

Но кто возьмется утверждать, что каждый, кто использует статистику, непременно умный и честный человек? Эта книга задумывалась как дань уважения классическому труду Дарелла Хаффа *How to Lie with Statistics* («Как лгать при помощи статистики»^{*}), который был впервые опубликован в 1954 году и разошелся тиражом свыше миллиона экземпляров. Да, реальность такова, что с помощью статистики *можно* вводить людей в заблуждение или совершать непреднамеренные ошибки. В любом случае математическая точность, сопутствующая статистическому анализу, может служить ширмой для откровенного бреда, которому пытаются придать некое научное обаяние. В своей книге я расскажу о наиболее характерных статистических

^{*} Издана на русском языке: Хафф Д. Как лгать при помощи статистики. — М. : Альпина Паблишер, 2015. *Прим. ред.*

ошибках и искажении фактов, чтобы вы могли распознать подобные случаи манипулирования статистикой (надеюсь, вы не станете сами пытаться ею манипулировать).

Итак, возвращаясь к названию этой главы, зачем нам изучать статистику? Это необходимо для того чтобы:

- обобщать огромные массивы данных;
- принимать более эффективные решения;
- находить ответы на важные социальные вопросы;
- распознавать ситуации, которые позволяют уточнить метод решения тех или иных задач, от продажи подгузников до поимки преступников;
- отслеживать мошенников и находить доказательства, помогающие изобличать преступников;
- оценивать эффективность полиции, тех или иных социальных программ, лекарственных препаратов, медицинских процедур и прочих инноваций;
- а также «вычислять» негодяев, которые используют мощные статистические инструменты для достижения своих неблагоприятных целей.

Если вам удастся делать все это и при этом превосходно выглядеть в костюме от Hugo Boss или черной мини-юбке, то вам ничто не мешает стать очередной звездой телешоу CSI: Regression Analysis.

Кто же все-таки лучший бейсболист всех времен и народов?

Давайте подумаем над двумя на первый взгляд не связанными между собой вопросами:

1. Что происходит с экономическим благополучием американского среднего класса?
2. Кого же все-таки считать лучшим бейсболистом всех времен и народов?

Первый вопрос крайне важен и, как правило, ложится в основу президентских кампаний и других социальных движений. Средний класс, если можно так выразиться, — это сердце Америки, поэтому его экономическое благополучие является индикатором общего экономического благосостояния страны. Второй вопрос тривиален (в буквальном смысле этого слова), однако любители бейсбола готовы до бесконечности спорить по этому поводу. Объединяет оба вопроса то, что они позволяют проиллюстрировать сильные и слабые стороны описательной статистики, которая представляет собой числа и вычисления, используемые для обобщения исходных данных.

Если я захочу продемонстрировать вам, что Дерек Джетер является великим игроком в бейсбол, то смогу описать каждый удачно посланный им мяч в каждом матче Высшей бейсбольной лиги, в котором он принимал участие. Это будут исходные данные, и, чтобы упорядочить их, потребуется какое-то время (с учетом того, что Джетер провел семнадцать сезонов в составе New York Yankees и за это время совершил 9868 удачных бросков).

Или я просто могу вам сказать, что к концу сезона 2011 года средний результат Дерек Джетера за всю его карьеру составлял 0,313. Это описательная, или «сводная» статистика.

Однако такой средний показатель — явное упрощение достижений Джетера за семнадцать сезонов игры в Высшей бейсбольной лиге. Да, он весьма

элегантен в своей простоте, но не отражает всех нюансов спортивной карьеры Джетера. В распоряжении экспертов по бейсболу есть целый арсенал описательных статистик, которые они считают более ценными, чем данный показатель. Я позвонил Стиву Мойеру, президенту Baseball Info Solutions (фирмы, которая предоставила большой объем исходных данных для спортивной драмы Moneyball*), чтобы задать ему два вопроса: 1) каковы самые важные статистические показатели для оценки бейсбольного таланта и 2) кто, по его мнению, величайший бейсболист всех времен и народов? Я познакомлю вас с ответами Стива, когда мы получим больше контекста.

А пока вернемся к менее тривиальному предмету — экономическому благополучию среднего класса. В идеале было бы желательно найти экономический эквивалент среднего показателя (или что-нибудь получше). Нас устроил бы какой-либо простой, но точный показатель того, как за последние годы изменилось экономическое благосостояние типичного американского рабочего. Стали ли люди, которых мы определяем как средний класс, богаче, беднее или в их финансовом положении ничего не изменилось? Подходящий вариант ответа на этот вопрос — который ни в коем случае нельзя рассматривать как «правильный» — рассчитать изменение дохода на душу населения в Соединенных Штатах на протяжении жизни одного поколения (примерно тридцать лет). Доход на душу населения вычисляется путем деления совокупного дохода на численность населения. Согласно этому показателю, средний доход в США повысился с 7787 долларов в 1980 году до 26 487 долларов в 2010-м (последний год, за который правительство располагает соответствующими данными)¹. Вот так-то! Принимайте поздравления.

Есть, правда, одна проблема. Мой быстрый подсчет технически правилен и совершенно неверен с точки зрения ответа на интересующий нас вопрос. Начнем хотя бы с того, что в приведенных выше цифрах отсутствует поправка на инфляцию. (Величина дохода на душу населения 7787 долларов в 1980 году составляет примерно 19 600 долларов в 2010-м.) Такой корректив внести относительно просто. Более серьезная проблема заключается в том,

* В российском прокате этот фильм вышел под названием «Человек, который изменил все». Фильм снят по книге Майкла М. Льюиса, изданной в 2003 году, о бейсбольной команде «Окленд Атлетикс» и ее генеральном менеджере Билли Бине. Его цель — создать конкурентоспособную бейсбольную команду, несмотря на отсутствие больших финансовых возможностей. Главную роль исполняет Брэд Питт. *Прим. перев.*

что средний доход в Америке не равняется доходу среднего американца. Попробуемся расшифровать это утверждение.

Чтобы вычислить величину дохода на душу населения, мы берем весь национальный доход и делим его на численность населения. Однако полученный таким образом показатель абсолютно ничего не говорит нам о том, кто и сколько при этом зарабатывает — хоть в 1980 году, хоть в 2010-м. Как сказали бы участники акции Оссури Wall Street, взрывообразный рост доходов 1% самых богатых людей Америки способен существенно повысить значение дохода на душу населения, ничего при этом не изменив в карманах остальных 99% американцев. Иными словами, средний доход может повышаться без помощи среднего класса.

Как и в случае бейсбольной статистики, мне хотелось узнать мнение авторитетного эксперта о том, как нам следовало бы измерять экономическое благосостояние американского среднего класса. Я спросил у двух известных специалистов по трудовым отношениям, в том числе у ведущего экономического советника президента Обамы, какие описательные статистики они использовали бы для оценки экономического благополучия типичного американца. Вы узнаете их ответы после того, как ознакомитесь с кратким обзором описательных статистик и лучше уясните их смысл.

Будь то бейсбол, доход или что-то еще, самая фундаментальная задача при работе с данными — обобщить их огромные массивы. Численность населения Соединенных Штатов составляет примерно 330 миллионов человек. Электронная таблица, в которой указывались бы фамилия и история доходов каждого американца, содержала бы всю информацию, которая могла потребоваться для оценки экономического благосостояния страны, однако эта информация была бы настолько громоздкой, что извлечь из нее хоть какую-то пользу было бы практически невозможно. Ирония судьбы заключается в том, что чем бóльшим количеством данных мы располагаем, тем труднее выделить в них главное. Поэтому мы вынуждены прибегать к упрощениям. Мы выполняем вычисления, которые сводят сложный массив данных к нескольким числам, описывающим эти данные, точно так же как пытаемся оценить разноплановую программу выступления гимнаста на Олимпийских играх одним числом: 9,8 балла.

Плюс состоит в том, что описательные статистики дают нам некое обобщенное и осмысленное представление исходного явления. О чем, собственно, и идет речь в этой главе. Минус же в том, что любое упрощение

порождает манипулирование. Описательные статистики можно сравнить с анкетами на сайтах знакомств: технически они точны и тем не менее сильно вводят в заблуждение.

Допустим, сидя на работе, вы от нечего делать бродите по интернету и наталкиваетесь на онлайн-дневник известной светской львицы Ким Кардашьян, в котором она рассказывает о своей «долгой» (целых семьдесят два дня!) супружеской жизни с профессиональным баскетболистом Крисом Хэмфри. И вот в тот самый момент, когда вы добрались до описания седьмого дня их супружеской жизни, в комнату неожиданно заходит ваш босс с двумя огромными папками данных. В одной из папок собрана информация о гарантийных претензиях по каждому из 57 334 лазерных принтеров, которые ваша фирма продала в прошлом году. (По каждому из проданных лазерных принтеров перечисляются все проблемы с качеством, зафиксированные в течение гарантийного периода.) В другой содержится такая же информация по каждому из 994 773 лазерных принтеров, которые продал за тот же период ваш главный конкурент. Босс хотел бы сравнить качество принтеров вашей компании с качеством принтеров конкурента.

К счастью, на компьютере, на котором вы почитывали дневник Кардашьян, установлен пакет основных статистических методов, но с чего в данном случае начать? Ваша интуиция, по-видимому, подсказывает вам правильное решение: первой описательной задачей зачастую становится поиск некоего показателя «середины» совокупности данных, или того, что статистики называют «центральной тенденцией». Что является типичным показателем качества для ваших принтеров по сравнению с принтерами конкурента? Обычно самым фундаментальным показателем «середины» какого-либо распределения считается среднее значение. В данном случае нам нужно определить среднее количество проблем с качеством на каждый проданный принтер для вашей фирмы и фирмы вашего конкурента. Вы могли бы просто подсчитать общее число выявленных проблем с качеством для всех принтеров в течение гарантийного периода, а затем разделить его на общее количество проданных принтеров. (Учтите, что в течение гарантийного периода в одном и том же принтере может возникнуть несколько проблем с качеством.) Эту операцию можно проделать для каждой компании, создав важную описательную статистику: среднее количество проблем с качеством на каждый проданный принтер.

Предположим, выяснилось, что среднее количество проблем с качеством в течение гарантийного периода у принтеров вашего конкурента равно 2,8 на каждый проданный принтер, тогда как соответствующий показатель для вашей фирмы составляет 9,1. Как видите, вывести среднее значение совсем не сложно. Вы просто использовали информацию для миллиона принтеров, проданных двумя разными компаниями, и извлекли из нее суть интересующей вас проблемы: ваши принтеры ломаются слишком часто. Похоже, самое время отправить боссу по электронной почте краткое уведомление с численным подтверждением столь тревожного факта, а затем вернуться к более увлекательному занятию: чтению дневника Ким Кардашьян.

А может, не стоит торопиться? Я ведь не зря выразился довольно туманно, упомянув о какой-то там «середине» распределения. В этом отношении у среднего значения есть определенные проблемы, а именно: оно подвержено существенным искажениям со стороны «отщепенцев», то есть значений, резко отклоняющихся от центра. Чтобы вам было легче уяснить эту концепцию, вообразите десяток парней, сидящих у стойки бара какого-нибудь питейного заведения в Сиэтле, рассчитанного на представителей среднего класса. Каждый из парней зарабатывает по 35 000 долларов в год; стало быть, средний годовой доход этой группы составляет 35 000 долларов. Внезапно в заведение входит Билл Гейтс с говорящим попугаем на плече (вообще-то в данном примере говорящий попугай не играет никакой особой роли; это не более чем деталь, призванная несколько оживить повествование и придать ему определенный колорит) и усаживается на одиннадцатый стул за стойкой бара; при этом средний годовой доход его завсегдатаев резко повышается до 91 миллиона долларов. Очевидно, что первые десять посетителей бара могут лишь мечтать о таком уровне годового дохода (хотя все они, наверное, надеются, что Билл Гейтс расщедрится и угостит их стаканчиком-другим). Если бы я написал, что средний годовой доход посетителей заведения составляет 91 миллион долларов, то данный вывод был бы статистически правильным, однако не имел бы ничего общего с реальным положением вещей. Этот бар отнюдь не относится к числу заведений, где коротают свободное время мультимиллионеры, — здесь обычно отдыхают молодые люди с относительно невысоким уровнем годовых доходов. Просто сегодня им повезло оказаться в компании с Биллом Гейтсом и его говорящим попугаем. Именно высокая чувствительность

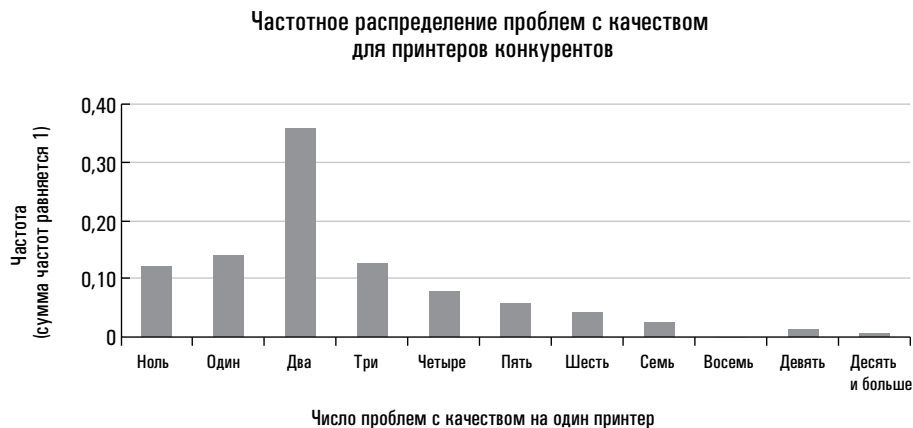
среднего значения к значениям, резко отклоняющимся от центра, не позволяет нам измерять экономическое благополучие среднего класса с помощью такого показателя, как величина дохода на душу населения. Поскольку в последнее время наблюдается резкий рост доходов в верхней части распределения — глав компаний, управляющих хедж-фондами и выдающихся спортсменов, таких как Дерек Джетер, — величина среднего дохода в США может быть сильно искажена, как в вышеупомянутом баре, где несколько парней с относительно скромными доходами случайно оказались в компании Билла Гейтса.

По этой причине нам приходится пользоваться еще одной статистикой, которая также является отражением «середины» распределения, однако делает это несколько иначе. Речь идет о так называемой медиане. Медиана — это точка, которая делит распределение пополам таким образом, что одна половина наблюдений располагается выше медианы, а другая половина — ниже. (При наличии четного количества наблюдений медиана представляет собой среднюю точку между двумя средними наблюдениями.) Если мы вернемся к примеру с баром, то срединный (медианный) годовой доход для десяти человек, сидевших поначалу за стойкой, равняется 35 000 долларов. Когда в заведении появился — и уселся на одиннадцатый стул — Билл Гейтс с говорящим попугаем, срединный годовой доход для одиннадцати человек по-прежнему составлял 35 000 долларов. Если представить, что посетители бара расселись за его стойкой в порядке возрастания их доходов, то доход посетителя, сидящего на шестом стуле, будет срединным для данной группы людей. Даже если бы в заведение зашел Уоррен Баффет и уселся рядом с Биллом Гейтсом на двенадцатый стул, медиана все равно осталась бы неизменной*.

В случае распределений без «отщепенцев» срединное (медиана) и среднее значения совпадают. Выше говорилось о гипотетической сводке данных, отражающих качество принтеров конкурирующей фирмы. В частности, я представил эти данные в виде так называемого частотного распределения (гистограммы). Число проблем с качеством на один принтер представлено

* После того как в баре оказалось бы двенадцать посетителей, медианой была бы средняя точка между доходом посетителя, сидящего на шестом стуле, и доходом посетителя, сидящего на седьмом стуле. Поскольку доход того и другого составляет 35 000 долларов, медиана равняется 35 000 долларов. Если бы доход одного из них равнялся 35 000, а доход другого — 36 000, то медиана для этой группы в целом равнялась бы 35 500 долларов.

на горизонтальной оси (внизу); высота каждого вертикального столбца соответствует проценту проданных принтеров, у которых наблюдалось такое число проблем с качеством. Например, у 36% принтеров конкурента в течение гарантийного периода возникало по две проблемы с качеством. Поскольку это распределение включает все возможные случаи проблем с качеством (в том числе и их отсутствие), сумма всех долей (процентов) должна равняться 1 (или 100%).



Поскольку такое распределение почти симметрично, среднее и срединное значения довольно близки друг к другу. Распределение слегка скошено вправо, что объясняется малым количеством принтеров, имеющих множественные дефекты. Эти «отщепенцы» слегка смещают среднее значение вправо, однако на медиану это не влияет. Допустим, что перед тем как составить для босса отчет о качестве принтеров, вы принимаете решение вычислить *медианы*, то есть число проблем с качеством для принтеров, проданных вашей и конкурирующей компанией. Нажав всего несколько клавиш, вы получите результат. Медиана проблем с качеством для принтеров конкурента равняется 2; а для принтеров вашей фирмы — 1.

Что из этого следует? Оказывается, медиана проблем с качеством на каждый принтер вашей фирмы фактически *меньше*, чем у вашего конкурента. Поскольку супружеская жизнь Ким Кардашьян становится однообразной, а полученный результат вас заинтриговал, вы распечатываете распределение частот проблем с качеством у принтеров, проданных вашей компанией.



Из приведенных выше гистограмм становится ясно, что для вашей компании нехарактерно равномерное распределение проблем с качеством. Напротив, у вас налицо проблема «лимона»*: у малого числа ваших принтеров наблюдается большое количество дефектов. Эти «отщепенцы» способствуют наращиванию среднего значения, тогда как медиана остается неизменной. Более важным с производственной точки зрения является то обстоятельство, что вам нет необходимости переоснащать весь производственный процесс; достаточно лишь определить, какое из предприятий компании выпускает некачественную продукцию, и исправить ситуацию**.

Вычисление среднего и медианы не представляет особых трудностей; самое главное в этом случае — определить, какой именно показатель «середины» более точен в каждой конкретной ситуации (именно этот фактор нередко используется для манипулирования средними показателями). Между тем у медианы имеются весьма полезные «родственники». Как указывалось выше, медиана делит любое распределение пополам. Затем его можно разбить на четверти, или, как их еще называют, квартили. Первый квартиль состо-

* «Лимонами» на американском сленге называют устройства с дефектами, которые проявляются уже после покупки. *Прим. ред.*

** Вот что удалось выяснить в ходе дальнейшего исследования проблемы. Оказалось, что почти все бракованные принтеры производились на заводе в Кентукки, где рабочие разобрали часть сборочного конвейера, чтобы создать подпольное предприятие по изготовлению виски. Постоянно пьяные рабочие и частично разобранный сборочный конвейер стали причиной резкого ухудшения качества выпускаемых заводом принтеров.

ит из нижних 25% наблюдений; второй из следующих 25% наблюдений и т. д. Еще один вариант — разделить распределение на децили, каждый из которых включает в себе 10% наблюдений. (Если ваш доход находится в верхнем дециле американского распределения доходов, то это означает, что вы зарабатываете больше, чем 90% ваших коллег-рабочих.) Можно пойти еще дальше и разбить распределение на сотые доли, или процентиля. Каждый процентиль представляет 1% распределения; таким образом, первый процентиль представляет нижний 1% данного распределения, а 99-й — его верхний 1%.

Преимущество описательных статистик такого рода заключается в том, что они указывают, где именно располагается то или иное конкретное наблюдение по сравнению с остальными. Например, информация, что ваш ребенок по результатам теста на понимание прочитанного материала получил третий процентиль, должна сказать вам о том, что вы уделяете недостаточно внимания совместному обсуждению книг, прочитанных вашим ребенком. Вам вовсе не обязательно знать подробности самого теста или точное количество вопросов, на которые ваш ребенок ответил правильно. Однако его попадание в определенный процентиль в любом случае говорит о том, насколько успешно ваш ребенок сдал этот тест по сравнению с другими его участниками. Если тест был сравнительно легким, то большинство его участников правильно ответят на подавляющее число вопросов, при этом количество правильных ответов у вашего ребенка все равно будет меньшим, чем у большинства других участников тестирования. Если же тест был очень трудным, то у всех его участников окажется малое число правильных ответов, однако и в этом случае «рейтинг» вашего ребенка будет несколько ниже, чем у остальных.

Сейчас самый подходящий момент познакомить вас с новой терминологией. «Абсолютная» сумма баллов, «абсолютный» показатель или «абсолютное» значение обладают неким внутренним, самостоятельным смыслом. Если я набираю 83 балла в результате бросков по восемнадцати лункам при игре в гольф, то речь идет об абсолютном показателе. Я мог бы продемонстрировать такой результат в день, когда температура достигала 41 градуса, что также является абсолютным показателем. Абсолютные показатели, как правило, можно интерпретировать без какого-либо контекста или дополнительной информации. Когда я сообщаю, что набрал 83 балла, вам, чтобы оценить достигнутый мною результат, вовсе не обязательно знать, сколько баллов набрали в тот день другие гольфисты. (Исключением может быть

ситуация, когда условия проведения игры особенно неблагоприятны или площадка для гольфа имеет очень сложный или, напротив, очень простой рельеф.) Если же по итогам турнира я оказался на девятом месте, то это относительная статистика. «Относительное» значение, или «относительный» показатель имеет смысл лишь в сравнении с чем-либо или в каком-либо более широком контексте, например в сравнении с восемью гольфистами, получившими более высокие баллы, чем я. Результаты большинства стандартизованных тестов тоже представляют интерес лишь как относительная статистика. Если я сообщу, что по итогам проведения единого экзамена штата Иллинойс ученик третьего класса одной из начальных школ штата набрал 43 балла из 60 возможных, то этот абсолютный показатель скажет вам не так много. Но если я преобразую его в процентиль — то есть помещу в некое распределение, содержащее показатели всех учеников третьих классов начальных школ штата Иллинойс, — то он обретет гораздо больший практический смысл. Поскольку 43 правильных ответа попадают в 83-й процентиль, знания этого ученика гораздо выше, чем у большинства его сверстников в штате Иллинойс. Если бы этот ученик оказался в 8-м процентиле, то уровень его знаний оценивался бы как весьма посредственный. В этом случае процентиль (относительный результат) несет в себе гораздо больше информации, чем количество правильных ответов (абсолютный показатель).

Еще одной статистикой, которая позволяет описывать большие нагромождения данных, является среднеквадратическое (или, как его еще называют, стандартное) отклонение — показатель разброса данных по отношению к их среднему значению. Другими словами, среднеквадратическое отклонение представляет собой показатель рассредоточенности наблюдений. Допустим, я собрал информацию о весе 250 человек, направляющихся на самолете в Бостон; кроме того, у меня есть данные о весе выборки (численность которой также составляет 250 человек) участников Бостонского марафона. Допустим также, что средний вес у членов обеих групп примерно одинаков и составляет 155 фунтов. Каждый, кому приходилось летать в забитом под завязку самолете, знает, что многие пассажиры типичного коммерческого рейса весят больше 155 фунтов. Однако завсегда так таких рейсов также хорошо известно, что среди пассажиров встречается немало число крикливых грудных младенцев и непоседливых детишек дошкольного и младшего школьного возраста, вес которых явно недотягивает до указанного значения. Когда нам приходится вычислять средний вес пассажиров самолета, то масса

320-фунтовых футболистов, сидящих по обе стороны от вашего кресла, наверняка компенсируется визгливым грудным младенцем, занимающим место с другой стороны прохода между креслами, и шестилетним мальчуганом, сидящим позади вас и пинающим ногами спинку вашего кресла.

На основе уже известных вам описательных инструментов мы приходим к выводу, что вес пассажиров самолета и участников марафона примерно одинаков. *Однако на самом деле это не совсем так.* Да, вес этих двух групп приблизительно одинаков «в среднем», но у пассажиров самолета гораздо больший разброс относительно этого среднего значения, то есть показатели их веса сильнее удалены от него. Мой восьмилетний сынишка сказал бы, что бегуны-марафонцы кажутся людьми, имеющими примерно одинаковый вес, тогда как среди пассажиров самолета встречаются как миниатюрные люди, так и настоящие здоровяки. Показатели веса пассажиров самолета характеризуются «бóльшим разбросом», что обязательно нужно учитывать при описании веса этих двух групп. Среднеквадратическое отклонение является описательной статистикой, которая позволяет выразить данный разброс по отношению к среднему значению единственным числом. Формулы для вычисления среднеквадратического отклонения и дисперсии (еще один широко распространенный показатель разброса, на основе которого вычисляется среднеквадратическое отклонение) включены в приложение, приведенное в конце этой главы. А сейчас давайте подумаем над тем, зачем нам измерять разброс.

Допустим, вы приходите в кабинет врача. С тех пор как вас выдвинули на руководящую должность, назначив главой Отдела борьбы за повышение качества североамериканских принтеров, вы чувствуете хроническую усталость. У вас берут кровь на анализ, и через пару дней ассистент врача отправляет вам на автоответчик сообщение о том, что некий показатель (назовем его условно $HCb2$) у вас в крови равняется 134. Вы быстро отправляете соответствующий поисковый запрос в интернет и выясняете, что величина $HCb2$ для людей вашего возраста составляет 122 (и медиана почти такая же). Черт побери! Случись нечто подобное со мной, я поспешил бы составить завещание — так, на всякий случай. Итак, вы пишете слезные письма родственникам, детям и близким друзьям. У вас возникает мысль прыгнуть напоследок с парашютом (ваша жизнь была так бедна на острые ощущения!) или попытаться как можно быстрее написать роман (а вдруг в вас скрывался недоюжинный писательский талант?). У вас даже может появиться желание

отправить по электронной почте письмо своему боссу, в котором вы сравните его с некой частью человеческого тела (и набрать весь текст письма ЗАГЛАВНЫМИ БУКВАМИ).

Между тем ничего этого вам, скорее всего, делать не следует (а идея с оскорбительным письмом боссу — глупая в любом случае). Когда вы повторно приходите к врачу, чтобы получить от него направление в хоспис, ассистент врача сообщает вам, что результаты вашего анализа крови находятся в пределах нормы. Как такое возможно? «Мой показатель HСb2 превышает среднее значение на целых 12 пунктов!» — недоумеваете вы.

«Среднеквадратическое отклонение для HСb2 равняется 18», — успокаивает вас ассистент врача.

Что все это значит?

Дело в том, что у HСb2, как и у большинства других биологических явлений (например, роста человека), существует вполне естественный разброс значений. В то время как среднее значение HСb2 действительно может составлять 122, у огромного числа здоровых людей оно может быть несколько выше или ниже. Опасность возникает только тогда, когда значение HСb2 *намного* выше или ниже указанного среднего значения. Но что именно следует понимать под «намного» в данном контексте? Как уже говорилось, среднеквадратическое отклонение является показателем разброса, то есть оно демонстрирует, насколько плотно группируются наблюдения вокруг среднего значения. Для многих типичных распределений данных высокая доля наблюдений располагается в пределах одного среднеквадратического отклонения от среднего значения (это означает, что они находятся в диапазоне, простирающемся от одного среднеквадратического отклонения ниже среднего значения до одного среднеквадратического отклонения выше среднего значения). Проиллюстрируем это на простом примере. Средний рост взрослого мужчины-американца равняется 5 футам 10 дюймам. Среднеквадратическое отклонение составляет примерно 3 дюйма. Рост значительной доли взрослых мужчин находится между 5 футами 7 дюймами и 6 футами 1 дюймом.

То же самое можно сформулировать несколько иначе: любой мужчина в этом диапазоне роста не считался бы слишком высоким или низким. Это возвращает нас к результатам количественного анализа HСb2, которые так нас встревожили. Да, значение HСb2 на 12 пунктов выше среднего, но это меньше, чем одно среднеквадратическое отклонение, что является аналогом роста, близкого к 6 футам, — следовательно, никакой особой аномалии

здесь не наблюдается. Разумеется, гораздо меньшее число наблюдений находится на расстоянии двух стандартных отклонений от среднего значения; еще меньшее число наблюдений находится на расстоянии трех или четырех стандартных отклонений. (Что касается роста, то американский мужчина выше среднего роста на три среднеквадратических отклонения достигал бы 6 футов 7 дюймов или был бы даже еще выше.)

Некоторые распределения более рассредоточены, чем другие. Следовательно, среднеквадратическое отклонение значений веса 250 пассажиров самолета будет выше, чем значений веса 250 бегунов-марафонцев. Распределение частот веса пассажиров самолета оказалось бы более «разбросанным», чем бегунов-марафонцев. После того как мы узнаем среднее значение и стандартное отклонение для какой-либо совокупности данных, мы получаем о ней весьма ценные сведения. Допустим, я сообщаю вам, что по результатам проведения единого экзамена по математике какого-либо штата средняя сумма баллов составила 500 при стандартном отклонении, равном 100. Как и в случае с ростом мужчин, большая часть учащихся, сдававших экзамен, продемонстрировала результаты в пределах одного среднеквадратического отклонения от среднего значения, то есть между 400 и 600 баллами. Сколько учеников, по вашему мнению, получили 720 и выше? Наверное, очень немногие, поскольку такой показатель превышает два среднеквадратических отклонения от среднего значения.

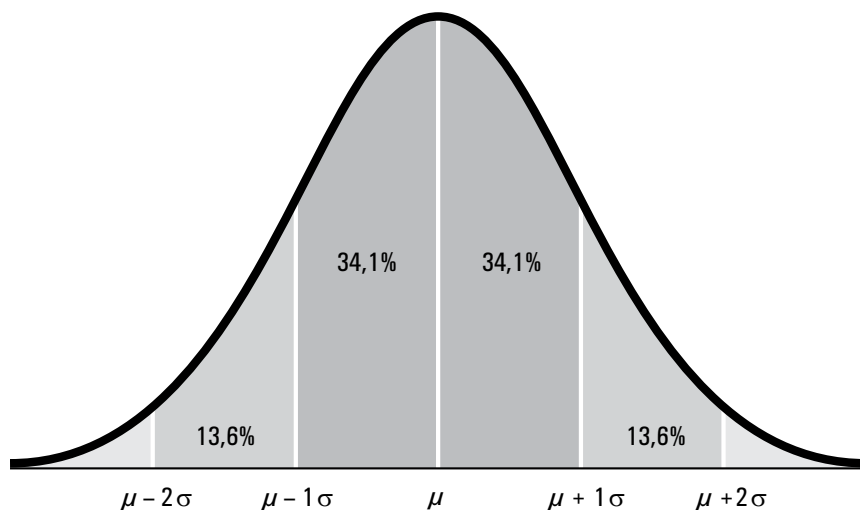
Теперь не мешало бы уточнить, *что* в данном случае имеется в виду под словами «очень немногие». Думаю, самое время познакомить читателей с одним из наиболее важных, полезных и распространенных распределений в статистике — нормальным распределением. Данные, которые распределены согласно этому закону, располагаются симметрично относительно своего среднего значения, причем это распределение имеет колоколообразную форму, которая наверняка вам хорошо знакома.

Нормальное распределение описывает многие явления, часто встречающиеся в жизни. Представьте себе распределение частот, описывающее, как стреляют зерна воздушной кукурузы (попкорна) на плите. Некоторые зерна начинают лопаться раньше остальных, издавая примерно один-два хлопка в секунду; через десять или пятнадцать секунд зерна уже взрываются как сумасшедшие. Постепенно количество хлопков в секунду сокращается приблизительно до частоты, наблюдавшейся в самом начале поджаривания. Значения роста мужчин-американцев распределены практически

в соответствии с законом нормального распределения, то есть расположены почти симметрично относительно среднего значения (5 футов 10 дюймов). Каждый тест SAT специально разрабатывается таким образом, чтобы обеспечить нормальное распределение результатов со средним значением 500 при среднеквадратическом отклонении, равном 100. Согласно Wall Street Journal, американцы даже склонны по закону нормального распределения парковать свои автомобили у крупных торговых центров: большинство автомобилей паркуются напротив центрального входа в торговый центр («вершина» кривой нормального распределения), а «хвосты» машин расходятся вправо и влево от центрального входа.

Красота нормального распределения — его мощь, изящество и элегантность — обусловлена тем, что нам по определению известно, какая именно доля наблюдений в нормальном распределении находится в пределах одного среднеквадратического отклонения от среднего значения (68,2%), двух среднеквадратических отклонений от среднего значения (95,4%), трех среднеквадратических отклонений от среднего значения (99,7%) и т. д. Хотя все это может показаться тривиальным, это именно тот фундамент, на котором строится значительная часть статистики. Мы вернемся к концепции нормального распределения чуть позже, чтобы рассмотреть ее подробнее.

Нормальное распределение



Средним значением является средняя линия, которую часто обозначают греческой буквой μ . Среднеквадратическое (стандартное) отклонение зачастую обозначают греческой буквой σ . Каждая вертикальная полоса на графике представляет одно среднеквадратическое отклонение.

Описательные статистики часто служат для сравнения двух значений или величин. Я на один дюйм выше своего брата; сегодня температура воздуха на девять градусов больше «исторического среднего» для этой даты и т. д. Такие сравнения имеют смысл, поскольку большинство из нас признают используемые в этих случаях шкалы единиц измерения. Один дюйм — не так много, когда речь идет о человеческом росте, поэтому вы можете заключить, что у нас с братом примерно одинаковый рост. И напротив, девять градусов — значительное отклонение температуры воздуха практически для любого климата в любое время года; поэтому, если в какой-то из дней было зафиксировано превышение средней температуры на девять градусов, это существенная аномалия. Но допустим, я сообщу, что хлопья Granola Cereal A содержат на 31 миллиграмм больше натрия, чем хлопья Granola Cereal B. Если вы не знакомы со специальной литературой, в которой рассматриваются последствия употребления в пищу натрия, и не знаете, о какой величине порции хлопьев идет в данном случае речь, на основе приведенной выше информации вы не сделаете полезных выводов. А если я скажу вам, что мой кузен Эл заработал в текущем году на 53 000 долларов меньше, чем в прошлом? Следует ли нам тревожиться за судьбу Эла? А что если он управляющий хедж-фонда, для которого сумма 53 000 долларов соизмерима с ошибкой округления при подсчете его годового дохода?

В примерах с содержанием натрия в хлопьях и доходом Эла отсутствует контекст, который позволил бы оценить масштаб проблемы, если таковая имеется. Самый простой способ придать смысл этим сравнениям — использовать процентные величины. Если бы я сообщил вам, что хлопья Granola Cereal A содержат на 50% больше натрия, чем хлопья Granola Cereal B, а доход моего кузена Эла сократился в прошлом году на 47%, это позволило бы вам сделать определенные выводы. Оценка тех или иных изменений в процентах предоставляет нам нечто наподобие шкалы.

Поскольку в школе вас наверняка научили вычислять проценты, не исключено, что у вас возникнет соблазн не читать нескольких следующих абзацев. Что ж, возможно, вы правы. Однако прежде чем принять окончательное решение, выполните одно простое упражнение. Допустим, в универмаге

продается платье за 100 долларов. Заместитель директора универмага решает снизить цену всех товаров на 25%. Но впоследствии его увольняют за то, что он зависает в баре с Биллом Гейтсом*, а новый заместитель директора распоряжается повысить все цены на 25%. Какой окажется окончательная цена платья? Если вы скажете (или подумаете), что 100 долларов, то вам лучше все же читать текст подряд.

В действительности окончательная цена платья составит 93,75 доллара. Этот нехитрый трюк принесет вам порцию аплодисментов и восхищение присутствующих на какой-нибудь вечеринке. Процентные величины — полезнейшая вещь, но подчас они порождают в головах людей путаницу и даже способны ввести в заблуждение. Формула для вычисления разности (или изменения) процентов такова: (новая величина – исходная величина) / исходная величина. Числитель (верхняя часть дроби) дает нам величину изменения в абсолютных значениях; знаменатель (нижняя часть дроби) помещает это изменение в контекст путем его сравнения с нашей исходной точкой. Поначалу это кажется очевидным, как в случае, когда заместитель директора универмага снижает цену платья (100 долларов) на 25%. Двадцать пять процентов от первоначальной цены (100 долларов) составляют 25 долларов; это скидка, в результате цена платья становится 75 долларов. Вы можете вставить соответствующие числа в указанную выше формулу и проделать простые вычисления, чтобы убедиться в правильности моих подсчетов: $(100 \text{ долл.} - 75 \text{ долл.}) / 100 \text{ долл.} = 0,25$, или 25%.

Платье продается за 75 долларов до тех пор, пока новый заместитель директора универмага не примет решение повысить цену на 25%. Именно в этом месте многие совершают ошибку, поскольку 25-процентное повышение цены вычисляется как процент от новой, сниженной цены платья, которая равняется 75 долларов. Повышение цены составит $0,25 \times 75 \text{ долл.} = 18,75 \text{ долл.}$; вот так и получается окончательная цена платья — $75 \text{ долл.} + 18,75 \text{ долл.} = 93,75 \text{ долл.}$ (а не 100 долларов). Дело в том, что любое процентное изменение всегда дает значение какого-то числа *относительно чего-либо еще*. Следовательно, нам нужно лучше понять, что же представляет собой это «что-то еще».

* Интересно отметить, что этот менеджер — один из тех десяти парней с годовым доходом 35 000 долларов, которые сидели в баре, когда туда вошел Билл Гейтс с говорящим попугаем на плече. Причуды судьбы!

Однажды я инвестировал деньги в компанию, основанную моим приятелем, с которым мы проживали в одной комнате студенческого общежития во время учебы в колледже. Поскольку это был частный бизнес, от его владельца не требовалось предоставлять акционерам строго определенный перечень сведений о его деятельности. В течение нескольких лет мне ничего не было известно о судьбе моей инвестиции — бывший приятель предпочитал не распространяться на сей счет. Наконец я получил по почте письмо, в котором говорилось, что прибыль компании выросла на 46% по сравнению с предыдущим годом. Какой была эта прибыль в абсолютных показателях, в письме не сообщалось, стало быть, я по-прежнему не имел ни малейшего представления об эффективности своих инвестиций. Допустим, в прошлом году эта фирма заработала 27 центов (то есть практически ничего), а в текущем — 39 центов (то есть опять-таки почти ничего). Тем не менее прибыль компании выросла с 27 центов до 39 центов, то есть на 47%! Очевидно, что рассылка такого письма акционерам — если бы в нем указывалось, что прибыль, накопленная фирмой за два года, меньше стоимости чашки кофе в сети Starbucks, — принесла бы им не радость, а жестокое разочарование.

К чести моего приятеля должен заметить, что в конечном счете он продал свою компанию за несколько сотен миллионов долларов, заработав для меня стопроцентную прибыль на вложенный капитал. (Поскольку вы не знаете, какую именно сумму я вложил в этот бизнес, вы не можете знать, сколько денег я в результате заработал. Впрочем, это лишь подтверждает правильность мыслей, высказанных мною выше.)

Читателям следует уяснить еще одно важное различие. Процентное изменение не следует путать с изменением, выраженным в процентных пунктах. Ставки зачастую выражаются в процентах. Ставка налога с продаж в штате Иллинойс равняется 6,75%. Я выплачиваю своему агенту 15% с авторских гонораров, которые получаю за свои книги. Эти ставки применяются к той или иной величине (например, к доходу в случае ставки подоходного налога). Очевидно, что ставки могут изменяться в ту или иную сторону. Менее очевидным является то обстоятельство, что такие *изменения* ставок можно описывать по-разному. Самым показательным примером в этом отношении может служить недавнее повышение ставки индивидуального подоходного налога в штате Иллинойс с 3% до 5%. Такое изменение налога можно выразить двумя способами, причем оба технически корректны. Представители Демократической партии США, которые инициировали это

повышение, объясняли (кстати говоря, совершенно правильно), что *ставка* подоходного налога в этом штате выросла на 2 *процентных пункта* (с 3% до 5%). Представители Республиканской партии США отмечали (также совершенно правильно), что подоходный налог в штате увеличился на 67%. [Это является весьма удобным способом проверки формулы, приведенной выше: $(5 - 3) / 3 = \frac{2}{3}$, что приблизительно соответствует 67%.]

Демократы сосредоточили внимание на абсолютном изменении налоговой ставки; республиканцы предпочли сфокусироваться на изменении величины налогового бремени. Как указывалось выше, оба описания правильны с технической точки зрения, хотя я настаиваю, что описание, предложенное республиканцами, более точно отражает влияние изменения этого налога, поскольку его величина, которую мне предстоит выплачивать государству — ведь именно она меня интересует, а вовсе не способ ее вычисления, — действительно повысится на 67%.

Многие явления окружающей нас действительности невозможно идеально описать посредством какой-то одной статистики. Допустим, куортербек Аарон Роджерс выполняет броски на 365 ярдов, которые, однако, не являются тачдаун-пасами. Между тем Пейтон Мэннинг совершает броски лишь на 127 ярдов — но с тремя тачдаун-пасами. Мэннинг зарабатывал больше очков, но, возможно, именно Роджерс приносил своей команде больше тачдаунов (то есть пересечений мячом или игроком с мячом линии зачетного поля соперника). Кого из них считать более ценным игроком? В главе 1 я обсуждал так называемый рейтинг распасовщика, который по идее должен решить эту статистическую проблему и широко применяется Национальной футбольной лигой. Рейтинг распасовщика — пример индекса, представляющего собой описательную статистику, составленную из других описательных статистик. После того как разные показатели эффективности действий куортербеков удалось объединить в один, такая статистика может использоваться для сравнения игры куортербеков в определенный день или даже на протяжении всей спортивной карьеры. Если бы единый индекс такого рода существовал в бейсболе, то вопрос о том, кого следует считать лучшим бейсболистом всех времен и народов, удалось бы давно решить, не так ли?

Преимущество любого индекса заключается в том, что он консолидирует в едином показателе большой объем сложной информации. После этого мы можем сопоставлять между собой вещи, которые в противном случае не поддаются простому сравнению (речь может идти о чем угодно, от сравнения

эффективности действий куортербеков до конкурсов красоты или работы разных колледжей). При проведении конкурса «Мисс Америка» победитель определяется по результатам пяти отдельных соревнований: личное интервью, купальник, вечернее платье, индивидуальные способности и вопрос на сцене. («Мисс конгениальность» выбирают сами участницы путем индивидуального голосования.)

Парадокс, но то, что любой индекс консолидирует в едином показателе большой объем сложной информации, является также его недостатком. Вывести единый показатель можно бесчисленным множеством способов, причем все они могут приводить к разным результатам. Малкольм Гладуэлл блестяще доказывает этот факт в одной из своих статей в еженедельнике *The New Yorker*, где высмеивает неизбытную тягу американцев к присвоению рейтингов буквально всему, что их окружает². (Особенно досталось от Малкольма тем, кто составляет рейтинги учебных заведений.) Гладуэлл приводит пример присвоения журналом *Car and Driver* («Автомобиль и водитель») рейтинга трем моделям спортивных автомобилей: *Porsche Cayman*, *Chevrolet Corvette* и *Lotus Evora*. Используя формулу, которая включает двадцать одну переменную, *Car and Driver* поставил на первое место *Porsche Cayman*. Однако Гладуэлл указывает, что в формуле *Car and Driver* такой показатель, как «дизайн кузова», оценивается всего в 4% от совокупного рейтинга, что для спортивного автомобиля смехотворно мало. Если бы «дизайн кузова» оценивался, к примеру, в 25%, то на первом месте оказался бы *Lotus Evora*.

Но это еще не все. Гладуэлл также отмечает, что в формуле *Car and Driver* такой показатель, как рекомендованная цена автомобиля, тоже имел ничтожный вес. Если бы этому важному показателю был присвоен больший вес (так, чтобы у цены, дизайна кузова и характеристик двигателя были одинаковые весовые коэффициенты), то на первом месте оказался бы *Chevrolet Corvette*.

Любой индекс очень чувствителен к описательным статистикам, которые включены в его состав, а также к весу, присваиваемому каждой из составляющих. В результате диапазон индексов простирается от полезных, но весьма несовершенных инструментов, до полнейших курьезов. Примером первого может служить так называемый индекс человеческого развития (*Human Development Index — HDI*), применявшийся ООН. HDI разрабатывался как более широкий показатель экономического благосостояния, чем доход как таковой. Доход является лишь одним из компонентов HDI, который

включает также показатели средней продолжительности жизни и уровня образования. По объему производства на душу населения Соединенные Штаты находятся на одиннадцатом месте в мире (пропустив вперед такие богатые запасами нефти страны, как Катар, Бруней и Кувейт), а по индексу человеческого развития занимают четвертое место в мире³. Правда, HDI-рейтинги слегка изменились бы в результате трансформации составных частей индекса, но вряд ли это бы привело к примерному равенству рейтингов Зимбабве и Норвегии. Иными словами, индекс HDI неплохо отражает текущую картину, касающуюся жизненных стандартов в разных странах мира.

Описательные статистики дают нам понимание сути интересующих нас явлений. Исходя из этого мы можем вернуться к вопросам, поставленным в начале главы. Кого же считать лучшим бейсболистом всех времен и народов? С точки зрения целей этой главы, гораздо важнее было бы выяснить, какие описательные статистики больше всего помогли бы нам ответить на этот вопрос. Согласно Стиву Мойеру, президенту Baseball Info Solutions, тройку ключевых статистик (кроме возраста) для оценивания эффективности действий любого игрока, за исключением питчера (подающего), составили бы следующие:

1. Процент попаданий в базу (on-base percentage — OBP), иногда называемый средним показателем попаданий в базу (on-base average — OBA). Оценивает процент успешных попаданий игрока в базу, в том числе и так называемые уоки (которые не учитываются в среднем показателе).
2. Процент отбивания (slugging percentage — SLG). Измеряет процент отбивания мячей путем вычисления совокупного количества попаданий в базу на каждый отбитый мяч. Одинарный оценивается в 1, двойной соответствует 2, тройной — 3, а хоумран — 4. Таким образом, процент отбивания у беттера (отбивающего), который отбил одинарный и тройной из пяти попаданий, составил бы $(1 + 3) / 5$, или 0,800.
3. Попадания (at bats — AB). Этот показатель помещает все сказанное выше в единый контекст. Любой игрок может продемонстрировать потрясающую статистику в одной-двух играх. Но лишь суперзвезда накапливает впечатляющие показатели на протяжении многих лет выступления за профессиональные бейсбольные команды.

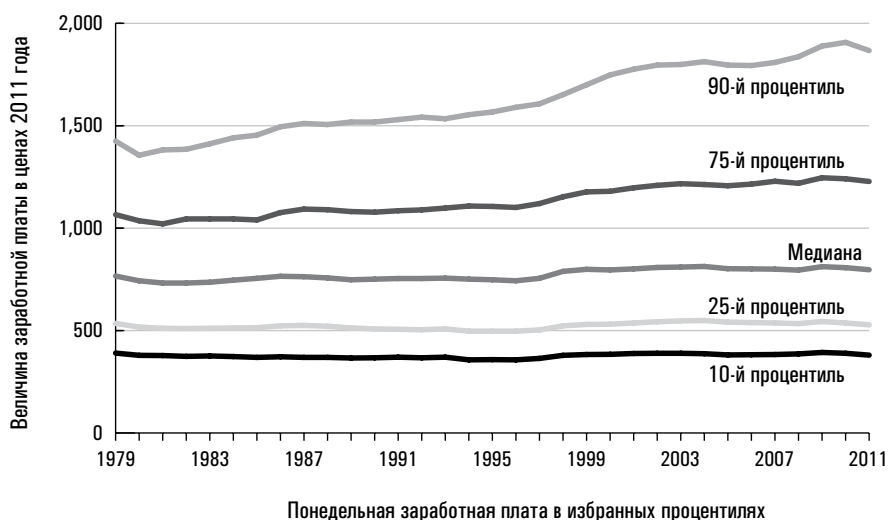
По мнению Стива Мойера (которое я полностью разделяю), лучшим бейсболистом всех времен и народов является Бейб Рут из-за его уникальной способности отбивать броски и выполнять точные подачи. Именно Бейбу Руту до сих пор принадлежит рекорд Высшей лиги «процент отбивания, достигнутый на протяжении всей карьеры бейсболиста»: 0,690⁴.

Теперь обратимся ко второму вопросу: что происходит с экономическим благополучием американского среднего класса? Как и в первом случае, я поинтересовался мнением экспертов, обратившись по электронной почте к Джеффу Гроггеру (моему коллеге по Чикагскому университету) и Алану Крюгеру (вы, наверное, помните: именно он изучал причины терроризма, а в настоящее время занимает пост председателя Совета экономических консультантов Барака Обамы). Ни тот ни другой не смог дать мне однозначного ответа на этот вопрос. Чтобы оценить экономическое благополучие американского среднего класса, нам следует проанализировать изменения медианной заработной платы (с поправкой на инфляцию) за последние несколько десятилетий. Кроме того, они порекомендовали проанализировать изменения величины заработных плат в 25-м и 75-м процентилях (есть все основания интерпретировать их как верхнюю и нижнюю границы для среднего класса).

Стоит также упомянуть еще об одном различии. При оценивании экономического благосостояния мы можем анализировать доход или заработную плату. Это не одно и то же. Заработная плата — это то, что нам платят за некое фиксированное количество труда (например, она может быть почасовой или понедельной). Доход представляет собой сумму всех платежей из разных источников. Если у работника есть вторая работа или он отработал большее количество часов, его доход может увеличиться, тогда как заработная плата останется прежней. (Именно поэтому доход может расти даже в случае, когда заработная плата снижается, — при условии, что работник трудится дольше.) Если, однако, работнику приходится больше работать, чтобы больше получать, то оценить, как это скажется на его благосостоянии, довольно сложно. Заработная плата является менее неоднозначным показателем того, как оплачивается труд американцев; чем она выше, тем больше человек получает за каждый час, проведенный на работе.

В дополнение к вышесказанному я привожу график заработной платы американцев за последние три десятилетия. Я также добавил 90-й процентиль, чтобы проиллюстрировать изменения заработной платы работников,

относящихся к среднему классу, в сравнении (за тот же период времени) с заработной платой работников, находящихся на вершине этого распределения.



Источник: Changes in the Distribution of Workers' Hourly Wages between 1979 and 2009, Congressional Budget Office, 16 февраля 2011 года. Данные для этой диаграммы можно найти на сайте <http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/120xx/doc12051/02-16-wagedispersion.pdf>

На основе этих данных можно сделать немало выводов. Они не позволяют получить единственный «правильный» ответ на вопрос о том, в какую сторону изменяется экономическое благополучие американского среднего класса, зато четко показывают, что типичный американский рабочий, получающий медианную заработную плату, на протяжении почти тридцати лет «топчется на месте». Работники в 90-м процентиле добились за это время гораздо больших успехов. Описательные статистики помогают очертить проблему. Какие именно действия мы предпримем в ответ на это (если вообще предпримем) — вопрос сугубо идеологический и политический.

* * *

Приложение к главе 2

Данные для графического отображения дефектов принтера

	0	1	2	3	4	5	6	7	8	9	10 или больше
Частота дефектов у конкурента	12	14	36	13	8	6	5	3	0	2	1
	0	1	2	3	4	5	6	7	8	9	10 или больше
Частота ваших дефектов	25	31	9	4	3	0	0	1	1	0	26

Формула для дисперсии и среднеквадратического отклонения

Дисперсия и среднеквадратическое отклонение — самые распространенные статистические механизмы для измерения и описания разброса того или иного распределения. Дисперсия, которая часто обозначается символом σ^2 , вычисляется путем определения, насколько далеко от среднего значения расположены наблюдения в рамках того или иного распределения. Однако фишка в данном случае состоит в том, что расстояние (разница) между каждым наблюдением и средним значением возводится в квадрат; сумма таких составляющих, возведенных в квадрат, затем делится на количество наблюдений.

А именно:

Для любой совокупности из n наблюдений $x_1, x_2, x_3 \dots x_n$
со средним значением μ
Дисперсия = $\sigma^2 = [(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots (x_n - \mu)^2] / n$

Поскольку разница между каждым членом и средним значением возводится в квадрат, формула для вычисления дисперсии присваивает определенный вес наблюдениям, которые расположены вдали от среднего значения (то есть «отщепенцам»), как показано в приведенной ниже таблице роста учащихся.

Группа 1	Рост ($\mu = 70$ дюймов)	Расстояние от среднего = абсолютному значению ($x_n - \mu$)*	($x_n - \mu$) ²	Группа 2	Рост ($\mu = 70$ дюймов)	Расстояние от среднего = абсолютному значению ($x_n - \mu$)*	($x_n - \mu$) ²
Ник	74	4	16	Сахар	65	5	25
Элана	66	4	16	Мэгги	68	2	4
Дайна	68	2	4	Фейсал	69	1	1
Ребекка	69	1	1	Тед	70	0	0
Бен	73	3	9	Джефф	71	1	1
Чару	70	0	0	Нарцисо	75	5	25
		Итого = 14	Итого = 46		Итого = 14		Итого = 56
			Дисперсия = = $46/6 = 7,7$				Дисперсия = = $56/6 = 9,3$
			Стандартное отклонение = = $\sqrt{7,7} = 2,8$				Стандартное отклонение = = $\sqrt{9,3} = 3$

* Абсолютное значение — это расстояние между двумя числами, независимо от знака разности между ними, то есть это значение всегда положительное. В данном случае оно представляет собой разницу в дюймах между ростом конкретного человека и средним значением.

Средний рост обеих групп учащихся составляет 70 дюймов. Суммы абсолютных отклонений от среднего значения в обеих группах также одинаковы — 14 дюймов. По этому показателю разброса указанные два распределения идентичны. Однако дисперсия для группы 2 оказалась выше из-за веса, присвоенного в формуле дисперсии значениям, которые расположены особенно далеко от среднего значения (в нашем случае эти значения относятся к Сахár и Нарцисо).

Дисперсия сама по себе редко используется в качестве описательной статистики. В наибольшей степени она полезна как один из шагов в направлении вычисления среднеквадратического (стандартного) отклонения интересующего нас распределения, которое, как описательная статистика, является более интуитивно понятным инструментом.

Среднеквадратическое отклонение для совокупности наблюдений представляет собой корень квадратный из дисперсии:

$$\begin{aligned} &\text{Для любой совокупности из } n \text{ наблюдений } x_1, x_2, x_3 \dots x_n \\ &\text{со средним значением } \mu \text{ среднеквадратическое отклонение} = \\ &= \sigma = \text{корню квадратному из этой величины} = \\ &\sqrt{[(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots (x_n - \mu)^2]/n}. \end{aligned}$$

**«Он — выдающаяся личность!»
и другие истинные, но вводящие
в заблуждение утверждения**

Каждого, кому когда-либо приходилось выбирать себе спутника жизни, фраза «Он — выдающаяся личность!» обычно заставляет насторожиться — и вовсе не потому, что такое описание не соответствует действительности, а потому, что за подобным заявлением человек может что-то *скрывать*, например факт отсидки в тюрьме или «не до конца» оформленный развод с бывшей женой. Мы не сомневаемся, что этот парень и впрямь выдающаяся личность, но беспокоимся о том, чтобы справедливое в принципе утверждение не использовалось в качестве ширмы с целью замаскировать информацию, выставляющую лицо, о котором идет речь, в неприглядном свете, и тем самым не вводило нас в заблуждение (предполагается, что большинство женщин предпочло бы не встречаться с бывшими уголовниками и брачными аферистами). Утверждение «Он — выдающаяся личность!» само по себе не является ложью (то есть это не повод обвинить в лжесвидетельстве), тем не менее оно может быть настолько неточным, что в конечном счете не будет соответствовать действительности.

То же самое касается и статистики. Несмотря на то что статистика как область знаний коренится в математике, а математика, как известно, относится к числу точных наук, использование статистики для описания сложных явлений не может быть точным. Это оставляет немалый простор для манипуляций и искажения реального положения вещей. Марк Твен сказал однажды фразу, ставшую впоследствии знаменитой: «Есть три вида лжи: ложь, наглая ложь и статистика»*. Как объясняется в предыдущей главе, большинство явлений можно описать множеством разных способов. Если существуют разные способы описания одного и того же явления (например,

* Марк Твен приписывал эти слова британскому премьер-министру Бенджамину Дизраэли; впрочем, каких-либо документальных свидетельств, подтверждающих авторство Дизраэли, не обнаружено.

«он — выдающаяся личность» или «он был осужден за мошенничество с ценными бумагами»), то описательные статистики, которые мы используем (или не используем) при этом, будут оказывать огромное влияние на итоговое впечатление. Кто-то из гнусных побуждений может обыграть даже самые невинные факты и численные показатели ради весьма сомнительных выводов, не имеющих ничего общего с реальной ситуацией.

Для начала давайте определим разницу между такими понятиями, как «точность» и «достоверность». Они не взаимозаменяемы. Словом «точность» мы обозначаем *математическую* точность того или иного явления. В описании протяженности вашего маршрута от дома до работы значение 41,6 мили будет более точным, чем «примерно 40 миль», которое, в свою очередь, намного точнее словосочетания «этот чертовски долгий путь на работу». Если вы спросите меня, как далеко до ближайшей автозаправки, я отвечу, что до нее 1,265 мили на восток. Это будет *точный* ответ. Но есть один нюанс: он может оказаться совершенно неточным, если вы ошибетесь в определении направления движения и поедете не строго на восток, а слегка отклонитесь. С другой стороны, если я скажу вам: «Едьте примерно десять минут, пока не увидите закусочную, а еще через пару сотен ярдов справа будет АЗС. Но если на вашем пути встретится ресторанчик Hooters, значит, вы уже проскочили автозаправку», то мой ответ окажется менее точным, чем «1,265 мили на восток», но более содержательным и полезным, поскольку я указал вам путь именно в направлении АЗС. Достоверность — это показатель того, соответствует ли истине рассматриваемое численное значение. Отсюда опасность путаницы между точностью и достоверностью. Если какой-либо ответ достоверный (правильный), то чем больше точность, тем, как правило, лучше. Однако даже самая высокая точность не в состоянии компенсировать недостоверности ответа.

На самом деле точность может маскировать — случайно или вполне намеренно — недостоверность, вызывая у нас ложное ощущение определенности. Паранойя, охватившая Джозефа Маккарти, сенатора от штата Висконсин и яркого антикоммуниста, достигла своего апогея в 1950 году, когда он не только утверждал, что в Госдепартамент США внедрились коммунисты, но и доказывал, что располагает поименным списком этих людей. Во время своего выступления в г. Уиллинг Маккарти потрясал в воздухе листком бумаги, заявляя: «Я держу в руке список из 205 фамилий членов Коммунистической партии. Они известны госсекретарю. Тем не менее эти люди

продолжают работать в Госдепе, более того, они формируют внешнюю политику страны!»¹ Впоследствии выяснилось, что Маккарти держал в руке чистый листок бумаги, однако указание точного числа (205) придавало словам сенатора бóльшую достоверность, несмотря на столь наглую ложь.

Я уяснил важное различие между точностью и достоверностью в менее негативном контексте. Однажды жена подарила мне на Рождество лазерный дальномер, чтобы я мог определять на поле для гольфа расстояние от мяча до лунки. Расстояние измеряется посредством лазерного луча: я становлюсь рядом с мячом на гладком поле (или неровной площадке) и навожу устройство на флажок, установленный на лужайке; при этом дальномер вычисляет расстояние, на которое мне предстоит отправить мяч. Это считается более удобным способом, чем стандартные маркеры, обозначающие расстояния в ярдах и только до центра лужайки (таким образом, маркеры позволяют получить правильный, но менее точный результат). С помощью дальномера я мог, например, узнать, что нахожусь в 147,2 ярда от лунки. Я рассчитывал, что точность, обеспечиваемая этой продвинутой технологией, улучшит мои результаты во время игры в гольф. Однако в действительности они заметно ухудшились.

У меня возникли две проблемы. Во-первых, я пользовался этим глупым устройством три месяца, прежде чем до меня дошло, что оно измеряет расстояния не в ярдах, а в метрах; таким образом, каждое точное вычисление (147,2) было неправильным. Во-вторых, иногда я непреднамеренно наводил лазерный луч на деревья позади лужайки, а не на флажок, отмечающий лунку, в результате чего мой «идеальный» удар преодолевал именно то расстояние, которое и должен был преодолеть: мяч пролетал над лужайкой и оказывался в лесу. Урок, который я извлек, касается всего статистического анализа и заключается в том, что даже самые точные измерения или вычисления не должны противоречить здравому смыслу.

Рассмотрим пример ситуации с более серьезными последствиями. Многие из моделей управления рисками, использовавшиеся на Уолл-стрит до финансового кризиса 2008 года, были довольно точными. Концепция «рисковой стоимости» (VaR) позволяла компаниям точно вычислить величину своего капитала, которая может быть потеряна в случае реализации тех или иных сценариев. Проблема состояла в том, что такие сверхсложные модели были эквивалентны настройке моего дальномера в метрах, а не в ярдах. Используемая в этом случае математическая модель была сложной и запутанной.

Ответы, которые можно было получить с ее помощью, казались обнадеживающе точными. Однако предположения относительно того, что может случиться с глобальными рынками, встроенными в эти модели, были изначально неверными, в результате чего выводы, полученные с помощью этих моделей, были совершенно неправильными, что привело к дестабилизации не только Уолл-стрит, но и всей мировой экономики.

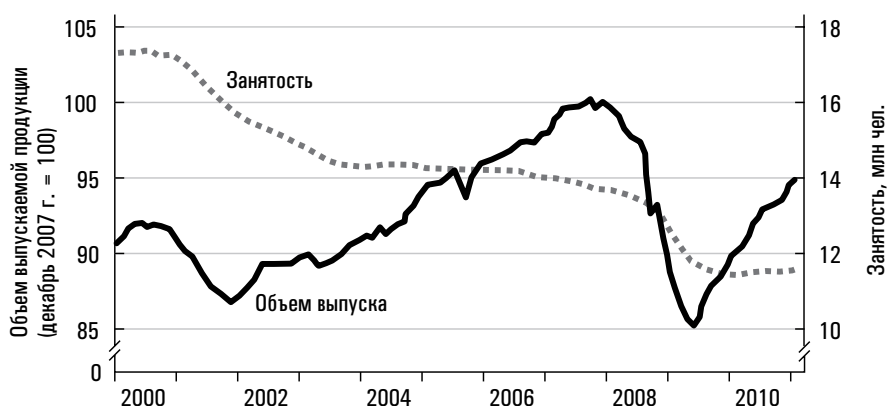
Даже самые точные описательные статистики могут стать жертвой более фундаментальной проблемы: недостаточной ясности того, что именно мы пытаемся определить, описать или объяснить. У статистических рассуждений и доказательств очень много общего с неудачными браками: участники дискуссии просто не понимают друг друга. Рассмотрим важный экономический вопрос: насколько успешны американские производственные отрасли? Нередко приходится слышать, что количество рабочих мест в них резко сокращается в результате появления новых рабочих мест в Китае, Индии и других странах с низким уровнем заработной платы. Также нередко приходится слышать, что в Соединенных Штатах высокотехнологичное производство по-прежнему процветает и Америка остается одним из ведущих мировых экспортеров товаров промышленного производства. Что же происходит на самом деле? Похоже, это тот случай, когда скрупулезный анализ надежных исходных данных мог бы примирить между собой эти противоречащие друг другу утверждения. Остаются ли американские производственные отрасли прибыльными и конкурентоспособными в глобальном масштабе или проигрывают в борьбе с сильными зарубежными конкурентами?

Верно и то и другое. Британскому журналу экономических новостей *The Economist* удалось примирить эти две полярные точки зрения на ситуацию в американских производственных отраслях с помощью приведенного ниже графика.

Кажущееся противоречие обусловлено разной трактовкой «благополучия» в американских производственных отраслях. По объему выпускаемой продукции — то есть общему количеству произведенных и проданных товаров — производственный сектор США демонстрировал неуклонный рост в начале 2000-х годов, затем испытал серьезный удар во время Великой рецессии, а теперь уверенно компенсирует потери. Это согласуется с данными, приведенными в справочнике-альманахе ЦРУ *The World Factbook* («Всемирная книга фактов»), которые показывают, что Соединенные Штаты являются

третьим по величине экспортером продукции промышленного производства в мире (после Китая и Германии), то есть по-прежнему остаются одним из мировых локомотивов производства.

Из статьи *The Rustbelt Recovery* (журнал *The Economist*, 10 марта 2011 года)



Но на графике, приведенном в журнале *The Economist*, есть еще одна линия, отражающая уровень *занятости* в производстве. Количество рабочих мест в производственной сфере США неуклонно сокращалось: за последнее десятилетие примерно на шесть миллионов. Указанные две тенденции — рост объема выпускаемой продукции и сокращение занятости — в совокупности объясняют реальную ситуацию в американских производственных отраслях. В производственной сфере Соединенных Штатов наблюдается стойкий рост производительности труда, а это означает, что заводы выпускают все больший объем продукции силами все меньшего числа работников. Это хорошо с точки зрения глобальной конкуренции, поскольку делает американскую продукцию более конкурентоспособной по сравнению с товарами, выпускаемыми в странах с низким уровнем заработной платы. (Одним из способов успешно конкурировать с компанией, выплачивающей работникам 2 доллара в час, является создание эффективного производственного процесса, где один работник, зарабатывающий 40 долларов в час, может делать в двадцать раз больше.) Однако это сопровождается сокращением рабочих мест в производственной сфере, что становится настоящим ударом для тех, кого уволили.

Поскольку моя книга посвящена статистике, а не проблемам в производственной сфере Соединенных Штатов, вернемся все же к главному вопросу: почему «благополучие» в американских производственных отраслях — показатель, который на первый взгляд не так уж сложно вычислить, — зависит от того, *чем* именно мы его определяем: объемом выпускаемой продукции или уровнем занятости? В данном случае (и во многих других) необходимо одновременно учитывать оба показателя, как и поступил The Economist, построив свой график.

Даже когда мы определяем единый показатель успеха (например результаты экзаменов в учебном заведении), все равно остается большой простор для статистических колебаний. Проверьте, можете ли вы примирить между собой два приведенных ниже гипотетических утверждения, причем оба вполне могут быть правильными.

Политик А (оппозиционер, критикующий существующее положение вещей): «Наша система школьного образования деградирует! Шестьдесят процентов наших школ продемонстрировали в этом году более низкие результаты экзаменов, чем в прошлом».

Политик В (должностное лицо, оправдывающее существующее положение вещей): «Наша система образования успешно развивается! Восемьдесят процентов наших учащихся продемонстрировали во время экзаменов в этом году более высокие результаты, чем в прошлом».

Подсказка: в разных школах обучается разное количество детей. Если взглянуть на эти утверждения, которые на первый взгляд кажутся взаимоисключающими, по-другому, то вы сразу заметите, что один политик использует в качестве *единицы анализа* школы («Шестьдесят процентов наших школ...»), а другой — учащихся («Восемьдесят процентов наших учащихся...»). Единица анализа — это объект, сравниваемый или описываемый посредством статистики; один из политиков говорит о деятельности школ, а другой — об успеваемости учащихся. Нет ничего нелогичного в том, что большинство учеников улучшают свои результаты, а большинство школ, напротив, ухудшают. Это происходит в случае, когда ученики, улучшающие свои результаты, обучаются в очень больших школах. Чтобы сделать этот пример более интуитивно понятным, выполним такое же по смыслу упражнение применительно к экономике американских штатов.

Политик А (популист): «Наша экономика катится в пропасть! В прошлом году уровень доходов в тридцати штатах сократился».

Политик В (противник популизма): «Наша экономика демонстрирует заметный рост: в прошлом году доходы семидесяти процентов американцев выросли».

Из этих утверждений напрашивается вывод, что самой благополучной можно считать экономику крупнейших штатов: Нью-Йорка, Калифорнии, Техаса, Иллинойса и т. п. Тридцатью штатами со снижающимся средним доходом, по-видимому, будут те, которые гораздо меньше по площади: Вермонт, Северная Дакота, Род-Айленд и т. п. Учитывая диспропорцию в величине штатов, вполне возможно, что экономическая ситуация во многих из них ухудшилась, тогда как доходы большинства американцев выросли. Главное — обратить внимание на единицу анализа. Кого именно (или что именно) мы пытаемся описать, и отличается ли этот «кто-то» (или это «что-то») от того, что пытается описать кто-то другой?

Хотя приведенные выше примеры относятся к категории гипотетических, ключевым здесь является отнюдь не гипотетический статистический вопрос: как влияет глобализация на неравенство доходов в мире в целом — в лучшую или в худшую сторону? По одной теории, глобализация лишь усугубляет существующее неравенство доходов: более богатые страны (если богатство измерять величиной ВВП [валовой внутренний продукт] на душу населения) демонстрировали более высокие темпы роста в период с 1980 по 2000 год, чем более бедные страны². Богатые страны становились еще богаче; из этого следовало, что торговля, аутсорсинг, зарубежные инвестиции и прочие компоненты глобализации — не что иное как инструменты, с помощью которых развитые страны укрепляют свою экономическую гегемонию. Короче говоря, долой глобализацию!

Но не будем торопиться с выводами. Те же данные можно (и нужно) интерпретировать совершенно по-другому, если изменить единицу анализа. Нас интересуют не бедные страны, а *бедные люди*. А самый высокий процент бедных людей в мире приходится на Китай и Индию. Китай и Индия — огромные страны (население каждой из них превышает миллиард человек); и обе были относительно бедными в 1980 году. В течение нескольких последних десятилетий они развивались ускоренными темпами, что в небольшой степени обуславливалось их возрастающей экономической интеграцией с остальным миром. The Economist описывает их как «быстрых глобализаторов». С учетом того, что наша цель — искоренить человеческую бедность, при анализе влияния глобализации на бедность нет смысла присваивать

Китаю (с населением 1,3 миллиарда человек) такой же вес, как Маврикию (с населением 1,3 миллиона человек).

Единицей анализа должны быть люди, а не страны. На самом деле то, что произошло в период с 1980 по 2000 год, во многом похоже на приведенный мною выше гипотетический пример со школами. Большая часть бедного населения планеты проживает в двух гигантских странах, которые в настоящее время бурно развиваются, все больше и больше интегрируясь в мировую экономику. Надлежащим образом выполненный анализ приводит нас к совершенно другому выводу относительно последствий глобализации для людей со скромными доходами. Как указывают авторы статьи в журнале *The Economist*, «если анализировать положение людей, а не стран, то глобальное неравенство стремительно сокращается».

Телекоммуникационные компании AT&T и Verizon недавно развязали «рекламную войну», в которой используется рассматриваемая нами двусмысленность в отношении того, что именно описывается. Обе компании предоставляют услуги сотовой связи. Одной из главных проблем для большинства пользователей мобильных телефонов является качество связи. Таким образом, кажется вполне логичным, чтобы в своих рекламных кампаниях оба мобильных оператора сравнивали масштаб и качество своих сетей связи. В то время как потребители просто заинтересованы в достойном качественном сервисе, AT&T и Verizon применяют *разные* показатели для оценивания этого несколько расплывчатого желания. Verizon запустила агрессивную рекламную кампанию, расхваливающую географическое покрытие, обеспечиваемое ее сетью связи: возможно, вы вспомнили географические карты Соединенных Штатов, показывающие, какую часть страны охватывает сеть связи этого оператора по сравнению с относительно скромным географическим покрытием, обеспечиваемым AT&T. Единицей анализа, выбранной Verizon, является площадь охватываемой ею территории, поскольку Verizon весьма преуспела именно в этом отношении.

AT&T ответила запуском рекламной кампании с другой единицей анализа. Билборды AT&T гласят, что «AT&T предоставляет услуги 97% американцев». Обратите внимание на использование слова «американцы», а не «Америка». AT&T сосредоточила внимание на том обстоятельстве, что большинство людей не проживает в сельскохозяйственном штате Монтана или в пустыне Аризона. Поскольку население неравномерно распределено по территории США, ключом к предоставлению качественных услуг сотовой связи

(как подразумевалось в данной рекламной кампании) является ее наличие в местах, где фактически проживают и работают потенциальные пользователи, а вовсе не обязательно там, где они могут проводить пару недель во время отпуска. Однако как человек, часто бывающий в сельскохозяйственном штате Нью-Гэмпшир, я отдаю свои симпатии компании Verizon.

Наши старые знакомые, среднее значение и медиана, также могут использоваться для всевозможных неблагоприятных целей. Как вы, наверное, помните из материала предыдущей главы, среднее значение и медиана — это показатели «середины» того или иного распределения, или его «центральная тенденция». Среднее значение — это просто арифметическое среднее: сумма наблюдений, поделенная на их количество (среднее значение чисел 3, 4, 5, 6 и 102 равняется 24). Медиана представляет собой среднюю точку распределения: половина наблюдений расположена над ней, а другая половина — под ней (медиана чисел 3, 4, 5, 6 и 102 составляет 5). Итак, умный читатель, наверное, обратил внимание на существенную разницу между 24 и 5. Если бы по какой-то причине я захотел описать эту группу чисел так, чтобы она показалась более внушительной, то отдал бы предпочтение среднему значению. Если же мне захотелось бы, чтобы она выглядела меньшей, то воспользовался бы медианой.

А теперь давайте посмотрим, как эти манипуляции осуществляются на практике. Рассмотрим снижение налогов, рекламируемое администрацией экс-президента Джорджа Буша как благо для большинства американских семей. Продвигая этот план, администрация Буша указывала, что для 92 миллионов американцев налоги в среднем уменьшатся на 1000 долларов (если быть более точным, то на 1083 доллара). Но является ли такая величина точной? Согласно The New York Times, «Эти данные не лгут, просто кое о чем умалчивают».

Снизилось бы налоговое бремя для 92 миллионов американцев? Да.

Уменьшились бы налоги для большинства из них примерно на 1000 долларов? Нет. Снижение налога, подсчитанное как медиана, оказалось бы меньше 100 долларов.

Сокращение налогов для относительно малого числа очень богатых людей оказалось бы очень существенным; именно эти большие числа искажают среднее значение, создавая иллюзию значительного снижения налогового бремени. В действительности величина такого снижения для большинства американцев оказалась бы гораздо меньшей. Медиана нечувствительна

к наблюдениям-«отщепенцам» и в данном случае стала бы более точным описанием того, как планируемые налоговые послабления сказались бы на типичной американской семье.

Разумеется, медиана также способна вводить в заблуждение — *именно потому, что нечувствительна к наблюдениям-«отщепенцам»*. Допустим, у вас обнаружили смертельную болезнь. Утешением для вас служит тот факт, что недавно появилось новое лекарство, излечивающее это заболевание. Плохо лишь то, что оно чрезвычайно дорогое и, кроме того, имеет множество опасных побочных эффектов. «Но поможет ли мне это лекарство?» — спрашиваете вы у врача. И он сообщает вам, что оно повышает медианную ожидаемую продолжительность жизни на... две недели. Подобная новость вряд ли добавит вам оптимизма (учитывая расходы на покупку лекарства и возможные побочные эффекты). К тому же ваша страховая компания отказывается оплачивать лечение по причине очень незначительного повышения медианной ожидаемой продолжительности жизни людей, страдающих вашим заболеванием.

Однако медиана вполне может оказаться весьма обманчивой статистикой в данном случае. Допустим, новое лекарство не помогает многим пациентам, однако немалое их число, скажем 30 или 40%, излечивается полностью. Этот процент успеха никак не сказывается на медиане (хотя средняя ожидаемая продолжительность жизни людей, принимающих новое лекарство, выглядела бы весьма впечатляюще). В этом случае наблюдения-«отщепенцы» — те, кому помогло новое лекарство, — должны сыграть важную роль в принятии вами окончательного решения. И это не просто некая гипотетическая ситуация. У Стефена Гоулда — ученого-биолога, занимающегося проблемами эволюции — была диагностирована форма рака, при которой медианная ожидаемая продолжительность жизни составляла восемь месяцев; спустя двадцать лет он умер от другого вида ракового заболевания, никак не связанного с предыдущим³. Гоулд впоследствии написал знаменитую статью под названием *The Median Isn't the Message* («Медиана — это не приговор»), в которой утверждал, что именно его научные познания в области статистики уберегли его от ошибочного заключения, будто он непременно умрет через восемь месяцев. Определение медианы говорит нам, что половина пациентов проживет по меньшей мере восемь месяцев — и, возможно, гораздо дольше этого срока. Распределение смертности «скошено вправо», а это — нечто гораздо большее, чем просто техническая подробность, когда речь идет о смертельной болезни⁴.

В данном примере определяющая характеристика медианы — то есть то, что она не присваивает наблюдениям весовые коэффициенты исходя из того, *насколько* они *отдалены* от средней точки, а лишь оценивает их в зависимости от того, где (выше или ниже) они расположены, — оказывается ее слабым местом. В отличие от медианы среднее значение *зависит* от разброса наблюдений. С точки зрения точности, ответ на вопрос «медиана или среднее значение» будет обусловлен тем, какое влияние оказывают наблюдения-«отщепенцы» в рассматриваемом нами распределении на описываемое явление: искажают его или, напротив, играют важную роль в уяснении нами его сути. (И снова здравое суждение берет верх над «голой» математикой.) Разумеется, ничто не скажет вам наверняка, чему именно следует отдать предпочтение — медиане или среднему значению. В любом комплексном статистическом анализе, скорее всего, будут задействованы оба показателя. Когда вы встречаете ссылку лишь на медиану или среднее значение, это наверняка было сделано из соображений краткости, хотя может указывать и на то, что кому-то очень хочется с помощью статистики «убедить» вас в чем-то.

Те из вас, кто достиг определенного возраста, возможно, помнят приведенный ниже обмен репликами между персонажами фильма *Caddyshack*, в роли которых выступают Чев Чейз и Тед Найт. Эти двое встречаются в раздевалке после игры в гольф.

Тед Найт: Сколько очков ты выбил?

Чев Чейз: Я не подсчитывал.

Тед Найт: Как же ты в таком случае сравниваешь себя с другими гольфистами?

Чев Чейз: По росту.

Я не буду объяснять, почему это должно быть смешно. Скажу лишь, что множество статистических манипуляций являются следствием сравнения «яблока и апельсинов». Допустим, вы пытаетесь сравнить цену гостиничного номера в Лондоне с ценой гостиничного номера в Париже и просите своего шестилетнего сынишку выполнить небольшое исследование в интернете, поскольку у него это получается гораздо быстрее, чем у вас. Спустя какое-то время сын докладывает, что гостиничные номера в Париже стоят дороже, примерно 180 за одну ночь; аналогичный номер в Лондоне обойдется приблизительно в 150 за одну ночь.

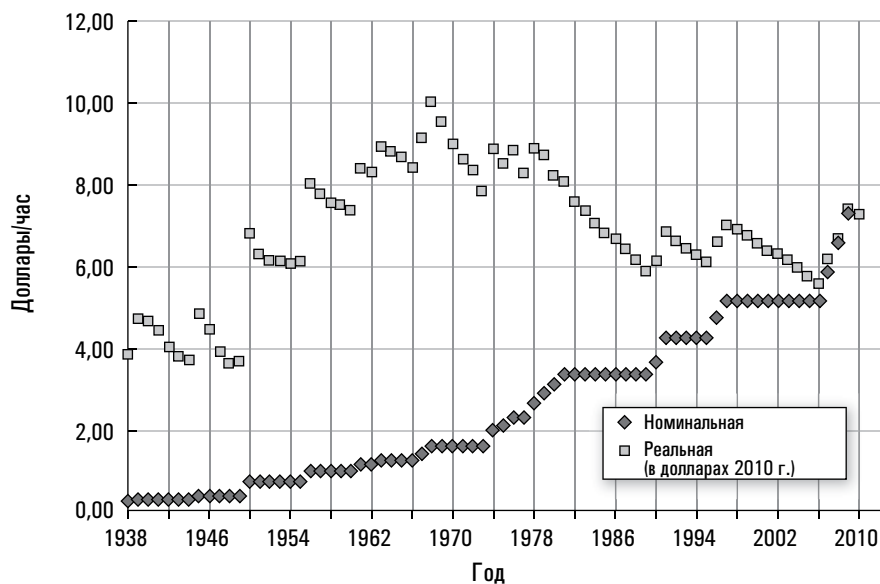
Скорее всего, вы объясните ребенку разницу между фунтами стерлингов и евро, а затем усадите его обратно за компьютер, чтобы выяснить обменные курсы этих валют и выполнить корректное сравнение цен. (Этот пример навеян моим собственным опытом: после того как я заплатил в Индии 100 рупий за чашку чая, моя дочь поинтересовалась, почему в Индии все настолько дорого.) Очевидно, сравнивать цены в разных странах, выраженные в соответствующих национальных валютах, бессмысленно, если не конвертировать их в сопоставимые денежные единицы. Каков обменный курс между фунтом стерлингов и евро или, в случае Индии, между долларом и рупией?

