

Чарльз Уилан

Голая статистика

Самая
интересная
книга
о самой
скучной
наук



[Почитать описание, рецензии и купить на сайте МИФа](#)

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>Волшебный 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

Введение

Почему я ненавижу 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 обще-школьные собрания таким образом, чтобы время их проведения совпадало с уроками по матанализу, 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 продемонстрирован 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 шансов 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 заболеть 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 десятиминуток? 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 уберезет 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 удается? 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.



[Почитать описание, рецензии
и купить на сайте](#)

Лучшие цитаты из книг, бесплатные главы и новинки:

