
ПЕРЕВОДЫ

П.Ф. Веллеман (*Итака, США*)

Л. Уилкинсон (*Чикаго, США*)

ТИПОЛОГИИ НОМИНАЛЬНЫХ, ОРДИНАЛЬНЫХ, ИНТЕРВАЛЬНЫХ И ОТНОСИТЕЛЬНЫХ ШКАЛ ВВОДЯТ В ЗАБЛУЖДЕНИЕ¹

Разработанная психофизиком С.С. Стивенсом типология измерительных шкал доминирует в социальной статистике на протяжении без малого 50 лет. В течение этого времени данная типология неизменно оставалась предметом активной полемики. Недавно типология шкал Стивенса пережила свой ренессанс, связанный с разработкой компьютерных статистических пакетов. Однако ныне терминологию Стивенса используют без учета как классической критики, сопровождавшей ее появление, так и достижений в области анализа данных, которые появились в последние десятилетия.

Ключевые слова: анализ данных, типы данных, измерительные шкалы, шкалирование.

Пол Ф. Веллеман – доцент департамента экономики и социальной статистики Корнелльского университета (Итака, штат Нью-Йорк), президент Data Description Inc.
Лиленд Уилкинсон – адъюнкт-профессор статистики Северо-западного университета, адъюнкт-профессор компьютерных наук Иллинойского университета в Чикаго, исполнительный вице-президент SYSTAT Software.

Перевод: **Всеволод Игоревич Хоменко** – аспирант Российского государственного гуманитарного университета, факультет социологии. E-mail: vskhomenko@gmail.com.

¹ The American Statistician. 1993. Vol. 47. No. 1 (Feb.). Статья опубликована в разделе «Комментарии», где обсуждаются вопросы статистического образования и статистической практики. Перевод выполнен в рамках конкурса переводов, объявленного журналом «Социология: 4М» в 2011 г.

В начале 1940-х гарвардский психолог С.С. Стивенс ввел термины *номинальная шкала*, *ординальная шкала*, *интервальная шкала* и *шкала отношений*, чтобы описать иерархию используемых в психологии измерительных шкал, и классифицировал статистические процедуры с точки зрения их «допустимости» для того или иного типа шкал. Затем эта таксономия вошла в некоторые важные учебники статистики и таким образом оказала влияние на статистическое мышление целого поколения. Несмотря на критику со стороны статистиков, авторы некоторых учебников до сих пор обращаются к стивенсовским категориям.

В последнее время, вследствие возникшего интереса к компьютерным программам, автоматизирующим статистический анализ, труды Стивенса вновь удостоились внимания. Компьютерные программы, разработанные для оказания помощи в выборе методов анализа данных, основаны на стивенсовских предписаниях. Более того, указанные предписания используются даже в некоторых программах общего назначения, структурируя взаимодействие с пользователем.

К сожалению, применение стивенсовских категорий при выборе или для рекомендации тех или иных методов статистического анализа неуместно и зачастую приводит к ошибкам. Категории Стивенса не описывают те свойства реальных данных, которые важны для обеспечения хорошего статистического анализа. Кроме того, эти категории не представляют собой классификационную схему, адекватную современным методам статистического анализа. Частично эти факты были обнаружены еще во времена работы самого Стивенса, частично – выявились в ходе развития новой философии и новых методов анализа данных.

В настоящей статье представлен обзор стивенсовской таксономии, приводятся определения терминов, которые использовались многими без экспликации их точного смысла. Использование подобных терминов в статистике и в различных ее приложениях анализируется с учетом классической критики работы Стивенса.

По тексту статьи приводятся ссылки, рассчитанные на заинтересованных читателей. Кроме того, описываются некоторые неудачи, порой возникающие при применении стивенсовской таксономии для классификации данных, а также исследуется природа этих неудач. Также рассматривается возможность классификации современных статистических методов в зависимости от подходящих им типов данных. Наконец, анализируются те идеи Стивенса, которые все же могут оказаться полезны для современного компьютеризированного статистического анализа.

Стивенсовская типология данных

В своей основополагающей работе «О теории измерительных шкал» [1] Стивенс представил иерархию шкал данных, основанную на инвариантности их значений для различных классов преобразований. Измерительные шкалы, не теряющие смысла при широком круге преобразований, в определенном отношении передают меньше информации, чем те, которые сохраняют смысл при более узком круге преобразований. Допустим, например, что шкала s использована для присвоения действительных чисел в \mathfrak{R} (числовой системе) элементам множества P так, что для всех i и j множества P верно $s(i) > s(j)$, если i «предпочтительнее» чем j . Если «предпочтительнее чем» обозначается символом \succ , то:

$$P^s \rightarrow \mathfrak{R} \text{ так, что} \\ i \succ j \leftrightarrow s(i) > s(j) \text{ для всех } i, j \in P. \quad (1)$$

Подобные шкалы Стивенс назвал *ординальными* при условии, что любое преобразование шкальных значений, сохраняющее их числовой порядок, создает новую шкалу, задающую то же взаимно однозначное отношение между сравнениями объектов (с использованием символа \succ) и сравнениями шкальных значений (с использованием символа $>$).