На первый взгляд это кажется совершенно очевидным, между тем попытки сопоставлять несопоставимое встречаются сплошь и рядом. Особенно это любят делать политики и студии Голливуда. Эти люди, конечно же, понимают разницу между фунтами стерлингов и евро, однако игнорируют менее очевидный пример «яблоко и апельсинов» — инфляцию. Нынешний доллар и доллар, каким он был шестьдесят лет назад, — это далеко не одно и то же: покупательная способность нынешнего доллара гораздо ниже. Вследствие инфляции товар, который стоил 1 доллар в 1950 году, стоил бы 9,37 доллара в 2011-м. В результате любые монетарные сравнения ситуации в 1950 году и в 2011 году без учета поправки на изменение стоимости доллара оказались бы даже менее точными, чем сравнение цен в фунтах стерлингов и евро, *поскольку фунты стерлингов и евро по своей стоимости сейчас гораздо ближе друг к другу, чем доллар 1950 и 2011 годов.*

Это настолько важное явление, что экономисты придумали специальные термины, указывающие, была ли внесена поправка на инфляцию или нет. Номинальные величины не скорректированы с учетом поправки на инфляцию. Сравнивая номинальную стоимость какой-либо государственной программы в 1970 году с номинальной стоимостью такой же государственной программы в 2011 году, мы просто сопоставляем величины чеков, выписанных Казначейством США в эти два года — без учета того обстоятельства, что покупательная способность доллара в 1970 году была выше, чем в 2011-м. Если в 1970 году мы потратили 10 миллионов долларов на некую программу оказания помощи ветеранам войны, а в 2011-м на такую же программу израсходовано 40 миллионов долларов, то в действительности это означает, что *федеральное правительство снизило выплаты по этой программе.* Да, суммы помощи повысились в номинальном выражении, однако это не отражает изменения стоимости долларов, затрачиваемых на ее оказание. Один

доллар в 1970 году эквивалентен 5,83 доллара в 2011-м. В 2011 году правительству нужно было бы потратить на реализацию программы помощи ветеранам войны 58,3 миллиона долларов, чтобы обеспечить им поддержку, сопоставимую с 1970 годом.

Реальные величины, в отличие от номинальных, учитывают поправку на инфляцию. Чаше величины приводят к какой-то одной единице, например долларам 2011 года, после чего становится возможным сравнение «яблок и апельсинов». На многих сайтах, включая сайт Бюро статистики труда (Министерства труда США), есть простые калькуляторы инфляции, которые позволяют сравнивать стоимость доллара в разные временные периоды*. Чтобы получить реальное представление о том, насколько может разниться статистика с поправкой и без поправки на инфляцию, рассмотрим приведенную ниже диаграмму изменения минимальной заработной платы на федеральном уровне США. На этой диаграмме представлены как номинальная величина минимальной заработной платы, так и ее реальная покупательная способность в долларах 2010 года.



Источник: <http://oregonstate.edu/instruct/anth484/minwage.html>.

* См. на сайте http://www.bls.gov/data/inflation_calculator.htm.

Минимальная заработная плата на федеральном уровне — показатель, который доводится до вашего сведения с помощью доски объявлений, вывешенной в каком-нибудь дальнем углу вашего офиса, — устанавливается Конгрессом США. Эта величина (в настоящее время составляющая 7,25 доллара) является номинальной. Ваш начальник не обязан гарантировать, что за 7,25 доллара вы купите такие же товары, как два года тому назад; он лишь должен гарантировать, что за каждый час работы вы получите не меньше этой суммы. Это лишь число, изображенное на чеке, а вовсе не то, что вы сможете приобрести за 7,25 доллара.

С течением времени инфляция снижает покупательную способность минимальной заработной платы (как и любой другой номинальной заработной платы; именно поэтому профсоюзы выступают за «поправки на рост стоимости жизни»). Если цены растут быстрее, чем Конгресс повышает минимальную почасовую заработную плату, ее реальная стоимость будет снижаться. Тем, кто обязан поддерживать минимальную заработную плату на должном уровне, следует учитывать ее реальную стоимость (поскольку закон должен гарантировать низкооплачиваемым работникам некий минимальный уровень потребления за каждый час работы), а не давать работнику чек, на котором указано некое число, не обеспечивающее ему даже минимальный прожиточный уровень. (В таком случае мы могли бы оплачивать труд низкооплачиваемых работников в рупиях.)

На мой взгляд, голливудские киностудии отличаются самым вопиющим игнорированием искажений, вносимых инфляцией, при сравнении доходов от разных фильмов в различные периоды времени (возможно, они делают это намеренно). Как, например, выглядит пятерка самых кассовых (на внутреннем рынке США) фильмов всех времен по состоянию на 2011 год?⁵

1. «Аватар» (2009)
2. «Титаник» (1997)
3. «Темный рыцарь» (2008)
4. «Звездные войны. Эпизод IV» (1977)
5. «Шрек 2» (2004)

Этот список не кажется вам несколько подозрительным? Все это вполне достойные фильмы — но «Шрек 2»? Неужели «Шрек 2» имел больший коммерческий успех, чем «Унесенные ветром», или «Крестный отец», или «Челюсти»? Нет, нет и еще раз нет! Голливуд хотел бы создать у нас впечатление,

что каждый его очередной блокбастер грандиознее и прибыльнее предыдущего. Один из способов сделать это — подсчитывать кассовые поступления в индийских рупиях, инспирируя таким образом газетные заголовки наподобие этого: «Недельный доход от проката Гарри Поттера составил 1,3 триллиона!» Но даже самые недалекие завсегдатаи кинотеатров с недоверием воспримут эти космические показатели дохода, потому что они выражаются в валюте с относительно низкой покупательной способностью (индийских рупиях). Несмотря на это, голливудские киностудии (и журналисты, освещающие их деятельность) просто используют номинальные величины, что создает впечатление необычайной коммерческой успешности последних кинолент Голливуда. Между тем впечатляющие показатели кассовых сборов, которыми так любит хвастаться Голливуд, в значительной мере объясняются тем, что нынешняя цена билета в кинотеатр существенно выше, чем, скажем, десять, двадцать или пятьдесят лет назад (когда в 1939 году «Унесенные ветром» впервые вышли на экраны страны, цена билета равнялась примерно 0,5 доллара). Наиболее точным способом сравнения коммерческого успеха фильмов, создававшихся в разные годы, было бы внесение в цену билета поправки на инфляцию. Добиться кассовых сборов порядка 100 миллионов долларов в 1939 году означает гораздо больший коммерческий успех, чем заработать 500 миллионов долларов в 2011-м. Как выглядела бы пятерка самых успешных с коммерческой точки зрения американских фильмов за всю историю существования кино в США *с поправкой на инфляцию*?⁶

1. «Унесенные ветром» (1939)
2. «Звездные войны. Эпизод IV» (1977)
3. «Звуки музыки» (1965)
4. «Инопланетянин» (1982)
5. «Десять заповедей» (1956)

В реальных величинах «Аватар» оказывается на 14-м месте, а «Шрек» опускается на 31-е.

Даже сравнение яблок с апельсинами оставляет значительный простор для манипуляций. Как отмечалось в предыдущей главе, одна из важных задач статистики — описание количественных изменений, происходящих с течением времени. Растут ли налоги? Сколько чизбургеров мы продаем по сравнению с прошлым годом? Насколько сократилось содержание мышьяка в питьевой воде? Чтобы отразить эти изменения, мы часто

используем процентные показатели, поскольку они создают у нас ощущение масштаба и контекста. Мы понимаем, что значит снизить содержание мышьяка в питьевой воде на 22%, тогда как лишь немногим из нас известно, можно ли считать существенным изменением уменьшение наличия мышьяка в воде на один микрограмм (абсолютное сокращение). Процентные показатели не лгут, но могут создавать излишне преувеличенную картину. Одним из способов сформировать у людей впечатление резкого роста чего-либо является использование процентного изменения, стартующего с очень низкой начальной точки. Я проживаю в округе Кук, штат Иллинойс. Однажды я испытал настоящее потрясение, узнав, что часть моих налогов, направляемую на борьбу с туберкулезом в нашем округе, планируется повысить на 527%! Однако узнав, что это изменение будет стоить мне меньше одного сэндвича с индейкой, решил отказаться от участия в массовом митинге против повышения налогов (к тому времени окончательное решение о его проведении еще не было принято). Количество заболеваний туберкулезом в нашем округе очень невелико, и средства, направляемые на борьбу с этим заболеванием также незначительны. В газете *Chicago Sun-Times* указывалось, что для типичного домовладельца налоговая декларация (счет) увеличится с 1,15 до 6 долларов⁷. Исследователи иногда квалифицируют тот или иной показатель роста, указывая, что он отсчитывается «от низкой базы»; это означает, что любое повышение при этом будет выглядеть довольно значительным.

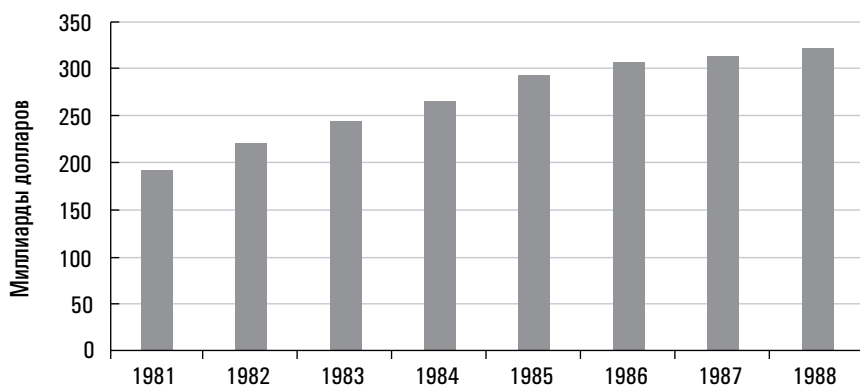
Очевидно, что справедливо и обратное. Даже небольшой процент от огромной суммы может выражаться большой абсолютной величиной. Допустим, министр обороны говорит, что расходы на его ведомство в этом году вырастут всего на 4%. Замечательно! Но не стоит радоваться, если принять во внимание, что бюджет Министерства обороны составляет примерно 700 миллиардов долларов. Четыре процента от этой суммы равны 28 миллиардам долларов — на такие деньги можно купить очень много сэндвичей с индейкой. В действительности это скромное на первый взгляд четырехпроцентное повышение бюджета оборонного ведомства *превышает бюджет НАСА и почти равняется совокупному бюджету Министерства труда и Казначейства США*.

Аналогично, ваш добросердечный и справедливый босс может объявить о повышении всем сотрудникам в этом году заработной платы на 10%. Какой великодушный жест! Правда, если вспомнить, что годовой доход вашего

босса составляет 1 миллион долларов, а ваш — 50 000 долларов, то окажется, что его годовой доход повысится на 100 000 долларов, а ваш — на 5000 долларов. Между тем заявление «В этом году заработная плата всех сотрудников повысится на 10%» звучит для вашего слуха гораздо приятнее, чем такие слова вашего босса: «Повышение моего годового дохода окажется в двадцать раз большим, чем вашего». Однако и то и другое не будет ложью.

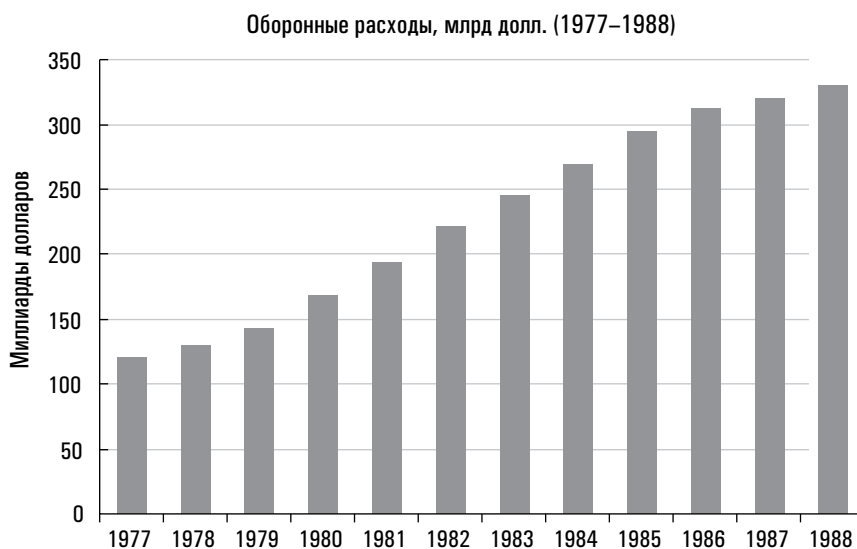
Любое количественное изменение, происходящее в течение какого-то времени, всегда имеет начальную и конечную точки. И этими точками можно манипулировать так, как того требуют обстоятельства. Один из моих преподавателей любил повторять, что у него есть «республиканские» и «демократические» слайды. Он имел в виду данные о расходах на оборону, а под слайдами подразумевал то, что готовясь к выступлению перед сторонниками Республиканской или Демократической партии, он мог организовать одни и те же данные таким образом, чтобы его выступление понравилось соответствующей аудитории. Выступая перед сторонниками Республиканской партии, он показал бы им слайды с данными о повышении оборонных расходов в период правления Рональда Рейгана. Разумеется, Рейган показал нам, насколько важно уделять должное внимание вопросам обороны и безопасности, что, в свою очередь, помогло нам выиграть холодную войну. Глядя на эти числа, невозможно не оценить по достоинству непреклонную решимость Рональда Рейгана запугать Советы.

Оборонные расходы, млрд долл. (1981–1988)



Для демократов мой бывший преподаватель использовал бы те же (номинальные) данные, но за более продолжительный отрезок времени. Выступая

перед этой аудиторией, он бы отметил, что именно Джимми Картер заслуживает уважения за наращивание оборонной мощи страны. Как следует из приведенного ниже «демократического» слайда, рост оборонных расходов с 1977 по 1980 год демонстрирует ту же базовую тенденцию, что и их рост в период правления Рональда Рейгана. Таким образом, нам остается лишь порадоваться тому, что Джимми Картер — выпускник Военно-морской академии США в Аннаполисе и бывший офицер ВМС — инициировал процесс возрождения военной мощи Америки!



Источник: <http://www.usgovernmentspending.com/spend.php?span=usgs302&year=1988&view=1&expand=30&expandC=&units=b&fy=fy12&local=s&state=US&pie=#usgs302>.

Хотя основная задача статистики — представить содержательную картину интересующих нас явлений, во многих случаях мы также рассчитываем опираться на эти данные. Командам NFL требуется какой-либо простой показатель эффективности действий куортербека, с помощью которого можно было бы находить и включать в свои составы талантливых игроков из высших учебных заведений. Компании оценивают эффективность действий своих работников, чтобы продвигать по службе тех, кто приносит им наибольшую пользу, и увольнять тех, кто абсолютно бесполезен. В бизнесе популярен такой афоризм: «Вы не можете управлять тем, что не в состоянии

измерить». И это действительно так. Однако вы должны быть твердо уверены в следующем: *то, что вы измеряете, действительно является тем, чем вы пытаетесь управлять.*

Рассмотрим вопрос качества школ, которое очень важно уметь измерять, поскольку, с одной стороны, это позволило бы поощрять и ставить в пример хорошие школы, а с другой — наказывать плохие и исправлять ситуацию в них. (А в рамках каждой школы перед нами стоит аналогичная цель — измерить качество преподавания.) Ключевым показателем качества школ и работы преподавателей являются результаты экзаменов. Если по итогам хорошо продуманного стандартизованного теста учащиеся демонстрируют впечатляющие баллы, то у нас есть все основания полагать, что учителя и школа отлично справляются со своей задачей. И наоборот, плохие результаты теста — это четкий сигнал о том, что многих преподавателей соответствующей школы нужно уволить, причем чем раньше, тем лучше. Такие статистические данные способны помочь улучшить государственную систему образования, не так ли?

Нет, не так. Любое оценивание школ и учителей, которое базируется исключительно на результатах экзаменов, представит очень неточную картину. У каждого учащегося свой жизненный опыт и способности, и они могут сильно разниться между собой. Нам известно, например, что уровень образования и величина дохода родителей ученика существенно влияют на его успеваемость, в какой бы школе он ни учился. Оказывается, только статистика, которой мы в данном случае не располагаем, а именно в какой мере успеваемость ученика (какой бы она ни была — хорошей или плохой) обуславливается происходящим в его школе (или классе, где он учится), имеет для нас значение.

Ученики из обеспеченных семей с высоким образовательным уровнем, как правило, демонстрируют хорошие результаты тестов буквально с первых и до последних дней учебы. Обратное также верно. Есть немало школ, где учатся исключительно дети из бедных семей и где преподаватели буквально творят чудеса, стремясь передать ученикам максимум знаний, но даже в этом случае трудно рассчитывать на хорошие баллы при сдаче тестов (хотя если бы в таких школах работали посредственные учителя, итоги экзаменов были бы еще плачевнее). Итак, нам необходим некий показатель «добавленной стоимости» на школьном уровне, или даже уровне класса. Нам не нужно знать абсолютный уровень успеваемости ученика — мы лишь хотим знать,

в какой мере его успеваемость определяется образовательными факторами, которые мы пытаемся оценить.

На первый взгляд это кажется довольно легкой задачей, поскольку мы можем просто предложить ученику сдать сначала предварительный тест, а затем — заключительный. Если нам известен результат предварительного теста, который сдается при поступлении в определенную школу или класс, то мы можем оценить успеваемость ученика в момент окончания учебы в этой школе или классе, а разницу результатов «до» и «после» объяснить полученными им знаниями.

Увы, опять ничего не выйдет. Ученики с разными способностями, к тому же выходцы из семей с полярно разными уровнями дохода и образования, могут и знания усваивать *с разной скоростью*. Кто-то из учеников схватывает материал буквально на лету, а кому-то приходится объяснять по несколько раз, причем все это не имеет никакого отношения к качеству преподавания. Таким образом, если ученики в Школе состоятельных родителей А и ученики в Школе бедных родителей В начинают изучать алгебру в одно и то же время и с одного и того же исходного уровня, то объяснить тот факт, что через год ученики школы А сдали экзамен по алгебре лучше, чем ученики школы В, можно либо тем, что в школе А преподают более квалифицированные учителя, либо тем, что в школе А учатся более способные ученики, быстрее усваивающие учебный материал, либо и тем и другим одновременно. Исследователи пытаются разработать статистические методы, которые при измерении качества преподавания учитывали бы способности учеников, а также материальное положение и образовательный уровень их родителей. А тем временем наши попытки выявить «наилучшие» школы могут оказаться до смешного несостоятельными.

Каждую осень несколько чикагских газет и журналов публикуют рейтинги лучших школ региона, основанные на результатах сдачи стандартизованного теста штата Иллинойс. Вот один из выводов, совершенно смехотворных с точки зрения статистики: поступление в несколько школ, постоянно занимающих высокие места в рейтинге, возможно лишь на конкурсной основе; для этого нужно предварительно подать соответствующие документы, причем в школу будет зачислена лишь малая часть из тех, кто их подал. Одним из важнейших критериев для поступления в такие школы являются результаты сдачи стандартизованных тестов. Итак, подведем итоги: 1) эти школы считаются «лучшими», потому что их ученики имеют высокие баллы

на экзаменах; 2) чтобы попасть в такую школу, нужно иметь высокие баллы стандартизованных тестов. Это, по сути, то же самое, как если бы вы наградили баскетбольную команду за то, что в ее составе выступают очень рослые ребята.

Даже при наличии надежного индикатора того, что вы пытаетесь измерить и чем пытаетесь управлять, проблемы не заканчиваются. Хорошей новостью будет то, что «управление посредством статистики» способно изменить к лучшему поведение соответствующего человека или учреждения. Если вы можете определить долю бракованных изделий, сходящих с производственного конвейера, и эти дефекты обусловлены ситуацией на заводе, то выплата работникам премии за сокращение количества бракованных изделий должна, по-видимому, надлежащим образом изменить их поведение. Каждый из нас реагирует на стимулы, даже если это просто похвала или предоставление более удобного места для парковки автомобиля. Статистика измеряет важные для нас результаты; стимулы подталкивают нас к их улучшению.

Или, в отдельных случаях, к приукрашиванию статистики. А вот это — плохо.

Если работа школьной администрации оценивается (и, возможно, даже оплачивается) исходя из процента учеников в определенном учебном округе, получивших аттестат об окончании школы, то ей следует сосредоточить усилия на увеличении количества выпускников. Разумеется, наряду с этим можно заняться и вопросом повышения доли учеников, окончивших школу (это не то же самое, что количество выпускников). Например, ученики, досрочно бросившие школу и не получившие аттестата, могут быть классифицированы как «сменившие место жительства», а не как бросившие учебу. Это вовсе не гипотетический пример; обвинение именно в таких манипуляциях было предъявлено бывшему министру образования Роду Пейджу во время его пребывания в должности школьного инспектора Хьюстона. Президент Джордж Буш назначил Рода Пейджа министром образования США под впечатлением его выдающихся успехов в Хьюстоне, суть которых заключалась в снижении доли учеников, досрочно бросивших школу, и резком улучшении результатов тестов.

Если вы коллекционируете афоризмы, могу поделиться собственным: «Если сегодня к вам в офис заглянули люди из программы 60 Minutes («60 минут»), то это определенно не лучший день в вашей жизни». Дэн Разер

и команда создателей программы 60 Minutes II побывали в Хьюстоне и пришли к выводу, что манипулирование статистикой в этом учебном округе производит гораздо большее впечатление, чем повышение уровня образования⁸. Учеников, бросающих учебу в школе, обычно включали в число тех, кто переводится в какую-то другую школу, возвращается к себе на родину (в другую страну) или желает получить General Equivalency Diploma (GED) — диплом об общем образовании, который выдается сдавшим тесты по программе средней школы. Ни один из этих вариантов не трактовался в официальной статистике как отказ от учебы в школе. В тот год администрация хьюстонского учебного округа рапортовала о снижении доли учеников, бросивших учебу в школе, до 1,5%. Хотя, согласно подсчетам 60 Minutes, этот показатель на самом деле находился между 25% и 50%.

Статистические манипуляции с тестовыми баллами были не менее впечатляющими. Один из способов добиться улучшения результатов тестов (в Хьюстоне или где-либо еще) — повысить качество образования, чтобы учащиеся углубляли свои знания и лучше сдавали экзамены. Это самый честный способ. Другой (менее честный) способ заключается в отстранении от их сдачи самых слабых учеников, поскольку в этом случае средний балл соответствующей школы или учебного округа повысится, даже если остальные ученики не продемонстрируют никакого прогресса. В Техасе единый тест штата проводится для десятиклассников. Есть свидетельства того, что руководство хьюстонских школ пыталось избавиться от отстающих учащихся еще до их перехода в десятый класс. В одном из особенно вопиющих случаев ученик провел три года в девятом классе, а затем его сразу перевели в одиннадцатый класс — такой вот хитроумный способ отстранения ученика от сдачи экзамена в десятом классе, не принуждая его бросить учебу (что плохо сказалось бы на другом статистическом показателе).

Был ли замешан Род Пейдж в этих статистических махинациях во время пребывания в должности школьного инспектора Хьюстона, выяснить не удалось, однако именно он добился внедрения программы строгой отчетности, которая предусматривала выплату денежных премий директорам школ, выполнявшим плановые показатели по результатам экзаменов и досрочному прекращению учебы в школе, и наказание вплоть до увольнения или понижения в должности директорам школ, не обеспечившим выполнение этих плановых показателей. Директора школ хорошо уяснили, что от них требуется, — и это должно послужить для нас еще одним важным уроком. Однако

нужно понимать, что те, чью деятельность пытаются оценивать подобными способами, не могут послушаться начальства, поскольку в противном случае рискуют предстать перед ним не в самом лучшем (со статистической точки зрения) виде.

Усвоение этой истины обошлось штату Нью-Йорк слишком дорого. Власти штата внедрили «оценочные таблицы», с помощью которых намеревались оценивать уровень смертности среди пациентов кардиохирургов, занимающихся коронарной ангиопластикой (восстановлением сосудов) — типичным способом лечения заболеваний сердца⁹. На первый взгляд такое использование описательной статистики кажется весьма разумным и полезным. Нам важно знать, какой процент пациентов кардиохирурга умирает в результате хирургической операции; государство должно иметь и обнародовать эту информацию, поскольку в противном случае у потенциальных пациентов не будет к ней доступа. Можно ли считать такую политику правильной? Да, если не принимать во внимание тот факт, что она способна убивать людей.

Кардиологи, конечно же, будут заботиться о состоянии своих «оценочных таблиц». Однако простейший способ, с помощью которого кардиохирург может сократить смертность, состоит вовсе не в стремлении сохранить жизнь как можно большему числу людей (у нас есть все основания полагать, что большинство врачей и без того делают в этом плане все от них зависящее), а в отказе оперировать самых тяжелых больных. Согласно результатам опроса, проведенного факультетом медицины и стоматологии Рочестерского университета, «оценочные таблицы», которые якобы служат благу пациентов, могут также приносить им вред: 83% опрошенных кардиохирургов сказали, что из-за оглашения данных о смертности часть пациентов, которые могли бы поправить здоровье с помощью ангиопластики, просто откажутся от такой операции; 79% кардиохирургов признались, что на некоторые их профессиональные решения повлияло знание того, что данные о смертности предаются огласке. Печальный парадокс этой, на первый взгляд полезной, описательной статистики заключается в том, что кардиохирурги реагировали на нее вполне рационально, отказываясь делать операции пациентам, которые больше всего в них нуждались.

Любой статистический индекс обладает всеми потенциальными подводными камнями, характерными практически для каждой описательной статистики, — плюс искажения, вносимые вследствие объединения нескольких индикаторов в единое обобщающее число. Любой индекс по определению

зависит от того, как именно он сконструирован; на него оказывает влияние и то, какие показатели в него входят, и то, какой весовой коэффициент присвоен каждому из этих показателей. Почему, например, рейтинг пасующего, которым принято оценивать эффективность пасующих в NFL, не включает какой-либо показатель «завершений с третьей попытки»? Если же мы говорим об индексе развития человеческого потенциала (Human Development Index), то каким должен быть вес уровня грамотности населения в этом индексе по сравнению с уровнем дохода на душу населения? И наконец, еще один немаловажный вопрос: должны ли простота и легкость применения, обеспечиваемые объединением многих показателей в одно число (индекс), иметь для нас большее значение, чем неточность, внутренне присущая такому объединению? Подчас приходится давать отрицательный ответ на этот вопрос, что возвращает нас (как и было обещано выше) к рейтингам высших учебных заведений, приведенным в журнале U.S. News & World Report (USNWR).

Для определения рейтингов USNWR используются шестнадцать показателей, с помощью которых оцениваются и распределяются в рейтинге по местам американские колледжи, университеты и профессиональные учебные заведения. Например, в рейтинге национальных университетов и гуманитарных колледжей за 2010 год на долю такого показателя, как «избирательный подход к приему в учебное заведение», приходилось 15% этого индекса; данный показатель, в свою очередь, вычислялся на основе нормы приема для той или иной школы, доли поступивших студентов, которые в выпускном классе своей школы входили в «лучшие 10%», а также средних баллов SAT* и АСТ** поступивших студентов. Преимущество рейтингов USNWR заключается в том, что они позволяют простым и доступным способом получить исчерпывающую информацию о тысячах учебных заведений. Даже критики вынуждены согласиться с тем, что большой объем информации об американских колледжах и университетах представляет немалую ценность. Потенциальные студенты должны знать о месте того или иного учебного заведения в рейтинге и средней величине учебной группы.

* SAT (Scholastic Aptitude Test) — тест на умение грамотно излагать свои мысли в устной форме и тест математических способностей, используемые при поступлении в американские колледжи. *Прим. перев.*

** АСТ (American College Testing) — стандартизированный тест для поступления в колледжи и университеты США. *Прим. перев.*

Разумеется, предоставление значимой информации имеет мало общего с ее объединением в общий индекс, который претендует на авторитетность. По мнению критиков, такие рейтинги неуклюже сконструированы, способны вводить в заблуждение и вредить долгосрочным интересам студентов. «Проблема в том, что это — не более чем перечень, в котором каждому учебному заведению присвоен определенный порядковый номер. Данные, на основе которых он определяется, не обеспечивают требуемой точности», — говорит Майкл Макферсон, бывший президент Макалистерского колледжа в Миннесоте¹⁰. Почему на долю пожертвований выпускников в пользу своего бывшего учебного заведения приходится 5% его рейтинга? И если этот показатель так важен, то почему на его долю не приходится, например, 10%?

Согласно U.S. News & World Report, «каждому показателю на основе наших собственных представлений о его важности присваивается определенный весовой коэффициент (выраженный в процентах)»¹¹. Представления — это что-то одно, а произвол и субъективизм — нечто другое. Показателем, имеющим в рейтинге национальных университетов и гуманитарных колледжей самый большой вес, является «научная репутация». Она определяется исходя из «опроса представителей научного сообщества» (анкету заполняют администраторы других колледжей и университетов) и опроса консультантов-психологов, работающих в школах. Критикуя рейтинги, публикуемые U.S. News & World Report, Малкольм Гладуэлл в пух и прах разносит (правда, с изрядной долей юмора) методологию «опроса представителей научного сообщества». Он цитирует опросник, разосланный бывшим председателем Верховного суда штата Мичиган примерно сотне юристов, в сопроводительном письме к которому тот просит юристов расположить десять юридических учебных заведений в порядке убывания их качества. Одним из таких вузов в списке был указан Пенсильванский университет: юристы поместили его примерно в середину перечня. Фишка в том, что *на тот момент Пенсильванский университет не имел юридического факультета*¹².

Какими бы обширными ни были собранные данные, вовсе не факт, что рейтинги USNWR измеряют именно то, что должно интересовать потенциальных студентов: какой объем знаний можно получить в том или ином учебном заведении. Футбольные болельщики могут спорить по поводу показателей, входящих в состав индекса расписовщика, но никто из них не станет отрицать того, что составные части этого индекса — коэффициент удачного завершения, среднее число ярдов на каждую попытку паса,

процент тачдаун-пасов на каждую попытку паса и коэффициент перехватов мяча — важная составляющая эффективности действий куортербека. Это вовсе не обязательно относится к критериям USNWR, большинство из которых фокусируется на исходных данных (например, какого рода учащихся принимают в учебное заведение, каков уровень оплаты преподавателей, какой процент штатных преподавателей), а не образовательных результатах. Двумя важными исключениями являются процент студентов первого курса, продолживших обучение, и процент выпускников, но даже они не позволяют оценить объем знаний, полученных студентом за время учебы. Как указывает Майкл Макферсон: «В действительности из рейтингов U.S. News & World Report невозможно понять, насколько за четыре года учебы в колледже или университете студенты фактически увеличили свои знания или развили способности».

Все это было бы довольно безобидным занятием, если бы его инициаторы не поощряли действия, не всегда направленные на благо учащихся и учебных заведений. Например, одним из статистических показателей, используемых для вычисления рейтингов, является величина финансовых ресурсов, выделяемых на обучение одного студента; проблема в данном случае — в отсутствии показателя, позволяющего оценить *эффективность использования* этих финансов. Учебное заведение, меньше расходующее средств в расчете на одного студента, автоматически ухудшает свой рейтинг даже в случае, если эти средства распределяются гораздо эффективнее, чем в других колледжах или университетах. Кроме того, колледжи и университеты заинтересованы в подаче документов как можно большим количеством абитуриентов — в том числе и теми, у кого практически нет шансов пройти по конкурсу, — поскольку высокий конкурс при поступлении также свидетельствует о престижности учебного заведения. Это оборачивается напрасной тратой ресурсов учебными заведениями и бесполезными затратами времени теми абитуриентами, которые изначально не имели никаких шансов на поступление.

Поскольку мы уже почти добрались до главы о вероятности, я готов биться об заклад, что с высокой вероятностью в обозримом будущем рейтинги, публикуемые U.S. News & World Report, по-прежнему будут пользоваться популярностью. Как сказал Леон Ботстейн, президент колледжа Bard College: «Люди предпочитают получать простые и легкие ответы. Какое место самое лучшее? Конечно же, первое»¹³.

Вывод, который можно сделать по прочтении этой главы, что «статистические преступления» не являются следствием математических ошибок. Скорее, наоборот: заумные математические расчеты подчас способны скрыть неблагоприятные намерения. То, что вы правильно подсчитали среднее значение, не отменяет тот факт, что медиана представляет собой более точный индикатор. Здравое суждение и честный подход к делу оказываются более важными условиями для выяснения истины. Глубокое знание статистики не мешает нечистым на руку людям манипулировать данными точно так же, как хорошее знание уголовного кодекса не мешает преступникам заниматься своими темными делишками. И в том и в другом случаях «плохие парни» зачастую очень хорошо понимают, *что* они делают!

Откуда Netflix известно, какие фильмы мне нравятся?

Netflix* утверждает, что мне точно понравится документальный фильм Bhutto, рассказывающий о жизни и трагической смерти бывшего пакистанского премьер-министра Беназир Бхутто. Возможно, мне действительно понравится этот фильм (я уже добавил его в список кинолент, которые собираюсь посмотреть). Прошлые рекомендации были просто потрясающими. К тому же когда Netflix советовала что-то из того, что я уже видел, то, как правило, фильм был из тех, которыми я действительно наслаждался.

Каким образом Netflix продельывает свои «фокусы»? Может быть, в штаб-квартире компании работает большое число стажеров, которые с помощью Google и опроса членов моей семьи и друзей «вычислили», что меня может заинтересовать документальный фильм о бывшем пакистанском премьер-министре? Конечно нет. Просто Netflix мастерски, со знанием дела использовала статистические данные. *Netflix не знакома со мной*. Но ей известно, какие фильмы мне понравились в прошлом (поскольку я выставлял им рейтинги). Воспользовавшись этой информацией наряду с рейтингами других кинозрителей и мощным компьютером, Netflix сумела сделать на удивление точные прогнозы относительно моих вкусов и предпочтений.

Я еще вернусь к алгоритму, который применила Netflix при составлении таких прогнозов, пока же достаточно будет сказать, что они основаны на корреляции. Netflix рекомендует фильмы, похожие на те, которые мне когда-то понравились или получили высокие оценки от других кинозрителей, чьи рейтинги подобны моим. Фильм Bhutto мне посоветовали потому, что в свое время я присвоил пятизвездочные рейтинги двум другим документальным фильмам: Enron: The Smartest Guys in the Room и Fog of War.

* Netflix — американская компания, поставщик фильмов и сериалов на основе потокового мультимедиа. *Прим. перев.*

Корреляция измеряет степень связи между двумя явлениями. Например, существует корреляция между летними температурами и продажей мороженого. Когда повышается температура, растут объемы продажи мороженого. Две переменные положительно коррелированы, если изменение одной переменной вызывает изменение другой в том же направлении, то есть в направлении увеличения или уменьшения (например, взаимосвязь между ростом и весом человека). У более высоких людей больший вес (в среднем); низкорослые люди весят меньше. Корреляция отрицательна, если положительное изменение одной переменной обуславливает отрицательное изменение другой (например, связь между регулярным выполнением физических упражнений и весом человека).

В зависимостях такого рода интересно то, что не каждое наблюдение вписывается в соответствующую схему. Иногда низкорослые люди весят больше, чем высокие. Иногда те, кто вообще не занимается спортом, бывают гораздо стройнее, чем те, кто регулярно выполняет физические упражнения. Тем не менее существует отчетливо выраженная связь между ростом и весом человека, а также между весом и физическими нагрузками.

Если построить диаграмму разброса данных, отражающих рост и вес произвольной выборки взрослых американцев, то получится примерно такая картина:



Если бы нам нужно было построить диаграмму разброса для данных о выполнении физических упражнений (количество минут, затрачиваемых на них каждую неделю) и данных о весе человека, то можно было бы ожидать отрицательной корреляции, причем те, кто занимается спортом больше времени, будут весить меньше. Однако картина в виде совокупности точек, разбросанных по определенной площади, представляет собой несколько неуклюжий инструмент. (Если бы Netflix попыталась предлагать мне какие-то фильмы, продемонстрировав диаграмму разброса рейтингов для тысяч кинолент, выставленных миллионами кинозрителей, то я посчитал бы такую рекомендацию просто неудачной шуткой.) Эффективность корреляции как статистического инструмента заключается в том, что мы можем выразить связь между двумя переменными с помощью одной описательной статистики — коэффициента корреляции.

Коэффициент корреляции обладает двумя чрезвычайно привлекательными характеристиками. Во-первых, в силу причин математического характера, которые мы обсудим в приложении, он представляет собой число в диапазоне от -1 до 1 . Корреляция, равная 1 (иногда ее называют идеальной корреляцией), означает, что каждому изменению одной переменной соответствует эквивалентное изменение другой переменной в том же направлении.

Корреляция, равная -1 (иногда ее называют идеальной отрицательной корреляцией), означает, что каждому изменению одной переменной соответствует эквивалентное изменение другой переменной в противоположном направлении.

Чем ближе корреляция к 1 или -1 , тем сильнее связь между переменными. Нулевая (или близкая к 0) корреляция говорит об отсутствии значимой связи между двумя переменными (например между результатом экзамена по математике и размером обуви экзаменуемого).

Второй привлекательной особенностью коэффициента корреляции является то, что с ним не связаны никакие единицы измерения. Мы можем рассчитать корреляцию между ростом и весом, несмотря на то что рост измеряется в дюймах, а вес — в фунтах. Мы можем даже вычислить корреляцию между количеством телевизоров, имеющихся дома у учеников, и результатами их экзаменов по математике (я почему-то уверен, что она окажется положительной). (Несколько ниже я остановлюсь подробнее на данной связи.) Коэффициент корреляции буквально творит чудеса: он сжимает сложное

сочетание данных, измеряемых в разных единицах (наподобие наших диаграмм разброса роста и веса), в единственную элегантную описательную статистику.

Как это удастся?

Как обычно, я привожу самую распространенную формулу для определения коэффициента корреляции в приложении, находящемся в конце этой главы. Это не та статистика, которую можно вычислить вручную. (После того как вы введете соответствующие данные, базовый программный пакет, например Microsoft Excel, рассчитает корреляцию между двумя соответствующими переменными.) Тем не менее на интуитивном уровне понять эту формулу несложно. Формула для вычисления коэффициента корреляции выполняет следующие операции:

1. Вычисляет среднее значение и стандартное (среднеквадратическое) отклонение для обеих переменных. Если вернуться к примеру с ростом и весом, то мы бы узнали средний рост людей в выборке, средний вес людей в той же выборке и стандартное отклонение для роста и веса.
2. Преобразует все данные таким образом, чтобы каждое наблюдение было представлено его расстоянием (в стандартных отклонениях) от среднего значения. Верьте мне, это совсем не сложно. Допустим, средний рост в выборке равняется 66 дюймам (при стандартном отклонении в 5 дюймов), а средний вес — 177 фунтов (при стандартном отклонении в 10 фунтов). Теперь предположим, что ваш рост — 72 дюйма, а вес — 168 фунтов. Мы можем также сказать, что ваш рост составляет 1,2 стандартного отклонения сверх среднего роста $[(72 - 66)/5] = 1,2$ и 0,9 стандартного отклонения ниже среднего веса, или $-0,9$ применительно к нашей формуле $[(168 - 177)/10] = -0,9$. *Да, это нетипично, когда рост человека выше среднего, а вес — ниже среднего, но поскольку вы уже заплатили неплохие деньги за эту книгу, то, как мне кажется, я должен в знак благодарности сделать вас высоким худощавым человеком.* Обратите внимание: ваш рост и вес, выражавшиеся поначалу в дюймах и фунтах, теперь выражаются абстрактными числами 1,2 и $-0,9$. Как видите, потребность в единицах измерения отпала.
3. Теперь я могу скрестить руки на груди и предоставить возможность компьютеру выполнить остальную работу. Формула вычисляет связь по всей выборке между ростом и весом, которые измеряются в стандартных единицах. Когда рост отдельных людей в выборке равняется,

к примеру, 1,5 или 2 стандартного отклонения выше среднего значения, какими должны быть значения их веса, *измеренные в стандартных отклонениях от среднего значения для веса*? А когда рост членов выборки близок к среднему значению, какими будут значения их веса, измеренные в стандартных единицах?

Если расстояние от среднего значения для одной переменной в целом соответствует — по величине и направлению — расстоянию от среднего значения для другой переменной (например, для людей, рост которых существенно отличается в ту или другую сторону от среднего значения роста, значения их веса, как правило, существенно отличаются от среднего значения веса, причем в том же направлении, что и рост), то у нас есть основания говорить о сильной положительной корреляции.

Если же расстояние от среднего значения для одной переменной в целом соответствует аналогичному расстоянию от среднего значения для другой переменной, *но в противоположном направлении* (например, у людей, которые чаще среднего занимаются физическими упражнениями, как правило, вес гораздо ниже среднего), то у нас есть основания говорить о сильной отрицательной корреляции.

Если две переменные в целом не отклоняются от среднего значения сколь-нибудь существенно (например, размер обуви и интенсивность занятий физическими упражнениями), то мы можем говорить о незначительной или нулевой корреляции.

Я чувствую, вы перенапряглись, читая этот раздел. Хочу вас утешить: вскоре мы вернемся к Netflix и тому, как ей удастся угадывать ваш интерес к тем или иным фильмам. Однако вначале поразмышляем над еще одним событием, где корреляция играет немаловажную роль, — SAT. Да, именно SAT, о котором говорилось в главе 3. Этот тест (первоначальное название — Scholastic Aptitude Test) представляет собой стандартизированный экзамен, состоящий из трех разделов: математика, чтение и письмо. Возможно, вам уже приходилось его сдавать (или придется сдавать в будущем). Не исключено, что вы особо не задумывались над тем, *почему* вам нужно его сдавать. Цель этого экзамена — оценить вашу способность к обучению и спрогнозировать вашу успеваемость в колледже или университете. Разумеется, у вас (и особенно у тех из вас, кому не нравятся стандартизированные тесты) может возникнуть резонный вопрос: уж не для этого ли предназначена средняя школа? Почему так важен какой-то там четырехчасовой тест, если члены

приемной комиссии колледжа могли бы просто ознакомиться с оценками, которые вы получали на протяжении *четырёх лет* учебы в старших классах школы?

Ответ на этот вопрос содержится в материале, с которым вы знакомились в главе 1 и 2. Оценки, которые выставляются ученикам в школе, представляют собой несовершенную описательную статистику. Ученик, получающий посредственные оценки при прохождении напряженной школьной программы для специализированных классов по математике и другим естественным наукам, может иметь большие академические способности и потенциал, чем ученик той же школы, предпочеший программу с гуманитарным направлением. Это объясняется тем, что гуманитарные предметы усваиваются, как правило, гораздо легче, и получить высокие оценки по ним не составляет особого труда. Очевидно, что между разными школами также существуют немалые различия, которые сказываются на оценках учеников. Согласно данным College Board (орган, который разрабатывает и управляет SAT), этот тест призван «демократизировать доступ к высшим учебным заведениям для всех учащихся». Что можно возразить против такого довода? Все справедливо! SAT предлагает стандартизированный показатель способностей, который позволяет сравнивать всех абитуриентов, поступающих в колледжи и университеты. *Но можно ли считать его достаточно надежным показателем способностей?* Если мы хотим показатель, который позволяет легко сравнивать способности учащихся, то мы могли бы также предложить всем выпускникам школы посоревноваться в забеге на 100 ярдов, что было бы гораздо дешевле и проще, чем администрировать SAT. Проблема, конечно же, в том, что результат, показанный в забеге, никоим образом не коррелирован с академической успеваемостью в колледжах и университетах. Данные о результатах забега получить легко, однако они не имеют ничего общего с интересующим нас вопросом.

Чем же SAT лучше в этом отношении? К большому разочарованию будущих поколений старшеклассников, SAT вполне достойно справляется с задачей прогнозирования успехов студентов-первокурсников, так что сдавать его придется. College Board публикует соответствующие показатели корреляции. На шкале от 0 (полное отсутствие корреляции) до 1 (идеальная корреляция) корреляция между средней оценкой ученика старших классов школы и средней оценкой студента-первокурсника равняется 0,56. (Чтобы было понятнее, *что* это означает, скажу, что корреляция между ростом и весом

взрослых мужчин в Соединенных Штатах составляет примерно 0,4.) Корреляция между комплексным результатом, показанным при сдаче SAT (чтение, математика и письмо), и средним баллом студента-первокурсника также 0,56¹. Это вроде бы говорит в пользу отказа от SAT, поскольку этот тест способен предсказать академическую успеваемость будущих студентов колледжей и университетов ничуть не лучше, чем средняя оценка ученика старших классов. По сути, самым надежным показателем будет комбинация баллов, полученных при сдаче SAT, и средней оценки ученика старших классов: корреляция между таким сочетанием и средним баллом студента-первокурсника составляет 0,64. Да, это действительно так.

Важным моментом в этом обсуждении является то, что корреляция не предполагает причинно-следственной связи: положительная или отрицательная корреляция между двумя переменными вовсе не обязательно означает, что изменения одной переменной вызывают изменения другой. Например, выше я указывал на вероятную положительную корреляцию между суммой баллов, полученных учащимся при сдаче SAT, и количеством телевизоров у него дома. Но это не значит, что родители могут существенно повысить результаты тестов своих детей путем покупки еще пяти телевизоров. Не говорит это, по-видимому, и о том, что сидение перед телевизором благотворно сказывается на академической успеваемости ученика.

Самым логичным объяснением такой корреляции может быть то, что высокообразованные родители могут себе позволить покупку нескольких телевизоров, что, однако, не мешает их детям сдавать экзамены с результатами, превышающими средний балл. Как количество телевизоров, так и экзаменационные оценки, по-видимому, обусловлены некой третьей переменной, коей является уровень образования родителей. Я не могу доказать наличие корреляции между количеством телевизоров в семье и количеством баллов, полученных при сдаче SAT (College Board не публикует соответствующих данных). Но готов доказать, что ученики из состоятельных семей демонстрируют в среднем более высокие результаты сдачи SAT, чем ученики из менее обеспеченных семей. Согласно данным, опубликованным College Board, учащиеся из семей с годовым доходом, превышающим 200 000 долларов, в среднем получают при сдаче математического раздела SAT 586 баллов, тогда как учащиеся из семей с годовым доходом, равным или меньшим 20 000 долларов, в среднем получают при сдаче того же математического раздела SAT лишь 460 баллов². Между тем, вполне вероятно и то, что в домах

семей с годовым доходом, превышающим 200 000 долларов, больше телевизоров, чем в домах семей с годовым доходом менее 20 000 долларов.

Я начал писать эту главу несколько дней назад. За это время у меня появилась возможность посмотреть фильм Bhutto. Он действительно замечательный. Полная версия фильма, в которой охватывается период с момента отделения Пакистана от Индии в 1947 году до убийства пакистанского премьер-министра Беназир Бхутто в 2007-м, производит сильное впечатление. Голос Бхутто искусно вплетается в сюжетную линию в форме выступлений и интервью. Как бы то ни было, я пометил эту киноленту пятью звездочками, что вполне соответствует прогнозу Netflix.

В своей деятельности компания Netflix использует концепцию корреляции. Все началось с того, что я выставил оценки ряду фильмов. Netflix сравнила их с рейтингами других кинозрителей, чтобы выявить тех, чьи рейтинги высоко коррелированы с моими. Этим кинозрителям, как правило, нравятся те же фильмы, что и мне. Установив данный факт, Netflix может рекомендовать мне фильмы, которые понравились моим единомышленникам и которых я еще не видел.

Это, так сказать, «картина в целом». Фактическая методология гораздо сложнее. Вообще говоря, в 2006 году Netflix инициировала конкурс, в рамках которого обычным гражданам было предложено разработать механизм, который бы повысил эффективность уже существующих рекомендаций Netflix по меньшей мере на 10% (это означает, что данная система стала бы на 10% точнее при прогнозировании того, как бы кинозритель оценил тот или иной фильм после просмотра). Победителю был обещан 1 миллион долларов.

Каждый человек или группа людей, зарегистрировавшихся для участия в конкурсе, получал «обучающие данные», состоящие из более чем 100 миллионов рейтингов, выставленных 18 000 фильмам клиентами Netflix (их общее количество составляло 480 000 человек). Отдельная совокупность из 2,8 миллиона рейтингов не разглашалась (то есть Netflix знала, как кинозрители оценили эти фильмы, но участникам конкурса такая информация не предоставлялась). Конкурсантов оценивали по тому, насколько успешно предложенные ими алгоритмы прогнозировали фактические оценки, выставленные зрителями этих «неразглашенных» фильмов. Спустя три года тысячи команд из более чем 180 стран представили на суд жюри свои предложения. К участникам конкурса предъявлялось два требования. Во-первых, победитель должен был уступить Netflix права на свой алгоритм.

И во-вторых, он должен был «объяснить миру, как ему удалось решить эту задачу и каким образом она работает»³.

В 2009 году Netflix объявила победителя. Им стала группа из семи человек, в состав которой входили статистики и программисты из США, Австрии, Канады и Израиля. Увы, я не могу описать здесь — даже в приложении — систему-победителя. Объяснение принципа ее действия занимает 92 страницы. Качество рекомендаций Netflix произвело на меня неизгладимое впечатление. Тем не менее система Netflix — просто супернавороченная вариация того, чем занимаются люди с момента появления кинематографа: найти кого-либо со схожими вкусами и попросить порекомендовать вам тот или иной фильм. Вам, как правило, нравятся те же фильмы, что и мне, и не нравятся те же фильмы, что и мне. Так что вы думаете о новом фильме Джорджа Клуни?

В этом и состоит суть корреляции.

Приложение к главе 4

Чтобы вычислить коэффициент корреляции между двумя совокупностями чисел, вы должны выполнить перечисленные ниже действия, каждое из которых иллюстрируется путем использования данных о значениях роста и веса для 15 гипотетических учащихся в приведенной ниже таблице.

1. Преобразуйте рост каждого учащегося в стандартные единицы: $(\text{рост} - \text{среднее значение}) / \text{стандартное отклонение}$.
2. Преобразуйте вес каждого из учащихся в стандартные единицы: $(\text{вес} - \text{среднее значение}) / \text{стандартное отклонение}$.
3. Для каждого учащегося вычислите произведение $(\text{вес в стандартных единицах}) \times (\text{рост в стандартных единицах})$. Вы должны увидеть, что это число будет самым большим по абсолютному значению, когда рост и вес ученика расположены относительно далеко от своих средних значений.
4. Коэффициент корреляции представляет собой сумму произведений, вычисленных выше, деленную на количество наблюдений (в нашем случае — 15).

Корреляция между ростом и весом для этой группы учащихся — 0,83. Учитывая, что коэффициент корреляции может находиться в диапазоне от -1 до 1 , это относительно высокая степень положительной корреляции, чего и следовало ожидать.

<i>A</i> <i>Учащийся</i>	<i>B</i> <i>Рост</i>	<i>C</i> <i>Вес</i>	<i>D</i> <i>Рост в стандартных единицах</i>	<i>E</i> <i>Вес в стандартных единицах</i>	<i>F</i> <i>(Вес в стандартных единицах) × (Рост в стандартных единицах)</i>
Ник	74	193	1,21	0,99	1,19
Элана	66	133	−0,63	−0,67	0,42
Дайна	68	155	−0,17	−0,06	0,01
Ребекка	69	147	0,06	−0,29	−0,02
Бен	73	175	0,98	0,49	0,48
Чару	70	128	0,29	−0,81	−0,24
Сахár	60	100	−2,00	−1,59	3,18
Мэгги	63	128	−1,32	−0,81	1,07
Фейсал	67	170	−0,40	0,35	−0,14
Тед	70	182	0,29	0,68	0,20
Нарцисо	70	178	0,29	0,57	0,17
Катрина	70	118	0,29	−1,09	−0,32
Си Джей	75	227	1,44	1,93	2,77
София	62	115	−1,54	−1,17	1,81
Уилл	74	211	1,21	1,49	1,80
Среднее значение	68,73	157,33			Итого = 12,39
Стандартное отклонение	4,36	36,12		Коэффициент корреляции = Итого/ $n = 12,39/15 = 0,83$	

Формула для вычисления коэффициента корреляции требует небольшого отступления, которое понадобится для того, чтобы объяснить систему обозначений, используемую в данном случае. Символ Σ часто применяется в статистике. Он обозначает суммирование величин, которые указаны после него. Если, например, имеется некая совокупность наблюдений x_1, x_2, x_3 и x_4 , то запись $\Sigma (x_i)$ говорит о том, что мы должны суммировать четыре наблюдения: $x_1 + x_2 + x_3 + x_4$. Таким образом, $\Sigma (x_i) = x_1 + x_2 + x_3 + x_4$. Наша формула для среднего значения совокупности из n наблюдений может быть представлена в следующем виде: среднее значение = $\Sigma (x_i)/n$.

Мы можем придать этой формуле еще более универсальный вид, записав ее как $\sum_{i=1}^n (x_i)$. Эта формула означает суммирование величин $x_1 + x_2 + x_3 + \dots + x_n$, или, другими словами, начиная с x_1 (поскольку $i = 1$) до x_n включительно

(поскольку $i = n$). Наша формула для среднего значения совокупности из n наблюдений может быть представлена в следующем виде:

$$\text{среднее значение} = \sum_{i=1}^n (x_i)/n.$$

С учетом этой универсальной системы обозначений формула вычисления коэффициента корреляции r для двух переменных x и y может выглядеть так:

$$r = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{\sigma_x} \frac{(y_i - \bar{y})}{\sigma_y},$$

где

n — количество наблюдений;

\bar{x}_x — среднее значение для переменной x ;

\bar{y}_y — среднее значение для переменной y ;

σ_x — стандартное отклонение для переменной x ;

σ_y — стандартное отклонение для переменной y .

Любая статистическая компьютерная программа может с помощью статистических инструментов вычислить коэффициент корреляции между двумя переменными. Использование Microsoft Excel в примере с ростом и весом учащихся позволяет получить такую же корреляцию между ростом и весом пятнадцати учащихся, что и вычисление, выполненное нами вручную на основе приведенной выше таблицы: 0,83.

Не покупайте расширенную гарантию для своего 99-долларового принтера

В 1981 году Joseph Schlitz Brewing Company потратила 1,7 миллиона долларов на необычайно смелую и рискованную маркетинговую кампанию для своего слабеющего бренда Schlitz. В перерыве матча за Суперкубок американского футбола Joseph Schlitz Brewing Company перед 100-миллионной зрительской аудиторией из разных стран мира провела в прямом эфире сравнительную дегустацию пива Schlitz Beer и его главного конкурента — пива Michelob¹, причем участвовали в ней не случайные люди, а сотня любителей пива Michelob. Это стало кульминацией маркетинговой кампании, проходившей в ходе игр плей-офф NFL². Всего транслировалось пять таких дегустаций; в каждой участвовало по 100 любителей конкурирующих сортов пива (Budweiser, Miller или Michelob), вслепую дегустировавших свой любимый сорт пива и Schlitz Beer. Каждый сеанс сопровождался рекламой, агрессивность которой не уступала агрессивности игр плей-офф (например, «Следите за сравнительной дегустацией Schlitz и Budweiser, проводимой во время игр плей-офф AFC»).

Маркетинговый месседж был совершенно четким: даже те, кому кажется, что они любят другой сорт пива, во время слепой дегустации отдают предпочтение пиву Schlitz. Во время матча за Суперкубок компания даже воспользовалась услугами одного из бывших судей NFL, который выступал в роли наблюдателя за ходом дегустации. Учитывая рискованный характер трансляции подобного действия в прямом эфире, вы, должно быть, подумали, что Joseph Schlitz Brewing Company выпустила потрясающее пиво?

Необязательно. В этом случае компании, которая знает толк в статистике, было достаточно предложить потребителям вполне ординарное пиво. Руководство Joseph Schlitz Brewing Company понимало, что такая уловка — слово, которым я стараюсь не злоупотреблять, даже когда речь идет о рекламе пива, — почти наверняка сработает. Большинство сортов пива в категории

Schlitz имеют примерно одинаковый вкус, и парадокс заключается в том, что Joseph Schlitz Brewing Company использовала в своей рекламной кампании именно это обстоятельство. Допустим, что рядовой любитель пива, так сказать «человек с улицы», неспособен отличить пиво Schlitz от Budweiser, Michelob или Miller. В таком случае сравнительная дегустация (выполняемая вслепую) любой пары сортов пива, по сути, равноценна подбрасыванию монеты. В среднем половина дегустаторов-любителей выберет Schlitz, а другая половина — пиво конкурента. Уже один этот факт, скорее всего, не позволил бы провести особо эффективную рекламную кампанию. («Вы не в состоянии уловить разницу, поэтому вполне могли бы пить Schlitz».) С другой стороны, Joseph Schlitz Brewing Company, безусловно, не хотелось бы проводить подобную акцию среди собственных лояльных клиентов, поскольку примерно половина из них предпочла бы какой-либо из конкурирующих сортов пива. А это очень плохо, когда приверженцы *вашего* бренда в ходе слепой дегустации выбирают какой-либо из сортов конкурентов, — но именно этот прием Joseph Schlitz Brewing Company пыталась против них применить.

Schlitz придумала умнейший ход. Гвоздем ее рекламной кампании стало проведение сравнительной дегустации среди любителей пива, отдающих предпочтение конкурирующим сортам этого напитка. Если слепая дегустация действительно равносильна подбрасыванию монеты, то примерно половина любителей Budweiser, Michelob или Miller обязательно выберет Schlitz. В результате Schlitz предстанет в более выгодном свете, чем конкуренты: *половине любителей Budweiser понравилось пиво Schlitz!*

Все это выглядит особенно убедительно в перерыве матча за Суперкубок, если дегустацией руководит бывший арбитр NFL (в соответствующей униформе). Вместе с тем это действие происходит в прямом телеэфире. Даже если статистикам из Joseph Schlitz Brewing Company предварительно удалось выяснить (в результате многочисленных дегустаций, проводившихся в частном порядке), что типичный любитель Michelob в половине случаев предпочтет пиво Schlitz, то как быть, если 100 любителей пива Michelob «начудят» во время дегустации в прямом эфире? Конечно, слепая дегустация равноценна подбрасыванию монетки, но все же как быть, если большинство дегустаторов *чисто случайно* выберут Michelob? В конце концов, если бы мы собрали тех же 100 парней и попросили их подбросить монетку, то вполне возможно, что в 85 или даже в 90 случаях из 100 выпала бы решка. Такая случайность во время дегустации в прямом эфире стала бы для бренда Schlitz

подлинной катастрофой (не говоря уж о потере 1,7 миллиона долларов, затраченных на покупку телевизионного рекламного времени).

Но на помощь пришла статистика! Если бы в нашем мире существовал некий супергерой статистики*, то именно сейчас должен был бы появиться в штаб-квартире Joseph Schlitz Brewing Company и поведать о том, что статистики называют биномиальным экспериментом (или экспериментом Бернулли). Ключевые характеристики биномиального эксперимента таковы: наличие некоего фиксированного числа испытаний (например, 100 дегустаторов), каждое с двумя возможными исходами (Schlitz или Michelob), и одинаковая вероятность «успеха» в каждом испытании. (Я предполагаю, что вероятность отдать предпочтение одному или другому сорту пива равна 50%; кроме того, я определяю успех как выбор дегустатором пива Schlitz.) Мы также исходим из того, что все испытания независимы, то есть решение одного дегустатора не оказывает влияния на решение другого.

Основываясь лишь на этой информации, наш статистический супергерой может вычислить вероятность всех исходов для 100 испытаний, например 52 Schlitz и 48 Michelob или 31 Schlitz и 69 Michelob. Те из нас, кто не претендует на звание супергероя, могут воспользоваться компьютером, чтобы получить тот же результат. Вероятность того, что все 100 дегустаторов выберут пиво Michelob, составляет 1 шанс из 1 267 650 600 228 229 401 496 703 205 376. Должен сказать, что вероятность того, что все дегустаторы погибнут вследствие падения астероида в перерыве матча за Суперкубок NFL, будет, пожалуй, даже меньше. Для нас гораздо важнее тот факт, что те же базовые вычисления позволяют определить суммарную вероятность для целого ряда исходов, например вероятность того, что 40 или меньше дегустаторов предпочтут пиво Schlitz. Эти подсчеты наверняка развеют опасения сотрудников маркетингового отдела Joseph Schlitz Brewing Company.

Предположим, что Joseph Schlitz Brewing Company вполне бы устроило, если бы по меньшей мере 40 из 100 дегустаторов выбрали пиво Schlitz — впечатляющий результат слепой дегустации, если принять во внимание, что все 100 дегустаторов — любители пива Michelob. Между тем вероятность подобного (очень и очень неплохого) исхода весьма высока. Если такая дегустация

* Я имею в виду «человека Шести Сигм». Строчной буквой греческого алфавита σ (сигма) обозначается среднеквадратическое отклонение. «Человек Шести Сигм» — это шесть среднеквадратических отклонений сверх нормы, выраженной в таких понятиях, как статистическая возможность, сила и ум.

действительно равноценна подбрасыванию монеты, то, согласно теории вероятностей, вероятность того, что по меньшей мере 40 из 100 дегустаторов выберут пиво Schlitz, равняется 98%, а вероятность того, что пиво Schlitz предпочтут как минимум 45 из 100 дегустаторов, — 82%*. Так что теоретически никакого особого риска в затее Joseph Schlitz Brewing Company не было.

Итак, чем же закончился этот трюк для Joseph Schlitz Brewing Company? В перерыве матча за Суперкубок NFL 1981 года в ходе слепой сравнительной дегустации ровно 50% любителей пива Michelob отдали предпочтение Schlitz.

Из этого примера следуют два важных урока: во-первых, вероятность — чрезвычайно мощный инструмент, и во-вторых, многие ведущие сорта пива в 1980-е годы были практически неотличимы друг от друга. Но в этой главе мы сосредоточимся именно на первом уроке.

Теория вероятностей — это наука о событиях и исходах, содержащих элемент неопределенности. Инвестирование на рынке ценных бумаг сопряжено с неопределенностью. То же касается и подбрасывания монетки, в результате которого может выпасть орел или решка. Подбрасывание монетки четыре раза подряд порождает дополнительные уровни неопределенности, поскольку каждое из четырех подбрасываний способно привести к выпадению орла или решки. Следовательно, вы не можете заранее знать исход этого эксперимента. Тем не менее я могу с некоторой долей уверенности говорить, что одни исходы (два раза орел, два раза решка) более вероятны, чем другие (четырежды орел). Как справедливо решили сотрудники Joseph Schlitz Brewing Company, выводы, основанные на теории вероятностей, могут оказаться чрезвычайно полезными. Вообще говоря, если вы поймете, почему вероятность выпадения орла четыре раза подряд равняется одному шансу из 16 (если, конечно, при этом не используется монетка со смещенным центром тяжести), то наверняка начнете понимать (приложив немного умственных усилий) буквально все, от принципа работы страховой индустрии до действий профессиональной футбольной команды в той или иной игровой ситуации (например, почему они совершили дополнительный удар после тагдауна или предпочли двухочковый переход).

Начнем с самого легкого: вероятности многих событий известны заранее. Вероятность выпадения орла при однократном подбрасывании «правильной» монетки равняется $\frac{1}{2}$, а единицы при однократном подбрасывании

* Для всех этих подсчетов я воспользовался очень удобным биномиальным онлайн-калькулятором с сайта <http://stattrek.com/Tables/Binomial.aspx>.

игральной кости — $\frac{1}{6}$. Выводы относительно вероятности наступления других событий можно сделать на основе прошлых данных. Вероятность успешного выбивания дополнительного очка после тачдауна в профессиональном футболе составляет 0,94; это означает, что бьющие по мячу игроки совершают в среднем 94 из каждых 100 дополнительных попыток. (Очевидно, что эта величина может несколько разниться у разных игроков; кроме того, она зависит от погодных условий и прочих сторонних факторов, однако не может существенно отклоняться от 0,94.) Наличие такого рода информации и умение правильно ее оценить зачастую облегчает принятие решений и позволяет лучше уяснить риски. Например, Австралийский совет по безопасности на транспорте опубликовал отчет о количественной оценке фатальных рисков при авариях на разных видах транспорта. Несмотря на широко распространенную боязнь летать самолетами, риски, связанные с пассажирскими авиаперевозками, ничтожно малы. Начиная с 1960-х годов в гражданской авиации Австралии не зафиксировано ни одной катастрофы со смертельным исходом; таким образом, коэффициент смертности в расчете на каждые 100 миллионов километров «налета», по сути, равен нулю. Для автомобильного транспорта он составил 0,5. В этом отчете впечатляет показатель для мотоциклистов. Для тех, кого вдохновляет идея стать донором органов, сообщаем: у мотоциклистов доля несчастных случаев со смертельным исходом в расчете на каждые 100 миллионов километров пробега в тридцать пять раз выше, чем у автомобилистов³.

В сентябре 2011 года 6,5-тонный спутник НАСА начал падение на Землю; ожидалось, что при прохождении плотных слоев атмосферы он распадется на части. Какой была вероятность того, что вам на голову упадет один из осколков этого спутника? Может быть, мне не следовало в те дни отправлять детей в школу? По оценке ученых-ракетчиков НАСА, вероятность попадания одного из фрагментов спутника в какого-то конкретного человека составляла 1 шанс из 21 триллиона. Тем не менее вероятность того, что кто-либо где-либо на Земле будет ушиблен куском этого спутника, оказалась не так уж мала — 1 шанс из 3200*. В конечном счете спутник действительно развалился на части, но куда именно они упали, науке до сих пор неизвестно⁴.

* Агентство НАСА также предупреждало граждан о том, что даже фрагменты упавшего на Землю спутника являются собственностью государства. Таким образом, каждый, кто найдет и спрячет их у себя (например для коллекции), будет считаться нарушителем закона — даже если найдет их в своем дворе.

Зато известно, что никто не обращался за медицинской помощью по причине такого рода травмирования. Вероятность не говорит нам о том, что случится наверняка; она лишь предупреждает, что может произойти *с высокой степенью вероятности* или *с менее высокой*. Здравомыслящие люди могут использовать эти данные у себя на работе или в повседневной жизни. Например, когда вы слышите по радио сообщение о том, что на Землю падает очередной спутник, вовсе не обязательно мчаться домой на мотоцикле, чтобы предупредить семью.

Когда речь заходит о риске, наши страхи не всегда бывают адекватны тому, что говорят нам числа о *реальной* опасности, которой мы подвергаемся, то есть о том, чего нам действительно следует бояться. Один из поразительных выводов сделали Стивен Левитт и Стивен Дабнер, авторы книги «Фрикономика»* (Freakonomics), заявив, что плавательный бассейн во дворе вашего дома гораздо опаснее, чем заряженный револьвер, хранящийся у вас в шкафу⁵. Левитт и Дабнер подсчитали: вероятность того, что ребенок в возрасте до десяти лет утонет в плавательном бассейне, в сто раз превышает вероятность того, что он случайно застрелится, играя с вашим револьвером (если, конечно, найдет его в шкафу)**. В интересной статье трех исследователей из Корнелльского университета — Гаррика Блалока, Вринды Кадияли и Дэниела Саймона — сообщается о том, что тысячи американцев, возможно, умерли после теракта 11 сентября *из-за страха летать самолетами*⁶. Мы никогда не узнаем подлинных рисков, связанных с терроризмом; однако нам доподлинно известно, что вождение автомобиля — опасное занятие. Когда после теракта 11 сентября американцы решили больше ездить наземным транспортом, чем летать, ежемесячное количество дорожно-транспортных происшествий в октябре, ноябре и декабре 2001 года,

* Левитт С., Дабнер С. Фрикономика. — М. : Манн, Иванов и Фербер, 2010.

** Левитт и Дабнер рассуждали примерно так. Каждый год тонут приблизительно 550 детей в возрасте до десяти лет, а 175 детей в возрасте до десяти лет погибают в результате неосторожного обращения с оружием. Левитт и Дабнер взяли за основу следующие коэффициенты смертности: один утонувший ребенок на каждые 11 000 плавательных бассейнов в сравнении с одним смертельным случаем в результате неосторожного обращения с оружием на каждые «миллион с хвостиком» единиц огнестрельного оружия. Что касается подростков, то указанные коэффициенты могут быть совершенно другими, во-первых, поскольку подростки лучше плавают и, во-вторых, могут гораздо чаще быть виновниками трагедии, если у них в руках случайно окажется огнестрельное оружие. Однако в моем распоряжении нет соответствующих данных.

согласно оценкам авторов данного исследования, увеличилось на 344 случая (с учетом среднего количества погибших и факторов, которые обычно способствуют ДТП, например погодных условий). Со временем — предположительно в результате уменьшения боязни терроризма — этот эффект сам по себе сошел на нет, но, по оценкам исследователей, теракты 11 сентября как таковые привели к более чем 2000 дорожно-транспортных происшествий со смертельным исходом.

Иногда вероятность может также говорить нам *постфактум*, что, по-видимому, произошло и что, по-видимому, не произошло — как в случае с анализом ДНК. Когда эксперты в телесериале CSI: Miami находят следы слюны на огрызке яблока рядом с жертвой преступления, в этой слюне нельзя обнаружить имя убийцы, даже если ее рассматривает через мощный микроскоп очень симпатичная девушка-эксперт. Однако эта слюна (или волос, или кусочек кожи или кости) содержит сегмент ДНК, в котором, в свою очередь, есть участки (локусы), специфические для каждого человека (за исключением однойцовых близнецов, имеющих одну и ту же ДНК). Когда медэксперт заключает, что у некоего образца ДНК выявлено совпадение, это лишь часть того, что предстоит доказать следствию. Да, определенные локусы на образце ДНК, взятом с места преступления, должны совпадать с соответствующими локусами на образце ДНК, взятом у подозреваемого. Тем не менее следователям также предстоит доказать, что такое совпадение *неслучайно*.

ДНК у разных людей бывают похожи, как и многие другие характеристики: размер обуви, рост, цвет глаз. (Свыше 99% ДНК у людей идентичны.) Если в распоряжении исследователей есть только малый образец ДНК, на котором можно проверить лишь пару-тройку участков, то вполне возможно, что у тысяч или даже миллионов людей окажется точно такой же генетический фрагмент. Следовательно, чем большее число локусов будет проверено и чем большее естественное генетическое отклонение будет в каждом из них обнаружено, тем определеннее окажется совпадение. Можно сказать и по-другому: тем меньше вероятность того, что данный образец ДНК совпадет с несколькими людьми⁷.

Чтобы лучше уяснить ситуацию с ДНК, представьте, что ваше «число ДНК» состоит из вашего телефонного номера, присоединенного к номеру вашей карточки социального страхования. Эта последовательность из девятнадцати цифр идентифицирует вас уникальным образом. Допустим, что

каждая такая цифра представляет собой «участок» с десятью возможностями: 0, 1, 2, 3 и т. д. Предположим также, что следователи обнаружили на месте преступления остаток некоего «числа ДНК»: __459__4_0_9817___. Оказалось, что этот фрагмент в точности совпадает с вашим «числом ДНК». Итак, вы — преступник?

Следует обратить внимание на три вещи. Во-первых, все, что меньше чем полное совпадение с полным геномом, оставляет некоторый простор для неопределенности. Во-вторых, чем больше локусов, которые мы можем проверить, тем меньше неопределенность. И в-третьих, важен контекст. Выявленное совпадение было бы чрезвычайно убедительным, если бы нашлись свидетели того, как вы пытались скрыться с места преступления, или если бы у вас в кармане обнаружили кредитную карточку жертвы.

Когда следователи располагают неограниченным временем и ресурсами, типичный процесс включает в себя проверку тринадцати разных локусов. Шансы, что профиль ДНК у двух разных человек совпадает по всем тринадцати локусам, чрезвычайно малы. Когда для идентификации останков, найденных во Всемирном торговом центре после терактов 11 сентября, использовался анализ ДНК, образцы, обнаруженные на месте трагедии, сравнивались с образцами, предоставленными членами семей жертв теракта. Вероятность, требовавшаяся для позитивной идентификации, равнялась один из миллиарда; то есть вероятность того, что останки принадлежат кому-то другому, а не идентифицируемой жертве, не превышает одного шанса из миллиарда. Впоследствии, по мере того как оставалось все меньше и меньше неидентифицированных жертв, с которыми могли бы быть спутаны останки, этот стандарт был ослаблен.

Если ресурсы ограничены или имеющийся образец ДНК слишком мал или загрязнен, чтобы можно было проверить тринадцать локусов, ситуация становится более запутанной и спорной. В 2008 году газета Los Angeles Times опубликовала серию материалов, посвященных использованию ДНК при расследовании преступлений⁸. В частности, издание задалось вопросом, не недооценена ли возможность случайных совпадений при использовании стандарта вероятности, определяемого законом. (Поскольку профиль ДНК всего населения не знает никто, то вероятности, на которые ссылаются в суде ФБР и другие правоохранительные органы, носят лишь оценочный характер.) Весьма неоднозначную реакцию в обществе вызвала информация

о том, что эксперт-криминалист из Аризоны, выполнявший тесты на основе базы данных ДНК этого штата, обнаружил совпадение ДНК на девяти локусах у двух опасных уголовных преступников, не являющихся родственниками; между тем, согласно ФБР, вероятность такого совпадения равна одному шансу из 113 миллиардов. Дальнейший поиск в других базах данных ДНК позволил выявить свыше тысячи пар людей с генетическими совпадениями на девяти и более локусах. Это может служить серьезным поводом к размышлению для правоохранительных органов и адвокатов. Пока же важный для нас урок заключается в том, что анализ ДНК, на который возлагаются столь большие надежды, хорош лишь настолько, насколько надежны значения вероятности, подкрепляющие его.

Зачастую бывает очень полезно знать вероятность одновременного наступления *нескольких* событий. Какова вероятность исчезновения электричества в сети и выхода из строя автономного генератора? Вероятность *одновременного* наступления двух независимых событий представляет собой произведение их соответствующих вероятностей. Другими словами, вероятность наступления события А и события В равна вероятности наступления события А, умноженной на вероятность наступления события В. Чтобы вам стало понятнее, приведу соответствующий пример. Если вероятность выпадения орла при однократном подбрасывании монетки составляет $\frac{1}{2}$, то вероятность его выпадения при подбрасывании такой же монетки два раза подряд равняется $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$; три раза подряд — $\frac{1}{8}$; четыре раза подряд — $\frac{1}{16}$ и т. д. (Понятно, что вероятность выпадения решки при подбрасывании монетки четыре раза подряд также составляет $\frac{1}{16}$.) Это объясняет, почему системный администратор в вашем учебном заведении или офисе постоянно напоминает вам о необходимости усложнить пароль. Если вы используете шестизначный пароль, состоящий только из цифр, мы можем подсчитать количество возможных паролей: $10 \times 10 \times 10 \times 10 \times 10 \times 10$, что равняется 10^6 , или 1 000 000. На первый взгляд, количество комбинаций настолько велико, что угадать пароль сложно, однако компьютер проверит все эти 1 000 000 вариантов за какую-то долю секунды.

Допустим, системный администратор убеждает вас включить в пароль буквы. На данном этапе для каждого из шести разрядов имеется 36 комбинаций: 26 букв английского алфавита и 10 цифр. Итак, количество возможных паролей возрастает до $36 \times 36 \times 36 \times 36 \times 36 \times 36$, или 36^6 , то есть свыше двух миллиардов. Если ваш системный администратор требует, чтобы пароль

состоял из восьми цифр, и призывает использовать символы #, @, % и !, как в Чикагском университете, то количество потенциальных паролей увеличивается до 46^8 , то есть свыше 20 триллионов.

Здесь нужно сделать одно важное замечание. Эта формула применима только если события независимы; иными словами, когда исход одного события не оказывает влияния на исход другого события. Например, вероятность того, что в результате первого подбрасывания монетки выпадет орел, не влияет на вероятность исхода второго подбрасывания той же монетки. С другой стороны, вероятность того, что сегодня пойдет дождь, не независима от того, был ли он вчера, поскольку грозовые фронты могут сохраняться на протяжении нескольких дней. Аналогично, вероятность того, что сегодня ваш автомобиль попадет в аварию, и того, что он попадет в нее в следующем году, также не независимы друг от друга. То, что привело к аварии вашего автомобиля в этом году, может спровоцировать ДТП и в следующем году: возможно, вы склонны садиться за руль в нетрезвом состоянии, или вам нравится устраивать гонки на дороге, или строчить эсэмэски во время вождения; наконец, не исключено, что вы просто плохой водитель. (Именно поэтому после каждого очередного ДТП ваша страховая ставка повышается; дело не столько в желании страховой компании компенсировать деньги, выплаченные ею согласно страховому договору, сколько в том, что теперь она располагает новой информацией о вероятности вашего попадания в дорожно-транспортные происшествия в дальнейшем, поскольку — после того как вы, заезжая в гараж, сильно поцарапали свой автомобиль — такая вероятность повысилась.)

Допустим, вас интересует вероятность наступления одного (исхода А) или другого (исхода В) события (опять же предполагая, что они независимы). В этом случае вероятность наступления события А или В равна сумме их индивидуальных вероятностей, то есть вероятность А *плюс* вероятность В. Например, вероятность выпадания 1, 2 или 3 в результате подбрасывания одной игральной кости равняется сумме их отдельных вероятностей: $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$. Это должно быть интуитивно понятно. При подбрасывании игральной кости есть шесть возможных исходов. Числа 1, 2 и 3 в совокупности составляют половину из них. Следовательно, вероятность выпадания 1, 2 или 3 вследствие подбрасывания одной игральной кости равняется 50%. Если вы играете в кости в Лас-Вегасе, то вероятность выпадания 7 или 11 в результате однократного подбрасывания равна количеству комбинаций,

составляющих в сумме 7 или 11, поделенному на общее число вариантов, которые могут выпасть в результате подбрасывания двух игральных костей, или $\frac{8}{36}$ *.

Вероятность также позволяет подсчитать математическое ожидание — чрезвычайно полезный инструмент, используемый при принятии любых управленческих решений, особенно в сфере финансов. Математическое ожидание — это среднее значение случайной величины. Математическое ожидание, или отдача (функция выигрыша) от некоторого события, например покупки лотерейного билета, представляет собой сумму всех разных исходов, весовыми коэффициентами при каждом из которых являются вероятность исхода и выигрыш. Как обычно, приведем пример, чтобы прояснить смысл сказанного. Допустим, вам предложили сыграть в кости, причем подбрасывается только одна игральная кость. Функция выигрыша в этой игре такова: 1 доллар, если у вас выпадает 1; 2 доллара, если у вас выпадает 2; 3 доллара, если у вас выпадает 3 и т. д. Каково математическое ожидание в случае однократного подбрасывания игральной кости? Вероятность каждого из возможных исходов равняется $\frac{1}{6}$, поэтому математическое ожидание вычисляется так:

$$\frac{1}{6}(\$1) + \frac{1}{6}(\$2) + \frac{1}{6}(\$3) + \frac{1}{6}(\$4) + \frac{1}{6}(\$5) + \frac{1}{6}(\$6) = \frac{21}{6}, \text{ или } \$3,50.$$

На первый взгляд, математическое ожидание 3,50 доллара кажется относительно бесполезной величиной. В конце концов, вы не можете фактически заработать 3,50 доллара в результате однократного подбрасывания игральной кости (так как ваш доход в любом случае должен равняться целому числу). На самом деле математическое ожидание представляет собой чрезвычайно мощный инструмент, поскольку он может сказать вам, является ли то или иное событие «справедливым», учитывая его цену и ожидаемый исход. Допустим, вам предлагают поучаствовать в описанной выше игре при ставке

* Существует шесть способов выбросить 7 при подбрасывании двух игральных костей: (1,6); (2,5); (3,4); (6,1); (5,2) и (4,3) и лишь два способа выбросить 11: (5,6) и (6,5).

Между тем есть 36 возможных вариантов результата подбрасывания двух игральных костей: (1,1); (1,2); (1,3); (1,4); (1,5); (1,6). И (2,1); (2,2); (2,3); (2,4); (2,5); (2,6). И (3,1); (3,2); (3,3); (3,4); (3,5); (3,6). И (4,1); (4,2); (4,3); (4,4); (4,5); (4,6). И (5,1); (5,2); (5,3); (5,4); (5,5); (5,6). И наконец, (6,1); (6,2); (6,3); (6,4); (6,5) и (6,6).

Следовательно, вероятность выпадения 7 или 11 равняется количеству возможных способов выбросить любое из этих двух чисел, деленное на общее количество возможных вариантов при подбрасывании двух игральных костей, то есть $\frac{8}{36}$. Между прочим, значительная часть ранних исследований вероятности выполнялась именно любителями азартных игр в попытках точно определить свои шансы.

3 доллара за каждое подбрасывание игровой кости. Имеет ли смысл соглашаться на такие условия? Да, поскольку математическое ожидание исхода (3,50 доллара) выше, чем стоимость игры (3,00 доллара). Это не означает, что вы обязательно заработаете деньги в результате однократного подбрасывания игровой кости, но помогает уяснить, на какой риск стоит пойти, а на какой — нет.

Этот гипотетический пример можно применить к профессиональному американскому футболу. Как указывалось ранее, после тачдауна команда может либо пробить и заработать дополнительное очко, либо попытаться выполнить двухочковую конверсию. Первый вариант предполагает такой удар по мячу с трехярдовой линии, в результате которого мяч должен пройти между стойками ворот; второй вариант предполагает пробежку или передачу мяча в концевую зону с трехярдовой линии, что значительно труднее. Команда может предпочесть более легкий вариант и заработать одно очко или выбрать более сложный вариант и заработать два очка. Как быть?

Возможно, статистики не играют в футбол и не назначают свиданий девушкам из группы поддержки, но они могут предоставить ценное статистическое руководство футбольным тренерам⁹. Как указывалось ранее, вероятность выполнения удара после тачдауна равняется 0,94. Это означает, что математическое ожидание попытки заработать одно очко после тачдауна также составляет 0,94, поскольку оно равняется «доходу» (1 очко), умноженному на вероятность успеха (0,94). Никакая команда не может заработать 0,94 очка, но эта величина помогает оценить данный вариант действий после тачдауна в сравнении с альтернативным вариантом (двухочковой конверсией). Математическое ожидание в случае «погони за двумя очками» оказывается гораздо меньшим: 0,74. Да, «доход» выше (2 балла), но вероятность успеха существенно ниже (0,37). Очевидно, если играть осталось совсем немного и для победы команде требуются два очка, то ей не остается ничего другого, как попытаться счастья с двухочковой конверсией. Но если цель команды — максимизация количества набранных очков, и она располагает для этого определенным запасом времени, то вариант с зарабатыванием одного очка для нее более приемлем.

Такой же базовый анализ может показать, почему не стоит покупать лотерейные билеты. В Иллинойсе вероятности, связанные с разными возможными выигрышами в лотерее, напечатаны на оборотной стороне каждого билета. Я купил за 1 доллар один билет мгновенной лотереи. (Интересно,

облагается ли эта сумма налогом?) На его оборотной стороне напечатаны — микроскопическим шрифтом — шансы выиграть различные денежные призы или получить еще один такой же билет (бесплатно): 1 шанс из 10 (бесплатный лотерейный билет); 1 шанс из 15 (2 доллара); 1 шанс из 42,86 (4 доллара); 1 шанс из 75 (5 долларов) и т. д. вплоть до 1 шанс из 40 000 — 1000 долларов. Я подсчитал ожидаемый доход для моего билета мгновенной лотереи, сложив все возможные варианты выигрыша денежного приза с весовыми коэффициентами, равными вероятности выигрыша каждого из этих денежных призов*. Оказалось, что ожидаемый доход для моего однодолларового лотерейного билета — примерно 0,56 доллара**. Таким образом, покупка такого билета — абсолютно бездарный способ потратить 1 доллар. Как назло, я выиграл 2 доллара.

Несмотря на мой неожиданный выигрыш, я все равно считаю, что покупка билета мгновенной лотереи — абсолютная глупость. Это один из важнейших уроков теории вероятностей. Хорошие решения — если их оценивать вероятностями, которые за ними кроются, — в действительности могут оказаться не такими уж хорошими. А плохие решения — например, покупка билета мгновенной лотереи в Иллинойсе — не такими уж плохими, по крайней мере на коротком отрезке времени. Но в конечном счете вероятность все равно торжествует. Важная теорема, известная как закон больших чисел, гласит, что по мере возрастания количества испытаний средний результат исходов все сильнее приближается к его математическому ожиданию. Да, я выиграл 2 доллара, купив сегодня билет мгновенной лотереи. И мог бы еще раз выиграть 2 доллара завтра. Но если я куплю тысячи однодолларовых лотерейных билетов, каждый с ожидаемым доходом 0,56 доллара, то я почти

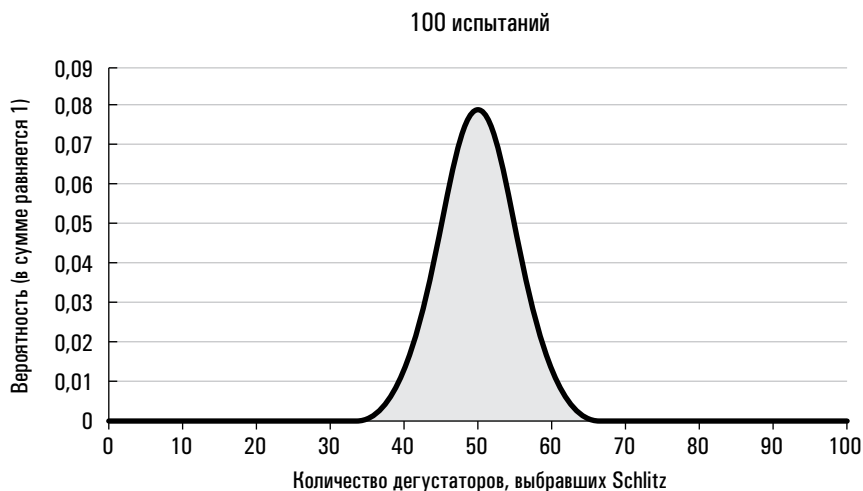
* Полное математическое ожидание для однодолларового билета мгновенной лотереи в штате Иллинойс (округленное до ближайшего цента) подсчитывается следующим образом: $1/15 \times (\$2) + 1/42,86 \times (\$4) + 1/75 \times (\$5) + 1/200 \times (\$10) + 1/300 \times (\$25) + 1/1589 \times (\$50) + 1/8000 \times (\$100) + 1/16\,000 \times (\$200) + 1/48\,000 \times (\$500) + 1/40\,000 \times (\$1000) = \$0,13 + \$0,09 + \$0,07 + \$0,05 + \$0,08 + \$0,03 + \$0,01 + \$0,01 + \$0,01 + \$0,03 = \$0,51$. Однако существует также шанс 1/10 получить в качестве выигрыша бесплатный лотерейный билет; ожидаемый доход этого варианта составляет 0,51 доллара; таким образом, ожидаемый доход в целом равняется $\$0,51 + 0,1 \times (\$0,51) = \$0,51 + \$0,05 = \$0,56$.

** Строго говоря, для правильного подсчета математического ожидания необходимо, чтобы сумма вероятностей всех возможных исходов равнялась 1. Здесь же сумма вероятностей представленных исходов составляет 0,2659. Однако, если принять, что с вероятностью $1 - 0,2659 = 0,7341$ выпадает билет без всякого выигрыша (то есть выигрыш равен 0), тогда математическое ожидание подсчитано правильно. *Прим. ред.*

наверняка останусь в проигрыше. К тому времени, когда я потрачу на покупку лотерейных билетов один миллион долларов, мой выигрыш составит сумму, очень близкую к 560 000 долларов.

Закон больших чисел объясняет, почему в долгосрочном периоде казино всегда выигрывают. Вероятности, связанные со всеми играми, которые практикуются в казино, благоприятствуют последнему (при условии, что казино способно помешать игрокам в блек-джек вычислять карты). Если в течение довольно продолжительного отрезка времени было сделано достаточное количество ставок, то казино обязательно получит больше, чем потеряет. Закон больших чисел также объясняет, почему вероятность того, что компания Joseph Schlitz Brewing Company добьется нужного ей результата, повышается при выполнении 100 слепых дегустаций, а не десяти. Взгляните на «функции плотности вероятности» для 10, 100 и 1000 слепых дегустаций пива. (Несмотря на свое мудреное название, функция плотности вероятности просто отображает упорядоченные исходы вдоль оси x и ожидаемую вероятность каждого исхода вдоль оси y ; в сумме эти вероятности дают 1.) Как и ранее, я предполагаю, что каждая дегустация эквивалентна подбрасыванию монетки, а каждый дегустатор выбирает пиво Schlitz с вероятностью 0,5. Как видно из приведенных ниже графиков, по мере увеличения количества дегустаторов ожидаемый исход все больше сосредоточивается в области выбора пива Schlitz половиной (50%) дегустаторов. В то же время вероятность получения исхода, который резко бы отклонялся от 50%, по мере роста числа испытаний резко падает.





Ранее я говорил, что руководство компании Joseph Schlitz Brewing Company было бы радо, если бы в ходе сравнительной слепой дегустации не менее 40% любителей пива Michelob выбрали пиво Schlitz. Приведенные ниже числа отражают вероятность достижения такого результата по мере увеличения количества дегустаторов:

10 дегустаторов: 0,83

100 дегустаторов: 0,98

1000 дегустаторов: 0,999999999

1 000 000 дегустаторов: 1

Сейчас интуиция должна подсказать вам смысл, заложенный в подзаголовке этой главы: «Не покупайте расширенную гарантию для своего

99-долларового принтера». Ладно, возможно, пока этот смысл для вас еще неочевиден. Вернемся к одному из предыдущих примеров. Вся страховая отрасль построена на вероятностях. (А гарантийное обязательство — одна из форм страхования.) Когда вы страхуете что-либо, вы заключаете договор на получение определенной компенсации при наступлении четко оговоренных обстоятельств. Например, страховка вашего автомобиля может предусматривать его замену в случае, если он будет украден или врежется в дерево. В обмен на эту гарантию вы соглашаетесь выплачивать определенную сумму за период, на который застраховали свое авто. Основная идея страхования заключается в том, что в обмен на регулярные и предсказуемые выплаты вы переносите на соответствующую страховую компанию риск того, что ваш автомобиль может быть похищен, или попасть в аварию, или даже прийти в полную негодность по причине вашего неумения хорошо водить.

Почему страховые компании готовы взять на себя такие риски? Потому что в долгосрочном периоде они заработают большие прибыли — если, конечно, правильно рассчитают величину своих страховых взносов. Разумеется, какие-то из автомобилей, застрахованных компанией Allstate Corporation, будут украдены. Другие придут в полную негодность, после того как их владельцы наедут, к примеру, на пожарный гидрант, как одна из моих старых приятельниц. (Кроме того, ей пришлось возместить полную стоимость устройства, что, между прочим, оказалось гораздо дороже, чем вы могли подумать.) Однако с большинством автомобилей, застрахованных Allstate Corporation или какой-либо другой компанией, серьезных неприятностей не случится. Чтобы получить прибыль, страховой компании нужно лишь позаботиться о том, чтобы сумма страховых взносов превышала возможные страховые выплаты. А для этого страховая компания должна иметь четкое представление о том, что в страховой отрасли принято называть «ожидаемыми потерями» на каждый страховой полис. Это в точности такая же концепция, что и математическое ожидание, но со «страховым уклоном». Если ваш автомобиль застрахован на 40 000 долларов, а вероятность того, что он будет украден в любом данном году, равняется 1 шансу из 1000, то годовые ожидаемые потери на ваш автомобиль составят 40 долларов. Величина годового страхового взноса для той части страхового покрытия, которая относится к угону автомобиля, должна быть *больше 40 долларов*. С этого момента страховая компания ничем, по сути, не отличается от казино или мгновенной лотереи в Иллинойсе. Да, иногда ей придется выплачивать

определенные суммы по страховым претензиям, но в долгосрочной перспективе поступления обязательно превысят эти выплаты.

Как потребитель, вы должны отдавать себе отчет, что в длительном периоде страховка не сэкономит вам деньги. Единственное, что она может для вас сделать, это предотвратить некоторые неприемлемо высокие убытки, компенсировав, например, потерю угнанного автомобиля стоимостью 40 000 долларов или сгоревшего дома за 350 000 долларов. Покупка страхового полиса с точки зрения статистики — «неудачная ставка», поскольку вы заплатите страховой компании в среднем больше, чем от нее получите. Тем не менее это все же вполне разумный способ защиты от исходов, которые в противном случае могли бы вас просто разорить. По иронии судьбы такие богачи, как Уоррен Баффет, могут экономить на страховке автомобиля, жилья или даже здоровья, потому что миллиардеры могут себе позволить практически любые несчастья, которые приключаются с людьми.

И вот тут мы наконец возвращаемся к пресловутому принтеру за 99 долларов. Предположим, вы купили замечательный новый лазерный принтер в какой-либо солидной торговой сети, например в Best Buy*. Когда вы подходите к кассовому аппарату, чтобы рассчитаться за покупку, продавец-консультант предлагает вам ряд вариантов продленного срока гарантии. Если вы заплатите дополнительно 25 или 50 долларов, Best Buy починит или заменит ваш принтер в случае его поломки в ближайшие год-два. Зная основы теории вероятностей, страхового дела и экономики, вы должны сразу же сделать следующие выводы: 1) Best Buy — коммерческая организация, которая стремится максимизировать свою прибыль; 2) продавец-консультант пытается навязать вам какой-либо из вариантов продленного срока гарантии; 3) исходя из пунктов 1) и 2) вы можете заключить, что стоимость такой гарантии будет выше, чем ожидаемая стоимость ремонта принтера для Best Buy (если бы это было не так, Best Buy вряд ли столь настойчиво вас бы уговаривала); 4) если ваш принтер за 99 долларов сломается и вам придется платить за его ремонт из собственного кармана, это никаким особым образом не повлияет на вашу жизнь.

* Ранее в этой книге я привел пример, в котором упоминалось о нетрезвых рабочих, выпускающих бракованные лазерные принтеры. Выбросьте его из головы: будем исходить из того, что компания, выпускающая лазерные принтеры, уже решила проблемы с качеством.

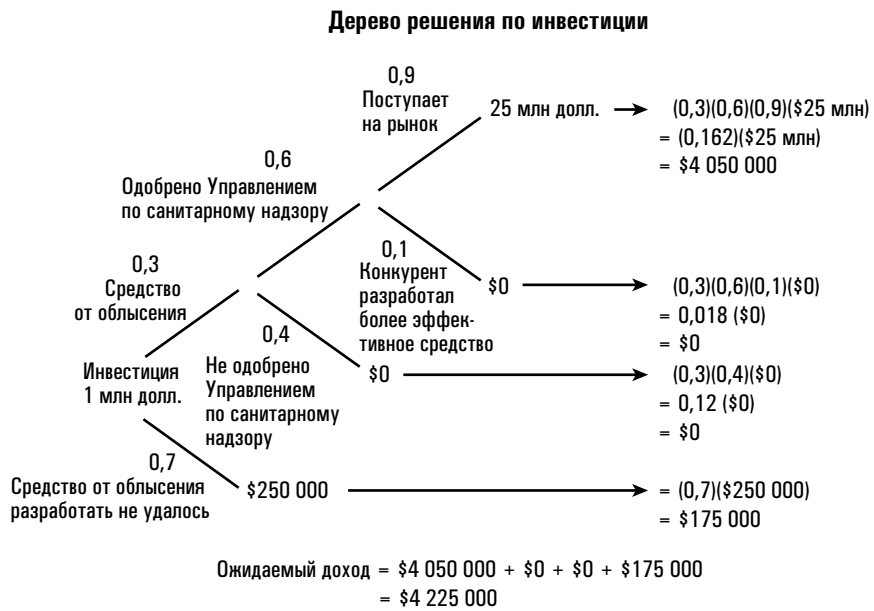
В среднем вы заплатите за продление гарантийного срока больше, чем пришлось бы выложить за ремонт принтера. Более универсальный урок — и один из основополагающих в деле личных финансов — заключается в том, что вы всегда должны страховать себя от любых неблагоприятных обстоятельств, которые могут внести *существенный* дискомфорт в вашу жизнь. Застраховываться от всего остального не имеет смысла.

Математическое ожидание также может помочь в принятии сложных решений, которые обуславливаются многими обстоятельствами в разные моменты времени. Допустим, кто-то из друзей попросил вас инвестировать один миллион долларов в исследовательский проект, связанный с разработкой новейшего средства от облысения. Вы, скорее всего, поинтересуетесь, каковы шансы проекта на успех, и получите весьма неоднозначный ответ. Так как речь идет об исследовательском проекте, вероятность того, что ученым удастся найти эффективное средство от облысения, составляет лишь 30%. В случае неудачи вам вернут только 250 000 долларов от вложенного миллиона, поскольку именно такая сумма была зарезервирована для вывода нового средства на рынок (тестирование, маркетинг и т. п.). Даже если исследователи добьются успеха, существует лишь 60-процентная вероятность того, что Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США одобрит это чудодейственное средство для медицинского применения. Но даже в том случае, если средство не только окажется эффективным, но и будет признано как безопасное для человека, существует 10-процентная вероятность того, что кто-либо из конкурентов предложит примерно в то же время еще более эффективный препарат, что лишит вас каких-либо надежд на получение прибыли. Но если удача будет во всем вам сопутствовать (ваше средство против облысения окажется эффективным и безопасным для человека, а ваши конкуренты не предложат ничего лучшего), оптимальная оценка доходности инвестиций составит 25 миллионов долларов.

Итак, имеет ли смысл вкладывать один миллион долларов в этот исследовательский проект?

Информация, которой вы располагаете, кажется весьма запутанной. Потенциальный доход выглядит довольно внушительно — в 25 раз больше вложенного капитала, — но и количество возможных ловушек велико. Эту информацию можно представить в виде дерева решений, которое — если вероятности, связанные с каждым исходом, соответствуют действитель-

ности, — даст вам вероятностную оценку того, как вам следует поступить. На дереве решений отображается каждый источник неопределенности, а также вероятности, связанные со всеми возможными исходами. Конец дерева указывает все возможные доходы, а также вероятность получения каждого из них. Если каждый такой доход умножить на весовой коэффициент, который равняется вероятности соответствующего дохода, и просуммировать все возможности, то мы получим математическое ожидание данной инвестиционной возможности. Как обычно, схематическое изображение способствует лучшему пониманию.



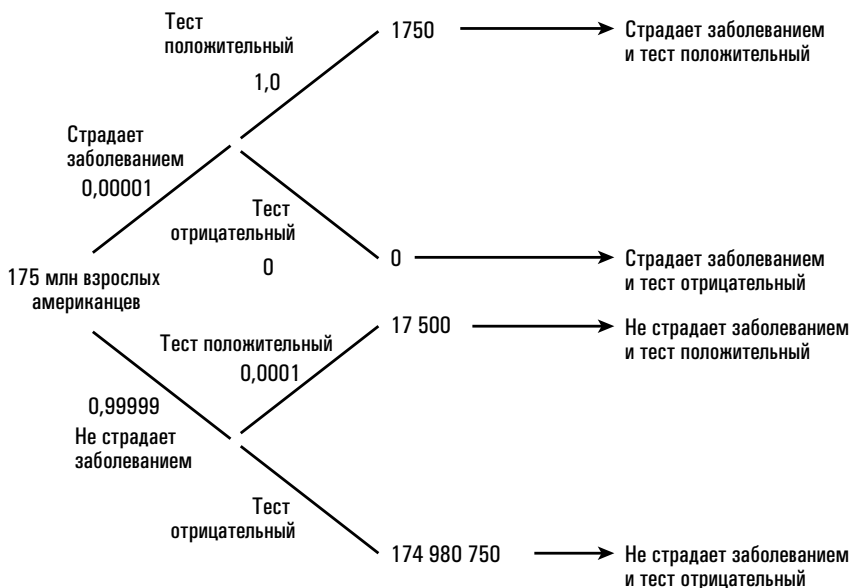
Эта конкретная возможность имеет привлекательное математическое ожидание. Величина ожидаемого дохода, полученная в результате суммирования всех возможных доходов с учетом их весовых коэффициентов, равняется 4,225 миллионам долларов. Тем не менее решение сделать такую инвестицию в исследовательский проект может оказаться не самым мудрым, если речь идет о вложении денег, накопленных вами на обучение детей в колледже. Дерево решений позволяет вам узнать, что ваш ожидаемый доход существенно выше суммы, которую вам предлагается инвестировать в данный проект. С другой стороны, наиболее вероятный исход — что

исследователям не удастся изобрести новое средство от облысения и вам вернут лишь 250 000 долларов. Ваша готовность к такой инвестиции может зависеть от вашей склонности к риску. Из закона больших чисел следует, что любая инвестиционная фирма или богатый человек вроде Уоррена Баффета должны выискивать сотни возможностей наподобие этой, с неопределенными исходами, но привлекательными величинами ожидаемой прибыли. Некоторые из них сработают; большинство наверняка нет. В среднем такие инвесторы заработают немало — точно так же как страховая компания или казино. Если величина ожидаемого дохода кажется вам привлекательной, то желательно, чтобы количество попыток было как можно бóльшим.

Аналогичный базовый процесс можно использовать для объяснения явления, которое на первый взгляд противоречит здравому смыслу. Иногда нет смысла проводить обследование всего населения с целью выявления какого-либо редкого, но серьезного заболевания, такого, скажем, как СПИД. Допустим, тестирование на какое-то редкое заболевание отличается высокой степенью точности. Предположим, что эта болезнь поражает одного из каждых 100 000 взрослых, а точность ее диагностирования составляет 99,9999%. Тест никогда не дает ложного отрицательного результата (то есть не пропускает человека, страдающего таким заболеванием); однако примерно в одном из 10 000 тестов, проведенных на здоровом человеке, будет зафиксирован ложный положительный результат (то есть тест укажет на наличие у человека данного заболевания, хотя на самом деле этот человек здоров). Парадоксальная особенность здесь состоит в том, что несмотря на впечатляющую точность теста, *большинство людей с положительным результатом тестирования в действительности оказываются не больны*. Но такой предварительный диагноз вызовет у них сильнейший стресс, пока не выяснится, что он ложный; кроме того, это может обусловить напрасное расходование средств на проведение повторных тестов и лечение людей, которые в действительности здоровы.

Если мы подвергнем тестированию все взрослое население Соединенных Штатов, то есть приблизительно 175 миллионов человек, то дерево решений примет следующий вид:

Широкомасштабное обследование на наличие редкого заболевания



$$\frac{\text{Люди, страдающие этим заболеванием}}{\text{Те, кому сообщили, что у них выявлено заболевание}} = \frac{1750}{1750 + 17\,500} = \frac{1750}{19\,250} = 0,09 = 9\%$$

Итак, только 1750 человек страдают этим заболеванием. У всех положительный результат теста. У остальных 174 с лишним миллионов взрослых этой болезни не выявлено. Для 99,9999% протестированных результат был определен правильно: они здоровы. Ложный положительный результат получили всего 0,0001% человек. Однако 0,0001% от 174 миллионов все же достаточно большое число. По сути, это в среднем 17 500 человек.

Попытаемся проанализировать, что это означает. В общей сложности 19 250 человек уведомляются о том, что они страдают данным заболеванием, и лишь 9% из них в действительности больны! А ведь речь идет о тесте с очень невысокой долей ложных положительных результатов. Не слишком отклоняясь от обсуждаемой темы, я привел этот пример, чтобы дать вам некоторое представление о том, почему методы сдерживания затрат в системе здравоохранения иногда предусматривают проведение обследования главным образом среди групп повышенного риска заболевания, а не среди

здорового населения. В случае таких заболеваний, как ВИЧ/СПИД, представители государственной системы здравоохранения зачастую рекомендуют обследовать группы повышенного риска, например гомосексуалистов или наркоманов.

* * *

Иногда вероятность сигнализирует нам об опасных ситуациях. В главе 1 рассказывалось о проблеме манипуляций со стандартизованными тестами и об одной из фирм, которая пыталась выявлять такие случаи, Caveon Test Security. Комиссия по ценным бумагам и биржам (Securities and Exchange Commission — SEC), государственное агентство, отвечающее за практическую реализацию федеральных законов, касающихся торговли ценными бумагами, применяет аналогичную методологию для обнаружения трейдеров-инсайдеров. (Инсайдерская торговля ценными бумагами связана с незаконным использованием конфиденциальной информации, такой как, скажем, знание юридической фирмой о предстоящем поглощении для торговли акциями и другими ценными бумагами компаний, участвующих в данном процессе.) SEC использует мощные компьютеры для анализа сотен миллионов операций купли-продажи ценных бумаг с целью выявления подозрительной активности, например крупной покупки акций компании непосредственно перед объявлением о ее поглощении или массовом «сбросе» акций компании буквально перед ее заявлением о резком сокращении прибыли¹⁰. SEC также расследует деятельность инвестиционных менеджеров с необычайно высокими прибылями на протяжении длительных периодов времени. (Как экономическая теория, так и исторические данные свидетельствуют, что отдельно взятому инвестору чрезвычайно трудно год за годом получать прибыль выше среднего уровня.) Разумеется, дальновидные инвесторы всегда пытаются прогнозировать хорошие и плохие новости и разрабатывать законные стратегии, которые позволяли бы неизменно достигать результата выше рыночного. Чтобы быть хорошим инвестором, вовсе не обязательно вступать в конфликт с законом. Как компьютер улавливает разницу между удачливыми инвесторами, действующими в рамках закона, и удачливыми инвесторами, преступившими закон? Я несколько раз звонил в отдел правоприменения SEC, чтобы выяснить это, однако сотрудники SEC не желали делиться со мной своими секретами.

В фильме 2002 года *Minority Report* Том Круз играет детектива, предотвращающего преступления. Его герой является сотрудником некоего бюро, которое использует определенную технологию для прогнозирования преступлений еще до того, как они будут совершены.

Нет, дорогие читатели, это уже не фантастика. В 2011 году в газете *The New York Times* вышла статья под заголовком: «Полиция прибывает на место до совершения преступления»¹¹. В ней рассказывалось, что специальная компьютерная программа предсказала высокую вероятность совершения краж из автомобилей в этот день на подземной парковке, расположенной в деловом районе города Санта-Круз. Когда туда приехали детективы, они обнаружили двух женщин, слишком уж пристально всматривающихся в окна автомобилей. Одна из них уже неоднократно задерживалась за воровство, а у другой нашли запрещенные наркотики.

Система, использовавшаяся в Санта-Круз, была разработана двумя математиками, антропологом и криминалистом. Отдел полиции в Чикаго создал у себя целое подразделение аналитиков-прогнозистов. Частично его формирование объяснялось тем, что банды, терроризировавшие город, действовали по определенным шаблонам. Книга *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis*, руководство по статистике для правоприменения, начинается со следующего бодрого заявления: «Теперь можно составлять прогноз в отношении уголовных преступлений; например выявлять тенденции развития преступности, прогнозировать “горячие точки” совершения преступлений, уточнять решения, касающиеся выделения ресурсов на те или иные цели, и обеспечивать максимальную защиту граждан наиболее эффективными способами». (Обратите внимание: я ознакомился со всей этой информацией, чтобы кратко изложить ее для вас и сэкономить вам таким образом время.)

«Прогнозная полиция» является частью более широкого движения, которое называется «предиктивным анализом». Уголовные преступления всегда включают в себя элемент неопределенности: вы никогда не сможете сказать заранее, кто разобьет ваш автомобиль или не выполнит обязательств по вашей ипотеке. Вероятность помогает нам ориентироваться в таких рисках. А информация позволяет уточнить понимание соответствующих вероятностей. Компании, сталкивающиеся в своей деятельности с неопределенностью, всегда пытаются получить количественную оценку рисков. Кредиторы интересуются такими вещами, как реальный доход получателя кредита и его

кредитный рейтинг. Однако эти несовершенные кредитные инструменты все больше напоминают нам некий эквивалент каменных орудий пещерного человека. Сочетание огромных объемов цифровых данных и дешевой вычислительной мощности позволяет нам гораздо лучше понимать поведение человека. Представители страховых компаний правильно описывают свой бизнес как «передача риска» — и поэтому им следует как можно точнее оценивать риски, переносимые на них. Особенности бизнеса, которым занимаются такие компании, как Allstate Corporation, заставляют их обращать самое пристальное внимание на вещи, которые стороннему наблюдателю могли бы показаться ничего не значащими случайностями:¹²

- в дорожно-транспортные происшествия со смертельным исходом чаще всего попадают водители в возрасте от двадцати до двадцати четырех лет;
- в штате Иллинойс чаще всего угоняют автомобили марки Honda Civic (а в штате Алабама — полноразмерные пикапы Chevrolet)*;
- хотя законом запрещена отправка SMS во время вождения, поскольку это часто приводит к ДТП, водителей это не останавливает. Более того, подобные законы могут даже усугублять ситуацию, заставляя водителей прятать мобильные телефоны и в результате отвлекаться от дороги во время набора сообщения.

Компании, выпускающие кредитные карточки, находятся на переднем крае такого анализа, поскольку они знают наши личные данные и покупательские привычки, а их модель ведения бизнеса сильно зависит от умения находить клиентов, кредитный риск (то есть риск неплатежа) которых сравнительно невелик. (Идеальные клиенты с точки зрения кредитного риска, как правило, расточительны, так как каждый месяц полностью оплачивают

* Так как я советовал вам с осторожностью относиться к описательным статистикам, я чувствую себя обязанным отметить, что автомобиль, который угоняют чаще всего, вовсе не обязательно является автомобилем, который угоняют вероятнее всего. Большое число автомобилей марки Honda Civic угоняют именно потому, что это самая распространенная марка, между тем как вероятность угона какого-либо отдельно взятого автомобиля марки Honda Civic (а именно это интересует страховые компании, страхующие от угона автомобилей) может оказаться весьма низкой. Напротив, даже если угоняют 99% всех автомобилей Ferrari, автомобиль этой марки не возглавил бы список «наиболее часто угоняемых», поскольку таких автомобилей сравнительно мало и, следовательно, их угоняют довольно редко.

свои счета; клиенты, располагающие крупными балансами с высокими процентными ставками, генерируют солидные прибыли — пока не окажутся неплатежеспособными.) Одно из самых интересных исследований того, кто, скорее всего, оплатит счет, а кто нет, было выполнено Дж. П. Мартином, любителем математики и одним из сотрудников компании Canadian Tire, крупной сети розничной торговли, специализирующейся на продаже широкого спектра автозапчастей и автомобильных аксессуаров, а также ряда других товаров¹³. Когда Мартин проанализировал каждую транзакцию, выполненную с помощью кредитной карточки Canadian Tire за предыдущий год, оказалось, что покупки, совершенные клиентами в прошлом, являются весьма точным предиктором их будущего «покупательского» поведения, если использовать этот показатель в сочетании с такими традиционными инструментами, как величина дохода и кредитная история.

Статья в The New York Times, озаглавленная «Что знает о вас компания, выпустившая вашу кредитную карточку?», содержала описание некоторых из самых интересных выводов Мартина: «Люди, которые покупают дешевые непатентованные моторные масла, с гораздо большей вероятностью уклонятся от платежей по кредитным карточкам, чем те, кто предпочитает дорогостоящие фирменные товары. Те, кто покупает датчики угарного газа для дома или мягкие войлочные подкладки для ножек стульев и табуреток, чтобы не царапать пол в комнате, почти никогда не увиливают от платежей. Практически каждый, кто купил какой-либо из дешевых автомобильных аксессуаров, впоследствии с большой долей вероятности не оплатит свой счет».

* * *

Теория вероятностей предоставляет нам инструменты для борьбы с неопределенностями жизни. Не стоит играть в лотерею. Но имеет смысл инвестировать в рынок ценных бумаг, если у вас длинный инвестиционный горизонт (поскольку доход от акций, как правило, достигает своих максимальных значений в долгосрочном периоде). Что же касается страхования, то здесь все зависит от того, что именно вы собираетесь застраховать. Учет фактора вероятности может даже помочь вам увеличить выигрыши в игровых шоу (я попытаюсь продемонстрировать это в следующей главе).

С учетом вышесказанного (точнее говоря, написанного) концепция вероятности не является детерминистской. Да, от покупки лотерейных билетов

следует воздержаться — тем не менее, купив лотерейный билет, вы *можете* выиграть деньги. Да, теория вероятностей может помочь нам поймать мошенников и уголовных преступников, но в случае ее неаккуратного использования за решеткой могут оказаться ни в чем не повинные люди. Все эти вопросы мы обсудим в главе 6.

«Загадка Монти Холла» — знаменитая задача по теории вероятностей, поставившая в тупик участников игрового шоу под названием *Let's Make a Deal* («Совершим сделку»), до сих пор популярного в ряде стран, премьера которого состоялась в Соединенных Штатах в 1963 году. (Помню, я всякий раз смотрел это шоу в детстве, когда не ходил в школу по причине болезни.) Во введении к книге я уже указывал, *что* в этом игровом шоу может быть интересно для статистиков. В конце каждого его выпуска участник, добравшийся до финала, становился вместе с Монти Холлом перед тремя большими дверями: Дверью № 1, Дверью № 2 и Дверью № 3. Монти Холл объяснял финалисту, что за одной из этих дверей скрывается очень ценный приз — например новый автомобиль, а за двумя другими — козел. Финалист должен был выбрать одну из дверей и получить то, что за ней находилось. (Я не знаю, был ли среди участников шоу хотя бы один человек, желающий получить козла, но для простоты рассуждений будем полагать, что подавляющее большинство участников мечтали о новом автомобиле.)

Начальную вероятность выигрыша определить довольно просто. Есть три двери, за двумя скрывается козел, а за третьей — автомобиль. Когда участник шоу вместе с Монти Холлом стоит перед этими дверями, у него есть один шанс из трех выбрать дверь, за которой находится автомобиль. Но, как отмечалось выше, в *Let's Make a Deal* кроется подвох, увековечивший эту телепрограмму и ее ведущего в литературе по теории вероятностей. После того как финалист шоу укажет на какую-то из трех дверей, Монти Холл открывает одну из двух оставшихся дверей, за которой всегда находится козел. Затем Монти Холл спрашивает финалиста, не желает ли он изменить свое решение, то есть отказаться от ранее выбранной им закрытой двери в пользу другой закрытой двери.

Допустим, ради примера, что участник указал на Дверь № 1. Затем Монти Холл открыл Дверь № 3, за которой скрывался козел. Две двери, Дверь № 1 и Дверь № 2, по-прежнему остаются закрытыми. Если бы ценный приз находился за Дверью № 1, финалист выиграл бы его, а если за Дверью № 2, то проиграл бы. Именно в этот момент Монти Холл обращается к игроку с вопросом, не желает ли он изменить свой первоначальный выбор (в данном случае отказаться от Двери № 1 в пользу Двери № 2). Вы, конечно, помните, что обе двери пока закрыты. Единственная новая информация, которую участник получил, состоит в том, что козел оказался за одной из двух дверей, которые он не выбрал.

Следует ли финалисту отказаться от первоначального выбора в пользу Двери № 2?

Отвечаю: да, следует. Если он будет придерживаться первоначального выбора, то вероятность выигрыша им ценного приза составит $\frac{1}{3}$; если же передумает и укажет на Дверь № 2, то вероятность выигрыша ценного приза будет $\frac{2}{3}$. Если не верите мне, читайте дальше.

Признаю, что такой ответ на первый взгляд далеко не очевиден. Кажется, что, какую бы из оставшихся двух дверей ни выбрал финалист, вероятность получения ценного приза в обоих случаях равняется $\frac{1}{3}$. Есть три закрытые двери. Поначалу вероятность того, что ценный приз скрывается за любой из них, составляет $\frac{1}{3}$. Разве имеет какое-то значение решение финалиста поменять свой выбор в пользу другой закрытой двери?

Безусловно, поскольку заковыка заключается в том, что Монти Холл знает, *что* находится за каждой дверью. Если финалист выберет Дверь № 1 и за ней действительно будет автомобиль, то Монти Холл может открыть либо Дверь № 2, либо Дверь № 3, чтобы продемонстрировать козла, скрывающегося за ней.

Если финалист выберет Дверь № 1, а автомобиль будет за Дверью № 2, то Монти Холл откроет Дверь № 3.

Если же финалист укажет на Дверь № 1, а автомобиль окажется за Дверью № 3, то Монти Холл откроет Дверь № 2.

Изменив свое решение после того, как ведущий откроет какую-то из дверей, финалист получает преимущество выбора двух дверей вместо одной. Я попытаюсь убедить вас в правильности этого анализа тремя разными способами.

Первый — эмпирический. В 2008 году колумнист газеты The New York Times Джон Тайерни написал материал о «феномене Монти Холла»¹. После этого сотрудники издания разработали интерактивную программу, которая позволяет вам сыграть в эту игру и самостоятельно принять решение, менять свой первоначальный выбор или нет. (В программе даже предусмотрены маленькие козлики и автомобильчики, которые появляются из-за дверей.) Программа фиксирует ваши выигрыши в случае, когда вы меняете свой первоначальный выбор, и в случае, когда остаетесь при своем мнении. Поэкспериментируйте сами*. Я заплатил одной из своих дочерей за то, чтобы она сыграла в эту игру 100 раз, каждый раз меняя первоначальный выбор. Я также заплатил ее брату, чтобы он тоже сыграл в эту игру 100 раз, каждый раз оставляя первоначальное решение. Дочь выиграла 72 раза; ее брат — 33 раза. Усилия каждого были вознаграждены двумя долларами.

Данные из эпизодов игры Let's Make a Deal свидетельствуют о такой же закономерности. Согласно Леонарду Млодинову, автору книги The Drunkard's Walk, те из финалистов, кто изменил свой первоначальный выбор, становились победителями примерно в два раза чаще, чем те, кто оставался при своем мнении².

Мое второе объяснение данного феномена основывается на интуиции. Допустим, правила игры слегка поменялись. Например, финалист начинает с выбора одной из трех дверей: Двери № 1, Двери № 2 и Двери № 3, как и было предусмотрено изначально. Однако затем, прежде чем открыть какую-то из дверей, за которой скрывается козел, Монти Холл спрашивает: «Согласны ли вы отказаться от своего выбора в обмен на открывание *двух оставшихся дверей?*» Таким образом, если вы выбрали Дверь № 1, вы можете передумать в пользу Двери № 2 и Двери № 3. Если сперва указали на Дверь № 3, можете выбрать Дверь № 1 и Дверь № 2. И так далее.

Для вас это было бы не особо трудным решением: совершенно очевидно, что вам следует отказаться от первоначального выбора в пользу двух оставшихся дверей, поскольку это повышает шансы на выигрыш с $\frac{1}{3}$ до $\frac{2}{3}$. Самое интересное, что *именно такой в сущности вариант предлагает вам Монти Холл в реальной игре, после того как откроет дверь, за которой скрывается козел*. Принципиальный факт заключается в том, что если бы вам была предоставлена возможность выбрать две двери, за одной из них в *любом* случае

* Вы можете сыграть в эту игру на сайте http://www.nytimes.com/2008/04/08/science/08monty.html?_r=2&oref=slogin&oref=slogin.

скрывался бы козел. Когда Монти Холл открывает дверь, за которой находится козел, и только после этого спрашивает вас, согласны ли вы изменить свой первоначальный выбор, он существенно повышает ваши шансы на выигрыш ценного приза! По сути, Монти Холл говорит вам: «Вероятность того, что ценный приз скрывается за одной из двух дверей, которые вы не выбрали с первого раза, составляет $\frac{2}{3}$, а это все-таки больше чем $\frac{1}{3}$!»

Это можно представить себе так. Допустим, вы указали на Дверь № 1. После этого Монти Холл дает вам возможность отказаться от первоначального решения в пользу Двери № 2 и Двери № 3. Вы соглашаетесь и получаете в свое распоряжение две двери, а это означает, что у вас есть все основания рассчитывать на выигрыш ценного приза с вероятностью $\frac{2}{3}$, а не $\frac{1}{3}$. А что было бы, если бы в этот момент Монти Холл открыл Дверь № 3 — одну из «ваших» дверей, — и за ней оказался бы козел? Поколебал бы этот факт вашу уверенность в принятом решении? Конечно же нет. Если бы автомобиль скрывался за Дверью № 3, Монти Холл открыл бы Дверь № 2! *Он бы ничего вам не показал.*

Когда игра идет по накатанному сценарию, Монти Холл действительно предоставляет вам выбор между дверью, которую вы указали поначалу, и двумя оставшимися дверями, за одной из которых *может* находиться автомобиль. Когда Монти Холл открывает дверь, за которой скрывается козел, он просто оказывает вам любезность, демонстрируя, за какой из двух других дверей нет автомобиля. Вы располагаете одинаковыми вероятностями выигрыша в обоих из указанных ниже сценариев.

1. Выбор Двери № 1, затем согласие «переключиться» на Дверь № 2 и Дверь № 3 еще до того, как будет открыта какая-либо дверь.
2. Выбор Двери № 1, затем согласие «переключиться» на Дверь № 2, после того как Монти Холл продемонстрирует вам козла за Дверью № 3 (или выбор Двери № 3, после того как Монти Холл продемонстрирует вам козла за Дверью № 2).

В обоих случаях отказ от первоначального решения обеспечивает вам преимущество двух дверей по сравнению с одной, и вы можете таким образом удвоить свои шансы на выигрыш: с $\frac{1}{3}$ до $\frac{2}{3}$.

Мой третий вариант представляет собой более радикальную версию той же базовой интуиции. Допустим, Монти Холл предлагает вам выбрать одну из 100 дверей (вместо одной из трех). После того как вы это сделаете,

скажем, указав на Дверь № 47, он открывает 98 оставшихся дверей, за которыми оказываются козлы. Теперь закрытыми остаются всего две двери: ваша Дверь № 47 и еще одна, например Дверь № 61. Следует ли вам отказаться от своего первоначального выбора?

Разумеется да! С 99-процентной вероятностью автомобиль находится за одной из дверей, которые вы не выбрали поначалу. Монти Холл оказал вам любезность, открыв 98 таких дверей, за ними автомобиля не было. Таким образом, существует лишь 1 из 100 шансов, что ваш первоначальный выбор (Дверь № 47) будет правильным. В то же время существует 99 из 100 шансов, что ваш первоначальный выбор неправильный. А если так, то автомобиль находится за оставшейся дверью, то есть Дверью № 61. Если вы хотите сыграть с вероятностью выигрыша в 99 случаях из 100, то вам следует «переключиться» на Дверь № 61.

Короче говоря, если вам когда-нибудь придется участвовать в игре Let's Make a Deal, вам, безусловно, нужно отказаться от своего первоначального решения, когда Монти Холл (или тот, кто будет его замещать) предоставит вам возможность выбора. Более универсальный вывод из этого примера состоит в том, что ваши интуитивные догадки относительно вероятности наступления тех или иных событий могут подчас вводить вас в заблуждение.

Проблемы с вероятностью

6

Как самоуверенные знатоки математики едва не разрушили глобальную финансовую систему

Статистика не может быть более совершенной, чем люди, которые ее используют. Но иногда она заставляет умных людей делать глупости. Одним из самых безответственных случаев применения статистики за последнее время стал механизм оценивания рисков на Уолл-стрит перед финансовым кризисом 2008 года. В то время компании, представляющие финансовый сектор, использовали общепринятый барометр риска — модель стоимости риска, или рискованной стоимости (Value-at-Risk — VaR). Теоретически VaR сочетала в себе элегантность индикатора (совмещающая обширную информацию в едином числовом показателе) с мощностью вероятности (присоединяя ожидаемую прибыль или убыток к каждому из активов или торговым позициям соответствующей фирмы). Такая модель исходила из того, что для каждой инвестиции компании существует определенный диапазон возможных исходов. Если, например, компания владеет акциями General Electric, то их стоимость может повышаться или понижаться. Когда VaR вычисляется для некоего короткого промежутка времени, например недели, то самым вероятным исходом станет то, что в конце данного периода у этих акций будет примерно такая же стоимость, как и в начале. Вероятность того, что их стоимость повысится или снизится на 10%, относительно невелика. Еще меньше вероятность того, что она повысится или снизится на 25%, и т. д.

На основе прошлых данных о движениях рынка «количественные» эксперты компании (в сфере финансов их часто называют «квантами» [от слова quantitative, то есть «количественный»], а во всех остальных отраслях — «богатырьскими ботанами») могли определить максимальную сумму в денежном выражении (например, 13 миллионов долларов), которую фирма может с 99-процентной вероятностью потерять на данной позиции в течение рассматриваемого периода времени. Другими словами, в 99 случаях из 100 компания не потеряет более 13 миллионов долларов на конкретной торговой позиции; а в 1 случае из 100 потеряет.

Запомните последнее утверждение, поскольку вскоре оно станет важным.

До финансового кризиса 2008 года фирмы охотно использовали модель VaR для оценки своего суммарного риска. Если у какого-либо отдельно взятого трейдера было 923 различных открытых позиции (инвестиций, стоимость которых могла расти или падать), то каждую из таких инвестиций можно было оценить, как описано выше для акций General Electric, и на основе этого вычислить совокупный риск портфеля данного трейдера. Формула даже учитывала корреляции между разными позициями. Если, например, ожидаемые доходности двух инвестиций отрицательно коррелированы между собой, то убыток по одной из них, скорее всего, будет компенсирован прибылью по другой; таким образом, две инвестиции в совокупности менее рискованны, чем каждая в отдельности. В целом глава торгового отдела должен знать, что, скажем, у Боба Смита, торгующего облигациями, 24-часовая VaR (стоимость риска в течение ближайших 24 часов) 19 миллионов долларов — как указывалось выше, с 99-процентной вероятностью. Максимум, что может потерять Боб Смит в течение ближайших 24 часов, это 19 миллионов долларов — в 99 случаях из 100.

К тому же, что еще лучше, в любой момент можно вычислить совокупный риск для компании, дополнив тот же самый базовый процесс. Математические механизмы, положенные в его основу, по-видимому, невероятно сложны, поскольку каждая из фирм располагает огромным массивом инвестиций в разных валютах, с разными величинами рычагов (по-другому, левверидж, или кредитное плечо; сумма, которая заимствована для выполнения соответствующей инвестиции), торгующихся на рынках с разными степенями ликвидности, и т. д. Несмотря на все это, менеджеры фирмы якобы располагали точным показателем величины риска, принимаемого ею на себя в любой момент времени. Как поясняет бывший экономический обозреватель The New York Times Джо Носера: «Огромная привлекательность VaR в глазах людей, которые не принадлежат к числу “квантов”, заключается в том, что она представляет риск в виде единого числа, в денежном выражении — и никак не меньше!»¹ В банке J. P. Morgan, где была разработана и неоднократно уточнялась модель VaR, ее ежесуточное вычисление носило название «отчет 4:15», так как результат этого расчета появлялся на рабочих столах высшего руководства каждый день в 16:15 — сразу же после закрытия в тот день американских финансовых рынков.

По-видимому, это был неплохой вариант, поскольку наличие дополнительной информации в любом случае лучше, особенно когда речь идет о риске. В конце концов, вероятность — довольно мощный инструмент. Разве эти вычисления принципиально отличаются от тех, которыми занималось руководство компании Joseph Schlitz Brewing Company, прежде чем потратить кучу денег на проведение слепой сравнительной дегустации пива в перерыве матча за Суперкубок?

Как сказать... Как только ни называли в свое время показатель VaR: и «потенциально катастрофический», и «надувательский», и... (да, были и другие эпитеты, которые вряд ли следует упоминать в такой солидной книге по статистике, как эта). В частности, именно эту модель обвиняли в наступлении финансового кризиса, разразившегося в 2008 году. Главной причиной критики в адрес VaR является то, что фундаментальные риски, связанные с финансовыми рынками, невозможно предсказать по аналогии с подбрасыванием монетки или слепой сравнительной дегустацией двух сортов пива. Ложное ощущение точности, встроенное в эти модели, породило ложное ощущение безопасности. Показатель VaR был похож на неисправный спидометр; пожалуй, это хуже, чем если бы его не было вообще. Понадеявшись на неисправный спидометр, вы перестанете обращать внимание на другие признаки того, что уже превысили допустимую скорость. В случае же отсутствия спидометра вам придется отслеживать признаки, указывающие на реальную скорость движения автомобиля.

Примерно в 2005 году, ориентируясь исключительно на показатели VaR, которые ежедневно появлялись на рабочих столах руководителей компаний ровно в 16:15, Уолл-стрит набрала скорость, существенно превышающую допустимую. К сожалению, с профилями риска, заложенными в моделях VaR, существовали две огромные проблемы. Во-первых, вероятности, на которых строились эти модели, исходили из прошлых движений рынка; однако на финансовых рынках (в отличие от дегустации пива) будущее вовсе не обязательно должно быть похожим на прошлое. Таким образом, не было никаких оснований полагать, что движения рынка в период с 1980 по 2005 год были наилучшим предиктором изменений на рынке после 2005 года. В какой-то степени этот недостаток воображения напоминает периодические ошибочные предположения генералов о том, что следующая война будет похожа на предыдущую. В 1990-е годы, а также в начале нулевых коммерческие банки широко применяли модели кредитования для жилищных ипотек,

согласно которым вероятность значительного снижения цен на жилье близилась к нулю². Цены на жилье никогда ранее не падали так сильно и так быстро, как это происходило с начала 2007 года. Однако случилось то, что случилось. Бывший глава Федеральной резервной системы Алан Гринспен, выступая впоследствии перед членами одного из комитетов Конгресса США, так объяснял этот факт: «Все это величественное интеллектуальное здание рухнуло летом 2007 года, поскольку данные, вводимые в модели управления риском, охватывали лишь два последних десятилетия, то есть период, когда всех нас захлестнула эйфория. Между тем, если бы мы использовали более подходящие модели, затрагивающие исторические периоды, характеризующиеся экономическим неблагополучием, то, как мне кажется, требования к капиталу оказались бы значительно выше, а финансовый мир чувствовал бы себя гораздо лучше»³.

Кроме того, даже если бы исходные данные могли точно прогнозировать будущий риск, 99-процентная гарантия, обещанная моделью VaR, была опасно бесполезной, *поскольку остающийся 1% действительно вводит в заблуждение*. Менеджер хеджевого фонда Дэвид Айнхорн поясняет: «Это как подушка безопасности, которая дает сбой именно в момент автокатастрофы». Если стоимость риска (VaR) какой-либо компании составляет 500 миллионов долларов, то это можно рассматривать как 99-процентную вероятность того, что на протяжении указанного периода фирма потеряет не более этой суммы. Но это также означает, что данная компания может с 1-процентной вероятностью потерять свыше 500 миллионов долларов (а при определенных обстоятельствах даже *значительно* больше). По сути, опираясь на эти модели, невозможно предусмотреть, насколько плохим может оказаться 1-процентный сценарий. Очень мало внимания уделялось так называемому хвостовому, то есть малому риску (производное от хвоста кривой распределения) какого-то катастрофического исхода. (Если вы возвращаетесь домой из ресторана за рулем своего автомобиля и уровень алкоголя в вашей крови равен 0,15 промилле, то вероятность того, что вы попадете в ДТП со смертельным исходом, наверное, будет менее 1%; тем не менее это не повод садиться за руль в нетрезвом виде.) Многие компании усугубили эту ошибку, сделав нереалистичное предположение о своей готовности к маловероятным событиям. Бывший глава Казначейства США Хэнк Поулсон пояснил, что большинство из них надеялись в крайнем случае привлечь денежные средства путем продажи активов⁴. Но во время кризиса деньги нужны *всем*,

поэтому все пытаются продать те или иные активы. С точки зрения управления рисками это равносильно тому, как если бы вы сказали: «Мне нет нужды запастись водой и продуктами питания, поскольку в случае стихийного бедствия я смогу пойти в супермаркет и купить все необходимое». Разумеется, после того как астероид упадет на ваш город, его пятьдесят тысяч жителей ринутся в супермаркеты, чтобы запастись водой и продуктами питания, но к тому моменту, когда вы доберетесь до ближайшего супермаркета, окна в нем будут разбиты, а полки пусты.

То обстоятельство, что вы никогда всерьез не рассматривали возможность падения на ваш город крупного астероида, в точности описывает проблему с VaR. Вот еще одна выдержка из статьи колумниста The New York Times Джо Носера, который подытоживает мысли Николаса Талеба, автора книги *The Black Swan: The Impact of the Highly Improbable** и яростного критика VaR: «Самые опасные — отнюдь не риски, которые вы можете увидеть и измерить, а риски, которые вы не можете увидеть и, следовательно, измерить. Это риски, находящиеся настолько далеко за пределами нормальной вероятности, что невозможно даже себе представить, что они могут произойти в вашей жизни, — хотя, конечно же, они случаются, и даже чаще, чем вы могли бы предположить».

В каком-то смысле фиаско VaR является полной противоположностью примера с компанией Joseph Schlitz Brewing Company, приведенного в главе 5. Данные о вероятности выбора пива в ходе слепой сравнительной дегустации, которыми располагала эта компания, позволили ей примерно предугадать поведение дегустаторов в ходе сравнительной дегустации, транслируемой в прямом эфире во время перерыва матча за Суперкубок. Компании даже удалось обернуть себе на пользу то обстоятельство, что в акции участвовали только любители *других* сортов пива. Даже если бы пиво Schlitz предпочли не более 25% любителей пива Michelob (практически нереальный исход), компания все равно могла бы сказать, что по крайней мере одному из каждых четырех любителей пива следовало бы переключиться на пиво Schlitz. И, возможно, самое важное: здесь речь шла лишь о пиве, а не о глобальной финансовой системе. «Кванты» с Уолл-стрит совершили три фундаментальные ошибки. Во-первых, они спутали точность с достоверностью. Модели VaR действовали подобно моему дальномеру, который был настроен

* Издана на русском языке: *Талеб Н. Черный лебедь. Под знаком непредсказуемости.* — М.: КоЛибри, 2009.

на измерение расстояний в метрах, а не в ярдах, в результате чего расстояния измерялись точно, но неправильно. Эта ложная точность заставила обитателей Уолл-стрит поверить, будто они контролируют риск, хотя на самом деле это было не так. Во-вторых, оценки вероятностей, положенные в основу вычислений согласно модели VaR, оказались ошибочными. Как указывал Алан Гринспен, выступая на слушаниях в одном из комитетов Конгресса США (цитату из его выступления я приводил чуть выше), относительно безмятежные и благополучные десятилетия до 2005 года не следовало брать за основу при построении распределений вероятностей, которые использовались для прогнозирования поведения рынков в предстоящие десятилетия. Это как если бы вы отправились в казино с твердой уверенностью, что сегодня выиграете в рулетку в 62 случаях из ста только потому, что именно так получилось вчера, когда удача сопутствовала вам. Подобная уверенность обошлась бы вам очень дорого! В-третьих, компании пренебрегли «хвостовым риском». Модели VaR прогнозируют, что должно произойти в 99 случаях из ста. Именно таков механизм действия вероятностей (во второй половине книги это обстоятельство будет подчеркиваться неоднократно). Между тем маловероятные события время от времени случаются. Более того, в долгосрочном периоде они не так уж и маловероятны. Иногда в людей попадает молния. Моя мать убедилась в этом на собственном опыте.

«Статистическое высокомерие», продемонстрированное коммерческими банками и на Уолл-стрит, в конечном счете сыграло ключевую роль в самом жестоком глобальном финансовом кризисе со времен Великой депрессии. Этот кризис, разразившийся в 2008 году, серьезно подорвал финансовое благополучие Соединенных Штатов, повысил уровень безработицы до более чем 10%, породил волну банкротств и отчуждений имущества и заставил многие государства, пытавшиеся минимизировать экономический ущерб, влезть в огромные долги. Подобный исход особенно печален потому, что столь изощренные инструменты, как VaR, обязаны были снизить угрозу риска.

Теория вероятностей предоставляет в наше распоряжение мощный и полезный набор инструментов, правильное использование которых поможет лучше уяснить ситуацию, складывающуюся в мире; а неправильное посеет в нем хаос. В русле метафоры «статистика как мощное оружие», которая неоднократно повторяется в этой книге, я хочу перефразировать любимое выражение сторонников свободной продажи огнестрельного оружия в нашей стране: ошибается не теория вероятностей, а люди, которые ею пользуются.

Далее в этой главе я перечислю ряд самых распространенных ошибок, заблуждений и этических дилемм, связанных с применением концепции вероятности.

Предполагается, что события независимы, тогда как на самом деле они зависимы друг от друга. Вероятность выпадания решки при подбрасывании «правильной» монетки равняется $\frac{1}{2}$. Вероятность двукратного (подряд) выпадания решки при подбрасывании такой же монетки составляет $(\frac{1}{2})^2$, или $\frac{1}{4}$, поскольку вероятность *одновременного* наступления двух независимых событий равняется произведению их индивидуальных вероятностей. Теперь, когда вы вооружены этим важным знанием, допустим, что вас назначили на должность начальника отдела управления рисками в крупной авиакомпании. Ваш заместитель сообщает вам, что вероятность выхода из строя по тем или иным причинам авиадвигателя во время трансатлантического перелета составляет 1 шанс из 100 000. Учитывая количество трансатлантических перелетов, этот риск нельзя считать приемлемым. К счастью, каждый современный самолет, совершающий такие перелеты, оснащен по меньшей мере двумя двигателями. Ваш заместитель подсчитал, что риск одновременного выхода из строя обоих во время трансатлантического перелета должен равняться $(1/100\,000)^2$, или 1 шансу из 10 миллиардов, что считается вполне приемлемым риском с точки зрения обеспечения безопасности полетов. Что же, сейчас самое время предложить вашему заместителю взять отпуск и подготовиться к увольнению. Поломка обоих авиадвигателей не относится к категории независимых событий. Если во время взлета самолет наталкивается на стаю гусей, то, вероятнее всего, оба двигателя выйдут из строя одинаковым образом. То же самое можно сказать о многих других факторах, влияющих на функционирование авиадвигателя, начиная с погодных условий и заканчивая небрежным выполнением своих обязанностей наземными службами техобслуживания. Если один двигатель выйдет из строя, то вероятность поломки второго будет значительно выше, чем 1 шанс из 100 000.

Это очевидно, не правда ли? Однако британским прокурорам это показалось не столь очевидным в 1990-е, когда они совершили серьезную судебную ошибку вследствие некорректного использования теории вероятностей. Как и в гипотетическом примере с авиадвигателями, ошибка заключалась в предположении о независимости нескольких событий (как с подбрасыванием монетки), хотя на самом деле они были зависимы (то есть когда какой-то

определенный исход повышает вероятность аналогичного исхода в будущем). Тем не менее эта теоретическая ошибка стоила свободы абсолютно невинным людям, которые в результате оказались за решеткой.

Эта история произошла в контексте так называемого синдрома внезапной смерти младенцев во время сна (СВСМ) — явления, когда вполне здоровый малыш умирает в своей кроватке. (У британцев СВСМ принято называть «смертью в колыбели».) Долгое время СВСМ оставался медицинской загадкой, которая привлекала к себе все большее внимание по мере снижения детской смертности по другим причинам*. Поскольку СВСМ настолько таинственен и малопонятен, его феномен породил всевозможные подозрения. Иногда они подтверждались. Время от времени ссылки на СВСМ использовались, чтобы скрыть факты небрежного выполнения родительских обязанностей или даже предумышленного убийства, так как вскрытие далеко не всегда позволяет отличить смерть в силу естественных причин от убийства. Британские прокуроры и суды были убеждены, что один из способов правильно определять причины СВСМ — повесить внимание к семьям с повторными случаями «смерти в колыбели». Сэр Рой Мидоу, известный британский педиатр, часто привлекался к рассмотрению подобных случаев в качестве эксперта. Как поясняется в британском журнале *The Economist*: «Мысль, которая пришла в голову Рою Мидоу и стала впоследствии известной как “закон Мидоу” (суть ее в том, что одна младенческая смерть — это трагедия, две смерти вызывают подозрение, а три — это убийство), основывается на том, что если какое-либо событие является достаточно редким, то два или большее число его наступлений в одной и той же семье настолько маловероятны, что нет никаких оснований считать это простой случайностью»⁵. Сэр Рой Мидоу объяснил присяжным, что вероятность внезапной смерти от естественных причин двух младенцев в одной семье чрезвычайно мала и равняется примерно одному шансу из 73 миллионов. Он толковал свои подсчеты так: поскольку случаи «смерти в колыбели» встречаются довольно редко (1 из 8500), вероятность наступления двух смертей в колыбели в одной и той же семье составляет $(1/8500)^2$, что равняется примерно одному шансу из 73 миллионов. Так что здесь явно пахнет предумышленным убийством. Руководствуясь этими доводами, присяжные выносили свои

* СВСМ по-прежнему остается медицинской загадкой, хотя многие из факторов риска, связанных с этим феноменом, удалось выявить. Например, смертность у младенцев можно резко снизить, если ребенка укладывать спать на спину.

вердикты. В результате, основываясь на статистике смертей в колыбели, присяжные отправили за решетку немалое число родителей (зачастую без учета каких-либо медицинских свидетельств, указывающих на их неумелое обращение с ребенком). В некоторых случаях у родителей, относительно которых возникали подозрения, вызванные необъяснимой смертью кого-либо из их детей в младенческом возрасте, последующих детей отбирали сразу же после рождения.

The Economist объясняет, каким образом неправильная трактовка статистической независимости могла привести к ошибочным выводам в докладе, с которым Мидоу выступал перед присяжными:

Как указывает Королевское статистическое общество (Royal Statistical Society), в рассуждениях Мидоу есть очевидный изъян. Выполненный им подсчет вероятности был бы правильным, если бы смерти в колыбели носили совершенно случайный характер и не были бы связаны с каким-то неизвестным фактором. Но когда речь идет о столь загадочном феномене, как смерть в колыбели, вполне возможно наличие какой-то связи, например некоего генетического фактора, вследствие действия которого угроза потерять по той же причине еще одного ребенка в семье, уже лишившейся одного малыша, гораздо выше, (а не ниже), чем в семьях, где таких случаев не зафиксировано. После того как в результате повторных смертей в колыбели многие родители оказались за решеткой, ученые поверили в реальность существования такой связи.

В 2004 году британское правительство объявило о предстоящем пересмотре 258 приговоров, согласно которым родители, обвинявшиеся в умышленном лишении жизни своих детей, отбывают тюремный срок.

Непонимание, когда события ДЕЙСТВИТЕЛЬНО независимы друг от друга. Еще одна разновидность ошибок возникает, когда события, действительно независимы друг от друга, рассматриваются как взаимосвязанные. Если вы когда-либо окажетесь в казино (место, в котором, с точки зрения статистики, вам лучше вообще не появляться), то обязательно увидите людей, вперившихся взглядом в игральные кости или карты и заявляющих, что они «ожидают должное». Если шарик рулетки пять раз подряд остановился на черном поле, то всякому здравомыслящему человеку понятно, что на следующий раз должно выпасть красное. Нет, нет и еще раз нет!

Вероятность того, что шарик остановится на красном поле, каждый раз будет одной и той же: $16/38$. Уверенность в том, что это вовсе не так, иногда называют «заблуждением игрока». В действительности, если «правильную» монетку подбросить 1 000 000 раз и каждый раз будет выпадать решка, то вероятность того, что на 1 000 001-й раз выпадет орел, по-прежнему останется $\frac{1}{2}$. Само определение статистической независимости двух событий заключается в том, что исход одного события никак не сказывается на исходе другого. Даже если статистика не убеждает вас, обратитесь к физике соответствующего явления: каким образом выпадание решки несколько раз подряд может повлиять на вероятность выпадания орла в результате следующего подбрасывания монетки?*

Даже в спорте представление о полосе удач и неудач может оказаться иллюзорным. В одной из самых знаменитых и интересных научных статей, посвященных вероятностям, опровергается общепринятое утверждение о том, что в течение одной игры у баскетболистов периодически возникает некая «полоса везения», когда один за другим следуют удачные броски по кольцу (в таких случаях говорят, что игрок «набил себе руку»). Несомненно, большинство спортивных болельщиков станут вас уверять, что игрок, попавший по кольцу, с большей вероятностью попадет по нему при выполнении следующего броска, чем игрок, «промазавший» перед этим. Однако исследование, проведенное Томасом Гиловичем, Робертом Валлоне и Амосом Тверски, которые протестировали феномен «набитой руки» тремя разными способами, говорит об обратном⁶. Во-первых, они проанализировали данные о результатах бросков, сделанных в ходе домашних игр командой НБА «Филадельфия Севенти Сиксерс» (сезон 1980–1981 годов). (На момент его проведения аналогичные данные для других команд НБА отсутствовали.) И «не обнаружили каких-либо свидетельств положительной корреляции между результатами бросков, следующих друг за другом». Во-вторых, они проделали такое же исследование относительно результатов штрафных бросков в команде «Бостон Селтикс» и пришли к аналогичным выводам. Наконец, они провели управляемый эксперимент с членами мужской и женской баскетбольных

* Вместе с тем в теории вероятностей доказан факт, что если достаточно долго подбрасывать монету, то будут наблюдаться периоды преобладания выпадания орла или решки. Это так называемый первый закон арксинуса. Этот закон не отменяет сказанного автором, а только показывает структуру исходов в испытаниях Бернулли. О данном феномене см., например, классическую книгу В. Феллер. Введение в теорию вероятностей и ее приложения. Т. 1. Глава III. *Прим. ред.*

команд Корнелльского университета, игроки которых в среднем попадали по кольцу с игры в 48 случаях из 100, когда предыдущий бросок игрока был удачным, и в 47 случаях из 100, когда предыдущий бросок был неудачным. Для четырнадцати игроков в возрасте 26 лет корреляция между результатом выполнения одного броска и результатом выполнения следующего броска оказалась отрицательной. Лишь у одного баскетболиста обнаружилась значительная положительная корреляция между результатом выполнения двух следующих друг за другом бросков.

Разумеется, такой результат полностью расходится с мнением любителей баскетбола. Например, 91% любителей баскетбола, опрошенных исследователями в Стэнфордском и Корнелльском университетах, согласились с утверждением, что вероятность попадания игроком по кольцу после того, как он выполнил перед этим два или три удачных броска, будет выше, чем в случае, если перед этим он два или три раза промазал. Важный вывод относительно феномена «набитой руки» заключается в наличии разницы между восприятием и эмпирической реальностью. Исследователи замечают, что «интуитивные представления людей о случайности или закономерности тех или иных событий систематически расходятся с положениями теории вероятностей». Нам подчас свойственно усматривать закономерности там, где их и в помине нет.

Как, например, в случае с раковыми кластерами.

Кластеры действительно встречаются. Вы, наверное, читали в газетах (или видели репортаж по телевизору) о том, что в некоем регионе отмечена повышенная заболеваемость редкой формой рака. Возможно, причиной тому является вода, расположенная поблизости атомная электростанция или вышка сотовой связи. Разумеется, любой из перечисленных факторов может реально обусловить развитие столь опасной болезни. (В последующих главах я постараюсь показать, как с помощью статистики можно идентифицировать подобные причинно-следственные связи.) Однако этот кластер (совокупность) случаев заболеваний также может оказаться результатом чистой случайности, даже когда количество заболевших подозрительно велико. Да, вероятность того, что пять человек в одном и том же учебном заведении, или церковном приходе, или на одном предприятии заболеют одной и той же редкой формой лейкемии, может составлять один шанс из миллиона, *однако не следует забывать, что существуют миллионы учебных заведений, церковных приходов и предприятий.* Не так уж маловероятно, что пять человек

могут заболеть одной и той же редкой формой лейкемии в одном из этих мест. Мы просто забываем о всех школах, церковных приходах и предприятиях, где этого не случилось. Возьмем другую разновидность того же исходного примера — вероятность выигрыша в мгновенной лотерее; хотя она может составлять 1 шанс из 20 миллионов, никто из нас не удивляется тому, что кому-то удастся выиграть; действительно, что же здесь удивительного, если были проданы миллионы билетов! (Несмотря на мое недоверие к лотереям в целом, меня восхищает лозунг иллинойской мгновенной лотереи: «Кто-то должен выиграть; возможно, этим человеком окажетесь вы!». И впрямь, почему бы и нет?)

Ниже описан эксперимент, который я провожу со своими студентами, чтобы подтвердить этот базовый постулат. Чем больше аудитория, тем лучше. Я предлагаю каждому из присутствующих вынуть монетку и встать. Затем все подбрасывают монетку, и те, у кого выпадает решка, садятся. Допустим, в аудитории находится 100 студентов; примерно 50 из них займут свое место после первого подбрасывания. Потом мы выполняем это упражнение еще раз, в результате чего останутся стоять примерно 25 студентов. И так далее. Чаще всего после пяти или шести подбрасываний остается всего один человек, у которого пять или шесть раз подряд выпал орел. Я спрашиваю этого уника: «Как вам это удалось?», или «Вам, наверное, известна какая-то особая методика тренировок, позволяющая достигать определенного результата?», или «Вы, возможно, придерживаетесь какой-то особой диеты, помогающей добиться такого исхода?». Все присутствующие, конечно, воспринимают это как шутку, поскольку наблюдали процесс подбрасывания монетки собственными глазами, к тому же неплохо знают друг друга и понимают, что у человека, которому удалось пять раз подряд поймать монетку орлом вверх, нет никаких особых талантов в этом занятии, а результат, которого он добился, не более чем случайное совпадение. Однако каждый раз, когда мы видим какое-либо аномальное событие вне конкретного контекста, в котором оно произошло, у нас поневоле возникает подозрение, что здесь, помимо чистой случайности, замешано что-то еще.

Ошибка прокурора. Допустим, в суде вы услышали показания, которые сводятся к следующему: 1) образец ДНК, найденный на месте преступления, совпадает с результатами анализа ДНК обвиняемого и 2) существует лишь один шанс из миллиона, что образец ДНК, найденный на месте преступления, совпадет с образцом ДНК, взятым у кого-либо другого (не у обвиняемого).

(Ради простоты будем полагать, что вероятности, на которые опирается обвинение, соответствуют действительности.) Готовы ли вы вынести вердикт «виновен» на основе таких доказательств?

Надеюсь, вы не станете торопиться.

Ошибки обвинения случаются, когда контекст статистических доказательств игнорируется. Ниже описаны два сценария, каждый из которых может объяснить доказательства виновности обвиняемого, базирующиеся на результатах анализа ДНК.

Обвиняемый 1. Этот обвиняемый — влюбленный, отвергнутый своей жертвой, — был схвачен полицией за три квартала от места преступления; при нем было найдено оружие убийства. После ареста у него был взят образец ДНК, который совпал с образцом ДНК, взятым с волоска, найденного на месте преступления.

Обвиняемый 2. Этот обвиняемый был осужден несколько лет назад за аналогичное преступление, совершенное в другом штате. Когда суд признал его виновным, у него взяли образец ДНК, который был включен в общенациональную базу данных ДНК (в ней хранятся образцы ДНК более миллиона опасных уголовных преступников). Образец ДНК, взятый с волоска, найденного на месте преступления, сравнили с образцами, хранящимися в базе данных, и обнаружили совпадение с ДНК обвиняемого 2. Однако следствию не удалось обнаружить какую-либо связь последнего с жертвой преступления.

Как указывалось выше, в обоих случаях прокурор может с полным основанием заявить, что образец ДНК, взятый с места преступления, совпадает с образцом ДНК обвиняемого, и подчеркнуть, что существует лишь один шанс из миллиона, что он может совпасть с образцом ДНК какого-либо другого человека. Однако когда речь идет об обвиняемом 2, вероятность того, что он может оказаться тем самым случайным «другим человеком», одним из миллиона, образец ДНК которого по чистой случайности похож на ДНК подлинного убийцы, весьма высока. *Поскольку шансы найти случайно совпадающий образец ДНК среди миллиона других образцов относительно высоки, если вы ищете его в базе данных, насчитывающей более миллиона образцов.*

Возврат к среднему. Возможно, вы слышали о так называемом проклятии Sports Illustrated, в результате которого спортсмены или команды, фотографии которых помещались на обложке журнала Sports Illustrated, впоследствии снижали свои спортивные достижения. Одно из объяснений этого феномена заключалось в том, что размещение фотографии спортсмена

на обложке издания неблагоприятно сказывается на его последующих спортивных показателях. Более правдоподобным, с точки зрения статистики, будет объяснение, что команды и спортсмены обычно появляются на обложке Sports Illustrated после того, как добьются выдающихся успехов (например станут олимпийскими чемпионами), поэтому вполне естественно, что, пройдя пик физической формы, они демонстрируют результаты, близкие к средним. Это явление называется возвратом к среднему. Теория вероятностей говорит о том, что любой «отщепенец» — наблюдение, существенно отклоняющееся от среднего значения в том или ином направлении, — зачастую сопровождается исходами, более близкими к долгосрочному среднему значению.

Тенденция возврата к среднему позволяет объяснить, почему Chicago Cubs* всегда платит огромные суммы за так называемых свободных агентов, которые впоследствии разочаровывают болельщиков вроде меня. Игроки могут выторговать у Chicago Cubs высокие зарплаты после одного-двух необычайно удачных для себя сезонов и, одевшись в форму Chicago Cubs, вовсе не обязательно начинают играть хуже (правда, я отнюдь не исключаю и такой вариант); скорее, Chicago Cubs платит за них огромные деньги по окончании какого-то особенно удачного для этих суперзвезд периода — года или двух, — после чего их спортивные результаты (уже в ходе выступлений за Chicago Cubs) возвращаются к неким средним показателям.

То же явление объясняет, почему когда некоторые учащиеся сдают какой-либо экзамен гораздо лучше, чем обычно, в ходе его повторной сдачи они демонстрируют худшие результаты, а у учащихся, которые сдают экзамен хуже обычного, при его повторной сдаче результаты оказываются лучше. Такая взаимосвязь наталкивает на мысль, что достижения — как интеллектуальные, так и физические — представляют собой сочетание труда (связанного со способностями данного человека) и некоторого элемента везения (или невезения). В любом случае можно допустить, что тем, кто длительное время демонстрировал высокие результаты, сопутствовала удача; а тем, у кого показатели были гораздо ниже среднего, наверное, в какой-то мере не везло. (Что касается экзаменов, то ученики иногда пытаются *угадать* правильный ответ — а здесь уже все полностью зависит от везения; когда речь идет о футболе, мяч, посланный нападающим в сторону ворот противника, может оказаться в воротах только потому, что по пути заденет ногу кого-либо

* Chicago Cubs — профессиональный бейсбольный клуб, выступающий в Центральном дивизионе Национальной бейсбольной лиги. *Прим. перев.*

из игроков команды противника.) Когда период сильного везения или невезения заканчивается — а рано или поздно это неизбежно происходит, — достигнутые результаты становятся ближе к среднему.

Представьте, что я пытаюсь сформировать команду подбрасывателей монет, основываясь на ошибочном предположении, что способности в этом деле играют большую роль. После того как я увидел студента, у которого шесть раз подряд выпал орел, я предлагаю ему десятилетний контракт на 50 миллионов долларов. Разумеется, я испытаю огромное разочарование, когда окажется, что на протяжении этих десяти лет выпадение орла придется лишь на 50% подбрасываний монетки.

На первый взгляд возврат к среднему вступает в противоречие с «заблуждением игрока». После того как у моего студента шесть раз подряд выпал орел, можно ли утверждать, что на седьмой раз он «обязан» выбросить решку? Вероятность того, что на седьмой раз выпадет решка, такая же, как и всегда, — $\frac{1}{2}$. То обстоятельство, что у студента несколько раз подряд выпал орел, вовсе не повышает шансы на выпадение решки. Каждое подбрасывание монетки является независимым событием. Однако мы вправе рассчитывать на то, что результаты последующих подбрасываний будут соответствовать не прошлой картине, а тому, что предсказывает нам теория вероятностей (то есть примерно одинаковые шансы на выпадение орлов и решек). Вполне возможно, что тот, у кого несколько раз подряд выпал орел, в ходе последующих 10, 20 или 100 подбрасываний начнет раз за разом выбрасывать решку. И чем больше подбрасываний он выполнит, тем ближе окончательный их результат будет к соотношению 50 на 50, то есть к среднему результату, который предсказывает нам закон больших чисел. В противном случае у нас будут все основания искать доказательства мошенничества.

Кстати, исследователи задокументировали так называемый феномен Businessweek. Когда главам компаний вручают престижные награды (в том числе еженедельник Businessweek присваивает звание «Лучший менеджер»), как правило, в течение трех последующих лет эти компании ухудшают показатели (в частности, такие как учетная прибыль и цена акций). Однако, в отличие от упоминавшегося выше эффекта Sports Illustrated, «феномен Businessweek» представляет собой нечто большее, чем возврат к среднему. По словам Ульрики Малмендьер и Джеффри Тейта, экономистов Калифорнийского университета в Беркли и UCLA соответственно, когда главы компаний обретают статус «суперзвезды», внезапно свалившаяся на них слава

начинает отвлекать их от дел⁷. Они пишут мемуары. Их приглашают в советы директоров других компаний. Они ищут для себя так называемых статусных (то есть молодых и эффектных) жен. (Упомянутые мною авторы предлагают лишь первые два объяснения, однако последнее мне также кажется вполне правдоподобным.) Малмендьер и Тейт пишут: «Полученные нами результаты свидетельствуют о том, что культура суперзвезд, искусственно формируемая средствами массовой информации, ведет к более глубоким изменениям поведения, чем обычный возврат к среднему». Иными словами, когда фотография главы компании появляется на обложке Businessweek, пиши пропало, то есть быстро продавай акции этой компании.

Статистическая дискриминация (установление различия в статистическом смысле). В каких случаях следует опираться на то, что подсказывает нам теория вероятностей, а в каких так поступать не стоит? В 2003 году Анна Диамантопуло, еврокомиссар по проблемам занятости и социальным вопросам, выступила с предложением запретить страховым компаниям применять разные ставки к мужчинам и женщинам, поскольку это нарушает принцип равноправия, исповедуемый Евросоюзом⁸. Однако страховые компании вовсе не рассматривают такие надбавки как гендерную дискриминацию — для них это всего лишь статистика. Мужчины обычно платят больше за автостраховку, поскольку чаще, чем женщины, попадают в аварии. Женщины платят больше за аннуитеты (финансовый продукт, который выплачивает фиксированную ежемесячную или ежегодную сумму), потому что живут дольше, чем мужчины. Очевидно, что многие женщины попадают в аварии чаще, чем многие мужчины, а многие мужчины живут дольше, чем многие женщины. Но, как объяснялось в предыдущей главе, страховым компаниям нет до этого никакого дела. Их интересуют лишь среднестатистические показатели, поскольку именно это позволяет им получать прибыль. Что же касается политики Еврокомиссии, запрещающей ставить страховые надбавки в зависимость от пола человека (эта политика вступила в силу в 2012 году), то интересно отметить, что они вовсе не утверждают, будто страхуемые риски никоим образом не связаны с полом человека; они лишь заявляют о неприемлемости увязки с полом ставок страхования*.

* Указанное изменение политики Еврокомиссии было в конечном счете разъяснено в особом постановлении Верховного суда Евросоюза от 2011 года. В этом постановлении было указано, что применение разных надбавок к мужчинам и женщинам представляет собой дискриминацию по половому признаку.

Поначалу это кажется всего лишь раздражающей данью политкорректности. Но после некоторого размышления я не стал бы торопиться с таким выводом. Помните впечатляющую информацию о предотвращении преступлений? В этом отношении теория вероятностей может завести нас в интересные, но весьма «проблемные» места. Как нам следует реагировать, когда вероятностные модели говорят о том, что мексиканские наркоторговцы чаще всего оказываются испаноязычными мужчинами в возрасте от восемнадцати до тридцати лет, перевозящими товар в красных грузовиках-пикапах где-то между девятью и двенадцатью часами ночи, если нам также известно, что подавляющее большинство испаноязычных мужчин, соответствующих такому профилю, не промышляют контрабандой метамфетамина? Да, я использовал слово «профиль», поскольку оно представляет собой менее эффективное описание предсказательной аналитики, о которой я так красочно рассказывал в предыдущей главе, или по крайней мере один ее потенциальный аспект.

Теория вероятностей указывает нам, какие события более вероятны, а какие — менее. Да, речь идет лишь о базовой статистике — инструментах, описанию которых я посвятил несколько последних глав. Но это также статистика с социальными последствиями. Если мы хотим поймать уголовных преступников, террористов и наркоторговцев, а также других лиц, представляющих угрозу для общества, то обязаны использовать для этого все имеющиеся в нашем распоряжении инструменты. Теория вероятностей может быть одним из таких инструментов. Было бы наивно полагать, будто пол, возраст, расовая принадлежность, национальность, вероисповедание и страна происхождения человека в своей совокупности не играют никакой роли в том, что касается правоприменения.

Однако вопрос, что мы можем или должны делать с информацией такого рода, предполагая, что она представляет собой какую-то прогностическую ценность, является философско-правовым (но отнюдь не статистическим). Буквально каждый день мы получаем все больше и больше информации о все более широком круге явлений. Разве мы не будем одобрять дискриминацию, если соответствующие данные говорят нам о том, что мы будем правы гораздо чаще, чем неправы? (Именно отсюда происходит термин «статистическая дискриминация», или «рациональная дискриминация».) Точно такой же анализ, как использовался для того, чтобы выяснить, что люди, покупающие корм для птиц, менее склонны увильнуть от оплаты по кредитным карточкам

(представьте, это действительно так!), может применяться ко всем остальным аспектам нашей жизни. Какая часть всего этого может быть приемлема для нас? Если нам удастся разработать модель, позволяющую выявлять наркоторговцев в 80 случаях из 100, что случится с беднягами, которые попадут в оставшиеся 20%, — *ведь для этих ни в чем не повинных людей наша модель несет вполне реальную угрозу!*

Проблема состоит в том, что наша способность анализировать данные развилась значительно больше, чем понимание того, как нам следует поступать с результатами этого анализа. Вы можете соглашаться или нет с решением Еврокомиссии, запрещающим применение страховых надбавок, связанных с полом человека, но я абсолютно уверен, что это далеко не последнее спорное решение такого рода. Нам нравится думать о числах как о «холодных, неумолимых фактах». Если вычисления выполнены правильно, то у нас должен получиться *правильный ответ*. Однако более интересная и опасная реальность заключается в том, что подчас мы можем правильно все рассчитать — и двинуться в опасном направлении. Мы можем разрушить финансовую систему или упечь за решетку двадцатидвухлетнего белого парня, которому не повезло оказаться в определенное время в определенном месте, потому что, согласно нашей статистической модели, он явился сюда затем, чтобы купить наркотики. Какой бы соблазнительной ни была элегантность и точность вероятностных моделей, они не заменят нам здравого размышления о сути и цели выполняемых вычислений.

Почему так важны данные

«Мусор на входе — мусор на выходе»

7

Весной 2012 года в популярном журнале Science вышел сенсационный материал. В нем говорилось, что на основании результатов одного исследования, опирающегося на последние достижения науки, ученые сделали вывод, что когда самка дрозофилы (плодовой мушки) категорически отвергает ухаживания самца, он впадает в отчаяние и начинает топить горе в алкоголе. Газета The New York Times так описывала этот эксперимент в своей передовице: «Это были молодые самцы, ищущие любовных приключений. Будучи многократно отвергнутыми группой привлекательных самок, порхавших неподалеку, они поступили так же, как и многие мужчины в аналогичном случае, — запили с горя, используя алкоголь как болеутоляющее средство от неразделенной любви»¹.

Это исследование дает нам возможность лучше уяснить работу системы вознаграждения, встроенную в мозг человека, что, в свою очередь, должно помочь в поиске новых стратегий борьбы с алкоголизмом и наркоманией. Один специалист по вопросам наркозависимости написал, что знакомство с результатами этого эксперимента позволило ему «заглянуть в далекое прошлое и увидеть там истоки системы вознаграждения, которая определяет фундаментальные модели поведения, такие как секс, еда и сон».

Поскольку я не являюсь экспертом в данной области, после прочтения публикации о неразделенной любви самцов дрозофилы у меня возникли две несколько различающиеся между собой реакции. Во-первых, появилось чувство ностальгии по временам студенческой молодости. Во-вторых, мой «внутренний исследователь» заинтересовался, каким образом самцам дрозофилы удавалось запить с горя. Может быть, где-то поблизости находился миниатюрный бар для мушек-дрозофил с широким ассортиментом фруктовых алкогольных напитков, которые подавал

бармен-дрозофил, всегда готовый выслушать вас и посочувствовать вашему горю? Наверное, в баре ненавязчиво звучала музыка в стиле кантри, вызывавшая сильный прилив чувств у мушек-дрозофил, чьи жизненные мечты потерпели крушение?

В действительности все оказалось гораздо проще и прозаичнее. Одной группе самцов дрозофилы исследователи предоставили возможность беспрепятственно спариваться с самками, еще не подобравшими себе пару. Другой разрешили попытать счастья среди самок, которые уже подобрали себе пару и по этой причине были равнодушны к ухаживаниям других самцов. Затем обеим группам самцов дрозофилы были предложены на выбор питательные соломинки, содержащие две разные «диеты»: стандартное питание мушек-дрозофил плюс закваска и сахар, а также кое-что «покрепче»: закваска, сахар и 15-процентный раствор спирта. Самцы, которые провели несколько дней в бесплодных попытках найти себе пару среди безразличных к ним самок, оказались более склонны к «горячительному».

Какими бы легкомысленными ни казались эти выводы, они очень важны для человека, поскольку указывают на связь между стрессом, ответными химическими реакциями в мозгу человека и тягой к алкоголю. Тем не менее эти результаты — не триумф статистики. Это триумф данных, сделавших возможным этот относительно фундаментальный статистический анализ. Изюминка исследования заключалась в нахождении пути формирования двух групп — сексуально удовлетворенных и сексуально неудовлетворенных самцов дрозофилы — и последующем поиске способа сравнения их тяги к спиртному. После того как ученым удалось реализовать свой замысел, обработка соответствующих данных оказалась не сложнее, чем написание школьного реферата по математике.

Данные для статистики — примерно то же самое, что для выдающегося куортербека мощная линия блокирующих игроков. Сами по себе они не представляют особого интереса для зрителей, но без них выдающийся куортербек не сможет проявить свои способности. В большинстве книг по статистике предполагается, что вы используете надежные данные, точно так же как в любой кулинарной книге предполагается, что для приготовления блюд вы не станете покупать тухлое мясо или гнилые овощи. Даже самый замечательный рецепт не сделает вкусным блюдо, приготовленное из некачественных продуктов. То же касается статистики: даже самый изощренный анализ не принесет никакой пользы, если за основу взяты сомнительные

данные. Отсюда выражение: «Мусор на входе — мусор на выходе»*. Данные заслуживают уважительного отношения — как и линия блокирующих игроков в американском футболе.

Как правило, данные выполняют одну из трех функций. Во-первых, нам может потребоваться определенная выборка данных, соответствующая характеристикам генеральной совокупности (так называемая репрезентативная выборка). Если наша задача — оценить отношение избирателей к конкретному политическому деятелю, нам понадобится опросить некоторую их часть (выборку) в соответствующем избирательном округе или в целом в стране. (Обратите внимание: нас не интересует выборка, которая представляла бы каждого, кто *проживает* на искомой территории; нам требуется выборка лиц, *планирующих участвовать в голосовании*.) Одно из самых фундаментальных положений статистики, более глубокому разъяснению которого мы уделим место в двух следующих главах, заключается в том, что выводы, сделанные на основе достаточно больших, надлежащим образом сформированных выборок, могут оказаться такими же точными, как и в случае, если бы мы попытались получить ту же информацию от *всего* населения.

Самый легкий способ собрать репрезентативную выборку большой генеральной совокупности — выбрать ее некоторое подмножество случайным образом. (Вы, наверное, испытаете потрясение, когда узнаете, что это называется простой случайной выборкой.) Ключом к данной методологии является то, что каждое наблюдение в соответствующей совокупности должно иметь одинаковые шансы на включение в выборку. Если вы собираетесь опросить случайную выборку, состоящую из 100 взрослых, на территории, где проживает 4328 взрослых, то ваша методология должна гарантировать, что у каждого из этих 4328 человек одинаковые шансы оказаться в числе той сотни, которую вы намерены опросить. Книги по статистике почти всегда иллюстрируют это положение вытаскиванием разноцветных шариков из урны. (В действительности это практически единственное место, где слово «урна» используется более или менее регулярно.) Если в какой-нибудь гигантской урне находится 60 000 голубых и 40 000 красных шариков, то наиболее вероятным составом выборки из 100 шариков, случайным образом вынутых из урны, было бы 60 голубых и 40 красных шариков. Если бы мы проделали

* Известный принцип программирования, в соответствии с которым неверные входные данные не могут привести к правильному результату. *Прим. перев.*

такой эксперимент несколько раз, то, разумеется, всякий раз наблюдались бы какие-то отклонения (например, в одной выборке оказалось бы 62 голубых и 38 красных шариков, а в другой — 58 голубых и 42 красных шарика). Однако вероятность вытащить какую-либо произвольную выборку, которая по своему составу существенно отличалась бы от общего соотношения голубых и красных шариков, крайне мала.

Правда, на практике возникают кое-какие проблемы. Большинство совокупностей, которые могут представлять для нас интерес в реальной жизни, как правило, гораздо сложнее, чем урна с разноцветными шариками. Как, например, получить случайную выборку взрослого населения Америки, которую можно было бы использовать для проведения телефонного опроса? Даже такое элегантное на первый взгляд решение, как устройство для произвольного набора телефонных номеров, не лишено определенных недостатков. У некоторых людей (например, у малоимущих) может не быть телефона. Другие (обычно люди с высоким уровнем дохода) зачастую бывают не очень-то склонны отвечать на телефонные звонки, не представляющие для них непосредственного интереса. В главе 10 я опишу ряд стратегий, используемых компаниями, проводящими опросы, для решения проблем подобного рода, возникающих при формировании выборки (с появлением мобильной связи эти проблемы еще больше усложнились). Ключевая идея заключается в том, что надлежащим образом сформированная выборка будет полностью отражать структуру той совокупности, из которой она извлечена. Интуитивно вы можете представить себе это на примере варки супа в большой кастрюле. Если предварительно хорошенько размешать суп, то одна его ложка позволит вам составить достаточно полное представление о его вкусовых качествах.

В любом учебнике по статистике вы встретите значительно больше подробностей относительно методов формирования выборки. Компании, проводящие опросы и выполняющие маркетинговые исследования, тратят немало времени на поиск наиболее эффективных с экономической точки зрения способов получения надежных репрезентативных данных из разных совокупностей. На данном этапе вам необходимо уяснить несколько принципиальных положений. 1) Репрезентативная выборка — чрезвычайно важная вещь, поскольку она позволяет вам воспользоваться рядом наиболее мощных инструментов, которые имеются в распоряжении статистики. 2) Получить хорошую выборку гораздо сложнее, чем может показаться

на первый взгляд. 3) Многие из самых ошибочных статистических утверждений обусловлены применением совершенно правильных статистических методов к плохим выборкам, а вовсе не наоборот. 4) Размер выборки имеет значение — чем она больше, тем лучше. Подробнее об этом мы поговорим в следующих главах, но уже сейчас вам должно быть интуитивно понятно, что крупная выборка дает возможность нивелировать любые аномальные отклонения. (Кастрюля супа, несомненно, более точно отражает его истинный вкус, чем ложка супа.) Важное предостережение: наращивание размера выборки не позволяет компенсировать ошибки, допущенные при выборе ее структуры (так называемую систематическую ошибку). Единственный способ устранения ошибок, порождаемых плохой выборкой, — использование хорошей выборки. Никакой суперкомпьютер или мудреная формула не помогут вам обеспечить правильность результатов *общенационального* телефонного опроса в преддверии президентских выборов, если в соответствующую выборку будут включены исключительно жители Вашингтона, потому что они обычно голосуют не совсем так, как остальная Америка, и даже если вы опросите не 1000, а 100 000 жителей этого округа, это не устраним данную фундаментальную проблему вашего опроса. Более того, применение крупной выборки, в которую вкралась систематическая ошибка, несомненно хуже небольшой выборки с такой же систематической ошибкой, поскольку это создает ложное мнение о надежности полученного результата.

Второе, что нам зачастую требуется от данных, — это чтобы они служили нам источником сравнения. Новое лекарство эффективнее нынешнего? Можно ли надеяться, что бывшие осужденные, освоившие в тюрьме какую-либо профессию, будут менее склонны к повторному совершению преступлений, чем бывшие осужденные, которые такую профессию не приобрели? Можно ли надеяться, что успеваемость учащихся престижных учебных заведений окажется лучше, чем учеников обычных государственных школ?

В подобных случаях наша задача — найти две группы субъектов, в целом похожих между собой — за исключением интересующего нас «параметра». В контексте социальных наук таким «параметром» может быть что угодно, от наличия у самца дрозофилы сексуальной неудовлетворенности до права налогоплательщика на скидку при уплате подоходного налога. Как и в случае любого применения научного метода, мы пытаемся изолировать влияние какого-то *одного конкретного стороннего воздействия или фактора*. В этом и состояла гениальность эксперимента с дрозофилами. Исследователям

удалось найти способ создания контрольной (самцов, которые уже нашли себе пару) и «подопытной» (отвергнутых самцов) группы, а последующую разницу в отношении самцов к спиртному можно было объяснить их принадлежностью к той или иной группе.

В физических и биологических науках формирование контрольной и подопытной группы не представляет особой проблемы. Химики могут добыть небольших вариаций в нескольких пробирках, а затем изучить разницу в полученных результатах. Биологи могут использовать ту же методологию с помощью чашек Петри. Хотя должен заметить, что большинство экспериментов с животными проще провести, чем заставить дрозофил пить спиртное. Одну группу крыс можно заставить регулярно тренироваться на «беговой дорожке», а другую — нет, а затем сравнить их способность ориентироваться в лабиринте. Но когда речь идет о людях, все существенно усложняется. Чтобы правильно выполнить статистический анализ, зачастую требуется сформировать контрольную и подопытную группы, однако далеко не всегда люди согласны делать то же, что и крысы. (К тому же многим не нравится проделывать подобные эксперименты даже с лабораторными крысами.) Не вызовут ли периодически повторяющиеся сотрясения мозга у спортсменов серьезные неврологические проблемы в дальнейшем? Это действительно очень важный вопрос. От ответа на него зависит будущее хоккея на льду (и, возможно, других видов спорта). Однако сделать это невозможно путем проведения соответствующих экспериментов над людьми. Следовательно, пока (или если) мы не научим дрозофил носить шлемы и своевременно уворачиваться от силовых приемов, нам придется изыскивать другие способы изучения долговременных последствий травм головы.

Неизменной проблемой, которую приходится решать исследователям, работающим с «человеческим материалом», является создание контрольной и подопытной групп, отличающихся между собой *только* тем, что над одной группой проводится соответствующий опыт, а над другой — нет. Именно поэтому «золотым стандартом» исследования стала рандомизация, то есть процесс, посредством которого объекты изучения (люди, школы, больницы и т. д.) произвольным образом распределяются либо в подопытную, либо в контрольную группу. Мы не исходим из того, что все испытуемые идентичны. Напротив, вероятность — в который раз! — становится нашим союзником, и мы предполагаем, что рандомизация более-менее равномерно поделит все значимые характеристики между этими двумя группами — как те,

которые нас интересуют, так и те, которые не интересуют или не подлежат измерению, например настойчивость или честность.

Третью причину сбора данных можно сформулировать так, как иногда любит объяснять свои поступки моя малолетняя дочь: «Потому что!» Иногда у нас нет четкого представления о том, для чего нам может понадобиться та или иная информация, но интуитивно мы предполагаем, что в какой-то момент она обязательно пригодится. Это похоже на работу следователя на месте преступления: ему необходимо собрать максимальное количество улик, чтобы впоследствии составить предельно полную картину преступления. Одни из этих материальных доказательств окажутся полезными, другие следствию не помогут. Если бы мы заранее знали наверняка, что именно нам пригодится, то предварительное расследование нам, наверное, было бы не нужно.

Вам, должно быть, известно, что курение и ожирение являются факторами риска, способствующими развитию сердечно-сосудистых заболеваний. Но, возможно, вы не знаете, что эту взаимосвязь помогло выявить обследование жителей города Фрамингема, проводившееся в течение длительного времени. Во Фрамингеме проживает около 67 000 человек, город расположен примерно в двадцати милях от Бостона. Обычным людям он известен как пригород Бостона с относительно дешевым жильем и удобным доступом к торговому центру Natick Mall, славящемуся своими высококачественными (и дорогостоящими) товарами. Что же касается ученых, то Фрамингем для них ассоциируется с исследованием под названием Framingham Heart Study — одним из самых успешных в истории современной науки, оказавшим огромное влияние на развитие медицины.

В ходе повторного исследования выполняется сбор информации о большой группе субъектов в разные моменты времени (например каждые два года). Одни и те же участники исследования могут периодически опрашиваться на протяжении десяти, двадцати или даже пятидесяти лет. Такой подход позволяет получить необычайно богатый материал для анализа. В случае фрамингемского исследования в 1948 году ученые собрали информацию о 5209 взрослых жителях города: их рост, вес, кровяное давление, уровень образования, состав семьи, типичные продукты питания, склонность к курению, употребление наркотиков и т. п. Важно то, что начиная с этого времени эти люди периодически повторно обследовались, а также собирались данные об их потомстве, чтобы выявить генетические факторы, связанные

с развитием сердечно-сосудистых заболеваний. Начиная с 1950 года фрамингемские данные использовались при написании более чем двух тысяч научных статей, причем около тысячи из них были написаны в период с 2000 по 2009 год.

Эти исследования позволили получить чрезвычайно важные для понимания механизмов развития сердечно-сосудистых заболеваний результаты, многие из которых кажутся нам сейчас очевидными: курение сигарет увеличивает риск сердечно-сосудистых заболеваний (1960 год); физическая активность снижает риск сердечно-сосудистых заболеваний, а ожирение, наоборот, повышает (1967 год); высокое кровяное давление увеличивает риск инсульта (1970 год); высокий уровень холестерина альфа-липопротеинов высокой плотности (известного с тех пор как «полезный холестерин») снижает риск смертельного исхода (1988 год); у лиц, родители и близкие родственники (родные братья и сестры) которых страдали сердечно-сосудистыми заболеваниями, риск их развития значительно выше (2004–2005 годы).

Данные повторных исследований являются чем-то вроде научного эквивалента Ferrari. Они представляют особую ценность, когда речь идет о выявлении причинно-следственных связей, картина которых разворачивается на протяжении нескольких лет или даже десятилетий. Например, исследование дошкольников под названием Perry Preschool Study началось в конце 1960-х годов и охватило группу из 123 афроамериканцев — выходцев из бедных семей в возрасте от трех до четырех лет. Эти дети были случайным образом распределены в группу, которая проходила интенсивное обучение по программе дошкольного образования, тогда как участники контрольной группы, использовавшейся для сравнения, его не проходили. *В течение следующих сорока лет* ученые анализировали различные результаты этого эксперимента, которые стали убедительным доводом в пользу раннего обучения детей. Учащиеся, которые прошли интенсивное обучение по программе дошкольного образования, уже в пятилетнем возрасте демонстрировали более высокие показатели IQ и имели больше шансов успешно окончить среднюю школу. В сорокалетнем возрасте у них были более высокие доходы. И напротив, многие из участников эксперимента, не получившие дошкольного образования, к сорокалетнему возрасту успевали по пять и более раз побывать за решеткой.

Нет ничего удивительного в том, что далеко не каждый из нас может позволить себе покупку Ferrari. Исследовательским эквивалентом этой машины является так называемый «поперечный срез», то есть совокупность данных, собранных в какой-то определенный момент времени. Если, например, эпидемиологи пытаются выявить причину какого-либо нового заболевания (или вспышки старого), они могут собрать данные обо всех, кто им страдает, в надежде получить картину, которая приведет к его источнику. Может быть, причина в том, что они едят? Или эти люди побывали в какой-то местности и заболели в результате этого? Что еще общего между людьми, страдающими этой болезнью? Кроме того, исследователи могут собрать данные о здоровых людях, чтобы выявить разницу между двумя группами.

На самом деле все эти оживленные дискуссии вокруг данных «поперечного среза» напоминают мне неделю перед моей свадьбой, когда я стал частью некой совокупности данных. В то время я работал в Катманду (столице Непала) и занемог желудком в результате малопонятной болезни под названием «сине-зеленые водоросли» (это заболевание было выявлено лишь в двух местах в мире). Ученым удалось выделить патоген, который был причиной болезни, но они все еще не были уверены, *что* собой представляет этот организм, поскольку на то время он еще не был известен науке. Когда я позвонил домой, чтобы рассказать невесте о моем диагнозе, я был вынужден признать, что это не самая приятная новость. Неизвестно, каким путем эта болезнь передается от человека к человеку и как ее лечить, к тому же она могла вызывать сильное утомление и прочие малоприятные побочные эффекты в период от нескольких дней до многих месяцев*. Учитывая, что до свадьбы оставалась всего неделя, это могло оказаться серьезной проблемой. Следовало ли мне держать себя в руках, когда я бродил по супермаркету? Пожалуй, да.

Но затем я попытался сосредоточиться на позитивной информации. Во-первых, «сине-зеленые водоросли» не считались смертным приговором. И во-вторых, специалисты по тропическим болезням из Бангкока (далековато!) проявили особый интерес к моему случаю. *Тебя не знобит?* (К тому же мне приходилось постоянно возвращать дискуссию к вопросу, волновавшему меня в тот момент больше всего, — к планированию предстоящей свадьбы: «Хватит о моей неизлечимой болезни. Поговорим лучше о цветах».)

* На тот момент средняя продолжительность этой болезни составляла сорок три дня со среднеквадратическим отклонением, равным двадцати четырем дням.

Свои последние часы в Катманду я провел, заполняя тридцатистраничную анкету, которая касалась буквально каждого аспекта моей жизни. Где я обедал и ужинал? Что именно я ел? Готовлю ли я дома, а если готовлю, то как? Приходилось ли мне плавать? Где и как часто? Все, у кого была обнаружена эта болезнь, заполняли точно такую же анкету. В конце концов патоген был выявлен: им оказались водные цианобактерии (*Cyanobacteria*). (Эти синие бактерии представляют собой единственный вид бактерий, получающих энергию из фотосинтеза; отсюда первоначальное описание заболевания как «сине-зеленые водоросли».) Оказалось, что эту болезнь можно лечить традиционными антибиотиками (но как ни странно, некоторые из новейших антибиотиков не действовали). Однако все эти открытия уже не могли мне помочь, потому что и без них я поправился достаточно быстро. Ко дню свадьбы моя пищеварительная система практически полностью восстановилась.

За каждым важным экспериментом кроются полезные данные, без которых он был бы невозможен. А за каждым неудачным... Одним словом, вам и так все понятно. Люди зачастую говорят о «лжи с помощью статистики». Я готов утверждать, что некоторые из самых вопиющих статистических ошибок обусловлены *ложью с помощью данных*; статистический анализ выполнен правильно, но данные, на основе которых он делался, неправильны или неуместны. Ниже приведено несколько типичных примеров из категории «мусор на входе — мусор на выходе».

Систематическая ошибка выбора. Говорят, что Паулина Кейл, кинокритик и давний сотрудник еженедельника *The New Yorker*, после того как Ричард Никсон стал президентом США, сказала: «Никсон не мог победить. Я не знаю ни одного человека, который бы за него проголосовал». Очень сомневаюсь, что Паулина Кейл могла такое сказать, но однако это весьма показательный пример того, как ничтожная выборка (группа либерально настроенных приятелей некоего человека) способна создать ложное представление о гораздо большем числе людей (всех американских избирателях). Отсюда вопрос, который всегда нас должен интересовать: как была сформирована выборка (или выборки) для оценивания? Если каждому члену генеральной совокупности не предоставлены равные шансы на включение в выборку, у нас наверняка возникнут проблемы с результатами, полученными на ее основе. Одним из ритуалов, связанных с проведением президентских выборов в Соединенных Штатах, является неофициальный, выборочный опрос

общественного мнения (так называемый соломенный опрос) в штате Айова. За год до президентских выборов, в августе, кандидаты от Республиканской партии собираются в городке Эймис штата Айова, чтобы набрать участников опроса. Каждый из желающих в нем участвовать должен заплатить 30 долларов. «Соломенный опрос» в штате Айова ничего не скажет нам наверняка о политическом будущем кандидатов от Республиканской партии. (Этот опрос точно предсказал лишь трех из последних пяти республиканских «номинантов».) В чем тут причина? Дело в том, что жители штата Айова, заплатившие 30 долларов, отличаются от других сторонников Республиканской партии в этом штате, не говоря уже о том, что ее сторонники в штате Айова отличаются от сторонников Республиканской партии в целом по стране.

Систематическая ошибка выбора может возникнуть при различных обстоятельствах. Опрос потребителей в аэропорту искажается тем фактом, что любители летать самолетами, как правило, более состоятельные люди, чем население в целом; в случае проведения опроса на площадке для отдыха возле автомагистрали Interstate 90 может сложиться противоположная ситуация. На результаты обоих опросов наверняка повлияет и то, что люди, готовые в них участвовать, отличаются от людей, предпочитающих не отвлекаться на подобные вещи. Если вы попросите 100 человек в каком-либо общественном месте заполнить совсем небольшую анкету, то *те 60, которые согласятся это сделать, наверняка будут существенно отличаться от остальных 40, которые вас проигнорируют.*

Один из самых известных статистических просчетов — опрос, проведенный еженедельником Literary Digest в 1936 году, — был обусловлен неправильно сформированной выборкой. В том году губернатор штата Канзас республиканец Алф Лэндон сражался за президентский пост с действующим президентом США Франклином Рузвельтом (демократом). Еженедельник Literary Digest, в то время весьма влиятельное издание, провел по почте опрос среди своих подписчиков, а также среди владельцев автомобилей и домашних телефонов, адреса которых редакции Literary Digest удалось получить из открытых источников. Опрос Literary Digest охватил 10 миллионов потенциальных избирателей, что представляло собой выборку поистине астрономического масштаба. При увеличении размера правильно сформированной выборки точность опроса повышается, поскольку сужается допустимый предел погрешности. Когда же увеличивается размер неправильно сформированной выборки, высота мусорной кучи также увеличивается,

а вонь от нее становится сильнее. Согласно прогнозу Literary Digest, победу на президентских выборах должен был одержать Алф Лэндон, получив 57% голосов избирателей. На самом же деле выиграл Франклин Рузвельт, получив 60% голосов избирателей, причем его победа была зафиксирована в сорока шести из сорока восьми штатов. Выборка, сформированная Literary Digest, оказалась пресловутым «мусором на входе»: подписчики еженедельника были более состоятельными людьми, чем средний американец, и, следовательно, были в большей степени склонны голосовать за республиканцев; то же самое можно сказать и о владельцах автомобилей и домашних телефонов (напомню, что выборы проводились в 1936 году)².

Мы можем столкнуться с такой же фундаментальной проблемой при сравнении исходов в подопытной и контрольной группах, если механизм включения в них участников не обеспечивает случайного выбора. Рассмотрим недавние выводы относительно побочных эффектов лечения рака простаты, опубликованные в медицинской литературе. Существует три широко распространенных метода лечения рака простаты: хирургическое удаление простаты, лучевая терапия и брахитерапия (которая предусматривает имплантацию радиоактивных «семян» вблизи раковой опухоли)³. Типичный побочный эффект лечения рака простаты — импотенция, поэтому исследователи документировали половую функцию мужчин, к которым применялся какой-либо из трех указанных методов лечения. Обследование 1000 мужчин показало, что через два года после лечения половые акты могли совершать 35% мужчин в «хирургической» группе, 37% — в «лучевой» и 43% — в «брахитерапийной».

Можно ли, глядя на эти результаты, утверждать, что брахитерапия в наименьшей степени сказывается на половой функции мужчин? Отнюдь! Авторы данного исследования специально подчеркивали, что для этого у них нет никаких оснований, поскольку мужчины, к которым применяется этот способ лечения, как правило, моложе и в целом физически крепче тех, кого лечили другими методами. Цель данного исследования заключалась лишь в том, чтобы задокументировать степень половых побочных эффектов для всех типов лечения.

Систематическая ошибка такого же рода, известная как систематическая ошибка самоотбора, возникает, когда люди сами напрашиваются в подопытную группу. Например, заключенные, которые добровольно соглашались на опробование какого-либо нового способа лечения от наркозависимости,

отличаются от других заключенных именно потому, что сами попросили об этом. Если окажется, что после освобождения из тюрьмы они будут попадать за решетку гораздо реже, чем остальные заключенные, то это, конечно, замечательно, но абсолютно ничего не говорит нам о ценности нового способа лечения от наркозависимости. Возможно, эти бывшие заключенные изменили свою жизнь именно потому, что в этом им помог новый способ лечения от наркозависимости. Не исключено также, что это произошло под воздействием каких-то других факторов, которые также повлияли на их желание стать участниками эксперимента (одним из таких факторов мог быть страх перед перспективой вновь оказаться за решеткой). Мы не можем отделить причинно-следственное влияние одного (экспериментальная методика лечения наркозависимости) от другого (человек, изъявивший желание участвовать в опробовании нового способа лечения от наркозависимости).

Систематическая ошибка публикации. Позитивные результаты обнаруживают охотнее, чем негативные. Допустим, вы только что завершили строго научное повторное исследование, которое показало, что увлечение видеоиграми не препятствует развитию рака толстой кишки. На протяжении двадцати лет вы обследовали репрезентативную выборку из 100 000 американцев и выяснили, что среди фанатов видеоигр и тех, кто совершенно ими не интересуется, заболеваемость раком толстой кишки находится примерно на одном уровне. Предположим, ваша методология безупречна. Какой из престижных медицинских журналов опубликует результаты данного исследования?

Думаю, никакой. И тому есть две причины. Во-первых, нет ни одной научно обоснованной причины полагать, что увлечение видеоиграми способствует развитию рака толстой кишки, а потому не совсем понятно, зачем вы проводили свое исследование. Во-вторых, тот факт, что некий фактор не препятствует заболеванию раком, не представляет научной ценности. В конце концов, многие вещи этому не препятствуют. Негативные результаты не производят особого впечатления — ни в медицине, ни где бы то ни было.

А теперь допустим, что один из ваших приятелей, вместе с которым вы оканчивали магистратуру, провел другое повторное исследование и выяснил, что среди тех, кто много играет в видеоигры, заболевание раком толстой кишки встречается реже. *А вот это уже гораздо интереснее!* Результаты именно такого рода привлекают повышенное внимание медицинских журналов, популярной прессы, блогеров и разработчиков компьютерных игр (которые не преминут их снабдить надписью о пользе игр для здоровья).

Пройдет совсем немного времени, и заботливые мамы по всей стране бросятся спасать своих детей от рака, выхватывая у них книги из рук и заставляя играть в видеоигры.

Разумеется, согласно важному положению в статистике, необычные явления происходят довольно редко и, как правило, в результате случайного стечения обстоятельств. В одном из 100 аналогичных исследований наверняка обнаружатся нелепые результаты типа взаимозависимости между увлечением видеоиграми и меньшей заболеваемостью раком толстой кишки. Проблема в том, что результаты 99 исследований, которые не выявили такую связь, опубликованы не будут, поскольку малоинтересны. А вот единственное исследование, которое ее обнаружит, попадет в печать и привлечет к себе повышенное внимание. Источником данной систематической ошибки является не исследование как таковое, а сомнительная информация, которая фактически становится достоянием широкого круга читателей. Тот, кого интересует литература о видеоиграх и заболеваемости раком толстой кишки, найдет публикацию лишь о единственном исследовании, и в ней будет утверждаться, что увлечение видеоиграми предотвращает заболевание раком. Хотя на самом деле результаты 99 исследований из 100 свидетельствуют об отсутствии какой-либо связи.

Да, мой пример несколько абсурден, однако данная проблема вполне реальна и довольно серьезна. Вот первое предложение из статьи в *The New York Times*, в которой говорится о систематической ошибке публикации, касающейся лекарств от депрессии: «Производители антидепрессантов, таких как Prozac и Paxil, никогда не публиковали результаты примерно трети испытаний своих лекарств, проводившихся, чтобы получить одобрение государственных контролирующих органов. Таким образом производители антидепрессантов вводили в заблуждение врачей и пациентов относительно подлинной эффективности этих препаратов»⁴. Оказывается, были обнародованы данные о 94% исследований с положительными результатами, касающимися эффективности этих лекарств, и лишь о 14% исследований с отрицательными результатами. Для пациентов, страдающих депрессией, это крайне важно. Если бы были оглашены результаты всех исследований, то оказалось бы, что в действительности антидепрессанты лишь немногим лучше любого плацебо.

Чтобы справиться с данной проблемой, теперь медицинские журналы, как правило, требуют зарегистрировать любое исследование в самом начале

проекта, если предполагается последующая публикация его результатов. Это предоставляет редакторам определенные свидетельства о соотношении позитивных и негативных исходов. Если, например, зарегистрировано 100 исследований по анализу влияния катания на роликовой доске (скейтборде) на развитие сердечно-сосудистых заболеваний, и лишь одно из них будет в конечном счете представлено для публикации с положительными результатами, то редакторы могут заключить, что в ходе других исследований получены отрицательные результаты (или по крайней мере проверить такую вероятность).

Систематическая ошибка памяти. Наша память — восхитительный дар, правда, далеко не всегда источник достоверных данных. Человеку свойственно воспринимать настоящее как логическое следствие прошедших событий. Иными словами, человек интуитивно пытается находить причинно-следственные связи. Проблема в том, что наша память оказывается «систематически хрупкой», когда мы пытаемся объяснить какой-либо особенно хороший или плохой результат в настоящем. Рассмотрим исследование, авторы которого пытаются выявить связь между рационом питания и раковыми заболеваниями. В 1993 году один из ученых Гарвардского университета собрал данные о группе женщин, страдающих раком груди, и информацию о группе женщин примерно того же возраста, у которых рак не был диагностирован. Женщин в обеих группах спросили об особенностях их рациона питания в молодые годы. Этот опрос выявил совершенно четкие результаты: женщины, страдающие раком груди, в молодости употребляли пищу с более высоким содержанием жиров.

Да, но это исследование вовсе не ставило перед собой задачу определить влияние рациона питания на вероятность развития раковых заболеваний! Оно было призвано определить, как заболевание раком влияет на память женщины о ее питании в прежние годы. Все женщины-участницы заполнили анкету, где спрашивалось об особенностях их питания, за много лет до того, как у кого-то из них был диагностирован рак. Самое интересное, что женщины, страдающие раком груди, вспомнили, что употребляли гораздо больше жиров, чем это было на самом деле, тогда как в воспоминаниях здоровых женщин существенных отклонений от реального рациона их питания в молодости не обнаружилось. Журнал The New York Times Magazine так описал скрытую природу этой систематической ошибки памяти:

Диагностирование рака груди не изменило настоящее и будущее женщины; оно изменило ее прошлое. Женщины, страдающие раком груди, решили (подсознательно), что пища с повышенным содержанием жиров, вероятно, обусловила их предрасположенность к раковым заболеваниям, и «вспомнили», что их рацион в молодости отличался повышенным содержанием жиров, хотя на самом деле это было не так. Подобный образ мышления знаком каждому, кто знает историю этого пользующегося дурной славой заболевания: эти женщины, как и тысячи женщин до них, искали в своих воспоминаниях причину заболевания и затем внедряли ее в память⁵.

Наличие такой систематической ошибки памяти — одна из причин, почему ученые чаще предпочитают проводить повторные исследования, а не исследования типа «поперечный срез». В случае повторного исследования сбор данных выполняется на протяжении всего времени его проведения. В пятилетнем возрасте участника спрашивают о его отношении к школе. Затем, спустя тринадцать лет, мы можем навестись к нему и выяснить, не бросил ли он школу досрочно. При проведении исследования «поперечный срез» все данные собираются одномоментно, и, спрашивая восемнадцатилетнего парня, бросившего школу, как он к ней относился в пятилетнем возрасте, мы вряд ли получим правдивый ответ.

Систематическая ошибка доживаемости до определенного возраста. Допустим, директор школы сообщает, что на протяжении четырех лет результаты экзаменов определенной группы учащихся неизменно улучшаются. Оценки второклассников лучше оценок первоклассников, а баллы третьеклассников еще лучше, чем у второклассников, ну и самые выдающиеся результаты демонстрируют ученики четвертого класса. Предполагается, что в данном случае отсутствуют какие-либо манипуляции с сознательным завышением оценок, а также с «творческим» применением описательных статистик. Каждый год эта группа становится все более успешной, какими бы показателями мы ни пользовались: средним, медианой, процентом учащихся, перешедших в следующий класс, и т. д.

Как бы вы поступили в подобном случае: а) присвоили директору этой школы звание «директор года» или б) потребовали бы от него дополнительных данных?

Лично я предпочел бы вариант б). У меня возникло сильное подозрение о наличии в данном случае систематической ошибки доживаемости до определенного возраста, которая возникает, когда какие-то из наблюдений выпадают из выборки, изменяя состав оставшихся наблюдений и тем самым сказываясь на результатах того или иного анализа. Допустим, что директор школы — никудышний администратор. Учебный процесс во вверенном ему заведении организован из рук вон плохо, учащиеся не приобретают никаких знаний, каждый год половина из них бросает учебу. Разумеется, это позитивно скажется на общих результатах экзаменов — притом что оценки каждого отдельно взятого учащегося не станут лучше. Если сделать вполне разумное допущение, что школу бросят самые нерадивые ученики (которые получали на экзаменах самые низкие оценки), то средний результат сдачи экзаменов оставшимися учащимися будет неуклонно повышаться по мере увеличения числа учеников, бросающих учебу. (Если собрать в одной комнате людей разного роста, а затем попросить «коротышек» выйти из комнаты, то средний рост оставшихся увеличится, хотя каждый из них в отдельности не стал выше.)

Индустрия взаимных фондов охотно ухватилась за систематическую ошибку доживаемости до определенного возраста, воспользовавшись ею для того, чтобы их прибыльность выглядела для инвесторов гораздо привлекательнее, чем на самом деле. Взаимные фонды обычно оценивают свою эффективность, сравнивая свои показатели с прибыльностью, обеспечиваемой одним из основных индексов, например Standard & Poor's 500 (индекс 500 ведущих акционерных компаний открытого типа в Америке)*. Если в прошлом году S&P 500 повысился на 5,3%, то считается, что некий взаимный фонд превзошел этот индекс, если его прибыльность оказалась выше, и наоборот, взаимный фонд завершил год хуже, если его прибыльность ниже этого индекса. Для инвесторов, которые не желают платить менеджеру взаимного фонда, одним из довольно дешевых и простых вариантов будет покупка акций S&P 500 Index Fund, который представляет собой взаимный фонд, просто приобретающий доли во всех 500 пакетах акций, представленных

* Standard & Poor's 500 — показательный пример того, что может и должен делать любой индекс. Этот индекс составлен из цен акций 500 ведущих американских компаний с учетом рыночной стоимости каждой из этих компаний (так, чтобы более крупные компании имели в этом индексе больший вес, чем мелкие). Данный индекс — простой и точный показатель того, что происходит с ценами акций крупнейших американских компаний в любой момент времени.

в S&P 500. Менеджеры взаимных фондов убеждают нас в своей дальновидности и умении использовать знания для выбора таких ценных бумаг, которые обеспечивают более высокую прибыльность, чем какой-нибудь простой индексный фонд. В действительности превзойти S&P 500 на достаточно продолжительном отрезке времени довольно трудно. (По сути, S&P 500 представляет собой среднее пакетов акций всех крупных компаний, которые торгуются на фондовой бирже, поэтому с математической точки зрения можно ожидать, что примерно половина активно управляемых взаимных фондов в данном году превзойдет S&P 500, а другая половина, наоборот, продемонстрирует более слабый результат, чем S&P 500.) Разумеется, недосчитаться прибыли, доверив свои деньги какому-нибудь бездумному индексу, который просто покупает 500 пакетов акций и держит их у себя, было бы не очень красиво. Никакого тебе анализа. Никакого мудреного макропрогнозирования. И — к немалому удовольствию инвесторов — никаких заоблачных выплат в пользу менеджеров взаимных фондов.

Чем занимается традиционная компания типа взаимного фонда? Манипулирует данными! Вот как они могут «превзойти рынок», в действительности не делая этого. Крупная компания типа взаимного фонда открывает много новых активно управляемых фондов (это означает, что эксперты подбирают ценные бумаги, зачастую руководствуясь определенной стратегией). Допустим, к примеру, что она открывает двадцать новых фондов, каждый из которых с 50-процентной вероятностью может в данном году превзойти S&P 500. (Это предположение вполне соответствует долгосрочным данным.) Итак, согласно теории вероятностей, в первый год лишь десять новых фондов компании превзойдут S&P 500; пять фондов превзойдут S&P 500 в течение двух лет подряд; а два или три фонда — в течение трех лет подряд.

Дальше наступает черед самой большой хитрости. В этот момент новые взаимные фонды, которые продемонстрировали не особо впечатляющие результаты по сравнению с S&P 500, потихому прикрываются (их активы включаются в другие существующие фонды). Затем компания может запустить массированную рекламу двух или трех новых фондов, которые «год за годом превосходят S&P 500», — даже если результат, достигнутый ими, такая же случайность, как выпадание решки три раза подряд. Дальнейшие показатели эффективности этих фондов наверняка приблизятся к среднему значению — правда, по пути они привлекут к себе толпы новых инвесторов.

На самом деле количество взаимных фондов или инвестиционных гуру, которые на протяжении достаточно продолжительного времени превосходят S&P 500, удручающе мало*.

Систематическая ошибка здорового человека. Те, кто заботится о наличии в своем рационе достаточного количества витаминов, как правило, отличаются крепким здоровьем — *поскольку это люди, потребляющие достаточное количество витаминов!* Играть ли какую-то роль в этом витамины — другой вопрос. Рассмотрим следующий мысленный эксперимент. Допустим, чиновники Министерства здравоохранения пропагандируют теорию, согласно которой всем молодым родителям следует укладывать своих детей в постель лишь в лиловых пижамах, поскольку это стимулирует умственное развитие ребенка. Спустя двадцать лет повторное исследование подтверждает, что те, кто в детстве спал в лиловых пижамах, достигли заметных успехов во взрослой жизни. Например, оказалось, что 98% студентов-первокурсников Гарвардского университета в детстве спали в лиловых пижамах (а многие и по сей день продолжают это делать), тогда как лишь 3% из тех, кто в детстве спал в пижамах лилового цвета, сидят в тюрьмах штата Массачусетс.

Разумеется, лиловые пижамы здесь абсолютно ни при чем, однако наличие родителей, которые заставляют своих детей спать в таких пижамах, *еще как «при чем»*. Даже когда мы пытаемся контролировать уровень образования родителей, нам все равно приходится иметь дело с не поддающимися наблюдению различиями между теми родителями, которые придают огромное значение цвету пижамы своего ребенка, и теми, кому это совершенно безразлично. Гэри Тобис, обозреватель The New York Times, специализирующийся на вопросах здоровья, объясняет: «Простому говоря, проблема в том, что те, кто с огромным энтузиазмом выполняет все рекомендации, которые кажутся им чрезвычайно полезными (неукоснительно принимают лекарства, прописанные врачом, или соблюдают диеты), принципиально отличаются от тех, кто к таким советам не считает нужным прислушиваться»⁶. Данный эффект способен обесценить любое исследование, пытающееся определить

* С очень интересным обсуждением того, почему следует отдать предпочтение покупке индексных фондов, вместо того чтобы пытаться превзойти рынок, можно ознакомиться в книге моего бывшего преподавателя, профессора Бертон Малкиела (Burton Malkiel) *A Random Walk Down Wall Street* (Случайная прогулка по Уолл-стрит. — Минск : Попурри, 2006).

реальную пользу действий, якобы благотворно влияющих на здоровье человека (например, регулярные занятия спортом или употребление в пищу листовой капусты). Мы полагаем, что сравниваем влияние на здоровье двух диет: с капустой и без нее. В действительности, если подопытная и контрольная группы сформированы случайным образом, мы сравниваем две диеты, которых придерживаются две разные категории людей. У нас есть подопытная группа, и она отличается от контрольной группы в *двух* аспектах, а не в одном.

Если статистика напоминает работу следователя, то данные являются аналогом вещественных улик. Моя жена год работала преподавателем в старших классах сельской школы штата Нью-Гэмпшир. Одного из ее учеников арестовали за ограбление магазина хозяйственных товаров. Полиции удалось быстро раскрыть это преступление, потому что 1) накануне кражи выпал снег и следы от магазина вели к дому, где проживал грабитель; и 2) в доме были найдены похищенные товары. Таким образом, надежные вещественные доказательства действительно помогли.

Цените надежные данные. Но для начала вам понадобится их добыть, а это гораздо труднее, чем может показаться на первый взгляд.

Центральная предельная теорема

Леброн Джеймс статистики

8

Порой статистика подобна магии. Она позволяет делать далекоидущие важные выводы на основе относительно небольшого объема данных. Каким-то образом нам удастся предсказать исход президентских выборов, опросив лишь тысячу избирателей. Или, проверив на птицефабрике сотню куриных тушек на наличие сальмонеллы, оценить, исходя из этой информации, общее санитарное состояние предприятия.

Что же является источником столь необычайной силы обобщения? Это центральная предельная теорема, значение которой для статистики соизмеримо со значением Леброна Джеймса* для профессионального баскетбола. Центральная предельная теорема — «источник энергии» для многих статистических действий, предполагающих использование той или иной выборки для получения выводов относительно некой более крупной совокупности данных (например, опрос населения или тест на наличие сальмонеллы). Хотя порой такого рода выводы могут казаться мистическими, фактически это просто сочетание двух инструментов, уже рассмотренных нами в этой книге: теории вероятностей и правильного формирования выборки. Прежде чем приступить к подробному рассмотрению механизма (на самом деле не такого уж сложного) центральной предельной теоремы, ознакомьтесь с примером, который поможет вам на интуитивном уровне понять, о чем пойдет речь.

Допустим, вы живете в городе, где будет проходить марафон. В нем примут участие бегуны со всего мира, а значит, многие из них не говорят по-английски. Чтобы своевременно и с максимальным комфортом доставить спортсменов к месту старта, всем участникам необходимо

* Леброн Рэймон Джеймс (LeBron Raymone James) — американский профессиональный баскетболист, играющий на позиции легкого и тяжелого форварда за команду НБА «Кливленд Кавальерс». *Прим. перев.*

зарегистрироваться утром в день забега, после чего их произвольным образом рассадят по автобусам и отвезут на старт. К сожалению, один из автобусов затерялся где-то в пути. (Ладно, вам придется предположить, что ни у одного из его пассажиров не было мобильного телефона, а у водителя не оказалось GPS-навигатора; если не хотите заниматься утомительными математическими выкладками, всегда держите мобильный телефон при себе.) Будучи одним из общественных активистов города, вы подключаетесь к поискам пропавшего автобуса.

Вам повезло: вы натываетесь на какой-то сломавшийся автобус неподалеку от своего дома; возле автобуса коротает время группа расстроенных пассажиров, причем ни один из них не говорит по-английски. Наверное, это и есть тот автобус, который вы разыскиваете! У вас появляется шанс стать героем дня. Правда, вас смущает одно обстоятельство: пассажиры автобуса — слишком упитанные люди. Окинув эту группу взглядом, вы заключаете, что средний вес ее пассажиров превышает 220 фунтов. Трудно представить, что в случайно сформированной группе бегунов-марафонцев могут оказаться столь колоритные экземпляры. Вы звоните по мобильному телефону в штаб-квартиру поисковой команды и сообщаете: «Мне кажется, это не тот автобус, который мы ищем. Продолжайте поиск».

Дальнейший анализ подтверждает ваше первоначальное предположение. Когда на место прибывает переводчик, оказывается, что сломавшийся автобус направлялся на Международный фестиваль любителей сосисок, который также проводится в вашем городе, причем в тот же день, что и марафонский забег. (Для большего правдоподобия замечу, что участники фестивалей любителей сосисок нередко ходят в спортивных брюках свободного покроя, которые не стесняют их движений.)

Примите мои поздравления! Если вам понятно, каким образом человек, просто окинув беглым взглядом группу пассажиров автобуса и оценив их вес, может прийти к выводу, что конечным пунктом назначения этого автобуса вряд ли может быть место старта марафонского забега, значит, на интуитивном уровне вы уже постигли базовую идею центральной предельной теоремы. И все, что вам остается, это уяснить некоторые детали. А если вы понимаете центральную предельную теорему, то и большинство форм статистических выводов наверняка покажутся вам интуитивно понятными.

Базовый принцип, лежащий в основе центральной предельной теоремы, заключается в том, что большая, надлежащим образом сформированная

выборка будет похожа на совокупность, из которой она извлечена. Разумеется, от выборки к выборке будут наблюдаться определенные вариации (например, группа пассажиров в каждом автобусе, направляющемся к месту старта марафонского забега, будет несколько отличаться от группы пассажиров в других автобусах), однако вероятность того, что какая-либо выборка будет *существенно* разниться с генеральной совокупностью, крайне низка. Именно эта логика позволила вам прийти к указанному выше интуитивному умозаключению, когда вы подошли к сломавшемуся автобусу и беглым взглядом оценили средний вес его пассажиров. Да, марафонскую дистанцию нередко бегут люди довольно плотного телосложения; среди участников каждого крупного марафона немало спортсменов, вес которых превышает 200 фунтов. Однако большинство бегунов-марафонцев — худощавые люди. Таким образом, вероятность того, что столь значительное число упитанных бегунов по случайному стечению обстоятельств окажется в одном автобусе, чрезвычайно мала. Вы могли бы вполне уверенно заключить, что встретившийся вам автобус перевозит не марафонцев. Конечно, не исключено, что вы ошибаетесь, однако, согласно теории вероятностей, шансы на ошибку в данном случае очень и очень невелики.

В этом и состоит интуитивная основа центральной предельной теоремы. Воспользовавшись кое-какими статистическими «прибамбасами», можно вычислить вероятность того, окажетесь ли вы правы или неправы. Например, мы можем подсчитать, что в случае, когда речь идет о 10 000 участниках марафонского забега, средний вес которых равняется 155 фунтов, вероятность того, что средний вес случайной выборки из 60 таких бегунов (примерная вместимость одного автобуса) окажется не ниже 220 фунтов, составляет менее одного шанса из 100. Давайте на данном этапе доверимся интуиции; впоследствии у нас будет немало возможностей выполнить соответствующие вычисления.

Центральная предельная теорема позволяет нам сделать перечисленные ниже выводы (их мы детально проанализируем в следующей главе).

1. Располагая подробными сведениями о какой-то совокупности, мы можем сделать далекоидущие выводы о любой надлежащим образом сформированной из нее выборке. Допустим, например, что у директора школы есть детальная информация о результатах сдачи стандартизованного теста всеми учащимися школы (среднее значение, среднеквадратическое отклонение и т. д.). Это значимые характеристики всей совокупности. Теперь предположим, что на следующей неделе

ожидается прибытие некоего чиновника окружного управления образования, который намерен провести такой же стандартизованный тест для 100 случайным образом отобранных учеников. Результаты, продемонстрированные этой выборкой учащихся, будут использованы для оценки качества преподавания в данной школе.

Может ли директор школы с уверенностью утверждать, что баллы этих 100 произвольно отобранных учеников будут точно отражать результаты всех учащихся данной школы при сдаче этого теста? Вполне. Согласно центральной предельной теореме, средний тестовый балл группы из 100 учащихся, как правило, не будет существенно отличаться от среднего балла всех учеников данной школы.

2. Располагая подробными сведениями о какой-либо надлежащим образом сформированной выборке (среднее значение и среднеквадратическое отклонение), мы можем сделать чрезвычайно точные выводы относительно совокупности, из которой эта выборка была получена. По сути, это обратный вариант ситуации, которую мы рассматривали в приведенном выше примере. Иными словами, мы должны поставить себя на место чиновника окружного управления образования, который оценивает школы в своем округе. В отличие от директора школы, этот чиновник не располагает результатами (или не доверяет им) сдачи стандартизованного теста всеми учащимися конкретной школы. Вместо этого он проводит в каждой школе аналогичный тест для произвольной выборки из 100 учеников.

Может ли этот чиновник быть уверен, что качество преподавания в какой-либо конкретной школе в целом можно точно оценить, основываясь на результатах сдачи стандартизованного теста группой из 100 учащихся соответствующей школы? Да, может. Центральная предельная теорема гласит, что достаточно большая выборка, как правило, не будет существенно отличаться от генеральной совокупности, а это означает, что результаты, продемонстрированные этой выборкой (то есть баллы 100 случайным образом отобранных учащихся), с достаточной степенью точности отражают результаты соответствующей совокупности в целом (то есть баллы всех учащихся конкретной школы). Разумеется, именно на таком принципе строятся все опросы. Методологически правильный опрос 1200 человек может многое поведать о настроениях всего населения страны.

Итак, если сказанное выше в п. 1 верно, то сказанное в п. 2 также должно быть верно, и наоборот. Если какая-то выборка, как правило, хорошо отражает совокупность, из которой она была сформирована, то верно и обратное: совокупность, как правило, будет похожа на выборку, сформированную из нее. (Если дети похожи на своих родителей, то и родители должны быть похожи на своих детей.)

3. Наличие данных о какой-то конкретной выборке и данных о какой-то конкретной совокупности позволяет определить, согласуется ли эта выборка с другой выборкой, которая, возможно, получена из той же совокупности. Здесь речь идет, по сути, о примере с пропавшим автобусом, приведенном в начале главы. Нам известен (приблизительно) средний вес участников марафона. Нам также известен (приблизительно) средний вес пассажиров сломавшегося автобуса. Центральная предельная теорема позволяет нам вычислить вероятность того, что конкретная выборка (упитанные люди в автобусе) была сформирована из данной совокупности (участники марафонского забега). Если эта вероятность невелика, то с высокой степенью уверенности можно заключить, что данная выборка сформирована не из интересующей нас совокупности (например, люди в автобусе отнюдь не похожи на группу бегунов-марафонцев, направляющихся к месту старта).
4. Наконец, если нам известны исходные характеристики двух выборок, то мы можем определить, сформированы ли они из одной и той же совокупности. Вернемся еще раз к становящемуся все более абсурдным примеру с автобусом. Теперь нам известно, что марафонский забег будет проводиться в данном городе — равно как и Международный фестиваль любителей сосисок. Допустим, что в обеих группах тысячи участников и обе наняли десятки автобусов, в каждый из которых поместили произвольные выборки либо бегунов-марафонцев, либо поглотителей сосисок. Допустим также, что при перевозке участников этих мероприятий столкнулись два автобуса. (Я уже признал абсурдность своего примера, поэтому сценарий развития событий не должен вас удивлять. Просто продолжайте спокойно читать дальше.) Будучи, как было сказано выше, одним из видных общественных активистов в городе, вы прибываете на место происшествия и пытаетесь определить, ехали ли оба автобуса на одно и то же мероприятие (фестиваль любителей сосисок или марафонский забег). К несчастью, никто

из пострадавших не говорит по-английски, но врачи скорой помощи, оперативно прибывшие на место происшествия, сообщают вам подробную информацию о весе каждого из пассажиров в столкнувшихся автобусах.

Основываясь лишь на этих сведениях, вы можете заключить, куда направлялись эти автобусы: на одно и то же мероприятие или на два разных. Как и прежде, положимся на интуицию. Допустим, что средний вес пассажиров в одном автобусе равняется 157 фунтам при среднеквадратическом (стандартном) отклонении 11 фунтов (это означает, что вес значительной части пассажиров находится в диапазоне от 146 до 168 фунтов). Теперь предположим, что средний вес пассажиров второго автобуса составляет 211 фунтов при среднеквадратическом отклонении 21 фунт (это означает, что вес значительной части пассажиров находится в диапазоне от 190 до 232 фунтов). Забудем на какое-то время о статистических формулах и будем опираться исключительно на логику: представляется ли вам вполне вероятным, что пассажиры обоих автобусов были случайным образом извлечены из одной и той же совокупности?

Вовсе нет. Более вероятным кажется то, что в одном из двух автобусов ехали бегуны-марафонцы, а в другом — любители сосисок. Помимо ощутимой разницы в показателях среднего веса пассажиров двух автобусов, нетрудно также заметить, что разброс в весе *между* этими двумя автобусами очень велик по сравнению с разбросом в весе *в* каждом из двух автобусов. Максимальный вес людей в «худощавом» автобусе (168 фунтов, что на одно среднеквадратическое отклонение больше среднего значения) меньше, чем минимальный вес людей в «упитанном» автобусе (190 фунтов, что на одно среднеквадратическое отклонение меньше среднего значения). Это верный признак (как со статистической, так и с логической точки зрения) того, что две выборки сформированы, скорее всего, из разных совокупностей.

Если на интуитивном уровне все это представляется вам вполне логичным, то вы уже на 93,2% приблизились к пониманию сути центральной предельной теоремы*. Чтобы придать этому интуитивному выводу некую

* Обратите внимание на весьма остроумное использование в данном случае ложной точности.

техническую солидность, нам необходимо продвинуться еще на один шаг вперед. Очевидно, когда вы заглядываете в поломанный автобус и видите там группу довольно упитанных людей в спортивных брюках свободного покроя, у вас тотчас же мелькает догадка, что вряд ли это бегуны на марафонские дистанции. Центральная предельная теорема позволяет нам подвести под свои предположения солидную теоретическую базу и придать им определенную степень уверенности.

Например, исходя из неких базовых вычислений я могу заключить, что в 99 случаях из 100 средний вес пассажиров любого случайным образом выбранного автобуса с бегунами будет отличаться не более чем на девять фунтов от среднего веса всех зарегистрированных участников марафона. Именно это служит статистическим подтверждением моей догадки, когда я натываюсь на поломанный автобус с людьми. Средний вес его пассажиров на двадцать один фунт превышает средний вес всех зарегистрированных участников марафона, а это значит, что вероятность принадлежности пассажиров этого автобуса к составу участников забега не превышает 1 шанс из 100. Это позволяет мне с 99-процентной уверенностью отвергнуть гипотезу о том, что встретившийся мне автобус перевозил спортсменов (иными словами, я могу рассчитывать на то, что сделанный мною вывод окажется правильным в 99 случаях из 100).

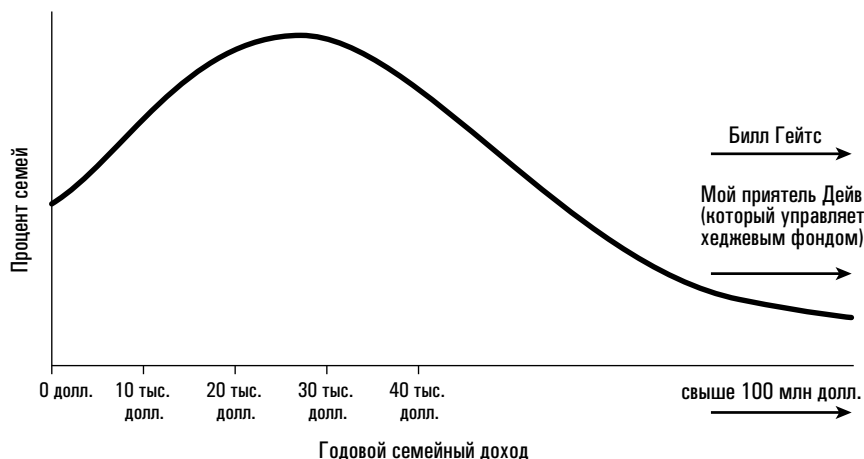
Правда, согласно теории вероятностей, в среднем я окажусь *неправ* в 1 случае из 100.

Анализ такого рода целиком следует из центральной предельной теоремы, которая, с точки зрения статистики, обладает такой же мощью и элегантностью, как действия Леброна Джеймса на баскетбольной площадке. Согласно центральной предельной теореме, средние значения выборок для любой совокупности будут распределены относительно ее среднего значения примерно по нормальному закону. Ниже я постараюсь разъяснить это положение.

1. Допустим, у нас есть некая совокупность, например все зарегистрированные участники марафона, и нас интересует вес каждого бегуна. Любая выборка участников марафона (например шестидесят бегунов, перевозимых каждым автобусом) будет характеризоваться средним значением их веса.

2. Если делать повторные выборки из всего состава зарегистрированных участников марафона, например формировать случайным образом группы из шестидесяти бегунов, то каждая из этих выборок будет характеризоваться собственным средним значением веса. Это и будут средние значения выборок.
3. Большинство этих средних значений будут очень близки к среднему значению веса для данной совокупности. Какие-то из них окажутся чуть больше, какие-то — чуть меньше. По чистой случайности лишь очень немногие из них будут существенно превышать или быть ниже среднего значения веса для данной совокупности.
Прислушайтесь к этой музыке, поскольку именно сейчас все звуки сливаются в мощное крещендо...
4. Центральная предельная теорема гласит, что эти средние значения выборок будут распределены относительно среднего значения совокупности примерно по нормальному закону. Нормальное распределение, как вы, наверное, помните из главы 2, представляет собой распределение колоколообразной формы (например, величины веса взрослых мужчин), в котором 68% наблюдений находятся на расстоянии одного среднеквадратического отклонения от среднего значения, 95% наблюдений — на расстоянии двух среднеквадратических отклонений и т. д.
5. Все эти утверждения будут истинными, как бы ни выглядело распределение исходной совокупности. Чтобы средние значения выборок были распределены по нормальному закону, вовсе не обязательно, чтобы совокупность, из которой получены эти выборки, имела нормальное распределение.

Рассмотрим реальные данные, например распределение семейного дохода в Соединенных Штатах. Семейный доход в США не распределен по нормальному закону, а, как правило, скошен вправо. В любом данном году никакая из семей не может заработать меньше 0 долларов, поэтому у данного распределения должна быть нижняя граница. Между тем, годовые доходы у какой-то небольшой группы семей могут быть очень велики — сотни миллионов, а в отдельных случаях даже миллиарды долларов. В результате можно ожидать, что распределение семейного дохода в стране будет характеризоваться длинным «хвостом» справа, нечто наподобие этого:



Медиана семейного дохода в Соединенных Штатах составляет примерно 51 900 долларов; средний семейный доход — 70 900 долларов¹. (Люди вроде Билла Гейтса сдвигают средний семейный доход вправо; вспомните последствия появления Билла Гейтса в баре, о которых рассказывалось в главе 2.) Теперь допустим, что мы берем случайную выборку из 1000 американских семей и собираем данные об их годовом семейном доходе. Что можно сказать об этой выборке, основываясь на приведенной выше информации и центральной предельной теореме?

Оказывается, довольно много. Прежде всего, можно подтвердить наше предположение о том, что среднее значение любой выборки будет равняться среднему значению совокупности, из которой такая выборка сформирована. Сущность репрезентативной выборки заключается в том, что она похожа на совокупность, из которой сформирована. Любая надлежащим образом созданная выборка не будет в среднем отличаться от Америки в целом. В такую выборку войдут и менеджеры хеджевых фондов, и бездомные, и полицейские, и все прочие основные группы населения, причем все они будут включены в выборку приблизительно в той пропорции, в какой представлены в соответствующей совокупности. Следовательно, можно ожидать, что средний семейный доход в репрезентативной выборке из 1000 американских семей приблизительно составит 70 900 долларов. Будет ли он в точности равен 70 900 долларам? Нет. Но существенно отличаться от этой суммы не будет.

Если мы возьмем несколько выборок из 1000 американских семей, то предположительно их средние значения будут группироваться вокруг

среднего значения данной совокупности, то есть 70 900 долларов. Можно ожидать, что некоторые из средних значений будут несколько выше этой суммы, а другие — несколько ниже. Может ли среди этих выборок оказаться такая, у которой средний семейный доход составит 427 000 долларов? Разумеется да, однако это очень и очень маловероятно. (Не забывайте, что мы используем правильную методологию формирования выборок, иными словами, не проводим опрос на парковке возле Greenwich Country Club.) Столь же маловероятно, что средний семейный доход в надлежащем образом сформированной выборке из 1000 американских семей составит 8000 долларов.

Все наши рассуждения основываются на простейшей логике. Центральная предельная теорема позволяет пойти еще дальше, описывая ожидаемое распределение средних значений разных выборок, группирующихся вблизи среднего значения генеральной совокупности. А именно, средние значения этих выборок вблизи среднего значения нашей совокупности (в данном случае 70 900 долларов) распределены по нормальному закону. Вспомните, что форма распределения исходной совокупности значения не имеет. Распределение семейного дохода в Соединенных Штатах характеризуется значительным скосом, *однако у распределения средних значений выборок скос отсутствует*. Если бы мы взяли 100 разных выборок, каждая из которых включает 1000 семей, и построили график частоты наших результатов, то можно было бы ожидать, что средние значения этих выборок образуют хорошо знакомое нам «колоколообразное» распределение в районе 70 900 долларов.

Чем больше количество выборок, тем точнее это распределение аппроксимируется нормальным распределением. А чем больше размер каждой выборки, тем такое распределение будет уже. Чтобы проверить этот результат, давайте проведем эксперимент с реальными данными о весе реальных американцев. Мичиганский университет выполнил повторное исследование под названием Americans' Changing Lives («Меняющаяся жизнь американцев»), которое предусматривает детальные наблюдения за несколькими тысячами взрослых американцев, в том числе и за их весом. Распределение веса слегка скошено вправо, поскольку биологически легче весить на 100 фунтов больше нормы, чем на 100 фунтов меньше нормы. Средний вес для всех взрослых в этом исследовании составляет 162 фунта.

С помощью компьютера и базового статистического программного обеспечения можно создать на основе данных Americans' Changing Lives

произвольную выборку из 100 человек. Вообще говоря, это можно делать многократно, чтобы увидеть, как полученные результаты согласуются с тем, что предсказывает нам центральная предельная теорема. Ниже приведен график распределения 100 средних значений выборок (с округлением до ближайшего фунта), сгенерированных случайным образом на основе данных *Americans' Changing Lives*.



Чем больше размер выборки и чем больше выборок, тем точнее распределение их средних значений аппроксимируется нормальным распределением. (Чтобы обеспечить применимость центральной предельной теоремы, желательно, чтобы размер выборки был не менее 30.) Это должно быть понятно на интуитивном уровне. Большой размер выборки в меньшей степени подвержен случайным отклонениям. Выборка же из 2 человек может быть сильно скошена, если в ней окажется человек с необычайно большим (или слишком малым) весом. Напротив, на выборку из 500 человек лишь очень незначительно повлияет наличие в ней нескольких человек с нестандартным весом.

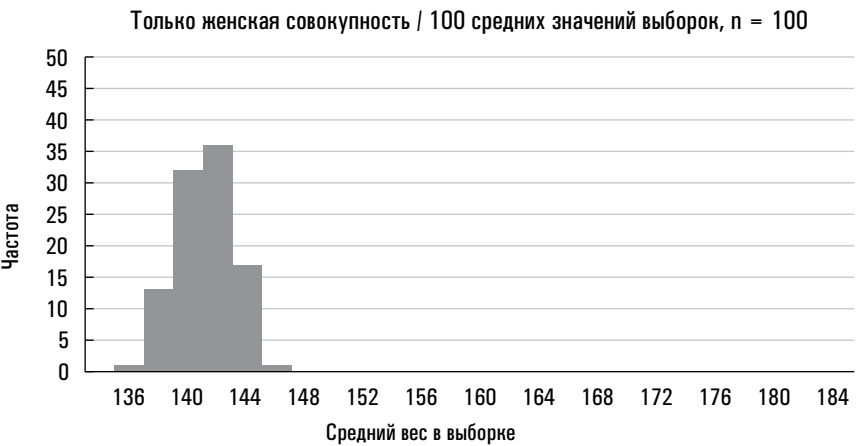
Итак, мы очень близки к тому, чтобы воплотить в жизнь все свои статистические мечты! Средние значения выборок распределены приблизительно по нормальному закону, как описано выше. Эффективность нормального распределения является следствием нашей информированности о том, какая примерно доля наблюдений окажется выше или ниже среднего значения на расстоянии, не превышающем одного среднеквадратического отклонения (68%); на расстоянии, не превышающем двух среднеквадратических отклонений (95%), и т. д. Это очень важная для нас информация.

Ранее в этой главе я указывал на возможность интуитивного вывода о том, что автобус с пассажирами, средний вес которых на двадцать пять фунтов превышает средний вес всех зарегистрированных участников марафона, вряд ли может быть потерявшимся автобусом с его участниками. Чтобы получить численное подтверждение своей интуитивной догадки — то есть иметь основания утверждать, что этот вывод окажется правильным в 95 (или в 99, или в 99,9) процентах случаев, — нам необходима еще одна техническая характеристика — стандартная (среднеквадратическая) ошибка.

Стандартная ошибка измеряет разброс средних значений выборок. Насколько предположительно близко они будут группироваться вокруг среднего значения совокупности? Здесь возможна некоторая путаница, поскольку вам уже известны два разных показателя разброса: среднеквадратическое (стандартное) отклонение и стандартная (среднеквадратическая) ошибка. Чтобы внести ясность в этот вопрос, нужно учитывать следующее.

1. Среднеквадратическое отклонение измеряет разброс в исходной совокупности. В данном случае оно может измерять разброс значения веса всех участников Framingham Heart Study, то есть разброс вблизи среднего значения для всех зарегистрированных участников марафона.
2. Стандартная ошибка измеряет разброс *средних значений выборок*. Если мы извлекли ряд выборок (в каждой по 100 значений) из Framingham Heart Study, то как будет выглядеть разброс их средних значений?
3. Вот что связывает между собой эти две концепции: стандартная ошибка является среднеквадратическим отклонением средних значений выборок! Замечательно, не правда ли?

Большая стандартная ошибка означает, что средние значения выборок разбросаны на значительных расстояниях от среднего значения совокупности; малая стандартная ошибка означает, что средние значения выборок располагаются относительно близко вокруг среднего значения совокупности. Ниже приведены три реальных примера на основе данных Americans' Changing Lives.



Второе распределение, размер выборки у которого больше, сгруппировано вблизи среднего значения плотнее, чем первое. Большой размер выборки снижает вероятность того, что ее среднее значение существенно отклонится от среднего значения совокупности. Последний набор средних значений выборок получен из подмножества рассматриваемой нами совокупности (в данном случае таким подмножеством являются женщины). Поскольку значения веса женщин в этой совокупности данных разбросаны в меньшей степени, чем значения веса всех лиц в рассматриваемой нами совокупности, вполне естественно, что вес выборок, сформированных исключительно из женской совокупности, должен быть менее разбросанным, чем выборки, извлеченных из всей совокупности Americans' Changing Lives. (Эти выборки также сгруппированы вблизи несколько отличающегося среднего значения совокупности, так как средний вес всех женщин в исследовании Americans' Changing Lives разнится со средним весом всей совокупности, охваченной данным экспериментом.)

Нарисованная мной картина носит универсальный характер. Средние значения выборок будут группироваться вблизи среднего значения совокупности *более плотно* по мере увеличения размера каждой выборки (например средние значения наших выборок группировались вблизи среднего значения совокупности более плотно, когда их размер увеличился с 20 до 100). И *менее плотно*, когда исходная совокупность окажется более «разбросанной» (например средние значения наших выборок для всей совокупности Americans' Changing Lives были более разбросанными, чем средние значения выборок лишь для женской совокупности).

Если вам до сих пор удавалось следить за логикой моего изложения, то формула для стандартной ошибки (SE) не потребует дополнительных разъяснений: $SE = s / \sqrt{n}$, где s — среднееквадратическое отклонение для совокупности, из которой сформирована данная выборка, а n — размер выборки. Не следует, однако, слишком уповать на формулы. Не забывайте привлекать на помощь интуицию. Стандартная ошибка будет большой, когда среднееквадратическое отклонение исходного распределения велико. Большая выборка, сформированная из сильно разбросанной совокупности, также, скорее всего, окажется сильно разбросанной; большая выборка, сформированная из совокупности, плотно сгруппированной вблизи среднего значения, также, скорее всего, окажется плотно сгруппированной вблизи среднего значения. Если вернуться к примеру с весом, то можно ожидать,

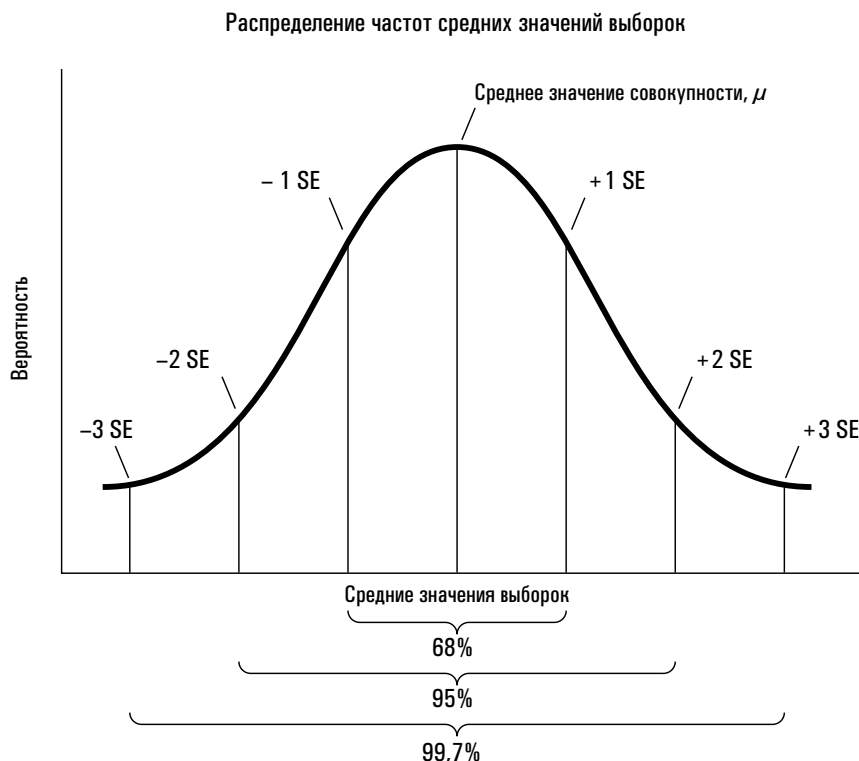
что стандартная ошибка для выборки, извлеченной из всей совокупности Americans' Changing Lives, будет большей, чем стандартная ошибка для выборки, состоящей только из мужчин в возрасте от двадцати до тридцати лет. Именно поэтому *среднеквадратическое отклонение (s) находится в числителе приведенной выше формулы.*

Аналогично можно ожидать, что стандартная ошибка будет уменьшаться по мере увеличения размера выборки, поскольку большие выборки в меньшей степени подвержены искажению со стороны экстремальных наблюдений («отщепенцев»). Именно поэтому размер выборки n находится в знаменателе формулы. (Разъяснение причины, по которой в формуле используется корень квадратный из n , мы оставим для более «продвинутых» учебников по статистике; в данном случае для нас важны базовые соотношения.)

В случае данных Americans' Changing Lives нам фактически известно среднеквадратическое отклонение этой совокупности, однако зачастую так не бывает. В отношении крупных выборок мы можем предположить, что их среднеквадратическое отклонение довольно близко к среднеквадратическому отклонению генеральной совокупности*.

Наконец, настало время подвести итог сказанному. Поскольку средние значения выборок распределены по нормальному закону (благодаря центральной предельной теореме), мы можем воспользоваться богатым потенциалом кривой нормального распределения. Мы рассчитываем, что примерно 68% средних значений всех выборок будут отстоять от среднего значения совокупности на расстоянии, не превышающем одной стандартной ошибки; 95% — на расстоянии, не превышающем двух стандартных ошибок; и 99,7% — на расстоянии, не превышающем трех стандартных ошибок.

* Когда среднеквадратическое отклонение соответствующей совокупности вычисляется на основании меньшей выборки, приведенная нами формула несколько видоизменяется: $SE = s / \sqrt{n - 1}$. Это помогает учесть то обстоятельство, что дисперсия в малой выборке может «недооценивать» дисперсию всей совокупности. Это не имеет особого отношения к более универсальным положениям, о которых идет речь в данной главе.



Теперь вернемся к отклонению (разбросу) в примере с пропавшим автобусом — правда, на этот раз призовем на помощь не интуицию, а числа. (Сам по себе этот пример остается абсурдным; в следующей главе мы рассмотрим множество более близких к реальности случаев.) Допустим, что организаторы исследования Americans' Changing Lives пригласили всех его участников на выходные в Бостон, чтобы весело провести время и заодно предоставить кое-какие недостающие данные. Участников распределяют произвольным образом по автобусам и отвозят в тестовый центр, где их взвешают, определяют рост и т. п. К ужасу организаторов мероприятия, один из автобусов пропадает где-то по пути в тестовый центр. Об этом событии оповещают в программе новостей местного радио и телевидения. Возвращаясь примерно в то же время в своем автомобиле с Фестиваля любителей сосисок, вы замечаете на обочине дороги сломавшийся автобус. Похоже, его водитель был вынужден резко свернуть в сторону, пытаясь уклониться от столкновения с лесом, неожиданно появившимся на дороге. От столь

резкого маневра все пассажиры потеряли сознание или лишились дара речи, хотя никто из них, к счастью, не получил серьезных травм. (Такое предположение понадобилось мне исключительно для чистоты приведенного здесь примера, а надежда на отсутствие у пассажиров серьезных травм объясняется моим врожденным человеколюбием.) Врачи кареты скорой помощи, оперативно прибывшие на место происшествия, сообщили вам, что средний вес 62 пассажиров автобуса составляет 194 фунта. Кроме того, оказалось (к огромному облегчению всех любителей животных), что лось, от столкновения с которым пытался увернуться водитель автобуса, практически не пострадал (если не считать легкого ушиба задней ноги), но от сильного испуга тоже потерял сознание и лежит рядом с автобусом.

К счастью, вам известен средний вес пассажиров автобуса, а также среднеквадратическое отклонение для всей совокупности Americans' Changing Lives. Кроме того, мы имеем общее представление о центральной предельной теореме и знаем, как оказать первую помощь пострадавшему животному. Средний вес участников исследования Americans' Changing Lives составляет 162 фунта; среднеквадратическое отклонение равняется 36. На основе этой информации вы можете вычислить стандартную ошибку для выборки из 62 человек (количество пассажиров автобуса, потерявших сознание): $s / \sqrt{62} = 36/7,9$, или 4,6.

Разница между средним значением этой выборки (194 фунта) и средним значением совокупности (162 фунта) равна 32 фунта, то есть значительно больше трех стандартных ошибок. Из центральной предельной теоремы вам известно, что 99,7% средних значений всех выборок будут отстоять от среднего значения совокупности на расстоянии, не превышающем трех стандартных ошибок. Таким образом, крайне маловероятно, что встретившийся вам автобус перевозит группу участников исследования Americans' Changing Lives. Будучи видным общественным активистом города, вы звоните организаторам мероприятия, чтобы сообщить, что в повстречавшемся вам автобусе, скорее всего, находится какая-то другая группа людей. Правда, в этом случае вы можете опираться на статистические результаты, а не свои «интуитивные догадки». Вы сообщаете организаторам, что отрицаете вероятность того, что найденный вами автобус именно тот, который они разыскивают, с 99,7% доверительным уровнем. А поскольку в данном случае вы разговариваете с людьми, знакомыми со статистикой, то можете не сомневаться, они понимают, что вы правы. (Всегда приятно иметь дело с умными людьми!)

Сделанные вами выводы находят дальнейшее подтверждение, когда врачи скорой помощи берут пробы крови у пассажиров автобуса и обнаруживают, что средний уровень холестерина в их крови превышает средний уровень холестерина в крови участников исследования *Americans' Changing Lives* на пять стандартных ошибок. Из этого следует, что впавшие в бессознательное состояние пассажиры — участники Фестиваля любителей сосисок. (Впоследствии это было неопровержимо доказано.)

[У этой истории оказался счастливый конец. Когда к пассажирам автобуса вернулось сознание, организаторы исследования *Americans' Changing Lives* посоветовали им проконсультироваться у специалистов-диетологов относительно опасности употребления в пищу продуктов с высоким содержанием насыщенных жиров. После таких консультаций многие из любителей сосисок решили порвать со своим позорным прошлым и вернуться к более здоровому рациону питания. Пострадавшего лося вывели в местной ветеринарной клинике и выпустили на свободу под одобрительные возгласы членов местного Общества защиты животных. Да, история почему-то умалчивает о судьбе водителя автобуса. Возможно, потому, что статистика не занимается судьбами отдельно взятых людей. Лось — совсем другое дело, замолчать его судьбу не удастся! В случае чего за него может вступить Общество защиты животных.]*

В этой главе я пытался говорить только об основах. Вы, наверное, обратили внимание, что центральная предельная теорема применима лишь в случаях, когда размер выборки достаточно велик (как правило, не менее 30). Кроме того, нам требуется относительно большая выборка, если мы намерены предположить, что ее среднеквадратическое отклонение будет примерно таким же, как и среднеквадратическое отклонение генеральной совокупности.

* Мой коллега из Чикагского университета, Джим Сэлли, сделал очень важное критическое замечание по поводу примеров с пропавшим автобусом. Он указал, что пропавший автобус — чрезвычайно большая редкость в наше время. Поэтому если нам придется искать какой-нибудь пропавший автобус, то любой встретившийся нам автобус, который окажется пропавшим или поломавшимся, наверняка будет именно тем автобусом, который нас интересует, *каким бы ни был вес пассажиров в этом автобусе*. Пожалуй, Джим прав. (Воспользуюсь такой аналогией: если вы потеряли в супермаркете своего ребенка и дирекция этого магазина сообщает по радио, что возле кассы номер шесть стоит чей-то потерявшийся ребенок, то вы наверняка сразу же решите, что речь идет именно о вашем ребенке.) Следовательно, нам не остается ничего другого, как дополнить наши примеры еще одним элементом абсурда, полагая, что пропажа автобуса является вполне рядовым событием.

Существует немало статистических поправок, которые можно применять в случае несоблюдения указанных условий, но все это похоже на сахарную глазурь на торте (и, возможно, даже на шоколадные крошки, которыми присыпают эту глазурь сверху). «Общая картина» здесь проста и чрезвычайно эффективна.

1. Если вы формируете на основе какой-либо совокупности большие (по объему) случайные выборки, то их средние значения будут распределены по нормальному закону вблизи среднего значения соответствующей совокупности (какой бы вид ни имело распределение исходной совокупности).
2. Большинство средних значений выборок будет расположено достаточно близко к среднему значению совокупности (что именно следует в том или ином случае считать «достаточно близким», определяется стандартной ошибкой).
3. Центральная предельная теорема говорит нам о вероятности того, что среднее значение выборки будет находиться не дальше определенного расстояния от среднего значения совокупности. Относительно маловероятно, что среднее значение выборки будет отстоять от среднего значения совокупности дальше, чем на расстояние двух стандартных ошибок, и крайне маловероятно, что среднее значение выборки будет отстоять от среднего значения совокупности дальше, чем на расстояние трех и более стандартных ошибок.
4. Чем меньше вероятность того, что какой-то исход оказался чисто случайным, тем больше мы можем быть уверены в том, что здесь не обошлось без воздействия какого-то другого фактора.

В этом по большому счету и заключается сущность статистического вывода. Центральная предельная теорема главным образом делает все это возможным. И до тех пор, пока Леброн Джеймс не станет столько раз чемпионом НБА, сколько Майкл Джордан (шесть), центральная предельная теорема будет производить на нас гораздо большее впечатление, чем знаменитый баскетболист.

Почему моему преподавателю статистики казалось, что я пытаюсь его обмануть

Весной, будучи уже в старших классах колледжа, я решил прослушать курс лекций по статистике. Вообще говоря, в то время я не испытывал особой любви ни к ней, ни к любым другим наукам, базирующимся на математике, но пообещал отцу, что прослушаю этот курс лекций с условием, что мне разрешат на десять дней поехать в СССР. Короче говоря, это был взаимовыгодный обмен, причем, как оказалось, статистика увлекла меня гораздо больше, чем я предполагал, к тому же мне удалось побывать в СССР весной 1988 года. Кто же тогда знал, что эта страна буквально через несколько лет расстанется со своим коммунистическим прошлым!

В действительности эта история имеет непосредственное отношение к материалу данной главы: дело в том, что в то время я не уделял изучению статистики должного внимания. Помимо всего прочего, я писал тогда дипломную работу, и до ее сдачи у меня оставалось не так уж много времени. По мере прохождения курса статистики мы периодически сдавали промежуточные экзамены, многие из которых я либо проваливал, либо попросту игнорировал. К середине семестра багаж знаний, полученных мною по данной дисциплине, был настолько скудным, что я мог рассчитывать исключительно на чудо. Но буквально за несколько недель до окончания семестра произошли два важных события. Во-первых, я дописал дипломную работу, в результате чего у меня появилось довольно много свободного времени. И во-вторых, осознал, что статистика не такая уж сложная наука, как мне казалось до этого. Я начал усиленно штудировать учебники по статистике, наверстывая упущенный материал. Итоговый экзамен по статистике я сдал на отлично.

Именно тогда преподаватель статистики (к сожалению, я забыл его имя) вызвал меня к себе в кабинет. Не помню точно, что он мне сказал, но это было нечто вроде: «Вы добились потрясающих успехов по сравнению с серединой

семестра». Однако его слова звучали отнюдь не как похвала моим достижениям. Напротив, в них мне послышалось скрытое обвинение в том, что во время сдачи экзамена я пользовался шпаргалками. Учитель не мог поверить в то, что студент, так «мелко плававший» на промежуточных экзаменах, способен на столь мощный рывок к концу семестра. Тогда мне было очень обидно, что меня подозревают в обмане, но со временем я понял этого человека и нисколько не осуждаю. Практически по всем предметам, изучаемым в колледже, наблюдается высокая степень корреляции между результатами, которые студенты демонстрируют на промежуточных и итоговых экзаменах. Очень маловероятно, что студент, получивший на промежуточных экзаменах оценку ниже средней, покажет блестящий результат во время сдачи итоговых экзаменов.

Я объяснил преподавателю, что, завершив написание дипломной работы, решил со всей серьезностью отнестись к изучению курса статистики (для этого мне понадобилось всего лишь читать рекомендованные им главы учебника и своевременно выполнять домашние задания). Мне показалось, что я его убедил, хотя его подозрение в том, что во время экзамена я пользовался шпаргалками (пусть даже не высказанное вслух), по-прежнему не давало мне покоя.

Хотите верьте, хотите нет, но этот случай воплощает в себе многое из того, что вам нужно знать о статистическом выводе, в том числе о его достоинствах и потенциальных недостатках. *Статистика не может ничего утверждать с определенностью.* Напротив, сила статистического вывода проистекает из наблюдения некой картины или исхода и последующего использования теории вероятностей для получения его (ее) самого вероятного объяснения. Допустим, в ваш город прибыл большой любитель азартных игр и предлагает вам пари: он выигрывает 1000 долларов, если в результате подбрасывания игральной кости выпадет шестерка; вы выигрываете 500 долларов, если выпадет любое другое число, — очень выгодный, на ваш взгляд, вариант. Затем в результате десяти подбрасываний игральной кости у него десять раз подряд выпадает шестерка. Вам не остается ничего другого, как уплатить 10 000 долларов.

Одно возможное объяснение этого феномена — необычайное везение вашего визави. Альтернативное объяснение — обман с его стороны. Вероятность того, что в ходе десяти подбрасываний «правильной» игральной кости десять раз подряд выпадет шестерка, равняется примерно 1 шансу

из 60 миллионов. Вы не можете доказать, что ваш партнер смошенничал, но вы должны по крайней мере проверить, все ли в порядке с игровой костью.

Разумеется, самое вероятное (правдоподобное) объяснение не всегда окажется правильным. Порой действительно случается то, что не должно было бы случиться. В Линду Купер из штата Южная Каролина четыре раза попадала молния¹. (По оценкам Федеральной службы чрезвычайных ситуаций, вероятность *однократного* попадания молнии в человека составляет 1 шанс из 600 000.) Страховая компания Линды Купер не может отказать ей в выплате компенсации лишь на основании того, что полученные ею травмы в результате попаданий молнии статистически невероятны. Что касается моего выпускного экзамена по статистике, то у преподавателя были причины подозревать меня в мошенничестве, поскольку он увидел крайне маловероятную картину (именно так следователи выявляют манипуляции с результатами стандартизованных тестов, а Комиссия по ценным бумагам и биржам — инсайдерские торговые операции с ценными бумагами). Но маловероятная картина остается просто маловероятной картиной, если наши подозрения не подтверждаются какими-либо дополнительными свидетельствами. Ниже мы обсудим ошибки, которые могут возникнуть в случаях, когда вероятность направляет нас по ложному пути.

На этом этапе нам следует уяснить, что статистический вывод использует данные для получения ответов на важные вопросы. Эффективно ли новое лекарство, предназначенное для лечения заболеваний сердца? Являются ли мобильные телефоны причиной развития раковых опухолей мозга? Обратите внимание: я вовсе не утверждаю, что статистика может ответить на такие вопросы *однозначно*. Статистический вывод говорит лишь о том, что вполне вероятно, а что — маловероятно или даже крайне невероятно. Исследователи не могут утверждать, что новое лекарство, предназначенное для лечения заболеваний сердца, действительно эффективно, даже располагая результатами его надлежащим образом проведенных клинических испытаний. В конце концов, вполне возможно, что при лечении пациентов в подопытной и контрольной группах появится случайное отклонение, никак не связанное с новым препаратом. То, что у 53 из 100 пациентов, принимающих это лекарство, наметилось существенное улучшение состояния здоровья, тогда как в группе пациентов, принимающих плацебо, такая картина наблюдается у 49 пациентов из 100, не дает нам права безапелляционно заявлять

об эффективности нового препарата. Такой исход можно объяснить случайным отклонением между двумя группами пациентов, а вовсе не действием нового лекарства.

Допустим, однако, что у 91 из 100 пациентов, принимающих новое лекарство, произошло существенное улучшение состояния здоровья, тогда как в контрольной группе здоровье значительно улучшилось только у 49 из 100 пациентов. Конечно, и на сей раз не исключено, что столь впечатляющий результат никак не связан с приемом нового препарата; возможно, пациентам в подопытной группе просто улыбнулась удача (а может, все дело в их жизнелюбии и оптимизме). *Однако в данном случае такое объяснение из разряда маловероятных.* На формальном языке статистического вывода, исследователи, скорее всего, заключили бы следующее. 1) Если бы экспериментальное лекарство никак не сказывалось на состоянии пациентов, то столь сильное отклонение в исходах между теми, кто его принимает, и теми, кто принимает плацебо, явилось бы большой редкостью. 2) Поэтому крайне маловероятно, что препарат не оказывает положительного воздействия на состояние пациентов. 3) Альтернативное — и более вероятное — объяснение полученной нами картины заключается в том, что экспериментальное лекарство оказывает положительный эффект.

Статистический вывод — это процесс, посредством которого данные позволяют нам делать обоснованные заключения. Именно в этом его достоинство! Задача статистики не в выполнении огромного множества строгих математических расчетов, а в том, чтобы помочь нам лучше разобраться в важных социальных (и не только) явлениях. Статистический вывод — это, по сути, союз двух уже обсуждавшихся нами концепций: данных и вероятности (с определенной помощью со стороны центральной предельной теоремы). В настоящей главе я воспользовался одним значимым методологическим упрощением: все приведенные мною примеры предполагают, что мы используем большую, надлежащим образом сформированную выборку. Это предположение означает возможность применения центральной предельной теоремы и то, что среднее значение и среднеквадратическое отклонение для любой выборки будет примерно таким же, как среднее значение и среднеквадратическое отклонение для совокупности, из которой она сформирована. Оба допущения делают наши расчеты проще.

Статистический вывод не зависит от этого упрощающего предположения, но систематизированные методологические уточнения, позволяющие

работать с малыми выборками или неполными данными, зачастую лишь препятствуют пониманию общей картины. Цель в данном случае — сделать так, чтобы читатель смог оценить важность и богатые возможности статистического вывода, а также механизм его действия. После того как вы уясните это, можно переходить на более высокий уровень сложности.

Одним из самых распространенных инструментов в статистическом выводе является проверка гипотез. Фактически я уже знакомил вас с этой концепцией — правда, не прибегая к использованию заумной терминологии. Как указывалось выше, сама по себе статистика не может ничего *доказать*; вместо этого мы применяем статистический вывод, чтобы принимать или отвергать объяснения на основе их вероятности. Точнее говоря, любой статистический вывод начинается с подразумеваемой или явно сформулированной основной (так называемой нулевой) гипотезы. Это наша начальная гипотеза, которая будет отвергнута или принята исходя из последующего статистического анализа. Если мы отвергаем нулевую гипотезу, то, как правило, принимаем какую-то альтернативную гипотезу, которая в большей степени соответствует наблюдаемым нами данным. Например, исходным предположением (или основной гипотезой) в суде является невиновность подсудимого (так называемая презумпция невиновности). Задача обвинения — убедить судью или присяжных в необходимости отклонить это предположение и принять альтернативную гипотезу, что подсудимый виновен. С точки зрения логики альтернативная гипотеза представляет собой заключение, которое должно быть истинным, если мы можем опровергнуть основную гипотезу. Рассмотрим несколько примеров.

Нулевая гипотеза: новый экспериментальный препарат не более эффективен для профилактики малярии, чем плацебо.

Альтернативная гипотеза: новый экспериментальный препарат способствует профилактике малярии.

Данные: члены случайным образом сформированной группы будут принимать новое экспериментальное лекарство, а контрольная группа будет принимать плацебо. По окончании определенного периода в группе, принимавшей новый препарат, было зафиксировано значительно меньше случаев заболевания малярией, чем в контрольной группе. Это было бы крайне маловероятно, если бы новое экспериментальное лекарство не оказывало медицинского воздействия. Таким образом, мы *отвергаем* нулевую гипотезу, что новый препарат не имеет медицинских последствий (конечно же,

помимо известного эффекта плацебо), и принимаем логическую альтернативу, то есть альтернативную гипотезу: новое экспериментальное лекарство способствует профилактике малярии.

Такой методологический подход достаточно необычен, поэтому приведу еще один пример. Опять же обратите внимание, что нулевая и альтернативная гипотезы логически дополняют друг друга. Если одна оказывается истинной, то другая таковой не является. Или если мы отвергаем одну гипотезу, то должны принять другую. Теперь еще один пример.

Нулевая гипотеза: лечение заключенных от наркозависимости не снижает вероятности их повторного ареста после выхода из тюрьмы.

Альтернативная гипотеза: лечение заключенных от наркозависимости снижает вероятность их повторного ареста после выхода из тюрьмы.

Данные (гипотетические): заключенных случайным образом разделили на две группы, «подопытная» группа проходила курс лечения от наркозависимости, а контрольная группа — нет. Через пять лет оказалось, что вероятность повторного ареста членов обеих групп примерно одинакова. То есть в этом случае мы не *можем отвергнуть* нулевую гипотезу*. Эти данные не дают нам повода отклонить исходное предположение о том, что лечение заключенных от наркозависимости не спасает их от повторного попадания за решетку.

Это может показаться нелогичным, но исследователи часто формулируют нулевую гипотезу в надежде, что им удастся отвергнуть ее. В обоих при-

* С точки зрения семантики мы еще не *доказали*, что нулевая гипотеза истинная (то есть что лечение заключенных от наркозависимости не имеет никакого эффекта). Такое лечение может оказаться чрезвычайно эффективным для какой-либо другой группы заключенных. Или, возможно, в этой подопытной группе значительно большее число заключенных совершили бы повторные преступления, если бы не прошли курс лечения от наркозависимости. В любом случае на основе собранных данных нам просто *не удалось отвергнуть* нулевую гипотезу. Существует аналогичная разница между «неспособностью отвергнуть» нулевую гипотезу и ее принятием. Сам по себе факт, что одному исследованию не удалось опровергнуть утверждение о том, что лечение от наркозависимости не помогает предотвратить повторный арест, еще не означает, что мы должны согласиться с тем, что лечение от наркозависимости бесполезно. С точки зрения статистики здесь имеет место существенная разница. С учетом сказанного следует отметить, что подобные исследования зачастую проводятся с целью информирования полиции, и тюремная администрация, которой приходится решать, как правильно распределить ресурсы, может считать лечение от наркозависимости неэффективным инструментом до тех пор, пока не убедится в обратном. В этом случае, как и в других при использовании статистических данных, следует полагаться на здравый смысл.

веденных выше примерах «успех» исследования (создание нового лекарства для профилактики малярии или снижение вероятности повторного ареста) подразумевал отказ от нулевой гипотезы. Сделать это на основе имеющихся данных удалось лишь в одном из случаев (лекарство для профилактики малярии).

В зале суда порогом для отмены презумпции невиновности является качественная оценка, что подсудимый «виновен ввиду разумных оснований для сомнения». Что именно означает в каждом конкретном случае такая формулировка, решает судья или присяжные заседатели. Статистика использует аналогичную основополагающую идею, но формула «виновен ввиду разумных оснований для сомнения» определяется не качественно, а количественно. Исследователи обычно спрашивают: если нулевая гипотеза истинна, то какова вероятность того, что мы наблюдаем такую картину данных по чистой случайности? Если мы воспользуемся приведенным в начале главы примером, то ученые-медики могут спросить: если это экспериментальное лекарство не способствует излечению сердечно-сосудистых заболеваний (нулевая гипотеза), то какова вероятность того, что состояние здоровья 91 из 100 пациентов, принимавших его, улучшилось, если учесть, что улучшение состояния здоровья было отмечено лишь у 49 из 100 пациентов, принимавших плацебо? Если имеющиеся в нашем распоряжении данные свидетельствуют о крайней маловероятности нулевой гипотезы (как в примере с экспериментальным лекарством), то мы должны отвергнуть ее и принять альтернативную гипотезу (о том, что экспериментальное лекарство способствует излечению от сердечно-сосудистых заболеваний).

С учетом этого давайте еще раз вернемся к скандалу, вызванному махинациями с результатами стандартизированных тестов в Атланте, о которых мы неоднократно упоминали в этой книге. Эти результаты привлекли к себе внимание контролирующих органов из-за высокого количества исправлений неправильных ответов на правильные. Понятно, что учащиеся, которым приходится сдавать стандартизированные тесты, время от времени исправляют свои ответы. Не исключено и то, что каким-то группам учащихся, прибегающих к таким исправлениям, особенно везет — и это вовсе не связано с какими-либо махинациями. Именно поэтому основная гипотеза сводится к тому, что результаты сдачи стандартизированных тестов в любом конкретном учебном округе правильны (с точки зрения закона) и что любые

исправления — не более чем продукт случайного стечения обстоятельств. Мы ни в коем случае не хотим наказывать учеников, преподавателей или администраторов из-за того, что необычайно высокий процент учащихся внесли в свои листы с ответами разумные исправления, сделав это буквально за несколько минут до окончания важного государственного экзамена.

Но словосочетание «необычайно высокий» отнюдь не описывает того, что происходило в Атланте. Количество исправлений неправильных ответов на правильные в листах с ответами некоторых классов превышало норму данного штата на 20–50 среднеквадратических (стандартных) отклонений. (Чтобы было понятнее, *что* это означает, вспомним, что большинство наблюдений в любом распределении, как правило, отклоняется от среднего значения не более чем на два среднеквадратических отклонения.) Так какова же вероятность того, что учащимся в Атланте удалось *по чистой случайности* исправить столь большое количество неправильных ответов на правильные? Официальный представитель Министерства образования, который проанализировал эти данные, описал вероятность того, что картина, зафиксированная в Атланте, сложилась исключительно в силу случайного стечения обстоятельств и вовсе не является результатом махинаций, как примерно равную вероятности появления на трибунах стадиона Georgia Dome 70 000 зрителей ростом свыше семи футов². Может такое случиться? Теоретически да, может. Насколько велика вероятность? Чрезвычайно мала!

Тем не менее власти штата Джорджия, столицей которого является Атланта, не смогли предъявить кому-либо обвинение в манипулировании результатами стандартизированных тестов, точно так же как мой преподаватель статистики не мог (и не должен был) вышвырнуть меня из школы только потому, что я сдал выпускной экзамен по статистике успешнее, чем промежуточный. *Властям штата Джорджия не удалось доказать факт мошенничества с оценками стандартизированных тестов.* Они, конечно, могли отвергнуть нулевую гипотезу, что эти результаты законны, причем «с высокой степенью уверенности» (это означало, что наблюдаемая ими картина была почти невозможной в обычных условиях), и принять альтернативную гипотезу, согласно которой результаты сдачи стандартизированных тестов в Атланте стали следствием махинаций. (В официальных документах они, наверное, использовали более дипломатичную формулировку.) В ходе дальнейшего расследования удалось выявить факты мошенничества с оценками стандартизированных тестов. В объяснительных записках преподавателями

приводились факты исправления ими неправильных ответов на правильные, заблаговременного ознакомления учащихся с правильными ответами, предоставления возможности отстающим ученикам списывать правильные ответы у отличников и даже указания учителем правильных ответов в тот момент, когда он останавливался возле парты ученика. Самым вопиющим примером махинаций было исправление ответов преподавателями непосредственно во время пикника, на который они собрались после экзаменов, прихватив с собой экзаменационные работы.

В примере с экзаменами в Атланте мы могли отвергнуть основную гипотезу («махинаций не было»), поскольку картина, зафиксированная в результате сдачи тестов, представлялась крайне маловероятной, если исходить из того, что обмана не было. Но насколько неправдоподобной должна быть нулевая гипотеза, чтобы мы могли ее отклонить и прибегнуть к какому-то альтернативному объяснению?

Одно из самых распространенных пороговых значений, используемых исследователями для отклонения нулевой гипотезы, — 5% (его нередко представляют в форме десятичной дроби: 0,05). Данная вероятность известна как *уровень значимости* и представляет собой верхнюю границу вероятности возникновения некоей картины данных в случае, если бы основная гипотеза оказалась верна*. Не спешите выражать свое возмущение: в действительности это не так сложно, как могло показаться на первый взгляд.

Что такое уровень значимости 0,05? Мы можем отвергнуть при нем основную гипотезу, если вероятность исхода, по крайней мере такого же экстремального, как тот, который мы наблюдали бы, если бы она была истинной, оказывалась меньше 5%. Попытаюсь объяснить это положение на простом примере. Хотя я себя и ругаю, но вынужден опять вернуться к нашему пресловутому пропавшему автобусу. Предположим, вам поручено пролить свет на очередную ситуацию, в которую он угодил (честь выполнить эту важную миссию вам оказана, в частности, с учетом героических усилий, приложенных в предыдущей главе). На сей раз вы прикомандированы к группе исследователей *Americans' Changing Lives*, которые предоставили вам чрезвычайно ценные данные, призванные помочь в выполнении важной миссии. В каждом из автобусов, арендованных организаторами исследования,

* В статистике уровнем значимости называют вероятность отклонить нулевую гипотезу при условии, что она истинна. Это так называемая ошибка первого рода. Об этой ошибке см. далее. *Прим. ред.*

находится примерно 60 пассажиров, поэтому мы можем рассматривать их как случайную выборку, сформированную из всей совокупности Americans' Changing Lives. Итак, вас разбудили рано утром, сообщив о захвате одного из автобусов группой террористов (ярких поборников прав людей, страдающих ожирением) в районе Бостона*. Ваша задача — спрыгнуть с вертолета на крышу движущегося автобуса, проникнуть внутрь через аварийный выход и тайком определить, основываясь исключительно на собственных оценках веса пассажиров, являются ли они участниками исследования Americans' Changing Lives. (Между прочим, этот сюжет ничуть не менее правдоподобен, чем сюжеты большинства приключенческих фильмов, зато гораздо более поучителен с образовательной точки зрения.)

После того как вертолет взлетает с базы войск спецназа, вам вручают автомат, несколько гранат, наручные часы (которые также могут выполнять функции видеокамеры с высоким разрешением) и вычисленные нами в предыдущей главе данные о среднем весе и стандартной ошибке для выборок, сформированных из участников исследования Americans' Changing Lives. Любая случайная выборка из 60 его участников будет иметь ожидаемый средний вес 162 фунта и среднеквадратическое отклонение 36 фунтов, поскольку именно таковы среднее значение и среднеквадратическое отклонение для всех участников исследования (генеральной совокупности). С помощью этих данных вы можете вычислить стандартную ошибку для среднего значения выборок: $s / \sqrt{n} = 36 / \sqrt{60} = 36 / 7,75 = 4,6$. В центре управления миссией представленное ниже распределение выводится на внутреннюю поверхность сетчатки вашего правого глаза, чтобы вы могли использовать его в качестве справочной информации, после того как проникнете в автобус и будете тайно прикидывать вес всех его пассажиров.

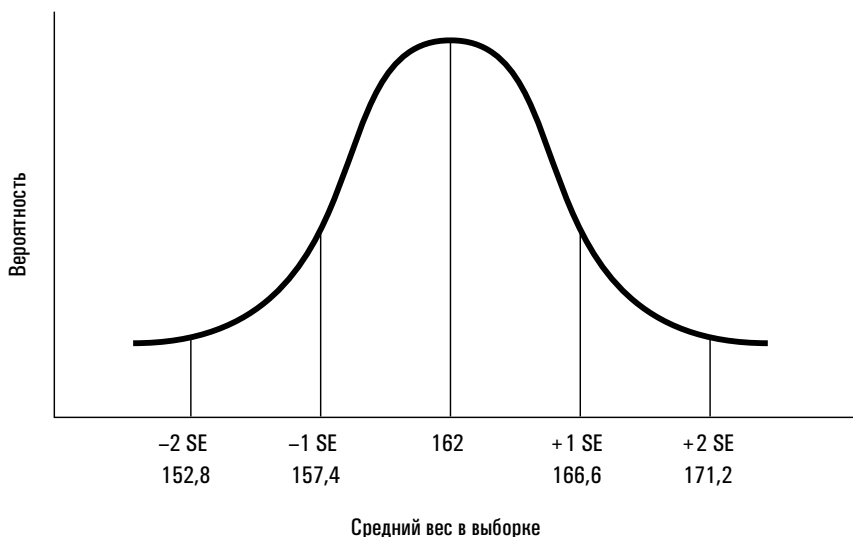
Как следует из представленного распределения, можно ожидать, что средний вес приблизительно 95% всех выборок из 60 человек, сформированных из участников исследования Americans' Changing Lives, будет отстоять от среднего значения совокупности не более чем на две стандартные ошибки, то есть находиться в пределах от 153 фунтов до 171 фунта**. И наоборот,

* Этот пример навеян реальными событиями. Понятное дело, многие подробности изменены исходя из соображений национальной безопасности. Что же касается меня, то я не могу ни подтвердить, ни отрицать в них своего участия.

** Точнее говоря, 95% средних значений всех выборок будут находиться в пределах 1,96 стандартной ошибки выше или ниже среднего значения совокупности.

лишь в 5 случаях из 100 средний вес выборки из 60 человек, сформированной случайным образом из участников исследования Americans' Changing Lives, окажется больше 171 фунта или меньше 153 фунтов. (Вы выполняете так называемую двустороннюю проверку гипотезы; разницу между «двусторонней» и «односторонней» проверками я разъясню в приложении, помещенном в конце главы.) Ваш руководитель из центра контртеррористических операций решил, что уровень значимости для вашей миссии равняется 0,05. Если средний вес 60 пассажиров в автобусе, захваченном террористами, окажется больше 171 фунта или меньше 153 фунтов, то вам придется отвергнуть нулевую гипотезу о том, что в автобусе едут участники исследования Americans' Changing Lives, и принять альтернативную гипотезу, что в автобусе находятся 60 человек, направляющихся в какой-то другой пункт назначения, и ждать дальнейших указаний.

Распределение средних значений выборок



Вы успешно проникаете в движущийся автобус и тайно определяете вес его пассажиров. Оказывается, что он составляет 136 фунтов, то есть меньше среднего значения совокупности более чем на две стандартные ошибки. (Еще одной важной подсказкой для вас становится то, что все пассажиры — дети, одетые в футболки с надписью «Глендейлский хоккейный лагерь».)

Руководствуясь инструкциями по выполнению вашей миссии, вы можете отклонить нулевую гипотезу, что этот автобус перевозит случайную выборку из 60 участников исследования *Americans' Changing Lives*. Это означает, что 1) средний вес пассажиров автобуса попадает в диапазон, который, согласно нашим ожиданиям, может наблюдаться лишь в 5 случаях из 100, если бы основная гипотеза была верна и автобус действительно перевозил участников исследования *Americans' Changing Lives*; 2) вы можете отвергнуть основную гипотезу с уровнем значимости 0,05 и 3) в среднем в 95 случаях из 100 ваше решение отклонить нулевую гипотезу окажется правильным, а в 5 случаях из 100 вы ошибетесь, то есть придете к заключению, что автобус не перевозил участников исследования *Americans' Changing Lives*, хотя на самом деле он их перевозил. Просто случилось так, что средний вес этой выборки участников исследования *Americans' Changing Lives* оказался существенно выше или ниже среднего значения для всех участников данного исследования.

Однако на этом ваша миссия не заканчивается. Ваш куратор из центра контртеррористических операций (в киноверсии данного примера эту роль играет Анджелина Джоли) просит вас вычислить *p*-значение для вашего результата. *p*-значение — это вероятность получения результата, по меньшей мере такого же экстремального, как тот, который мы наблюдали бы, если бы нулевая гипотеза была верна. Средний вес пассажиров автобуса равняется 136 фунтов, что меньше среднего значения для данной совокупности (участников исследования *Americans' Changing Lives*) на 5,7 стандартной ошибки. Вероятность получения результата, по меньшей мере такого же экстремального, если бы это действительно была выборка из участников исследования *Americans' Changing Lives*, не превышает 0,0001. (На языке, принятом у исследователей, это было бы обозначено как $p < 0,0001$.) Завершив свою миссию, вы выпрыгиваете из движущегося автобуса и оказываетесь на пассажирском сиденье автомобиля с откидным верхом, движущегося параллельным автобусу курсом.

[Эта история со счастливым концом. После того как террористы, выступающие за права тучных людей, узнали о том, что в вашем городе проводится Международный фестиваль любителей сосисок, они сразу же согласились отпустить заложников и пообещали продолжать борьбу за права людей, страдающих ожирением, исключительно мирными средствами, пропагандируя и организуя фестивали любителей сосисок по всему миру.]

* * *

Если уровень значимости 0,05 кажется вам в какой-то мере произвольным, то вы абсолютно правы: так оно и есть! Не существует единого стандартизированного статистического порога для отказа от нулевой гипотезы. Значения 0,01 и 0,1 тоже широко используются для выполнения описанного выше анализа.

Очевидно, что отказ от нулевой гипотезы с уровнем значимости 0,01 (то есть когда наши шансы наблюдать какой-либо результат в этом диапазоне, если бы нулевая гипотеза была верна, составляют менее 1 из 100) обладает большим статистическим весом, чем отказ от нулевой гипотезы с уровнем значимости 0,1 (то есть когда наши шансы наблюдать данный результат, если бы нулевая гипотеза была верна, составляют менее 1 из 10). Плюсы и минусы тех или иных уровней значимости будут обсуждаться ниже в этой главе. Пока же запомним следующее: когда мы можем отвергнуть основную гипотезу с некоторым разумным уровнем значимости, соответствующие результаты считаются «статистически значимыми».

В реальной жизни это означает вот что. Когда вы читаете в газете, что те, кто съедает двадцать булочек из отрубей в день, реже страдают раком толстой кишки, чем те, кто не употребляет пищу из отрубей в больших количествах, научные исследования, на основании которых сделан этот вывод, вероятнее всего, выглядели примерно так. 1) Исходя из некоторой объемной совокупности данных ученые определили, что те, кто ежедневно съедает по меньшей мере двадцать булочек из отрубей, болеют раком толстой кишки реже, чем те, кто не увлекается пищей из отрубей. 2) Основная гипотеза исследователей звучала так: поедание булочек из отрубей не влияет на заболеваемость раком толстой кишки. 3) Разницу в заболеваемости раком толстой кишки между теми, кто ест булочки из отрубей, и теми, кто игнорирует их, невозможно объяснить чистой случайностью. Точнее говоря, если поедание булочек из отрубей не связано с заболеваемостью раком толстой кишки, то вероятность *чисто случайного* наблюдения столь большой разницы должна быть ниже некоторого порога, например 0,05. (Этот порог исследователи должны устанавливать до выполнения статистического анализа, чтобы избежать его выбора постфактум, что бывает очень удобно, когда полученным результатам требуется придать значимость.) 4) Соответствующая научная публикация, наверное, содержит примерно такой вывод: «Мы обнаружили статистически значимую зависимость между ежедневным употреблением

двадцати и более булочек из отрубей и снижением заболеваемости раком толстой кишки. Эти результаты значимы на уровне 0,05».

Когда я впоследствии читал об этом исследовании в газете Chicago Sun-Times, по привычке завтракая ветчиной и яйцами, заголовок статьи «20 булочек из отрубей в день уберегут вас от рака толстой кишки!» чрезвычайно меня заинтересовал. Хотя он показался мне гораздо интереснее самой статьи, на мой взгляд, он грешил существенной неточностью. В действительности исследователи вовсе не заявляли, будто поедание булочек из отрубей снижает риск заболевания раком толстой кишки; они лишь продемонстрировали наличие отрицательной корреляции между употреблением булочек из отрубей и заболеваемостью раком толстой кишки в одной объемной совокупности данных. Но такой статистической связи недостаточно, чтобы доказать, что булочки из отрубей послужили *причиной* «улучшения состояния здоровья». В конце концов, те, кто ест булочки из отрубей (особенно если это целых двадцать штук за день!), наверняка делают много чего другого, чтобы снизить риск заболевания раком толстой кишки, например практически не употребляют красного мяса, регулярно занимаются физическими упражнениями, периодически обследуются и т. п. (Это так называемая систематическая ошибка здорового человека, о которой рассказывалось в главе 7.) В чем же состоит подлинная причина снижения риска заболевания раком толстой кишки: в употреблении булочек из отрубей, каких-то других особенностях поведения или личных качествах, характерных для любителей таких булочек? Это различие между корреляцией и причинно-следственной связью очень важно для правильной интерпретации статистических результатов. Чуть позже мы еще вернемся к утверждению о том, что «корреляция и причинно-следственная зависимость — не одно и то же».

Кроме того, должен отметить, что статистическая значимость ничего не говорит о *степени* связи. У тех, кто употребляет много булочек из отрубей, заболеваемость раком толстой кишки действительно может оказаться ниже — но насколько ниже? Разница в заболеваемости раком толстой кишки между теми, кто ест много булочек, и теми, кто их не ест, может быть очень несущественной; выяснение статистической значимости лишь означает, что наблюдаемый нами эффект, каким бы ничтожным он ни был, по-видимому, не является чистой случайностью. Допустим, вы узнали, что результаты надлежащим образом организованного и проведенного исследования продемонстрировали наличие статистически значимой положительной связи

между поеданием банана перед сдачей школьного экзамена по математике и получением по нему более высокой оценки. Прежде всего вас интересует, насколько силен этот эффект. Если, например, средняя оценка за экзамен по математике составляет 500 баллов, то ее повышение на 0,9 балла вряд ли радикально изменит вашу жизнь. В главе 11 мы еще вернемся к разнице между *степенью* и *значимостью*, когда будем говорить об интерпретации статистических результатов.

Однако вывод об «отсутствии статистически значимой связи» между двумя переменными означает, что любую связь между этими переменными можно объяснить исключительно чистой случайностью. Газета The New York Times недавно провела собственное расследование относительно правдивости заявлений некоторых компаний, занимающихся разработкой программного обеспечения, о том, что их продукты повышают успеваемость учащихся. Желание изобличить обман у сотрудников The New York Times возникло после того, как в их руки попали данные, свидетельствующие об обратном³. В материале, опубликованном The New York Times, утверждалось, что Университет Карнеги-Меллон продает компьютерную программу под названием Cognitive Tutor, сопровождаемую лозунгом «Революционный курс математики! Революционные результаты!» Между тем, оценка Cognitive Tutor, проведенная Министерством образования США, показала, что данный продукт «не оказывает никакого заметного влияния на результаты экзаменов по математике в старших классах. (The New York Times считает, что в соответствующей маркетинговой кампании следовало бы использовать более скромные заявления, например: «Заурядный курс математики. Сомнительные результаты».) В действительности анализ десяти программных продуктов, предназначенных для обучения математике или чтению, показал, что девять из них «не оказывают статистически значимого влияния на итоги сдачи экзаменов». Иными словами, любые различия в успеваемости между учащимися, которые пользуются и не пользуются этими программными продуктами, вполне могут быть обусловлены чистой случайностью.

Сейчас я сделаю небольшую паузу, чтобы напомнить вам, почему все это для нас так важно. В мае 2011 года в газете The Wall Street Journal вышла статья под заголовком «Причина аутизма в размере мозга». Это был настоящий прорыв, поскольку причины аутизма до сих пор не установлены. В первом же предложении этой статьи, в которой кратко излагался материал, опубликованный ранее в журнале Archives of General Psychiatry, сообщалось: «У детей, страдающих

аутизмом, объем мозга больше, чем у здоровых детей, причем, согласно результатам нового исследования, обнаруженным в понедельник, увеличение объема мозга, по-видимому, происходит в возрасте до двух лет»⁴. На основе томографического обследования 59 детей, страдающих аутизмом, и 38 здоровых детей ученые из Университета Северной Каролины пришли к выводу, что объем мозга у детей-аутистов на 10% больше, чем у их здоровых сверстников.

Возникает естественный медицинский вопрос: существует ли какая-либо физиологическая особенность у мозга ребенка, страдающего аутизмом? Если да, то это может помочь нам понять причины развития аутизма, а также найти способы его лечения или профилактики.

Появляется и соответствующий статистический вопрос: могут ли исследователи делать далекоидущие выводы относительно общих причин аутизма, основываясь на обследовании сравнительно небольшой группы детей, страдающих аутизмом (59), и еще меньшей контрольной группы (38) — то есть всего 97 участников обследования? Ответ: да, могут. Ученые пришли к заключению, что вероятность наблюдения различий в общем объеме мозга, которые они обнаружили в двух своих выборках, составляла бы 2 из 1000 ($p = 0,002$), если на самом деле в совокупности в целом не существует никакой разницы в объеме мозга между детьми-аутистами и здоровыми детьми.

Я обратился к оригинальному исследованию, результаты которого были опубликованы в журнале *Archives of General Psychiatry*⁵. Методы, использованные в нем, ничуть не сложнее уже освоенных нами концепций. Приведу краткий обзор подоплеку этого социально и статистически значимого результата. Во-первых, вы должны признать, что каждая группа детей, 59 из которых страдают аутизмом, а 38 здоровы, представляет собой довольно крупную выборку, сформированную из соответствующих им совокупностей, то есть всех детей-аутистов и всех здоровых детей. Эти выборки достаточно большие для того, чтобы можно было применить центральную предельную теорему. Если вы уже подзабыли, в чем ее суть, я вам напомню: 1) средние значения выборок из какой-либо совокупности будут распределены примерно по нормальному закону вблизи среднего значения соответствующей совокупности; 2) можно ожидать, что среднее значение и среднеквадратическое (стандартное) отклонение выборки будут примерно равняться среднему значению и среднеквадратическому отклонению совокупности, из которой выборка извлечена; и 3) примерно 68% средних значений выборок будут отстоять от среднего значения соответствующей совокупности на расстояние,

не превышающее одной стандартной ошибки, примерно 95% — на расстояние, не превышающее двух стандартных ошибок, и т. д.

Проще говоря, любая выборка должна быть очень похожа на совокупность, из которой она сформирована. Несмотря на то что все выборки несколько отличаются друг от друга, среднее значение надлежащим образом сформированной выборки довольно редко будет значительно отклоняться от среднего значения генеральной совокупности. Аналогично, можно ожидать, что две выборки, извлеченные из одной и той же совокупности, будут очень похожи друг на друга. Или, если представить ситуацию несколько иначе: две выборки со средними значениями, сильно разняющимися между собой, с наибольшей вероятностью сформированы из разных совокупностей.

Вот краткий пример, который должен быть понятен на интуитивном уровне. Допустим, ваша нулевая (основная) гипотеза гласит, что средний рост профессиональных баскетболистов равен среднему росту остальной части взрослого мужского населения. Вы формируете произвольным образом выборку из 50 профессиональных баскетболистов и выборку из 50 взрослых мужчин-неспортсменов. Допустим, что средний рост членов первой группы (баскетболисты) составляет 6 футов и 7 дюймов, а второй (небаскетболисты) — 5 футов и 10 дюймов (разница — 9 дюймов). Какова вероятность зафиксировать столь большую разницу между значениями среднего роста у этих двух выборок, если бы действительно (как мы предположили) средний рост профессиональных баскетболистов и всего остального взрослого мужского населения страны не отличался? «Нетехнический» ответ: чрезвычайно низкая*.

Базовая методология, использовавшаяся при выполнении исследования аутизма, точно такая же. В упомянутой нами статье сравниваются несколько показателей объема мозга у разных выборок детей. (Измерения выполнялись по методу визуализации с помощью магнитного резонанса у детей в возрасте двух, четырех и пяти лет.) Я сосредоточусь лишь на одном показателе:

* Существуют две возможные альтернативные гипотезы. Первая заключается в том, что профессиональные баскетболисты выше, чем мужское население в целом. Вторая — что средний рост профессиональных баскетболистов отличается от среднего роста мужского населения в целом (при этом не будем забывать о вероятности того, что рост профессиональных баскетболистов может в действительности быть меньшим, чем у некоторых обычных мужчин). Это различие не играет большой роли при выполнении проверки по критерию значимости и вычислении р-значения. Соответствующее объяснение можно найти в более подробных учебниках по статистике, однако это не играет особой роли для нашего обсуждения, имеющего более общий характер.

общем объеме мозга. Нулевая гипотеза исследователей, скорее всего, заключалась в том, что анатомические различия в головном мозге детей-аутистов и здоровых детей отсутствуют. Альтернативная гипотеза — что головной мозг детей-аутистов существенно отличается от головного мозга здоровых детей. Вывод, к которому пришли ученые, по-прежнему оставляет много вопросов, однако указывает, в каком направлении должны проводиться дальнейшие эксперименты.

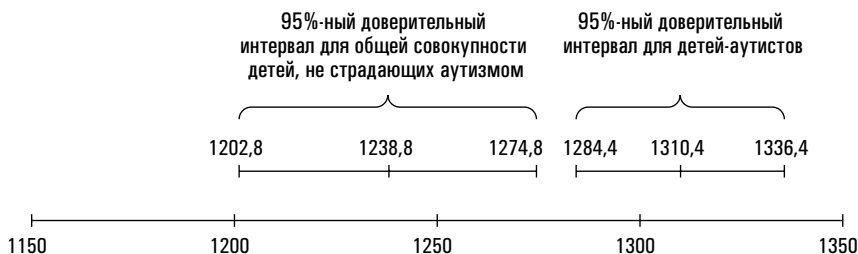
В рассматриваемом нами исследовании средний объем головного мозга детей, страдающих аутизмом, составляет 1310,4 кубических сантиметра; средний объем головного мозга детей в контрольной группе равен 1238,8 кубических сантиметра. Таким образом, разница в среднем объеме головного мозга у этих двух групп составит 71,6 кубических сантиметра. Какова вероятность наблюдения такого результата, если бы на самом деле разницы в среднем объеме головного мозга у детей-аутистов и здоровых детей *во всей совокупности* не было?

Из материала предыдущей главы вы, возможно, помните, как вычислить стандартную ошибку для каждой выборки: s / \sqrt{n} , где s — среднеквадратическое отклонение данной выборки, а n — количество наблюдений. Соответствующие величины приведены в рассматриваемой нами статье. Стандартная ошибка для общего объема головного мозга 59 детей в выборке детей-аутистов составляет 13 кубических сантиметров, а 38 детей в контрольной группе — 18 кубических сантиметров. Согласно центральной предельной теореме, для 95 выборок из 100 среднее значение выборок будет отстоять от истинного среднего значения совокупности на расстояние, не превышающее двух стандартных ошибок (в ту или другую сторону).

Таким образом, на основании нашей выборки можно заключить, что в 95 случаях из 100 интервал 1310,4 кубических сантиметра ± 26 (что равняется двум стандартным ошибкам) будет содержать средний объем головного мозга для *всех* детей, страдающих аутизмом. Это выражение называется доверительным интервалом. Мы можем с 95%-ной уверенностью утверждать, что диапазон от 1284,4 до 1336,4 кубических сантиметра содержит средний общий объем головного мозга для детей-аутистов в их общей совокупности.

Используя ту же методологию, мы можем с 95%-ной уверенностью утверждать, что интервал $1238,8 \pm 36$, или диапазон от 1202,8 до 1274,8 кубических сантиметра, будет включать средний объем головного мозга для здоровых детей в генеральной совокупности.

Да, вас, наверное, утомило обилие числовых показателей. Возможно, вы уже зашвырнули книгу в дальний угол*. Если же еще нет (или раскаялись и возобновили чтение), то должны были обратить внимание на то, что наши доверительные интервалы не перекрываются. *Нижняя граница* 95%-ного доверительного интервала для среднего объема головного мозга детей-аутистов в общей совокупности (1284,4 кубических сантиметра) все же выше, чем *верхняя граница* 95%-ного доверительного интервала для среднего объема головного мозга здоровых детей в общей совокупности (1274,8 кубических сантиметра), что иллюстрируется приведенной ниже диаграммой.



Это первый намек на вероятность существования какой-то анатомической особенности в головном мозге детей, страдающих аутизмом. Однако это всего лишь подсказка. Ведь сделанные заключения основываются на данных, описывающих небольшое число детей (менее 100 человек). Нельзя исключать вариант, что мы имеем дело с какими-то аномальными выборками.

Одна финальная статистическая процедура способна внести ясность в ситуацию. Если бы статистика была одним из олимпийских видов спорта, например фигурным катанием, то это было бы последним видом программы выступлений, после которой преданные болельщики бросают на лед букеты цветов. Мы можем точно вычислить вероятность наблюдения по меньшей мере столь же значительной разницы средних значений (1310,4 кубических сантиметра в сравнении с 1238,8 кубическими сантиметрами), если действительно между объемом головного мозга детей-аутистов и всех остальных детей в общей совокупности никакого отличия нет. Мы можем найти р-значение для наблюдаемой разницы в средних значениях.

Чтобы вы прямо сейчас не зашвырнули эту книгу в самый дальний угол комнаты, соответствующая формула будет приведена в приложении.

* Сознаюсь, что однажды в отчаянии я изорвал одну книгу по статистике.

Впрочем, на интуитивном уровне все должно быть достаточно понятно. Если мы извлекаем две большие выборки из одной и той же совокупности, то можно ожидать, что их средние значения будут очень близки между собой. Более того, в идеале они должны быть одинаковы. Если бы, например, средний рост выбранных мною 100 баскетболистов из НБА составлял 6 футов и 7 дюймов, то я был бы вправе ожидать, что в какой-нибудь другой случайной выборке 100 баскетболистов из НБА средний рост игроков будет близок к 6 футам и 7 дюймам. Ладно, возможно, средний рост игроков в этих двух выборках будет отличаться на один-два дюйма. Однако вероятность того, что он будет разниться на 4 дюйма, окажется низкой, а того, что на 6 или 8 дюймов, будет еще ниже. Мы можем вычислить стандартную ошибку для разности между средними значениями двух выборок, которая может служить мерой ожидаемого разброса (но в среднем) при вычитании среднего значения одной выборки из среднего значения другой. (Как указывалось ранее, соответствующая формула приводится в приложении к этой главе.) Важно то, что мы можем использовать эту стандартную ошибку для определения вероятности того, что две выборки сформированы из одной и той же совокупности. Принцип действия этого механизма таков.

1. Если две выборки сформированы из одной и той же совокупности, мы имеем все основания предполагать, что разница между их средними значениями равна нулю.
2. Согласно центральной предельной теореме, в повторных выборках разница между этими двумя средними значениями будет распределена примерно по нормальному закону. (Итак, вы уже влюбились в центральную предельную теорему или еще нет?)
3. Если обе выборки действительно сформированы из одной и той же совокупности, то приблизительно в 68 случаях из 100 разница между их средними значениями будет отличаться от нуля не более чем на одну стандартную ошибку, в 95 случаях из 100 — не более чем на две стандартные ошибки, а примерно в 99,7 случая из 100 — не более чем на три стандартные ошибки. Так вот что побудило исследователей сделать вывод, о котором мы узнали из статьи об аутизме, опубликованной в *The Wall Street Journal*.

Как указывалось ранее, разница в среднем объеме головного мозга между выборкой детей-аутистов и контрольной группой составляет 71,6 кубичес-

ких сантиметра. Стандартная ошибка для этой разницы — 22,7. Это означает, что разница между средними значениями двух выборок больше нуля на три стандартные ошибки. Можно ожидать, что столь (или еще более) экстремальный исход окажется возможным лишь в 2 случаях из 1000, если эти выборки сформированы из одной и той же совокупности.

Как отмечалось выше, авторы статьи, опубликованной в Archives of General Psychiatry, сообщают о *p*-значении, равном 0,002. Теперь вы понимаете, откуда взялась эта величина.

Несмотря на все достоинства статистического вывода, он не лишен недостатков. И они становятся очевидны из примера, приведенного в начале главы. Если вы помните, в нем речь шла о моем преподавателе статистики, заподозрившем меня в обмане. Процесс статистического вывода основывается на понятии вероятности, а вовсе не на абсолютной и не вызывающей ни малейшего сомнения достоверности. Таким образом, когда речь идет о проверке той или иной гипотезы, мы имеем дело с фундаментальной дилеммой.

Эта статистическая реальность заявила о себе во весь голос в 2011 году, когда Journal of Personality and Social Psychology готовился опубликовать одну научную статью, которая на первый взгляд ничем особенным не выделялась⁶. Некий профессор Корнелльского университета предложил нулевую гипотезу, а затем, на основе полученных им экспериментальных результатов, отверг ее с уровнем значимости 0,05. Этот результат произвел настоящий фурор в научных кругах, а также широко освещался в ведущих средствах массовой информации, таких как The New York Times.

Достаточно сказать, что статьи в Journal of Personality and Social Psychology обычно не привлекают к себе внимания СМИ. Что же вызвало на сей раз столь повышенный интерес прессы? Упомянутый мной исследователь проверял способность человека к экстрасенсорному восприятию (Extra Sensory Perception — ESP). Основная гипотеза ученого отрицала существование ESP; альтернативная подтверждала. Чтобы изучить вопрос, исследователь предложил большой выборке людей, которых он пригласил поучаствовать в эксперименте, рассмотреть два «занавеса», представленных на экране монитора. Компьютерная программа случайным образом помещала некое эротическое изображение то за одним, то за другим «занавесом». В ходе повторяющихся попыток испытуемым удалось правильно выбрать «занавес», за которым скрывалось эротическое изображение, в 53 случаях из 100, тогда как, согласно теории вероятностей, это должно происходить лишь в 50 случаях из 100.

Достаточно большой размер выборки позволил ученому отклонить нулевую гипотезу и принять альтернативную. Решение опубликовать статью об этом эксперименте подверглось широкой критике на том основании, что какое-то одно статистически значимое событие вполне может оказаться следствием чистой случайности, особенно при отсутствии каких-либо других свидетельств, подтверждающих или даже объясняющих полученный результат. Статья в *The New York Times* так резюмировала критические высказывания: «Утверждения, которые бросают вызов практически всем законам науки, по определению являются экстраординарными и, как правило, требуют экстраординарных, неопровержимых доказательств. Нежелание учитывать это обстоятельство — как того требует общепринятый научный метод — делает результаты многих исследований гораздо значимее, чем они есть на самом деле».

Одним из достойных ответов на подобную критику был бы выбор более жесткого порога для определения статистической значимости, например 0,001*. Однако это порождает собственные проблемы. Выбор надлежащего уровня значимости в любом случае предполагает определенный компромисс.

Если наше «бремя доказательства», которое позволило бы отвергнуть основную гипотезу, будет чересчур низким (например 0,1), то нам придется периодически отклонять нулевую гипотезу, хотя на самом деле она верна (я подозреваю, что именно так и произошло при исследовании ESP). На языке статистики это называется *ошибкой первого рода*. Рассмотрим пример из судебной практики в США, где нулевая гипотеза заключается в том, что подсудимый (ответчик) невиновен, а порогом, когда она отвергается, является «критерий доказанности при отсутствии обоснованного сомнения» (то есть подсудимый признается виновным при отсутствии обоснованного сомнения в его невиновности). Допустим, мы решили ослабить этот порог, обозначив его, например, как «сильное подозрение, что подсудимый все же совершил данное преступление». Это должно гарантировать, что за решеткой окажется большее число настоящих преступников — а вместе с ними и большее число ни в чем не повинных людей. В статистическом контексте это эквивалентно использованию относительно низкого уровня значимости (например 0,1).

Ладно, «в 1 случае из 10» — не такое уж маловероятное событие. Рассмотрим эту проблему в контексте утверждения нового лекарства от рака.

* Еще одним ответом могла бы стать попытка повторить полученные результаты в дополнительных исследованиях.

На каждые десять препаратов, которые мы одобряем с этим относительно низким «бременем статистического доказательства», один на практике оказывается неэффективным, а в процессе тестирования показывает обнадеживающие результаты лишь по чистой случайности. (Или, если воспользоваться примером из судебной практики, из каждых десяти подсудимых, признанных виновными, один фактически невиновен.) Ошибка первого рода заключается в ошибочном отказе от основной гипотезы. Иногда это называют «ложным позитивом», хотя употребление такого термина кажется несколько парадоксальным. Вот один способ примириться с подобным жаргоном. Когда вы приходите к врачу, чтобы выяснить, не страдаете ли вы какой болезнью, основная гипотеза заключается в том, что вы ею не страдаете. Если результаты анализов позволяют отвергнуть нулевую гипотезу, то врач говорит, что у вас положительный результат анализов. А если у вас положительный результат анализов, хотя в действительности вы не больны, то это и есть случай «ложного позитива».

Как бы то ни было, чем ниже «статистическое бремя» для отклонения нулевой гипотезы, тем выше вероятность «ложного позитива». Очевидно, что мы предпочли бы не утверждать неэффективные лекарства от рака и не отправлять невинных людей за решетку.

Но здесь есть один нюанс. Чем выше порог для отказа от нулевой гипотезы, тем вероятнее, что нам не удастся отвергнуть ту нулевую гипотезу, которую на самом деле следовало было отвергнуть. Если бы нам потребовалось не менее пяти свидетелей, чтобы признать виновным каждого обвиняемого, то на свободе оказалось бы немалое число настоящих преступников. (Разумеется, при этом за решетку не угодили бы многие невинные люди.) Если при клинических испытаниях всех новых лекарств от рака мы примем уровень значимости 0,001, то мы действительно минимизируем утверждение неэффективных препаратов. (В этом случае будет лишь 1 шанс из 1000 ошибочно отвергнуть нулевую гипотезу, которая заключается в том, что испытываемое лекарство эффективно не более чем плацебо.) Однако при этом возникает риск не допустить на рынок много эффективных лекарств, поскольку мы установили очень высокую планку для их утверждения. На языке статистики это называется ошибкой второго рода, или «ложным негативом»*.

* Ошибка второго рода — это вероятность принятия нулевой гипотезы тогда, когда она неверна. *Прим. ред.*

Какая же из двух ошибок хуже? Это зависит от конкретных обстоятельств. Самое важное — что вы признаете необходимость компромисса. В статистике «бесплатный завтрак» невозможен. Рассмотрим перечисленные ниже нестатистические ситуации, каждая из которых предполагает достижение определенного компромисса между ошибками первого и второго рода.

1. **Спам-фильтры.** Основная гипотеза: любое конкретное сообщение, приходящее по электронной почте, не спам. Ваш спам-фильтр отыскивает признаки, которые могут использоваться для отказа от нулевой гипотезы для того или иного конкретного сообщения, например огромные списки рассылки или наличие фраз типа «удлинение пениса». Ошибка первого рода предполагает отбраковку сообщения, которое на самом деле не является спамом («ложный позитив»). Ошибка второго рода предполагает пропуск спама через фильтр и его попадание в ваш почтовый ящик («ложный негатив»). Сравнивая последствия от потери важного сообщения и незначительное раздражение, вызванное получением совершенно не интересующего вас письма, содержащего, скажем, рекламу БАДов, большинство людей, скорее всего, предпочтут терпеть неудобства, обусловленные ошибками второго рода. Оптимально разработанный спам-фильтр должен требовать относительно высокой степени определенности, прежде чем отвергнуть нулевую гипотезу и заблокировать соответствующее сообщение.
2. **Проверка на наличие раковых заболеваний.** Существуют многочисленные тесты для раннего выявления раковых заболеваний, например маммография (рак молочной железы), ПСА-тест (рак простаты) и даже магнитно-резонансная визуализация (МРТ) всего тела для выявления всего, что может вызывать подозрения. Основная гипотеза для каждого, кто проходит такое обследование, заключается в том, что он не болен раком. Проверка на наличие раковых заболеваний используется для того, чтобы отвергнуть нулевую гипотезу, если результаты тестирования вызывают подозрения. Соответствующее предположение всегда исходит из того, что ошибка первого рода («ложный позитив», что в конечном счете означает отсутствие заболевания) безусловно предпочтительнее ошибки второго рода («ложный негатив», который означает, что диагностирование не выявило

заболевания, которое на самом деле имеется). Проверка на наличие раковых заболеваний является полной противоположностью примеру со спам-фильтром. Врачи и пациенты готовы мириться с умеренным количеством ошибок первого рода («ложный позитив»), чтобы избежать вероятности появления ошибок второго рода («ложный негатив»), когда пациенту не диагностируется раковое заболевание, хотя в действительности он болен. Впрочем, в последнее время специалисты в области политики охраны здоровья подвергают сомнению такой подход из-за высоких издержек и побочных эффектов, связанных с «ложными позитивами».

3. **Поимка террористов.** В этой ситуации неприемлема ни ошибка первого, ни ошибка второго рода. Именно поэтому в обществе продолжают дебаты, связанные с поиском подходящего баланса между борьбой с терроризмом и защитой гражданских прав. Основная гипотеза в данном случае заключается в том, что человек не террорист. Как и в обычном уголовном контексте, нам не хотелось бы совершать ошибки первого рода и отправлять невиновных в тюрьму Гуантанамо. Однако в мире, где накоплено большое количество оружия массового поражения, даже одного террориста опасно оставлять на свободе (ошибка второго рода), поскольку это может повлечь за собой поистине катастрофические последствия. Именно поэтому — нравится вам это или нет — власти Соединенных Штатов удерживают в Гуантанамо людей, подозреваемых в терроризме, основываясь при этом даже на меньшей доказательной базе, чем могло бы потребоваться для вынесения им обвинительного приговора в обычном уголовном суде.

Статистический вывод — это не волшебная палочка и отнюдь не безошибочный метод. Тем не менее это замечательный инструмент для осмысления мира. Мы можем глубже понять многие явления нашей жизни лишь путем нахождения им наиболее вероятного объяснения. Многие из нас делают это постоянно (например, мы говорим: «Мне кажется, этот молодой человек, развалившийся на полу в окружении множества пустых банок из-под пива, хватил лишку», а не «Мне кажется, что этого молодого человека, развалившегося на полу в окружении множества пустых банок из-под пива, отравили террористы»).

Статистический вывод лишь формализует процесс.

Приложение к главе 9

Вычисление стандартной ошибки для разности средних значений

Формула для сравнения двух средних значений

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

где \bar{x} — среднее значение выборки x

\bar{y} — среднее значение выборки y

s_x — среднеквадратическое отклонение выборки x

s_y — среднеквадратическое отклонение выборки y

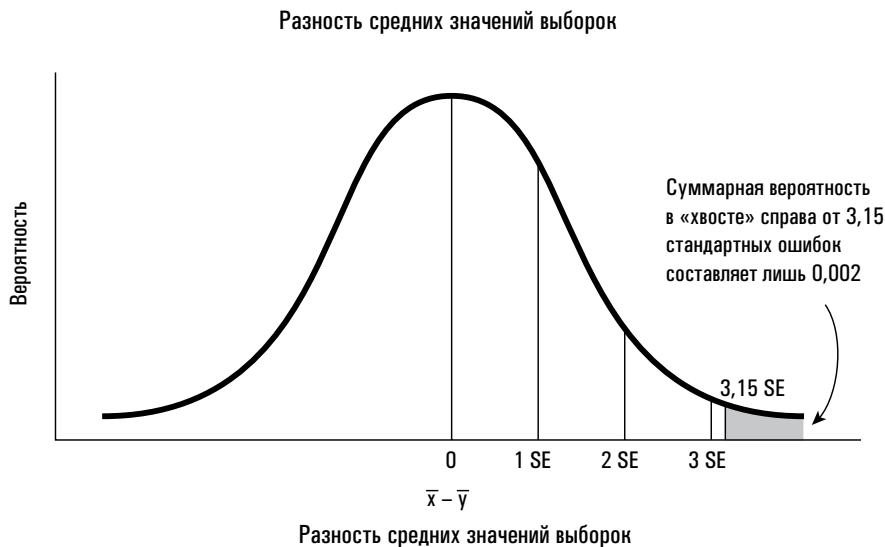
n_x — количество наблюдений в выборке x

n_y — количество наблюдений в выборке y

(В числителе вычисляется разность двух средних значений; в знаменателе — стандартная ошибка для разности двух средних значений разных выборок.)

Нулевая гипотеза: средние значения этих двух выборок одинаковы. Приведенная выше формула вычисляет наблюдаемую разность средних значений относительно величины стандартной ошибки для разности средних значений. Как и прежде, мы предполагаем, что имеем дело с нормальным распределением. Если средние значения исходной совокупности действительно одинаковы, то можно ожидать, что разность средних значений двух выборок окажется меньше одной стандартной ошибки в 68 случаях из 100 и меньше двух стандартных ошибок в 95 случаях из 100 (и т. д.).

В приведенном примере с аутизмом разность средних значений двух выборок составляла 71,6 кубических сантиметра при стандартной ошибке 22,7. Отношение этой наблюдаемой разности равняется 3,15; это означает, что средние значения двух указанных выборок отстоят друг от друга более чем на три стандартные ошибки. Как уже отмечалось, вероятность получения выборок со столь различающимися средними значениями в случае, если средние значения исходных совокупностей одинаковы, чрезвычайно низкая. Точнее говоря, вероятность наблюдения разности средних значений, составляющей не менее 3,15 среднеквадратических ошибок, равняется 0,002.



Проверка гипотез с одно- и двусторонним критерием

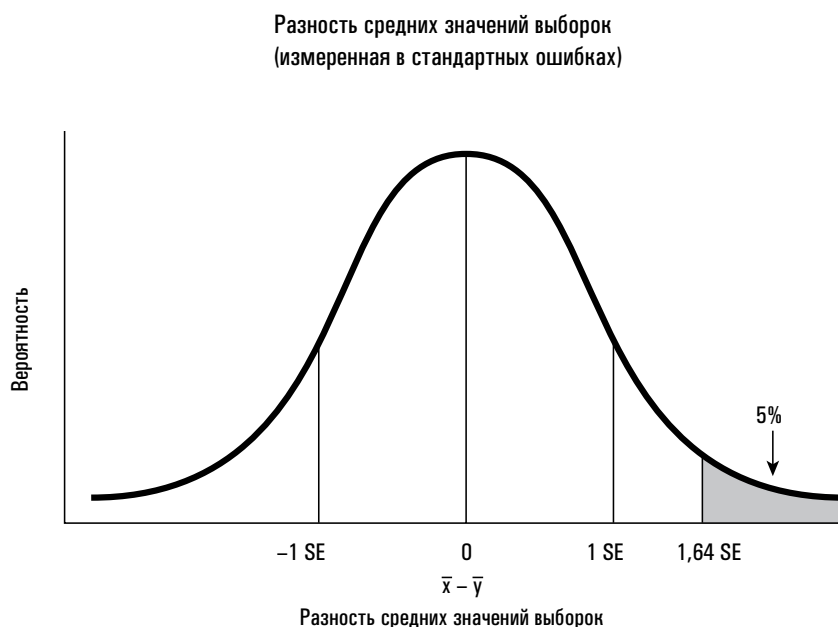
Когда мы использовали пример со сравнением роста профессиональных баскетболистов с ростом мужского населения в целом, я сознательно упустил одну маленькую деталь. Наша нулевая гипотеза заключалась в том, что рост профессиональных баскетболистов такой же, как средний рост мужского населения в целом. Однако я не указал, что в действительности у нас есть две возможные альтернативные гипотезы.

Одна заключается в том, что средний рост профессиональных баскетболистов отличается от среднего роста мужского населения: они могут быть выше или ниже, чем другие мужчины в совокупности. Именно таким подходом вы воспользовались, когда проникли в автобус, угнанный террористами, и определили вес пассажиров, чтобы выяснить, являются ли они участниками исследования Americans' Changing Lives. Вы могли отвергнуть нулевую гипотезу, что пассажиры угнанного автобуса являются участниками этого исследования, если бы их средний вес был значительно больше, чем средний вес участников исследования, *или* значительно меньше (как и оказалось на самом деле). Вторая альтернативная гипотеза заключается в том, что средний рост профессиональных баскетболистов превышает средний рост остального мужского населения. В этом случае нам пригодится обычный жизненный опыт, который подсказывает, что рост профессиональных

баскетболистов не может быть меньше, чем средний рост остального мужского населения. Различие между этими двумя альтернативными гипотезами определяет, выполняем ли мы проверку гипотез с односторонним или двусторонним критерием.

В обоих случаях мы исходим из того, что будем выполнять проверку значимости на уровне 0,05. Мы отвергнем нулевую гипотезу, если будем наблюдать разницу в росте между указанными двумя выборками хотя бы в 5 случаях из 100, притом что рост в обеих выборках действительно одинаков. Пока все идет нормально!

Однако с этого момента появляются небольшие нюансы. Когда альтернативная гипотеза гласит, что средний рост профессиональных баскетболистов превышает средний рост остального мужского населения, мы будем выполнять *проверку гипотез с односторонним критерием*. Мы измерим разницу среднего роста между выборкой профессиональных баскетболистов и выборкой обычных лиц мужского пола. Мы знаем, что в случае, если наша нулевая гипотеза верна, мы будем наблюдать разницу не меньше 1,64 стандартной ошибки лишь в 5 случаях из 100. Мы отвергнем нулевую гипотезу, если полученный результат попадает в диапазон, указанный на приведенном ниже графике.

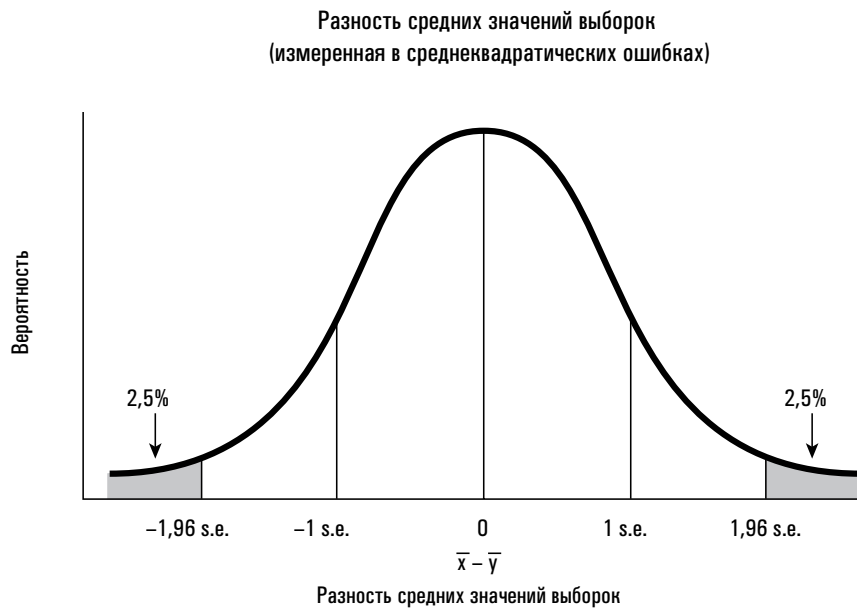


А теперь вернемся к другой альтернативной гипотезе, которая заключается в том, что средний рост профессиональных баскетболистов может быть больше или меньше среднего роста других мужчин в совокупности. Наш общий подход остается неизменным. Как и прежде, мы отвергнем нулевую гипотезу, гласящую, что рост профессиональных баскетболистов такой же, как средний рост мужского населения в целом, если получим результат, который будет наблюдаться не чаще чем в 5 случаях из 100, притом что действительно разницы в росте между этими двумя выборками никакой нет. Различие, однако, состоит в том, что на сей раз мы должны допустить и вероятность того, что рост профессиональных баскетболистов меньше среднего роста других мужчин в совокупности. Таким образом, мы отвергнем основную гипотезу, если средний рост выборки профессиональных баскетболистов окажется значительно больше *или меньше* среднего роста выборки «обычных» мужчин. Для этого нам понадобится выполнять *проверку гипотез с двусторонним критерием*. Граница, по достижении которой мы отклоняем нулевую гипотезу, будет другой, поскольку на сей раз мы должны учитывать вероятность большой разницы в средних значениях выборок в обоих направлениях: положительном и отрицательном. Точнее говоря, диапазон, в котором мы отвергнем нулевую гипотезу, разделится между двумя «хвостами». Мы по-прежнему отвергнем основную гипотезу, если получим исход, встречающийся не более чем в 5% случаев, если рост профессиональных баскетболистов окажется таким же, как у других мужчин в совокупности; правда, на этот раз существуют два разных варианта, при которых мы можем отказаться от основной гипотезы.

Мы отклоним нулевую гипотезу, если средний рост выборки профессиональных баскетболистов окажется настолько больше среднего роста выборки «обычных» мужчин, что мы наблюдали бы такой исход лишь в *2,5 случаях из 100*, если средний рост профессиональных баскетболистов действительно не отличается от среднего роста «обычных» мужчин.

Кроме того, мы отвергнем нулевую гипотезу, если средний рост выборки профессиональных баскетболистов окажется настолько меньше среднего роста «обычных» мужчин, что мы наблюдали бы такой исход лишь в *2,5 случаях из 100*, если средний рост профессиональных баскетболистов действительно не отличается от среднего роста «обычных» мужчин.

В совокупности эти две ситуации дают 5%, как следует из приведенного ниже графика.



Чтобы решить, какой вариант проверки гипотез — с одно- или двусторонним критерием — больше подходит для того или иного анализа, понадобится рассудительность.

Откуда нам известно, что 64% американцев поддерживают смертную казнь (ошибка выборки $\pm 3\%$)

В конце 2011 года в газете The New York Times вышла передовая статья, в которой сообщалось, что «страну охватило чувство сильной тревоги и неуверенности в будущем»¹. Авторы публикации всесторонне исследовали психологическое состояние американцев, выяснив общественное мнение по широкому кругу вопросов, от оценки эффективности администрации Обамы до отношения населения к распределению общественного богатства страны. Ниже приведена небольшая выдержка мнений американцев, которые высказывались осенью 2011 года.

- Около 89% американцев (шокирующий показатель!) заявили, что не доверяют правительству (самый высокий уровень недоверия к власти за все время проведения подобных опросов).
- Две трети опрошенных считают, что общественное богатство страны должно распределяться среди граждан более равномерно.
- Сорок три процента жителей страны сказали, что в целом согласны со взглядами участников движения Occupy Wall Street (довольно аморфное протестное движение, стартовавшее в Нью-Йорке вблизи Уолл-стрит и впоследствии охватившее другие города страны)*.

* Согласно сайту движения Occupy Wall Street, это народное движение, которое возникло 17 сентября 2011 года в Либерти-сквер, финансовый округ Манхэттена, и распространилось на более чем 100 городов Соединенных Штатов, а также инициировало акции протеста в более чем 1500 городах по всему миру. Occupy Wall Street выступает против засилья крупных банков и транснациональных корпораций, оказывающих разлагающее влияние на демократический процесс, и против роли Уолл-стрит в создании экономического коллапса, который породил тяжелейшую рецессию за все время существования человечества. Это движение вызвано народными волнениями в Египте и Тунисе и ставит своей задачей показать, как 1% самых богатых людей диктуют правила несправедливой глобальной экономики, которая становится непреодолимым препятствием на нашем пути в будущее.

Чуть больше опрошенных, 46%, заявили, что взгляды участников движения Оссуру Wall Street «в целом отражают взгляды большинства американцев».

- Сорок шесть процентов американцев одобрили деятельность Барака Обамы на посту президента США — и такие же 46% выразили неудовлетворенность тем, как он справляется со своими обязанностями.
- Лишь 9% жителей страны поддерживают деятельность Конгресса США.
- Несмотря на то что президентские праймериз должны были начаться только через два месяца, примерно 80% избирателей, во время праймериз обычно голосующих за республиканцев, заявляли, что «еще слишком рано говорить о том, кого именно они будут поддерживать».

Впечатляющие данные, приведенные выше, давали политическим аналитикам обильную пищу для изучения настроений американцев за год до президентских выборов. Правда, возникает резонный вопрос: откуда все это известно? Как удалось сделать столь далекоидущие выводы о настроениях сотен миллионов взрослых американцев? И почему мы должны верить, что эти выводы верны?

Ответ очевиден: это результат опросов общественного мнения. К тому же в приведенном выше примере их проводили The New York Times и CBS News. (То обстоятельство, что две конкурирующие новостные организации совместно реализовывали проект, подобный этому, является первым указанием на то, что такие исследования довольно затратны.) Я не сомневаюсь, что вы знакомы с результатами этих опросов. Возможно, не столь явно выраженным кажется тот факт, что методология их проведения представляет собой всего лишь еще одну форму статистического вывода. Опрос общественного мнения — это получение выводов о настроениях определенной совокупности людей, основанных на мнениях, высказанных некоторой выборкой, сформированной из генеральной совокупности.

Эффективность опросов обуславливается использованием того же источника, что и в предыдущих примерах с выборками, — центральной предельной теоремы. Если мы опрашиваем достаточно большую репрезентативную выборку избирателей (или любую другую группу), то у нас есть все основания полагать, что она будет очень похожа на совокупность, из которой извлечена. Если ровно половина взрослых американцев не одобряют

однополые браки, то мы вполне можем рассчитывать, что в репрезентативной выборке из 1000 американцев примерно половина ее членов также выступят против однополых браков.

И наоборот (что гораздо важнее для проведения опросов общественного мнения), если в репрезентативной выборке из 1000 американцев удалось выявить определенные настроения, например 46% недовольны деятельностью Обамы на посту президента США, то это дает веский повод думать, что среди населения в целом — примерно в такой же пропорции — также присутствуют подобные настроения. Вообще говоря, мы можем рассчитать вероятность того, что результаты, полученные с помощью нашей выборки, будут значительно отклоняться от доминирующих настроений в обществе. Когда вы читаете, что статистическая погрешность составляет $\pm 3\%$, в действительности речь идет о том же 95%-ном доверительном интервале, который мы вычисляли в предыдущей главе. Наш «95%-ный доверительный интервал» означает, что если бы мы провели 100 разных опросов общественного мнения в выборках, сформированных из одной и той же совокупности, то, предположительно, полученные ответы в 95 из 100 опросов отличались бы (в ту или другую сторону) от истинных настроений этой совокупности не более чем на 3%. В контексте вопроса об оценке деятельности Обамы на посту президента США, фигурировавшего в опросе, проводившемся The New York Times и CBS News, мы могли на 95% быть уверены, что истинная доля американцев, не одобряющих его деятельность, находится в диапазоне $46 \pm 3\%$, то есть от 43% до 49%. Если вы прочитаете сопроводительный текст к опросу, набранный мелким шрифтом (между прочим, я бы настоятельно рекомендовал вам всегда это делать), то увидите, что его смысл заключается в следующем: «Теоретически в 19 случаях из 20 результаты, базирующиеся на таких выборках, будут отличаться не более чем на 3% (в ту или другую сторону) от результатов, которые были бы получены в ходе опроса *всех* взрослых американцев».

Одно из фундаментальных отличий опросов общественного мнения от других форм использования метода выборки состоит в том, что интересующим нас статистическим показателем выборки будет не среднее значение (например, 187 фунтов веса), а некий процент или доля (например, 47% избирателей, или 0,47). В остальном же процессы идентичны. При наличии крупной репрезентативной выборки (опрос общественного мнения) можно ожидать, что доля респондентов, охваченных определенными настроениями

(например, 9% респондентов в этой выборке одобряют деятельность Конгресса США), примерно равна доле американских избирателей в целом, испытывающих аналогичные настроения. Это в принципе ничем не отличается от предположения о том, что средний вес выборки из 1000 мужчин-американцев должен примерно равняться среднему весу всех мужчин-американцев. Тем не менее мы допускаем вероятность какого-то разброса от выборки к выборке доли тех, кто одобряет деятельность Конгресса США, точно так же как у нас есть все основания ожидать какого-то разброса в средних значениях веса при использовании разных произвольных выборок из 1000 мужчин-американцев. Если бы The New York Times и CBS News провели еще один опрос — задавая те же вопросы другой выборке из 1000 взрослых американцев, — то очень маловероятно, что его результаты полностью бы совпали с результатами первого опроса. С другой стороны, можно ожидать, что ответы, полученные в ходе первого и второго опросов, будут незначительно отличаться между собой. (Воспользуюсь метафорой, к которой уже прибегал в этой книге: если вы попробуете ложку супа из кастрюли, затем хорошенько перемешаете суп и попробуете ложку супа еще раз, то его вкус, скорее всего, покажется вам примерно таким же.) Стандартная ошибка — вот что указывает на то, какого разброса результатов от выборки к выборке (в данном случае от опроса к опросу) мы можем ожидать.

Формула расчета стандартной ошибки в случае, когда речь идет о процентной величине или доле, несколько отличается от формулы, с которой вы уже познакомились; впрочем, интуитивные соображения остаются такими же. Для любой произвольной выборки, сформированной надлежащим образом, стандартная ошибка равняется $\sqrt{p(1-p)/n}$, где p — доля респондентов, выражающих определенную точку зрения, $(1-p)$ — доля респондентов, имеющих противоположную точку зрения, а n — общее количество респондентов в выборке. Обратите внимание, что стандартная ошибка будет уменьшаться с увеличением размера выборки, поскольку n находится в знаменателе. Стандартная ошибка также будет уменьшаться с увеличением разности между p и $(1-p)$. Например, стандартная ошибка будет меньше в случае опроса, в ходе которого 95% респондентов выражают определенную точку зрения, чем в случае опроса, в котором мнения респондентов разделяются примерно 50 на 50. Это чисто математический результат, поскольку $0,05 \times 0,95 = 0,047$, тогда как $0,5 \times 0,5 = 0,25$; меньшая величина в числителе формулы ведет к уменьшению стандартной ошибки.

Допустим, что в результате проведения простого экзитпола репрезентативной выборки из 500 избирателей выяснилось, что 53% проголосовали за кандидата от республиканцев, 45% — за кандидата от демократов и 2% поддержали независимого кандидата. Если использовать кандидата от республиканцев как интересующую нас долю, то стандартная ошибка для этого экзитпола составит: $\sqrt{(0,53)(1 - 0,53)/500} = \sqrt{(0,53)(0,47)/500} = \sqrt{0,25/500} = \sqrt{0,0005} = 0,02236$.

Для упрощения округлим стандартную ошибку для этого экзитпола до 0,02. Пока это всего лишь некое число. Подумаем, почему оно так важно для нас. Предположим, избирательные участки только что закрылись, и вашему работодателю (коим является некая телекомпания) не терпится объявить победителя выборов еще до того, как станут известны официальные результаты. Вам как человеку, уже прочитавшему две трети этой книги, поручено заниматься обработкой данных, полученных в ходе экзитпола. Ваш начальник желает знать, можно ли на их основании назвать победителя выборов.

