



TIM HARFORD

HOW TO MAKE THE WORLD ADD UP

ТИМ ХАРФОРД

ЛОЖЬ, НАГЛАЯ ЛОЖЬ И СТАТИСТИКА

ПРИЕМЫ, КОТОРЫЕ ПОМОГУТ ВИДЕТЬ
ПРАВДУ ЗА ЦИФРАМИ

УДК 311
ББК 60.6
Х22

HOW TO MAKE THE WORLD ADD UP by

Tim Harford

© Tim Harford, 2020

В оформлении обложки использована иллюстрация:

hauvi / Shutterstock / FOTODOM

Используется по лицензии от Shutterstock / FOTODOM

Харфорд, Тим.

Х22 Ложь, наглая ложь и статистика : приемы, которые помогут видеть правду за цифрами / Тим Харфорд ; [перевод с английского Е. А. Левиной]. — Москва : Эксмо, 2024. — 336 с. — (Книги-Драйверы).

ISBN 978-5-04-172278-4

С помощью статистики можно доказать, что детей приносят аисты. Или убедить всех, что Земля плоская! И только 10 простых правил Тима Харфорда помогут не попасться на удочку искажений истины. Автор уверен, что восприятие информации сильно зависит от наших эмоций, предубеждений и политических взглядов. Не включая критическое мышление, мы бездумно репостим сообщения о новых невероятных открытиях «британских ученых».

Эта книга, словно рентгеновский аппарат, поможет разглядеть, как устроен удивительный мир статистики.

Вы узнаете про:

- эффект страуса и эффект публикации;
- иллюзию глубины понимания;
- принципы здравого смысла.

И научитесь:

- противостоять когнитивным ловушкам;
- разбираться в данных;
- принимать верные решения.

УДК 311
ББК 60.6

ISBN 978-5-04-172278-4

© Левина Е.А., перевод на русский язык, 2024
© Оформление. ООО «Издательство «Эксмо», 2024

ОГЛАВЛЕНИЕ

Вступление КАК ЛГАТЬ ПРИ ПОМОЩИ СТАТИСТИКИ	7
Правило первое. ПРИСЛУШАЙТЕСЬ К ГОЛОСУ СЕРДЦА	27
Правило второе. УЧИТЫВАЙТЕ СВОЙ ЛИЧНЫЙ ОПЫТ	56
Правило третье. НЕ СПЕШИТЕ С ПОДСЧЕТАМИ	75
Правило четвертое. ЧТОБЫ УВИДЕТЬ ВСЮ КАРТИНУ, ОТСТУПИТЕ НА ШАГ НАЗАД	97
Правило пятое. УЗНАЙТЕ ПРЕДЫСТОРИЮ	115
Правило шестое. СПРОСИТЕ, КОГО НЕ ХВАТАЕТ	148
Правило седьмое. ТРЕБУЙТЕ ПРОЗРАЧНОСТИ ОТ КОМПЬЮТЕРА ...	167
Правило восьмое. ЦЕНИТЕ КРАЕУГОЛЬНЫЙ КАМЕНЬ СТАТИСТИКИ	201
Правило девятое. ПОМНИТЕ, ЧТО ДЕЗИНФОРМАЦИЯ ТОЖЕ БЫВАЕТ ПРИВЛЕКАТЕЛЬНОЙ	231
Правило десятое. НЕ БОЙТЕСЬ ИЗМЕНИТЬ СВОЕ МНЕНИЕ	258
Золотое правило. СОХРАНЯЙТЕ ЛЮБОЗНАТЕЛЬНОСТЬ	284
Примечания.	300
Благодарности.	327
Об авторе.	329
Алфавитный указатель	330

Вступление

КАК ЛГАТЬ ПРИ ПОМОЩИ СТАТИСТИКИ

Настоящая же проблема... заключается не в том, чтобы доказать, что ложь — это ложь, а в том, чтобы доказать истинность истинного объекта.

Умберто Эко¹

Знаете старую байку про то, что детей приносят аисты? Это чистая правда.

Сейчас я докажу это при помощи статистики.

Возьмите число аистов в каждой стране и сравните его с числом рождений за год. По всей Европе эти два числа на удивление четко коррелируют. Больше аистов — больше младенцев; меньше аистов — меньше младенцев. Закономерность достаточно заметная, чтобы пройти отсев в научном журнале. Есть даже опубликованная научная работа под названием «Аисты приносят детей ($p = 0.008$)». Не будем вдаваться в детали, скажу лишь, что все эти нули говорят о том, что это не простое совпадение².

Вы, наверное, уже догадались, в чем тут подвох. В крупных европейских странах вроде Германии, Польши и Турции много младенцев и много аистов. В небольших странах, таких как Албания и Дания, и младенцев, и аистов меньше. Между числом тех и других есть взаимосвязь, но это не означает, что младенцы появляются благодаря аистам.

Создается впечатление, что при помощи статистики можно «доказать» все что угодно — даже то, что детей приносят аисты.

Если вы читали «Как лгать при помощи статистики», вы наверняка пришли к такому выводу. Эту остроумную и циничную книжку выпустил в 1954 малоизвестный тогда американский журналист-фрилансер Дарелл Хафф. The New York Times немедленно написала на нее восторженный отзыв, и с тех пор было продано более миллиона экземпляров этой книги, что, возможно, сделало ее самой популярной книгой по статистике на свете.

Книга заслуживает и славы, и похвалы. Это — настоящее чудо популярной статистики. Среди любителей чисел Дарелл стал легендой. Бен Голдакр, эпидемиолог и автор бестселлера «Обман в науке», восхищается тем, какую «бомбу» написал «брюзга Хафф». Американский писатель Чарльз Уилан говорит, что его книга «Голая статистика» вдохновлена «классической» работой Хаффа. Через 50 лет со дня выхода книги Хаффа известный журнал «Статистическая наука» организовал в его честь ретроспективу.

Я и сам раньше так думал. Подростком я зачитывался книгой «Как лгать при помощи статистики». Эта блестящая, хлесткая, да еще и забавно проиллюстрированная книга позволила мне проникнуть за завесу статистических манипуляций и научила меня распознавать обман.

Хафф использует море примеров. Для начала он задается вопросом, сколько зарабатывают выпускники престижного Йельского университета. Если верить опросу, проведенному в 1950 году, выпускники 1924 года зарабатывали в среднем что-то вроде 500 000 долларов в год в пересчете на сегодняшние деньги. Ну что ж, можно в это поверить — в конце концов, речь идет о Йеле, — но полмиллиона долларов в год — это целое состояние... неужели и правда это средний доход?

Вовсе нет. Хафф объясняет, что источник этой «невероятно привлекательной» информации — ответы самих выпускников, а как же не приукрасить рассказ о себе любимом и не сказать, что зарабатываешь больше, чем на самом деле? Кроме того, опрос учитывает только тех, кто взял на себя труд ответить на вопросы — и только

тех, кого исследователи смогли найти. А кого легче всего найти? Конечно, людей богатых и известных. «Что же это за заблудшие овечки, которые в списках выпускников значатся под рубрикой «адрес неизвестен?» — спрашивает Хафф. Йель следит за успехами выпускников-миллионеров, но их менее успешные товарищи могли не попасть в его поле зрения. Все это говорит о том, что в результатах опроса зарплата намного выше, чем в реальности. Хафф галопом пробегается по галерее статистических преступлений — от рекламы зубной пасты, подтвержденной лишь искусно подобранными данными, до схем, которые, в зависимости от того, как их раскрасить, меняют значение. Хафф выразился вот как: «Пройдохам эти фокусы давно известны. Люди честные должны о них знать, чтобы их не обвели вокруг пальца».

Если вы прочитаете «Как лгать при помощи статистики», вы поймете, как часто вас обманывают числа. Это умная и полезная книга.

Но я больше десятилетия рассказываю людям о статистике и проверяю заявления, основанные на числовой информации, — и чем дальше, тем больше у меня сомнений по поводу книжки «Как лгать при помощи статистики» и того, что она собой представляет. Что можно подумать о статистике — и о нас самих — если самая известная книга на эту тему от корки до корки состоит из рассказов о ложной информации?

Дарелл Хафф напечатал свою книгу в 1954 году. Но в том же году случилось кое-что другое: британские ученые Ричард Долл и Остин Брэдфорд Хилл выпустили одно из первых убедительных исследований, демонстрирующих, что курение сигарет приводит к раку легких³.

Без статистики Долл и Хилл никогда не пришли бы к этому выводу. За 15 лет число случаев рака легких в Соединенном Королевстве выросло в шесть раз. Уже в 1950 году страна лидировала по числу случаев заболеваний, а число смертей от рака легких впервые превысило число смертей от туберкулеза. Статистический подход нужен был уже хотя бы для того, чтобы заметить эти изменения. Доктора-одиночки располагали информацией лишь о единичных случаях.

Статистика была необходима и чтобы доказать, что убивают именно сигареты. Многие тогда полагали, что рак легких возникает из-за автомобилей. В этом была своя логика. В первой половине двадцатого века автомобили стали привычным явлением, а вместе с ними появились и выхлопные газы, и всепроникающий запах новых асфальтовых дорог. В то же самое время выросло и число случаев рака легких. Чтобы понять, в чем дело, чтобы догадаться, что не машины, а сигареты вызывают рак легких, недостаточно просто оглядеться по сторонам. Для этого нужны ученые, которые тщательно считают и сравнивают данные. Короче говоря — нужна статистика.

Многие скептически относились к гипотезе о сигаретах, хотя она была и не нова. К примеру, нацистская Германия активно пыталась доказать вред сигарет, а Адольф Гитлер относился к ним с отвращением. Не сомневаюсь, что фюрер был доволен, когда немецкие врачи доказали, что курение приводит к раку. Но нацисты — так себе пример для подражания, и табак не терял своей популярности.

Поэтому Долл и Хилл решили провести свои собственные статистические изыскания. Ричард Долл был красивым, тихим и безупречно вежливым молодым человеком, участником Второй мировой. У него была масса идей, как использовать статистику для революции в медицине. Его наставник, Остин Брэдфорд Хилл, служил пилотом в Первой мировой, а потом едва не умер от туберкулеза*. Харизматичный и острый на язык Хилл считался одним из лучших специалистов по медицинской статистике в двадцатом веке⁴. Их совместная работа по анализу данных спасла не одну жизнь.

Коллеги начали свое исследование первого января 1948 года. Под предводительством Долла они исследовали порядка 20 больниц на северо-западе Лондона. Всякий раз, когда в больницу прибывал пациент с раком, медсестра должна была подобрать ему в пару человека того же пола и примерно того же возраста — руководствуясь

* В качестве акта сладкой мести Хилл позже показал, как вылечить туберкулез, в ходе того, что принято считать первым строго рандомизированным клиническим исследованием.

одной лишь случайностью. И больных раком, и их «напарников» подробно опрашивали о том, где они живут и работают, как проводят свое время, что едят, а также как долго и как часто курят. Понемножку, неделя за неделей, месяц за месяцем набиралось все больше ответов.

В октябре 1949 года, меньше чем через два года после начала исследований, Долл бросил курить. Ему было 37, и всю свою взрослую жизнь он был курильщиком. Что же обнаружили Долл и Хилл? Курение не просто удваивает, утраивает или учетверяет риск рака легких. Курение увеличивает этот риск в 16 раз⁵.

Результаты исследования Хилл и Долл опубликовали в сентябре 1950 года и тут же принялись за более масштабное, длительное и амбициозное исследование. Хилл написал каждому врачу в Соединенном королевстве — а было их 59 600 — и попросил заполнить «опросник» касательно здоровья и курения. Рассуждали они так: врачи с легкостью смогут описать, что и когда они курили. Они внесены в медицинский реестр, значит, их всегда можно будет найти без труда. А когда доктор умирает и устанавливают причину его смерти, этой информации точно можно доверять. Доллу и Хиллу просто нужно было немножко подождать.

На призыв Хилла откликнулись более 40 тысяч врачей, но рады ему были далеко не все. Следует понимать, что курение в то время было крайне распространено. Неудивительно, что в первоначальной выборке Долла и Хилла 85% врачей-мужчин курили. Кому приятно слушать, что ты медленно, но верно ведешь себя к верной гибели — тем более если твой метод самоубийства вызывает такую сильную зависимость?

