Путеводитель по лжи

Автор:
Дэниел Левитин

Путеводитель по лжи

Дэниел Левитин

Информацию сегодня можно получить практически мгновенно, но стоит ли ей доверять? Как определить, не содержатся ли в новостях псевдофакты, искажения данных или откровенная ложь? Чтобы найти истину, нужно мыслить критически. Именно этому – критическому мышлению – и учит Дэниел Левитин. Он дает набор инструментов, овладев которыми, вы сможете обрабатывать входящую информацию и выявлять ложь.

Это книга для всех, кто хочет отличать правду от лжи и свести к минимуму вероятность неверных решений и ошибочных выводов.

Дэниел Левитин

Путеводитель по лжи. Критическое мышление в эпоху постправды

Библиотека фонда «Эволюция»

Научный редактор Надежда Шихова

Издано с разрешения Insula Corporation c/o The Wylie Agency (UK) LTD

Все права защищены.

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

© A Field Guide to Lies, 2016, Daniel Levitin. All rights reserved

© Перевод на русский язык, издание на русском языке, оформление. ООО «Манн, Иванов и Фербер», 2018

* * *

Посвящается моей сестре Шари, чей пытливый ум заставил меня мыслить лучше

О фонде «Эволюция»

Просветительский фонд «Эволюция» основан в 2015 году сообществом российских просветителей. Цель фонда – популяризация научного мировоззрения, продвижение здравомыслия и гуманистических ценностей, развитие науки и образования. Одно из направлений работы фонда – поддержка издания научно-популярных книг. Каждая книга, выпущенная при содействии фонда «Эволюция», тщательно отбирается серьезными учеными. Критерии отбора – научность содержания, увлекательность формы и значимость для общества. Фонд сопровождает весь процесс создания книги – от выбора до выхода из печати. Поэтому каждое издание библиотеки фонда – праздник для любителей научно-популярной литературы.

Больше о работе просветительского фонда «Эволюция» можно узнать по адресу www.evolutionfund.ru (http://www.evolutionfund.ru/).

Фонд «Эволюция» выражает огромную благодарность Евгению и Алине Дябиным за поддержку издания этой книги.

Введение. Критическое мышление

Для начала я скажу две вещи, которые обязательно кого-нибудь да разозлят. Во-первых, язык, который мы используем, стал стирать грань между выдумкой и фактами. Во-вторых, это становится опасным побочным эффектом недостаточного образования, которое повлияло на целое поколение. В силу этих двух фактов ложь разрослась до неимоверных масштабов, и нам становится все сложнее принимать правильные решения, касающиеся нас самих, а также наших сограждан.

Так что же случилось с нашим языком? В 2016 году словом года в Оксфордском словаре стало понятие «постправда», которое определяют как «нечто, относящееся к обстоятельствам или обозначающее обстоятельства, в которых объективные факты влияют на формирование общественного мнения меньше, чем обращение к эмоциям и личным убеждениям». Это слово стало победителем, так как количество его использований в течение года взлетело до небес. Уверен, нам нужно снова вернуться к старому доброму слову «правда» – и поскорее. А также нужно отказаться от мысли о том, что правды больше не существует.

Нам всем свойственно более чем осторожно относиться к фальши. Возможно, пытаясь избежать личных разногласий и «просто поладить», мы стали использовать эвфемизмы, говоря о том, что иначе чем чистым безумием не назовешь. Ложная информация о том, что в пиццерии Comet Ping Pong в Вашингтоне велась сексуальная работорговля под руководством Хиллари Клинтон, привела к тому, что 28-летний житель Солсбери Эдгар Уэлч проехал 560 километров от своего дома до Вашингтона и в воскресенье 4 декабря (через несколько дней после того, как «постправда» стала словом года) открыл внутри пиццерии огонь из полуавтоматического оружия. Нью-йоркская газета Daily News назвала эту ложь «странной теорией». Теория, между прочим, не просто идея – это идея, выведенная путем тщательного анализа данных. И не просто данных, а только тех, что имеют отношение к рассматриваемой проблеме, отобранных строго и беспристрастно.

Другие эвфемизмы, используемые для лжи, – «контрзнания», «полуправда», «экстремистские взгляды», «альт-истина», «теория заговора» и недавнее название «выдуманные новости».

Словосочетание «выдуманные новости» звучит слишком игриво, это слишком похоже на ситуацию, когда школьник прикидывается больным, чтоб пропустить контрольную. Эти эвфемизмы скрывают тот факт, что история про сексуальное рабство – беспардонная ложь. Те, кто об этом писал, знали, что это неправда. Не может быть двух мнений, если одно из них – заведомая ложь. Журналисты, да и все мы, должны перестать уделять внимание суждениям, если они не подкреплены фактами. Две точки зрения возможны тогда, когда каждая из них опирается на объективные данные. Разумные люди могут спорить относительно того, как оценивать эти данные и какие выводы из них следуют. Каждый, конечно, имеет право на собственное мнение, но не имеет права на собственные факты. Ложь – это отсутствие фактов, а во многих случаях и прямое противоречие им.

Истина имеет значение. Эпоха постправды – это эпоха умышленной иррациональности, сводящая на нет все великие достижения человечества. Возможно, журналисты не называют «выдуманные новости» ложью, потому что не хотят оскорблять лжецов. А я говорю: не бойтесь их оскорбить! Вызовите их на ковер.

Возможно, лучше сформулировать вопрос иначе: «Что происходит с системой образования и институтами в преддверии наступления эпохи постправды?» Количество книг, которые читают студенты, в среднем уверенно снижается каждый год начиная со второго класса. Пятнадцать лет назад Департамент образования Соединенных Штатов обнаружил, что один из пяти взрослых американцев не может даже найти информацию в тексте, не говоря уже о том, чтобы мало-мальски проанализировать его и сделать выводы. По-видимому, мы не сумели обучить наших детей тому, что такое доказательство и как его оценивать. Наше возмущение оправданно. Эдгар Уэлч, стрелок из Comet Ping Pong, рассказал властям, что поверил в эту историю после того, как прочитал о ней в интернете. Наша информационная инфраструктура могущественна. Она может приносить пользу, но может и причинять вред. И каждый из нас должен знать, где одно, а где другое.

Возможно, Уэлч относился к ситуации исходя из этой амбивалентности, однако никаких доказательств, что какое-либо расследование вообще имело место, нет.

Похоже, этот невежда даже не знает, что значит собирать и оценивать доказательства. В таком случае кто-нибудь мог намеренно поискать связь между Хиллари Клинтон и рестораном, увидеть в поведении Клинтон интерес к ведению секс-бизнеса и даже мотив, подтолкнувший ее извлечь из этого выгоду (конечно, она не могла быть финансовой, если вспомнить недавнюю суматоху, связанную с гонораром за ее выступления). Он мог заметить, выходили ли дети, занимавшиеся проституцией, и их клиенты из заведения. А за неимением необходимого склада ума и образования для проведения собственного расследования можно почитать профессионалов, занимающихся расследовательской журналистикой, и поинтересоваться, как они видят ситуацию. Ни один преданный профессии журналист не поверит в эту историю, что уже говорит о многом. При этом я прекрасно понимаю, что есть люди, которые считают журналистов коррумпированными и утверждают, будто те продались правительству.

Бюро трудовой статистики Соединенных Штатов сообщает, что в стране насчитывается 45 790 репортеров и корреспондентов. По оценкам Американского общества новостных редакторов, являющегося независимой торговой группой, 32 900 репортеров работают в 1400 газетных изданиях, выходящих ежедневно. Некоторые журналисты могут быть коррумпированы, но при таком количестве весьма маловероятно, что все.

Facebook прилагает все усилия, чтобы оправдать свои социальные обязанности как источник информации, «с помощью которого 1,8 миллиарда пользователей могут сообщать выдуманные новости». Иными словами, распространять ложь.

Возможно, когда-нибудь в будущем и другие социальные сети будут играть стратегически более важную роль. По крайней мере мы можем надеяться, что их роль в том, что ложь используют в качестве оружия, уменьшится.

Многие новостные организации изучали, откуда взялась история о растлении малолетних в той пиццерии. Компания NBC сообщила о процветающем сообществе, распространяющем «выдуманные новости», в Велесе, Македония, – оно вполне могло быть источником. До 1991 года этот регион относился к коммунистической Югославии. Новостная медиакомпания BuzzFeed совместно с газетой Guardian обнаружила более сотни появившихся там поддельных названий новостных доменов. Молодые люди из Велеса, не имеющие никакого отношения к политическим партиям Соединенных Штатов, сочиняют лживые истории, реклама которых хорошо оплачивается по принципу «пенни за клик» на

таких платформах, как Facebook. Подростки могут зарабатывать тысячи долларов в городах с небольшими экономическими возможностями. Должны ли мы винить их за выстрелы в пиццерии? Или это вина социальных сетевых платформ? Или виновата образовательная система, воспитавшая граждан, не умеющих осмысливать заявления, с которыми мы сталкиваемся каждый день?

Вы можете возразить: «Но я не обязан критически оценивать статистику. Есть газеты, блогеры, правительство, Википедия, которые должны делать это за нас». Да, должны, но не всегда могут, им становится все труднее и труднее поспевать проверять информацию, так как ложь размножается быстрее, чем они могут с ней справиться. Это как компьютерная игра «Убей крота». Репортаж о пиццагейте собрал более миллиона просмотров, в то время как материалы по разоблачению этой истории, собранные сайтом Snopes, просмотрели менее 35 тысяч человек. Нам очень повезло, что в США свободная пресса, – исторически у большинства наций было намного хуже. Но мы не должны принимать свободу, честность и принципиальность СМИ как должное. Журналисты и те компании, которые им платят, будут и дальше помогать нам выявлять и разоблачать обман, но они не смогут делать это в одиночку – ложь победит, если доверчивое и неподготовленное общество окажется готово эту ложь принять.

Конечно, большинство из нас не поверили бы, что Хиллари Клинтон могла управлять подобным притоном в Вашингтоне. Данная книга не только о такого рода нелепостях. Действительно ли вам нужно это новое лекарство или же миллиардная маркетинговая кампания, стоящая за ним, раскручивает вас на покупку с помощью тщательно отобранных псевдоданных? Как нам узнать, действительно ли виновен человек, оказавшийся в суде? Как оценить ту или иную инвестицию или объем противоречивых данных, полученных в результате голосования? Что остается за пределом наших когнитивных способностей просто потому, что нам предоставили недостаточно информации?

Лучшая защита против хитрых и изворотливых людей – умение критически мыслить. Нам не удалось объяснить детям, как бороться со склонностью к доверчивости, подаренной нам эволюцией. Мы – социальный вид и склонны верить тому, что нам говорят. Наш мозг – такой механизм, который прекрасно придумывает и рассказывает истории: если нам дать оригинальную идею, мы можем легко придумать заковыристое объяснение, почему она хороша. Но в этом и заключается различие между образным и критическим мышлением, между ложью и истиной: истина подкрепляется фактами, объективными

доказательствами. Некоторые утверждения могут быть истиной, но истинные утверждения истинны всегда.

В ходе изучения Стэнфордским университетом умения пользователей рассуждать относительно информации из интернета за полтора года (вплоть до июня 2016-го) было протестировало более 7800 учеников старших классов школы и студентов колледжа. Исследователи ссылаются на «потрясающую и вместе с тем тревожную тенденцию. В целом умение молодых людей рассуждать об информации из интернета можно обозначить одним словом: “никакое”». Они с трудом могли отличить новости высокого качества от лжи. Мы должны начать учить их этому незамедлительно. И пока часть из нас занята этим, другая часть может пойти на курсы повышения квалификации. К счастью, мышление, в основе которого лежит работа с фактами, доступно для большинства 12-летних.

Многие говорили, что история про пиццагейт была одним из примеров «выдуманных новостей», но давайте называть вещи своими именами – это была наглая ложь. В выдуманных новостях нет «новости» как таковой. Вера в ложь может быть безобидной, так же как и вера в Санта-Клауса или в то, что новые джинсы меня стройнят. Не СМИ и не Facebook делают ложь инструментом воздействия на умы. Опасность заключается в силе убеждения – в чрезмерной, безоговорочной уверенности в том, что это правда.

Критическое мышление учит нас смотреть на ситуацию немного со стороны, оценивать факты и делать выводы, основываясь только на них. То, что заставило Уэлча расчехлить огнестрельное оружие в вашингтонской пиццерии, стало следствием его полной неспособности понять, что его взгляд на проблему был в корне неправильным. Самый важный компонент критического мышления, которого так не хватает в нашем обществе, – смирение. Это простая, но очень глубокая мысль: если мы поймем, что знаем далеко не всё, то сможем узнать больше. Если мы будем думать, что знаем всё, научиться чему-нибудь будет невозможно. Так получилось, что наша система образования, равно как и наша зависимость от интернета, привели к появлению поколения, которое не знает, что оно не знает. Если мы смиримся с этим и примем как правду, то сможем воспитать американский ум, вернуть былую образованность и обезвредить огромное количество лжи, ставшей орудием воздействия на умы и угрожающей нашему миру. Это единственная возможность добиться процветания демократии.

Три вида стратегической защиты

Я начал писать эту книгу в 2001 году, пока читал в колледже курс по критическому мышлению. Я усердно работал над ней в период с 2014 по 2016 год, книга вышла под названием A Field Guide to Lies и имела несколько другое введение. С тех пор, однако, опасность, равно как и охват действия лжи, стали просто ошеломляющими. Теперь уже ложь – это не то, о чем можно говорить с сарказмом в голосе или над чем можно похихикать. Она стала оружием. И опасность эта, возможно, будет только усиливаться и приведет к проблемам, которых мы не знали на протяжении многих поколений. Может, однако, случиться и так, что она пройдет стороной и не будет иметь никаких серьезных последствий. В любом случае инструменты, предложенные на страницах этой книги, – такие же, как и в первом издании: они не зависят от политических, социальных и экономических веяний.

Часто проблема заключается в том, каким источником вы пользуетесь. В старину бумажные книги и новостные статьи вызывали доверие, в отличие от отсебятины, которую какой-нибудь псих мог напечатать в подвале на домашней пишущей машинке. Интернет, конечно же, все изменил. Фальшивый сайт может выглядеть так же подлинно, как и авторитетный, проверяющий факты, – я приведу позже несколько примеров. Дезинформация чертовски плотно переплетена с информацией настоящей, из-за чего становится трудно отличить одно от другого. А еще дезинформация не структурирована – она объединяет людей самых разных социальных слоев, имеющих самое разное образование, и появляется в местах, где вы и не ожидали ее увидеть. Ложные данные переходят от одного человека к другому, публикуются в Twitter, Facebook, Snapchat, Instagram, Tumblr и других социальных сетях, получая таким образом глобальное распространение. Дезинформация может завладеть миром – и получится так, что люди вдруг начнут верить в то, что на самом деле неправда.

Это книга о том, как разрешать с помощью тех фактов, которые вам встречаются, затруднения, которые могут привести вас к неправильным выводам. Иногда те, кто предлагает вам какие-то факты, так и ждут, что вы сделаете неверное заключение. Иногда они и сами не видят разницы. Сегодня информацию можно получить практически мгновенно, государственные лидеры появляются в ваших социальных сетях, сигналы об «экстренных сообщениях» привлекают ваше внимание ежедневно, даже ежечасно. Но где найти время, чтобы определить, не содержатся ли в новостях псевдофакты, искажения данных или откровенная ложь? Нам нужны эффективные стратегии, чтобы

понимать, стоит ли доверять тому, что нам говорят.

За последние пять лет мы создали больше искусственной информации, чем за всю предшествующую историю человечества. К сожалению, наряду с истинным знанием есть много того, что знанием назвать нельзя: сайты, видео, книги и социальные сети. И это не просто новая проблема. Дезинформация известна человечеству на протяжении тысячелетий, о ней говорили в библейские времена, существовала она и в Классической Греции (V–IV вв. до н. э.)[1 - Дезинформация известна человечеству на протяжении тысячелетий… – Авраам специально дезинформировал правителя Авимелеха, говоря о своей жене Сарре, чтобы защитить себя. Троянский конь тоже был своего рода дезинформацией: он появился как подарок, но внутри сидели воины противника.]. Уникальная проблема, с которой мы столкнулись сегодня, состоит в том, что дезинформация очень быстро распространяется, а ложь может стать мощным инструментом в формировании социальной и политической стратегии, ошибочных путей развития которых мы могли бы избежать.

В последующих главах я объединил эти стратегии в категории. Часть 1 этой книги посвящена числовой дезинформации. Она показывает, как неправильное обращение со статистикой и графиками искажает перспективу и заставляет нас делать ошибочные выводы (а также принимать необоснованные решения). В части 2 я исследую ошибочные аргументы, показывая, как легко быть убедительным, рассказывать истории, которые, пусть ненавязчиво, уводят от фактов. Кроме того, в книге описываются способы, помогающие критически оценивать новости, рекламу, отчеты. Последняя часть раскрывает, что лежит в основе нашей способности определять, что есть истина, а что ложь: научный метод. Это лучший из когда-либо созданных инструментов для разгадывания самых сложных тайн. Он отсылает нас к таким великим мыслителям в истории человечества, как Аристотель, Бэкон, Галилей, Декарт, Земмельвайс и Поппер. В этой части книги я говорю о границах наших знаний, о том, что мы можем и чего не можем знать, в том числе о том, что мы знаем на данный момент и чего пока еще не знаем. Я предлагаю ряд конкретных примеров применения логического мышления в совершенно разных условиях: во время дачи свидетельских показаний в зале суда, при принятии медицинских решений, в магии, в отношении современной физики и теорий заговора.

Мыслить критически не значит ставить под сомнение все подряд – это значит пытаться различать те утверждения, что имеют под собой доказательства, и те, что не обоснованы.

Фанаты статистики и графиков умеют обманывать с их помощью, потому что знают: большинство людей не станут вникать в то, как работают диаграммы. Возможно, этим людям даже кажется, что они недостаточно умны. Но любой человек может с этим разобраться, и как только вы поймете самые базовые принципы работы с графиками, диаграммы быстро раскроют вам свою изящность – или уродство.

Возьмем приведенный ранее пример о том, что количество книг, которые читают учащиеся, начиная со второго класса неуклонно снижается с каждым годом. Как следствие можно говорить о том, что наша система образования испортилась: у детей не вырабатывается полезная привычка к обучению, их не волнует саморазвитие, они не вовлечены в процесс интеллектуально. А теперь остановитесь и задайте себе вопрос: а можно ли считать количество прочитываемых книг адекватным мерилом? Можем ли мы с его помощью делать выводы на эту тему? Второклассники обычно читают очень короткие рассказы, и объем книг увеличивается с возрастом. К средней школе дети могут прочесть «Повелителя мух» (две сотни страниц), а к колледжу – «Войну и мир» (1225 страниц). Возможно, следует оценивать количество прочитанных страниц или то время, которые было проведено за чтением. К моменту окончания школы или уже работая в таких сферах, как юриспруденция, управление, промышленность, финансы и наука, люди могут читать меньше художественных книг, но больше сухих научных статей. Если чиновник не читал никаких книг, но тратил время на изучение Конституции, законов, читал журналы и газеты, можно ли сказать, что он не вовлечен в процесс интеллектуально? Тот факт, что в каком-то материале приводятся статистические данные, не означает, что они имеют отношение к теме статьи. Более того, это исследование, похоже, было проведено компанией, разрабатывающей и продающей программное обеспечение для улучшения читательских навыков, – ребята извлекли выгоду из сообщений о низком уровне чтения. Вот оно, критическое мышление в действии.

Обнаружение ошибочных аргументов, представляющих собой часть каких-то историй, поможет вам оценить, ведет ли цепь рассуждений к обоснованным выводам. Информационная грамотность включает знание о том, что существует иерархия источников, что псевдофакты могут легко маскироваться под истины, а предвзятое отношение – искажать информацию, из-за чего мы делаем неверные выводы и приходим не к тем результатам. Иногда доказательства опираются на числа, и тогда мы задаемся вопросом: «Откуда они взялись? Как они были собраны?» Бывает, цифры выглядят нелепо, но для того чтобы это увидеть, нужно хорошенько над ними поразмыслить. Иногда утверждения кажутся обоснованными, однако исходят из источника, который не заслуживает

доверия, как в случае с человеком, заявляющим, что он свидетель преступления, хотя на самом деле его там не было. Благодаря этой книге вы научитесь распознавать неправду[2 - …научитесь распознавать неправду. – По мотивам книги Д. Хаффа (1954/1993) How to Lie with Statistics[268 - Издана на русском языке: Хафф Д. Как лгать при помощи статистики. М.: Альпина Паблишер, 2015.]. New York: W.W. Norton, p. 19. И, как вы узнаете дальше, он, вероятно, вторит Марку Твену, или Джошу Биллингсу, или Уиллу Роджерсу, а может, и еще кому-нибудь.]. И остановите пройдох, ставших на путь обмана.

Часть 1. Оценка цифр

Проблемы вам создает не то, чего вы не знаете, а скорее то, в чем вы абсолютно уверены, тогда как на самом деле заблуждаетесь.

Марк Твен

Правдоподобие

Статистика – это цифры, а потому она кажется нам холодной и жесткой. Есть ощущение, что она отражает факты, которые нам дала сама природа, и наше дело всего лишь их найти. Но важно помнить, что собирают статистические данные люди, и никто иной. Люди выбирают, что считать, как считать, какими результатами делиться, какими словами их описывать и как интерпретировать числа[3 - Люди выбирают, что считать… – практически прямая цитата из Best, J. (2005). Lies, calculations and constructions: beyond How to Lie with Statistics. Statistical Science, 20(3), 210–214.]. Статистика – это не факты, это интерпретация. И ваша интерпретация может быть такой же хорошей (равно как и такой же плохой), как и та, что вам предлагает другой человек.

Числа не всегда верны, и для начала проще всего быстренько проверить их на правдоподобие. Даже если они прошли проверку, у вас могут возникнуть вопросы трех типов: как данные были собраны, как они были интерпретированы и как представлены графически. Ответы на них помогут вам сформулировать

правильные выводы.

Вы можете проверить (в большинстве случаев это возможно), правдоподобен ли факт, в уме или на оборотной стороне конверта. Не принимайте все на веру, попытайтесь разобраться.

Когда мы проводим подобную проверку, точность цифр не очень важна, как бы парадоксально это ни звучало. Достаточно просто здравого смысла: если Берт говорит, что хрустальный бокал упал со стола на ковер и не разбился, это кажется правдоподобным. Если Эрни скажет, что бокал упал с высоты 40-этажного здания на тротуар, не разбившись, это уже будет неправдоподобно. Тут вам помогут знания о том, как устроен мир, а также элементарный жизненный опыт. Подобным же образом, если кто-то вам скажет, что ему 200 лет, или что он постоянно выигрывает в рулетку в Лас-Вегасе, или что он может пробежать 40 миль за час, – все это будет маловероятно и очень неправдоподобно.

Что вы скажете о следующем заявлении?

Все 35 лет после того, как в Калифорнии перестал действовать закон о марихуане, число курильщиков марихуаны удваивается с каждым годом.

Звучит правдоподобно? Давайте разберемся, но с чего же начать? Предположим, что 35 лет назад в Калифорнии был только один курильщик марихуаны – конечно, очень заниженная оценка (в 1982 году по всей стране было полмиллиона арестов за курение марихуаны). Если ежегодно удваивать это число на протяжении 35 лет, получим 17 миллиардов человек – это больше, чем население всего земного шара. (Попробуйте подсчитать сами, и вы увидите, что ежегодное удвоение в течение 21 года приведет вас к числу больше миллиона: 1; 2; 4; 8; 16; 32; 64; 128; 256; 512; 1024; 2048; 4096; 8192; 16 384; 32 768; 65 536; 131 072; 262 144; 524 288; 1 048 576.) Таким образом, это утверждение не то чтобы неправдоподобно – оно попросту невозможно. К сожалению, не у всех получается мыслить четко и ясно, когда речь заходит о числах: многие их просто боятся. Но, как видите, для подобных подсчетов хватит и школьного уровня арифметики плюс здорового скепсиса.

Вот еще один пример. Вас только что взяли в отдел продаж по телефону, и вам нужно обзванивать ничего не подозревающих (и, без сомнения, раздраженных) потенциальных клиентов. Ваш босс, пытаясь вас мотивировать, говорит:

Наш лучший специалист продавал тысячу товаров в день.

Правдоподобно ли это? Попробуйте сами набрать телефонный номер – самое меньшее вам понадобится пять секунд. Плюс еще пять секунд на то, чтобы дозвониться. Теперь давайте предположим, что каждый звонок действительно заканчивается продажей, – это, понятное дело, не очень реалистично, но давайте представим идеальный вариант, чтобы посмотреть, что получится. Добавьте десять секунд: вы проговорите предложение о продаже, а потенциальный клиент его примет. Затем еще 40 секунд, чтобы узнать у него адрес и записать номер кредитной карточки. Это дает один звонок в минуту (5 + 5 + 10 + 40 = 60 секунд), или 60 продаж в час, или 480 продаж за очень напряженный восьмичасовой рабочий день, без перерывов. Так что тысяча товаров, проданных за день, – это нереально, даже при самом оптимистичном раскладе.