Вы объясняете, что ответ на этот вопрос зависит от того, насколько уверенной хочет быть телекомпания в правильности своего заявления — или, точнее говоря, какой риск она готова принять на себя, если оно окажется ошибочным. Вспомните: стандартная ошибка дает нам представление о том, как часто можно ожидать, что доля в выборке (экзитпол) окажется достаточно близкой к истинной доле в совокупности (результат голосования). Нам известно, что примерно в 68% случаев мы можем ожидать, что доля в выборке — в данном случае 53% избирателей, которые утверждают, что проголосовали за кандидата от республиканцев, — отстоит от истинного окончательного результата голосования не более чем на одну стандартную ошибку. Таким образом, вы говорите начальнику «с 68%-ной уверенностью», что ваша выборка, которая показывает, что кандидат от республиканцев получил голоса 53% избирателей $\pm 2\%$, то есть между 51 и 55%, соответствует истинному достигнутому им результату. Между тем, согласно тому же экзитполу, за кандидата от демократов отдали голоса 45% избирателей. Если предположить, что итог голосования за кандидата от демократов имеет ту же стандартную ошибку (упрощение, суть которого я объясню ниже), то с 68%-ной уверенностью можно утверждать, что наша выборка (экзитпол), которая показывает, что за кандидата от демократов проголосовали 45% избирателей $\pm 2\%$, то есть между 43 и 47%, заключает в себе истинный результат этого кандидата. Согласно этому подсчету, победителем становится кандидат от республиканцев.

Группа графического дизайна бросается строить красочную трехмерную диаграмму, чтобы вы могли отобразить ее на экранах ваших телевизоров:

Представитель Республиканской партии	53%
Представитель Демократической партии	45%
Независимый кандидат	2%
(Предел погрешности	2%)

Поначалу ваш босс приходит в восторг — главным образом потому, что диаграмма представлена в трехмерном виде, насыщена яркими красками и даже может вращаться на экране вокруг вертикальной оси. Однако когда вы объясняете, что примерно в 68 случаях из 100 результаты экзитпола будут отличаться от действительных результатов выборов не более чем на одну стандартную ошибку, ваш начальник, которому уже не раз приходилось посещать курсы аутотренинга и управления негативными эмоциями, указывает на совершенно очевидную вещь: в 32 случаях из 100 результаты экзитпола будут отличаться от действительных результатов выборов *более* чем на одну стандартную ошибку. И что тогда?

Вы объясняете, что есть два варианта: 1) кандидат от республиканцев мог получить *даже больше* голосов, чем предсказывал экзитпол, тогда все равно вы назвали бы победителя правильно; 2) но существует достаточно высокая вероятность того, что кандидат от демократов набрал гораздо больше голосов, чем предсказывал экзитпол; в этом случае ваша восхитительная красочная вращающаяся трехмерная диаграмма объявит победителя неправильно.

Босс запускает чашкой с кофе в стену, из чего вы делаете вывод, что посещение курсов аутотренинга и управления негативными эмоциями не пошло ему на пользу. Между тем, начальник продолжает бушевать: «Как, черт бы вас побрал, мы можем быть уверены в правильности результата, показанного на вашей ...ной диаграмме?»

Понимая кое-что в статистике, вы указываете ему, что не можете быть уверены в каком-либо результате до тех пор, пока не будут подсчитаны все голоса. И предлагаете в качестве критерия уверенности воспользоваться 95-процентным доверительным интервалом. В данном случае ваша восхитительная красочная вращающаяся 3D-диаграмма предскажет победителя неправильно в среднем лишь в 5 случаях из 100.

Начальник закуривает сигарету и пытается успокоиться. Вы решаете не напоминать ему о запрете курения на рабочем месте, несмотря на уча-

стившиеся в последнее время случаи пожаров в офисах, однако все же отваживаетесь поделиться кое-какими плохими новостями: единственный способ, позволяющий вашей телекомпании повысить уверенность в результатах экзитпола, — расширить предел погрешности, но тогда однозначно назвать победителя выборов будет невозможно. После этого вы показываете начальнику новую 3D-диаграмму:

Представитель Республиканской партии	53%
Представитель Демократической партии	45%
Независимый кандидат	2%
(Предел погрешности	4%)

Из центральной предельной теоремы вам известно, что приблизительно 95% пропорций выборки будут отстоять от истинной пропорции доли голосов совокупности на расстоянии, не превышающем *двух стандартных ошибок* (в данном случае 4%). Таким образом, если мы хотим обеспечить большую уверенность в результатах экзитпола, то нам придется умерить свои амбиции в том, что касается точности прогноза. Как следует из приведенной выше пропорции доли голосов (к сожалению, мы не можем показать здесь соответствующую красочную вращающуюся 3D-диаграмму), ваша телекомпания может, при 95%-ном доверительном уровне, объявить о том, что кандидат от республиканцев получил 53% голосов избирателей $\pm 4\%$, то есть между 49 и 57% голосов избирателей, а кандидат от демократов — $45\% \pm 4\%$, то есть между 41 и 49% голосов избирателей.

Правда, теперь вы сталкиваетесь с новой проблемой. При 95%-ном доверительном уровне вы не можете отвергнуть вероятность того, что каждый из кандидатов мог набрать по 49% голосов избирателей. Это неизбежный компромисс; единственная возможность обрести большую уверенность в том, что результаты вашего экзитпола будут соответствовать истинным результатам выборов *без использования новых данных*, — обуздать свои амбиции относительно точности прогнозов. Подумайте об этом вне статистического контекста. Допустим, вы говорите приятелю, что «почти не сомневаетесь» в том, что Томас Джефферсон был третьим или четвертым президентом США. Каким образом вы можете обрести большую уверенность в своих исторических познаниях? Снизив категоричность утверждений. Можно, например, сказать, что вы «абсолютно уверены» в том, что Томас Джефферсон был одним из первых пяти президентов США.

Ваш начальник предлагает вам заказать пиццу и быть готовым к тому, что придется поработать вечером (или даже всю ночь). На этот раз статистические боги оказываются к вам милостивы. Вам на стол кладут данные второго экзитпола, для проведения которого использовалась выборка из 2000 избирателей. Его результаты таковы: кандидат-республиканец — 52% голосов, кандидат-демократ — 45% голосов, независимый кандидат — 3% голосов. На этот раз ваш босс совершенно взбешен, поскольку эти данные показывают, что разрыв между кандидатами сократился, а это еще больше затрудняет своевременное предсказание итогов голосования. Но не нужно спешить с выводами! Вы указываете (стараясь сохранять присутствие духа), что размер второй выборки (2000) в четыре раза больше первой, которая использовалась при проведении первого экзитпола. Таким образом, стандартная ошибка существенно уменьшилась. Новая стандартная ошибка для кандидата от республиканцев равняется $\sqrt{0,52(0,48)/2000}$, что составляет 0,01.

Если вашего начальника по-прежнему устраивает 95%-ный доверительный интервал, то вы можете объявить победителем кандидата от республиканцев. С учетом вашей новой стандартной ошибки 0,01 95%-ные доверительные интервалы для кандидатов таковы: кандидат-республиканец: 52 ± 2 , или между 50 и 54% голосов избирателей; кандидат-демократ 45 ± 2 , или между 43 и 47% голосов избирателей. Теперь между этими двумя доверительными интервалами нет никакого взаимного перекрытия. Вы можете в прямом эфире сообщить, что на выборах победил кандидат от республиканцев; такой прогноз окажется правильным более чем в 95 случаях из 100*.

Но это даже лучше. Из центральной предельной теоремы вам известно, что в 99,7% случаев пропорция долей выборки будет отстоять от истинной пропорции долей совокупности на расстоянии, не превышающем трех стандартных ошибок. В нашем примере с выборами 99,7%-ные доверительные интервалы для двух кандидатов таковы: кандидат от республиканцев: $52 \pm 3\%$, или между 49 и 55% голосов избирателей; кандидат от демократов

* Можно ожидать, что истинный процент голосов избирателей, отданных за кандидата от республиканцев, окажется за пределами доверительного интервала экзитпола приблизительно в 5 случаях из 100. В таких случаях истинный процент голосов избирателей, отданных за кандидата республиканцев, окажется меньше 50% или больше 54%. Если, однако, он получит больше 54% голосов избирателей, ваша телекомпания не ошибется, назвав его победителем (просто его победа окажется еще более убедительной, чем вы предсказывали). Таким образом, вероятность того, что проведенный вами экзитпол заставит вас ошибочно объявить победителем кандидата-республиканца, составляет лишь 2,5%.

$45 \pm 3\%$, или между 42 и 48% голосов избирателей. То есть после того как вы объявите победителем выборов кандидата-республиканца, благодаря новой выборке из 2000 избирателей останется лишь ничтожная вероятность того, что вы вместе со своим начальником будете уволены.

Вы, наверное, обратили внимание, что использование большей по объему выборки снижает стандартную ошибку. Именно за счет этого крупные общенациональные опросы позволяют получить необычайно точные результаты. В то же время выборки меньшего размера увеличивают величины стандартных ошибок и, следовательно, доверительный интервал (или «предел ошибки выборочного исследования», как принято говорить среди специалистов по проведению опросов общественного мнения). Текст, набранный мелким шрифтом в опросе The New York Times / CBS News, гласит, что предел погрешности для вопросов по поводу праймериз республиканцев составляет 5 процентных пунктов в сравнении с 3 процентными пунктами для других вопросов, включенных в опрос общественного мнения. Эти вопросы задавались лишь тем, кто сам назвал себя сторонником Республиканской партии, и тем, кто участвовал в голосованиях на закрытых собраниях ее членов, поэтому размер выборки для данной подгруппы вопросов снизился до 455 (общее количество избирателей, участвовавших в опросе, составило 1650).

Как обычно, примеры, приведенные в этой главе, «грешат» многими упрощениями. Вы, наверное, обратили внимание, что в примере с выборами у кандидатов от Республиканской и Демократической партий должна быть своя собственная стандартная ошибка. Вернемся еще раз к приведенной выше формуле: $SE = \sqrt{p(1-p)/n}$. Размер выборки n один и тот же для обоих кандидатов, однако p и $(1-p)$ будут несколько различаться. Во втором экзитполе (когда размер выборки был увеличен до 2000 избирателей) стандартная ошибка для кандидата от Республиканской партии составила $\sqrt{0,52(0,48)/2000} = 0,01117$; для кандидата от Демократической партии — $\sqrt{0,45(0,55)/2000} = 0,01112$. Разумеется, какими бы ни были наши намерения и цели, эти два числа должны быть одинаковы*. Именно поэтому я остановил свой выбор на общепринятом соглашении: из двух значений стандартной ошибки использовать большее значение для всех кандидатов. В любом

* Неравенство стандартных ошибок здесь обусловлено наличием третьего, «независимого» кандидата и, соответственно, процентом избирателей, отдавших ему свои голоса. Если было бы только два кандидата, то стандартные ошибки для каждого из них были бы всегда равны. *Прим. ред.*

случае такой подход вносит в доверительные интервалы небольшую дополнительную меру предосторожности.

При проведении многих общенациональных опросов общественного мнения, включающих в себя большое число вопросов, идут еще дальше. В случае опроса The New York Times / CBS News для каждого вопроса должна быть, строго говоря, своя стандартная ошибка (в зависимости от ответа). Например, стандартная ошибка, относящаяся к ситуации, когда 9% участников опроса одобряют деятельность Конгресса США, должна быть меньше стандартной ошибки, относящейся к ситуации, когда 46% участников опроса одобряют деятельность Обамы на посту президента США, поскольку $0,09 \times 0,91$ меньше, чем $0,46 \times 0,54$: 0,0819 против 0,2484. (Интуитивные соображения, на которых основывается эта формула, объясняются в приложении к настоящей главе.)

Поскольку использование собственной стандартной ошибки для каждого вопроса было бы неудобным и вносило бы излишнюю путаницу, при проведении подобных опросов общественного мнения обычно предполагается, что доля выборки для каждого вопроса равняется 0,5 (или 50%) — что порождает максимально возможную стандартную ошибку для любого размера выборки, — и именно такая стандартная ошибка используется при вычислении предела ошибки выборки для опроса в целом*.

При соответствующей организации опросы общественного мнения становятся поистине замечательными инструментами. Согласно Фрэнку Ньюпору, главному редактору Gallup Organization, опрос 1000 человек позволяет с высокой степенью точности оценить настроения в обществе в целом. С точки зрения статистики Фрэнк Ньюпор, несомненно, прав. Но чтобы получить столь значимые и точные данные, мы должны надлежащим образом провести опрос, а затем правильно интерпретировать его результаты, что порой намного легче *сказать*, чем *сделать*. Неправильные результаты опросов обычно обусловлены не ошибкой в математических расчетах при вычислении стандартных ошибок, а являются следствием некорректно сформированной выборки, или неправильно сформулированных вопросов, или того

* Формула для вычисления стандартной ошибки опроса, которую я использовал в данном случае, предполагает, что опрос проводится в произвольной выборке из соответствующей совокупности. Организации, специализирующиеся на проведении опросов общественного мнения, могут отходить от этого метода проведения выборочных исследований; в таком случае формула для вычисления стандартной ошибки опроса также несколько изменяется. Однако базовая методика остается той же.

и другого. Выражение «мусор на входе — мусор на выходе» полностью применимо к проведению социологических опросов. Ниже перечислены ключевые методологические вопросы, которые необходимо задать при проведении любого опроса общественного мнения или оценивании чьей-то работы.

Действительно ли данная выборка является репрезентативной (представительной) из совокупности, настроения которой мы пытаемся выяснить? Многие типичные проблемы, связанные с данными, уже обсуждались в главе 7. Тем не менее мне придется еще раз указать на опасность систематической ошибки выбора, особенно систематической ошибки самоотбора. Любой опрос, результаты которого зависят от людей, попадающих в выборку по собственной инициативе, например в ходе ток-шоу на радио или при проведении добровольных интернет-опросов, будет отражать мнения лишь тех, кто сам пожелал его высказать. В подобных случаях мы узнаем лишь мнения людей, которые проявляют повышенный интерес к рассматриваемому вопросу или располагают избытком свободного времени. Очевидно, что ни та ни другая группа не может отражать *общие* настроения общества. Однажды я сам участвовал в ток-шоу на радио в качестве гостя. Один из слушателей программы, ехавший в это время в автомобиле по каким-то своим делам, позвонил на радиостанцию и выразил категорическое несогласие с моим мнением. Мои взгляды возмутили его до такой степени, что он не поленился свернуть с автомагистрали к телефонной будке, которую заметил возле обочины, чтобы позвонить в радиостудию. Хотелось бы верить, что те слушатели, которые во время этого ток-шоу не свернули с автомагистрали, разделяли мои взгляды.

Любой метод выяснения мнений, который систематически исключает какой-либо сегмент совокупности, также приводит к ошибке выбора. Например, широкое распространение мобильной связи породило множество новых методологических сложностей. Организации, специализирующиеся на проведении социологических опросов, делают все от них зависящее, чтобы опросить репрезентативную выборку соответствующей совокупности. Опрос The New York Times / CBS News базировался на телефонных интервью, проводившихся на протяжении шести дней с 1650 взрослыми американцами, 1475 из которых сообщили, что зарегистрированы для участия в голосовании.

Относительно остальной части методологии, применявшейся при проведении этого опроса, я могу лишь догадываться, но большинство опросов,

которые проводятся социологическими организациями, используют тот или иной вариант описанных ниже методов. Чтобы гарантировать, что люди, поднявшие трубку, отражают мнение совокупности в целом, данный процесс начинается с использования теории вероятностей — нечто наподобие вытаскивания шариков из урны. Компьютер случайным образом выбирает некую совокупность номеров коммутационных станций стационарной телефонной связи. (Номер коммутационной станции стационарной телефонной связи представляет собой код региона плюс первые три цифры телефонного номера.) За счет случайного выбора 69 000 номеров коммутационных станций стационарной телефонной связи в Соединенных Штатах, каждый в пропорции к своей доле во всей совокупности телефонных номеров, данный опрос в целом, по-видимому, отразит географическое распределение соответствующей совокупности. Как поясняется в тексте, набранном мелким шрифтом, «номера коммутационных станций стационарной телефонной связи были выбраны таким образом, чтобы каждый регион страны был представлен в пропорции к его доле во всей совокупности телефонных номеров». К каждому выбранному номеру компьютер добавил четыре случайные цифры. Таким образом, в окончательном списке домохозяйств, которые предстояло обзвонить в ходе опроса, оказались как фактически используемые, так и неиспользуемые телефонные номера. Кроме того, этот опрос предусматривал «случайный набор номеров мобильных телефонов».

Для каждого набираемого телефонного номера один взрослый член семьи назначался респондентом посредством некой «произвольной процедуры» (например, телефонную трубку предлагалось взять самому молодому из взрослых членов семьи). Этот процесс был усовершенствован, чтобы получить выборку респондентов, отражающую возрастной и половой состав взрослого населения страны. Самое главное — интервьюер будет пытаться сделать несколько звонков в разное время суток, чтобы дозвониться на каждый из выбранных телефонных номеров. Эти неоднократные попытки — до десяти или двенадцати звонков на один и тот же телефонный номер — являются важным условием получения правильной выборки. Очевидно, было бы дешевле и проще звонить на разные телефонные номера до тех пор, пока достаточно большая выборка взрослых не подойдет к телефонам и не ответит на соответствующие вопросы. Однако такая выборка допустила бы сильный крен в пользу тех, кто большую часть времени проводит дома, а в это число входят главным образом безработные,

пенсионеры, инвалиды и т. д. Такой вариант опроса был бы вполне уместен, если бы вы намеревались квалифицировать его результаты следующим образом: деятельность Обамы на посту президента США одобряют 46% безработных, пенсионеров и прочих слоев населения, с готовностью отвечающих на телефонные опросы общественного мнения.

Одним из показателей достоверности опроса является так называемый процент ответивших, то есть доля респондентов, выбранных для проведения опроса и в конечном счете ответивших на его вопросы. Низкий процент ответивших может указывать на неправильное формирование выборки. Чем больше респондентов отказались отвечать на поставленные вопросы (или до них просто не удалось дозвониться), тем выше вероятность, что эта значительная группа людей в чем-то весьма существенно отличается от тех, кто согласился участвовать в опросе. Организаторы опроса могут выполнить тест на «систематическую ошибку отсутствия ответа», проанализировав имеющиеся в их распоряжении данные о респондентах, с которыми им не удалось установить контакт. Возможно, они проживают в каком-то специфическом регионе, или не желают отвечать на вопросы в силу какой-то особой причины, или принадлежат к какой-то расовой или этнической группе, или имеют какой-то определенный уровень дохода. Анализ такого рода зачастую помогает выяснить, повлияет ли низкий процент ответивших на результаты опроса в целом.

Позволяет ли формулировка вопросов получить точную информацию по интересующим нас темам? Чтобы выяснить настроения в обществе, необходимо учитывать гораздо больше нюансов, чем при оценивании экзамена или измерении веса респондентов. Результаты социологического опроса во многом зависят от правильности формулировки задаваемых вопросов. Рассмотрим пример, который на первый взгляд кажется довольно простым: какой процент американцев поддерживает смертную казнь? Как следует из названия этой главы, это заведомое большинство американцев. Согласно опросу, проведенному Институтом Гэллапа, начиная с 2002 года свыше 60% американцев ежегодно заявляют, что поддерживают применение смертной казни в отношении лиц, осужденных за убийство. Процент американцев, выступающих за смертную казнь, колеблется в относительно узком диапазоне, от высоких 70% в 2003 году до более низких 64% в отдельные годы. Эти данные позволяют сделать однозначный вывод: заведомое большинство американцев выступают за смертную казнь.

Или такой вывод слишком поспешен? Поддержка американцами смертной казни падает, *когда в качестве альтернативы предлагается пожизненное тюремное заключение без права условно-досрочного освобождения*. Опрос, проведенный Институтом Гэллапа в 2006 году, показал, что лишь 47% американцев считают смертную казнь справедливой карой за убийство, тогда как 48% высказываются за пожизненное тюремное заключение². Это не просто некий статистический парадокс, которым можно удивить гостей, пришедших к вам на вечеринку; фактически это уже означает отсутствие в стране большинства, поддерживающего применение смертной казни при наличии альтернативы в виде пожизненного тюремного заключения. Когда мы пытаемся выяснить отношение общества к той или иной проблеме, важнейшую роль играют формулировка вопроса и выбор языка.

Политики зачастую стараются сыграть на этом обстоятельстве, используя опросы общественного мнения и фокус-группы для тестирования «слов, которые приносят нужный результат». Например, избиратели в большей степени склонны поддерживать формулировку «снижение налогового бремени», чем «урезание налогов», несмотря на то что обе формулировки, по сути, описывают одно и то же действие. Аналогично, избирателей меньше волнует «изменение климата», чем «глобальное потепление», несмотря на то что глобальное потепление — лишь одна из форм изменения климата. Очевидно, политики пытаются манипулировать ответами избирателей путем использования «не нейтральных» слов. Если социологическая организация стремится создать себе репутацию «честной», то есть выдающей результаты, заслуживающие доверия, она должна отказаться от употребления языка, способного повлиять на точность собираемой информации. Точно так же если по истечении какого-то времени предполагается сравнивать результаты опросов (например, как потребители оценивают нынешнее состояние экономики в сравнении с тем, как они оценивали его год назад), то вопросы, позволяющие получить требуемую информацию, в том и другом случае должны быть одинаковыми — или по крайней мере очень похожими.

Организации по исследованию общественного мнения (например Gallup Organization) зачастую проводят так называемое тестирование расщепленной выборки, когда разные варианты одного и того же вопроса тестируются на разных выборках, чтобы оценить, как незначительные изменения в формулировке вопроса влияют на ответы респондентов. Для таких экспертов,

как Фрэнк Ньюпор из Gallup Organization, ответы буквально на каждый вопрос несут в себе значимую информацию, даже когда они кажутся несовместимыми³. Тот факт, что отношение американцев к смертной казни резко меняется, когда в качестве альтернативы предлагается пожизненное тюремное заключение без права условно-досрочного освобождения, свидетельствует о чем-то важном. По мнению Ньюпора, результаты любого опроса общественного мнения необходимо рассматривать в общем контексте. Никакой отдельно взятый вопрос или опрос не в состоянии охватить всей глубины настроений общества, когда речь идет о какой-либо сложной проблеме.

Говорят ли респонденты правду? Опрос общественного мнения, как и знакомство в интернете, предполагает некоторое «пространство для маневра». Нам известно, что люди не всегда говорят правду, особенно когда им приходится отвечать на разного рода затруднительные и щекотливые вопросы. Респонденты могут завышать свой доход или преувеличивать свои возможности, когда у них спрашивают, например, о том, сколько раз в месяц они занимаются сексом. Они могут сообщить, что пойдут голосовать, хотя на самом деле предпочтут какой-то другой вид досуга. Они могут бояться выражать непопулярную или социально неприемлемую точку зрения. Именно по этим причинам даже идеально продуманный и организованный опрос зависит от того, насколько правдивы ответы респондентов.

При проведении опросов, касающихся выборов, очень важно заранее отсортировать тех, кто не придет на избирательные участки, от тех, кто намерен голосовать. (Если наша цель — определить вероятного победителя выборов, то какое нам дело до мнения тех, кто не собирается его избирать?) Люди часто говорят, что примут участие в голосовании, только потому, что им кажется, будто именно такого ответа от них ждут. Результаты исследований, в ходе которых сравнивалось количество избирателей, фактически пришедших на избирательные участки, с количеством тех, кто обещал прийти, показали, что от одной четверти до трети респондентов, утверждавших, что будут участвовать в выборах, не голосовали⁴. Один из способов минимизировать искажения, вносимые неправдивыми ответами, — выяснить, участвовал ли данный респондент в голосовании на прошлых или нескольких предыдущих выборах. Респонденты, регулярно игнорирующие выборы, скорее всего, не станут участвовать в них и в дальнейшем. Аналогично, если есть опасения, что респонденты не решатся дать социально неприемлемый ответ на поставленный вопрос (например выразить отрицательное отношение

к какой-либо расовой или этнической группе), то можно попытаться его более тонко сформулировать (например спросить, «придерживаются ли такого мнения знакомые вам люди»).

Одним из самых щекотливых за все время стало исследование, проведенное Национальным центром исследования общественного мнения (National Opinion Research Center — NORC) при Чикагском университете. Полное название исследования было таким: «Социальная организация сексуальности: половая жизнь в Соединенных Штатах»; Впрочем, довольно быстро за ним закрепилось более краткое название: «Исследование секса»⁵. Формальное описание исследования включало такие фразы: «организация моделей поведения, на которых строятся половые контакты» и «выбор сексуальных партнеров и сексуальное поведение на протяжении жизни». Я слишком упрощаю, говоря, что исследователи пытались задокументировать «кто, как, с кем и как часто». Целью данного исследования, результаты которого были опубликованы в 1995 году, было не просто просветить нас относительно сексуального поведения соседей (хотя об этом тоже шла речь), но и оценить, как сексуальное поведение американцев влияет на распространение ВИЧ/СПИД.

Если уж американцы не решаются признаться, что не будут голосовать, то можно только представить, насколько они горят желанием описывать свое сексуальное поведение, если под ним могут, в частности, подразумеваться какие-либо предосудительные действия (например супружеская неверность) или даже склонность к половым извращениям. В данном исследовании использовалась впечатляющая методология. Оно основывалось на собеседованиях с репрезентативной выборкой взрослого населения США, включающей 3342 человека. Каждое собеседование занимало примерно 90 минут. Почти 80% респондентов заполнили соответствующую анкету, что позволило авторам исследования сделать вывод о том, что его результаты достаточно точно отражают сексуальное поведение американцев в целом (по крайней мере, в 1995 году).

Поскольку вы уже одолели большую часть книги и, в частности, главу, посвященную методологии проведения опросов общественного мнения, вы имеете право вкратце ознакомиться с выводами авторов «Исследования секса» (должен заранее вас разочаровать: ничего особенно шокирующего в них нет). Как заметил один из обозревателей, «секс занимает в жизни американцев гораздо меньше места, чем можно было бы предположить»⁶.

- Люди, как правило, занимаются сексом с теми, кто им близок по тем или иным признакам. Девяносто процентов пар относятся к одной и той же расе, религии, социальному классу и возрастной группе.
- Типичный респондент занимался сексом «пару-тройку раз в месяц» (правда, разброс по этому показателю весьма значителен). Количество сексуальных партнеров после достижения восемнадцатилетнего возраста колеблется от нуля до 1000 (и более).
- Примерно 5% мужчин и 4% женщин сообщили о том или ином числе сексуальных контактов с партнерами своего пола.
- У 80% респондентов в предыдущем году был либо один, либо ни одного сексуального партнера.
- Респонденты, имеющие одного сексуального партнера, оказались более счастливы по сравнению с теми, у кого вообще не было сексуального партнера или у кого их было много⁷.
- Четверть женатых мужчин и 10% замужних женщин сообщали о наличии у них внебрачных половых связей.
- Большинство людей занимаются «этим» по старинке: вагинальный половой акт оказался самым привлекательным способом половых контактов для мужчин и женщин.

В одном из обзоров «Исследования секса» было высказано простое, но важное критическое замечание, что точность этого опроса отражает действительные сексуальные практики взрослого населения Соединенных Штатов и «предполагает, что респонденты являются частью населения, от которого эти ответы были получены, и что эти люди честно отвечали на поставленные вопросы»⁸. Данное высказывание также может служить выводом для всей этой главы. На первый взгляд, самым подозрительным в любом опросе может показаться то, что мнения столь небольшого числа людей способны отражать мнения населения всей страны. *Но в этом-то как раз ничего удивительного или подозрительного нет.* Один из самых фундаментальных статистических принципов заключается в том, что надлежащим образом сформированная выборка способна точно отражать совокупность, из которой она извлечена. Реальных проблем проведения опросов общественного мнения две: 1) определение правильной выборки и выход на нее и 2) получение информации от этой репрезентативной группы таким образом, чтобы она точно отражала мнения ее членов.

Приложение к главе 10

Почему стандартная ошибка оказывается больше, когда p и $(1 - p)$ близки к 50%

Здесь излагаются интуитивные соображения, объясняющие, почему стандартная ошибка оказывается самой большой, когда доля ответивших определенным образом (p) близка к 50% (что с математической точки зрения означает, что величина $(1 - p)$ также будет близка к 50%). Давайте представим: вы проводите два опроса в штате Северная Дакота. Первый опрос призван оценить соотношение сторонников Республиканской и Демократической партии в этом штате. Допустим, истинное их соотношение примерно 50 на 50, однако в ходе проведения вашего опроса выяснилось, что 60% населения штата поддерживают республиканцев, а 40% — демократов. Ваши результаты отличаются от реального положения вещей на 10%, что является достаточно большой погрешностью. Однако столь существенная ошибка у вас получилась несмотря на то, что вы не допустили невообразимо большой ошибки при сборе данных. Вы завысили долю сторонников Республиканской партии по сравнению с их настоящей долей в населении штата на 20% [$(60 - 50)/50 = 0,2$], при этом занизив долю сторонников Демократической партии в штате также на 20% [$(40 - 50)/50 = 0,2$]. Это могло случиться даже при использовании весьма эффективной методологии проведения опроса.

Ваш второй опрос призван определить долю коренных жителей Америки в населении штата Северная Дакота. Допустим, их истинная доля равняется 10%, а доля некоренных жителей — 90%. Теперь рассмотрим, насколько неправильными должны оказаться собранные вами данные, чтобы погрешность вашего опроса составила целых 10%. Это могло бы произойти в двух случаях. Первый: ваш опрос мог показать, что доля коренных жителей Америки в населении Северной Дакоты составляет 0%, то есть все население штата — некоренные американцы. Второй: согласно опросу, доля коренных жителей Америки в населении Северной Дакоты составляет 20%, а доля некоренных жителей — 80%. Итак, в первом случае вы упустили из виду *всех коренных жителей Америки*, а во втором *в два раза завысили их долю в населении штата* по сравнению с истинным положением дел. Это, конечно, серьезные ошибки проведения выборочного исследования. В обоих случаях

ошибка вашей оценки составила 100%: либо $[(0 - 10)/10 = -1]$, либо $[(20 - 10)/10 = 1]$. А если бы вы упустили из виду лишь 20% коренных жителей Америки — именно такую ошибку вы допустили в опросе, касающемся соотношения сторонников Республиканской и Демократической партии в Северной Дакоте, — то в результате вашего опроса оказалось бы, что доля коренных жителей Америки в Северной Дакоте составляет 8%, а доля некоренных жителей — 92%, то есть в этом случае вы ошиблись бы всего на 2% по сравнению с истинным соотношением коренных и некоренных жителей Америки в населении Северной Дакоты.

Когда p и $(1 - p)$ близки к 50%, относительно небольшие ошибки выборочного исследования трансформируются в крупные абсолютные ошибки результатов опроса.

И наоборот, когда p или $(1 - p)$ близки к нулю, даже относительно крупные ошибки выборочного исследования трансформируются в небольшие абсолютные ошибки результатов опроса.

Одна и та же 20-процентная ошибка выборочного исследования исказила результат опроса, касающегося соотношения сторонников Республиканской и Демократической партии, на 10%, исказив лишь на 2% результат опроса о соотношении коренных и некоренных жителей Америки в населении Северной Дакоты. Поскольку стандартная ошибка любого опроса измеряется в абсолютных значениях (например, $\pm 5\%$), из нашей формулы следует, что эта ошибка приблизится к своему максимальному значению, когда p и $(1 - p)$ окажутся близки к 50%*.

* По-видимому, самое простое доказательство, что функция $f(p) = p(1 - p) = p - p^2$ принимает максимальное значение при $p = 0,5$, — это математическое доказательство. Находим производную $f'(p) = 1 - 2p$, приравниваем ее к нулю и получаем уравнение $1 - 2p = 0$. Решением этого уравнения будет $p = 0,5$. Что и требовалось доказать. (О том, что это максимум, свидетельствует вторая производная $f''(p) = -2$.) Прим. ред.

Может ли стресс на работе стать причиной вашей смерти? Да, вполне. Существуют убедительные доказательства того, что суровые условия на работе могут привести к преждевременной смерти, особенно в результате развития сердечно-сосудистых заболеваний. Однако это не тот вид стресса, о котором вы, наверное, подумали. Главы компаний, которым буквально каждый день приходится принимать чрезвычайно сложные и ответственные решения, определяющие дальнейшую судьбу их бизнеса, рискуют значительно *меньше*, чем их секретарши, бесконечно отвечающие на телефонные звонки, параллельно выполняя множество других задач, предусмотренных должностной инструкцией. Как такое может быть? Оказывается, самый опасный вид стресса на работе обусловлен невозможностью человека в достаточной степени контролировать способы и условия выполнения поставленных задач. Ряд исследований, проводившихся (по заказу правительства) в отношении тысяч британских мелких чиновников, показал, что от них практически не зависит, чем именно им предстоит заниматься и как именно это выполнять, что и является причиной их высокой смертности по сравнению с чиновниками более высоких рангов, ответственных за принятие важных решений. Согласно результатам исследования, человека убивает не стресс, связанный с повышенной ответственностью, а стресс, вызванный необходимостью делать работу, не имея возможности решать, как и когда.

Но не пугайтесь, эта глава не о стрессе на работе, сердечно-сосудистых заболеваниях или государственных служащих Британии. Нас прежде всего интересует, как ученые приходят к подобным выводам. Очевидно, что это не результат рандомизированного эксперимента. Мы не можем произвольно поручать людям некую работу, заставляя их долгие годы ею заниматься, а затем выяснять, кто из них раньше умер. (Случайным образом поручая людям выполнение тех или иных задач, мы рискуем нанести огромный

вред государственной службе Британии, не говоря уже об этической стороне дела.) Вместо этого исследователи собирали о тысячах государственных служащих Британии подробные повторные данные, анализ которых позволяет выявить определенные связи, например между невозможностью человека в достаточной степени контролировать способы и условия выполнения своей работы и развитием сердечно-сосудистых заболеваний.

Но такой связи мало для того, чтобы сделать вывод о вреде тех или иных видов работ для здоровья человека. Если мы просто замечаем, что мелкие государственные служащие в иерархии британской государственной службы страдают сердечно-сосудистыми заболеваниями чаще других, то полученные нами результаты будут искажаться действием ряда других факторов. Например, можно было бы ожидать, что уровень образования мелких чиновников окажется ниже, чем у чиновников более высоких рангов. Может также выясниться, что среди мелких государственных служащих больше курящих (не исключено, что это объясняется их неудовлетворенностью работой). Вполне вероятно, что у этих людей было трудное детство, и это сузило перспективы их будущего карьерного роста. Или их невысокий уровень доходов не позволяет им уделять должное внимание своему здоровью. И так далее. Дело в том, что любое сравнительное исследование — изменение состояния здоровья у большой группы британских работников или какой-то другой крупной группы населения — не позволяет нам сделать далекоидущие выводы. Возможно, что другие источники изменения полученных нами данных внесут искажения в интересующую нас связь. Можем ли мы быть уверены в том, что именно невозможность человека в достаточной степени контролировать содержание, способы и условия выполнения своей работы является подлинной причиной развития у него сердечно-сосудистых заболеваний? Или истинная причина — в сочетании действия ряда факторов, которые оказались общими для данной категории людей?

Статистический инструмент под названием регрессионный анализ помогает решить данную проблему. А если конкретнее, то регрессионный анализ позволяет нам измерить величину зависимости между какой-то переменной и интересующим нас исходом, *зафиксировав действие всех прочих факторов*. Другими словами, мы можем вычленить влияние одной переменной (например, занятие определенным родом деятельности), сохраняя на постоянном уровне действие других переменных. Регрессионный анализ использовался при проведении упоминавшегося нами исследования, которое проводилось

по заказу британского правительства и имело своей целью оценить, как невозможность человека в достаточной степени контролировать содержание, способы и условия выполнения своей работы сказывается на состоянии здоровья людей, схожих во всех остальных отношениях, например курильщиков. (Рядовые работники действительно курят больше своих начальников; это объясняет относительно малую величину разброса в сердечно-сосудистых заболеваниях во всей иерархии британской государственной службы.)

Большинство исследований, о которых вам приходилось читать в прессе, основываются на регрессионном анализе. Когда ученые приходят к выводу, что у детей, посещавших детсад, чаще возникают проблемы с успеваемостью в начальной школе, чем у детей, которые воспитывались дома, это вовсе не означает, что они случайным образом сформировали выборку из нескольких тысяч детей, одну половину которых отправили в детсады, а другую оставили на попечении родителей. Это также не означает, что исследователи просто сравнили успеваемость в начальной школе детей, посещавших детсад, и детей, находившихся дома, не отдавая себе отчета в том, что эти две группы детей фундаментально разнятся между собой по ряду других показателей. В разных семьях принимаются разные решения относительно воспитания детей *именно потому, что эти семьи — разные*. В одних семьях детей воспитывают оба родителя, в других — только один. Есть семьи, где работают оба родителя, а есть — где только один. Какие-то семьи более состоятельны и образованны, какие-то менее. Все эти факторы так или иначе сказываются на принятии решений относительно воспитания детей *и не могут не влиять на их успеваемость во время учебы в начальной школе*. В случае надлежащего выполнения регрессионный анализ помогает оценить влияние воспитания, исключив из рассмотрения другие факторы воздействия на детей: семейный доход, структуру семьи, образование родителей и т. п.

В приведенном выше предложении есть два ключевых словосочетания. Первое: «в случае надлежащего выполнения». Сегодня при наличии соответствующих данных и доступа к персональному компьютеру даже шестилетний ребенок может воспользоваться какой-либо статистической программой для получения результатов регрессионного анализа, поскольку это не потребует практически никаких умственных усилий. Проблема не в выполнении регрессионного анализа как такового, главная трудность — определить, какие именно переменные следует рассматривать в этом анализе и как это лучше всего сделать. Регрессионный анализ подобен многим современным

универсальным электромеханическим инструментам: им относительно легко пользоваться, но трудно это делать эффективно, не говоря уже о том, что при ненадлежащем использовании, то есть неумелом обращении, он оказывается потенциально опасен.

Второе важное словосочетание: «помогает оценить». Наше исследование воспитания детей не дает нам «правильного» ответа относительно зависимости между способом воспитания ребенка (в детсаду или дома) и его успеваемостью в начальной школе. Вместо этого оно оценивает величину этой связи *у конкретной группы детей на определенном отрезке времени*. Можем ли мы сделать выводы, применимые к более широкой совокупности? Да, но при этом нам придется иметь дело с такими же ограничениями и условиями, с какими мы сталкиваемся, делая любой другой статистический вывод. Во-первых, используемая нами выборка должна быть репрезентативной, то есть представлять всю интересующую нас совокупность. Исследование 2000 детей в Швеции не позволит нам прийти к сколь-нибудь значимым выводам относительно оптимальных методов дошкольного образования детей в сельскохозяйственных районах Мексики. И во-вторых, не следует забывать о существовании разброса между выборками. Если мы выполняем ряд исследований, касающихся детей и их воспитания, то их результаты будут несколько отличаться между собой, даже если используемые при этом методологии будут одинаковы и совершенно надежны.

Регрессионный анализ подобен проведению опросов общественного мнения. Обнадешивает то, что при применении крупной репрезентативной выборки и правильной методологии наблюдаемая взаимосвязь между данными выборки не должна существенно отличаться от истинной взаимосвязи для совокупности в целом. Если у 10 000 человек, занимающихся спортом не менее трех раз в неделю, уровень заболеваемости сердечно-сосудистой системы значительно ниже, чем у 10 000 человек, не занимающихся спортом (но не отличающихся от первых 10 000 человек во всех остальных отношениях), то весьма высока вероятность того, что мы будем наблюдать аналогичную связь между регулярными занятиями спортом и уровнем заболеваемости сердечно-сосудистой системы для более широкой совокупности. Именно поэтому мы выполняем исследования такого рода. (Задача ученых вовсе не в том, чтобы по завершении исследования упрекнуть тех, кто не занимается спортом и имеет проблемы с сердцем, что в свое время им не следовало игнорировать эти занятия.)

Плохо, однако, то, что мы не можем с полной уверенностью утверждать, что занятия спортом предотвращают возникновение сердечно-сосудистых заболеваний. Вместо этого мы отвергаем нулевую гипотезу о том, что занятия спортом никак не связаны с болезнями сердца. Отвергнуть ее нам позволяет достижение определенного статистического порога, выбранного еще до начала выполнения исследования. Если конкретнее, то авторы данного исследования должны были бы указать, что в случае, если занятия спортом никак не связаны с сердечно-сосудистыми заболеваниями, вероятность наблюдения столь заметной разницы в уровне заболеваемости сердечно-сосудистой системы между теми, кто регулярно занимается спортом, и теми, кто им не занимается, в этой крупной выборке должна быть менее 0,05 или ниже какого-то другого порога статистической значимости.

Давайте остановимся на мгновение и помашем нашим первым гигантским желтым флагом*. Допустим, что в этом конкретном исследовании сравнивалась большая группа людей, регулярно играющих в сквош, с людьми из такой же по величине группы, которые вообще не занимаются спортом. Игра в сквош обеспечивает неплохую нагрузку на сердечно-сосудистую систему. Однако нам также известно, что игроки в сквош — достаточно состоятельные люди, чтобы быть членами клубов, располагающих хорошими сквош-кортами. Богатые люди могут себе позволить уделять должное внимание здоровью, что также способствует снижению заболеваемости их сердечно-сосудистой системы. Если выполненный нами анализ страдает небрежностями, то хорошее состояние здоровья можно объяснить игрой в сквош, хотя на самом деле оно объясняется высокими доходами, которые дают человеку возможность играть в сквош (в таком случае даже увлечение игрой в поло можно при желании связать с хорошим состоянием здоровья, если, конечно, закрыть глаза на то, что во время игры в поло большая часть физической работы выполняется лошадью).

Ничто не мешает нам также предположить, что причинно-следственные связи имеют противоположную направленность. Может быть, здоровое сердце является «причиной» того, что человек занимается спортом? Почему бы и нет! Те, кто не блещет здоровьем, — особенно люди с врожденными заболеваниями сердца, — не могут полноценно заниматься спортом, что вполне

* Согласно Международному своду сигналов, поднятый желтый флаг означает карантин. Таким образом автор предостерегает читателя об «опасности» дальнейшего текста, где описывает возможные «ловушки» регрессионного анализа. *Прим. ред.*

понятно. Вряд ли они в состоянии регулярно играть в сквош. Опять-таки, если выполненный нами анализ сделан небрежно или чрезмерно упрощен, утверждение о том, что занятия спортом способствуют улучшению здоровья, может лишь отражать то обстоятельство, что тем, кто им не блещет, бывает очень нелегко заниматься спортом. В этом случае игра в сквош никоим образом не улучшает состояние здоровья — а лишь отделяет здоровых от больных.

Существует так много потенциальных «регрессионных ловушек», что я решил посвятить их рассмотрению всю следующую главу. Пока же будем считать, что на нашем пути ни одна из них не встретится. Регрессионный анализ обладает замечательным свойством вычленять в каждом отдельном случае статистическую связь, которая представляет для нас интерес, например связь между невозможностью человека в достаточной степени контролировать содержание, способы и условия выполнения своей работы и развитием сердечно-сосудистых заболеваний, учитывая при этом другие факторы, которые могут внести в нее искажения.

Как действует данный механизм? Если нам известно, что мелкие государственные служащие Британии курят чаще, чем их начальники, то как нам определить, в какой мере плохое состояние их сердечно-сосудистой системы обусловлено спецификой работы, а в какой — этой пагубной привычкой? Оба фактора кажутся неразрывно связанными между собой.

Регрессионный анализ (выполненный надлежащим образом!) позволяет разделить эти факторы. Чтобы объяснить процесс на интуитивном уровне, мне придется начать с базовой идеи, лежащей в основе всех форм регрессионного анализа, от простейших статистических связей до сложных моделей, разработанных лауреатами Нобелевской премии. По своей сути регрессионный анализ стремится найти «наилучшее приближение» линейной зависимости между двумя переменными. Простой пример — зависимость между ростом и весом людей. Те, кто выше ростом, как правило, весят больше, хотя эта закономерность соблюдается не всегда. Если бы мы построили диаграмму разброса для роста и веса группы студентов-выпускников, то получили бы нечто наподобие того, что уже видели в главе 4.

Если бы вас попросили описать получившуюся картину, вы бы наверняка сказали что-то вроде: «Вес, по-видимому, увеличивается пропорционально росту». Такую догадку вряд ли можно назвать озарением. Регрессионный анализ позволяет нам пойти дальше и «провести линию», которая точнее всего отражает линейную зависимость между этими двумя переменными.

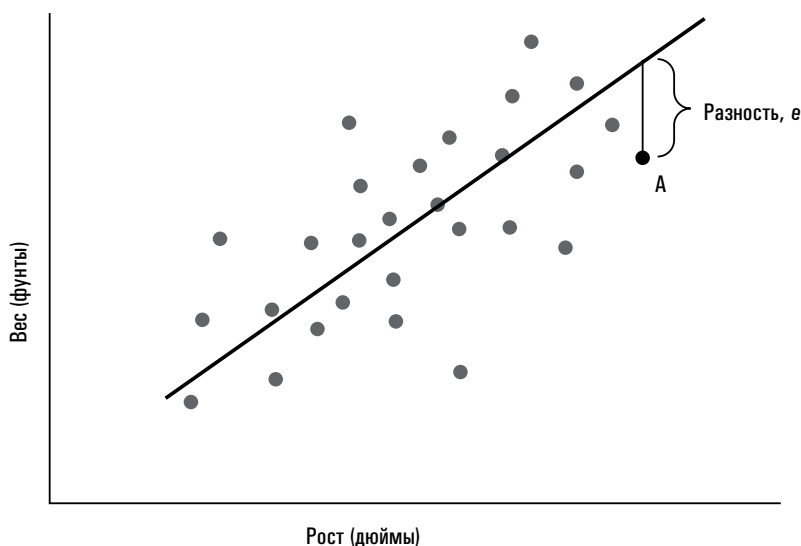


Можно провести множество линий, которые будут отражать соотношение между ростом и весом. Но как знать, какая из них это делает *точнее всего*? К тому же посредством какого критерия мы определяем эту линию? Регрессионный анализ обычно использует методологию под названием стандартный метод наименьших квадратов, МНК. Если читателя интересуют его технические подробности и он хочет узнать, почему МНК обеспечивает «наилучшее приближение», ему придется обратиться к более солидным учебникам по статистике. Ключевыми словами в названии МНК являются «наименьшие квадраты»: МНК определяет линию, минимизирующую сумму квадратов разностей. Это не настолько сложно, как может показаться на первый взгляд. Каждое наблюдение в нашей совокупности данных «рост/вес» характеризуется разностью, которая представляет собой его расстояние по вертикали от линии регрессии; это не относится к наблюдениям, расположенным непосредственно на линии: для них разность равняется нулю. (На представленной ниже диаграмме разброса разность отмечена для некоего гипотетического лица А.) На интуитивном уровне должно быть понятно, что чем больше сумма разностей в целом, тем худшее приближение обеспечивает данная линия. Единственное, что может быть непонятно в МНК на интуитивном уровне, это то, что в соответствующей формуле суммируются *квадраты* каждой разности (тем самым увеличивается весовой

коэффициент, назначаемый наблюдениям, которые расположены особенно далеко от линии регрессии, то есть «отщепенцам»).

Обычный метод наименьших квадратов позволяет определить линию, которая минимизирует сумму квадратов разностей, как показано ниже.

Линия наилучшего приближения для роста и веса



Если технические подробности вызывают у вас головную боль, можете не обращать на них внимания. Важно запомнить главное: стандартный метод наименьших квадратов позволяет получить наилучшее описание линейной зависимости между двумя переменными. В результате мы получаем не только линию как таковую, но и — как вы, наверное, помните из курса геометрии в средней школе — уравнение, описывающее ее. Оно известно как уравнение регрессии и имеет следующий вид: $y = a + bx$, где y — вес в фунтах, a — отрезок, отсекаемый этой линией на оси Y (то есть значение y , когда $x = 0$), b — коэффициент наклона линии, а x — рост в дюймах. Коэффициент наклона b найденной нами линией описывает «наилучшую» линейную зависимость между ростом и весом для соответствующей выборки, как определяется стандартным методом наименьших квадратов.

Линия регрессии, конечно, не описывает идеальным образом каждое наблюдение в соответствующей совокупности данных. Но как бы то ни было, это лучшее из возможных описаний зависимости между весом и ростом

человека. Это также означает, что каждое наблюдение можно объяснить как $\text{Вес} = a + b(\text{Рост}) + e$, где e — «разность», представляющая собой отклонение веса для каждого человека, которое не объясняется его ростом. Наконец, это означает, что наше оптимальное предположение относительно веса какого-либо человека в рассматриваемой совокупности данных будет иметь такой вид: $a + b(\text{Рост})$. Несмотря на то что большинство наблюдений не лежат непосредственно на линии регрессии, *ожидаемая* величина разности все же равняется нулю, поскольку вероятность того, что вес любого человека в выборке окажется больше, чем прогнозирует уравнение регрессии, равна вероятности того, что его вес окажется меньше, чем прогнозирует уравнение регрессии.

Впрочем, довольно теоретического жаргона! Давайте посмотрим на реальные данные роста и веса из исследования Americans' Changing Lives. Правда, вначале мне придется прояснить кое-какую базовую терминологию. Переменная, которая подлежит объяснению, — в нашем случае это вес — называется зависимой переменной, так как она зависит от других факторов. Переменные, используемые для объяснения зависимой переменной, называются объясняющими переменными, поскольку они объясняют интересующий нас результат. (Чтобы еще больше запутать мозги, объясняющие переменные иногда называют независимыми или управляющими переменными.) Начнем с использования роста, чтобы объяснить вес участников исследования Americans' Changing Lives, а впоследствии добавим другие потенциальные объясняющие факторы*. В исследовании Americans' Changing Lives участвуют 3537 взрослых. В нашем случае это количество наблюдений, или n . (Иногда в научных статьях это обозначается так: $n = 3537$.) Когда мы выполняем простую регрессию по отношению к данным Americans' Changing Lives, где вес — зависимая переменная, а рост — единственная объясняющая переменная, то получаем следующие результаты:

Вес = -135 + 4,5 × Рост в дюймах

$a = -135$. Это не что иное, как отрезок, отсекаемый линией регрессии на оси Y ; никакого специального объяснения у этой величины нет. (Если

* Это упражнение следует рассматривать как «игру с данными», а вовсе не как заслуживающее доверия исследование каких-либо зависимостей, описанных в последующих уравнениях регрессии. Наша цель — предоставить читателям интуитивно понятный пример того, как «работает» регрессионный анализ, а не выполнить строго научное исследование, касающееся веса американцев.

интерпретировать ее буквально, то получается, что человек с нулевым ростом весил бы -135 фунтов [отрицательная величина]; очевидно, что это нонсенс с любой точки зрения.) Эту величину также называют константой, поскольку она является отправной точкой для вычисления веса всех наблюдений в исследовании.

$b = 4,5$. Наша оценка для b (4,5) называется коэффициентом регрессии или, на статистическом жаргоне, «коэффициентом по росту», поскольку такой коэффициент служит наилучшей оценкой зависимости между ростом и весом участников исследования Americans' Changing Lives. У коэффициента регрессии имеется удобная интерпретация: увеличение на одну единицу независимой переменной (рост) ассоциируется с увеличением на 4,5 единицы зависимой переменной (вес). Для нашей выборки данных это означает, что увеличение роста на один дюйм сопряжено с увеличением веса на 4,5 фунта. Таким образом, если бы мы не располагали никакой другой информацией, то нашим оптимальным предположением относительно веса участника исследования Americans' Changing Lives, рост которого составляет 5 футов и 10 дюймов (то есть 70 дюймов), было бы $-135 + 4,5 \times 70 = 180$ фунтов.

Это наша победа, поскольку нам удалось получить численное выражение наилучшего приближения линейной зависимости между ростом и весом участников исследования Americans' Changing Lives. Те же самые базовые инструменты можно использовать для исследования более сложных зависимостей и получения ответов на более социально значимые вопросы. При любом коэффициенте регрессии вас, по сути, будут интересовать три вещи: *знак, величина и значимость*.

Знак. Знак (положительный или отрицательный) при коэффициенте для независимой переменной указывает направление его связи с зависимой переменной (исход, который мы пытаемся объяснить). В рассматриваемом нами случае коэффициент по росту является положительным. Более высокие люди, как правило, имеют больший вес. Некоторые зависимости действуют в противоположном направлении. Скажем, можно ожидать, что связь между занятиями спортом и весом будет отрицательной. Если бы в исследовании Americans' Changing Lives фигурировали, например, данные о «количестве миль, пробегаемых участником за один месяц», то я бы несколько не сомневался, что коэффициент по «количеству пробегаемых миль» будет отрицательным: чем большее количество миль вы ежемесячно пробегаете, тем меньше ваш вес.

Величина. Насколько велика наблюдаемая нами зависимость между независимой и зависимой переменными? Можно ли считать ее величину существенной для нас? В рассматриваемом нами случае увеличение роста человека на дюйм ассоциируется с прибавкой веса на 4,5 фунта; в процентном выражении это значительная доля массы тела типичного человека. В объяснении того, почему одни люди весят больше, чем другие, рост, несомненно, является важным фактором. В других исследованиях мы можем обнаружить объясняющую переменную, которая оказывает статистически значимое влияние на интересующий нас исход (это означает, что наблюдаемый эффект вряд ли объясняется чистой случайностью), но оно порой бывает настолько малым, что может считаться несущественным, или *незначимым*. Например, допустим, что мы исследуем определяющие факторы дохода. Объясняющими переменными здесь могут быть образование, стаж работы и т. п. При использовании достаточно крупного набора данных ученые также могут прийти к выводу, что люди с более белыми зубами зарабатывают на 86 долларов в год больше, чем остальные работники, *ceteris paribus*. (*Ceteris paribus* по-латыни означает «при прочих равных условиях».) Положительный и статистически значимый коэффициент по переменной «белые зубы» предполагает, что те, кого мы сравниваем, в остальном (по уровню образования, рабочему стажу и т. п.) не различаются между собой. (Ниже я объясню, каким образом мы можем выполнить это условие.) Наш статистический анализ продемонстрировал, что более белые зубы ассоциируются с 86-долларовой прибавкой к годовому доходу и что этот эффект вряд ли объясняется чистой случайностью. Это означает, что 1) мы с достаточно высокой степенью уверенности отвергли основную (нулевую) гипотезу, гласящую, что наличие у человека белых зубов никак не связано с уровнем его годового дохода; и 2) если мы проанализируем другие выборки данных, то наверняка обнаружим аналогичную связь между хорошо выглядящими зубами и повышенным уровнем дохода.

Что же из этого следует? Мы выявили статистически значимый результат, хотя для нас он практически бесполезен. Начнем с того, что прибавка в 86 долларов к годовому доходу вряд ли существенно изменит уровень жизни человека. С экономической точки зрения она вряд ли оправдывает регулярное выполнение процедур по отбеливанию зубов, поскольку такие процедуры наверняка обойдутся гораздо дороже, поэтому нам не имеет смысла рекомендовать подобные инвестиции молодым работникам. И, несколько

забегая вперед, я озабочился бы также рядом серьезных методологических проблем. Например, идеальный вид зубов может ассоциироваться с другими чертами характера человека, обуславливающими более высокий уровень его доходов: то есть дело не в зубах как таковых, а в том, что люди с высоким уровнем доходов, как правило, заботятся об их состоянии. Пока же для нас важно обратить внимание на степень (величину) наблюдаемой нами связи между объясняющей переменной и интересующим нас исходом.

Значимость. Является ли наблюдаемый нами результат заблуждением, обусловленным нерепрезентативной выборкой данных, или он отражает реально существующую связь, которая, скорее всего, будет присуща всей соответствующей совокупности? Это тот же самый фундаментальный вопрос, на который мы пытаемся ответить на протяжении нескольких последних глав. Можно ли ожидать в контексте роста и веса, что мы будем наблюдать аналогичную положительную ассоциацию в других выборках, которые являются репрезентативными по отношению к данной совокупности? Чтобы ответить на этот вопрос, используем уже знакомые вам базовые инструменты статистического вывода. Наш коэффициент регрессии основывается на наблюдаемой зависимости между ростом и весом для определенной выборки данных. Если бы мы тестировали более крупную выборку, то почти наверняка выявили бы несколько иную зависимость между ростом и весом и, следовательно, другой коэффициент регрессии. Зависимость между ростом и весом, наблюдаемая в данных, полученных британским правительством (напоминаю, что они касаются государственных служащих Британии), безусловно, будет отличаться от зависимости между ростом и весом для участников исследования *Americans' Changing Lives*. Однако из центральной предельной теоремы следует, что среднее значение для большой, надлежащим образом сформированной выборки, как правило, не будет существенно отклоняться от среднего значения для генеральной совокупности. Аналогично мы можем предположить, что наблюдаемая зависимость между переменными, такими как рост и вес, тоже не будет значительно разниться от выборки к выборке, если, конечно, эти выборки будут достаточно крупными и надлежащим образом сформированными из одной и той же совокупности.

Вы должны понимать это на интуитивном уровне. Весьма маловероятно (хотя в принципе возможно), что, обнаружив зависимость между каждым дополнительным дюймом роста и дополнительными 4,5 фунта веса участников исследования *Americans' Changing Lives*, мы в то же время не выявили бы

никакой зависимости между ростом и весом в какой-то другой репрезентативной выборке, состоящей из 3000 взрослых американцев.

Это должно дать вам первый намек на то, как мы будем проверять, являются ли результаты нашей регрессии статистически значимыми. Для коэффициента регрессии, как и для опросов общественного мнения и других форм статистического вывода, мы можем вычислить стандартную ошибку, которая представляет собой показатель вероятного разброса, наблюдаемый нами в значениях этого коэффициента в случае, если бы мы выполнили регрессионный анализ по нескольким выборкам, сформированным из одной и той же совокупности. Если бы мы измерили рост и вес в какой-то другой выборке, состоящей из 3000 взрослых американцев, то последующий анализ мог бы показать, что каждый дополнительный дюйм роста ассоциируется с дополнительными 4,3 фунта веса. Если бы мы проделали те же самые действия в отношении еще одной выборки из 3000 взрослых американцев, то могли бы обнаружить, что каждый дополнительный дюйм роста связан с дополнительными 5,2 фунта веса. И здесь на помощь снова приходит нормальное распределение. При использовании больших выборок данных можно предположить, что полученные нами разные коэффициенты регрессии будут распределены по нормальному закону вблизи «истинной» зависимости между ростом и весом в совокупности взрослых американцев. В таком предположении мы можем вычислить стандартную ошибку для коэффициента регрессии, что позволит составить представление о том, насколько большой разброс коэффициентов регрессии следует ожидать от выборки к выборке. Я не буду здесь вдаваться в подробное объяснение формулы для вычисления стандартной ошибки, поскольку для этого пришлось бы прибегнуть к множеству математических выкладок и к тому же все базовые статистические пакеты программного обеспечения вычислят ее за вас.

Однако должен предупредить, что при использовании небольшой выборки данных — например группы из 20 взрослых американцев вместо группы из более чем 3000 участников исследования *Americans' Changing Lives* — нормальное распределение на помощь нам уже не придет. В частности, если мы будем то и дело выполнять регрессионный анализ в отношении разных малых выборок, то уже не сможем исходить из того, что полученные нами разные коэффициенты регрессии будут распределены по нормальному закону вблизи «истинной» зависимости между ростом и весом в совокупности взрослых американцев. Вместо этого они будут распределены вблизи

«истинной» зависимости между ростом и весом в совокупности взрослых американцев по закону, известному как *t*-распределение, или распределение Стьюдента. (Вообще говоря, *t*-распределение характеризуется большей степенью разброса, чем нормальное распределение, и, следовательно, имеет «более толстые хвосты».) Все прочее остается неизменным; любые базовые статистические пакеты программного обеспечения без проблем справятся с дополнительной сложностью, связанной с использованием *t*-распределений. Поэтому более подробное объяснение *t*-распределения приведено в приложении к этой главе.

Пока же будем исходить из того, что имеем дело с большими выборками (и с нормальным распределением). Самое главное сейчас — понять, почему для нас так важна стандартная ошибка. Как и в случае с опросами общественного мнения и другими формами статистического вывода, мы ожидаем, что более половины наблюдаемых коэффициентов регрессии будут отстоять от истинного параметра* совокупности на расстояние, не превышающее одной стандартной ошибки. Примерно 95% коэффициентов регрессии будут отстоять от истинного параметра совокупности на расстояние, не превышающее двух стандартных ошибок. И так далее. Учитывая сказанное, можно считать, что мы почти у цели, так как теперь можем выполнить небольшую проверку гипотез. (А вы и в самом деле полагали, что с проверкой гипотез покончено?) Поскольку у нас уже есть коэффициент и стандартная ошибка, мы можем проверить основную гипотезу, которая заключается в том, что между объясняющей и зависимой переменной на самом деле никакой зависимости нет (а это, в свою очередь, означает, что истинная зависимость между ними в данной совокупности равна нулю).

В нашем простом примере с ростом и весом мы можем проверить, какова вероятность обнаружить, что в выборке *Americans' Changing Lives* каждый дополнительный дюйм роста ассоциируется с 4,5 дополнительными фунта веса, если на самом деле во всей совокупности зависимость между ростом и весом отсутствует. Я вычислил соответствующую регрессию, воспользовавшись одним из распространенных статистических пакетов; стандартная ошибка по коэффициенту роста составила 0,13. Это означает, что в случае

* «Параметр» — это термин, обозначающий любую статистику, которая описывает ту или иную характеристику какой-либо совокупности; средний вес для всех взрослых мужчин — параметр соответствующей совокупности. То же можно сказать о среднеквадратическом отклонении. В приведенном примере истинная связь между ростом и весом для данной совокупности является параметром этой совокупности.

многократного выполнения такого анализа (скажем, с сотней разных выборок) можно было бы ожидать, что наш наблюдаемый коэффициент регрессии будет отстоять от истинного параметра совокупности на расстояние, не превышающее двух стандартных ошибок, примерно в 95 случаях из 100.

Следовательно, это позволяет нам выразить полученные результаты двумя разными, но взаимосвязанными между собой способами. Первый — это построить 95%-ный доверительный интервал. Мы можем утверждать, что в 95 случаях из 100 доверительный интервал (который составляет $4,5 \pm 0,26$) будет включать истинный параметр совокупности. Это диапазон от 4,24 до 4,76. Любой из статистических пакетов также вычислит этот интервал. Второй — отвергнуть основную гипотезу об *отсутствии* зависимости между ростом и весом для совокупности в целом на 95%-ном доверительном уровне, видя, что наш 95%-ный доверительный интервал для истинной зависимости между ростом и весом не включает нуль. Этот результат можно также выразить как статистически значимый на уровне 0,05: существует лишь 5%-ная вероятность того, что мы ошибочно отвергли основную гипотезу.

На самом деле наши результаты еще более убедительны, чем кажется на первый взгляд. Стандартная ошибка (0,13) очень мала по сравнению с величиной коэффициента (4,5). Практика показывает, что этот коэффициент можно считать статистически значимым, когда его величина по меньшей мере в два раза превышает величину стандартной ошибки*. Любой из базовых статистических пакетов также вычисляет р-значение, которое в данном случае равняется 0,000; это означает, что если в действительности зависимости между ростом и весом в совокупности в целом нет, то вероятность получить столь необычный результат, какой нам удалось наблюдать, по сути, равна нулю. Не забывайте, что мы вовсе не *доказали*, что более рослые люди весят больше во всей совокупности, а лишь показали, что если бы это было не так, то наши результаты для выборки Americans' Changing Lives были бы крайне маловероятными.

Базовый регрессионный анализ дает еще одну статистику, заслуживающую внимания, R^2 , которая предстает собой показатель суммарной

* Когда нулевая гипотеза заключается в том, что коэффициент регрессии равняется нулю (а это имеет место в большинстве случаев), отношение наблюдаемого коэффициента регрессии к стандартной ошибке называется *t-статистикой*. Это также объясняется в приложении к данной главе.

величины разброса, объясняемого уравнением регрессии*. Нам известно, что в выборке Americans' Changing Lives наблюдается широкий разброс веса. Многие члены выборки весят больше среднего веса для данной группы в целом; многие — меньше. Величина R^2 говорит нам, какая доля этого разброса вокруг среднего значения ассоциируется лишь с различиями в росте. В нашем случае эта доля составляет 0,25, или 25%. Более значимым может быть то обстоятельство, что 75% этого разброса в весе для нашей выборки остаются необъясненными. Есть очевидные факторы, помимо роста, которые могут нам помочь их объяснить. Ситуация становится интереснее.

В начале этой главы я объявил регрессионный анализ чудодейственным эликсиром для социальных исследований. До сих пор я использовал некий базовый статистический пакет и впечатляющие данные, чтобы продемонстрировать тот факт, что рослые люди, как правило, весят больше коротышек. Краткая прогулка по какому-нибудь супермаркету наверняка убедила бы вас в том же. Теперь пора оценить реальные возможности регрессионного анализа. Иными словами, пора пересаживаться с детского трехколесного велосипеда на велосипед для взрослых!

Как я уже говорил, регрессионный анализ позволяет распутывать сложные взаимосвязи, в которых многие факторы оказывают влияние на интересующий нас исход, например доход, или результаты экзамена, или развитие сердечно-сосудистых заболеваний. Когда мы включаем в уравнение регрессии несколько переменных, анализ дает оценку линейной зависимости между *каждой* объясняющей и зависимой переменной, оставляя при этом неизменными другие зависимые переменные (то есть «контролируя» их). Давайте на какое-то время сосредоточимся на весе. Мы выявили зависимость между ростом и весом, а также знаем о существовании других факторов (возраст, пол, режим питания, занятия спортом и т. п.), которые могут помочь объяснить вес. Посредством регрессионного анализа (часто называемого множественным регрессионным анализом, если в нем задействовано несколько объясняющих переменных, или многофакторным регрессионным анализом) можно вычислить некий коэффициент регрессии для каждой объясняющей переменной, задействованной в уравнении регрессии. Скажем, какова зависимость между возрастом и весом *среди людей одного и того же пола и роста*. Когда нам приходится иметь дело с несколькими

* В статистике этот показатель называется коэффициентом детерминации. *Прим. ред.*

объясняющими переменными, соответствующие данные уже невозможно отобразить на двумерной диаграмме. (Попытайтесь представить себе диаграмму, которая отображает вес, пол, рост и возраст каждого участника исследования Americans' Changing Lives.) Тем не менее базовая методология остается той же, что и в примере с ростом и весом. При добавлении объясняющих переменных статистический пакет будет вычислять коэффициенты регрессии, которые минимизируют общую сумму квадратов разностей для соответствующего уравнения регрессии.

Пока ограничимся данными исследования Americans' Changing Lives, а затем я вернусь и предложу интуитивно понятное объяснение того, как действует этот механизм. Мы можем начать с добавления в уравнение регрессии еще одной переменной, которая объясняет вес участников Americans' Changing Lives, — «возраст». Когда мы вычислим уравнение регрессии, включающее рост и возраст в качестве объясняющих переменных, то получим вот что:

$$\text{Вес} = -145 + 4,6 \times (\text{Рост в дюймах}) + 0,1 \times (\text{Возраст в годах})$$

Коэффициент возраста равняется 0,1. Это можно интерпретировать так: каждый дополнительный год к возрасту человека ассоциируется с 0,1 дополнительных фунта к весу человека *при неизменном росте*. Для любой группы людей одного и того же роста те, кто на десять лет старше, весят в среднем на один фунт больше. Как видим, влияние возраста на вес человека не так уж велико, но это соответствует тому, что мы обычно наблюдаем в реальной жизни. Данный коэффициент является значимым на уровне 0,05.

Возможно, вы заметили, что коэффициент для роста несколько увеличился. После того как мы включили в нашу регрессию возраст, у нас появилось уточненное понимание зависимости между ростом и весом. Среди людей одного возраста в выборке (иными словами, при фиксированном возрасте) каждый дополнительный дюйм роста ассоциируется с дополнительными 4,6 фунта веса.

Теперь давайте добавим еще одну переменную — пол. Тут есть один нюанс: пол может принимать лишь два значения (мужской и женский). Как вставить эти «М» и «Ж» в регрессию? Благодаря использованию так называемой двоичной, или фиктивной переменной. Вводим в нашей совокупности данных 1 для участников-женщин и 0 — для участников-мужчин. (Дорогие мужчины, пожалуйста, не обижайтесь!) При этом коэффициент пола

можно интерпретировать как влияние на вес того обстоятельства, что данный участник является женщиной — при прочих равных условиях (*ceteris paribus*). Этот коэффициент составляет $-4,8$, что не должно вызывать у вас удивления. Это можно истолковать так: когда речь идет об участниках одного и того же роста и возраста, женщины обычно весят на 4,8 фунта меньше мужчин. Теперь вам уже должны быть в какой-то мере ясны богатые возможности множественного регрессионного анализа. Нам известно, что женщины обычно ниже мужчин, и наш коэффициент учитывает это обстоятельство, поскольку мы уже контролируем рост (мы его «зафиксировали»). В данном случае мы рассматриваем влияние пола — точнее говоря, женского пола. Новая регрессия принимает следующий вид:

$$\text{Вес} = -118 + 4,3 \times (\text{Рост в дюймах}) + \\ + 0,12 \times (\text{Возраст в годах}) - 4,8 (\text{Если пол женский})$$

Наша «наилучшая» оценка веса пятидесятитрехлетней женщины, рост которой равен 5 футов и 5 дюймов, такова: $-118 + 4,3 \times 65 + 0,12 \times 53 - 4,8 = 163$ фунта.

Наша «наилучшая» оценка веса тридцатипятилетнего мужчины, рост которого составляет 6 футов и 3 дюйма, такова: $-118 + 4,3 \times 75 + 0,12 \times 35 = 209$ фунтов. Мы опускаем последний член ($-4,8$) при вычислении результата регрессии, поскольку рассматриваемый нами человек не является женщиной.

Теперь давайте приступим к проверке более интересных и менее предсказуемых вещей. Что можно сказать по поводу образования? Как оно может влиять на вес? Я бы выдвинул гипотезу, что более образованные люди в большей степени заботятся о своем здоровье и, следовательно, весят меньше. Кроме того, мы еще не проверяли влияние занятий спортом; я полагаю, что при прочих равных условиях члены нашей выборки, регулярно занимающиеся спортом, весят меньше.

А что можно сказать по поводу бедности? Не сказываются ли низкие доходы части американцев на их весе? В исследовании *Americans' Changing Lives* есть вопрос о том, получает ли его участник продовольственные талоны. (Продовольственные талоны в Соединенных Штатах выдаются только малоимущим гражданам.) Наконец, меня интересует расовая принадлежность человека. Нам известно, что люди разных рас в США имеют разный жизненный опыт *именно вследствие своей расовой принадлежности*. С той

или иной расой в Соединенных Штатах ассоциируются определенные культурные факторы и места компактного проживания. Все эти факторы могут оказывать влияние на вес человека. Многие города Америки характеризуются высокой степенью расовой сегрегации: афроамериканцы чаще других американских граждан проживают в так называемых продовольственных пустынях, то есть территориях с ограниченным доступом к продовольственным магазинам, где продаются свежие фрукты, овощи и другая свежая продукция.

Регрессионный анализ можно использовать для обособления независимого влияния каждого из потенциальных объясняющих факторов, описанных выше. Например, мы можем вычленировать связь между расовой принадлежностью и весом человека, сохраняя постоянными другие социально-экономические факторы, такие как уровень образования и бедность. *Существует ли статистически достоверная связь между весом человека и его принадлежностью к негроидной расе, если речь идет о людях, окончивших среднюю школу и имеющих право на получение продовольственных талонов?*

В данном случае уравнение регрессии окажется таким длинным, что было бы весьма проблематично привести его здесь полностью. Научные статьи обычно включают огромные таблицы, обобщающие результаты разных уравнений регрессии. В приложении к этой главе вы найдете таблицу с полными результатами этого уравнения регрессии. Между тем, я могу подсказать, что произойдет, если мы добавим в уравнение такие факторы, как уровень образования человека, его склонность к занятиям спортом, показатель бедности (исходя из которого определяется его право на получение продовольственных талонов) и расовая принадлежность.

Все наши исходные переменные (рост, возраст и пол) по-прежнему остаются значимыми. При добавлении объясняющих переменных несколько изменяются коэффициенты. Новые переменные являются статистически значимыми на уровне 0,05. Значение R^2 для этой регрессии повысилось с 0,25 до 0,29. (Вспомните: нулевая величина R^2 означает, что уравнение регрессии прогнозирует вес любого человека в данной выборке ничуть не лучше, чем среднее значение; если же R^2 равно 1, то наше уравнение регрессии идеально прогнозирует вес каждого человека в данной выборке.) Существенная доля разброса величин веса среди членов данной выборки остается необъясненной.

Как я и предполагал, зависимость между образованием и весом человека оказалась отрицательной. Среди участников исследования Americans' Changing Lives каждый дополнительный год образования ассоциируется с –1,3 фунта веса.

Неудивительно, что физические упражнения также отрицательно связаны с весом человека. Исследование Americans' Changing Lives включает индекс, который оценивает каждого участника исследования с точки зрения уровня его физической активности. Те, кто находится в нижнем квинтиле* склонности к регулярным занятиям спортом, весят в среднем на 4,5 фунта больше, чем другие взрослые в этой выборке, *ceteris paribus*. И примерно на 9 фунтов больше, чем взрослые в верхнем квинтиле склонности к регулярным занятиям спортом.

Вес тех, кто получает продовольственные талоны (что служит показателем бедности в этой регрессии), больше, чем у других взрослых. Получатели продовольственных талонов весят в среднем на 5,6 фунта больше, чем другие участники исследования Americans' Changing Lives, ceteris paribus.

Переменная расовой принадлежности представляет особый интерес. Даже если мы зафиксируем все остальные вышеперечисленные переменные, расовая принадлежность сыграет довольно важную роль в объяснении веса. Неиспаноязычные взрослые негроидной расы в выборке Americans' Changing Lives весят в среднем примерно на 10 фунтов больше, чем другие взрослые в выборке. Десять фунтов — весьма существенная прибавка в весе как в абсолютном выражении, так и по сравнению с влиянием других объясняющих переменных в нашем уравнении регрессии. И это вовсе не какой-то случайный «выверт» данных. *p*-значение по фиктивной переменной для неиспаноязычных взрослых негроидной расы равняется 0,000, а 95%-ный доверительный интервал охватывает величины веса от 7,7 фунта до 16,1 фунта.

Что же происходит? Честно говоря, не имею понятия. Могу лишь повторить замечание, сделанное мною выше в одной из сносок: я лишь экспериментирую с данными, чтобы проиллюстрировать принцип действия

* Квинтиль — это квантиль порядка 0,2. Если выборочные значения организовать в порядке возрастания, то квинтили делят эту выборку на пять равных (по количеству) частей. В данном случае «нижний квинтиль склонности к регулярным занятиям спортом» — это группа наименее склонных к регулярным занятиям спортом, составляющая пятую часть из совокупности лиц, регулярно им занимающихся.
Прим. ред.

регрессионного анализа. Представленные здесь аналитические материалы призваны подтвердить результаты научного исследования значения дворового хоккея для НХЛ. (Шутка.) Если бы это был реальный исследовательский проект, то для подтверждения правильности его выводов понадобились бы недели и даже месяцы аналитической работы. Могу лишь сказать, что я продемонстрировал вам, почему множественный регрессионный анализ — лучший из имеющихся в нашем распоряжении инструмент для поиска существенных закономерностей в больших и сложных совокупностях данных. Мы начали со смехотворно банального упражнения: поиска численного выражения связи между ростом и весом, а затем перешли к рассмотрению вопросов, имеющих реальное социальное значение.

В этом ключе я могу предложить вам реальное исследование, в котором регрессионный анализ использовался для решения социально значимой проблемы — дискриминации по половому признаку на рабочем месте. Такую дискриминацию, как правило, трудно наблюдать непосредственно. Никто из работодателей не скажет вам напрямую, что тому или иному работнику платят меньше только по причине его расовой или половой принадлежности или что кого-то не приняли на работу по каким-либо дискриминационным соображениям (в результате чего этот человек, наверное, нашел другую работу, но с более низкой заработной платой). Однако на практике мы наблюдаем различия в зарплате по расовому или половому признаку, которые могут быть следствием дискриминации: белые зарабатывают больше, чем черные; мужчины — больше, чем женщины, и т. д. Методологическая проблема заключается в том, что эти различия могут также оказаться результатом других различий между работниками, которые не имеют ничего общего с дискриминацией (например, женщины зачастую предпочитают работать неполный рабочий день). В какой мере имеющаяся разница в оплате труда обусловлена факторами, связанными с производительностью на работе, а в какой — с дискриминацией работников (если таковая вообще присутствует)? Никто не станет утверждать, что этот вопрос относится к разряду тривиальных.

Регрессионный анализ может помочь нам на него ответить. Однако в этом случае наша методология будет несколько более «окольной», чем в примере с анализом, объясняющим вес. Поскольку дискриминация не поддается непосредственному измерению, нам придется исследовать другие факторы (например образование, производственный стаж, род занятий и т. п.),

которые традиционно объясняют уровень заработной платы. Мы можем действовать методом исключения: если после фиксации этих факторов все же останется существенная разница в зарплате, то дискриминация на работе, по-видимому, имеет место. Чем больше необъясненная доля разницы в заработной плате, тем сильнее подозрения в наличии дискриминации на рабочем месте. Рассмотрим статью трех экономистов, исследующих траектории заработной платы в выборке, состоящей примерно из 2500 мужчин и женщин — выпускников Booth School of Business Чикагского университета (все они обладатели степени MBA)¹. Сразу после выпуска средний начальный уровень заработной платы у мужчин и женщин приблизительно одинаков: 130 000 долларов у мужчин и 115 000 долларов у женщин. Однако через десять лет образуется огромный разрыв: женщины в среднем зарабатывают на целых 45% меньше, чем их бывшие однокурсники-мужчины: 243 000 долларов против 442 000 долларов. В более широкой выборке, включающей свыше 18 000 выпускников (обладающих степенью MBA), которые приступили к работе в период с 1990 по 2006 год, у женщин на 29% ниже зарплаты, чем у мужчин. Что же происходит с женщинами, после того как они выходят на рынок труда?

Согласно авторам данного исследования (Марианна Бертран из Booth School of Business, Клаудиа Голдин и Лоуренс Кац из Гарвардского университета), дискриминация не является вероятным объяснением большей доли разрыва в зарплатах. Причем разрыв по половому признаку исчезает, когда авторы добавляют в анализ дополнительные объясняющие переменные. Например, при прохождении программы MBA мужчины посещают дополнительные курсы финансов и на выпускных экзаменах получают в среднем более высокие оценки. Когда эти данные используются в уравнении регрессии в качестве управляющих переменных, необъясненная доля разрыва в уровнях зарплаты мужчин и женщин снижается до 19%. Когда же в это уравнение включаются переменные, позволяющие учитывать рабочий стаж после окончания университета, необъясненная доля разрыва в уровнях зарплаты мужчин и женщин снижается до 9%. А когда в уравнение добавляются объясняющие переменные для других характеристик (например, тип работодателя и количество реально отработанных часов), необъясненная доля разрыва в уровнях зарплаты мужчин и женщин снижается до менее 4%.

Что касается работников, стаж которых превышает десять лет, то авторы исследования могут в конечном счете объяснить все, кроме 1%-ного разрыва

в уровнях зарплаты мужчин и женщин, факторами, не имеющими никакого отношения к дискриминации на работе*. Авторы пришли к следующему выводу: «Мы выявили три непосредственные причины существования большого увеличивающегося разрыва в уровнях зарплаты мужчин и женщин: разница в уровнях знаний, полученных в высшем учебном заведении; разница, обусловленная большими перерывами в стаже у женщин; разница в количестве реально отработываемых часов в неделю. Эти три детерминанта могут объяснить львиную долю разрыва в уровнях зарплаты мужчин и женщин по окончании ими вуза и после начала трудовой деятельности».