Однажды на лондонской вечеринке Хилла остановил какой-то сердитый врач и выдал: «Так это вы хотите, чтобы мы бросили курить». «Вовсе нет, — ответил Хилл, который в то время еще курил трубку. — Если вы продолжите курить, мне любопытно знать, как вы умрете. Если бросите курить — мне тоже любопытно, как

вы умрете. Так что выбор — курить или не курить — за вами. Мне это безразлично. И так и так все подсчитаю»⁶.

Не правда ли, обаятельный парень? Я забыл упомянуть, что Хилл учился на экономиста.

Исследование продолжалось на протяжении десятилетий, но довольно быстро Долл и Хилл набрали достаточно данных, чтобы с уверенностью заявить: курение вызывает рак легких, и чем больше вы курите, тем больше риск. Более того — курение также вызывает сердечные приступы (это была новая гипотеза).

Врачи не лыком шиты. Когда в 1954 году результаты исследования были опубликованы в «Британском медицинском журнале», профессиональном журнале медиков, они сделали соответствующие выводы. В тот же год Хилл бросил курить, а с ним и многие его коллеги. Врачи стали первой социальной группой в Великобритании, которая массово бросала курить.

Получается, что в один и тот же год, в 1954 году, в мире было два разных подхода к статистике. Многочисленные читатели книги Дарелла Хаффа «Как лгать при помощи статистики» считали статистику игрой, популярной среди шулеров и жуликов, — и развлекались попытками раскусить негодников. Но для Остина Брэдфорда Хилла и Ричарда Долла статистика была делом серьезным. Они понимали: в их игре на кону самый большой куш на свете, и, если играть усердно и по-честному, они смогут спасти сотни жизней.

К весне 2020 года стало ясно, насколько велико значение тщательных, своевременных и верных статистических подсчетов. Мир захватил новый коронавирус.

Давно уже политикам не приходилось принимать такие масштабные решения — а принимать их надо было в спешке. Эти ключевые решения основывались на исследовательской работе эпидемиологов, специалистов по медицинской статистике и экономистов. Речь шла о десятках миллионов жизней и рабочих местах целых миллиардов людей. Специалисты по анализу данных отчаянно пытались понять, с чем же столкнулась наша планета, но данные приходили намного позже, чем нужно. Эпидемиолог Джон Иоаннидис писал в марте 2020 года, что

Covid-19 — «провал доказательной медицины — возможно, крупнейший на нашем веку»⁷. Приходилось на ощупь разбираться в вопросах жизни и смерти. Однако через несколько недель аналитики данных начали набрасывать портрет вируса и вызываемой им болезни. Обнаружилось, что многие становятся заразными еще до проявления симптомов. Обнаружилось, что некоторые вообще не испытывали симптомов болезни. Надо было понять, сколько таких людей. (Сейчас полагают, что около 25%. Но если поразмыслить, вы поймете, что без тщательной аналитической работы ответить на этот вопрос нелегко.) Мы быстро смекнули, что для пожилых риск намного выше, чем для молодежи. Удалось даже прикинуть долю смертности от инфекции. В богатых странах с большой долей пожилых людей она составляла что-то около одного процента. Не одно копье сломалось об это число, а тем временем оно изменяется, ведь медицинская наука движется вперед, а вирус мутирует. Но эти ранние оценки оказались на удивление живучи.

Хоть кое-что так и продолжает оставаться тайной, с течением времени картина прояснялась. Нам удалось расширить программы тестирования, проанализировать клинические случаи и даже отследить генетические мутации вируса. Не зря было столько шума по поводу тестирования: оно помогло увидеть невидимого врага при помощи сбора данных — особенно в случаях, когда тесты проводились и результаты анализировались по правилам, как это делала, например, Национальная статистическая служба Соединенного королевства. В дни пандемии статистика играла роль радара.

Я начал писать эту книгу задолго до того, как вирус Sars-Cov-2 инфицировал свою первую жертву. Смертельная пандемия вновь и вновь подтверждала правильность трех моих главных тезисов, но это меня не радовало.