Стивенс использовал термин *допустимость* для описания набора преобразований, не разрушающих отношение порядка

для числового отображения в (1). А именно, преобразование f допустимо для ординальной шкалы, если и только если:

$$s(i) > s(j) \Rightarrow f[s(i)] > f[s(j)]. \quad (2)$$

Любое монотонное преобразование значений $s(i)$ и $s(j)$ допустимо для ординальных шкальных данных. Таким образом, мы можем брать логарифм, извлекать квадратный корень из значений (если они не отрицательные) или осуществлять линейное преобразование, складывая значение с константой или умножая его на (положительную) константу.

Стивенс разработал сходные аргументы и для трех других типов шкал. Интервальные шкалы вместо порядка ($>$) задействуют разность ($-$), так что набор допустимых преобразований для интервальных шкал сохраняет отношения разности. Так, преобразование f допустимо для интервальных шкал в том случае, если существует константа c , при которой оказывается верным следующее выражение:

$$s(i) - s(j) \Rightarrow c\{f[s(i)] - f[s(j)]\}. \quad (3)$$

Таким образом, для интервальных шкал допустимыми являются линейные преобразования, при которых мы прибавляем одну и ту же константу к каждому значению и/или умножаем каждое значение на константу, но мы не можем взять логарифм каждого из исходных значений. Иными словами, в этом случае мы имеем более узкий набор классов допустимых преобразований, чем с ординальными данными, а это означает, что в определенном смысле интервальные данные несут больше информации.

Шкалы отношений сохраняют отношения между числовыми значениями так, что допустимые преобразования удовлетворяют следующему:

$$\frac{s(i)}{s(j)} \Rightarrow \frac{cf[s(i)]}{f[s(j)]} \quad (4)$$

для некоторой константы c .

Таким образом, данные, полученные по шкале отношений, допустимо умножать на константу, но брать логарифм или прибавлять константу нельзя. Данные шкалы отношений имеют абсолютный ноль, который не может быть изменен.

Номинальные шкалы находятся на противоположном конце иерархии. Они даже не требуют присвоения числовых значений – достаточно присвоения некоторых уникальных идентификаторов/означающих (чисел, букв, цветов). Они инвариантны при любых преобразованиях, сохраняющих отношения между индивидами и их идентификаторами. Таким образом, выполнение практически любых операций со значениями допустимо, пока мы не совмещаем или смешиваем тождественные значения. Когда данные являются числовыми, эти операции могут включать любые функции, с помощью которых исходное множество чисел взаимно однозначно отображается в новое множество. Если используемые значения нечисловые, то допустимы операции по реконфигурации значений. Разумеется, только самые малоинформативные данные могут без ущерба претерпеть подобные произвольные преобразования.

Теоретики измерения называют вопросы, связанные с присвоением наблюдениям шкальных значений, как показано в (1), проблемой *репрезентации*. Инвариантность шкал при преобразованиях, как (2), (3) или (4), они называют проблемой *единственности*. Определение истинности или ложности суждений, основанных на сравнении приписанных шкальных значений, называется проблемой *осмысленности* [2]. Эта последняя проблема, связанная с осмысленностью эмпирического шкалирования и основанного на нем анализа, все еще остается предметом статистической полемики.

Предписывающая и запрещающая статистики

В своей статье «Математика, измерение и психофизика» [3] Стивенс вышел за пределы своей элементарной типологии. Он

классифицировал не только простые операции, но также и статистические процедуры с точки зрения их «допустимости» для тех или иных шкал. Шкала, не теряющая смысл при определенном классе преобразований, должна, по мнению Стивенса, описываться только статистиками, значение которых не изменяется при любых из этих преобразований, примененном к данным.

Поэтому, например, анализ номинальных данных должен быть сведен к обобщающим статистикам, таким как количество случаев, мода, коэффициенты сопряженности, для которых достаточно лишь соблюдения тождественности значений. Допустимые статистики для ординальной шкалы включают перечисленное выше плюс медиану, перцентили, ранговые корреляции – это статистики, сохраняющие смысл шкалы в случае применения монотонных преобразований. Вдобавок интервальные данные допускают вычисление среднего, стандартного отклонения (несмотря на то что не все часто употребляемые статистики имеют дело со стандартным отклонением), корреляции как произведения моментов, так как интерпретации этих статистик не изменяются при линейных преобразованиях данных. Наконец, шкала отношений допускает все перечисленное плюс геометрическое среднее и коэффициенты вариации, неизменные при изменении шкалы данных.