Некоторые утверждения оценивать сложнее. Вот, например, заголовок из журнала Time за 2013 год:

Людей с мобильными телефонами больше, чем тех, у кого есть туалет[4 - Людей с мобильными телефонами больше, чем тех, у кого есть туалет. – Wang, Y. (2013, 25 марта). More people have cell phones than toilets. По материалам исследований ООН. http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows/ (http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows/).].

И как быть с таким утверждением? На ум приходят, с одной стороны, жители развивающихся стран, не имеющие водопровода, а с другой стороны – те многочисленные жители процветающих стран, у кого больше одного мобильного телефона. Кажется, что заявление вполне правдоподобно, – это не означает, однако, что мы должны его принять. Скорее, мы не можем отвергать его просто потому, что оно нелепо. Нам потребуются другие техники для его оценки, но

тест на правдоподобие оно прошло.

Иногда нельзя оценить, правдиво ли утверждение, не проведя собственного исследования. Да, конечно, газеты и сайты в интернете должны бы делать это за вас, но так бывает не всегда – вот тогда статистика идет вразнос. Несколько лет назад было очень распространено вот такое утверждение, основанное на статистических данных:

Каждый год в США от анорексии умирает 150 тысяч девушек и молодых женщин[5 - Каждый год в США от анорексии умирает 150 тысяч девушек и молодых женщин. – Steinem, G. (1992). Revolution from Within. New York: Little, Brown. Wolf, N. (1991). The Beauty Myth. New York: William Morrow.].

Хорошо, давайте проверим, насколько этот факт правдоподобен. Согласно данным американских Центров по контролю и профилактике заболеваний, ежегодное количество смертей девушек и молодых женщин в возрасте от 15 до 25 от всех видов заболеваний – 8500. Добавьте сюда женщин от 25 до 45 – показатель все равно достигнет только 55 тысяч[6 - Добавьте сюда женщин от 25 до 45 – показатель все равно достигнет только 55 тысяч. – Этот пример зацепил меня, когда я читал Best, J. (2005). Lies, calculations and constructions: beyond How to Lie with Statistics, 20(3), 210–214. Статистику можно посмотреть на сайте www.cdc.gov (http://www.cdc.gov/).]. Количество случаев смерти от анорексии за год не может превышать в три раза количество всех смертей[7 - Количество случаев смерти от анорексии за год не может превышать в три раза количество всех смертей. – Возможно, вы работаете в бухгалтерии крупной компании. Один из сотрудников потребовал возместить ему расходы на покупку бензина для служебной машины, 5 тысяч долларов за апрель. Призовите на выручку свой жизненный опыт: на сегодняшний день многие машины могут проехать более 70 километров на 10 литрах топлива (а то и в несколько раз больше). Кроме того, вы знаете, что максимально разрешенная скорость 100 км/час, и если бы вам приходилось быть за рулем по десять часов в день, вы бы проезжали 1000 километров. Умножьте это число на стандартный месяц, состоящий из 21,5 рабочего дня, – и вы получите 21 500 километров. Когда речь идет о приблизительных подсчетах, обычно используют круглые числа, чтобы было легче, поэтому давайте считать, что это 21 тысяча километров. Поделите это на расход топлива в 7 км/л – и, если грубо, вашему сотруднику понадобилось бы 3 тысячи литров топлива. Вы выясняете среднюю цену на бензин по стране в

апреле – это 0,64 доллара за литр. Ну давайте скажем, что 3 литра можно купить на 2 доллара (снова округляем и при этом даем вашему сотруднику презумпцию невиновности – ему ведь могло и не посчастливиться заправляться по самой лучшей цене). Итак, если 3 литра можно купить на 2 доллара, то на 3 тысячи литров потребуется 2 тысячи долларов. Расходы в количестве 5 тысяч долларов в отчете о расходах теперь даже отдаленно не выглядят правдоподобными. Даже если бы ваш сотрудник проводил за рулем по 20 часов в день, расходы не были бы такими высокими. https://www.fueleconomy.gov/feg/best/bestworstNF.shtml (https://www.fueleconomy.gov/feg/best/bestworstNF.shtml), на 1 августа 2015 г.].

В своей статье для журнала Science Луи Поллак и Ганс Вайс сообщили, что с момента образования Communication Satellite Corp.

…расходы на телефонные разговоры снизились на 12 тысяч процентов[8 - …расходы на телефонные разговоры снизились на 12 тысяч процентов. – Pollack, L., & Weiss, H. (1984). Communication satellites: countdown for Intelsat VI. Science, 223(4636), 553.].

Если расходы уменьшаются на 100 %, они падают до нуля (и не важно, какими они были изначально). Если же расходы падают на 200 %, это значит, что кто-то платит вам ту же самую сумму, которую когда-то платили вы ему, чтобы получить его продукт. Снижение на 100 % случается крайне редко, а снижение на 12 тысяч процентов кажется и вовсе маловероятным[9 - …снижение на 12 тысяч процентов кажется и вовсе маловероятным. – Полагаю, можно выдумать историю, которая докажет, что это так. Может, какая-нибудь штука стоила раньше один доллар, но теперь, решив сделать ей рекламу, компания не хочет отдавать ее вам бесплатно и платит вам 11,999 доллара, чтобы она стала вашей (а это снижение цены на 12 тысяч процентов). Такое случается на рынке недвижимости и в крупном бизнесе. Бывает, нужно снести какой-нибудь старый захудалый дом, прежде чем на его месте построить новый. Может, владелец должен будет выплатить огромный налог на собственность; расходы на снос дома тоже могут быть высокими, и потому он решает заплатить кому-то, чтобы только не решать вопрос самому. В какой-то момент в конце 1990-х годов несколько крупных отягощенных долгами фирм звукозаписи «продавались» за 0 долларов при условии, что новый владелец возьмет на себя их долги.]. В статье в профессиональном издании Journal of Management Development утверждалось, что количество клиентских жалоб сократилось на 200 % в

результате перехода компании на новую политику поддержки клиентов[10 - …количество клиентских жалоб сократилось на 200 %… – Bailey, C, & Clarke, M. (2008). Выстраивание бизнес-стратегии в соответствии с потребностями компании: цена дискриминации. Journal of Management Development, 27(9), 912–934. Другие примеры снижения на 200 %: Rajashekar, B. S., & Kalappa, V. P. (2006). Effects of planting seasons on seed yield & quality of tomato varieties resistant to leaf curl virus. Seed Research, 34(2), 223–225. http://www.bostoncio.com/AboutRichardCohen.asp (http://www.bostoncio.com/AboutRichardCohen.asp).].

Писатель Дэн Кеппел даже назвал свою книгу Get What You Pay For: Save 200 % on Stocks, Mutual Funds, Every Financial Need («Получите то, за что вы платите: экономьте 200 % на бирже, инвестиционных фондах открытого типа, на любой финансовой потребности»). У Кеппела есть степень MBA. Ему следует лучше разбираться в вопросе. Конечно, чтобы аккуратно сравнивать проценты, их нужно брать от одного и того же базового показателя. Нельзя вернуться к изначальному уровню зарплаты, сокращенной на 50 %, увеличив на 50 % вашу новую, более низкую зарплату[11 - …зарплаты, сокращенной на 50 %… – Illustration © 2016 by Dan Piraro based on an example from Huff, ibid.].

Проценты кажутся простыми и логичными, но иногда они могут и запутать. Если процентная ставка увеличивается с 3 до 4 процентов, то она увеличивается на 1 процентный пункт, или на 33 % (так как увеличение на 1 процентный пункт отсчитывается от базового показателя 3; это увеличение на

/

= 0,33 от 3). Если же процентная ставка упадет с 4 до 3 процентов, то она уменьшится на 1 процентный пункт. При этом она уменьшится не на 33 %, как в предыдущем случае, а на 25 %, потому что уменьшение на 1 процентный пункт отсчитывается от базового показателя 4 (1 – это

/

, или 25 %, от 4). Исследователи и журналисты не всегда щепетильны в этом вопросе и не видят порой разницы между процентными пунктами и процентами, но вы не должны их путать[12 - …не видят порой разницы между процентными пунктами и процентами… – Выражаю благодарность Джеймсу Скэнлану, судебному поверенному из Вашингтона, который ответил на мой запрос в Американскую статистическую ассоциацию и предоставил мне этот пример неверного использования.].