Первый тезис такой: наши эмоции, предубеждения и политические взгляды могут сильно исказить то, как мы интерпретируем полученные данные. Это ключевая проблема книги, и ей посвящена первая глава.

Это легко заметить на примере того, как США отреагировали на пандемию. В конце февраля 2020 года президент Трамп заявил: «Она пройдет. Наступит день, когда она пройдет — просто сама по себе». Выдавать желаемое за действительное — отличный прием. Прошло четыре недели, а мистер Трамп продолжал воодушевленно рассказывать, как уже на Пасху все смогут пойти в церковь⁸. К моменту, когда он оставил свой пост, умерли 400 000 американцев.

Если дело касается Трампа, люди всегда четко делятся на два лагеря. Подозреваю даже, что, скажи он: «Как приятно в жаркий день съесть эскимо», — часть американцев станет питаться одним лишь эскимо, а другая устроит шумные демонстрации возле кафе-мороженых. То же было и с ковидом. Многие сторонники Трампа гордились, точно медалью, своей безрассудной и извращенной привычкой не надевать маску ни при каких обстоятельствах. Его противники ударились в другую крайность: много шума наделал твит американского журналиста, который писал, что пандемия в Британии «вышла из-под контроля», так как люди, гуляя в парке, не надевали маски. Британца такое заявление приводит в замешательство, но и только. Все данные показывают, что риск передать вирус во время прогулки на открытом воздухе очень низок, и неважно, есть на вас маска или нет. К тому же в конце января 2021 года, когда этот твит был опубликован, ситуацию в Британии неподконтрольной никак нельзя было назвать: число случаев заболеваний стремительно падало.

Этот твит имеет смысл расценивать исключительно как политический выпад в битве двух непримиримых сторон касательно правильного ношения маски. Ни те, ни другие не были заинтересованы в поиске истины. Как это ни странно, но фанатическую приверженность к той или иной группировке проще заметить, если сами вы в этом не замешаны. Если вы воюете на стороне республиканцев или демократов, вы слишком погружены в борьбу и неспособны как следует рассуждать. А если ты стоишь в сторонке и с интересом наблюдаешь за сражением, ясную голову сохранить гораздо проще.

Во-вторых, от решений политиков зависит, какие статистические данные мы собираем и распространяем, а также какие данные игно-

рируются или скрываются. К примеру, в самом начале ковидного кризиса геополитическая борьба осложнила распространение достоверной статистической информации — к этой проблеме мы еще вернемся в Главе 8. Тайвань жаловался, что в конце декабря 2019 года он предоставил Всемирной организации здравоохранения важные свидетельства того, что вирус способен передаваться от человека к человеку. Но даже в середине января ВОЗ все еще с оптимизмом заявляла, что Китай не нашел свидетельства о передаче вируса от человека к человеку. (Тайвань не состоит в ВОЗ, так как Китай считает его своей территорией и требует, чтобы остров не рассматривали как независимое государство. Не исключено, что именно это мешало своевременному распространению информации — если эта задержка и правда имела место)⁹.

А так ли это важно? Скорее всего, да. Поскольку количество случаев удваивалось каждые несколько дней, кто знает, как развивались бы события, узнай мы об этом на пару недель раньше.

В-третьих, статистика позволяет нам увидеть вещи, которые в противном случае так и остались бы незамеченными. Это не просто способ приукрасить газетную статью или колынуть противника в политическом споре. Получаем ли мы надежную статистическую информацию или искаженные, неполные данные — вопрос жизни и смерти.

Когда весной 2020 года я работал над черновиком первого издания «Как посчитать все на свете», многое было неизвестно, много информации было недоступно. Каждую пару дней рос уровень смертности — не было времени ждать у моря погоды. Политические лидеры погрузили экономику в искусственную кому: в течение одной лишь недели в конце марта более трех миллионов американцев называли себя безработными, что в пять раз превышает предыдущий «рекорд». Следующая неделя была еще хуже: безработными называли себя шесть с половиной миллионов. Так ли ужасны были перспективы здравоохранения, чтобы лишить заработка стольких людей? Тогда считали, что да, — но у эпидемиологов было очень мало информации, и они должны были действовать наугад.