Льюс, суммируя сказанное, заключил: «... тип шкалы устанавливает [ограничения] на статистики, которые, можно осмысленно использовать. Если интерпретация определенной статистики или статистического теста изменяется при использовании допустимого для шкалы преобразования, то наши содержательные выводы будут зависеть от того, какую из произвольных репрезентаций мы задействовали в подсчетах. Большинство ученых, осознав эту проблему, чувствуют, что подобной статистики следует избегать, и полагаются только на те статистики, которые демонстрируют должную инвариантность для взятого шкального типа. Как геометрические, так и арифметические средние оказываются правомочны в этом смысле для шкал отношений (единица произвольна),

только последнее уместно для интервальных шкал (единица и ноль произвольны) и ни одно из средних не годится для ординальных шкал» [4, р. 84].

Авторы учебников быстро адаптировали эти идеи (например: [5, 6]), вероятно, по той причине, что намеревались в том числе предоставить простое руководство и оградить наивного аналитика от ошибок при применении статистики. К сожалению, в то время как определение типа шкалы не составляет труда, обоснование этого определения в терминах классов преобразований представляется весьма изощренным и обычно остается непонятным для начинающих студентов и, более того, как будет показано ниже, тип шкалы может оказаться и вовсе не ясен.

Стало привычным встречать таблицы (зачастую внутри задней стороны обложки), в которых читатель может найти «подходящий тест», основанный на количестве и шкальных типах переменных. Также идеи Стивенса повлияли на методологов социальных наук на более высоком уровне. Эндрюс, Клем, Дэвидсон, О'Малли и Роджерс [7] вывели расширенную таксономию одномерных и многомерных статистических процедур, основанных на шкалах Стивенса. Их древовидная система была использована по крайней мере в одной компьютерной программе, претендующей быть статическим советником, основанным на искусственном интеллекте.

Недавно таксономия Стивенса была положена в основу интерфейса некоторых компьютерных статистических пакетов общего назначения. Перед началом анализа эти пакеты требуют определить тип измерительной шкалы каждой переменной. Затем они автоматически выбирают «подходящий» вид анализа, определенный в соответствии с пользовательским описанием взаимосвязей данных. Анализ, определенный Стивенсом в качестве недопустимого для определенной шкалы, не может быть выполнен без предварительного изменения типа шкалы.

Классическая критика ограничений Стивенса

Критика работы Стивенса сосредоточилась на трех моментах. Во-первых, сведение выбора только к тем статистическим методам, которые «демонстрируют инвариантность, подходящую для данного типа шкалы», представляется опасным для анализа данных практикой. Во-вторых, его таксономия слишком строга, чтобы ее возможно было применять для реальных данных. В-третьих, стивенсовские ограничения часто ведут к понижению уровня данных через их преобразование в ранги и последующее ненужное обращение к непараметрическим методам.

Лорд [8] критиковал аргументы Стивенса в занятой и простой для понимания форме, показав, что выбор допустимых статистических тестов для некоторого набора данных не зависит от проблем репрезентации или единственности, а зависит от осмысленности. Лорд доказывал, что осмысленность статистического анализа зависит от вопроса, для ответа на который он разработан. Для этого Лорд изобразил профессора, ведавшего распределением номеров на футбольных майках, которого обвинили в том, что он раздал первокурсникам слишком уж низкие по величине номера. И хотя профессор полагал, что футбольные номера составляют лишь номинальную шкалу, статистик, который его консультировал, не затруднился провести с ними операцию сложения, возвести их в квадрат, затем вычислить среднее, а также произвести все прочие необходимые для применения неравенства Чебышева операции (не опираясь на предположение о нормальности), чтобы протестировать обоснованность обвинения в раздаче первокурсникам «чересчур малых» номеров. Когда профессор возразил, что футбольные номера относятся к номинальной шкале, статистик заметил, что номерам об этом не известно. В ретроспективе это замечание кажется слишком несерьезным при всей существенности лордовской критики.

Бейкер, Хардик и Петринович [9], Боргатта и Борнштедт [10] подчеркнули тот факт, что следование стивенсовским ограниче-

ниям часто заставляет исследователей прибегать к ранговому упорядочению данных и тем самым отказываться от использования параметрических тестов. Их аргументы базировались на центральной предельной теореме и методе Монте-Карло и были направлены на то, чтобы показать, что в случае с типичными данными не стоит переживать о том, являются шкалы «ординальными» или «интервальными». Их аргументация носила несколько *ad hoc* характер и, к сожалению, завершалась рекомендацией использовать стандартные параметрические процедуры вместо того, чтобы связываться с проблемой робастности. Тем не менее они высветили недостатки обсуждения Стивенсом «допустимой» арифметики.

Гуттман [11] в более общем смысле доказывал, что статистическая интерпретация данных зависит от того, какой вопрос обращен к данным и какое доказательство мы готовы принять в ответ на этот вопрос. Он определил это доказательство в терминах функции потерь, выбранной для проверки качества модели. В любом случае одни и те же данные могут быть интерпретированы по-разному посредством выбора такой функции потерь: «Для анализа данных не требуется никаких разрешений. Что требуется, так это минимизация функции потерь. Практики любят спрашивать об априорных правилах, как о том, что “разрешено делать” с их неупорядоченными, упорядоченными или числовыми данными, без какого-либо упоминания функции суммарных потерь применительно к интересующей их проблеме. А в действительности они должны сказать математику, что здесь они имеют функцию потерь, и спросить, как возможно минимизировать потери. Минимизация может потребовать обращения с числовыми данными, как с лишенными порядка, или с неупорядоченными данными, как с данными числовыми. Если математик дает “разрешение” или отказывает в нем вне какой-либо связи с функцией потерь, то он, наверное, помогает практику избежать прояснения исследовательской проблемы».

Джон Тьюки также критиковал стивенсовские ограничения как опасные для хорошего статистического анализа. Подобно Лорду и Гуттману, Тьюки отметил важность смысла данных при определении и шкалы, и подходящего способа анализа. Поскольку шкальные типы Стивенса абсолютны, в ситуации когда, например, данные нельзя считать полностью интервальными, их следует понизить в ранге до ординальных. Он доказывал, что это неправильное использование статистик – полагать, что статистические методы также должны быть абсолютны. Тьюки, обращаясь к приведенной выше мысли Льюса, писал: «Взгляд, обобщенный Льюсом, представляется опасным... Причина чувств тех, кто верит, что определенный тип шкалы должен ограничивать использование статистик, наверное заключается в популярном взгляде на статистические процедуры как на ритуал освящения и печать окончательного одобрения. Результаты, основанные на приблизительных основаниях, должны быть использованы с сознанием этой приблизительности. Те, кто ищет определенности скорее чем истины, фактически избегают этого соображения. Но какое знание не основано на некоторой приблизительности? И какой прогресс возможен без подобного знания?» [12, р. 245–246].

Даже сам Стивенс оговаривался, замечая: «Фактически большая часть шкал, широко и эффективно применяемых психологами, – это шкалы порядка. Обычные статистики, включая средние и стандартные отклонения, при строгом подходе не должны использоваться при работе с этими шкалами... Однако такому неправоначальному использованию может быть дано известное прагматическое оправдание: во многих случаях оно приводит к плодотворным результатам» [3, с. 56].

Дискуссия о статистиках и шкальных типах

В общем, статистики отвергли запрет на методы, основанный на ограничениях, связанных с допустимыми преобразованиями.

Теоретики измерения достигли немалых результатов в области формализации (см., например: [13–16]). Многие из этих авторов преимущественно занимались статистикой, заключая, как правило, что выбор статистических методов должен ограничиваться шкальным типом данных. (См., например: [14, ch. 20–22].) У Зумбо и Зиммермана [17] можно найти исчерпывающий обзор и объемную библиографию.

Порой дебаты принимали довольно жесткую форму. Так, Гейто [18] делал саркастические выпады в адрес теории измерения, Таунсенд и Эшби [19], защищая последнюю, отвечали тем же. К сожалению, как отметил Мичел [20], часто выпады обеих сторон приходились мимо цели.

Мы не предлагаем окончить эти яростные споры, продолжавшиеся почти половину века. Вместо этого мы концентрируем внимание на определенном аспекте применения теории измерений в статистике: использовании шкальных типов для выбора или предписания статистических методов. Несмотря на то что мы предлагаем много аргументов, ключевой аргумент против использования предписания статистик на основе шкального типа гласит: это не работает!

Различие в точках зрения частично проистекает из фундаментальной разницы между математикой и наукой. Тьюки отметил эту разницу, проведя разграничение между анализом данных и математической статистикой. «Существуют различные взгляды на то, что создает науку, но три компонента признаются повсеместно: (a1) интеллектуальное содержание, (a2) приведение в доступный пониманию вид, (a3) опора на проверку опытом как окончательный критерий достоверности. В последнем смысле математика не есть наука, так как такими критериями достоверности в ней является согласие относительно логической непротиворечивости и доказуемости» [21, p. 397].

Аксиоматическая теория измерения – скорее математика, чем наука. Предписываемые ею ограничения на использование опреде-

ленных статистических методов не удовлетворяют критерию, предложенному Тьюки (а3): в широком диапазоне ситуаций опыт показывает, что применение запрещенных статистик к данным приводит к научно значимым результатам, важным при принятии решений и ценным для дальнейших исследований.

Альтернативные шкальные таксономии

Несколько авторов предложили альтернативные таксономии типов данных (обычно, однако, не предполагающих предписывание или ограничение использования тех или иных статистических методов и часто не претендующих на исчерпывающий характер). Один такой побуждающий к размышлениям список был представлен Мостеллером и Тьюки [22, гл. 5]. В него вошли следующие наименования:

- *классы* (упорядоченные наименования, такие как первокурсник, второкурсник, третьекурсник, дипломник),
- *ранги* (начинающиеся с единицы, которая может быть наибольшим или наименьшим значением),
- *доли* (в пределах от нуля до единицы, включая, например, проценты),
- *счетные числа* (неотрицательные целые числа),
- *величины* (неотрицательные действительные числа),
- *соотношения или веса* (не ограниченные, положительные или отрицательные значения).