The New York Times сообщила о закрытии текстильной фабрики в Коннектикуте и ее переезде в Вирджинию[13 - …о закрытии текстильной фабрики в Коннектикуте и ее переезде в Вирджинию. – Этот пример был взят из Spirer, L., Spirer, H. F., & Jaffe, A. J. (1987). Misused Statistics, New York: Marcel Dekker, p. 194, Miller, J. (1996, Dec. 29). High costs are blamed for the toss of a mill. New York Times, Connecticut Section. N. a. (1997, Jan. 12). Correction, New York Times, Connecticut Section.]. Причиной такого решения стали возросшие расходы на сотрудников. По сведениям газеты, «фонд заработной платы, все виды компенсации сотрудникам, а также пособие по безработице в Коннектикуте в 20 раз выше, чем в Вирджинии». Правдоподобно ли это? Если бы это было так, вы бы, наверное, ожидали массового исхода в Вирджинию – все компании, а не только эта фабрика, захотели бы переехать, и вы бы уже знали об этом. На самом деле все это неправда, и Times пришлось опубликовать опровержение. Как же такое могло произойти? Дело в том, что журналистка просто неправильно прочла отчет компании. Один показатель – пособие по безработице – на самом деле обходился компании в 20 раз дороже в Коннектикуте, чем в Вирджинии, но с учетом остальных показателей в Коннектикуте все расходы на содержание штата были в целом выше в 1,3 раза, а не в 20 раз. У автора статьи не было образования в сфере бизнес-администрирования – и мы не вправе ожидать этого. Чтобы отследить такого рода ошибку, нужно просто спокойно все обдумать. Это под силу каждому (а журналистка и ее редакторы просто обязаны были это сделать).

В Нью-Джерси одобрили новую законодательную инициативу, согласно которой матери, находящиеся на социальном обеспечении, не получали никаких дополнительных льгот[14 - …одобрили новую законодательную инициативу… – McLarin, K. J., (1993, Dec. 5). New Jersey welfare’s give and take; mothers get college aid, but no extra cash for newborns. New York Times. См. также: Henneberger, M. (1995, April 11). Rethinking welfare: deterring new births – a special report; state aid is

capped, but to what effect? New York Times.]. Некоторые члены законодательного органа посчитали, что женщины в Нью-Джерси специально рожали детей, чтобы увеличить ежемесячное пособие, получаемое от государства. Через два месяца законодатели заявили, что им удалось решить этот вопрос, так как уровень рождаемости снизился на 16 %. Вот что писала New York Times:

Всего два месяца спустя государство опубликовало данные о том, что количество новорожденных детей у женщин, уже находящихся на социальном обеспечении, уменьшилось на 16 %[15 - …количество новорожденных детей у женщин, уже находящихся на социальном обеспечении, уменьшилось на 16 %. – McLarin, K. J., (1993, Dec. 5). New Jersey welfare’s give and take; mothers get college aid, but no extra cash for newborns. New York Times.]. Власти поздравляют себя с потрясающими результатами, которых они добились в такие короткие сроки.

Обратите внимание, что учитывались не беременности, а количество родов. Что же здесь не сходится? Так как беременность длится девять месяцев, никакие изменения за последние два месяца нельзя связывать напрямую с законом. Скорее всего, тут свою роль играют обычные колебания рождаемости (ведь известно, что уровень рождаемости – дело сезонное).

Есть в этом вопросе и другие неточности, которые нельзя обнаружить простой проверкой на правдоподобие:

…с течением времени эти 16 % сократились до 10 %. Дело в том, что государству стало известно о родах, о которых не сообщалось ранее. Оказалось, что многие роженицы не считали нужным сообщить о своих новорожденных детях, так как их социальные дотации на период ухода за ребенком никак не увеличивались[16 - …не считали нужным сообщить о своих новорожденных детях. – McLarin, K. J., (1993, Dec. 5). New Jersey welfare’s give and take; mothers get college aid, but no extra cash for newborns. New York Times.].

Вот вам пример того, с какими проблемами можно столкнуться, собирая статистические данные: оказывается, мы учитываем не всех людей, хотя думаем, что охвачены все. Одни ошибки в рассуждениях заметить проще, другие сложнее, но со временем мы лучше научимся их распознавать. Для начала

давайте взглянем на простой инструмент, который часто используют неверно.

С помощью круговой диаграммы легко представить себе процентные соотношения – то, каким образом распределены разные части единого целого. Например, вы хотите узнать, какой процент школьного бюджета тратится на зарплаты учителям, на учебные материалы, на ремонт. Или же вам хочется выяснить, какая часть денег, ассигнованных на учебные материалы, идет на естественные науки, язык, физкультуру, музыку и т. д. Главное правило круговых диаграмм – сумма процентов во всех секторах должна быть равна 100. Представьте себе пирог: ведь круговая диаграмма – не что иное, как пирог, поделенный на кусочки. Если девять человек хотят разделить его поровну, мы не можем порезать его на восемь частей. И тут ничего нельзя поделать.

Fox News, однако, это не смутило, и они опубликовали вот такую диаграмму:

Главное правило круговых диаграмм: сумма процентов во всех секторах должна быть равна 100. (Fox News, 2010)

Можно легко объяснить, как такое могло произойти. У избирателей есть возможность отдать свой голос более чем за одного кандидата. Однако в таком случае нельзя представлять результаты в виде круговой диаграммы.

Чехарда со средними

Среднее значение может быть весьма полезно, да и разобраться с ним проще, чем с круговой диаграммой. Оно позволяет нам охарактеризовать огромное количество информации одним-единственным числом. Например, мы хотим узнать среднее благосостояние людей в комнате, чтобы понять, принесет ли встреча с ними какую-нибудь пользу нашим фандрайзерам[17 - Фандрайзер (fundraiser, англ.) – тот, кто предоставляет услуги по привлечению ресурсов самостоятельно или по контракту. Здесь и далее, если не указано иное, прим.

ред.] или менеджерам по продажам. Другой пример: мы хотим узнать среднюю цену на бензин, чтобы оценить, во сколько обойдется поездка на машине из Ванкувера в Банф. Однако средние могут быть обманчиво сложными.

Есть три вида средних, и они могут выражаться разными числами. Поэтому те, кто всерьез занимается статистикой, избегают слова «среднее», отдавая предпочтение другим, более точным терминам, как то: среднее арифметическое, медиана или мода. И только так. Иногда все эти величины совпадают, но чаще они различаются. Если вам встретилось слово «среднее», оно, как правило, означает «среднее арифметическое», но нельзя быть в этом абсолютно уверенным.

Чаще других из этих трех встречается среднее арифметическое; оно равно сумме всех данных, поделенной на их количество. Например, среднее благосостояние всех людей, находящихся в комнате, будет равно их общему благосостоянию, поделенному на количество человек. Если в комнате находится десять человек, состояние каждого из которых оценивается в 100 тысяч долларов, то общее богатство составит миллион. Отсюда легко вычислить среднее арифметическое (даже доставать калькулятор не нужно): 100 тысяч долларов. А если состояние каждого присутствующего будет варьироваться от 50 тысяч до 150 тысяч долларов, но общее количество будет по-прежнему миллион, то среднее арифметическое по-прежнему будет 100 тысяч долларов (потому что мы просто разделим миллион на десять, не принимая во внимание, сколько денег на счете у каждого).

Медиана – это число в середине упорядоченного набора чисел (статистики называют его выборкой): половина данных находится ниже этого значения, а половина выше. Как вы помните, смысл среднего значения в том, чтобы охарактеризовать весь объем данных одним-единственным числом. Медиана лучше с этим справляется, если некоторые из ваших данных уж очень отличаются от большинства, статистики называют такие значения выбросами.

Представим себе комнату, в которой находятся девять человек; состояние восьмерых из них равно примерно 100 тысяч долларов, а один находится на грани банкротства, его долг равен 500 тысячам долларов. Вот что у нас получится:

Человек 1: –500 тыс. долл.

Человек 2: 96 тыс. долл.

Человек 3: 97 тыс. долл.

Человек 4: 99 тыс. долл.

Человек 5: 100 тыс. долл.

Человек 6: 101 тыс. долл.

Человек 7: 101 тыс. долл.

Человек 8: 101 тыс. долл.