Я надеюсь, что убедил вас в полезности множественного регрессионного анализа, особенно в возможности делать выводы по результатам исследований путем обособления влияния какой-то одной объясняющей переменной и фиксации («контроля») других факторов, способных вносить искажения в выводы. Я еще не предложил вам интуитивно понятного объяснения того, как этот статистический «волшебный эликсир» работает. Когда мы используем регрессионный анализ для оценивания зависимости между образованием и весом человека, *ceteris paribus*, как применяемый нами статистический пакет контролирует такие факторы, как рост, пол, возраст и доход, когда нам доподлинно известно, что участники исследования *Americans' Changing Lives* вовсе не идентичны в других отношениях?

Чтобы уяснить, каким образом можно изолировать влияние на вес какой-либо отдельно взятой переменной, например образования, давайте представим следующую ситуацию. Допустим, что все участники исследования *Americans' Changing Lives* собрались в каком-то одном месте, например во Фрамингеме. Теперь предположим, что мы отделили мужчин от женщин, а затем распределили их по росту. В одном помещении собрали всех мужчин, рост которых равняется шести футам; в соседнем — рост которых равняется шести футам и одному дюйму и т. д. для представителей обоих полов. Если в нашем исследовании участвует достаточно много людей, мы можем разбить их на группы по уровню дохода и распределить по разным комнатам. В каждой комнате будут находиться люди, идентичные во всех отношениях,

* Более широкие силы дискриминационного характера могут влиять на выбор женщинами той или иной служебной карьеры или на тот факт, что женщинам гораздо чаще, чем мужчинам, приходится брать отпуск по уходу за детьми. Однако эти важные вопросы не следует путать с более узким вопросом, платят ли женщинам меньше, чем мужчинам, за одну и ту же работу.

за исключением образования и веса, которые и являются двумя интересующими нас переменными. В результате описанного распределения обязательно окажется комната, где соберутся сорокапятилетние мужчины ростом 5 футов и 5 дюймов, годовой доход которых составляет от 30 000 до 40 000 долларов. В соседней комнате будут находиться сорокапятилетние женщины ростом 5 футов и 5 дюймов и годовым доходом от 30 000 до 40 000 долларов. И так далее.

В каждой комнате все же будет наблюдаться некоторый разброс величин веса: вес людей одного пола и роста, имеющих примерно одинаковый доход, будет разным, хотя, наверное, в этом случае эта разница будет гораздо меньшей, чем в выборке в целом. Сейчас наша цель — увидеть, какую долю остающегося разброса величин веса в каждой комнате можно объяснить уровнем образования. Иными словами, какова «наилучшая» линейная связь между образованием и весом в каждой комнате?

Конечная проблема, однако, заключается в том, что мы не хотели бы использовать разные коэффициенты для каждой комнаты. Весь смысл этого упражнения — рассчитать единственный коэффициент, который бы наилучшим образом отражал связь между образованием и весом для рассматриваемой нами выборки в целом — при неизменности других факторов. Мы хотели бы определить единый коэффициент для образования, который можно было бы использовать *в каждой комнате*, чтобы минимизировать сумму квадратов разностей для совокупности всех комнат. Какой коэффициент для образования минимизирует квадрат необъясненного веса для каждого человека по всем комнатам? Этот коэффициент становится нашим коэффициентом регрессии, поскольку является наилучшим объяснением линейной зависимости между образованием и весом для данной выборки при неизменности таких факторов, как пол, рост и доход.

Данный пример позволяет понять, почему так полезны большие совокупности данных. Они дают нам возможность контролировать многие факторы, располагая при этом большим количеством наблюдений в каждой «комнате». Очевидно, компьютер может выполнить соответствующие вычисления буквально за доли секунды, не распределяя тысячи людей по разным комнатам.

Завершу главу тем же, с чего начал, — зависимостью между стрессом на работе и развитием сердечно-сосудистых заболеваний. Цель исследований, выполняемых по заказу британского правительства в отношении государственных служащих, заключалась в том, чтобы определить связь между

невозможностью человека в достаточной степени контролировать содержание, способы и условия выполнения своей работы и развитием сердечно-сосудистых заболеваний за определенный период времени. В ходе одного из первых исследований, проводившегося на протяжении семи с половиной лет, использовалась выборка из 17 530 государственных служащих². Авторы исследования пришли к следующему заключению: «Служащие (мужчины) низшего ранга, как правило, ниже ростом, полнее, имеют проблемы с артериальным давлением, больше курят и меньше занимаются спортом, чем чиновники более высоких рангов. Даже после внесения поправки, учитывающей влияние на уровень смертности всех этих факторов плюс содержание холестерина в крови, отрицательная закономерность между рангом госслужащего и уровнем смертности от сердечно-сосудистых заболеваний оставалась достаточно сильной». Упомянутая «поправка» вносится посредством регрессионного анализа*. Результаты исследования демонстрируют, что при фиксации остальных факторов здоровья (включая рост, который является надежным показателем здоровья и качества питания в раннем детстве) работа на «низких» должностях может в буквальном смысле вас убить.

Скептицизм — вполне разумная первая реакция. В начале главы я написал, что невозможность человека в достаточной степени влиять на содержание, способы и условия выполнения своей работы отрицательно сказывается на его здоровье. Это может быть (или не быть) синонимом пребывания работника на нижних ступенях административной иерархии. Дальнейшее исследование, в ходе которого использовалась вторая выборка

* Эти исследования несколько отличаются от уравнений регрессии, о которых рассказывалось выше в настоящей главе. В этих исследованиях интересующий нас исход, или независимая переменная, являются двоичными. За время исследования у его участника либо возникло то или иное заболевание сердца, либо нет. Таким образом, исследователи используют инструмент под названием многомерная логистическая регрессия. основополагающая идея остается такой же, как и в случае обычных моделей наименьших квадратов, описанных в настоящей главе. Каждый коэффициент выражает влияние конкретной объясняющей переменной на зависимую переменную при неизменности влияния других переменных в данной модели. Ключевая разница заключается в том, что все переменные в нашем уравнении влияют на вероятность наступления некоторого события, например на вероятность сердечного приступа за период проведения исследования. Например, в этом исследовании вероятность возникновения за период его проведения каких-либо проблем с сердцем у работников, входящих в состав контрольной группы с низкими должностями, в 1,99 раза выше, чем у работников, входящих в состав контрольной группы с высокими должностями, после фиксации всех остальных «сердечных факторов риска».

из 10 308 британских государственных служащих, было призвано более глубоко уяснить эту разницу³. Работников еще раз разделили на административные ранги — высокий, промежуточный и низкий, — но на сей раз предложили заполнить анкету из пятнадцати пунктов, чтобы оценить уровень «диапазона принятия решений или контроля» работника. Анкета содержала вопросы типа: «Можете ли вы выбирать, как именно будете выполнять порученную вам работу?»; кроме того, предлагались разные варианты ответа (от «никогда» до «часто») на утверждения наподобие: «Я могу самостоятельно решать, когда устроить себе перерыв». Исследователи пришли к выводу, что за время проведения эксперимента у работников с «низким уровнем контроля» риск развития сердечно-сосудистых заболеваний был значительно выше, чем у работников с «высоким уровнем контроля». Вместе с тем ученые обнаружили, что риск развития сердечно-сосудистых заболеваний у служащих с жесткими требованиями к выполняемой работе ничуть не выше, чем у работников с низким уровнем социальной поддержки на работе. Похоже, что невозможность человека в достаточной степени влиять на содержание, способы и условия выполнения поставленных задач убивает его в буквальном смысле этого слова.

Упомянутое нами исследование британских служащих обладает двумя характеристиками, типичными для таких солидных экспериментов. Во-первых, его результаты подтверждены аналогичными исследованиями в других странах. В медицинской литературе представление о «низком контроле» (то есть недостаточной возможности человека влиять на содержание, способы и условия выполнения своей работы) привело к появлению термина «переутомление на работе», который характеризует должности с «высокой психологической нагрузкой» и «недостаточностью полномочий для принятия решений». В период с 1981 по 1993 год были опубликованы результаты тридцати шести исследований по этому вопросу; в большинстве из них найдена значительная положительная взаимосвязь между переутомлением на работе и развитием сердечно-сосудистых заболеваний⁴.

Во-вторых, исследователи выявили дополнительные биологические свидетельства, объясняющие механизм, посредством которого этот особый вид стресса на работе приводит к ухудшению здоровья работника. Условия работы, предусматривающие строгие требования, но не позволяющие человеку влиять на процесс выполнения поставленных задач, могут вызывать физиологические реакции (например выделение гормонов, связанных

со стрессом), повышающие риск развития сердечно-сосудистых заболеваний в долгосрочной перспективе. Раскрыть этот механизм помогают даже опыты над животными: у обезьян и павианов, занимающих низкий статус (и имеющих немало общего с мелкими государственными служащими), есть физиологические отличия от их высокостатусных сородичей, причем эти отличия обуславливают их бóльшую склонность к сердечно-сосудистым заболеваниям⁵.

При прочих равных условиях лучше, конечно, не становиться низкостатусным павианом (именно эту мысль я пытаюсь как можно чаще доносить до сознания своих детей — особенно сына). Более значительный месседж заключается в том, что регрессионный анализ, пожалуй, — самый важный из имеющихся в распоряжении исследователей инструментов для поиска значимых закономерностей и связей в крупных совокупностях данных. Как правило, у нас нет возможности проводить управляемые эксперименты для получения данных о дискриминации на работе или выявления факторов, вызывающих развитие сердечно-сосудистых заболеваний. Источником наших представлений об этих и многих других социально значимых проблемах являются статистические инструменты, о которых шла речь в этой главе. В сущности, не будет преувеличением сказать, что значительная часть всех важных исследований, выполненных в области социальных наук за последние полстолетия (особенно после появления сравнительно недорогих компьютеров), проводилась с применением регрессионного анализа.

Регрессионный анализ представляет собой важную разновидность научного метода исследований; благодаря ему мы стали более здоровыми, защищенными и информированными людьми.

Какие же потенциальные ловушки подстерегают нас при использовании столь мощного и впечатляющего инструмента? Об этом я расскажу в следующей главе.

Приложение к главе 11

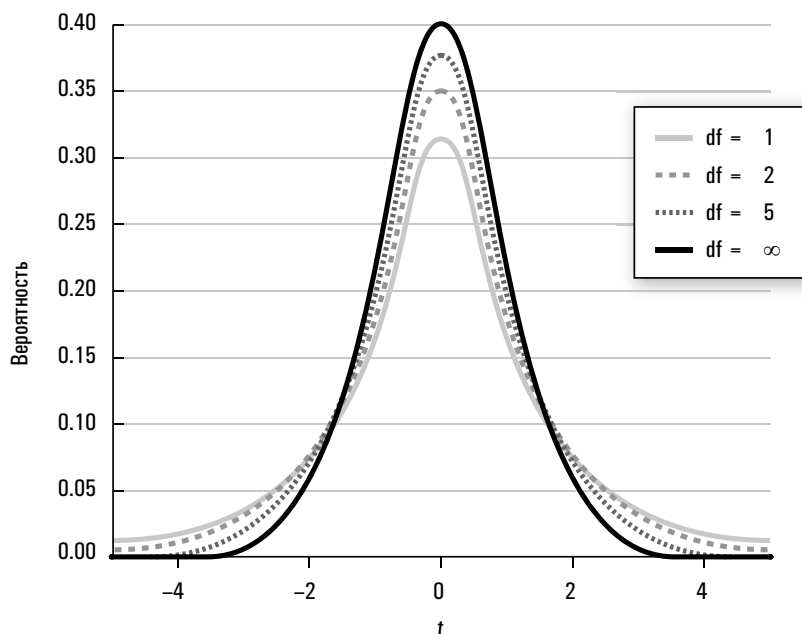
t-распределение

Жизнь несколько усложняется при выполнении регрессионного анализа (или других видов статистического вывода) с малой выборкой данных. Допустим, нам нужно проанализировать зависимость между весом и ростом на основе выборки, состоящей всего из 25 взрослых, вместо того чтобы использовать огромный набор данных, как в исследовании *Americans' Changing Lives*. Логика подсказывает, что надо с меньшей уверенностью обобщать полученные результаты на все взрослое население, если выборка состоит не из 3000 взрослых, а лишь из 25. Одно из положений, которые неоднократно подчеркивались в этой книге, заключается в том, что меньшие выборки, как правило, порождают больший разброс исходов. Выборка из 25 взрослых по-прежнему обеспечивает значимые результаты, как обеспечивала бы выборка из 10 и даже 5 человек, но насколько значимыми они являются?

На этот вопрос ответит *t*-распределение. При анализе зависимости между ростом и весом для нескольких выборок из 25 взрослых уже нельзя исходить из того, что разные коэффициенты регрессии, которые мы получаем, будут распределены по нормальному закону вблизи «истинного» коэффициента регрессии для взрослого населения в целом. Они по-прежнему будут распределяться вблизи «истинного» коэффициента для взрослого населения в целом, но *формой этого распределения уже не будет хорошо нам знакомая колоколообразная кривая нормального распределения*. Вместо этого мы должны предположить, что многие выборки, состоящие лишь из 25 взрослых, будут порождать больший разброс вблизи истинного коэффициента совокупности и, следовательно, это распределение будет с «более толстыми хвостами». А многие выборки из 10 взрослых будут порождать еще больший разброс и, соответственно, распределение с еще более толстыми хвостами. По сути, *t*-распределение представляет собой некую совокупность, или «семейство», функций плотности вероятности, которые варьируются в зависимости от величины выборки. В частности, чем больше данных содержится в выборке, тем больше «степеней свободы»* у нас имеется при определении подходящего распределения, которое служит нам эталоном для оценки

* Степень свободы и в русской статистической литературе обозначается как *df* (от англ. *degrees of freedom*). См. ниже диаграмму. *Прим. ред.*

результатов. Если вы решите изучать более продвинутый курс статистики, то узнаете, как именно вычисляются степени свободы; пока же можем считать, что они примерно равны количеству наблюдений в выборке. Например, регрессионный анализ с выборкой, размер которой составляет 10, и с единственной объясняющей переменной, имеет 9 степеней свободы. Чем больше степеней свободы, тем больше уверенность, что выборка представляет истинную совокупность, и тем «плотнее» будет распределение, как следует из приведенной ниже диаграммы*.



Когда число степеней свободы увеличивается, t -распределение сходится к нормальному распределению. Именно поэтому при работе с большими совокупностями данных вы можете использовать для соответствующих вычислений нормальное распределение.

t -распределение лишь добавляет определенные нюансы в тот же процесс статистического вывода, который мы неоднократно использовали в этой книге. Мы по-прежнему формулируем нулевую гипотезу, а затем проверяем ее на наблюдаемых нами данных. Если эти данные крайне маловероятны

* Для тех, кто еще не догадался: t -распределение — это распределение Стьюдента. В русской литературе чаще всего оно называется именно так. *Прим. ред.*

в случае правильности нулевой гипотезы, то она отвергается. Единственное, что изменяется при использовании t -распределения, — это основные вероятности для оценивания наблюдаемых исходов. Чем «толще» хвост у конкретного распределения вероятностей (например, t -распределение для восьми степеней свободы), тем больший разброс следует ожидать в наблюдаемых данных и, следовательно, тем меньше уверенность в правильности отказа от нулевой гипотезы.

Допустим, мы решаем уравнение регрессии и, согласно нулевой гипотезе, коэффициент при какой-то конкретной переменной равняется нулю. После того как мы получим результаты вычислений, мы могли бы рассчитать t -статистику, которая представляет собой отношение наблюдаемого коэффициента к стандартной ошибке для этого коэффициента*. Эта t -статистика затем оценивается с точки зрения величины выборки данных, для которой подходит t -распределение (поскольку именно это в значительной мере определяет число степеней свободы). Когда t -статистика достаточно велика, то есть наблюдаемый коэффициент далек от того, что предсказывает нулевая гипотеза, мы можем отвергнуть нулевую гипотезу на некотором уровне статистической значимости. Опять-таки это тот же самый базовый процесс статистического вывода, с которым мы неоднократно сталкивались в этой книге.

Чем меньше степеней свободы (и, следовательно, чем «толще» хвосты у соответствующего t -распределения), тем больше должна быть t -статистика, чтобы мы могли отвергнуть нулевую гипотезу на некотором заданном уровне статистической значимости. Если бы в описанном выше гипотетическом примере регрессии было четыре степени свободы, то нам понадобилось бы, чтобы t -статистика была не менее 2,13: только в этом случае мы могли бы отвергнуть нулевую гипотезу на доверительном уровне 0,05 (при использовании одностороннего критерия).

Если бы у нас было 20 000 степеней свободы (что вполне позволяет использовать нормальное распределение), то для того чтобы отвергнуть нулевую гипотезу на доверительном уровне 0,05 (при использовании того же одностороннего критерия), необходимо, чтобы t -статистика равнялась всего 1,65.

* Более общая формула для вычисления t -статистики имеет следующий вид:

$$t_b = \frac{b - b_0}{SE_b}$$

где b — наблюдаемый коэффициент, b_0 — нулевая гипотеза для этого коэффициента, а SE_b — стандартная ошибка для наблюдаемого коэффициента b .

Уравнение регрессии для веса

<i>Переменная</i>	<i>Коэф- фициент</i>	<i>Стан- дартная ошибка</i>	<i>t-ста- тис- тика</i>	<i>p-значение (двусто- ронний критерий)</i>	<i>95%-ный довери- тельный интервал</i>
Рост	4,4	0,2	21,4	0,000	От 4,0 до 4,8
Возраст	0,08	0,03	2,2	0,026	От 0,01 до 0,2
Пол	-5,7	1,7	-3,4	0,001	От -9,0 до -2,4
Годы от получения образования	-0,7	0,2	-3,5	0,000	От -1,1 до -0,3
Нижний квинтиль занятий спортом	3,7	1,4	2,6	0,009	От 0,9 до 6,5
Фиктивная переменная «получение продоволь- ственных талонов»	5,6	2,1	2,7	0,007	От 1,5 до 9,7
Неиспано- язычный афро- американец	9,7	1,3	7,2	0,000	От 7,0 до 12,3
Отрезок, отсекаемый на оси Y	-117				

При проведении исследований, предполагающих выполнение регрессионного анализа, вы должны помнить одну очень важную вещь: постарайтесь никого не убить. Можете даже приклеить скотчем к монитору своего компьютера листочек с надписью: «Твои исследования не должны убивать людей». Дело в том, что подчас даже самые умные люди непреднамеренно нарушают это важное правило.

Начиная с 1990-х годов в системе здравоохранения возобладала концепция, согласно которой пожилые женщины должны принимать эстрогенные добавки, чтобы защититься от сердечно-сосудистых заболеваний, остеопороза и прочих недугов, связанных с менопаузой¹. К 2001 году эстрогенные добавки были предписаны примерно 15 миллионам женщин в надежде, что это снизит риск развития перечисленных заболеваний. На чем основывалась эта надежда? На проводившихся в то время исследованиях — с применением базовой методологии, описанной в предыдущей главе, — согласно которым прием эстрогенных добавок считался разумной медицинской стратегией. В частности, повторное исследование 122 000 женщин (так называемое Nurses' Health Study) продемонстрировало наличие отрицательной зависимости между приемом эстрогенных добавок и сердечными приступами. Риск возникновения последних у женщин, принимающих эстроген, составлял примерно одну треть от соответствующего риска у женщин, которые его не принимали. Исследование проводилось, конечно, не парой подростков, использующих отцовский компьютер для просмотра порнофильмов и попутного решения уравнений регрессии, а Гарвардской медицинской школой и Гарвардской школой общественного здравоохранения.

Между тем, ученые и практикующие врачи выдвинули теорию, объясняющую, почему гормональные добавки могут быть полезны для здоровья женщин. В пожилом возрасте женские яичники вырабатывают меньше эстрогена, а поскольку он необходим женскому организму, то восполнение

его дефицита в пожилом возрасте укрепляет здоровье женщины в долгосрочной перспективе. Отсюда и название метода: терапия путем замещения гормона. Некоторые исследователи рекомендовали эстрогенное стимулирование даже пожилым мужчинам².

А затем, после того как миллионам женщин была предписана заместительная гормонотерапия, эстроген подвергли более строгой форме научного исследования — клиническим испытаниям. Вместо того чтобы искать статистические взаимосвязи (которые могут выражать (или не выражать) реальную связь причины и следствия) в большой совокупности данных наподобие той, которая использовалась в ходе исследования Nurses' Health Study, клинические испытания предусматривают проведение управляемого эксперимента. Одна выборка получает лечение (например, в виде терапии путем замещения гормона), а другая принимает плацебо. Клинические испытания показали, что у женщин, принимающих эстроген, более высокий уровень сердечно-сосудистых заболеваний, инсультов, образования тромбов, чаще диагностируется рак груди и наблюдаются прочие неблагоприятные для здоровья исходы. Эстрогенные добавки приносят определенную пользу, однако она полностью нивелируется дополнительными рисками. Начиная с 2002 года врачам было рекомендовано не назначать эстроген пожилым пациенткам. Остается только гадать, скольких женщин постигла преждевременная смерть, у скольких случился инсульт или развился рак груди из-за приема таблеток, которые якобы должны были укрепить их здоровье.

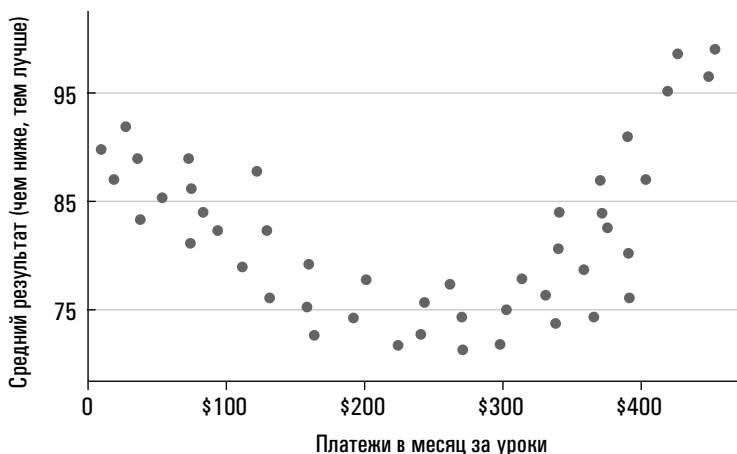
Вполне возможно, что их количество исчисляется десятками тысяч³.

Регрессионный анализ — это своего рода водородная бомба в арсенале статистики. Каждый владелец персонального компьютера и большой совокупности данных может стать исследователем, не выходя из дома или не покидая стен офиса. В чем же причина проблем с регрессионным анализом? Таких причин очень много. Регрессионный анализ позволяет получить *точные* ответы на сложные вопросы, но они могут быть правильными или неправильными. В неумелых руках регрессионный анализ даст результаты, которые способны ввести в заблуждение или попросту оказаться неверными. И, как показывает пример с эстрогеном, *даже в умелых руках* этот мощный статистический инструмент может направить по ложному — и опасному! — пути. Задача настоящей главы — объяснить самые типичные «ошибки» регрессии. Слово «ошибки» я заключил в кавычки по той причине, что, как и в случае с другими видами статистического анализа, ловкие люди могут совершенно осознанно использовать их в неблагоприятных целях.

Ниже перечислены семь самых типичных злоупотреблений этим замечательным инструментом.

Использование регрессии для анализа нелинейной связи*. Приходилось ли вам читать предостережение, которое обычно наносится на корпус фена для волос: «Не пользоваться во время мытья в ванне»? Читая эти слова, вы, наверное, думали: «Какой болван может до такого додуматься?» *Ведь это электроприбор, им нельзя пользоваться в воде.* Электроприборы для этого не предназначены. Если бы регрессионный анализ снабжался подобным предостережением, то оно должно было бы гласить: «Не пользоваться, когда между анализируемыми переменными существуют нелинейные зависимости». Запомните: коэффициент регрессии описывает степень наклона «линии наилучшего приближения» для рассматриваемых вами данных; не прямая линия будет характеризоваться разными степенями наклона в разных точках. Рассмотрим, например, следующую гипотетическую связь между числом уроков игры в гольф, которые я беру в течение месяца (объясняющая переменная), и моим средним результатом для восемнадцатилункового раунда за тот же месяц (зависимая переменная). Как нетрудно заметить из приведенной ниже диаграммы разброса данных, в этом случае отсутствует устойчивая *линейная* зависимость.

Влияние уроков игры в гольф на результаты



* Чтобы приспособить регрессионный анализ для использования данных с нелинейными связями, существуют более сложные методы. Однако прежде чем их применять, вам нужно уяснить, почему использование обычного метода наименьших квадратов с нелинейными связями лишено смысла.

Итак, мы видим некую картину, которую невозможно описать с помощью одной прямой линии. Первые несколько уроков игры в гольф, похоже, привели к быстрому улучшению моих показателей (количество очков уменьшилось — в гольфе это считается положительным результатом). На этом отрезке времени наблюдается отрицательная зависимость между уроками и набранным мною количеством очков; наклон линии отрицательный. Чем больше уроков, тем меньше очков.

Но когда я начинаю тратить на уроки игры в гольф от 200 до 300 долларов в месяц, это, по-видимому, не оказывает на мои результаты вообще никакого влияния. На данном отрезке времени не наблюдается какой-либо четкой взаимосвязи между дополнительными уроками и моими результатами; наклон линии — нулевой.

Наконец наступает момент, когда уроки становятся контрпродуктивными. Если сумма, потраченная на уроки игры в гольф, достигает 300 долларов в месяц, дополнительные уроки ассоциируются с бóльшим количеством набранных мною очков; на этом отрезке времени наблюдается положительный наклон линии. (Ниже в этой главе мы обсудим вероятность того, что плохие результаты игры в гольф могут стимулировать брать дополнительные уроки, а не наоборот.)

Самое важное здесь то, что с помощью единственного коэффициента регрессии мы не можем точно выразить зависимость между уроками и результатами. Наилучшей интерпретацией описанной выше картины будет то, что уроки игры в гольф характеризуются несколькими линейными связями с моими результатами. Вы можете видеть это, а пакет статистического программного обеспечения — нет. Если вы введете эти данные в уравнение регрессии, то компьютер выдаст вам единственный коэффициент. И он не будет точно отражать истинную взаимосвязь между интересующими нас переменными. Полученные результаты будут представлять собой статистический эквивалент использования фена для волос во время принятия ванны.

Регрессионный анализ предназначен для использования в случае линейной зависимости между переменными*. В солидных учебниках по статисти-

* Необходимо уточнить, что метод наименьших квадратов (МНК), который автор объявил основой регрессионного анализа, действительно можно использовать только для линейных уравнений регрессии. Но линейных относительно коэффициентов регрессии, а не переменных. Поэтому МНК вполне можно применять и для нелинейных (по переменным) уравнений регрессии, которые, однако, являются линейными относительно коэффициентов регрессии либо становятся таковыми после

ке указаны также другие базовые условия его применения. Как и при использовании любого другого инструмента, чем больше вы отклоняетесь от заранее оговоренных условий его применения, тем менее эффективным — и даже потенциально опасным — он становится.

Корреляция и причинно-следственные зависимости — не одно и то же. Регрессионный анализ может лишь продемонстрировать взаимосвязь между двумя переменными. Как я уже упоминал, с помощью только статистики невозможно доказать, что изменение одной переменной обусловило изменение другой переменной. Вообще говоря, неправильное уравнение регрессии может указать на существование внушительной и статистически значимой зависимости между двумя переменными, которые в действительности между собой никак не связаны. Допустим, мы планируем выявить потенциальные причины роста числа случаев аутизма в Соединенных Штатах за последние два десятилетия. Наша зависимая переменная — исход, который мы хотели бы объяснить, — могла бы служить показателем заболеваемости аутизмом, таким как количество диагностированных случаев на каждых 1000 детей определенного возраста. Если бы мы включили в качестве объясняющей переменной годовой доход на душу населения в Китае, то почти наверняка выявили бы положительную и статистически значимую зависимость между повышением доходов в Китае и ростом заболеваемости аутизмом в США за последние два десятилетия.

Чем это объясняется? Всего лишь тем, что оба показателя резко увеличились за указанный период. Между тем, я очень сомневаюсь, что наступление экономической рецессии в Китае приведет к снижению заболеваемости аутизмом в США. (Справедливости ради должен заметить, что если бы я наблюдал четкую связь между быстрым экономическим ростом в Китае и заболеваемостью аутизмом *только в Китае*, то я, возможно, приступил бы к поиску какого-либо фактора окружающей среды, связанного с экономическим ростом (например, загрязнение окружающей среды отходами промышленного производства), который мог бы объяснить подобную зависимость.)

Только что продемонстрированный мной род ложной зависимости между двумя переменными — лишь один пример более универсального явления, известного как фиктивные причинно-следственные связи. Существует

преобразований. Также отметим, что в арсенале регрессионного анализа есть методы, отличные от МНК, которые предназначены для нахождения коэффициентов регрессии в существенно нелинейных уравнениях. *Прим. ред.*

несколько других вариантов, когда связь между А и В может быть неправильно интерпретирована.

Обратная причинно-следственная зависимость. Статистическая зависимость между А и В не доказывает, что А является причиной В. Вообще говоря, не исключено, что В — это причина А. Я указывал на такую вероятность ранее в примере с уроками игры в гольф. Допустим, что когда я построил сложную модель, чтобы объяснить свои результаты в гольфе, оказалось, что переменная, обозначающая количество уроков игры в гольф, демонстрирует четкую взаимосвязь с ухудшением моих показателей. Чем больше уроков я беру, тем хуже результаты! Одним из объяснений может быть то, что мне попался очень плохой тренер. Более правдоподобное объяснение: я обычно беру дополнительные уроки, когда начинаю плохо играть, то есть плохие результаты являются причиной увеличения количества уроков, а не наоборот. (Существует ряд простых методологических исправлений проблем такого рода. Например, я мог бы включить количество уроков игры в гольф в одном месяце в качестве объясняющей переменной для моих показателей в следующем месяце.)

Как указывалось выше в этой главе, причинно-следственные связи могут действовать в обоих направлениях. Допустим, согласно проводимому вами исследованию, штаты, которые тратят больше денег на школьное образование, демонстрируют более высокие темпы экономического роста, чем штаты, вкладывающие в школьное образование меньше денег. Наличие положительной и значимой зависимости между этими двумя переменными ничего нам не говорит о *направлении* этой зависимости. Инвестиции в программу школьного образования могут вызывать экономический рост. С другой стороны, штаты, демонстрирующие более высокие темпы экономического роста, могут себе позволить больше инвестировать в школьное образование; стало быть, сильная экономика может быть причиной увеличения расходов на образование. Другой вариант: дополнительные траты на школьное образование могут стимулировать экономический рост, что позволяет вкладывать больше средств в образование, то есть причинно-следственные связи могут носить двусторонний характер.

Следовательно, мы не должны использовать объясняющие переменные, зависящие от исхода, который мы пытаемся объяснить, — в противном случае результаты могут оказаться безнадежно запутанными. Например, было бы неуместно использовать коэффициент безработицы в уравнении

регрессии, объясняющем рост ВВП, поскольку совершенно очевидно, что уровень безработицы зависит от темпов роста ВВП. Или, иначе говоря, результат регрессионного анализа, заключающийся в том, что снижение безработицы обусловит рост ВВП, представляется совершенно бессмысленным, потому что именно рост ВВП обычно приводит к снижению безработицы.

У нас должны быть все основания полагать, что наши объясняющие переменные влияют на зависимую переменную, а не наоборот.

Систематическая ошибка, вызванная пропущенной переменной. Увидев в газете броский заголовок: «Игроки в гольф чаще болеют сердечно-сосудистыми заболеваниями, раком и артритом!», не относитесь к нему серьезно. Я не был бы удивлен, если бы это было так. Я также подозреваю, что гольф полезен для здоровья, поскольку обеспечивает не только возможность социализации, но и умеренную физическую нагрузку. Как совместить оба утверждения? Очень просто! Любое исследование, измеряющее влияние игры в гольф на состояние здоровья человека, должно надлежащим образом контролировать возраст. Вообще говоря, гольфом в большей степени увлекаются люди старших возрастов — особенно пенсионеры. Любой анализ, не принимающий во внимание возраст как объясняющую переменную, упускает из виду тот факт, что гольфисты в среднем — более пожилые люди, чем те, кто в него не играет. Не гольф убивает людей, а старость. Так уж случается, что гольф предпочитают именно пожилые люди. Я полагаю, что при использовании возраста в регрессионном анализе в качестве управляющей переменной мы получим другой результат: *для людей одного и того же возраста* игра в гольф может стать профилактикой серьезных заболеваний. Это весьма существенная разница.

В данном примере возраст — важная «пропущенная переменная». Когда мы не учитываем его в уравнении регрессии, объясняющем развитие сердечно-сосудистых заболеваний или какие-то другие исходы, неблагоприятные для здоровья человека, переменная «увлечение игрой в гольф» исполняет *две объясняющие роли*, а не одну. Она говорит о влиянии игры в гольф на состояние сердечно-сосудистой системы и о влиянии старости на состояние сердечно-сосудистой системы. На языке статистики это будет звучать примерно так: переменная «увлечение игрой в гольф» *подхватывает* (учитывает) влияние возраста. Проблема заключается в том, что эти два влияния объединяются. В лучшем случае наши результаты оказываются весьма запутанными. В худшем мы приходим к ошибочному выводу, что

гольф плохо сказывается на здоровье человека, хотя на самом деле вероятнее обратное утверждение.

Результаты регрессии будут вводить нас в заблуждение и страдать неточностью в случае отсутствия в уравнении регрессии какой-либо важной объясняющей переменной, особенно если другие переменные в этом уравнении «подхватывают» данный эффект. Допустим, мы пытаемся объяснить качество школ. Нам очень важно понять, что именно делает школы хорошими. Нашей зависимой переменной — численным показателем качества — будут, вероятнее всего, результаты экзаменов. Мы почти наверняка станем рассматривать расходы школы как одну объясняющую переменную в надежде найти численное выражение связи между расходами и результатами экзаменов. Можно ли утверждать, что школы, у которых больше расходы, добиваются лучших результатов? Если бы расходы школы были единственной объясняющей переменной, я не сомневаюсь, что нам удалось бы выявить четкую и статистически значимую зависимость между ними и итогами экзаменов. Однако такой вывод, а также вытекающее из него следствие, будто улучшить качество школ можно путем повышения расходов, глубоко ошибочны.

Здесь есть немало потенциально значимых пропущенных переменных, однако важнейшей из них будет уровень образования родителей. Высокообразованные семьи, как правило, проживают в престижных районах. А расположенные в этих районах школы обычно расходуют немалые средства. К тому же дети в таких семьях демонстрируют хорошие результаты на экзаменах (тогда как баллы детей из малоимущих семей гораздо хуже). Если у нас нет какого-либо показателя социально-экономического статуса учащихся, который можно было бы использовать в качестве управляющей переменной, то результаты нашей регрессии наверняка укажут на четкую положительную зависимость между расходами школы и итогами экзаменов, тогда как в действительности эти результаты могут быть функцией социально-экономического положения учащихся школы, а не суммы денег, израсходованных ею.

Я помню, как один из преподавателей нашего колледжа утверждал, что результаты школьных экзаменов высоко коррелированы с количеством автомобилей, которыми владеет семья. Этим он как бы намекал на несправедливость школьных тестов и невозможность использовать их итоги в качестве основного критерия при поступлении в колледж. Разумеется, система школьных экзаменов не лишена недостатков, но корреляция между их результатами и количеством автомобилей в семье вовсе не то, что тревожит

меня больше всего. Меня мало волнует, что богачи могут устроить своих детей в колледж, купив еще три автомобиля. Количество автомобилей в семейном гараже является показателем дохода соответствующей семьи, уровня образования ее членов и прочих признаков их социально-экономического статуса. То обстоятельство, что дети из состоятельных семей сдают экзамены успешнее их менее зажиточных сверстников, не новость. (Как отмечалось ранее, средний балл сдачи стандартизированного теста по чтению у учащихся из семей, совокупный доход которых превышает 200 000 долларов, на 134 балла выше, чем средний результат сдачи такого же теста детьми из семей, совокупный доход которых не превышает 20 000 долларов.)⁴ Гораздо больше меня интересует вероятность улучшить результаты сдачи стандартизированного теста путем «натаскивания» ученика. Насколько ученик может их улучшить, воспользовавшись услугами частных репетиторов? Очевидно, у состоятельных семей гораздо больше возможностей нанять для своих детей хороших репетиторов. Любое улучшение результатов сдачи экзаменов учащимися, занимающимися с репетиторами (если, конечно, это не чистая случайность), говорит в пользу детей из состоятельных семей по сравнению с их менее зажиточными сверстниками, *даже если способности тех и других совершенно одинаковы*, — ведь ученики из малообеспеченных семей тоже могли бы улучшить свои результаты, если бы воспользовались услугами частных репетиторов (однако им это не по карману).

Сильно коррелированные объясняющие переменные (мультиколлинеарность). Если уравнение регрессии включает две объясняющие переменные (или даже больше), сильно коррелированные между собой, то анализ вполне может не выявить истинной зависимости между каждой из этих переменных и исходом, который мы пытаемся объяснить. Приведу соответствующий пример. Допустим, мы хотим измерить влияние противозаконного использования наркотиков на результаты сдачи экзаменов. В частности, мы располагаем данными о том, употребляли ли когда-либо участники нашего исследования кокаин и «баловались» ли когда-либо героином. (Будем исходить из того, что в нашем распоряжении есть и много других управляющих переменных.) Каково влияние употребления кокаина на результаты сдачи экзаменов (при условии неизменности всех остальных факторов, включая употребление героина)? А каково влияние употребления героина на итоги экзаменов (при условии неизменности всех остальных факторов, включая употребление кокаина)?

Вполне возможно, что коэффициенты по употреблению героина и кокаина не смогут ответить на интересующие нас вопросы. Методологическая проблема в данном случае заключается в том, что те, кто «баловался» героином, наверняка употребляли и кокаин. Если поместить в уравнение обе переменные, то число тех, кто употреблял один из этих наркотиков, но не употреблял другой, окажется очень незначительным. Это оставит нам довольно мизерное расхождение в данных, на основании которого мы могли бы вычислить их независимые влияния. Вспомните мысленный эксперимент, который мы провели в предыдущей главе, чтобы объяснить регрессионный анализ. Мы распределили выборку данных по разным комнатам, в которых каждое наблюдение идентично за исключением одной переменной, что позволяло затем вычленить влияние этой переменной, параллельно контролируя другие факторы, потенциально способные сказываться на интересующем нас исходе. В нашей выборке может быть 692 человека, которые употребляли и кокаин, и героин. Но у нас может быть и всего три человека, которые употребляли только кокаин, и два человека, употреблявших только героин. Любой вывод относительно независимого влияния лишь одного или другого наркотика будет основываться на этих крошечных выборках.

Вряд ли нам удастся получить достоверные коэффициенты регрессии по какой-либо из этих двух переменных (кокаин или героин); мы можем также проигнорировать более сильную и важную зависимость между результатами экзаменов и употреблением какого-то одного из этих наркотиков. Когда две объясняющие переменные сильно коррелированы между собой, исследователи обычно используют в уравнении регрессии какую-то одну из них; как вариант, они могут создать некую составную переменную, например «употреблял кокаин или героин». Если же исследователи хотят контролировать в целом социально-экономическое положение учащегося, они могут включить переменные «образование матери» и «образование отца», поскольку это обеспечивает важное указание на уровень образования соответствующей семьи в целом. Однако если цель регрессионного анализа — вычленить влияние либо образования отца, либо образования матери, то включение в уравнение обеих переменных скорее запутает вопрос, чем внесет в него ясность. Корреляция между уровнями образования мужа и жены столь высока, что мы не можем полагаться на то, что регрессионный анализ даст нам коэффициенты, которые позволят надлежащим образом вычленить

влияние образования кого-либо из родителей (это так же трудно, как обособить влияние употребления кокаина от влияния употребления героина)*.

Экстраполяция за границы имеющихся данных. Регрессионный анализ, как и все формы статистического вывода, помогает нам лучше понять окружающий мир. Мы пытаемся выявить закономерности, которые будут общими и для более крупной совокупности. *Однако наши результаты будут справедливы лишь для совокупности, подобной выборке, в отношении которой выполнялся анализ.* В предыдущей главе я создал уравнение регрессии, позволяющее предсказывать вес, основываясь на ряде независимых переменных. Значение R^2 в моей окончательной модели равнялось 0,29; это означает, что оно дает возможность объяснить разброс веса для крупной выборки людей, если все они оказались взрослыми.

Итак, что же произойдет, если мы воспользуемся нашим уравнением регрессии для предсказания вероятного веса новорожденного младенца? Давайте проверим. При рождении рост моей дочери составлял 21 дюйм. Допустим, ее возраст в момент рождения равнялся нулю; у нее, конечно же, не было образования и она не занималась спортом. Она относилась к белой расе и была женского пола. Уравнение регрессии, основанное на данных America's Changing Lives, предсказывает, что ее вес при рождении должен иметь отрицательную величину: -19,6 фунта. (В действительности она весила 8,5 фунта.)

Авторы одного из исследований, выполнявшихся по заказу британского правительства (мы упоминали о них в предыдущей главе), сделали совершенно четкий вывод: «Неспособность работника влиять на свою рабочую среду ассоциируется с повышенным риском развития заболеваний сердечно-сосудистой системы *среди государственных служащих*»⁵ (курсив мой).

Интеллектуальный анализ (слишком много переменных). Если игнорирование важных переменных представляет собой потенциальную проблему, то, может быть, ее возможным решением будет максимальное наращивание количества объясняющих переменных в уравнении регрессии? Отнюдь! Ваши результаты могут быть поставлены под угрозу, если вы включите в уравнение регрессии чересчур большое число переменных, особенно если

* Проще говоря (так, как принято в этой книге), мультиколлинеарность заключается в наличии сильной линейной (статистической) зависимости внутри некоторой группы объясняющих переменных. Это порождает вычислительные сложности или вообще невозможность рассчитать коэффициенты функции регрессии. *Прим. ред.*

речь идет о дополнительных объясняющих переменных без какого-либо теоретического обоснования такого решения. Например, *не следует* разрабатывать стратегию исследования, построенную на следующей предпосылке: поскольку нам неизвестно, что вызывает аутизм, нужно включить в уравнение регрессии как можно больше потенциальных объясняющих переменных, чтобы увидеть, что именно может оказаться статистически значимым; затем, возможно, мы сумеем получить кое-какие ответы. Если вы включите в уравнение регрессии достаточно большое число лишних переменных, то одна из них, по чистой случайности, обязательно достигнет порога статистической значимости. Еще одна опасность заключается в том, что лишние переменные порой не так-то легко распознать именно как *лишние*. Опытные исследователи могут всегда обосновать теоретически, *постфактум*, почему та или иная необычная переменная, которая в действительности совершенно бессмысленна, оказывается статистически значимой*.

Чтобы доказать это, я нередко проделываю то же упражнение с подбрасыванием монетки, которое приводил при обсуждении вероятностей. В аудитории примерно из сорока студентов я предлагаю каждому подбросить монетку. Все, у кого выпадает решка, выводятся из игры; остальные продолжают подбрасывание. Во втором раунде те, у кого выпадает решка, снова выводятся из игры. Я продолжаю раунды до тех пор, пока у кого-то из студентов пять или шесть раз подряд не выпадет орел. Наверняка вам придется на память глупые вопросы, которые обычно задают в таких случаях: «В чем ваш секрет? Вы достаете этих орлов из рукава? Можете ли вы научить нас подбрасывать монетку так, чтобы каждый раз выпадал орел? Может быть, все дело в фирменной футболке Гарвардского университета, в которой вы пришли сегодня на лекцию?»

Разумеется, череда следующих друг за другом выпадений орлов — чистая случайность: студенты, присутствовавшие в аудитории, были свидетелями происходящего. Однако полученный результат мог по-разному интерпретироваться в научном контексте. Вероятность пятикратного (подряд) выпада-

* Еще одной проблемой «лишних» переменных является мультиколлинеарность (описанная выше), вероятность которой резко возрастает при внесении в уравнение регрессии дополнительных переменных, не прошедших специальной проверки. С другой стороны отметим, что в регрессионном анализе развиты средства отбраковки лишних незначимых объясняющих переменных. Простейшим из которых является так называемый скорректированный коэффициент детерминации, рассчитываемый на основе параметра R^2 . *Прим. ред.*

ния орлов равняется $\frac{1}{32}$, или 0,03. Это существенно ниже порога 0,05, который мы обычно используем, чтобы отвергнуть основную гипотезу. Наша нулевая гипотеза в данном случае заключается в том, что этот студент не обладает особым талантом подбрасывать монетку. Тем не менее удачная череда выпаданий орлов (которая обязательно произойдет по крайней мере у одного студента, если этот эксперимент будет проводиться с достаточно большим количеством участников) позволяет нам отклонить нулевую гипотезу и принять альтернативную гипотезу, утверждающую, что данный студент обладает особым талантом подбрасывать монетку так, чтобы каждый раз выпадал орел. После того как он достиг этого впечатляющего результата, мы можем подвергнуть его более детальному изучению в надежде выявить причины столь блестящих достижений: методика подбрасывания монетки, особая физическая подготовка, умение полностью сконцентрироваться на монетке, пока она вращается в воздухе, и т. п. Все это совершеннейшая чепуха!

Подобное явление способно расстроить даже безупречно организованное исследование. Считается, что нулевую гипотезу следует отвергнуть, когда мы наблюдаем нечто, что должно было бы произойти по чистой случайности не чаще, чем в 1 случае из 20, если бы наша основная гипотеза была верна. Разумеется, если мы проведем 20 исследований или включим в одно уравнение регрессии 20 лишних переменных, то в среднем получим один ложный статистически значимый результат. Журнал *The New York Times* блестяще выразил это противоречие, процитировав Ричарда Пето, медицинского статистика и эпидемиолога: «Эпидемиология так восхитительна и позволяет получить столь важные представления о жизни и смерти человека! Удручает лишь невероятное количество никому не нужных, бестолковых публикаций»⁶.

Даже к результатам клинических испытаний, которые обычно представляют собой статистические эксперименты и, следовательно, являются «золотым стандартом» медицинских исследований, следует относиться с изрядной долей скептицизма. В 2011 году газета *The Wall Street Journal* разместила на первой странице материал, который охарактеризовала как один из «грязных маленьких секретов» медицинских исследований: «Большинство результатов, в том числе и публикуемых в солидных научных периодических изданиях, рецензируемых коллегами авторов статей, невозможно воспроизвести повторно»⁷. (Речь идет о публикациях, предварительно проверяемых с точки зрения их методологической надежности другими экспертами в той же области; лишь после такой проверки материал отправляется в печать. Такие

публикации принято считать заслуживающими особого доверия с научной точки зрения.) Одна из причин этого «грязного маленького секрета» — систематическая ошибка позитивной публикации, описанная в главе 7. Если исследователи и медицинские журналы склонны обращать внимание на позитивные результаты и игнорировать негативные, то они вполне могут опубликовать итоги исследования, свидетельствующие об эффективности некоего лекарства, и проигнорировать девятнадцать других исследований, доказывающих его бесполезность. Некоторые клинические испытания могут также основываться на небольших выборках (что бывает обусловлено объективными факторами, например редко встречающейся болезнью), что повышает вероятность того, что случайное отклонение в данных привлечет к себе больше внимания, чем оно того заслуживает. Самое главное — у исследователей может быть предубеждение (осознаваемое или нет), вызванное или непоколебимой уверенностью в чем-либо, или пониманием того, что позитивный результат будет способствовать их научной карьере. (Никто еще не разбогател и не стал знаменитым, доказав, что то или иное лекарство не излечивает от рака.)

В силу всех перечисленных причин количество экспертных исследований, результаты которых оказались ошибочными, очень велико. Джон Иоаннидис, греческий врач-эпидемиолог, проанализировал итоги сорока девяти исследований, опубликованных в трех солидных медицинских журналах⁸. Каждое из них цитировалось в медицинской литературе не менее тысячи раз. Тем не менее примерно треть результатов впоследствии была опровергнута дальнейшими экспериментами. (Например, некоторые из исследований, проанализированных Иоаннидисом, доказывали эффективность упоминавшейся выше терапии путем замещения эстрогена.) По оценкам д-ра Иоаннидиса, выводы примерно половины опубликованных научных статей в конце концов оказываются ошибочными⁹. Его исследование было опубликовано в *Journal of the American Medical Association*, одном из журналов, в которых печатались проанализированные им статьи. Из этого следует забавный парадокс: если исследование д-ра Иоаннидиса верно, то вполне вероятно, что его исследование ошибочно.

Регрессионный анализ по-прежнему остается потрясающим статистическим инструментом. (Похоже, мои эпитеты в его адрес заставляют относиться к нему как к «волшебному эликсиру», о котором я упоминал в предыдущей главе. Разумеется, мои слова не лишены некоторого преувеличения.)

Он позволяет выявлять важные закономерности в крупных совокупностях данных, которые зачастую оказываются ключом к серьезным исследованиям в медицине и социальных науках. Статистика предоставляет нам объективные стандарты для оценивания этих закономерностей. Регрессионный анализ, при надлежащем использовании, — значимая составляющая научного метода. Считайте эту главу предупреждением, к которому обязательно нужно прислушаться.

Все конкретные предостережения, о которых шла речь в этой главе, можно свести к двум ключевым положениям. Во-первых, создание эффективного уравнения регрессии — то есть определение, какие переменные нужно проанализировать и что должно быть источником соответствующих данных, — важнее самих статистических вычислений. Этот процесс называется оцениванием адекватности уравнения или выбором правильного уравнения регрессии. Лучшие исследователи — те, кто может путем логических умозаключений решить, какие переменные включить в уравнение регрессии, какие проигнорировать и как следует интерпретировать конечные результаты.

Во-вторых, как и большинство других статистических выводов, регрессионный анализ выстраивает лишь некую версию, основанную на косвенных доказательствах. Зависимость между двумя переменными подобна отпечаткам пальцев, оставленным на месте преступления. Она указывает на преступника, но одних лишь отпечатков недостаточно, чтобы осудить человека. (К тому же они могут ему не принадлежать.) Любой регрессионный анализ нуждается в теоретическом обосновании. Почему в уравнение регрессии включены именно эти объясняющие переменные? Какие явления из других областей могут объяснить наблюдаемые результаты? Например, почему мы считаем, что красные туфли у экзаменуемых способствуют значительному улучшению результатов сдачи школьных экзаменов или что употребление попкорна помогает предотвратить рак простаты? Соответствующие результаты должны быть повторно воспроизводимыми или по крайней мере не должны противоречить итогам других исследований.

Даже волшебный эликсир может не оказать должного эффекта, если не пользоваться им так, как предписано.

Изменит ли вашу жизнь поступление в Гарвардский университет

Блестящие исследователи в области социальных наук блестящие вовсе не потому, что умеют выполнять в уме сложные вычисления и выигрывают в телевикторине Jeopardy* больше денег, чем обычные исследователи (хотя, возможно, они преуспели и в том и в другом). Блестящие исследователи — это те, кто существенно меняет наши знания и представления о мире и находит творческие способы проведения управляемых экспериментов. Чтобы измерить чье-либо влияние, нам требуется нечто такое, относительно чего мы будем выполнять измерение. Как сказалось бы на вашей жизни поступление в Гарвардский университет? Чтобы ответить на этот вопрос, вам нужно знать, что произойдет после того, как вы поступите в Гарвардский университет, и что произойдет *после того, как вы в него не поступите*. Очевидно, вы не можете располагать данными для *обоих* случаев. Тем не менее умные исследователи находят возможность сравнить то или иное «воздействие» (например поступление в Гарвардский университет) с его противоположным сценарием.

Чтобы проиллюстрировать это положение, давайте поразмыслим над казалось бы простым вопросом: приведет ли к снижению преступности увеличение количества полицейских на улицах? Это социально значимый вопрос, поскольку преступность обходится обществу слишком дорого. Если рост числа полицейских на улицах позволит ее снизить (либо потому, что окажется сдерживающим фактором для преступников, либо за счет поимки большего количества «плохих парней»), то инвестиции в наращивание численности полицейских могут обернуться большой выгодой для общества. С другой стороны, рост числа полицейских — весьма дорогостоящее удовольствие; и если эта мера не даст нужного результата или он будет совсем незначительным, то общество может пожалеть о том, что не нашло

* Русский аналог этой телевикторины называется «Своя игра». *Прим. перев.*

более эффективного применения своим ресурсам (например, внедряя современные технологии борьбы с преступностью, такие как видеокамеры наблюдения).

Проблема в том, что ответить на этот вроде бы простой вопрос о последствиях наращивания численности полицейских на улицах не так уж просто. На основании прочитанного к данному моменту материала вы должны это признать, по крайней мере мы не можем ответить на него исходя из информации о тех населенных пунктах, где число полицейских на душу населения существенно превышает средний показатель. Цюрих — не Лос-Анджелес. Даже сравнение крупных американских городов окажется совершенно некорректным: Лос-Анджелес, Нью-Йорк, Хьюстон, Майами, Детройт и Чикаго — слишком разные города с разным демографическим составом населения и разными проблемами, порождающими преступность.

Нашим обычным подходом было бы попытаться определить уравнение регрессии, в котором учитывались бы все эти различия. Но увы, даже множественный регрессионный анализ здесь не поможет. Если мы попробуем объяснить уровень преступности (нашу зависимую переменную) путем использования числа полицейских на душу населения в качестве объясняющей переменной (наряду с другими объясняющими переменными), то возникнет серьезная проблема с обратной причинно-следственной зависимостью. У нас есть надежное теоретическое основание полагать, что увеличение количества полицейских на улицах приведет к снижению преступности, но возможно и обратное: уровень преступности «обуславливает» рост численности полицейских, то есть в городах с повышенной криминогенной обстановкой будет больше служителей порядка. Мы легко можем обнаружить положительную — но вводящую в заблуждение — взаимосвязь между уровнем преступности и количеством полицейских: в районах, где больше всего полицейских, как правило, самый высокий уровень преступности. Аналогично, там, где больше всего медицинских работников, обычно высокий уровень заболеваемости. Разумеется, люди там болеют вовсе не потому, что там слишком много врачей; просто врачи сконцентрированы в местах, где в них существует особая потребность (с другой стороны, больные люди направляются в места, где они могут получить надлежащую медицинскую помощь). Подозреваю, что во Флориде непропорционально большое число онкологов и кардиологов; но даже если выслать половину из них за пределы Флориды, проживающие в этом штате пенсионеры здоровее не станут.

Итак, добро пожаловать в программы статистического оценивания, представляющие собой процесс, посредством которого мы пытаемся измерить результат того или иного воздействия, коим может быть все что угодно, от нового лекарства от рака до программы обеспечения занятости тех, кто бросил школу. Воздействие, о котором я веду речь, обычно называется «активированием»*, хотя в статистическом контексте это слово используется в более широком значении, чем в повседневной жизни. Активирование может быть воздействием в буквальном смысле (то есть медицинским вмешательством того или иного рода), или чем-то наподобие поступления в колледж, или обучением какой-либо профессии после выхода из тюрьмы. Дело в том, что мы пытаемся изолировать влияние этого единственного фактора; в идеале нам хотелось бы знать, как чувствует себя группа, получающая такое «активирование», по сравнению с абсолютно идентичной группой, которая его не получает.

Программы статистического оценивания предлагают совокупность инструментов, позволяющих обособить влияние активирования, когда невозможно установить причину и следствие. Ниже описано, как Джонатан Клик и Александер Табаррок, исследователи из Пенсильванского университета и Университета Джорджа Мейсона, изучали влияние наращивания численности полицейских на улицах на уровень преступности. Стратегия их исследования предусматривала использование системы оповещения о терроризме (terrorism alert system). Если конкретнее, то полицейское управление Вашингтона в дни «повышенной опасности терроризма» выводит на улицы определенных районов города дополнительные наряды полиции, поскольку столица США является естественной мишенью терроризма. Мы можем предположить, что между уличной преступностью и угрозой терроризма нет никакой зависимости, поэтому такое увеличение количества полицейских на улицах Вашингтона не связано с уровнем обычной преступности, то есть обусловлено «внешними» причинами. Самым ценным стало то, что исследователи на основе естественного эксперимента смогли ответить на вопрос: что происходит с обычной преступностью в дни «повышенной опасности терроризма»?

* В оригинале приведено слово treatment, которое имеет множество значений. Эти значения: обработка, решение, лечение, трактовка, активизация и др. Мы выбрали слово «активирование» как наиболее подходящее по смыслу для использования в данном тексте. *Прим. ред.*

Ответ оказался таким. Количество преступлений, совершенных в дни «оранжевой» угрозы (высокая опасность и использование дополнительных нарядов полиции), было примерно на 7% ниже, чем в дни, когда уровень террористической угрозы был «желтым» (повышенная опасность, но никаких дополнительных мер по обеспечению правопорядка не принимается). Авторы также пришли к выводу, что самое резкое снижение уровня преступности наблюдалось в полицейском округе, который пользуется самым пристальным вниманием со стороны полиции в дни высокой опасности терроризма (этот округ включает Белый дом, Капитолий и Эспланаду*). Важный вывод заключается в том, что мы можем ответить на сложные социально значимые вопросы, если подходить к этому делу с умом. Ниже перечислены некоторые из наиболее типичных подходов к обособлению влияния активирования.

Статистический управляемый эксперимент. Самый простой способ создать подопытную (по какому-либо методу активирования) и контрольную группы — это... создать подопытную и контрольную группы. Но у этого подхода есть две крупные проблемы. Во-первых, существует много видов экспериментов, которые мы не можем проводить на людях. И это ограничение (я надеюсь) в обозримом будущем никуда не денется. Таким образом, мы можем проводить управляемые эксперименты на людях лишь тогда, когда у нас есть основания полагать, что соответствующее активирование принесет потенциально положительный результат. Зачастую у нас такой уверенности *нет*, именно поэтому нам необходимы стратегии, о которых будет рассказано в этой главе.

Во-вторых, люди отличаются между собой гораздо больше, чем лабораторные крысы. На эффект воздействия, который мы проверяем, вполне могут наложиться другие различия в испытуемой и контрольной группе: там обязательно окажутся высокие люди и «коротышки», больные и здоровые, мужчины и женщины, преступники и законопослушные граждане, алкоголики и трезвенники, банкиры и малообеспеченные люди и т. п. Как мы можем гарантировать, что различия по этим и другим характеристикам не скажутся на результатах тестирования? У меня есть для вас хорошая новость: это один из тех редких случаев жизни, когда наилучший подход предполагает минимум усилий! Оптимальный способ создания любой подопытной группы, которая подвергается активированию, и контрольной группы заключается

* Эспланада — отрезок музейно-парковой зоны в центре Вашингтона между Капитолием и памятником Джорджу Вашингтону. *Прим. перев.*

в случайном (рандомизированном) распределении по ним участников исследования. Прелесть рандомизации в том, что она приведет к более или менее равномерному распределению между этими двумя группами переменных, не связанных с активированием, — как очевидных характеристик, таких как пол, расовая принадлежность, возраст и образование, так и ненаблюдаемых характеристик, которые могли бы исказить интересующие нас результаты.

Это можно представлять себе так. Если в нашей большой выборке присутствуют 1000 женщин, то после того как мы произвольно разделим ее на две группы, в каждой из групп, скорее всего, окажется по 500 женщин. Разумеется, утверждать наверняка это нельзя, но и здесь теория вероятностей придет нам на помощь. Вероятность того, что в какой-то из двух групп будет непропорционально большое число женщин (или непропорционально большое число людей с какой-либо другой характеристикой), очень мала. Если, например, в выборке из 1000 человек половину составляют женщины, то вероятность того, что в какой-то из двух групп окажется менее 5 женщин, будет меньше 1%. Понятно, что чем больше выборка, тем эффективнее (с точки зрения создания похожих, в широком смысле, групп) рандомизация.

Медицинские испытания, как правило, стремятся проводить в духе рандомизированных управляемых экспериментов. В идеале они представляют собой двойное контрольное (слепое) испытание; это означает, что ни пациент, ни врач не знают, кто получает лечение, а кто — плацебо. Разумеется, двойное контрольное испытание невозможно, когда речь идет о хирургических процедурах (надеюсь, кардиохирургу будет заранее известно, к каким из пациентов применяется операция шунтирования). Хотя даже в этом случае иногда удастся держать пациентов в неведении относительно того, к какой именно группе (подопытной или контрольной) они относятся. Одно из моих любимых исследований предполагало оценивание определенного вида коленной хирургии, призванной облегчить боль. Участникам «лечебной» группы делали соответствующую хирургическую операцию на колене. А членам контрольной группы хирург, имитируя выполнение операции, делал три небольших надреза в области колена пациента*. Оказалось, что реальная хирургическая операция была ненамного эффективнее, чем мнимая¹.

Рандомизированные статистические исследования могут применяться для тестирования некоторых интересных явлений. Например, улучшают ли

* Участники этого эксперимента знали, что участвуют в клиническом испытании и что им могут сделать фиктивную хирургическую операцию.

постхирургические исходы молитвы людей, незнакомых с прооперированными? Разумные люди по-разному относятся к религии, но авторы исследования, результаты которого были опубликованы в *American Heart Journal*, провели управляемый эксперимент, который должен был ответить на вопрос, будет ли меньше послеоперационных осложнений у больных, перенесших операцию шунтирования на сердце, если большая группа совершенно незнакомых им людей будет молиться за их скорейшее выздоровление². В исследовании участвовали 1800 пациентов и членов трех религиозных общин со всей страны. Пациентов, перенесших операцию коронарного шунтирования, разделили на три группы: за одну группу никто не молился; за вторую молились, и им сообщили об этом; за третью молились, но ее участникам сказали, что за них могут молиться или не молиться (таким образом обеспечивалась управляемость плацебо-эффекта молитвы). Тем временем членов религиозных конгрегаций попросили молиться за определенных пациентов по их имени и первой букве фамилии (например, Чарли У.). Какие именно молитвы они будут произносить, прихожане решали самостоятельно; единственное условие, чтобы молитва содержала фразу «за успешную хирургическую операцию с быстрым выздоровлением и без осложнений».

Ну и?.. Станут ли молитвы экономически эффективным решением проблем здравоохранения в Америке? Наверное, нет. Исследователи не обнаружили какой-либо разницы в частоте осложнений в течение тридцатидневного послеоперационного периода между теми, за кого молились, и теми, за кого не молились. Критики этого эксперимента указывали на переменную, которую не учли ученые: молитвы, исходившие от других источников. Как резюмировала газета *The New York Times*: «Эксперты сказали, что это исследование не смогло преодолеть самое, пожалуй, крупное препятствие к изучению эффективности молитв: неизвестный объем молитв, получаемых каждым участником эксперимента от своих друзей, родственников, членов семьи и конгрегаций по всему миру, которые ежедневно молятся за страждущих».

Проведение экспериментов на людях может закончиться арестом или даже международным трибуналом. О такой «перспективе» никогда не следует забывать. Однако в области социальных наук всегда найдется место для статистических управляемых экспериментов с участием людей. Одним из самых знаменитых стало исследование *Tennessee's Project STAR*, в ходе которого изучалось влияние уменьшения количества учащихся в группах на степень усвоения ими знаний. В наши дни буквально все страны борются

за повышение качества своих систем образования. Если уменьшение количества учащихся в группах способствует более эффективному обучению, то при прочих равных условиях (*ceteris paribus*) общество должно направлять дополнительные средства на подготовку большего числа преподавателей, которые при этом понадобятся. В то же время дополнительные преподаватели — это дополнительные расходы (и весьма немалые); если учащиеся в небольших группах демонстрируют лучшие результаты по причинам, не *зависящим от размера этих групп*, то государственные деньги окажутся попросту выброшенными на ветер.

Связь между величиной учебной группы и эффективностью усвоения учащимися материала, как ни странно, изучить не так просто. Учебные заведения, формирующие компактные группы студентов, как правило, имеют больше ресурсов; это означает, что учащиеся и преподаватели в них отличаются от учащихся и преподавателей в учебных заведениях с большими размерами учебных групп. К тому же меньшие учебные группы обычно формируются в силу определенных причин. Например, директор может создать такую группу для отстающих учеников (в этом случае может наблюдаться ложная отрицательная зависимость между небольшим количеством учеников в классе и их успеваемостью). Еще один вариант: опытные преподаватели могут отдать предпочтение небольшим группам; в этом случае преимущество последних будет следствием выбора преподавателей.

Начиная с 1985 года в ходе исследования Tennessee's Project STAR был проведен управляемый эксперимент по выявлению последствий использования меньших учебных групп³. (В то время губернатором штата Теннесси был Ламар Александер, будущий министр образования в правительстве Джорджа Буша.) Ученики из семидесяти девяти разных подготовительных школ* были случайным образом распределены либо в небольшой класс (13–17 учеников), либо в обычный класс (22–25 учеников), либо в обычный класс с обычным преподавателем и помощником преподавателя. Учителей также распределили случайным образом по разным классам. До третьего класса включительно ученики оставались в тех классах, в которые они изначально попали. В процессе рандомизации влияние тех или иных жизненных реалий в какой-то мере нивелировалось. Одни ученики входили в эту систему в середине эксперимента, другие «выходили из игры». Кого-то из детей

* В Соединенных Штатах в подготовительных школах учатся дети пяти-шести лет.
Прим. перев.

переводили из одного класса в другой по дисциплинарным соображениям; кому-то из родителей удалось перевести своих чад в меньшие по численности классы. И так далее.

Тем не менее исследование Tennessee's Project STAR остается единственным рандомизированным тестом статистически и социально значимых результатов сокращения численности учебных групп. В целом при сдаче стандартизированных экзаменов успеваемость учеников в меньших по размеру классах оказалась на 0,15 среднеквадратических отклонений лучше, чем в обычных классах; а успеваемость чернокожих учащихся меньших по численности классов — в два раза выше, чем в обычных классах. А теперь плохая новость. Эксперимент Tennessee's Project STAR обошелся примерно в 12 миллионов долларов. Стоимость исследования влияния молитв на послеоперационные осложнения вылилась в 2,4 миллиона долларов. Самые точные и полезные исследования характеризуются тем же, что и все «точное и полезное», — очень высокими затратами.

Натурный эксперимент. Далеко не каждый располагает несколькими лишними миллионами долларов, которые он готов потратить на проведение крупномасштабного статистического исследования. А поскольку жизнь иногда совершенно случайно создает подопытную и контрольную группы, ученые стараются воспользоваться столь привлекательной ситуацией. Эта более экономичная альтернатива называется натурный эксперимент, яркий пример которого — наш рассказ о полицейских Вашингтона, приведенный в начале главы. Рассмотрим интересную, но сложную взаимосвязь между образованием и долголетием. Более образованные люди обычно живут дольше, даже если зафиксировать такие факторы, как уровень дохода и доступ к медицинским услугам. Как отмечала The New York Times: «Единственным социальным фактором, который, по признанию ученых, безусловно связан с продолжительностью жизни человека во всех странах, где изучалось его влияние, является образование. Уровень образования более важен, чем принадлежность к той или иной расе; он нивелирует любое влияние дохода»⁴. Правда, до сих пор речь шла лишь о корреляции. Но можно ли утверждать, что более высокий уровень образования, *ceteris paribus*, способствует улучшению здоровья? Если образование как таковое представить как «активирование», то можно ли утверждать, что повышение его уровня приведет к увеличению продолжительности вашей жизни?

Поначалу кажется, что исследовать этот вопрос практически невозможно, поскольку люди, стремящиеся повысить свой уровень образования,

отличаются от тех, кто этого не хочет. Разница между выпускниками средней школы и выпускниками колледжей не только в том, что вторые учились на четыре года дольше. У людей, испытывающих тягу к знаниям, вполне могут оказаться какие-то ненаблюдаемые (то есть не поддающиеся наблюдению) общие черты, которые могут объяснять их более высокую продолжительность жизни. Если это действительно так, то предлагать повышать образовательный уровень тем, кто к этому не склонен, бессмысленно: это не улучшит состояния их здоровья. То есть крепкое здоровье не зависит от уровня образования и может быть присуще той категории людей, для которых стремление учиться *естественно*.

Мы не можем провести рандомизированный эксперимент, чтобы решить эту головоломку, так как это предусматривало бы уход некоторых его участников из школы раньше, чем им хотелось бы. (Попробуйте кому-то объяснить, что он не сможет — никогда! — поступить в колледж, поскольку находится в контрольной группе!) Единственно возможным тестом причинно-следственной связи между уровнем образования и продолжительностью жизни мог бы стать эксперимент, в рамках которого достаточно большой сегмент совокупности оставался бы в школе дольше, чем это было бы в противном случае. Такой вариант по крайней мере приемлем с этической точки зрения, потому что мы рассчитываем на положительный эффект исследования. Тем не менее мы не можем заставлять детей оставаться в школе — это не американский стиль жизни.

Правда, элемент принуждения все же присутствует в реальной жизни. В каждом штате действуют законы, предусматривающие обязательный минимальный уровень образования, причем на разных исторических этапах *эти законы менялись*. Такого рода экзогенные (внешние) изменения обязательного минимального уровня образования — настоящий подарок для ученых. Адриана Ллерас-Муни, выпускница Колумбийского университета, усмотрела значительный исследовательский потенциал в том, что разные штаты вносили изменения в законы в разное время. Она выполнила исследование в историческом аспекте и изучила связь между периодами, когда штаты вносили изменения в законы об обязательном минимальном уровне образования, и последующими изменениями продолжительности жизни в штатах (для этого ей пришлось перелопатить горы данных по переписям населения). Тем не менее перед Ллерас-Муни оставалась серьезная методологическая проблема: увеличение продолжительности жизни жителей

какого-то штата после повышения обязательного минимального уровня образования нельзя объяснить именно повышением последнего. Продолжительность жизни людей со временем увеличивается (это связано с достижениями науки и медицины, улучшением качества питания и т. п.). В 1900 году люди жили дольше, чем в 1850-м, независимо от того, какие законы принимались теми или иными штатами.

Однако в распоряжении Ллерас-Муни был естественный управляющий фактор: штаты, в которых законы об обязательном минимальном уровне образования не менялись. Ее работа соответствует гигантскому лабораторному эксперименту, в котором жители Иллинойса были обязаны учиться в школе не менее семи лет, тогда как их соседи из штата Индиана имели право бросить учебу через шесть лет. Разница в том, что этот управляемый эксперимент стал возможен в силу исторической случайности — отсюда термин «натурный эксперимент».

Что же оказалось в результате? Продолжительность жизни взрослых, достигших тридцатипятилетнего возраста, увеличилась примерно на полтора года только за счет того, что они проучились в школе один дополнительный год⁵. Результаты, полученные Ллерас-Муни, удалось воспроизвести в других странах, где изменения законов об обязательном школьном образовании создавали аналогичные натурные эксперименты. Однако у нас остаются основания для некоторого скептицизма в отношении итогов этого эксперимента. Нам по-прежнему непонятен механизм, посредством которого продолжительность учебы повышает продолжительность жизни.

Неэквивалентный контроль. Иногда оптимальным вариантом для изучения эффекта активирования является создание нерандомизированных подопытной и контрольной групп. Наша надежда/ожидание заключается в том, что обе группы будут похожи в широком смысле слова даже несмотря на то, что обстоятельства не позволяют нам такой «статистической роскоши», как рандомизация. Радует уже то, что у нас есть подопытная и контрольная группы. Хотя плохо, что неслучайный характер назначения в эти группы создает как минимум возможность ошибки. Между подопытной и контрольной группами могут существовать ненаблюдаемые различия, связанные с распределением участников. Отсюда и название — «неэквивалентный контроль».

Тем не менее неэквивалентная контрольная группа представляет собой чрезвычайно полезный инструмент. Давайте поразмышляем над вопросом,

поставленным в подзаголовке к этой главе: обеспечивает ли значительное жизненное преимущество учеба в каком-либо из элитных колледжей или университетов? Очевидно, что выпускники Гарвардского, Принстонского и Дартмутского университетов имеют все основания рассчитывать на успех. В среднем они зарабатывают больше и имеют более широкие перспективы, чем студенты, посещающие менее престижные учебные заведения. (Исследование, выполненное PayScale.com в 2008 году, показало, что средняя заработная плата выпускников Дартмутского университета с рабочим стажем от десяти до двадцати лет составляет 134 000 долларов; это самая высокая заработная плата среди выпускников высших учебных заведений, обладающих степенью бакалавра. На втором месте по этому показателю — выпускники Принстонского университета, их средняя заработная плата 131 000 долларов.)⁶ Надеюсь, вы понимаете, что столь впечатляющие показатели средней заработной платы абсолютно ничего не говорят нам о ценности образования, получаемого в Дартмутском или Принстонском университете. Студенты этих учебных заведений, несомненно, талантливы — *именно поэтому им удалось туда поступить*. Но они наверняка преуспели бы в жизни независимо от того, в каком университете или колледже учились.

Нам неизвестно, каков эффект от учебы в таких учебных заведениях, как Гарвардский или Йельский университет. Их выпускники преуспевают в жизни потому, что просто необыкновенно талантливы? Или это университеты и колледжи сделали этих изначально талантливых молодых людей еще более продуктивными? Или, может быть, сыграло свою роль то и другое?

Мы не можем провести рандомизированный эксперимент, чтобы ответить на этот вопрос. Вряд ли среди выпускников средней школы найдется много желающих попасть в какой-либо выбранный наугад колледж; весьма сомнительно и то, что Гарвардский и Дартмутский университеты обрадуются идее произвольного набора студентов. Похоже, у нас нет какого-либо механизма проверки ценности эффекта рассматриваемого нами активирования. Талант под угрозой! Но экономисты Стейси Дейл и Алан Крюгер нашли способ ответить на этот вопрос, воспользовавшись* тем

* Исследователям нравится слово «воспользоваться» (exploit). Оно, в частности, применяется в значении «воспользоваться какой-либо возможностью, связанной с данными». Например, когда исследователи обнаруживают какой-либо натурный эксперимент, который создает подопытную и контрольную группу, они пишут, как собираются «воспользоваться разбросом в соответствующих данных».

фактом, что многие из выпускников средней школы подают документы сразу в несколько колледжей⁷. Некоторых выпускников принимают в то или иное элитное учебное заведение, и они решают поступить именно в него; другие в аналогичной ситуации выбирают менее престижный колледж или университет. Замечательно! Теперь у нас есть подопытная группа (студенты, которые предпочли элитные колледжи и университеты) и неэквивалентная контрольная группа (*достаточно талантливые студенты, тем не менее решившие поступить в менее престижный колледж или университет*)*.

Дейл и Крюгер изучили повторные данные о доходах в обеих группах. Это нельзя назвать идеальным сравнением яблок с яблоками, а уровень доходов, конечно же, не самое главное жизненное достижение человека, однако результаты, полученные Дейлом и Крюгером, должны развеять тревоги переутомленных напряженной учебой старшеклассников и их родителей. Выпускники элитных колледжей зарабатывают приблизительно столько же, сколько выпускники менее престижных учебных заведений — если, конечно, речь идет о людях примерно одинаковых способностей. Единственное исключение — студенты из малоимущих семей, которые зарабатывали больше, если учились в элитных колледжах или университетах. Подход Дейла и Крюгера кажется мне весьма элегантным способом

* Здесь существует вероятность ошибки. Обе группы студентов достаточно талантливы для того, чтобы быть принятыми в один из элитных колледжей или университетов. Однако одна группа студентов решила поступить в элитное учебное заведение, а другая предпочла менее престижный колледж или университет. Вторая группа студентов может быть менее мотивирована, менее трудолюбива или может отличаться в каких-то других, ненаблюдаемых отношениях. Если бы Дейл и Крюгер обнаружили, что студенты, поступившие в элитные учебные заведения, впоследствии зарабатывали больше, чем студенты, принятые в одно из элитных учебных заведений, но выбравшие менее престижный вуз, мы все же не могли бы быть уверены, что разница в их будущих доходах объясняется учебой в элитном учебном заведении, а не особенностями человека, получившего шанс поступить в элитное учебное заведение и воспользовавшегося им. Но в исследовании Дейла и Крюгера эта потенциальная ошибка не играет существенной роли. Дейл и Крюгер обнаружили, что студенты, которые поступили в элитные учебные заведения, впоследствии зарабатывали ненамного больше тех, кто выбрал какой-либо другой вариант продолжения учебы, *несмотря на то обстоятельство, что студенты, отказавшиеся поступить в элитные учебные заведения, могли обладать другими (помимо образования) особенностями, которые мешали им зарабатывать больше*. Как бы то ни было, упомянутая мною ошибка заставляет авторов данного исследования скорее *преувеличивать* денежные выгоды учебы в элитных колледжах и университетах, которые в любом случае оказываются несущественными.

отделить эффект активирования (четырёхлетняя учеба в элитном учебном заведении) от эффекта престижности учебного заведения (в эти учебные заведения принимают самых способных студентов). Подводя итоги исследования в статье для *The New York Times*, Алан Крюгер косвенно ответил на вопрос, поставленный в подзаголовке этой главы: «Вам придется признать тот факт, что ваша собственная мотивация, амбиции и способности будут определять ваш успех в большей степени, чем название колледжа в вашем дипломе»⁸.

Разница в различиях. Один из наилучших способов наблюдать причину и следствие — это сделать что-то, а затем посмотреть, к чему это действие приведет. Именно так маленькие дети (а подчас и взрослые) познают окружающий мир. Например, мои дети довольно быстро поняли, что если бросать куски пищи из одного конца кухни в другой (причина), то собака будет носиться как угорелая по кухне за этими кусками (следствие). Вполне возможно, что та же сила наблюдения помогает человеку в течение жизни. Если сокращение налогов приводит к оздоровлению экономики, значит, оно могло исправить в ней ситуацию.

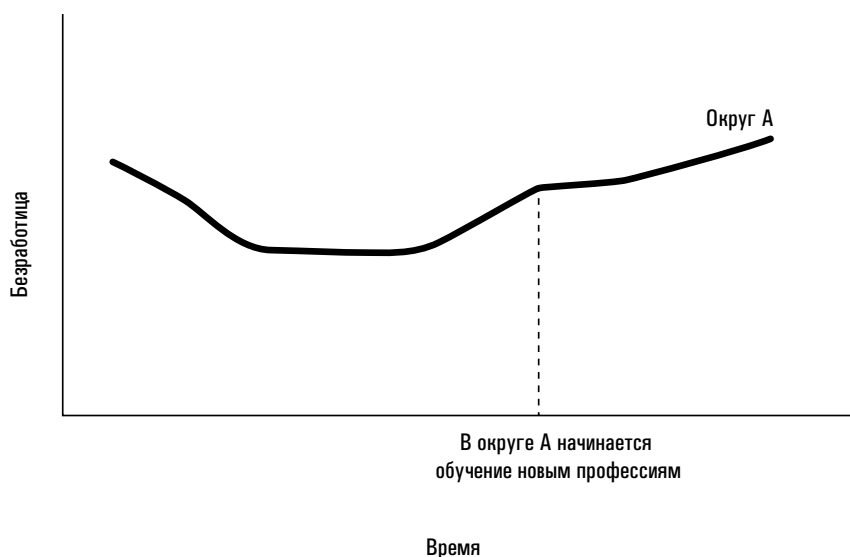
Может быть, может быть... Весьма внушительный подводный камень такого подхода заключается в том, что жизнь, как правило, гораздо сложнее швыряния обглоданных куриных ножек из одного конца кухни в другой. Да, на каком-то этапе мы могли сократить налоги, однако примерно на том же этапе разворачивались другие «активизации». В колледжи могло поступить большее число женщин, интернет и другие технологические новшества привели к повышению производительности американских рабочих, стоимость китайской национальной валюты оказалась заниженной, профессиональный бейсбольный клуб *Chicago Cubs* уволил своего генерального менеджера и т. д. Все, что произойдет после сокращения налогов, нельзя приписывать исключительно сокращению налогов. Проблема любого анализа, который можно отнести к категории «до и после», заключается в том, что если одно событие (событие А) произошло после какого-то другого (события В), то это отнюдь не означает, что событие В послужило причиной события А.

Подход, который мы обозначили как «разница в различиях», может помочь идентифицировать последствия некоторой активизации при выполнении двух условий. Во-первых, мы анализируем данные «до» и «после» для той группы или юрисдикции (подведомственной области), которые

получили соответствующее активирование, например данные по безработице для округа, реализовавшего программу обучения новым профессиям. Во-вторых, сравниваем эти данные с показателями безработицы за тот же период времени для какого-либо похожего округа, который не реализовал данную программу.