Мостеллер и Тьюки использовали эти типы, чтобы предложить «первую помощь» для преобразования данных, включая преобразования, изменяющие тип данных. Ни в коем случае они не имели в виду, что наш выбор анализа или даже преобразований должен быть ограничен этими категориями. Также указанные типы не предлагались в качестве типов измерительных шкал в смысле аксиоматических аргументов Льюса и соавторов [14].

Список Мостеллера и Тьюки показывает, что стивенсовские типы не исчерпывают всех возможностей даже для простых дан-

ных. Куда, к примеру, в его классификации следует поместить доли (такие как проценты), имеющие верхний и нижний пределы, и, таким образом, не выдерживающие даже произвольных сдвигов шкалы?

Ограничивая преобразования

Многие авторы отмечали, что простые преобразования могут привести данные в более удобный для качественного анализа вид. Большинство из участвовавших в дискуссии рекомендовали такую практику. Мостеллер и Тьюки, предложив свой список типов данных, рекомендовали преобразование данных – зачастую изменяющее «тип» значений в их собственном списке.

Преобразование значений данных для упрощения структуры (например, для большей симметричности распределения, для приближения разброса между группами к постоянной величине, для увеличения линейности взаимосвязей, повышения соответствия факторных экспериментов аддитивным моделям) имеет длинную и отнюдь не бесславную историю в статистике (см., например: [23–25].) Эти и другие авторы доказали, что инструменты для хорошего статистического анализа включают подобные преобразования. Наиболее популярные и полезные преобразования включают логарифмические и простые экспоненциальные, а также корни, которые монотонны, но не линейны (в противном случае они не могли бы упростить структуру). Но таксономия Стивенса разрешает такие преобразования только для номинальных и ординальных шкал – шкал, для которых такие понятия, как линейность, гомоскедастичность, аддитивность и симметричность полагаются бессмысленными.

Тьюки предложил мысленный эксперимент, в котором у обыкновенных почтовых весов шкала оказалась не откалибрована, в результате чего измерение веса достигает правильного упорядочивания объектов, но не является шкалой отношений [12, р. 250]. Он утверждал, что хотя экспериментальное доказательство и покажет, что вес не является измерением на уровне

шкалы отношений, лучше будет преобразовать «веса» обратно в более простую шкалу.

Нет оснований полагать, что данные, которые мы получаем, измерены наилучшим образом. Хоаглин обратил внимание на множество повседневных примеров данных, обычным образом преобразуемых посредством некоторой (как правило монотонной) функции [26].

Эйбелсон и Тьюки отображали ординальные шкалы в интервальные и обсуждали размер вероятной ошибки, вызванной такой процедурой. Они критиковали тенденцию выбирать непараметрические статистики из-за шкалы не из-за того, что первым недостает мощности, а потому, что «они слабо приспособлены к разнообразному использованию, необходимому для хорошего интуитивного проникновения в закономерности массива данных» [27, p. 407].

Шепард [28], Крускал [29], Гуттман [30] и другие авторы разработали процедуры многомерного шкалирования, которые могут использоваться для перевода ординальных, согласно стивенсовскому определению, измерений в шкалы отношений. Такие результаты могут быть связаны с многообразием предназначенных для «отношений» статистических процедур (например, пространственные статистики), которые инвариантны относительно монотонных преобразований изначально ранжированных данных (коль скоро они не влияют на результаты многомерного шкалирования). Эта двухэтапная процедура нарушает стивенсовское предписание относительно того, что статистики типа t и F невалидны для порядковых данных, однако она оказалась полезной многим аналитикам.

Хороший анализ данных не основан на допущениях о типе данных

Некоторые авторы замечали, что в анализе данных «вещи редко являются тем, чем кажутся». Так, Джойнер [31] обратил

внимание на примеры, в которых данные, принадлежащие, как кажется, одному типу, в действительности скрывают другую информацию (прячущиеся переменные, в его терминологии). Например, идентификационный номер розничной торговой точки должен быть, по всей видимости, номинальным. Тем не менее следует иметь в виду, что номера могли присваиваться последовательно в ходе расширения сети, поэтому можно учитывать возможность корреляции номера и других важных переменных (продажи, прибыль).

Джойнер ссылаясь на пример, в котором располагавшиеся высоко на стене клетки с подопытными животными имели совсем иное окружение в сравнении с другими, расположенными ближе к полу. В другом эксперименте животные были (некорректно) распределены по клеткам (и соответственно по условиям эксперимента) на основе принадлежности к помёту, а не посредством правильной процедуры рандомизации. Это показывает, что внимательному аналитику не следует полагать, что шкальный тип переменной таков, каким он кажется, даже в том случае, если относительно данных имеются однозначные гарантии.

Стивенсо́вские категории не описывают фиксированных свойств данных

Довольно легко сконструировать ситуации, в которых шкальный тип зависит от их интерпретации или от наличия дополнительной информации. На приёме, который спонсировался секциями статистических вычислений и статистической графики Американской статистической ассоциацией, всем участникам на входе последовательно выдавались лотерейные билеты, пронумерованные с единицы. Когда выигравший номер – 126 – был объявлен, одна из участниц сравнила этот номер с номером своего билета, чтобы узнать, не выиграла ли она, при этом верно рассматривая «126» в номинальном ключе. Затем она окинула

зал взглядом и заметила, что едва ли в нем могло набраться 126 человек, вновь верно проинтерпертировав ту же величину (но уже с учетом дополнительной информации о последовательной нумерации билетов начиная с единицы), как принадлежащую шкале отношений. Один из авторов сравнил номер своего билета (56) с выигравшим номером и осознал, что он приехал слишком рано, т.е. проинтерпретировал значение как относящееся к ординальной шкале. Будь ему доступны дополнительные данные о частоте и регулярности прибытия гостей, он мог бы попытаться оценить, на сколько позже ему следовало явиться, чтобы покрыть эту разницу в 70 билетов, воспринимая, таким образом, номер билета как значение на интервальной шкале.

Обычный набор данных сообщает факты об автомобилях. Например, количество цилиндров двигателя. В определенных случаях число цилиндров анализируется как номинальная категория: «есть ли значимые различия по расходу топлива между машинами с восьми-, шести- и четырехцилиндровыми двигателями?» Разумеется, эти категории безусловно упорядочены, поэтому статистики для ординальных данных также будут применимы. Но можно спросить и о среднем количестве цилиндров, скажем, в американских автомобилях, и озадачиться вопросом: не уменьшилось ли это среднее за последние годы? А это требует, чтобы значения (все из которых целые числа) выступали как значения интервальной шкалы, в чем, собственно, также нет затруднения: разница между восемью и шестью цилиндрами равняется разнице между шестью и четырьмя. Наконец, мы можем учесть размер каждого цилиндра и подсчитать отношения объемов двигателей, что есть вполне приемлемая операция для данных шкалы отношений.

Цель всех этих примеров – показать, что утверждение, общее для многих традиционных статистических текстов – «значения данных бывают номинальными, ординальными, интервальными или измеренными по шкале отношений», – упрощает положение дел вплоть до полного его искажения. Шкальный тип, описанный

Стивенсом, не является свойством данных, он зависит от вопроса, который мы ставим, и от дополнительной информации, которой мы можем обладать. Он способен меняться по причине преобразования данных, по причине появления новой информации, позволяющей по-разному трактовать данные, или попросту по причине смены исследовательского вопроса.

Розенбум доказывал сходную точку зрения: «Если мы однако обнаруживаем интерпретативную значимость в некоторой статистике, запрещенной для того типа данных, к которому, как считается, она относится, то шкальный “тип” тогда расширяется до требуемого этой интерпретацией уровня» [32, р. 197].

Категории Стивенса недостаточны для описания шкал данных

Шкалы могут быть многомерными. Здесь приведена частично упорядоченная двоичная шкала.

Таблица 1

ЧАСТИЧНО УПОРЯДОЧЕННАЯ ДВОИЧНАЯ ШКАЛА

Левая			Правая	Сумма по строке	
		1111		4	Больше
		1110	0111	3	
	1100	0110	0011	2	
1000	0100	0010	0001	1	
		0000		0	Меньше

В этой шкале горизонтальное измерение содержит качественную (номинальную) шкалу признаков, вертикальное – количественную (ординальную, интервальную или отношений). Например, каждый профиль может отображать наличие или отсутствие каждого из четырех симптомов у пациента. В этом случае

вертикальная шкала может быть связана с тяжестью болезни, а горизонтальная – с различными синдромами. Эти шкалы обсуждались Гудманом [33] и Гуттманом (в сборнике Шайя) [34]. Если бы мы использовали стивенсовскую иерархию как руководство для анализа этих структур, то мы попросту скрыли бы их существование, так как различие шкальных типов в строках и столбцах не позволяет определить общий шкальный тип. Область неметрических совместных измерений также посвящена многомерному шкалированию «номинальных» и «ординальных» данных [35].

Андерсон показал, что одни те же данные могут быть измерены по различным шкалам, относящимся к одному типу, которые тем не менее будут приводить к разным статистическим результатам. Он ссылается на пример выбора между измерением продолжительности и быстроты процесса. То и другое – валидные интервальные шкалы, при этом статистики, вычисленные в одной форме, могут совершенно отличаться от вычисленных в другой. Андерсон отметил: «Очевидно, наличие интервальной шкалы не гарантирует инвариантности статистик интервальных шкал» [36, р. 31].

Статистические процедуры не могут классифицироваться по критерию Стивенса

Это было верно даже в момент появления статьи Стивенса, однако стало еще более очевидно с развитием робастных методов. Представим, например, линейную оценочную функцию для территории:

$$L = \sum a_i x_i, \quad (5)$$

где x_i – статистика i -го порядка для выборки размера n . Допустим, a_i – однородные веса, приписанные так, чтобы добиться симметрично цензурированной линейной оценки. Тогда некоторые веса на каждом конце последовательности равняются нулю. Если

мы используем веса $1/n$ без цензуры, то L становится средним. Если цензуются менее 50% значений с каждой стороны, то L становится медианой. Оценивание, таким образом, находится между ординальной и интервальной стивенсовской категорией. Естественно, невозможно ввести в категорию «тип» данных, для которых приемлемо частичное цензурирование (хотя исследования показали, что при многих обстоятельствах подобное оценивание оказывается уместным).

В определенном смысле кажется, что цензурированное среднее делит данные на центральный массив «интервальных» значений и хвосты «ординальных». Если настаивать на категоризации более робастных мер, то в соответствии со стивенсовскими типами получится, что в середине они работают с данным как с интервальными, на хвостах – как с ординальными, а в предельных случаях – как с номинальными. В исследовании с реальными данными Ф. Хампель, Э. Рончетти, П. Рауссеу и В. Штаэль [37] заметили, что подобное оценивание применимо к существенной доле реальных данных. Должны ли мы принимать во внимание, что многие данные могут быть описаны как относящиеся к нескольким шкальным типам одновременно?

Если мы ищем простые правила для идентификации шкальных типов, робастные меры будут смущать нас и дальше. Приписывание значений к «середине» или «хвосту» распределения переменчиво и зависит от наблюдаемых значений данных. Добавление даже одного нового наблюдения способно изменить это приписывание. А для многих мер переход от хвоста к середине гладок и вообще не поддается точному определению.

Шкальные типы — не точные категории

Во многих обсуждениях и практически при всякой математической работе шкальные типы понимаются как абсолютные категории. Предполагается, что данные соответствуют той или

иной категории. Если не удастся достигнуть определенного уровня измерения, то считается, что следует осуществить понижение уровня. Как бы то ни было, реальные данные не удовлетворяют требованиям шкальных типов. Тьюки [12] подчеркнул, что когда интервальное измерение производится с систематической ошибкой калибровки, которая зависит от измеряемых значений (что нередко случается), полученные значения не образуют действительной интервальной шкалы. Разница между двумя измеренными значениями на одном конце шкалы не будет полностью соответствовать разнице между двумя значениями на другом конце шкалы. Тем не менее, если ошибки относительно измеряемых значений малы, мы пожертвуем большей частью информации, содержащейся в данных, если «разжалуем» их до ordinalной шкалы. Например, такое понижение не позволит нам узнать даже о равенстве дисперсий в двух совокупностях, измеренных подобным образом. Тьюки заключил: «Чрезмерно упрощенный и рафинированный взгляд на то, что такое измерение, не может диктовать, как анализировать данные» [12, p. 247].

Шкалы и анализ данных

Обсуждение статистик в терминах шкальных типов (например, Льюс и соавторы [14, ch. 22]) подразумевает, что особенности измерения определяют тип данных и задают возможности адекватной постановки (и проверки) гипотезы. Современные подходы к анализу данных, как, например, разведочный статистический анализ [38–40], прояснили известный ученым-прикладникам факт, что зачастую гипотеза не предшествует анализу данных.

Как было показано выше, шкальный тип данных может частично обуславливаться тем вопросом, который ставит исследователь, или целями, для которых он предназначен. Так, лордовский вымышленный профессор обосновал интервальный характер шкалы номеров футбольных игроков, когда он задался вопросом:

выше ли значения номеров у второкурсников, чем у первокурсников? В описанной лотерее, чтобы определить, кто же победил, номера билетов понимались как номинальные. Чтобы выяснить, сколько человек пришло на конференцию, те же самые значения были поняты в ключе шкалы отношений, что также было вполне адекватно.

Хороший анализ данных редко следует формальной парадигме проверки гипотезы. Открытие непредвиденных взаимосвязей происходит с помощью общего анализа данных в поисках паттернов. Конечно, подобный анализ становится невозможен, если утверждается, что данные принадлежат такому шкальному типу, который не допускает даже рассмотрения некоторых паттернов, но это – очевидно ненаучный подход. Ученый должен быть открыт для *любого интересного* паттерна. Подходы к статистике, начинающиеся с априорного шкального типа и затем исключающие различные гипотезы или статистические методы и тесты для данного шкального типа, есть негодная наука и плохой анализ данных.

Именно в этом духе выдающиеся статистики критиковали стивенсовские ограничения. Например, И.Р. Сэвидж в критическом обзоре работы Сигела [42] утверждал: «Я не знаю причины ограничивать статистические процедуры аутентичными операциями, совместимыми со шкалой наблюдаемых количественных значений» [41, p. 247].

Осмысленность

Определения стивенсовских шкал в первой части статьи следуют традиционной идее, что суждение о данных для разных шкал сохраняет смысл при допустимых преобразованиях. Однако термин *осмысленность* остался без определения. Теория измерений сообщает, что осмысленность – это «то, что сохраняется при допустимых преобразованиях». А отсюда один шаг до ограничения статистик, использующих запрещенные операции, так как

эти операции уничтожают осмысленность (для примера можно посмотреть приведенную выше цитату Льюса).