Человек 9: 104 тыс. долл.

Теперь складываем все показатели и получаем общую сумму в 299 тысяч долларов. Разделим это число на общее количество участников, девять, и получится, что среднее арифметическое равно 33 222 долларам. Создается, однако, впечатление, что среднее арифметическое – не лучший способ охарактеризовать данные о присутствующих. Смею предположить, что фандрайзер не захочет наносить им визит, если среди них найдется человек с показателем-выбросом, который тянет вниз всю группу. В этом и заключается вся трудность работы со средним арифметическим: оно слишком чувствительно к выбросам.

Медиана здесь равна 100 тысячам долларов: четверо зарабатывают меньше этой суммы, а четверо – больше. Мода равна 101 тысяче долларов – это та цифра, которая появляется намного чаще других. И медиана, и мода в этом конкретном примере оказываются гораздо показательнее.

Можно по-разному использовать средние, особенно если вы хотите, чтобы кто-то увидел в ваших данных то, что нужно вам.

Давайте представим, что вы с двумя друзьями запустили стартап – небольшую компанию, в которой работают пять человек. Сейчас конец года, вы собираетесь подвести финансовые итоги и рассказать сотрудникам, как у компании идут дела, чтобы они почувствовали удовлетворение от долгих часов, проведенных в офисе, и от холодной пиццы, съеденной за это время. А еще вы хотите привлечь инвесторов. Скажем, четверо сотрудников – все программисты – за год заработали по 70 тысяч долларов, а один – офис-менеджер – 50 тысяч. Это даст среднее арифметическое зарплат, равное 66 тысячам долларов в год: сумма (4 ? 70 000) + (1 ? 50 000), поделенная на 5. Вы и двое ваших друзей принесли домой по 100 тысяч долларов – это ваша зарплата. Следовательно, общая сумма выплаченных зарплат составит (4 ? 70 000) + (1 ? 50 000) + (3 ? 100 000) = 630 тысяч долларов. Кроме того, пусть ваша компания принесла 210 тысяч долларов чистого дохода, который вы разделили поровну между собой и соучредителями в качестве бонусов, это дает каждому из вас по 100 тысяч + 70 тысяч. Как вы им об этом сообщите?

Вы могли бы сказать:

средняя зарплата сотрудников составляет 66 тысяч долларов;

средняя зарплата + прибыль владельцев составляет 170 тысяч долларов.

И хотя это правда, вряд ли новость понравится кому-нибудь, кроме вас и вашей мамы. Если ваши подчиненные прознают об этом, то решат, что им существенно недоплатили. А потенциальные инвесторы сочтут, что учредителям платят слишком много. Поэтому вы можете оформить отчет по-другому:

средняя зарплата сотрудников составляет 66 тысяч долларов;

средняя зарплата владельцев составляет 100 тысяч долларов;

прибыль: 210 тысяч долларов.

Для потенциальных инвесторов это выглядит убедительнее, тем более что им можно и не говорить о том, что вы поделили прибыль между совладельцами. А сотрудникам последнюю строку можно и вовсе не показывать. Каждый из четырех программистов будет думать, что он на хорошем счету и его ценят, так как зарабатывает он больше, чем большинство. Единственный, кто будет не очень доволен, – ваш офис-менеджер. Но ведь девушка и раньше понимала, что программисты зарабатывают больше нее. Теперь представим: вы чувствуете, что перегружены работой, и хотите уговорить ваших партнеров, которые не сильны в критическом мышлении, что вам нужно нанять еще сотрудников. Вы можете поступить так же, как в таком случае делают многие компании, и заявить о «прибыли на одного сотрудника», поделив прибыль компании, равную 210 тысячам долларов, на пятерых:

средняя зарплата сотрудника: 66 тысяч долларов;

средняя зарплата владельцев: 100 тысяч долларов;

годовая прибыль на одного сотрудника: 42 тысячи долларов.

Теперь вы можете заявить, что 64 % зарплаты, которую вы выплачиваете своим сотрудникам (42 000 / 66 000), возвращаются к вам в виде прибыли, и это означает, что в итоге, получив прибыль, вы должны будете выплатить только 36 % их зарплат. Конечно, эти данные не свидетельствуют о том, что, наняв еще сотрудников, вы увеличите прибыль. Но в глазах того, кто не очень силен в критическом мышлении, это выглядит как весомый довод для увеличения штата.

А что, если вы хотите выглядеть невероятно честным и справедливым работодателем и показать, что разница между вашей прибылью и зарплатой ваших сотрудников довольно разумна? Возьмите прибыль в 210 тысяч долларов и разделите часть этой суммы, 150 тысяч долларов, в качестве бонуса между собой и своими партнерами. Об оставшихся 60 тысячах вы скажете позже, что это и есть «прибыль». На этот раз подсчитайте среднюю зарплату, включив в эти подсчеты себя и своих партнеров вместе с бонусами:

средняя зарплата: 97 500;

средняя прибыль владельцев компании: 20 тысяч.

А теперь повеселимся по-настоящему:

общие выплаты зарплат плюс бонусы: 840 тысяч;

зарплаты: 780 тысяч;

прибыль: 60 тысяч.

Теперь все выглядит разумно, правда? Из всей суммы в 840 тысяч долларов, включающей зарплату и прибыль, только 60 тысяч, или 7 %, составили личную прибыль владельцев. Ваши сотрудники будут думать, что вы безупречны, – кто станет обвинять владельца компании в том, что он присвоил 7 %? По сути, это ведь не так уж много: 7 % делится между владельцами компании поровну, и каждый получает по 2,3 %. Да тут даже возразить нечего!

А можно придумать и кое-что получше. Представьте, что в первый год существования вашей компании у вас были только сотрудники, работавшие неполный день. Они зарабатывали по 40 тысяч в год. На второй год у вас были только сотрудники, работавшие полный день. И они получали 66 тысяч, о которых говорилось выше. В таком случае вы со всей уверенностью можете заявить, что в среднем заработок каждого сотрудника увеличился на 65 %. Вы – великий предприниматель! Правда, вы замалчиваете тот факт, что сравниваете две несопоставимые вещи: работу на неполный и полный рабочие дни. Могу сказать, что в этом вы не первый: американская корпорация по производству стали U.S. Steel додумалась до этого еще в 1940-х годах.

В уголовном судопроизводстве то, как представлена информация, т. е. фрейминг[18 - Изложение логически эквивалентной информации различными способами.], оказывает сильное воздействие на мнение присяжных относительно виновности подсудимого. Хотя математически эти два утверждения эквивалентны[19 - Хотя математически эти два утверждения эквивалентны… – Koehler, J. J. (2001). The psychology of numbers in the courtroom: how to make DNA-match statistics seem impressive or insufficient. Southern California

Law Review, 74, 1275–1305. And Koehler, J. J. (2001). When are people persuaded by DNA match statistics? Law and Human Behavior, 25(5), 493–513.], фраза: «Вероятность того, что обнаруженная на месте преступления кровь совпадет с кровью подозреваемого, если только это действительно не его кровь, составляет всего 0,1 %» (один к тысяче) гораздо убедительнее, чем заявление: «Кровь одного человека из каждой тысячи жителей Хьюстона тоже соответствует найденной».

Средние часто используют для того, чтобы рассказать о результатах, например «один брак из X случаев заканчивается разводом». Но это не означает, что статистика применима к вашей улице, к вашему бридж-клубу или к вашему знакомому. Брак либо закончится разводом, либо нет, но нужно знать определенные факторы уязвимости, чтобы предсказать, кто действительно разведется, а кто нет.

Еще один пример: вы можете прочитать, что один из пяти новорожденных детей – китаец. Вы подмечаете, что у шведского семейства, живущего на вашей улице, уже есть четверо детей, а сейчас они ждут пополнения. Но это не означает, что в семье родится маленький китаец. Среднее значение вычислено по всем рождениям в мире, а не в конкретной семье, в конкретном доме, в конкретном районе или даже стране.

Будьте осторожны со средними, а также с тем, как их интерпретируют. Один из способов ввести в заблуждение, используя средние, – усреднять данные по выборкам из несопоставимых совокупностей. Этот способ может привести к абсурдным выводам, как то:

В среднем у каждого человека одно яичко[20 - В среднем у каждого человека одно яичко. – Приписывается профессору математики Десмонду Макхейлу из Университетского колледжа Корка, Ирландия.].