Предполагается (что довольно важно), что две группы, используемые для такого анализа, в целом сопоставимы — за исключением активирования; таким образом, любое значимое различие в исходах между этими группами может быть обусловлено оцениваемой нами программой или политикой. Допустим, что один округ в штате Иллинойс реализует программу обучения новым профессиям в надежде снизить уровень безработицы. Однако в течение двух последующих лет безработица продолжает расти. Означает ли это, что программа провалилась? Как это выяснить?

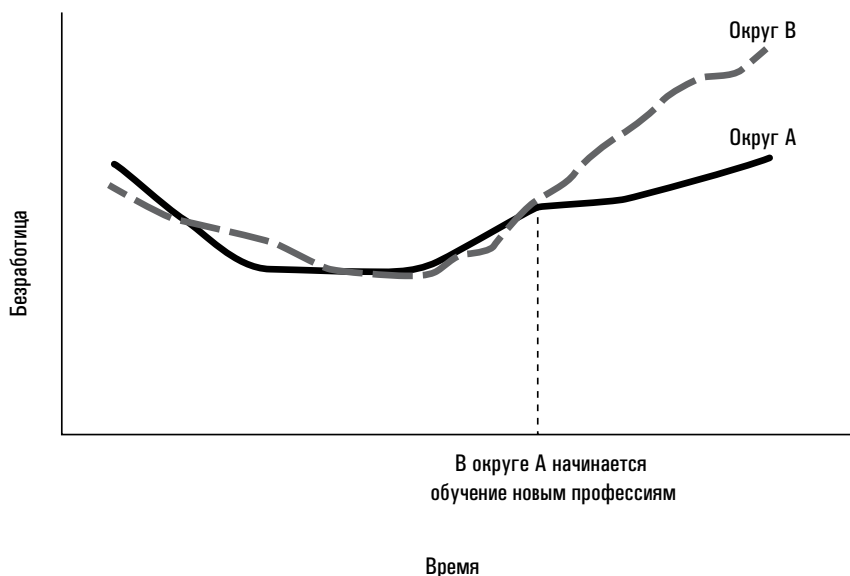
Влияние обучения новым профессиям на уровень безработицы в округе А



Разумеется, на этой арене могут действовать множество экономических сил, в том числе и вероятность продолжительного экономического спада. Подход «разница в различиях» сравнил бы изменение уровня безработицы в течение какого-то времени в оцениваемом нами округе с уровнем безработицы в каком-либо из соседних округов, где не реализуется программа обучения новым профессиям. Эти два округа должны быть похожи во всех

остальных важных аспектах, таких как структура промышленности, демографическая картина и т. п. Как уровень безработицы в округе, где реализуется программа обучения новым профессиям, изменяется со временем *в сравнении с округом, в котором такая программа не внедрена*? Мы можем оценить эффект данной программы, сравнив изменения в этих округах за время, в течение которого проводилось данное исследование («разница в различиях»). Другой округ в этом исследовании действует, по сути, как контрольная группа, что позволяет нам воспользоваться данными, собранными до и после воздействия. Если контрольная группа выбрана правильно, то она будет подвергаться воздействию тех же более широких сил, что и подопытная группа. Подход «разница в различиях» может быть особенно поучителен, когда активирование сперва оказывается неэффективным (после начала реализации программы обучения новым профессиям уровень безработицы повышается), хотя контрольная группа показывает, что в отсутствие активации эта тенденция могла быть еще хуже.

Влияние обучения новым профессиям на уровень безработицы в округе А, для сравнения используется округ В



Анализ на основе разрывности. Один из способов создать подопытную и контрольную группы — сравнить исходы для какой-то группы, которая едва подходит для воздействия или активирования, с исходами для группы,

которая едва превысила порог непригодности и не получает активирования. Те, кто окажется чуть выше или чуть ниже некоторого случайно выбранного предела, например результата сдачи экзаменов или минимального дохода семьи, будут почти идентичны во многих важных отношениях; то обстоятельство, что одна группа получила активирование, а другая — нет, является, по сути, делом случая. В итоге мы можем сравнить их исходы таким способом, который позволял бы сделать значимые выводы относительно эффективности соответствующего воздействия.

Допустим, какому-то учебному округу требуется летняя школа, в которой могли бы заниматься отстающие ученики. Руководство округа хотело бы знать, представляет ли собой какую-либо долгосрочную учебную ценность такая программа летнего обучения. Как обычно, простое сравнение учеников, посещающих и не посещающих эту школу, было бы не только бесполезно, но даже вредно. Ученики, которые посещают такую школу, ходят туда *именно потому, что являются отстающими*. Даже если программа обучения в этой школе весьма эффективна, успеваемость ее учеников в долгосрочной перспективе все же, по-видимому, будет хуже, чем успеваемость учеников, не нуждающихся в дополнительных занятиях. Мы хотим знать, какой будет успеваемость отстающих учеников после посещения летней школы *по сравнению с их успеваемостью в случае, если бы они ее не посещали*. Да, мы могли бы провести нечто вроде управляемого эксперимента, в ходе которого случайным образом отобрали бы отстающих учеников для посещения летней школы, и, соответственно, отстающих учеников, которые не будут в нее ходить (контрольная группа). Однако это бы предполагало отказ контрольной группе в доступе к программе обучения в летней школе, между тем как мы предполагаем, что обучение в летней школе будет полезно *всем* отстающим ученикам.

Вместо этого подопытная и контрольная группы создаются путем сравнения учеников, которые оказались чуть ниже порога для летней школы, с учениками, которые находятся чуть выше этого порога. Это можно представлять себе так: ученики, провалившие экзамены по итогам полугодия, существенно отличаются от учеников, которые успешно их сдали. Но ученики, набравшие 59% от максимальной суммы баллов (что не дает им права перейти в следующий класс), не *столь уж сильно* отличаются от учеников, у которых 60% от максимальной суммы баллов (проходной балл). Если для тех, кто провалил экзамены по итогам полугодия, организовать некую программу

обучения (например, обязательная подготовка к выпускным экзаменам под руководством преподавателя), то у нас появились бы приемлемые подопытная и контрольная группы, где бы мы сравнили результаты выпускных экзаменов тех, кто недобрал самую малость до проходного балла (и получил возможность готовиться под руководством преподавателя), с результатами тех, кто едва превысил проходной балл в ходе сдачи экзаменов по итогам полугодия (и не пользовался помощью преподавателя).

Этот подход был применен, чтобы определить эффективность тюремного заключения малолетних преступников в качестве сдерживающего фактора от повторного совершения преступлений. Очевидно, при выполнении такого анализа нельзя просто сравнивать уровень повторной преступности среди несовершеннолетних, получавших тюремные сроки, и тех, кто отделался более легким наказанием. *Малолетние преступники, которых сажают в тюрьму, как правило, совершают более серьезные преступления, чем малолетние преступники, получающие более легкие приговоры; именно поэтому их и сажают за решетку.* Естественно, мы не можем создать подопытную и контрольную группы путем случайного назначения тюремных сроков (если не хотите на двадцать пять лет угодить за решетку за то, что повернули на красный свет). Исследователь Рэнди Хьялмарссон, которая в настоящее время работает в Лондонском университете, воспользовалась статистикой вынесенных приговоров для малолетних преступников в штате Вашингтон, чтобы понять, как влияет тюремное заключение на будущее криминальное поведение этих людей. В частности, она сравнила уровень повторной преступности среди несовершеннолетних, получивших тюремный срок буквально «на грани» действующего уголовного законодательства, с уровнем повторной преступности среди тех малолетних преступников, кто избежал тюремного срока также буквально «на грани» (то есть «отделался выплатой штрафа или условным сроком»)⁹.

Система уголовных наказаний в штате Вашингтон создает для каждого осужденного некую систему координат, используемую для администрирования меры наказания преступника. Ось X измеряет признанные судом прежние преступления данного осужденного. Например, каждое прежнее тяжкое уголовное преступление оценивается в один балл; каждое прежнее уголовное преступление, не относящееся к категории тяжких, оценивается в четверть балла. Сумма баллов округляется до ближайшего меньшего целого значения (почему это важно, станет понятно из дальнейшего объяснения).

По оси Y измеряется тяжесть текущего преступления по шкале от Е (наименее тяжкое) до А+ (самое тяжкое). Приговор осужденному вычисляется (в буквальном смысле) путем нахождения подходящей клетки в такой системе координат. Осужденный, предыдущие преступления которого оцениваются в два балла, совершивший тяжкое уголовное преступление класса В, получает от пятнадцати до тридцати шести месяцев тюремного заключения. Осужденный, предыдущие преступления которого оцениваются лишь в один балл, совершивший такое же уголовное преступление, не отправляется за решетку. Именно такая разрывность шкалы наказания определила выбор стратегии исследования. Рэнди Хьялмарссон сравнила исходы для осужденных преступников, которые оказались чуть выше и чуть ниже порога получения тюремного срока. Как объясняет Хьялмарссон в своей статье: «Если есть два преступника с текущим преступлением класса С+ и суммами баллов за предыдущие приговоры $2\frac{3}{4}$ и 3, то лишь последний из этих двоих будет отправлен в тюрьму».

С точки зрения данного исследования эти два человека, по сути, одинаковы — пока один из них не попадет за решетку. С этого момента их поведение резко разнится. Вероятность осуждения малолетних преступников, попавших за решетку, за какое-либо другое преступление после их освобождения из тюрьмы оказывается существенно ниже.

Нас интересуют прежде всего действенные способы анализа. Это касается медицины, экономики, бизнеса, уголовного судопроизводства и всего остального. Все же причинно-следственные зависимости — крепкий орешек, который не так-то просто расколоть даже в случаях, когда причина и следствие кажутся совершенно очевидными. Чтобы уяснить подлинную эффективность того или иного активирования, нужно знать, «что было бы в противном случае», то есть в отсутствие соответствующего активирования, или воздействия. Зачастую узнать это не так-то легко, а порой и невозможно. Рассмотрим пример не из области статистики. Оказалась ли Америка в большей безопасности после вторжения в Ирак?

Существует лишь один честный ответ на этот вопрос: мы никогда не узнаем это наверняка. А не узнаем потому, что нам не дано знать, что случилось бы, если бы Америка не вторглась в Ирак. Да, Соединенные Штаты не нашли в Ираке оружия массового поражения. Но не исключено, что на следующий день после того, как США *не вторглись в Ирак*, у Саддама Хусейна во время

принятия душа могли бы возникнуть следующие мысли: «А не обзавестись ли мне водородной бомбой? Может быть, Северная Корея продаст мне парочку?» Кто знает, что случилось бы потом...

Правда, на следующий день после того как Соединенные Штаты *не вторглись в Ирак*, Саддам Хусейн во время принятия душа мог бы подумать: «А не обзавестись ли мне...», после чего ему на голову могла бы упасть отвалившаяся от стены кафельная плитка и он бы умер, так и не додумав до конца интересную мысль о возможности покупки водородной бомбы.

В этом случае мир избавился бы от Саддама Хусейна, не понеся колоссальных издержек, связанных с вторжением Америки в Ирак. Короче говоря, вариантов — множество, но никто никогда не сможет сказать наверняка, что случилось бы, если бы Америка не вторглась в Ирак.

Задача любой программы статистического оценивания — узнать, «что было бы в противном случае». Только так мы можем измерить эффективность того или иного активирования, или воздействия. В случае статистического управляемого эксперимента роль «что было бы в противном случае» исполняет контрольная группа. В случаях, когда проведение управляемого эксперимента нецелесообразно или невозможно по этическим соображениям, нужно найти какой-то другой способ приближения того, «что было бы в противном случае». Наше понимание окружающего мира зависит от нахождения разумных способов решения этой задачи.

Заключение

Пять вопросов, на которые поможет ответить статистика

Еще не так давно информацию было гораздо труднее собирать и гораздо дороже анализировать. Вообразите, каких титанических усилий стоило изучить данные о миллионе транзакций по кредитным картам при отсутствии (а ведь это было всего лишь несколько десятков лет тому назад!) персональных компьютеров, позволяющих оперативно это сделать, и наличии исключительно бумажных товарных чеков. Во время Великой депрессии не существовало официальной статистики, помогающей оценить глубину экономических проблем. Правительство не собирало официальную информацию о валовом внутреннем продукте (ВВП) и безработице, поэтому руководители государства ориентировались в экономической ситуации примерно так же, как человек, не имеющий компаса, в незнакомом лесу. В 1930 году Герберт Гувер, исходя из неточных и устаревших данных, имевшихся в его распоряжении, объявил об окончании Великой депрессии. В своем послании о положении в стране он сообщил ее гражданам о двух с половиной миллионах безработных, хотя на самом деле их было пять, причем это число каждую неделю увеличивалось на сто тысяч. Недавно Джеймс Суrowецки написал в еженедельнике *The New Yorker*, что «Вашингтон вершил тогда свою политику втемную»¹.

Сейчас у нас нет недостатка в данных. И по большому счету это благо. Статистические инструменты пригодятся вам для решения ряда самых острых социальных проблем. Именно поэтому я счел уместным завершить книгу вопросами, а не ответами. Когда будете анализировать огромные объемы информации, не забывайте о пяти важных (и на первый взгляд не связанных друг с другом) вопросах, получение социально значимых ответов на которые предполагает использование многих инструментов, описанных в этой книге.

Какое будущее ждет американский футбол?

В 2009 году Малкольм Гладуэлл в одной из статей, опубликованных в *The New Yorker*, поставил вопрос, который сначала мне показался нарочито сенсационным и провокационным: чем отличаются собачьи бои от футбола?² Связь между тем и другим проистекает из того обстоятельства, что куортербек Майкл Вик, в свое время отсидевший в тюрьме за участие в организации собачьих боев, был восстановлен в Национальной футбольной лиге после того, как стало известно, что травма головы, полученная им во время футбольного матча, впоследствии могла вызвать депрессию, частичную потерю памяти, слабоумие и прочие неврологические проблемы. Ключевая мысль Малкольма Гладуэлла, которую он хотел подчеркнуть, состояла в том, что общим для профессионального футбола и собачьих боев является оказываемое ими разрушительное воздействие на психику их участников. Дочитав статью Гладуэлла до конца, я пришел к выводу, что ее автор поднял очень интересную тему.

Вот что об этом известно. Существуют многочисленные доказательства, что сотрясения мозга и другие травмы, связанные с игрой в футбол, могут вызывать серьезные и долговременные неврологические расстройства. (Аналогичные явления наблюдаются у боксеров и хоккеистов.) Многие из известных профессиональных футболистов, завершивших карьеру, рассказывают о таких «постфутбольных» недугах, как депрессия, частичная потеря памяти и слабоумие. Пожалуй, наиболее впечатляющей можно считать историю Дейва Дайерсона, бывшего ключевого игрока команды *Chicago Bears*, который покончил жизнь самоубийством, выстрелив себе в грудь. Он завещал своей семье отдать его мозг на экспертизу. В ходе телефонного опроса тысячи случайно выбранных бывших игроков НФЛ старше пятидесяти лет (отыгравших в лиге не менее трех сезонов) 6,1% из них сообщили, что у них диагностировали слабоумие, болезнь Альцгеймера или какое-то другое заболевание, связанное с частичной потерей памяти. Это более чем в пять раз превышает соответствующий средний показатель в США для данной возрастной группы. Среди более молодых игроков частота подобного диагноза в девятнадцать раз превышает соответствующий средний показатель. Сотни бывших игроков НФЛ обвиняют руководство лиги и производителей футбольной экипировки (особенно это касается изготовителей шлемов) в сокрытии информации о реальной опасности травм головы³.

Энн Макки — один из исследователей, изучающих влияние травм мозга, — руководит лабораторией нейропатологии в госпитале для ветеранов в Бедфорде. (Кроме того, Энн выполняет исследования в области нейропатологии в рамках Framingham Heart Study.) В ходе экспериментов д-р Макки выявила в мозге спортсменов (в частности, боксеров и профессиональных футболистов), перенесших травму мозга, накопление аномальных белков под названием тау, что приводит к состоянию, известному как хроническая травматическая энцефалопатия, или СТЕ, — прогрессирующему неврологическому заболеванию, у которого много таких же проявлений, как у болезни Альцгеймера.

Другие исследователи обнаружили связь между занятиями профессиональным футболом и травмами головного мозга. Кевин Гускевич, руководитель исследовательской программы Sport Concussion Research Program в Университете Северной Каролины, установил датчики внутри шлемов футболистов, выступающих за команду Северной Каролины, чтобы определить силу и природу получаемых игроками ударов по голове. Согласно данным Гускевича, футболисты регулярно получают удары по голове с силой, эквивалентной удару водителя головой о лобовое стекло при столкновении автомобиля с препятствием на скорости двадцать пять миль в час.

А вот то, чего мы не знаем. Являются ли имеющиеся сегодня сведения о травмах мозга репрезентативными с точки зрения долгосрочных неврологических рисков, с которыми приходится сталкиваться всем профессиональным футболистам? Или речь идет лишь о каком-то «кластере» неблагоприятных исходов, который представляет собой не что иное, как статистическую аберрацию? Даже если футболисты действительно подвержены повышенному риску развития неврологических заболеваний по завершении профессиональной карьеры, нам все же необходимо проверить наличие соответствующей причинно-следственной связи. Не может ли так случиться, что мужчины, играющие в футбол (а также те, кто занимается боксом или хоккеем), склонны к развитию таких болезней в силу определенных особенностей своего организма? А вдруг еще какие-то факторы, например использование стероидов, влияют на подобный исход?

Если накопленная нами информация подтвердит наличие четкой причинно-следственной связи между занятиями профессиональным футболом и долгосрочными травмами мозга, то игрокам (а также родителям молодых игроков), тренерам, юристам, руководителям НФЛ и даже, возможно,

государственным регуляторным органам придется найти ответ на следующий ключевой вопрос: нельзя ли изменить правила игры в футбол так, чтобы минимизировать (или вообще исключить) риск получения травмы головы? Если нет, то как тогда быть? Вот что в действительности скрывается за сравнением футбола и собачьих боев, которым меня поразил Малкольм Гладуэлл. Он объясняет, что собачьи бои вызывают протест общества потому, что владелец собаки осознанно подвергает своего питомца испытанию, кульминацией которого являются страдания, боль и даже смерть. «Для чего все это нужно? — спрашивает автор. — На потеху зрителям и ради шанса сорвать куш? В XIX столетии собачьи бои пользовались большой популярностью у американской публики. Но в наше время этот вид развлечения неприемлем по этическим соображениям».

Сегодня практически каждый тип статистического анализа, описанный в этой книге, используется для того, чтобы выяснить, есть ли будущее у профессионального американского футбола в его нынешнем виде.

Что вызывает резкий рост заболеваемости аутизмом в наши дни (если такая причина вообще существует)?

В 2012 году Centers for Disease Control сообщил, что у одного ребенка из каждых 88 американских детей диагностируется расстройство аутистического спектра (РАС) (этот вывод был сделан на основе данных за 2008 год)⁴. Таким образом, за менее чем десять лет частота диагностирования аутизма значительно (почти в два раза) повысилась по сравнению с «1 из 110» в 2006 году и «1 из 150» в 2002-м. Заболевания аутизмом (autism spectrum disorders — ASD) представляют собой группу заболеваний вследствие порока развития головного мозга, которые характеризуются проблемами с социализацией и общением, а также атипичным поведением. Слово spectrum («спектр») в названии этой группы заболеваний указывает на то, что аутизм охватывает широкий спектр состояний, определяемых поведением⁵. Частота диагностирования РАС у мальчиков в пять раз выше, чем у девочек (это означает, что заболеваемость аутизмом у мальчиков даже выше, чем «1 из 88»).

Первый интригующий статистический вопрос таков: с чем мы имеем дело в данном случае? С эпидемией аутизма, «эпидемией диагностирования» или и тем и другим, вместе взятыми?⁶ В предыдущие десятилетия дети, страдающие аутизмом, имели симптомы, которые вообще могли

не диагностироваться, а проблемы в развитии таких детей зачастую описывались как «неспособность к обучению». В наши дни врачи, родители и преподаватели гораздо больше осведомлены о симптомах РАС, что, естественно, приводит к увеличению случаев диагностирования аутизма независимо от того, действительно ли он стал встречаться гораздо чаще по сравнению с предыдущими десятилетиями.

В любом случае необычайно высокая заболеваемость аутизмом в наши дни представляет серьезную проблему для семей, школ и общества в целом. Средняя стоимость решения проблем, связанных с РАС, для отдельно взятого человека составляет 3,5 миллиона долларов⁷. Несмотря на очевидно эпидемический характер данного заболевания, нам на удивление мало известно о причинах его возникновения. Томас Инзел, директор National Institute of Mental Health, говорит: «Может быть, виной всему мобильные телефоны? Ультразвук? Низкокалорийные газированные напитки? У каждого из родителей есть на сей счет собственная теория. Наука пока не может ответить на этот вопрос»⁸.

Чем особенным или уникальным характеризуется нынешняя жизнь детей, страдающих аутизмом? Каковы самые существенные физиологические различия между детьми-аутистами и обычными детьми? Является ли заболеваемость аутизмом разной в разных странах? Если да, то почему? Традиционная статистическая «детективная» работа позволяет найти ключи к ответам на эти вопросы.

Недавнее исследование, выполненное учеными Калифорнийского университета в Дэвисе, выявило в Калифорнии десять мест, где заболеваемость аутизмом вдвое превышает соответствующий показатель в соседних регионах; каждый из этих кластеров аутизма представляет собой территорию с плотной концентрацией белых родителей с высоким образовательным уровнем⁹. Является ли это ключом к разгадке причин аутизма или речь идет о случайном совпадении? Или это отражает тот факт, что лишь относительно привилегированные семьи могут себе позволить затраты, связанные с диагностированием аутизма? Те же ученые проводят исследование, в ходе которого собирают образцы пыли в домах 1300 семей, где есть дети-аутисты, чтобы проверить эту пыль на наличие химических веществ или других загрязнений окружающей среды, которые могут вызывать заболевание аутизмом.

Другие исследователи идентифицировали возможный генетический компонент аутизма, исследовав РАС среди однояйцевых и двуяйцевых

близнецов¹⁰. Вероятность того, что двое детей в одной и той же семье больны аутизмом, выше среди однояйцевых близнецов (у которых одинаковая организация генетического материала), чем среди двуяйцевых (генетическое подобие которых такое же, как у обычных братьев или сестер). Это открытие не исключает влияния существенных факторов окружающей среды или возможного взаимодействия между ними и генетическими факторами. В конце концов, в заболеваниях сердечно-сосудистой системы присутствует значительная генетическая составляющая, но совершенно очевидно, что важную роль также играют курение, режим питания, физическая активность и многие другие поведенческие модели и факторы окружающей среды.

Одним из самых осязаемых вкладов, вносимых статистическим анализом, до сих пор считалась возможность развенчать ложные причины, многие из которых возникли из-за путаницы между корреляцией и причинно-следственными зависимостями. Аутизм зачастую внезапно проявляется между первым и вторым днями рождения ребенка. Это привело к широко распространенному мнению, будто вакцинация детей — особенно тройная вакцина от кори, свинки и краснухи (КСК) — является причиной роста заболеваемости аутизмом. Дэн Бартон, член Конгресса от штата Индиана, в интервью газете *The New York Times* рассказал: «Мой внук получил девять прививок в один день; семь из них содержали тимеросал, который, как известно, состоит на 50% из ртути. Спустя непродолжительное время у внука был диагностирован аутизм»¹¹.

Ученые полностью опровергли ложную зависимость между тимеросалом и РАС. Когда тимеросал изъяли из вакцины КСК, заболеваемость аутизмом не снизилась; к тому же в странах, где эта вакцина никогда не применялась, заболеваемость аутизмом не ниже, чем в странах, где она регулярно используется. Тем не менее эта ложная связь укоренилась в сознании людей, что заставило некоторых родителей отказаться от вакцинации своих детей. Ирония судьбы в том, что это не обеспечивает защиты от аутизма, но зато подвергает детей риску подхватить другие серьезные болезни (и способствует их распространению среди остального населения).

Аутизм — одна из величайших медицинских и социальных проблем нашего времени. А мы так мало знаем о нем по сравнению с его огромным (и, возможно, растущим) влиянием на наше общее благополучие. Исследователи используют буквально каждый из инструментов, описанных в этой книге (а также многие другие), чтобы изменить нынешнее положение вещей.

Как выявить и наградить хороших преподавателей и хорошие школы?

Безусловно, нам нужны хорошие школы, так же как нам нужны и хорошие преподаватели, без которых такие школы невозможны. Следовательно, нам нужно вознаграждать хороших преподавателей и хорошие школы, а также увольнять плохих преподавателей и закрывать плохие школы.

Но как выявить хороших преподавателей и хорошие школы?

Результаты сдачи экзаменов могут служить объективным показателем успеваемости учащихся. Тем не менее мы знаем, что некоторые ученики сдают стандартизованные тесты лучше своих сверстников по причинам, не имеющим ничего общего с тем, что происходит в стенах конкретного класса или школы. Казалось бы, самым простым решением здесь стало бы оценивание школ и преподавателей на основе *прогресса*, которого добиваются учащиеся за определенный период времени. Как изменились их знания с момента начала их занятий в определенном классе у конкретного преподавателя? Какими они были год назад? Разница между этими объемами знаний представляет собой «добавленную стоимость» соответствующего класса.

Чтобы получить более точное представление об этой «добавленной стоимости», мы можем воспользоваться статистическими инструментами, приняв во внимание демографические характеристики учащихся в конкретном классе, такие как расовая принадлежность, уровень дохода в семье и успеваемость по другим тестам (что может быть показателем сообразительности). Если преподаватель добился значительных успехов в работе с учащимися, которые не блистали знаниями в прошлом, то его следует считать весьма эффективным.

Вот так-то! Теперь мы можем оценить качество учителя со статистической точностью. А хорошие школы — это, безусловно, школы, где работают такие учителя.

Как эти удобные статистические оценки функционируют на практике? В 2012 году городские власти Нью-Йорка проявили инициативу и опубликовали рейтинги всех 18 000 преподавателей государственных школ на основе «оценки добавленной стоимости», измеренной по результатам сдачи экзаменов учениками с учетом их различных характеристик¹². Газета Los Angeles Times опубликовала аналогичную совокупность рейтингов преподавателей Лос-Анджелеса в 2010 году.

И в Нью-Йорке, и в Лос-Анджелесе реакция на эти публикации была бурной и противоречивой. Арни Дункан, министр образования США, в целом поддержал такого рода оценки на основе «добавленной стоимости», заявив, что они хороши хотя бы тем, что позволяют получить какую-то информацию в областях, где раньше ее вообще не было. После публикации данных по Лос-Анджелесу Арни Дункан сказал газете *The New York Times*: «Молчание — это вообще не вариант». Администрация Обамы предоставила штатам финансовые средства для разработки показателей добавленной стоимости, на основе которых можно было бы оплачивать труд преподавателей и продвигать их по службе. Сторонники данных способов оценивания подчеркивают, что это огромный шаг вперед по сравнению с системами, в которых труд всех преподавателей оплачивается в соответствии с единой шкалой окладов, не учитывающей такие «мелочи», как профессионализм.

С другой стороны, многие эксперты предупреждают, что подобные системы оценивания характеризуются большими допустимыми пределами погрешности и могут давать ошибочные результаты. Профсоюз преподавателей Нью-Йорка потратил более 100 000 долларов на рекламную кампанию в газетах, проводившуюся под лозунгом «Так оценивать работу преподавателей нельзя»¹³. Оппоненты утверждают, что «оценивание добавленной стоимости» создает ложное впечатление высокой точности, причем такие оценки могут вольно трактоваться родителями и государственными чиновниками, не понимающими ограничений подобного оценивания.

Похоже, это тот самый случай, когда правы — в определенной степени — и те и другие. Даг Стайгер, экономист Дартмутского колледжа, широко использующий в работе данные «добавленной стоимости» преподавателей, предостерегает, что в этих данных, в силу самой их природы, многовато «шума». Результаты по конкретному учителю зачастую основываются на результатах одного экзамена. При этом действие множества разнообразных факторов — от особенно «трудной» группы учащихся до сломавшегося в день сдачи экзамена кондиционера в классе — может вести к флуктуациям. Корреляция в эффективности отдельно взятого преподавателя, которая использует эти индикаторы, от года к году составляет лишь 0,35. (Интересно, что корреляция в эффективности игроков Высшей бейсбольной лиги от года к году также составляет примерно 0,35; для хиттеров она измеряется средним уровнем достижений, а для питчеров — средним числом зачетных пиребежек.)¹⁴

По мнению Стайгера, данные об эффективности учителей полезны, но это лишь один из инструментов, применяемых в процессе оценивания их профессионализма. Эти данные оказываются «менее шумными», когда исследователи располагают информацией по конкретному преподавателю за много лет, особенно если он работал с разными группами учащихся (точно так же, как мы можем сделать более достоверные выводы о достижениях спортсмена, имея данные о большем числе игр и сезонов, в которых он участвовал). В случае рейтингов преподавателей Нью-Йорка руководство системы образования предупреждало о необходимости правильного использования данных «добавленной стоимости» и о присущей им ограниченности. Однако широкой общественности не было известно об этих предупреждениях. В результате предложенные оценки учителей зачастую рассматриваются как определитель «хороших» и «плохих» преподавателей, заслуживающий полного доверия. Нам вообще нравятся рейтинги — вспомните хотя бы рейтинги колледжей, опубликованные в U.S. News & World Report, — даже когда соответствующие данные не поддерживают такую точность.

Стайгер делает предупреждение несколько иного рода: нам следовало заранее удостовериться в том, что оцениваемые нами исходы (например результаты стандартизированного теста) действительно соответствуют тому, что нас интересует в долгосрочной перспективе. Некоторые уникальные данные, которые приводит Академия ВВС США, свидетельствуют о том, что блестящие результаты экзаменов в будущем могут такими не казаться. Академия ВВС США, подобно другим военным учебным заведениям, произвольно распределяет своих кадетов по разным секциям стандартизированных базовых курсов, таких как введение в математический анализ. Подобная рандомизация устраняет любое влияние потенциального отбора при сравнении эффективности преподавателей; спустя какое-то время мы можем исходить из того, что все преподаватели получают студентов с примерно одинаковыми способностями (в отличие от большинства университетов, где студенты с разными способностями могут выбирать для себя те или иные курсы). Кроме того, Академия ВВС США использует в каждой части конкретного курса одинаковые учебные программы и экзамены. Скотт Каррелл и Джеймс Уэст, профессора Калифорнийского университета в Дэвисе и Академии ВВС США, воспользовались этой элегантной системой организации процесса подготовки студентов, чтобы ответить на один из самых важных вопросов в высшем образовании: какие из профессоров самые эффективные?¹⁵

Ответ: *профессора с меньшим опытом преподавательской работы и меньшим числом научных степеней от новомодных университетов*. Студенты таких профессоров, как правило, очень позитивно отзываются о них и демонстрируют более высокие результаты сдачи стандартизированных тестов по вводным курсам. Понятно, что эти молодые и мотивированные преподаватели относятся к своей работе с большим энтузиазмом, чем старые, «замшелые» профессора с докторскими степенями от престижных университетов (например Гарвардского). Не исключено, что эти «старые зубры» все еще пользуются теми же пожелтевшими от времени конспектами лекций, что и в далеком 1978 году; возможно, они полагают, что PowerPoint — это нечто вроде энергетического напитка (правда, они могут и не знать, что это такое). Очевидно, эти данные говорят нам о том, что этих «динозавров» давно пора уволить или по крайней мере предоставить им шанс красиво уйти со сцены.

Впрочем, не будем торопиться. И спешить с увольнениями. Опыт Академии ВВС США позволил сделать еще один важный вывод — об успеваемости студентов *на более протяженном горизонте*. Скотт Каррелл и Джеймс Уэст выяснили, что в области математики и точных наук студенты, у которых были более опытные (и более титулованные) преподаватели вводных курсов, *демонстрируют лучшую успеваемость в последующих обязательных курсах*, чем студенты, обучавшиеся у менее опытных преподавателей вводных курсов. Одно из логических объяснений заключается в том, что менее опытные преподаватели в большей степени склонны «натаскивать на экзамен» по соответствующему вводному курсу. Это приносит впечатляющие результаты на экзаменах и моральное удовлетворение студентам, что выливается в высокие оценки, которые они выставляют своим преподавателям.

Между тем, старые замшелые профессора (которых мы уже были готовы уволить) уделяют меньше внимания экзаменам, сосредоточиваясь на важных положениях своего курса, что положительно влияет на успеваемость студентов в процессе прохождения ими последующих обязательных курсов, а также в их практической деятельности по завершении учебы в Академии ВВС США.

Разумеется, мы должны оценивать деятельность школьных учителей и профессоров. Нам нужно лишь убедиться, что мы делаем это *правильно*. Задача на перспективу (коренящаяся в статистике) состоит в разработке такой системы, которая вознаграждала бы реальную добавленную стоимость преподавателя, создаваемую им для учеников.

С помощью каких инструментов лучше всего бороться с глобальной бедностью?

Мы очень мало знаем о том, как сделать бедные страны богаче. Да, мы понимаем, что отличает богатые страны от бедных, например уровень образования и качество кадров, управляющих бедными странами. Правда и то, что за последние несколько десятилетий такие страны, как Индия и Китай, совершили радикальные преобразования в области экономики. Но даже при наличии этих знаний мы не можем с уверенностью сказать, какие шаги следует предпринять, чтобы повысить уровень благосостояния таких стран, как Мали и Буркина-Фасо. С чего же начать?

Французский экономист Эстер Дафло трансформирует знания о глобальной бедности путем использования старого инструмента для решения новых задач. Речь идет о статистическом управляемом эксперименте. Дафло — преподаватель Массачусеттского технологического института — в буквальном смысле проводит эксперименты с разными формами воздействия, пытаясь улучшить жизнь бедняков в развивающихся странах. Например, одна из давних проблем школ в Индии — прогулы учителей; особенно это касается небольших провинциальных школ с единственным учителем. Дафло и ее соавтор Рема Ханна провели тестирование на основе использования технологий на произвольной выборке из 60 школ с единственным учителем в индийском штате Раджастан¹⁶. Учителям в этих 60 экспериментальных школах был предложен определенный бонус за работу без прогулов. Вот креативная часть этого проекта: учителям были вручены фотокамеры с защитой от несанкционированного вмешательства и временными отметками. Это позволяло учителям каждый день доказывать свое присутствие на уроках, делая групповую фотографию с учениками¹⁷.

В результате прогулы среди учителей в экспериментальных школах сократились наполовину по сравнению с учителями в произвольно выбранной контрольной группе из 60 школ. Результаты сдачи экзаменов учениками тоже заметно улучшились: в следующий класс перешло намного больше учеников. (Полагаю, групповые фотографии также были очаровательны!)

Один из экспериментов, проведенных Дафло в Кении, предполагал выдачу произвольно выбранной группе фермеров небольшой субсидии на покупку удобрений непосредственно *после* уборки урожая. Предыдущая практика показывала, что внесение удобрений существенно повышает урожайность. Фермеры знали об этом, но каждый раз, когда наступало время сева,

у них не оставалось достаточно денег от продажи прошлого урожая, чтобы закупить удобрения. В результате фермеры попадали в так называемую ловушку бедности: они были слишком бедны, чтобы что-то сделать для повышения своего благосостояния. Дафло и ее соавторы пришли к выводу, что даже очень небольшая субсидия — бесплатная раздача удобрений, — предлагаемая фермерам, когда у них еще остаются деньги после уборки урожая, повышала использование удобрений на 10–20% по сравнению с использованием удобрений в контрольной группе¹⁸.

Эстер Дафло даже умудрилась ввязаться в «войну полов». Кто проявляет большее чувство ответственности в деле управления семейными финансами — мужчины или женщины? В богатых странах супружеские пары нередко ссорятся по этому поводу, в связи с чем им подчас даже приходится обращаться за услугами к консультантам-психологам. В бедных странах от этого может зависеть (в буквальном смысле), будут ли сыты дети. Многочисленные свидетельства, дошедшие до нас из глубин истории человеческой цивилизации, указывают на то, что женщины считают здоровье и благополучие детей одним из важнейших приоритетов, тогда как мужчины в большей степени склонны пропивать зарплату в местном пивбаре (или проводить время за каким-нибудь другими, не менее приятными занятиями, которые мужчины находили для себя даже в каменном веке). В худшем случае эти факты лишь подкрепляют давно сложившиеся стереотипы. В лучшем это не так-то легко доказать, поскольку семейные финансы в той или иной степени находятся в совместном распоряжении мужа и жены.

Но этот деликатный вопрос не смущает Эстер Дафло¹⁹. Напротив, она выявила впечатляющий натурный эксперимент. В Кот-д'Ивуаре муж и жена, как правило, несут общую ответственность за выращивание определенных сельскохозяйственных культур. В силу давно устоявшихся традиций мужчины и женщины также выращивают разные товарные культуры. (Для мужчин это обычно какао, кофе и некоторые другие сельскохозяйственные культуры.) С точки зрения исследователя такой подход к ведению сельского хозяйства интересен тем, что урожаи «мужских» и «женских» сельскохозяйственных культур в разные годы бывают разными, что определяет появление у мужчин или женщин дополнительных денег от продажи «своего» урожая. В годы, когда удается собрать хороший урожай кофе или какао, дополнительные деньги заводятся в карманах мужчин. В годы, урожайные для кокосов и бананов, дополнительный заработок появляется у женщин.

Теперь нам впору задать деликатный вопрос: в какие годы дети в таких семьях живут сытнее? В годы, когда удастся собрать хороший урожай «мужских» сельскохозяйственных культур, или когда обильный урожай собирают женщины?

К большому стыду мужчин, ответ будет следующим: когда хороший урожай собирают женщины, они тратят часть своих дополнительных доходов на бюджет семьи, чего не скажешь о мужчинах. Увы!

В 2010 году Дафло наградили медалью John Bates Clark Medal; ее вручает Американская экономическая ассоциация лучшему экономисту не старше сорока лет*. Среди экономистов эта награда считается даже более престижной, чем Нобелевская премия в области экономики, поскольку она вручается лишь раз в два года. (Правда, после получения этой медали Эстер Дафло в 2010 году награждение теперь проводится ежегодно.) Как бы то ни было, присуждение John Bates Clark Medal для людей, которые носят очки с толстыми линзами (как вам моя метафора?!), аналогично признанию «самым ценным игроком» в спорте. Дафло занимается программами оценивания. Ее работа, как и работа тех, кто сейчас использует ее методы, меняет жизнь малоимущих людей в лучшую сторону. С точки зрения статистики работа Дафло побуждает нас шире применять статистические управляемые эксперименты — которые давно считались преимущественной сферой деятельности лабораторных наук — для выявления причинно-следственных связей во многих областях нашей жизни.

Кто и что знает о вас?

Прошлым летом я нанял для своих детей новую няню. Когда она к нам пришла, я решил рассказать ей немного о нашей семье: «Я — профессор, моя жена — преподаватель...»

«Спасибо, мне уже все известно, — махнула няня рукой. — Я вас прогуглила».

С одной стороны, я испытал некоторое облегчение, поскольку мне не нужно было произносить заранее заготовленную речь. С другой — меня несколько насторожило то, как много можно обо мне узнать, пошарив часок-другой

* Я не имел права на получение этой медали за 2010 год, поскольку к тому времени мне уже было больше сорока лет. К тому же я не сделал ничего, что давало бы мне право на получение такой награды.

в сети. Наша нынешняя способность собирать и анализировать огромные объемы данных, появившаяся в результате объединения цифровой информации с дешевой вычислительной мощностью и интернетом, представляет собой поистине уникальное явление в истории человечества. Для этой новой эпохи необходим ряд новых правил.

Чтобы сегодня оценить потенциал имеющихся у нас данных, достаточно рассмотреть пример сети розничной торговли Target. Подобно большинству компаний, Target стремится повысить прибыль за счет лучшего понимания своих клиентов. Для этого она нанимает специалистов по статистике, перед которыми ставится задача выполнить своего рода «упреждающий анализ» (о нем уже упоминалось в этой книге); они используют данные продаж в сочетании с другой информацией о потребителях, чтобы выяснить, кто, что и почему покупает. Ничего изначально плохого во всем этом нет: просто это означает, что ближайший к вам магазин Target хочет знать, *что* вам как потребителю может понадобиться.

Но давайте рассмотрим хотя бы один пример того, что могут «вычислить» специалисты по статистике, обосновавшиеся в плотно зашторенных комнатах цокольного этажа корпоративной штаб-квартиры. Target выяснила, что беременность — особенно важное время с точки зрения построения моделей покупательского поведения. У беременных женщин вырабатываются определенные «отношения с розничной торговлей», которые могут длиться десятилетиями. В результате Target хочет выявить беременных женщин — особенно тех, кто на четвертом–шестом месяцах, — и заманить их в свои магазины, чтобы они посещали их как можно чаще. Один из журналистов The New York Times Magazine наблюдал за тем, как бригада специалистов по упреждающему анализу компании Target пыталась это сделать²⁰.

Первая часть задачи решалась достаточно просто. У Target есть специальный реестр, в котором беременные женщины регистрируются, чтобы еще до рождения ребенка приобрести все товары, необходимые для новорожденного. Эти женщины уже являются клиентками Target, и они, по сути, сами сообщают магазину о своей беременности. Но здесь есть одна статистическая уловка: *компания «вычислила», что другие женщины, которые демонстрируют аналогичное покупательское поведение, также, вероятно, беременны.* Например, беременные женщины зачастую переходят на использование неароматизированных лосьонов, начинают покупать витаминные добавки, огромные упаковки ватных тампонов. Специалисты по упреждающему

анализу компании Target выявили двадцать пять наименований товаров, которые в совокупности составляют «признак, позволяющий предсказать беременность». Цель этого анализа заключалась в том, чтобы послать беременным женщинам соответствующие купоны в надежде сделать их постоянными покупательницами Target.

Насколько эффективной оказалась данная стратегия? The New York Times Magazine поведал своим читателям историю о мужчине из Миннеаполиса, который зашел в один из магазинов Target, чтобы поговорить с менеджером. Мужчина был разгневан тем, что Target буквально бомбардировал его дочь-старшеклассницу купонами для беременных женщин. «Она еще учится в школе, а вы присылаете ей купоны на детские кроватки и одежду для новорожденного! Может быть, вы предлагаете ей побыстрее забеременеть?!» — возмущался мужчина.

Менеджер извинялся и заверял мужчину, что это не более чем досадная ошибка. Через несколько дней менеджер еще раз позвонил этому мужчине, чтобы извиниться. На сей раз мужчина вел себя гораздо спокойнее и в свою очередь тоже посчитал нужным извиниться. «Понимаете, у меня в доме небольшой переполох, — сказал он. — Одним словом, она должна родить в августе».

Статистики Target узнали, что его дочь беременна, раньше, чем он!

Может быть, они суют нос не в свои дела — или все-таки это *их* дела? Иногда складывается впечатление, что их внимание к нам чрезмерно. Именно поэтому некоторые компании пытаются делать вид, что знают о нас гораздо меньше, чем им известно на самом деле. Если, например, вы — женщина и находитесь на четвертом–шестом месяце беременности, то в вашем почтовом ящике могут появиться купоны на детскую кроватку и бумажные полотенца — вместе с предложением скидки на газонокосилку и купоном на бесплатные носки для боулинга (при условии покупки любой пары обуви для боулинга). Появление в вашем почтовом ящике купонов для беременных в сочетании с рекламой других товаров, не имеющих никакого отношения к беременности, может показаться вам чистой случайностью. В действительности компания знает, что вы не увлекаетесь боулингом и не стрижете лужайку перед домом. Просто она пытается «замести следы», делая вид, что знает о вас гораздо меньше, чем ей известно на самом деле.

Facebook, компания, у которой практически нет физических активов, стала одной из самых дорогих в мире. С точки зрения инвесторов (в отличие

от пользователей), Facebook располагает одним колоссальным активом — данными. Инвесторам не нравится Facebook, поскольку из-за этой компании они иногда наталкиваются в сети на своих бывших одноклассниц. Инвесторам нравится Facebook, так как каждый щелчок мышью приносит им данные о месте проживания пользователей, о магазинах, где они обычно совершают покупки, о том, какие покупки они делают, и о том, кого они знают и как проводят свободное время.

Крис Кокс, производственный вице-президент Facebook, сказал в интервью *The New York Times*: «Проблема информационной эпохи заключается в том, как распорядиться информацией»²¹.

Вот так-то.

Что же касается публичной сферы, то слияние данных и технологий порождает еще бóльшие проблемы. Во многих крупных городах мира в общественных местах установлены тысячи камер видеонаблюдения, многие из которых вскоре будут обладать способностью распознавать лица людей. Правоохранительные органы могут отслеживать маршрут движения любого автомобиля, куда бы он ни направлялся (и сохранять соответствующую информацию в архивной памяти), прикрепляя к автомобилю то или иное устройство глобального позиционирования, а затем отслеживая его перемещения с помощью спутника. Является ли это достаточно дешевым и эффективным способом слежения за действиями преступных элементов? А может быть, это не что иное, как использование государством современных технологий для ограничения нашей личной свободы? В 2012 году Верховный суд США единогласно постановил, что это действительно посягательство на нашу свободу, и запретил правоохранительным органам прикреплять устройства слежения на личные транспортные средства без соответствующего ордера*.

Между тем, многие государства обзавелись огромными базами данных ДНК, которые являются мощным инструментом, позволяющим раскрывать уголовные преступления. Чьи ДНК должны храниться в таких базах данных? Всех осужденных преступников? Каждого, кто подвергался аресту (даже если впоследствии был признан невиновным)? Или всех граждан без исключения?

Мы лишь приступаем к решению проблем, которые находятся на пересечении технологий и персональных данных. Ни одна из них не была настолько

* Судебный процесс *The United States vs. Jones*.

актуальна, когда соответствующая информация хранилась на пыльных стеллажах в подвалах серьезных государственных учреждений, а не в цифровых базах данных, в которые в принципе может забраться любой желающий. Статистика в наши дни играет даже более важную роль, чем когда-либо прежде, поскольку сейчас у нас появилось больше возможностей для эффективного использования данных. Однако сами по себе формулы не подскажут нам наилучшие способы их использования. Иными словами, математика не может заменить суждение.

Учитывая вышесказанное, давайте завершим эту книгу, попытавшись найти связь между следующими словами: огонь, ножи, автомобили, крем для удаления волос. Каждая из этих вещей служит важной цели. Каждая делает нашу жизнь лучше. И каждая может создать серьезные проблемы в случае неосторожного с ней обращения.

Теперь вы можете добавить статистику в этот список. Она наверняка поможет вам лучше понять многие явления нашей жизни при условии, что вы будете пользоваться статистическими инструментами разумно и по назначению!

Приложение

Статистическое программное обеспечение

Подозреваю, что вы не будете выполнять статистический анализ с помощью карандаша, бумаги и карманного калькулятора. Ниже приведен краткий обзор программных пакетов, наиболее широко используемых для решения задач, описанных в этой книге.

Microsoft Excel

Microsoft Excel — пожалуй, самая широко используемая программа для вычисления простых статистических показателей, таких как среднее значение и среднеквадратическое (стандартное) отклонение. Кроме того, с помощью Excel можно выполнять базовый регрессионный анализ. Большинство компьютеров комплектуется пакетом Microsoft Office, поэтому Excel, скорее всего, уже есть на жестком диске вашего ПК. В сравнении с более сложными статистическими программными пакетами Excel довольно дружелюбна к пользователю. Основные статистические вычисления можно выполнять с помощью строки формул.

Excel не умеет решать ряд более сложных задач, рассчитанных на применение более специализированных программ. Однако вы можете купить расширения Excel (впрочем, некоторые из них можно загрузить бесплатно), которые существенно повышают статистические возможности этой программы. Огромное преимущество Excel — в том, что эта программа обеспечивает простые способы отображения двумерных данных посредством весьма наглядной графики, которая легко импортируется в Microsoft PowerPoint и Microsoft Word.

Stata*

Stata — статистический пакет, используемый специалистами-исследователями во всем мире; его интерфейс отличается серьезным, научным видом. Stata

* См. <http://www.stata.com/>.

обладает широким спектром функций для решения базовых статистических задач, таких как создание таблиц данных и вычисление описательных статистик. Разумеется, университетские профессора и другие ученые отдадут предпочтение Stata не только по этой причине. Это программное обеспечение предназначено для проведения сложных статистических испытаний и моделирования данных, которые выходят далеко за рамки задач, описанных в этой книге.

Stata представляет собой идеальный инструмент для тех, кто обладает всесторонним знанием статистики (знание основ программирования также не повредит) и кому не требуется изощренное форматирование — лишь ответы на статистические вопросы. Впрочем, Stata окажется не самым идеальным инструментом, если ваша цель — оперативно строить графики на основе имеющихся данных. Опытные пользователи утверждают, что хоть Stata и умеет это делать, для этой цели удобнее пользоваться Excel.

Разработчики Stata предусмотрели несколько разных самостоятельных пакетов программного обеспечения. Вы можете купить бессрчную или годовую лицензию на этот продукт (в последнем случае через год ПО будет заблокировано на вашем компьютере). Один из самых дешевых вариантов — Stata/IC, предназначенный для «студентов и исследователей, оперирующих наборами данных среднего объема». Предусмотрена скидка для работников сферы образования. Однако даже в таком случае однопользовательская годовичная лицензия на Stata/IC обойдется вам в 295 долларов, а за бессрчную лицензию придется уплатить 595 долларов. Если же вы собираетесь запустить спутник на Марс, в связи с чем вам предстоит выполнить по-настоящему серьезные научные вычисления, то у вас есть возможность воспользоваться более «продвинутыми» пакетами Stata, стоимость которых исчисляется в тысячах долларов.

SAS*

Ввиду наличия у SAS широкого спектра аналитических способностей, этот статистический пакет привлекателен не только для профессиональных исследователей, но и для бизнес-аналитиков и инженеров. У SAS есть два разных статистических пакета. Первый называется SAS Analytics Pro и может считывать данные практически в любом формате, а также выполнять их сложный анализ. В этом пакете также предусмотрены хорошие инструменты визуализации

* См. <http://www.sas.com/technologies/analytics/statistics/>.

данных; в частности он обладает расширенными возможностями отображения. Пакет не из дешевых. Даже работникам сферы образования и государственных учреждений покупка одной коммерческой или индивидуальной лицензии на него обойдется в 8500 долларов — плюс плата за годовую лицензию.

Второй статистический пакет SAS называется SAS Visual Data Discovery. Он снабжен удобным интерфейсом, который не требует специальных знаний кодирования или программирования и обладает весьма широкими функциями анализа данных. Как следует из названия пакета, он позволяет легко исследовать данные с помощью интерактивной визуализации. Вы можете также экспортировать анимации своих данных в презентации, веб-страницы и прочие документы. Этот пакет также недешев: покупка одной коммерческой или индивидуальной лицензии на него обойдется в 9810 долларов — плюс плата за годовую лицензию.

SAS предлагает ряд специализированных инструментов управления. К их числу относится, например, продукт, использующий статистику для выявления фактов мошенничества и финансовых преступлений.

R

Возможно, это название напоминает вам звуки, издаваемые каким-нибудь персонажем из фильмов о Джеймсе Бонде. На самом же деле R — это бесплатный (с открытым исходным кодом) популярный статистический пакет. Его можно загрузить и установить на компьютер в течение буквально нескольких минут. Кроме того, уже сформировалось так называемое R-сообщество, члены которого готовы поделиться с вами соответствующим кодом и предоставить всю необходимую помощь.

Прелесть пакета R не только в его бесплатности, но и в необычайной пластичности и гибкости. В зависимости от поставленных вами целей эта гибкость может либо разочаровать вас, либо обеспечить немалыми дополнительными возможностями. Если вы только приступаете к использованию статистического программного обеспечения, то эта программа не предоставит вам практически никакой структуры. Мало в чем поможет вам и ее интерфейс. С другой стороны, для программистов (и даже тех, кто знаком лишь с основами программирования) такое отсутствие структуры может оказаться благом, раскрепощающим их творчество. Пользователи вольны попросить программу сделать практически все, что они пожелают; в том числе заставить ее взаимодействовать со сторонними программами.

IBM SPSS*

IBM SPSS есть что предложить как «зубрам» статистики, так и бизнес-аналитикам, менее сведущим в вопросах статистики. Пакет IBM SPSS также хорош для начинающих, поскольку в нем предусмотрен интерфейс, управляемый системой меню. Кроме того, в IBM SPSS имеется ряд инструментов, или «модулей», предназначенных для выполнения специализированных функций, например IBM SPSS Forecasting (прогнозирование), IBM SPSS Advanced Statistics (расширенная статистика), IBM SPSS Visualization Designer (дизайнер визуализации) и IBM SPSS Regression (регрессионный анализ). Эти модули продаются по отдельности или пакетами.

Самым базовым из предлагаемых вариантов IBM SPSS является IBM SPSS Statistics Standard Edition (стандартная версия статистики), который позволяет рассчитывать простые статистические закономерности и выполнять базовый анализ данных, такой как выявление тенденций и построение прогнозных моделей. Одну коммерческую лицензию, рассчитанную на фиксированный срок, можно приобрести за 2250 долларов. Премиум-пакет, который включает в себя большинство упомянутых выше модулей, обойдется в 6750 долларов. Скидки предоставляются работникам сферы образования.

(От научного редактора. Здесь уместно привести хотя бы краткие характеристики статистических пакетов, которые распространены на российском рынке. Кроме перечисленных автором, отметим еще один зарубежный статистический пакет, получивший широкое распространение в России. Это универсальный статистический пакет STATISTICA, который может служить не только эффективным инструментом для научных исследований, но и чрезвычайно удобной средой для обучения методам статистического анализа. Из российских разработок отметим пакеты STADIA, «ЭВРИСТА», «МЕЗОЗАВР», «САНИ», «СТАТЭксперт» и др. Советуем обратить внимание на удивительно компактный пакет STADIA. Кроме набора современных и эффективных методов статистического анализа, этот пакет имеет полный комплект научной, деловой и многомерной графики, а также понятную систему интерпретации результатов анализа.)

* См. <http://www-01.ibm.com/software/analytics/spss/products/statistics/>.

Примечания

Глава 1

1. Central Intelligence Agency, The World Factbook, <https://www.cia.gov/library/publications/the-world-factbook/>.
2. Steve Lohr, For Today's Graduate, Just One Word: Statistics, New York Times, August 6, 2009.
3. Там же.
4. Baseball-Reference.com, <http://www.baseball-reference.com/players/>
5. Trip Gabriel, Cheats Find an Adversary in Technology, New York Times, December 28, 2010.
6. Eyder Peralta, Atlanta Man Wins Lottery for Second Time in Three Years, NPR News (блог), November 29, 2011.
7. Alan B. Krueger, What Makes a Terrorist: Economics and the Roots of Terrorism (Princeton: Princeton University Press, 2008).

Глава 2

1. U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements, <http://www.census.gov/en.html>.
2. Malcolm Gladwell, The Order of Things, The New Yorker, February 14, 2011.
3. CIA, World Factbook, и United Nations Development Program, 2011 Human Development Report, <http://hdr.undp.org/en/statistics/>.
4. Baseball-Reference.com.

Глава 3

1. Robert Griffith, The Politics of Fear: Joseph R. McCarthy and the Senate, 2nd ed. (Amherst: University of Massachusetts Press, 1987), p. 49.

2. Catching Up, Economist, August 23, 2003.
3. Carl Bialik, When the Median Doesn't Mean What It Seems, Wall Street Journal, May 21–22, 2011.
4. Stephen Jay Gould, The Median Isn't the Message, с предисловием и заключением Стива Данна (Steve Dunn), http://cancerguide.org/median_not_msg.html.
5. См. <http://www.movieweb.com>.
6. Box Office Mojo (boxofficemojo.com), June 29, 2011.
7. Steve Patterson, 527% Tax Hike May Shock Some, But It's Only About \$5, Chicago Sun-Times, December 5, 2005.
8. Rebecca Leung, The 'Texas Miracle': 60 Minutes II Investigates Claims That Houston Schools Falsified Dropout Rates, CBSNews.com, August 25, 2004.
9. Marc Santora, Cardiologists Say Rankings Sway Surgical Decisions, New York Times, January 11, 2005.
10. Интервью на National Public Radio, August 20, 2006, <http://www.npr.org/templates/story/story.php?storyId=5678463>.
11. См. <http://www.usnews.com/education/articles/2010/08/17/frequently-asked-questions-college-rankings#4>.
12. Gladwell, Order of Things.
13. Интервью на National Public Radio, February 22, 2007, <http://www.npr.org/templates/story/story.php?storyId=7383744>.

Глава 4

1. College Board, FAQs, http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2010/correlations-of-predictors-with-first-year-college-grade-point-average.pdf.
2. College Board, 2011 College-Bound Seniors Total Group Profile Report, <http://research.collegeboard.org/programs/sat/data/archived/cb-seniors-2011>.
3. См. <http://www.netflixprize.com/rules>.

Глава 5

1. David A. Aaker, Managing Brand Equity: Capitalizing on the Value of a Brand Name (New York: Free Press, 1991).

2. Victor J. Tremblay and Carol Horton Tremblay, *The U.S. Brewing Industry: Data and Economic Analysis* (Cambridge: MIT Press, 2005).
3. Australian Transport Safety Bureau Discussion Paper, *Cross Modal Safety Comparisons*, January 1, 2005.
4. Marcia Dunn, 1 in 21 Trillion Chance Satellite Will Hit You, *Chicago Sun-Times*, September 21, 2011.
5. Steven D. Levitt and Stephen J. Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything* (New York: William Morrow Paperbacks, 2009).
6. Garrick Blalock, Vrinda Kadiyali, and Daniel Simon, *Driving Fatalities after 9/11: A Hidden Cost of Terrorism* (неопубликованная рукопись, December 5, 2005).
7. Источником общей информации о генетическом тестировании является Human Genome Project Information, DNA Forensics, http://www.ornl.gov/sci/techresources/Human_Genome/elsi/forensics.shtml.
8. Jason Felch and Maura Dolan, FBI Resists Scrutiny of 'Matches', *Los Angeles Times*, July 20, 2008.
9. David Leonhardt, In Football, 6 + 2 Often Equals 6, *New York Times*, January 16, 2000.
10. Roger Lowenstein, The War on Insider Trading: Market Beaters Beware, *New York Times Magazine*, September 22, 2011.
11. Erica Goode, Sending the Police before There's a Crime, *New York Times*, August 15, 2011.
12. Источниками данных о страховании рисков являются: Teen Drivers, Insurance Information Institute, March 2012; Texting Laws and Collision Claim Frequencies, Insurance Institute for Highway Safety, September 2010; Hot Wheels, National Insurance Crime Bureau, August 2, 2011.
13. Charles Duhigg, What Does Your Credit Card Company Know about You? *New York Times Magazine*, May 12, 2009.

Глава 5½

1. John Tierney, And behind Door No. 1, a Fatal Flaw, *New York Times*, April 8, 2008.
2. Leonard Mlodinow, *The Drunkard's Walk: How Randomness Rules Our Lives* (New York: Vintage Books, 2009).

Глава 6

1. Joe Nocera, Risk Mismanagement, New York Times Magazine, January 2, 2009.
2. Robert E. Hall, The Long Slump, American Economic Review 101, no. 2 (April 2011): 431–69.
3. Alan Greenspan, Testimony before the House Committee on Government Oversight and Reform, October 23, 2008.
4. Hank Paulson, Speech at Dartmouth College, Hanover, NH, August 11, 2011.
5. The Probability of Injustice, Economist, January 22, 2004.
6. Thomas Gilovich, Robert Vallone, and Amos Tversky, The Hot Hand in Basketball: On the Misperception of Random Sequences, Cognitive Psychology 17, no. 3 (1985): 295–314.
7. Ulrike Malmendier and Geoffrey Tate, Superstar CEOs, Quarterly Journal of Economics 124, no. 4 (November 2009): 1593–638.
8. The Price of Equality, Economist, November 15, 2003.

Глава 7

1. Benedict Carey, Learning from the Spurned and Tipsy Fruit Fly, New York Times, March 15, 2012.
2. Cynthia Crossen, Fiasco in 1936 Survey Brought ‘Science’ to Election Polling, Wall Street Journal, October 2, 2006.
3. Tara Parker-Pope, Chances of Sexual Recovery Vary Widely after Prostate Cancer, New York Times, September 21, 2011.
4. Benedict Carey, Researchers Find Bias in Drug Trial Reporting, New York Times, January 17, 2008.
5. Siddhartha Mukherjee, Do Cellphones Cause Brain Cancer? New York Times, April 17, 2011.
6. Gary Taubes, Do We Really Know What Makes Us Healthy? New York Times, September 16, 2007.

Глава 8

1. U.S. Census Bureau.

Глава 9

1. John Friedman, *Out of the Blue: A History of Lightning: Science, Superstition, and Amazing Stories of Survival* (New York: Delacorte Press, 2008).
2. Low Marks All Round, *Economist*, July 14, 2011.
3. Trip Gabriel and Matt Richtel, Inflating the Software Report Card, *New York Times*, October 9, 2011.
4. Jennifer Corbett Dooren, Link in Autism, Brain Size, *Wall Street Journal*, May 3, 2011.
5. Heather Cody Hazlett et al., Early Brain Overgrowth in Autism Associated with an Increase in Cortical Surface Area before Age 2 Years, *Archives of General Psychiatry* 68, no. 5 (May 2011): 467–76.
6. Benedict Carey, Top Journal Plans to Publish a Paper on ESP, and Psychologists Sense Outrage, *New York Times*, January 6, 2011.

Глава 10

1. Jeff Zeleny and Megan Thee-Brenan, New Poll Finds a Deep Distrust of Government, *New York Times*, October 26, 2011.
2. Lydia Saad, Americans Hold Firm to Support for Death Penalty, *Gallup.com*, November 17, 2008.
3. Телефонное интервью с Фрэнком Ньюпором, November 30, 2011.
4. Stanley Presser, Sex, Samples, and Response Errors, *Contemporary Sociology* 24, no. 4 (July 1995): 296–98.
5. Эти результаты были опубликованы в двух разных формах, одна из которых более академическая, чем другая. Edward O. Lauman, *The Social Organization of Sexuality: Sexual Practices in the United States* (Chicago: University of Chicago Press, 1994); Robert T. Michael, John H. Gagnon, Edward O. Laumann, and Gina Kolata, *Sex in America: A Definitive Survey* (New York: Grand Central Publishing, 1995).
6. Kaye Wellings, book review in *British Medical Journal* 310, no. 6978 (February 25, 1995): 540.
7. John DeLamater, The NORC Sex Survey, *Science* 270, no. 5235 (October 20, 1995): 501.
8. Presser, Sex, Samples, and Response Errors.

Глава 11

1. Marianne Bertrand, Claudia Goldin, and Lawrence F. Katz, Dynamics of the Gender Gap for Young Professionals in the Corporate and Financial Sectors, NBER Working Paper 14681, January 2009.
2. M. G. Marmot, Geoffrey Rose, M. Shipley, and P. J. S. Hamilton, Employment Grade and Coronary Heart Disease in British Civil Servants, *Journal of Epidemiology and Community Health* 32, no. 4 (1978): 244–49.
3. Hans Bosma, Michael G. Marmot, Harry Hemingway, Amanda C. Nicholson, Eric Brunner, and Stephen A. Stansfeld, Low Job Control and Risk of Coronary Heart Disease in Whitehall II (Prospective Cohort) Study, *British Medical Journal* 314, no. 7080 (February 22, 1997): 558–65.
4. Peter L. Schnall, Paul A. Landesbergis, and Dean Baker, Job Strain and Cardiovascular Disease, *Annual Review of Public Health* 15 (1994): 381–411.
5. M. G. Marmot, H. Bosma, H. Hemingway, E. Brunner, and S. Stansfeld, Contribution of Job Control and Other Risk Factors to Social Variations in Coronary Heart Disease Incidence, *Lancet* 350 (July 26, 1997): 235–39.

Глава 12

1. Gary Taubes, Do We Really Know What Makes Us Healthy? *New York Times Magazine*, September 16, 2007.
2. Vive la Difference, *Economist*, October 20, 2001.
3. Taubes, Do We Really Know?
4. College Board, 2011 College-Bound Seniors Total Group Profile Report, <http://research.collegeboard.org/programs/sat/data/archived/cb-seniors-2011>.
5. Hans Bosma et al., Low Job Control and Risk of Coronary Heart Disease in Whitehall II (Prospective Cohort) Study, *British Medical Journal* 314, no. 7080 (February 22, 1997): 564.
6. Taubes, Do We Really Know?
7. Gautam Naik, Scientists' Elusive Goal: Reproducing Study Results, *Wall Street Journal*, December 2, 2011.
8. John P. A. Ioannidis, Contradicted and Initially Stronger Effects in Highly Cited Clinical Research, *Journal of the American Medical Association* 294, no. 2 (July 13, 2005): 218–28.
9. Scientific Accuracy and Statistics, *Economist*, September 1, 2005.

Глава 13

1. Gina Kolata, Arthritis Surgery in Ailing Knees Is Cited as Sham, New York Times, July 11, 2002.
2. Benedict Carey, Long-Awaited Medical Study Questions the Power of Prayer, New York Times, March 31, 2006.
3. Diane Whitmore Schanzenbach, What Have Researchers Learned from Project STAR? Harris School Working Paper, August 2006.
4. Gina Kolata, A Surprising Secret to a Long Life: Stay in School, New York Times, January 3, 2007.
5. Adriana Lleras-Muney, The Relationship between Education and Adult Mortality in the United States, Review of Economic Studies 72, no. 1 (2005): 189–221.
6. Kurt Badenhausen, Top Colleges for Getting Rich, Forbes.com, July 30, 2008.
7. Stacy Berg Dale and Alan Krueger, Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables, Quarterly Journal of Economics 117, no. 4 (November 2002): 1491–527.
8. Alan B. Krueger, Children Smart Enough to Get into Elite Schools May Not Need to Bother, New York Times, April 27, 2000.
9. Randi Hjalmarsson, Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality? Journal of Law and Economics 52, no. 4 (November 2009): 779–809.

Заключение

1. James Surowiecki, A Billion Prices Now, The New Yorker, May 30, 2011.
2. Malcolm Gladwell, Offensive Play, The New Yorker, October 19, 2009.
3. Ken Belson, N.F.L. Roundup; Concussion Suits Joined, New York Times, February 1, 2012.
4. Shirley S. Wang, Autism Diagnoses Up Sharply in U.S., Wall Street Journal, March 30, 2012.
5. Catherine Rice, Prevalence of Autism Spectrum Disorders, Autism and Developmental Disabilities Monitoring Network, Centers for Disease Control and Prevention, 2006, <http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5810a1.htm>.
6. Alan Zarembo, Autism Boom: An Epidemic of Disease or of Discovery? latimes.com, December 11, 2011.

7. Michael Ganz, The Lifetime Distribution of the Incremental Societal Costs of Autism, *Archives of Pediatrics & Adolescent Medicine* 161, no. 4 (April 2007): 343–49.
8. Gardiner Harris and Anahad O'Connor, On Autism's Cause, It's Parents vs. Research, *New York Times*, June 25, 2005.
9. Julie Steenhuysen, Study Turns Up 10 Autism Clusters in California, *Yahoo! News*, January 5, 2012.
10. Joachim Hallmayer et al., Genetic Heritability and Shared Environmental Factors among Twin Pairs with Autism, *Archives of General Psychiatry* 68, no. 11 (November 2011): 1095–102.
11. Gardiner Harris and Anahad O'Connor, On Autism's Cause, It's Parents vs. Research, *New York Times*, June 25, 2005.
12. Fernanda Santos and Robert Gebeloff, Teacher Quality Widely Diffused, Ratings Indicate, *New York Times*, February 24, 2012.
13. Winnie Hu, With Teacher Ratings Set to Be Released, Union Opens Campaign to Discredit Them, *New York Times*, February 23, 2012.
14. T. Schall and G. Smith, Do Baseball Players Regress to the Mean? *American Statistician* 54 (2000): 231–35.
15. Scott E. Carrell and James E. West, Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors, *National Bureau of Economic Research Working Paper* 14081, June 2008.
16. Esther Duflo and Rema Hanna, Monitoring Works: Getting Teachers to Come to School, *National Bureau of Economic Research Working Paper* 11880, December 2005.
17. Christopher Udry, Esther Duflo: 2010 John Bates Clark Medalist, *Journal of Economic Perspectives* 25, no. 3 (Summer 2011): 197–216.
18. Esther Duflo, Michael Kremer, and Jonathan Robinson, Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya, *National Bureau of Economic Research Working Paper* 15131, July 2009.
19. Esther Duflo and Christopher Udry, Intrahousehold Resource Allocation in Côte d'Ivoire: Social Norms, Separate Accounts and Consumption Choices, *Working Paper*, December 21, 2004.
20. Charles Duhigg, How Companies Learn Your Secrets, *New York Times Magazine*, February 16, 2012.
21. Somini Sengupta and Evelyn M. Rusli, Personal Data's Value? Facebook Set to Find Out, *New York Times*, February 1, 2012.

От автора

Написанием этой книги я хотел отдать дань уважения классическому труду Дарелла Хаффа *How to Lie with Statistics* («Как лгать при помощи статистики»), выпущенному тиражом свыше миллиона экземпляров издательством W. W. Norton в 1950-е годы и, что примечательно, раскупленному полностью. Дарелл Хафф, как и я, пытался убедить рядового читателя, что непонимание им истинной сути чисел, время от времени появляющихся в заголовках новостей, *может* нанести ему немалый вред. Я надеюсь, что стал достойным продолжателем дела Дарелла Хаффа. Как бы то ни было, я пришел бы в восторг, если бы проданся миллион экземпляров моей книги!

Я искренне признателен издательству W. W. Norton, и в частности Дрейку Макфили, за предоставленную возможность публиковать книги, освещающие весьма актуальные темы на доступном для понимания широким кругом читателей уровне. Более десяти лет Дрейк Макфили показывает себя как настоящий друг и помощник.

Заслуги Джеффа Шриви из W. W. Norton в публикации этой книги трудно переоценить. Впервые встретив Джеффа, вы наверняка подумали бы, что он, пожалуй, слишком мягкий человек для того, чтобы жестко контролировать соблюдение сроков подготовки книги к печати. И, увы, оказались бы неправы. Джефф действительно очень милый человек, но как-то так получилось, что именно его интеллигентная манера ведения дел как нельзя лучше способствовала выполнению всех запланированных работ. (Например, этот раздел должен быть готов к завтрашнему утру.) Мне очень повезло, что рядом со мной был такой «надсмотрщик», как Джефф Шриви.

Я поистине в неоплатном долгу перед теми, кто выполнял важную исследовательскую и аналитическую работу, описанную мною в книге. Я ведь не специалист по статистике и не исследователь, а всего лишь литературный интерпретатор интересной и нужной работы, проделанной другими людьми.

Я надеюсь, что с помощью своей книги смогу убедить читателей в важности глубоких исследований и всестороннего анализа, делающих каждого из нас здоровее, богаче, защищеннее и информированнее.

В частности, мне хотелось бы отметить масштабную работу экономиста Принстонского университета Алана Крюгера, который внес огромный вклад в решение широкого круга проблем, начиная с выявления первопричин терроризма и заканчивая экономической отдачей от высшего образования (сделанные им выводы по обоим проблемам весьма неожиданны и порой парадоксальны). Еще более значимым (для меня) оказалось то, что во время моей учебы в магистратуре профессор Алан Крюгер был одним из моих преподавателей статистики: на меня всегда производила большое впечатление его способность успешно сочетать исследовательскую, преподавательскую и общественную деятельность.

Джим Сэлли, Джефф Гроггер, Патти Андерсон и Артур Майнетц ознакомились с черновыми набросками этого манускрипта и внесли многочисленные — и весьма ценные — предложения по улучшению текста. Друзья, вы спасли меня от самого себя! Фрэнк Ньюпор из Института Гэллага и Майк Кэйги из *The New York Times* были столь любезны, что ознакомили меня с методологическими тонкостями проведения опросов общественного мнения. Несмотря на все их усилия, ошибки, которые вы, возможно, найдете в этой книге, остаются целиком на моей совести.

Кейти Уэйд была моим неутомимым помощником в исследованиях (мне всегда хотелось вставить в свой текст слово «неутомимый», и вот наконец мне подвернулся идеальный, на мой взгляд, контекст), став неисчерпаемым источником историй и примеров из жизни, которые прекрасно иллюстрировали концепции, представленные в книге. Без Кейти вы многого бы не узнали.

Еще со школьных лет я мечтал писать книги. Человеком, подарившим мне такую возможность (а также способ зарабатывать этим на жизнь), является мой агент Тина Беннетт. Тина воплощает в себе все лучшее, что присуще издательскому бизнесу. Она испытывает истинное удовольствие, когда важная, по ее мнению, работа приносит свои плоды; при этом Тина неустанно отстаивает интересы своих клиентов.

Наконец, самых высоких похвал заслуживают члены моей семьи за ангельское терпение, проявленное во время подготовки книги к публикации. (Листок бумаги со сроками сдачи в печать каждой главы прикреплялся магнитом к холодильнику.) Есть неопровержимые свидетельства того, что я становлюсь

в этот период на 31% раздражительнее и на 23% истощеннее. Моя жена Лия исполняет роль первого, самого строгого и важного редактора всех моих текстов. Спасибо тебе, Лия, за это, а также за то, что всегда остаешься умным и интересным партнером, поддерживающим меня в любых начинаниях.

Эта книга посвящается моей старшей дочери Кэтрин. Мне с трудом верится, что когда я готовил к публикации книгу *Naked Economics*, Кэтрин была грудным младенцем, а сейчас уже читает главы моей новой книги и даже делает время от времени ценные замечания. Кэтрин, ты — чудо-ребенок, как, впрочем, и мои младшие, Софи и Си-Джей, которые вскоре тоже будут читать мои книги и давать полезные советы.

Предметно-именной указатель

Т

t-распределение 254, 268

t-статистика 270

А

Активирование 291

Анализ на основе разрывности 303

Б

Бертран Марианна 262

Блалок Гаррик 106

В

Вариан Хол 20

Величины

номинальные 72

реальные 73

Вероятность

выигрыша в лотерею 113

выигрыша у Монти Холла 127

выпадения орла 104

изменения стоимости акций 133

наступления нескольких событий 109

наступления одного из нескольких
событий 110

совпадения ДНК 107

Выборка

размер 180

репрезентативная 153

формирование 154

Выборочное исследование 23

Г

Гипотеза

альтернативная 195

нулевая 195

Гистограмма 40

Голдин Клаудиа 262

Гладуэлл Малкольм 53

Д

Дабнер Стивен 106

Данные 152

выборка 153

для проверки гипотез 195

контрольная и подопытная
группы 156

использование 23

определение 21

«поперечного среза» 159

Дециль 43

Диамантопуло Анна 148

Дисперсия

вычисление 57

Доверительный интервал 208

Е

Единица анализа 66, 67, 68

З

Загадка Монти Холла 127

Закон

арксинуса 142

Мидоу 140

И

Игра в казино 24

Индекс человеческого развития 53

Инфляция 74

Иоаннидис Джон 286

Исследования

повторные 157

поперечный срез 159

секса американцев 236

К

Кац Лоуренс 262

Квартиль 42

Корреляция 89

коэффициент 91

определение 90

Коэффициент
Джини 18, 19
детерминации R^2 255
корреляции 91
корреляции, вычисление 98
регрессии 250
Крюгер Алан 30

Л

Левитт Стивен 106
Ллерас-Муни Адриана 297

М

Макки Энн 311
Мартин Дж. П. 125
Математическое ожидание 111
вычисление 111
применение 118
Медиана 40, 41, 70
определение 69
семейного дохода 179
Метод наименьших квадратов 247
Мультиколлинеарность 281
Модель стоимости риска VaR 133

Н

Натурный эксперимент 296
Неэквивалентный контроль 298
Нормальное распределение 47, 177
средних выборок 189
Ньюпор Фрэнк 230

О

Описательная статистика 35
абсолютные показатели 43
достоверность 62
индексы 52
медиана 40
процентные шкалы 49
среднее значение 38
среднеквадратическое отклонение 44
точность 62
центральная тенденция 38
Опрос общественного мнения 23, 221
доверительный интервал 223

методологические вопросы 231
определение 222
репрезентативная выборка 222
создание репрезентативной
выборки 231
экзитпол 225
эффективность 222
Оценивание риска 134
Оценочные таблицы 83

П

Переменная
двоичная 257
зависимая 249
независимая 249
объясняющая 249
управляющая 249
фиктивная 257
Показатель «добавленной
стоимости» 79
Поправка на инфляцию 36
Проверка гипотез 195
двухсторонний критерий 217
доверительный интервал 208
односторонний критерий 217
ошибка второго рода 213
ошибка первого рода 212, 213
уровень значимости 199
Процентиль 43

Р

Разер Дэн 81
Рамсфелд Дональд 32
Рандомизация 156, 293
Регрессионный анализ 241
адекватность уравнения 287
коэффициент R^2 255
коэффициент регрессии 250
метод наименьших квадратов 247
многофакторный 256
множественный 256
мультиколлинеарность 281
интеллектуальный анализ 283
ограничение 30

- определение 28, 242
- типичные ошибки 273
- уравнение 248
- Рейтинг
 - USNWR 84
 - национальных университетов 84
 - распасовщика 17, 18
 - спортивных автомобилей 53
 - школ 80
- Распределение средних выборок 200
- Репрезентативная выборка
 - значение 154, 179
 - опросы общественного мнения 222
 - создание 153
- Риски
 - банковские 63
 - в бизнесе 24
 - кредитные 124
 - страхования 25, 116, 124
 - финансовые 133
- С**
- Синдром внезапной смерти
 - младенцев 140
- Систематическая ошибка
 - выбора 160
 - доживаемости до определенного возраста 166
 - здорового человека 169
 - памяти 165
 - публикации 163
 - самоотбора 162
- Среднее
 - выборки 179
 - семейного дохода 179
 - совокупности 179
- Среднее значение
 - определение 69
- Среднеквадратическое отклонение 181
 - вычисление 57
 - определение 44
 - совокупности 182
- Стандартная ошибка 182
 - выборки 182
 - вычисление 184
 - для разности средних 216
 - коэффициента регрессии 253
 - опроса общественного мнения 224, 230, 238
 - экзитпола 229
- Статистическая дискриминация 148
- Статистические выводы 191
 - недостатки 211
 - определение 194
 - опросы общественного мнения 222
 - проверка гипотез 195
- Статистика
 - задачи 33
 - описательная 21, 35
- Статистическое оценивание 289
 - активирование 291
 - анализ на основе разрывности 303
 - контрольные и подопытные группы 292
 - натурный эксперимент 296
 - неэквивалентный контроль 298
 - определение 291
 - программы 291
 - разница в различиях 301
 - рандомизация 293
 - управляемый эксперимент 292
- Статистическое программное
 - обеспечение 327
 - IBM SPSS 330
 - Microsoft Excel 327
 - R 329
 - SAS 328
 - STADIA 330
 - Stata 327
 - STATISTICA 330
- Статистический управляемый эксперимент 292
- Степень вероятности 106
- Страхование 25, 116, 124

Т

Тaleb Николас 137

Теория вероятностей

основы 101

ошибки применения 139

применение 138

Тест

АСТ 84

SAT 84, 93, 95

У

Уравнение регрессии 248

Уровень значимости 199

Центральная предельная теорема

возможность применения 194

применение 228

Ф

Финансовый кризис 2008 г. 133

Х

Хафф Даррелл 32

Ц

Центральная предельная теорема 171

нормальное распределение 177, 185

основной принцип 172

размер выборки 181

Э

Экзитпол 225

Экспертиза данных 25

Научно-популярное издание

Чарльз Уилан

Голая статистика

Самая интересная книга о самой скучной науке

Главный редактор *Артем Степанов*

Ответственный редактор *Наталья Шульпина*

Арт-директор *Алексей Богомолов*

Литературный редактор *Татьяна Сквородникова*

Дизайн переплета *Ксения Белоброва*

Верстка *Лариса Чернокозинская*

Корректоры *Надежда Болотина, Екатерина Лебедева*