

ПРАВДА, НАГЛАЯ (ЧЁРТОВА) ПРАВДА И СТАТИСТИКА

Пол Ф. Веллеман

Корнельский университет, Итака, Нью Йорк, США
Departments of Statistical Science and of Social Statistics of the Cornell University,
Ithaca, New York, U.S.A.

АННОТАЦИЯ

Статистики-практики и преподаватели статистики часто вынуждены отталкиваться от пространенного мнения, что статистика учит, как обращаться с данными. Те, кто ошибочно полагает, что статистика является исключительно разделом математики (и, следовательно, алгоритмизирована), часто рассматривают использование суждений [принятия решений] в статистике как доказательство того, что мы действительно манипулируем нашими результатами.

В стремлении дать корректные формулы и определения мы можем не заметить важную роль, которую играют суждения. Мы должны учить наших студентов тому, что они несут личную ответственность за принятие статистических решений. Но мы также должны предложить некое руководство для [построения] статистических правил принятия решений. Такое руководство требует, чтобы мы признали роль этики в статистике. Принцип, определяющий эти правила, должен заключаться в честном поиске правды о мире, а принцип поиска такой истины, должен занимать центральное место в курсах статистики.

Замечание, приписываемое Дизраэли^a, будет часто применяться с известной степенью справедливости и силы: «Есть три вида лжи: ложь, наглая ложь и статистика» (Марк Твен – Twain, Mark).

Это быть может моя наименее любимая цитата о статистике. Но я хотел бы остановиться на том, что лежит как в её основе, так и на радостной готовности многих, кто вообще ничего не знает о статистике, цитировать её так, будто она оправдывает их низкое мнение о дисциплине.

Эта цитата проникла в дискуссии по многим дисциплинам. Наверняка вам ответят, что вы достаточно глупы, чтобы признаться в приличной компании, что вы преподаете статистику.

Найджел Рис^{1,b} утверждает, что это единственное часто используемое замечание в британских СМИ². Поиски книг в Google [по мему] «ложь, наглая ложь и статистика» показывает 495 книг, а общий поиск в Google находит «около 207 000» цитат. Небольшая (неслучайная) выборка этих ссылок показывает, что большинство предназначено для того, чтобы предлагать нечестные манипуляции и интерпретации.

Ключевые слова: наглая ложь; Твен; этика; статистика образования.

1. ИСТОЧНИКИ

Происхождение [мема] «ложь, проклятая ложь...» неясно. Онлайн-материалы по истории статистики, опубликованные Университетом Йорка³, дают хорошую основу для дискуссии. Ясно две вещи: Твен не является его автором (и, разумеется, он это не утверждал), и он, скорее всего, ошибочно приписал его Дизраэли. Источник Твена не известен. Наиболее вероятным источником может быть поздравительный адрес, зачитанный в Саратога-Спрингс (Saratoga Springs) в 1895 году (недалеко от дома Твена в Эльмире, Нью-Йорк [Elmira, N. Y]) Леонардом Генри Кортни^c (1832-1918), британским экономистом и политиком, который сказал:

В конце концов, факты есть факты, и, хотя мы можем со смехом цитировать друг другу слова Мудрого государственного деятеля: «Ложь – наглая ложь – и статистика», тем не менее есть некоторые простые цифры, которые поймет любой проstack, а хитрец не сможет из них выбрать-ся.^{4, p. 25}

Вполне вероятно, что Твен услышал или прочитал речь и ошибочно предположил, что «Мудрый государственный деятель» – Дизраэли. Лорд Кортни ясно указывает на ценность статистики, несмотря на забавный комментарий, который он приводит, что неудивительно, поскольку он был президентом Королевского статистического общества с 1897 по 1899 год. Но, как мы увидим, это не противоречит тому, что на самом деле хотел сказать Твен.

Velleman P.F. Truth, damn truth, and statistics // J. Stat. Education. 2008. V. 16, No. 2. P. 1-14 (перевод Г.С. Розенберга). Комментарии [под буквами – a, b, c...] см. в следующей статье.

Есть несколько упоминаний, которые предшествовали как Твену, так и Кортни. Пожалуй, самым интригующим является сообщение в «*Manchester Guardian*» 1892 года (*Wednesday, June 29, Page 5. – Г.Р.*) о политической речи Артура Джеймса Бальфура^d. Бальфур использует эту фразу именно так, как она использовалась бы сегодня, чтобы обвинить политического оппонента в манипулировании статистикой, не замечая меняющегося общего знаменателя:

Но улучшение в Ирландии с 1886 по 1892 год было неудобным фактом, и поэтому Гладстонцы принялись за работу, чтобы доказать, что этот факт вовсе не был фактом... Были сделаны определенные предположения, настолько очевидные для каждого человека, который знал факты, что было напрасно выставлять их напоказ; они не представили готовую статистику, потому что она была бы оскорбительна, но манипулировали ей, на глазах любой аудитории в стране. (Слушайте, слушайте!..) Профессор Мунро^e напомнил мне старую поговорку, которую он довольно неохотно предложил повторить в этой компании: имеет место три градации неискренности – была ложь, отвратительная (*чёртова, d – d. – Г.Р.*) ложь, и статистика (Смех).

По словам г-на Мунро, средние поступления от пассажиров в период с 1881 по 1885 годы, когда г-н Гладстон^f находился в офисе (*был в Правительстве. – Г.Р.*), составляли 1 098 [фунтов] за милю, в то время как средние поступления в период с 1886 по 90 годы, когда г-н Бальфур был в офисе, были только 1092 за милю, показывая снижение на 6 за милю... [Однако] в Ирландии было большое расширение скоростных железных дорог, и в результате... эффект, конечно, постепенно увеличивался... количество миль, по которым должны были быть получены средние транспортные потоки легко рассчитывается (1892, р. 5-6).

Это, конечно, повышает вероятность того, что именно Бальфур, один из лидеров Консервативной партии в течение 50 лет, а не Дизраэли, был «мудрым государственным деятелем», который заслужил упоминание лорда Кортни⁵. Если, как утверждает Бальфур, *bon mot* – это «старая поговорка», мы никогда не узнаем её истинное происхождение.

2. ИСТИНА

Конечно, можно использовать инструменты статистики, чтобы ввести в заблуждение. Но даже те, кто повторяет цитату, не верят, что

цель статистики как научной дисциплины состоит в том, чтобы ввести в заблуждение или что в статистике есть что-то принципиально нечестное. Статистика не испытывает недостатка в уважении только потому, что люди считают статистиков мошенниками. Вы бы купили подержанную машину у статистика? Почему же такое оскорбление так широко распространено? И что мы, как преподаватели статистики, можем извлечь из этой повсеместно употребляемой фразы?

Давайте подойдем к этим вопросам, сначала подумав об истине. Философы обсуждали природу (и даже существование) истины на протяжении веков. Я хочу, как можно дольше избегать этих дебатов, поэтому изложу свои аксиомы и предлагаю не обсуждать их здесь.

- Мир существует вне нас.
- Есть факты об этом мире, которые правдивы в обычном смысле этого слова. То есть они правдивы, независимо от того, что вы, я или кто-либо еще, может в них верить.
- Чтобы быть полезными, теории, модели и объяснения мира должны учитывать наблюдаемые факты.
- Научное (и, следовательно, статистическое) мышление ищет пути, в которых существующее понимание может быть неправильным, потому что такие исключения освещают путь к лучшему пониманию истины о мире.
- Одна из основных целей статистики – фактически, основная цель статистики, – облегчить обнаружение, понимание, количественную оценку, моделирование и распространение фактов о мире.

Другие дисциплины видят статистику иным образом. Некоторые считают статистический анализ, своего рода, *привратником*. Статистическая значимость является первым требованием для публикации во многих общественных науках. Медицинские дисциплины обсуждают «доказательную медицину», что означает выбор методов лечения, основанных на лучших научных (обычно статистически обоснованных) данных на сегодняшний день.

Эти проблемы входят в наше обучение на курсах начальной статистики. Рубрики для теста AP Statistics настаивают на том, чтобы последним этапом вопроса со свободным ответом было одно или два предложения, связывающих результат с реальным вопросом, который мотивировал исследование. Такая позиция (мотивация для статистического анализа – это вопрос о мире, а результаты анализа – это утверждение, касающееся этого вопроса), – становится центральной в нашем исследовании.

3. ФАКТЫ И ПРОЦЕСС

Термин «статистика» имеет здесь два значения, которые нас беспокоят. Это может относиться к изолированным фактам, или это может относиться к анализу и процессам, с помощью которых мы работаем с этими фактами, чтобы достичь более глубокого понимания.

Те, кто выбирает, определяет и обобщает факты, должны их интерпретировать. Изменяющийся знаменатель, на который указывал Бальфур, типичен для суждений, которые могут потребовать даже простые факты. Риск сделать неправильное суждение или быть обманутым кем-то давно признан. По словам Уильяма Гая^е, (см. «Протоколы Британской ассоциации»), одного из основателей Статистического общества Лондона (предшественника Королевского статистического общества) в 1834 г., целью Общества должно было стать приобретение, организация и публикация

...фактов, рассчитанных на иллюстрацию условий и перспектив общества..., первое и самое важное правило [заключалось в том, чтобы] тщательно исключать все мнения (интерпретации) из протоколов и публикаций – строго ограничивать свое внимание фактами... (р. 482).

Гай отмечает невозможность и нежелательность достижения этой цели, а также тот очевидный факт, что ведущие члены Общества почти сразу же её (цель) проигнорировали.

Для нашего понимания нам достаточно признать, что даже самая простая статистика включает суждения (интерпретации). Таким образом, нет никакой практической разницы между двумя значениями при их зависимости от суждений.

4. ЧЁРТ, ПРАВДА

Джон Тьюки^h учил, что статистика – это больше [самостоятельная] наука, чем раздел математики. Чтобы математическая теорема была элегантной, достаточно, чтобы она была красивой и правдивой. Но статистика подчиняется дополнительному стандарту, навязанному наукой⁶. Модель данных, независимо от того, насколько она элегантна или правильно получена, должна быть отброшена или пересмотрена, если она не соответствует данным, или, когда найдены новые или более качественные данные и они «не работают», не соответствуют модели. Томас Хакслиⁱ назвал это

Великая научная трагедия – уничтожение прекрасной гипотезы уродливым фактом. «Биогенез и Абиогенез»^j (1870).

Статистика живет на эмпирической, а не на теоретической стороне науки. Статистики име-

ют дело с данными, а также строят и исследуют модели для этих данных. Когда модель и данные расходятся, это скорее признак прогресса, чем неудачи, на чем мы часто концентрируем наше внимание. Как напоминает нам Айзек Азимов^k

Самая захватывающая фраза, которую можно услышать в науке, предвещающая новые открытия, это не «Эврика!», а «это смешно...»⁷.

Эта постоянная ссылка на понимание мира – один из моментов, который делает курс по статистике более трудным для преподавания, чем курс по математике на сопоставимом уровне технической сложности. Наша настойчивость состоит в том, что студенты должны писать одно или два предложения, связывая свои расчеты с представлениями о мире, что иногда удивляет и расстраивает их, так как они думали, что записались на уроки математики (и, конечно, радует других, которые боялись, что это так и есть).

Большую часть времени мы не можем знать правду, которую надеемся описать или смоделировать. Какова истинная доля взрослых, которые курят в США, имеют детей или курят в присутствии своих детей? Лучшее, на что мы можем надеяться, – это данные, которые представляет население, что и позволяет сделать разумные выводы.

Когда цель регулирует практику, предоставляя цель, к которой стремится практика, философы называют её *регулирующим идеалом*. Тот факт, что мы никогда не узнаем правду, не должен удерживать статистиков от использования её в качестве маяка для принятия решений, когда мы стремимся к лучшему и лучшему пониманию. Нельзя достичь цели, служащей регулирующим идеалом. Например, идеальный результат в гольфе – 18.

5. ЧЕСТНОСТЬ СТАТИСТИКОВ

Можно утверждать, что, признавая нашу неопределенность и определяя её количественно, статистики в некотором смысле более честны в своих высказываниях о мире, чем другие, которые делают абсолютные заявления. В конце концов, мы, статистики, не утверждаем, что знаем то, чего не можем знать. Вместо этого, например, мы предлагаем интервал вероятных значений для неизвестного параметра. Не удовлетворенные этим, мы тратим больше усилий на то, чтобы точно описать, насколько мы не уверены, что даже этот интервал охватывает истинное значение и то, что мы должны предположить о неизвестных и непознаваемых особенностях мира, чтобы эти оценки были правильными⁸.

Каков же тогда источник представления «чёртова ложь»? Статистики, очевидно, очень

стараятся быть честными и с готовностью признают свою неопределенность. Лжецы обычно уверенно лгут, стремясь верить. И когда вывод статистика оказывается неверным, ошибка не рассматривается как преднамеренный обман. Другая случайная выборка может дать другой результат, но статистику не обвиняют в неудаче сбора или нарушении этических норм при анализе данных.

Я думаю, что проблема заключается в другом аспекте неопределенности в статистическом анализе; тот факт, что статистика, как наука, не является алгоритмической или детерминированной. Проблема не в том, что другая выборка может дать другой ответ, а в том, что другой статистик, работающий с той же выборкой, может дать другой ответ⁹.

Например, мы читаем у американского юмориста Эвана Эсара (Evan Esar; 1899-1995) в его комическом словаре (Esar, 1943), что

Статистика [это] единственная наука, которая позволяет различным экспертам, использующим одни и те же цифры, делать разные выводы.

Кажется, что мы не даём понять другим – и особенно не нашим студентам, – что хороший статистический анализ включает в себя суждения (интерпретации), и мы не учим наших студентов, как делать такие суждения.

6. СУЖДЕНИЕ

Понимание того, что статистический анализ требует принятия решений, не ново. По иронии судьбы, это можно найти в комментариях статистиков, которые разработали статистические методы, которые чаще всего способствуют неправильному пониманию статистики как алгоритмического приложения математики. Именно в области проверки гипотез мы часто видим, что люди слепо применяют статистические методы, надеясь на статистически значимый [уровень достоверности] $P < 0,05$, но пренебрегая своим суждением. (Действительно, в некоторых дисциплинах суждение в таких анализах не рекомендуется).

Легко найти блок-схемы для «процесса проверки гипотез», иногда называемого «научный метод». Но создатели стандартных [вычислительных] методов были явными сторонниками суждений (интерпретаций). Эгон Пирсон¹, работая с Ежи Нейманом^m, предложил частотный статистический вывод и в то же время писал

Мы оставили в нашей математической модели пробел для осуществления более интуитивного процесса личного принятия

решений (цитируется [Abelson, 1995, p. 395]).

По иронии судьбы, процедуры логического вывода, разработанные Нейманом и Пирсоном, относятся к числу статистических методов, которые, скорее всего, следует рассматривать и преподавать как механические манипуляции с данными, которые приводят к автоматическому решению принять или отклонить гипотезу.

Сэр Рональд Фишерⁿ был основателем планирования эксперимента и дисперсионного анализа, и ему, обычно, приписывают предложение величины 0,05 в качестве приемлемого значения P для статистической значимости, – возможно, окончательного произвольного решения, принятого большинством пользователей статистики¹⁰. Тем не менее, он тоже считал принятие решений важным в достижении выводов о данных. Абельсон^o отмечает, что

Сэр Рональд Фишер обвинил Неймана и Пирсона в том, что они высказали чрезмерные механические рекомендации, подчеркнув, что экспериментирование – это непрерывный процесс, требующий сообщества свободных умов, принимающих свои собственные решения на основе общей информации (Abelson, 1995, p. 2).

Важной особенностью суждений, необходимых для статистического анализа, является то, что они основаны на наших знаниях о мире. «Сообщество свободных умов, принимающих... решения на основе общей информации», на которое ссылается Фишер, делится информацией об эволюционирующем понимании того, как устроен мир. Статистическое принятие решений должно основываться на таком понимании. Вот почему статистика является обязательным курсом во многих дисциплинах; «общая информация» научной дисциплины обеспечивает основу для принятия решений, что необходимо делать на основе анализа данных.

К счастью для нашего преподавания, многие из этих суждений проявляются в начале первого курса статистики. Например:

- следует ли считать гистограмму перекошенной или симметричной,
- является ли отдельная точка выбросом,
- является ли второй максимум реальным режимом (и данные, возможно, не однородны)^p,
- показывает ли график рассеяния достаточно прямую картину для соответствия корреляции и регрессии,
- показывает ли график нормальной вероятности или гистограмму невязок распределение, достаточно близкое к нормальному для вывода на основе t -критерия (или раз-

- мер выборки достаточно велик, чтобы это предположение не имело значения),
- могут ли данные наблюдений двух групп правильно считаться парными (или, альтернативно, могут ли группы считаться независимыми),
 - если мы отклонили нулевую гипотезу, достаточно ли велика величина эффекта, чтобы иметь смысл (и важность понимания разницы),
 - когда нужно повторно выразить переменную и какое повторное выражение выбрать,
 - является ли наша выборка достаточно представительной (или, наоборот, можем ли мы думать о важных источниках предвзятости),
 - являются ли случаи в наших данных взаимно независимыми (как требуется почти всеми нашими процедурами вывода), и
 - являются ли остатки регрессии действительно гомоскедастичными⁴?

По мере того, как ученики становятся более искусными, мы можем обсуждать предвзятость в сборе данных, неудачи независимости, скрывающиеся переменные и необходимость ясности в точном определении того, что и какую информацию мы можем получить. Каждый из них также обычно включает суждения, которые содержат знания о реальном мире и о характере данных.

7. BSR

Иногда статистика используется для лжи. Необходимость принятия решений открывает дверь к неэтичному искажению результатов, смещенному сбору данных и частичному представлению результатов или манипулированию результатами с целью ввести в заблуждение. Американская статистическая ассоциация публикует этические руководящие принципы для статистической практики¹¹, в которых содержится призыв (среди прочего):

Избегание любой тенденции подгонять статистическую работу к предопределенным результатам.

А также

Разоблачение нечестного или некомпетентного использования статистики.

Некоторые люди используют статистику как часть преднамеренной лжи. Они знают, что выбрали предвзятую выборку или лучшие результаты с низкими значениями *P*. Поскольку они знают, что результаты ложные, но, тем не менее, пытаются убедить других, они, по словам Франкфурта⁸,

«...мошенники и обманщики, которые пытаются, как они говорят, манипулировать... мнениями. Поэтому прежде всего

их волнует, насколько эффективно то, что они говорят, для выполнения этих манипуляций. Соответственно, они более или менее равнодушны к тому: верно ли то, что они говорят, или ложно» (Frankfurt, 2005).

Использование статистики – это то, что Франкфурт определил, как *чушь собачью*.

Тем не менее, другие науки (в том числе и социальные) могут и были искажены для обоснования ложных утверждений (если только мы не будем называть это *наглой ложью*). Мы слышим, как случайный эксперт, получивший диплом, отрицает эволюцию, заявляет, что ВИЧ не вызывает СПИД, не поддерживает регрессивную терапию, и утверждает, что рабство было полезно для рабов. Эти взгляды меньшинства не заставляют большинство ученых или информированных мирян называть целые дисциплины *наглой ложью*. Даже когда эксперты в других дисциплинах, которые не выходят за рамки основного течения, не соглашаются публично, мы не считаем их различные суждения *ложью*.

8. МАГИЯ

Почему тогда статистический анализ считается еще хуже, чем *наглая* (чёртова) *ложь*? Я думаю, что проблема заключается в следующем. Для многих математика, космология и ядерная физика столь же загадочны, как и статистика, и выходят за рамки обычного понимания. Но, честно говоря, большинству людей все равно, что они (ученые) делают; воображаемые миры, или слишком далеко, или слишком малы, чтобы казаться важными. А вот выводы статистики непосредственно касаются нашего мира. В редакционной статье в журнале «Лансет – Lancet» от 2 января 1937 г. читаем:

«Статистика... имеет тенденцию вызывать сильную эмоциональную реакцию в нематематических умах. Это потому, что статистику применяют к проблемам, в которых мы заинтересованы, методами, которые мы не понимаем».

Это ставит статистику в область Третьего Закона Артура Кларка (Arthur Clarke's Third Law)¹, который говорит нам, что

Любая достаточно развитая технология неотличима от магии.

«Черты будущего».

Статистика – это передовая технология, относящаяся к нашему повседневному миру, поэтому она использует магию, которую волшебники мифа и вымысла применяют для изменения повседневных обстоятельств.

Правила для этого вида магии хорошо известны в традициях и литературе:

- Вы должны сотворить заклинание точно так, как указано в неизвестных знаниях.
- Вам не нужно понимать заклинание (которое может быть на тайном языке). В самом деле, нет никакого поощрения или преимущества для этого – это должно быть неясным.
- Тем не менее, если вы поймете что-то неправильно, то это просто не сработает или, как ученик колдуна, вы можете получить результат, который вы не хотели.

Важно отметить, что здесь нет места для принятия решений. Однако сила магии дает определенную степень власти тем, кто может (или, кажется, способен) контролировать её (и кто может затем применить запрет на принятие решений, чтобы заставить замолчать сомневающийся).

Конечно, это противоположность научного мышления, которое включает в себя принятие решений. В этом и заключается конфликт. Есть, например, те, кто хочет просто произнести заклинания, сделать соответствующие движения мыши, а не палочки, достичь статистической значимости и заявить об авторитете магии в качестве доказательства того, что их выводы должны быть правильными. Они думают – они надеются, – что наука продвигается вперед, проводя данные через статистические программы и отсеивая для значений P менее 0,05. Принимая этот подход, они поощряют страх и неуважение к дисциплине. Это не что иное, как сценическая магия, и те, кто может видеть сквозь их слабые рассуждения, могут также обвинять свои инструменты. Как только вы видите проволоку, поддерживающую парящую даму, естественно отклонить весь магический акт как мошенничество.

9. ...И СТАТИСТИКА

Некоторые потребители статистики хотят верить в волшебство статистики. Они рассматривают статистику как разновидность математики и обижаются, узнав, что принятие решений (интерпретация) является её составной частью. В конце концов, по их мнению, *теорема Пифагора* не обсуждается. Сам факт, что суждение является частью статистического анализа, открывает вопрос о честности статистики.

Суждения о том, как собирать данные, о качестве данных и о том, что можно честно сделать из этих данных, являются важными частями статистического анализа. И все же может показаться, что тем, кто занимается статистикой, стыдно признать, что они принимали ре-

шения в ходе своей работы. Скорее, им должно быть стыдно скрывать это от открытого взгляда и рассмотрения.

Многие пользователи статистики предпочитают представлять свои выводы как абсолютно объективные и свободные от суждений. Существует соблазн одеть мантию математической объективности, полагая, что полученные выводы почему-то менее открыты для спора. Когда необходимы суждения (интерпретации), существует традиция во многих областях знания писать «пассивным голосом – (writing in passive voice)». Решения о дизайне исследования «приняты»; анализы «были выполнены» с использованием методов, которые «было решено использовать». Трудно найти ученых и статистиков, принимающих решение, которых бы не обвиняли в фальсификации с «помощью заклинаний – with the spells». Но статистика без принятия решений – это магия, а не логика.

10. МОДЕЛИРОВАНИЕ МИРА КАК РЕГУЛЯТИВНОГО ИДЕАЛА

В принятии статистических решений нам требуется беспристрастное стремление к честным суждениям. Наш руководящий принцип должен заключаться в том, что мы ищем правду о мире. Решения, основанные на этом принципе, могут помочь нам сформулировать вопросы, задаваемые статистическим анализом. Пока мы не знаем, почему мы проводим анализ, трудно понять, как честно искать соответствующие ответы.

Уильям Хантер^u советует:

Прежде всего, наиболее важные вопросы, которые должен задать статистик: какова цель этого исследования?

«Практическая статистика: реальный мир – идея, время которой пришло».

Если мы знаем, в чем заключаются проблемы, это помогает нам делать здравые суждения, и мы должны делать это беспристрастно.

Что мы хотим от принятия решений, чтобы они следовали руководящему этическому принципу: мы ищем правду о мире. Конечно, честность в этом контексте не означает, что наши исследования обязательно скажут правду, и мы даже не поверим, что они это сделают. Скорее, мы прилагаем беспристрастные усилия, чтобы узнать правду. Эти усилия могут включать поддержание альтернативных и даже противоречивых моделей мира, если все они хорошо учитывают имеющиеся у нас данные и факты, которые мы знаем. Лучший результат часто получается из дарвиновской конкуренции среди альтернативных моделей. (Джон Уокер^v предложил назвать это соревнование «выжива-

нием наиболее приспособленных»). По мере проведения анализа каждая модель порождает новые вопросы, прогнозы и остатки, которые нужно исследовать, в результате чего одни линии рассуждений процветают, а другие вымирают. Это также может привести к тому, что некоторые люди, которые не понимают, как на самом деле работает статистика, считают, что статистики играют в правду, а не ищут её. Конечно, они думают, что нет смысла поддерживать две конфликтующие модели, когда они не могут быть правдой. Статистики (которые понимают ценность рандомизированных плацебо-контролируемых экспериментов) понимают, что знание того, что (по крайней мере) одна модель является ложной, не то же самое, что знать, какая именно.

В частности, статистические рассуждения скептически. Структура логического вывода требует, чтобы мы отвергли нулевую гипотезу, если мы хотим добиться прогресса. Поэтому мы часто работаем с моделями, которые могут быть ложными, и даже с теми, которые, как мы надеемся, являются ложными. Мы не можем просто заявить, что считаем правдой, и собрать доказательства, подтверждающие это. Это просто здравые научные рассуждения. Это отмечал (почти 400 лет назад) и основатель современного научного мышления сэр Фрэнсис Бэкон^W:

И потому правильно ответил тот, который, когда ему показали выставленные в храме изображения спасшихся от кораблекрушения принесением обета и при этом добивались ответа, признает ли теперь он могущество богов, спросил в свою очередь:

«А где изображения тех, кто погиб, после того как принес обет?». Таково основание почти всех суеверий – в астрологии, в сновидениях, в поверьях, в предсказаниях и тому подобном.

Novum Organum
(Урбахский перевод)^{12, X}

Древние корни этой борьбы не облегчают обучение. Однако, это позволяет нам указать, что, хотя нашей дисциплине едва ли исполнилось столетие в её основных частях, основополагающая философия открытия истины о мире посредством наблюдения за ней и сложность необходимых суждений намного старше.

11. ОБУЧЕНИЕ

Наша основная надежда на изменение представлений о статистике, как о наглой (чёртовой) лжи, – это образование. Если мы будем преподавать статистику как механическую путаницу магических методов, наши ученики са-

ми сделают вывод, что это чёртова ложь. Но мы можем поступить лучше:

1. Мы должны сказать нашим студентам, что для использования статистики они должны научиться принимать решения и что нет никакого гарантированного способа достичь истины. Это будет огорчать тех, кто надеялся просто вставить новые цифры из упражнений и экзаменационных вопросов в алгоритмы и формулы, найденные в малых рамках учебника.
2. Мы должны посоветовать студентам знать мотивирующую причину анализа, потому что это поможет им в принятии этих решений. Они должны знать, кто (или какие) случаи содержатся в данных, что было измерено или записано о них (и в каких единицах) и когда это было сделано. Даже определения, которые кажутся разумными, должны быть подвергнуты сомнению.
3. Мы должны учить студентов тому, что руководящим принципом при составлении статистических суждений является поиск истины о мире. Столкнувшись с призывом к принятию решений, мы делаем выбор, который наилучшим образом поддерживает наши усилия по моделированию или пониманию мира таким, какой он есть. Там, где этот выбор не ясен, мы делаем честную попытку сделать лучший выбор. Можно использовать альтернативные или противоречивые модели, если нет данных, позволяющих нам выбирать между ними.
4. Мы должны научить студентов не спешить с выводами, экстраполировать и предлагать объяснения ассоциаций, которые предполагают причинность. И мы должны научить их скептически относиться к любым статистическим отчетам, которые не соответствуют этим высоким стандартам.
5. Самое главное, мы должны представить весь предмет как поиск понимания мира, когда у нас есть данные, чтобы создать у других принципов основу, на которую можно было бы опереться.

Насколько хорошо мы делаем это сейчас?

Большинство современных определений «статистики» трактуют статистику способами, совместимыми с принципом (5). Вот несколько примеров:

- [Статистика это] искусство и наука планирования исследований и анализа данных. Конечной целью является перевод данных в знания и понимание окружающего нас мира. Короче говоря, статистика – это искусство и наука изучения данных» (Agresti, Franklin, 2007).

- Статистика – это способ мышления, а также набор инструментов и методов, призванных помочь нам понять мир (Deveaux, Velleman, Bock, 2009).
- Статистика – это наука о данных (Moore, 2007).
- Статистика гораздо больше относится к научному методу, чем к чему-либо еще, определяя вопросы исследования; планирования исследований; организацию, обобщение и анализ данных; интерпретацию результатов; и принятие решений (Rumsey, 2003).
- Статистика – это набор процедур и принципов для сбора данных и анализа информации, помогающий нам принимать решения в условиях неопределенности (Utts, Neckard, 2007).

Сохраняют ли эти определения – и мы, как учителя, – цели изучения мира и принятие обоснованных решений на основе данных, которые являются центральными для дискуссий на протяжении всего курса? Существует неумолимая тенденция сосредотачиваться на определениях и методах, и уделять меньше внимания более широкой цели. Это может быть особенно сложно и опасно, когда мы переходим к обсуждению выводов. Методы – это техническая часть статистики, принятие решений следует за многими [методическими] шагами; всего этого слишком много для большинства наших студентов, чтобы помнить начало (основу) – выводы о мире делаются почти по сценарию.

Мы все напоминаем студентам, чтобы они наносили на график данные и остатки, но когда курс становится техническим и концептуально сложным, трудно посмотреть на него глобально. Если мы рассмотрим принятие решений, нам нужно будет обсудить, как бороться с потерей независимости [наблюдений], ненормально-распределенными остатками или случайными выбросами. Проще обойтись без всех этих соображений и вместо этого поговорить о формуле стандартной ошибки или о том, сколько у нас степеней свободы.

Кроме того, что немаловажно, трудно писать и оценивать упражнения, связанные с реальными проблемами, и требовать выводов, касающихся этих проблем. Как заманчиво просто проверить, является ли P -значение правильным или было ли отклонено значение NULL в соответствии с ключом. Издатели не хотят публиковать приложения [с исходными данными], необходимыми для длинных ответов в конце книги, поэтому у нас могут быть только крат-

кие числовые решения без реальных объяснений.

Тем не менее, мы можем предложить студентам такие упражнения, и мы можем представить упражнения с простыми суждениями, которые я перечислил ранее. Методы вывода требуют предположений о данных и о мире, который генерировал эти данные. О некоторых из этих допущений, таких как независимость и рандомизация, мы судим по пониманию того, как были собраны данные, кто и что измерял, а затем используя наш здравый смысл и знания о том, как устроен мир. Даже когда у нас нет данных, достаточно просто подумать об обстоятельствах, чтобы решить, искажено ли распределение или нелинейные отношения. Другие предположения, такие как симметрия, нормальность и линейность, мы оцениваем, отображая данные обсуждая то, что видим.