Этот пример наглядно показывает разницу между средним арифметическим, медианой и модой. Так как женщин в мире несколько больше, чем мужчин, медиана и мода будут равны нулю, в то время как среднее арифметическое будет близко к единице (возможно, оно будет равно 0,98 или около того).

Кроме этого, нужно быть внимательным и помнить, что среднее ничего не говорит о размахе значений. Средняя годовая температура в Долине Смерти

в Калифорнии равна 25 °C, что считается комфортным. Но размах может быть просто убийственным, с колебанием температуры от – 9 до 57 °C, – факт, зафиксированный приборами[21 - …с колебанием температуры от – 9 до 57 °C. – http://en.wikipedia.org/wiki/Death_Valley (http://en.wikipedia.org/wiki/Death_Valley).].

Или… Я мог бы вам сказать, что в среднем благосостояние сотни людей, находящихся в комнате, составляет колоссальную сумму: 350 миллионов долларов. Вы, наверное, думаете: вот бы отправить туда моих лучших менеджеров по продажам. Но в комнате могут находиться Марк Цукерберг (его состояние оценивается в 25 миллиардов долларов[22 - На момент издания книги почти 65 млрд. Прим. науч. ред.]) и 99 бедняков. Таким образом, средний показатель может размыть разницу в важных показателях.

Если вы работаете со средними, остерегайтесь еще бимодального распределения. Вспомните, мода – это то значение, которое встречается чаще всего. Во многих наборах данных – биологических, физических, социальных – у распределения может быть два или больше пиков. А это значит, что два или больше показателей встречаются чаще других.

Например, подобный график может отображать сумму, потраченную на обеды в неделю (ось X), и количество людей, потративших такую сумму (ось Y)[23 - …сумму, потраченную на обеды в неделю… – В качестве примера предположим, что шестеро взрослых людей тратят на обед следующие суммы: 12, 10, 10, 12, 11, 11 долларов; расходы у детей выглядят следующим образом: 4, 3,85, 4,15, 3,50, 4,50, 4 доллара. Медиана (в нашем случае четное число наблюдений, поэтому медиана вычисляется как среднее арифметическое двух срединных значений 4,5 и 10) равняется 7,25 доллара. Среднее арифметическое и медиана отражают ту цифру, которую никто никогда не тратит на самом деле.]. Представьте, что вы изучали две группы людей: детей (левый горб) – они покупают школьные обеды – и руководителей компаний (правый горб) – они ходят в дорогие рестораны. Среднее арифметическое и медиана в данном случае – это числа где-то между этими двумя горбами, и они ничего не скажут нам о том, что происходит на самом деле, – ведь во многих случаях среднее арифметическое и медиана отражают ту сумму, которую никто не тратит. Подобный график говорит лишь о том, что в вашем примере имеет место

неоднородность – вы сравниваете яблоки с апельсинами. В таком случае лучше сразу сказать, что вы имеете дело с бимодальным распределением, и сообщить о двух модах. А еще лучше разделить группу на две подгруппы и собрать статистические данные для каждой.

Будьте осторожны, когда будете делать выводы об отдельных людях и о группах, основываясь на средних данных. Тут можно легко наткнуться на определенные подводные камни, которые даже получили собственные названия: «экологическая ошибка» и «ошибка исключения». Экологическая ошибка возникает, если мы делаем выводы об отдельном элементе, основываясь на совокупных данных (таких как средняя величина группы), а ошибка исключения – если делать все ровно наоборот.

Представьте себе, например, два маленьких городка, в каждом из которых живет всего по сотне человек. Девяносто девять жителей города А зарабатывают по 80 тысяч долларов в год, а на земле одной женщины было найдено месторождение нефти, и теперь она одна получает 5 миллионов долларов в год. В городе Б живут 50 человек, которые зарабатывают по 100 тысяч долларов в год, а также 50 человек, которые получают по 140 тысяч долларов. Средний арифметический доход в городе А составляет 129 тысяч долларов, а в городе Б – 120 тысяч долларов. И хотя средняя величина доходов города А больше, в 99 случаях из 100 доход любого жителя города Б, которого вы выберете наугад, будет выше дохода любого случайно выбранного жителя города А. Экологическую ошибку совершают те, кто считает, что если выбрать наугад человека из группы с более высоким средним доходом, то следует ожидать, что и у него доход будет выше. Самое замечательное в этом примере то, что в городе А выше среднее арифметическое, а мода выше в городе В (так бывает не всегда).

Вот еще один пример: считается, что состоятельные люди скорее проголосуют за республиканца, но, как показывает практика, более состоятельные штаты обычно голосуют за демократов. Дело в том, что общая картина благосостояния жителей процветающих штатов может быть немного перекошена из-за суперсостоятельных индивидуумов. Во время президентских выборов 2004 года за кандидата от Республиканской партии Джорджа Буша проголосовали 15 самых бедных штатов, а за кандидата от Демократической партии Джона Керри – девять из 11 самых состоятельных[24 - Во время президентских выборов 2004 года… – См. Gelman, A. (2008). Red State, Blue State, Rich State, Poor State. Princeton, NJ: Princeton University Press.]. Если же изучить вопрос более детально,

мы увидим, что за Буша отдали голоса 62 % тех, чей годовой доход составляет более 200 тысяч долларов, а за Керри – 36 % голосующих, зарабатывающих в год 15 тысяч долларов или меньше.

Чтобы понять, что такое ошибка исключения, давайте представим себе: вы прочли, что машины марки Volvo считаются самыми надежными, и решили купить такой автомобиль. По дороге в офис этой компании вы проходите мимо механика Volvo и парковки, заполненной машинами этой марки, которые ждут, пока их отремонтируют. Если вы измените свое решение о покупке машины этой марки, основываясь на том, что только что увидели, значит, вы по небольшому числу исключительных данных формируете выводы о целой группе. Никто ведь не говорил, что Volvo вообще не нуждается в ремонте, – скорее речь шла о следующем: вероятность того, что им может понадобиться техническая поддержка, гораздо меньше (отсюда и набившее оскомину предупреждение, которым заканчивается любой рекламный ролик, что «каждая машина индивидуальна»). Обратите внимание, что на вас оказывают огромное влияние и другим образом: единственное место, где должны стоять машины Volvo, требующие ремонта, – автосервис Volvo. Ваш «базовый показатель» сдвинулся, и вы не можете уже считать этот пример случайным.

Сейчас, когда вы уже эксперт по средним, вас не удивит заявление, что 100 лет назад наши предки жили меньше, чем мы сегодня. Возможно, вы читали, что продолжительность жизни в наше время сильно увеличилась. Продолжительность жизни тех, кто родился в 1850 году, была 38 лет для мужчин и 40 для женщин, а у тех, кто родился в 1990-м, она составляет 72 года и 79 лет соответственно[25 - Продолжительность жизни тех, кто родился в 1850 году, была 38 лет для мужчин и 40 для женщин, а у тех, кто родился в 1990-м, она составляет 72 года и 79 лет соответственно. – Эти данные приводятся для белых мужчин и женщин. Данные для цветного населения на 1850 год недоступны. http://www.infoplease.com/ipa/A0005140.html (http://www.infoplease.com/ipa/A0005140.html). Ситуация осложняется тем, что, согласно Бюро переписи населения США, данные на 1850 год были приведены только для штата Массачусетс.]

Конец ознакомительного фрагмента.

notes

Сноски

1

Дезинформация известна человечеству на протяжении тысячелетий… – Авраам специально дезинформировал правителя Авимелеха, говоря о своей жене Сарре, чтобы защитить себя. Троянский конь тоже был своего рода дезинформацией: он появился как подарок, но внутри сидели воины противника.

2

…научитесь распознавать неправду. – По мотивам книги Д. Хаффа (1954/1993) How to Lie with Statistics[268 - Издана на русском языке: Хафф Д. Как лгать при помощи статистики. М.: Альпина Паблишер, 2015.]. New York: W.W. Norton, p. 19. И, как вы узнаете дальше, он, вероятно, вторит Марку Твену, или Джошу Биллингсу, или Уиллу Роджерсу, а может, и еще кому-нибудь.

3

Люди выбирают, что считать… – практически прямая цитата из Best, J. (2005). Lies, calculations and constructions: beyond How to Lie with Statistics. Statistical Science, 20(3), 210–214.

4

Людей с мобильными телефонами больше, чем тех, у кого есть туалет. – Wang, Y. (2013, 25 марта). More people have cell phones than toilets. По материалам исследований ООН. http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows/ (http://newsfeed.time.com/2013/03/25/more-people-have-cell-phones-than-toilets-u-n-study-shows/).

5

Каждый год в США от анорексии умирает 150 тысяч девушек и молодых женщин. – Steinem, G. (1992). Revolution from Within. New York: Little, Brown. Wolf, N. (1991). The Beauty Myth. New York: William Morrow.

6

Добавьте сюда женщин от 25 до 45 – показатель все равно достигнет только 55 тысяч. – Этот пример зацепил меня, когда я читал Best, J. (2005). Lies, calculations and constructions: beyond How to Lie with Statistics, 20(3), 210–214. Статистику можно посмотреть на сайте www.cdc.gov (http://www.cdc.gov/).

7

Количество случаев смерти от анорексии за год не может превышать в три раза количество всех смертей. – Возможно, вы работаете в бухгалтерии крупной компании. Один из сотрудников потребовал возместить ему расходы на покупку бензина для служебной машины, 5 тысяч долларов за апрель. Призовите на выручку свой жизненный опыт: на сегодняшний день многие машины могут проехать более 70 километров на 10 литрах топлива (а то и в несколько раз больше). Кроме того, вы знаете, что максимально разрешенная скорость 100 км/час, и если бы вам приходилось быть за рулем по десять часов в день, вы бы проезжали 1000 километров. Умножьте это число на стандартный месяц,

состоящий из 21,5 рабочего дня, – и вы получите 21 500 километров. Когда речь идет о приблизительных подсчетах, обычно используют круглые числа, чтобы было легче, поэтому давайте считать, что это 21 тысяча километров. Поделите это на расход топлива в 7 км/л – и, если грубо, вашему сотруднику понадобилось бы 3 тысячи литров топлива. Вы выясняете среднюю цену на бензин по стране в апреле – это 0,64 доллара за литр. Ну давайте скажем, что 3 литра можно купить на 2 доллара (снова округляем и при этом даем вашему сотруднику презумпцию невиновности – ему ведь могло и не посчастливиться заправляться по самой лучшей цене). Итак, если 3 литра можно купить на 2 доллара, то на 3 тысячи литров потребуется 2 тысячи долларов. Расходы в количестве 5 тысяч долларов в отчете о расходах теперь даже отдаленно не выглядят правдоподобными. Даже если бы ваш сотрудник проводил за рулем по 20 часов в день, расходы не были бы такими высокими. https://www.fueleconomy.gov/feg/best/bestworstNF.shtml (https://www.fueleconomy.gov/feg/best/bestworstNF.shtml), на 1 августа 2015 г.

8

…расходы на телефонные разговоры снизились на 12 тысяч процентов. – Pollack, L., & Weiss, H. (1984). Communication satellites: countdown for Intelsat VI. Science, 223(4636), 553.

9

…снижение на 12 тысяч процентов кажется и вовсе маловероятным. – Полагаю, можно выдумать историю, которая докажет, что это так. Может, какая-нибудь штука стоила раньше один доллар, но теперь, решив сделать ей рекламу, компания не хочет отдавать ее вам бесплатно и платит вам 11,999 доллара, чтобы она стала вашей (а это снижение цены на 12 тысяч процентов). Такое случается на рынке недвижимости и в крупном бизнесе. Бывает, нужно снести какой-нибудь старый захудалый дом, прежде чем на его месте построить новый. Может, владелец должен будет выплатить огромный налог на собственность; расходы на снос дома тоже могут быть высокими, и потому он решает заплатить кому-то, чтобы только не решать вопрос самому. В какой-то момент в конце 1990-х годов несколько крупных отягощенных долгами фирм звукозаписи

«продавались» за 0 долларов при условии, что новый владелец возьмет на себя их долги.

10

…количество клиентских жалоб сократилось на 200 %… – Bailey, C, & Clarke, M. (2008). Выстраивание бизнес-стратегии в соответствии с потребностями компании: цена дискриминации. Journal of Management Development, 27(9), 912–934. Другие примеры снижения на 200 %: Rajashekar, B. S., & Kalappa, V. P. (2006). Effects of planting seasons on seed yield & quality of tomato varieties resistant to leaf curl virus. Seed Research, 34(2), 223–225. http://www.bostoncio.com/AboutRichardCohen.asp (http://www.bostoncio.com/AboutRichardCohen.asp).

11

…зарплаты, сокращенной на 50 %… – Illustration © 2016 by Dan Piraro based on an example from Huff, ibid.

12

…не видят порой разницы между процентными пунктами и процентами… – Выражаю благодарность Джеймсу Скэнлану, судебному поверенному из Вашингтона, который ответил на мой запрос в Американскую статистическую ассоциацию и предоставил мне этот пример неверного использования.

13

…о закрытии текстильной фабрики в Коннектикуте и ее переезде в Вирджинию. – Этот пример был взят из Spirer, L., Spirer, H. F., & Jaffe, A. J. (1987). Misused Statistics, New York: Marcel Dekker, p. 194, Miller, J. (1996, Dec. 29). High costs are blamed for the toss of a mill. New York Times, Connecticut Section. N. a. (1997, Jan. 12). Correction, New York Times, Connecticut Section.

14

…одобрили новую законодательную инициативу… – McLarin, K. J., (1993, Dec. 5). New Jersey welfare’s give and take; mothers get college aid, but no extra cash for newborns. New York Times. См. также: Henneberger, M. (1995, April 11). Rethinking welfare: deterring new births – a special report; state aid is capped, but to what effect? New York Times.

15

…количество новорожденных детей у женщин, уже находящихся на социальном обеспечении, уменьшилось на 16 %. – McLarin, K. J., (1993, Dec. 5). New Jersey welfare’s give and take; mothers get college aid, but no extra cash for newborns. New York Times.

16

…не считали нужным сообщить о своих новорожденных детях. – McLarin, K. J., (1993, Dec. 5). New Jersey welfare’s give and take; mothers get college aid, but no extra cash for newborns. New York Times.

17

Фандрайзер (fundraiser, англ.) – тот, кто предоставляет услуги по привлечению ресурсов самостоятельно или по контракту. Здесь и далее, если не указано иное, прим. ред.

18

Изложение логически эквивалентной информации различными способами.

19

Хотя математически эти два утверждения эквивалентны… – Koehler, J. J. (2001). The psychology of numbers in the courtroom: how to make DNA-match statistics seem impressive or insufficient. Southern California Law Review, 74, 1275–1305. And Koehler, J. J. (2001). When are people persuaded by DNA match statistics? Law and Human Behavior, 25(5), 493–513.

20

В среднем у каждого человека одно яичко. – Приписывается профессору математики Десмонду Макхейлу из Университетского колледжа Корка, Ирландия.

21

…с колебанием температуры от – 9 до 57 °C. – http://en.wikipedia.org/wiki/Death_Valley (http://en.wikipedia.org/wiki/Death_Valley).

22

На момент издания книги почти 65 млрд. Прим. науч. ред.

23

…сумму, потраченную на обеды в неделю… – В качестве примера предположим, что шестеро взрослых людей тратят на обед следующие суммы: 12, 10, 10, 12, 11, 11 долларов; расходы у детей выглядят следующим образом: 4, 3,85, 4,15, 3,50, 4,50, 4 доллара. Медиана (в нашем случае четное число наблюдений, поэтому медиана вычисляется как среднее арифметическое двух срединных значений 4,5 и 10) равняется 7,25 доллара. Среднее арифметическое и медиана отражают ту цифру, которую никто никогда не тратит на самом деле.

24

Во время президентских выборов 2004 года… – См. Gelman, A. (2008). Red State, Blue State, Rich State, Poor State. Princeton, NJ: Princeton University Press.

25

Продолжительность жизни тех, кто родился в 1850 году, была 38 лет для мужчин и 40 для женщин, а у тех, кто родился в 1990-м, она составляет 72 года и 79 лет соответственно. – Эти данные приводятся для белых мужчин и женщин. Данные для цветного населения на 1850 год недоступны. http://www.infoplease.com/ipa/A0005140.html (http://www.infoplease.com/ipa/A0005140.html). Ситуация осложняется тем, что, согласно Бюро переписи населения США, данные на 1850 год были приведены

только для штата Массачусетс.

268

Издана на русском языке: Хафф Д. Как лгать при помощи статистики. М.: Альпина Паблишер, 2015.

Купить: https://tellnovel.com/ru/levitin_deniel/putevoditel-po-lzhi

Текст предоставлен ООО «ИТ»
Прочитайте эту книгу целиком, купив полную легальную версию: Купить