Мы не должны делать это просто еще одним шагом в сценарии, которому ученики учатся следовать. Мы должны обсудить этику честного поиска истины о мире в рамках представления статистики студентам. Мы должны напомнить нашим студентам, что, поскольку статистика касается моделирования и понимания мира с помощью данных, мы обязаны использовать нашу цель в адекватных моделях в качестве руководящего принципа при вынесении суждений.

Когда мы концентрируем вводный курс статистики на поиске правды о мире, мы отвечаем на вопрос, который возникает у большинства наших студентов в самом начале: «Почему я выбрал этот курс?» Наш традиционный ответ: «Потому что это необходимо», не очень хорош. Ответ, который поддерживается современной статистической образовательной практикой: «Чтобы помочь вам научиться понимать мир на основе данных» более убедительный и интригующий. Ответ, который я предлагаю, «Потому что вы должны принимать решения, а они зависят от других знаний, которые вы получаете в процессе обучения», интегрирует весь курс обучения студентов.

Более того, предоставляя центральную тему, которая действует на протяжении всего вводного курса, мы можем уберечь её от того, чтобы не стать изумительным рассказом, который переходит от графики к случайности, к выборке, к вероятности и к выводам без особого объединительного сюжета. Вместо этого мы можем представить последовательную историю, в которой каждый новый метод добавляет понимание и соответствует другим.

12. ПРАВДА, ТВЕН И УЧЕНИЕ

Давайте приведем цитату, которая открыла эту дискуссию в её более полном контексте. Вот что сказал Марк Твен:

Исходя из вышеизложенного, я пришел к выводу, что в течение этих тридцати шести лет я неуклонно замедлялся, но я чувствую, что в моей статистике есть недостаток: 3000 слов весной 1868 года, когда я работал семь, восемь или девять часов за один присест, имеет мало или вообще не имеет преимуществ перед сегодняшними занятиями (занимая половину времени и производя половину результата). Цифры часто вводят меня в заблуждение, особенно когда я сам их использую; в этом случае справедливо и убедительно звучит высказывание, приписываемое Дизраэли: «Существует три вида лжи: ложь, наглая ложь и статистика» (Twain, 1924, p. 246).

Твен занимается некоторым простым статистическим анализом. Он отмечает, что, хотя он производит меньший объем письменной работы за день, фактически он поддерживает свою производительность около 375 слов в час. Он просто работает меньше.

Он отмечает, что у его первоначальной статистики есть недостаток, потому что они не учитывают, как долго он теперь работает каждый день. И это заставляет его задуматься над тем, как статистика может ввести в заблуждение.

Но что Твен имел в виду под ложью, наглой ложью? Просто шутка? Чтобы понять его, нам нужно поразмышлять. Как писатель, он тщательно подбирал слова. Это делает его выбор «обманчиво» интригующим, потому что слово имеет несколько альтернативных значений.

Если бы он имел в виду значение «обмануться», чтобы «приятно провести время», он предложил бы нам представить, как он «лепит» ложь из цифр для собственного удовольствия. Это кажется сомнительным даже для циничного «возрастного» Твена.

Скорее всего, он имел в виду «вводить в заблуждение или дезориентировать». Действительно, остальная часть цитаты подсказывает именно это направление. Если бы Твен просто сказал, что «цифры часто обманывают меня; я согласен с Дизраэли, когда он говорит...», это было бы естественной интерпретацией. Но что нам делать с квалификацией Твена «особенно, когда я сам их использую» в сочетании с его демонстрацией понимания правды о его [писательской] продуктивности путем вычисления

среднего значения? Он мягко упрекает статистические данные за то, что они вводят его в заблуждение, и признает, что это часто случается с ним, когда он сам «их использует». Нет никакого намерения обмануть; это, скорее, легкое «проскальзывание» в неправильном истолковании, которое проявляется [ловится] при более тщательном анализе данных. Это не несет обвинения в преднамеренной фальсификации или обмане, что, как представляется, цитирующие эту фразу, считают пониманием Твена. Скорее, это предостережение: внимательно смотрите на ваши данные, потому что (не будучи простыми фактами без интерпретации) они могут ввести вас в заблуждение¹³.

Я хотел бы думать, что Твен также имел в виду и третье значение слова «обманывать»: «завоевывать и удерживать чье-то внимание, интерес или преданность; очаровывать или отвлекать». В этой интерпретации Твен мог бы иметь в виду, что, когда ему дают возможность работать («расставлять», «упорядочивать») с числами для себя, он находит это увлекательным, потому что часто обнаруживает истины, которые, поначалу могли быть скрыты в данных или в упрощенной статистике, обобщая их.

Учитывая широкий кругозор [интеллект] Твена, его острый цинизм в отношении человечества и его блестящее владение английским языком, это может быть одной из интерпретаций обсуждаемой нами фразы. И если это так, то это настоящий призыв к нам научить студентов делать здравые, этические суждения, когда они интерпретируют свои данные так, чтобы тоже могли быть обмануты статистикой⁵.

Конечные заметки

¹ <http://www1c.btwebworld.com/quote-unquote/>.

² «Единственное, что необходимо для победы зла, – это чтобы хорошие люди ничего не делали»; приписывается Edmund Burke², но без твердой уверенности.

³ <http://www.york.ac.uk/depts/maths/histstat/lies.htm>.

⁴ Некоторые [исследователи] предполагают, что пунктуация показывает, что лорд Кортни (Lord Courtney) перечислил только две альтернативы: ложь – наглую ложь и статистику, но остальная часть фразы говорит об обратном.

⁵ Лорд Кортни, будучи секретарем казначейства в 1882 г., был одним из тех «гладстонцев», против которых выступал Бальфур, но к 1892 г. он отошел от либералов и достиг весьма высоких партийных позиций.

⁶ В этом эссе термин «наука» следует понимать, как включающий в себя социальные (гумани-

тарные) науки, а также любую дисциплину, развивающуюся научным методом.

⁷ Эта цитата широко и правдоподобно приписывается Азимову, но, как я обнаружил, не имеет достоверных источников.

⁸ Наша одержимость нашим невежеством на этом не заканчивается. Документы, которые обсуждают, настолько хороши или настолько плохи, что мы можем оценить размер доверительного интервала.

⁹ Один пример того, как два статистика, каждый со здоровой (но слегка различающейся) интерпретацией, могут прийти к противоположным выводам из одних и тех же данных, известен как «Парадокс Лорда» (Lord, 1967), «Парадокс в интерпретации групповых сравнений», *Psych. Bull.* 68, 304-305. Хорошее объяснение см. у (Wainer, Brown, 2004). Два статистических парадокса в интерпретации различий в группах: проиллюстрировано на основании данных о поступлении в медицинскую школу и лицензировании (*American Statistician*, 58, 117-123, 2004).

¹⁰ Говард Уэйнер отмечает, что Фишер выбрал 0,05, чтобы удовлетворить небольшие предварительные исследования в контексте текущих исследований. Его идея заключалась в том, что начальные исследования, как правило, невелики, поэтому наличие слишком строгого критерия усложнит поиск. Стоимость исправления ошибки I-го типа невелика (обычно это происходит при первой репликации), тогда как стоимость повторного обнаружения пропущенного эффекта намного выше. Отсюда и щедрый критерий один на двадцать (Wainer, Robinson, 2003).

¹¹ <http://jse.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics>.

¹² Марк Твен предлагает аналогичное понимание в своих записных книжках (*Mark Twain's Notebook*):

Разумен ли человек? Нет, он думает, размышляет, размышляет, но не рассуждает... То есть в двух вещах, которые являются своеобразной областью сердца, а не разума, – политике и религии. Он не хочет знать другую сторону. Ему нужны аргументы и статистика для его собственной стороны, и ничего более.

¹³ Как и у Бальфура, проблема Твена в меняющемся общем знаменателе.

СПИСОК ЛИТЕРАТУРЫ

Abelson, R.P. (1995), *Statistics as principled argument*, Hillsdale, N.J., L. Erlbaum Associates.

American Statistical Association, *Ethical Guidelines for Statistical Practice*, <http://jse.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics>.

Agresti, A., and Franklin, C. (2007), *Statistics: the Art and Science of Learning from Data*, Upper Saddle River, Pearson Prentice Hall.

Bacon, Sir Francis, (1610) *The Novum Organum; with other parts of, The great instauration* translated and edited by Peter Urbach and John Gibson (1994), Chicago, Open Court.

Balfour, Arthur James, (1892) "Mr. Balfour's Reply to Professor Munro," *Manchester Guardian*, Wednesday, June 29, 1892 P. 5.

Clarke, Arthur, (1973) *Profiles of the Future*. New York, Harper & Row.

Courtney, Leonard Henry, (1895) "To my Fellow Disciples at Saratoga Springs". *The National Review* (London) 26, 21–26.

Deveaux, Velleman, Bock, (2009) *Intro Stats. Third Edition*. Boston, Addison Wesley.

Esar, Evan, (1943) *Esar's Comic Dictionary*, Harvest House.

Frankfurt, Harry G., (2005), *On Bullshit*, Princeton, NJ, Princeton University Press.

Guy, William A, (1865) "On the Original and Acquired Meaning of the term 'Statistics,' and on the Proper Functions of a Statistical Society: also on the Question whether there be a Science of Statistics and, if so, what is its Relation to Political Economy and 'Social Science'". *Journal of the Statistical Society of London*, 28, No. 4, Dec, pp. 478-493.

Hunter, William G. (1981), "The Practice of Statistics: The Real World is an Idea Whose Time Has Come". *The American Statistician*, 35: 2, 72-76.

Huxley, Thomas Henry (1959), "Biogenesis and Abiogenesis" in *Collected Essays*. New York, Harper.

Lord, F.M. (1967), "A Paradox in the Interpretation of Group Comparisons". *Psych. Bull.* 68, 304-305.

Moore, D.S. (2007), *The Basic Practice of Statistics, Fourth Edition*. New York, W.H. Freeman.

Pearson, E.S. (1962). "Some Thoughts on Statistical Inference". *Annals of Mathematical Statistics*, 33, 394-403.

Rumsey, D. (2003), *Statistics for Dummies*. Indianapolis, Wiley Publ.

Twain, Mark, (1924), *Autobiography*, A.B. Paine (ed.). New York and London: Harper Brothers. Vol. I.

Mark Twain's Notebook, A. B. Paine (ed.). New York, Cooper Square Publishers (1972).

Utts, J.M. and Heckard, R.F. (2007), *Mind on Statistics*. Belmont, Duxbury, Thomson Brooks/Cole.

Wainer, H. and Brown, L.M. Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *The American Statistician*, 58, 117-123, 2004.

Wainer, H. and Robinson, D. (2003). Shaping Up the Practice of Null Hypothesis Significance Testing. *Educational Researcher*, 32 (7), 22-30.