Сторонники этого подхода считают осмысленность абсолютной. Например, Таунсенд и Эшби утверждали: «Как, вероятно, очевидно, осмысленность – это “все-или-ничего” понятие. Поэтому суждение не может быть почти осмысленным» [19, р. 394].

В науке, как и в анализе данных, смысл и осмысленность не столь просты. Наука развивается посредством проведения измерений – неизбежно содержащих ошибки, и построения теорий – предположительно неверных (притом, что они могут быть лучшим, на что мы сегодня способны), чтобы затем попытаться прийти к лучшему. Если бы наука была ограничена доказуемо осмысленными суждениями, она не смогла бы развиваться. Нам следует относиться к нашим несовершенным описаниям мира с уважением. Как отмечал Френсис Бэкон, «истина скорее возникает из ошибки, чем из неясности» [43, р. 210].

В статистическом анализе смысл происходит не только из данных, но и из изучаемого вопроса, моделей, открытых в ходе анализа, и дополнительных данных, которые могут быть доступны. В примере Лорда величина номеров на футбольных майках была важна не в привычном смысле, а в смысле, возникшем в связи с шутками второкурсников над первокурсниками и последовавшим желанием первокурсников вернуть свои деньги, что привело к тому, что профессор был вынужден разобраться в ситуации. В примере с лотереей абсолютная величина победного номера не имела никакого значения, так как номер победителя был выбран произвольно, но абсолютная величина обрела смысл, когда была использована для оценки количества посетителей приема.

Споры вокруг осмысленности могут частично происходить из безответственного расширения границ термина, изначально предназначенного для специального понятия. Математики часто используют обыкновенные слова для обозначения четко определенных концептов. Наименование концепта некоторым словом не дает переопределения

слову. Как «значимые» статистические показатели не обязаны иметь особую теоретическую значимость, «нормальное» распределение редко встречается, а «мощный» критерий не связан с мощностью в ваттах, так и осмысленность оказывается более богатым понятием, чем заложено аксиомами теории измерений.

Роль типов данных

Было бы ошибкой полагать, что типы данных не имеют значения. Конечно, в планировании любого эксперимента должно учитываться различие между категориальными факторами, которые в терминологии Стивенса обычно являются номинальными или ординальными, и непрерывными ковариатами, обычно измеряемыми на уровне шкалы интервалов или отношений. Понятие типа шкалы важно, а терминология Стивенса зачастую бывает удобна. Действительно, значительная часть дискуссии, представленной в настоящей статье, была бы невозможна без этих понятий. Нужно иметь в виду, однако, что шкальный тип не есть фундаментальное свойство данных, а скорее производное от того, как данные были измерены и того, какой на их основе должен быть сделан вывод.

При любом анализе данных есть возможность ставить бессмысленные вопросы. Понимание шкалирования порой может помочь избавиться от бессмыслицы, но рассуждение должно строиться в верном порядке. Вместо того чтобы основывать выбор статистических методов на типе шкалы, нужно начинать с самих данных и наших теорий о том, что стоит за этими данными. Мы направляем анализ данных в соответствии с тем, что мы хотим узнать с помощью этих данных. Производство вывода без сомнения требует, чтобы данные поддерживали тот или иной тип измерения. Когда потенциальная модель вывода построена, следует проверить, можно ли считать имеющиеся данные соответствующими той измерительной шкале, которая требуется для производства этого вывода. Если нет (например, мы были уверены,

что номера клеток с животными – номинальные, но в дальнейшем выяснили, что существует корреляция с зависимой переменной), то мы должны найти объяснение. Меньшее следует признать безответственной наукой.

Ограничить исследование только гипотезой и вычислениями, дозволенными априорной моделью типа шкалы, будет еще более безответственно. Как заметил Кун, «открытие начинается с осознания аномалии, т.е. с установления того факта, что природа каким-то образом нарушила навеянные парадигмой ожидания, направляющие развитие нормальной науки» [44, с. 82].

Анализ данных должен быть открыт аномалии, если он направлен на развитие науки. Попытки сузить количество возможных взаимосвязей, которые могут быть учтены, запретить преобразования, которые могут быть применены, или ограничить статистики, которые могут быть вычислены, снижают нашу способность к обнаружению аномалии. Учебники и компьютерные программы, навязывающие подобный подход к данным, вводят читателей и пользователей в заблуждение.

Один из источников сложностей с компьютерными программами может быть обнаружен в том, что программисты обычно устанавливают типы для переменных, разделяя, например, действительные числа, целые числа или текстовые строки. Для разработчиков компьютерного софта это может быть естественно – приписывать данным типы, но тем не менее нет причин навязывать их пользователям. Многие из современных статистических методов, бросающих вызов типизации шкал данных, стали практически осуществимы только благодаря компьютерам. То, как мы их используем, зачастую зависит от того, каким образом эти методы реализованы на компьютерах. Нам следует проявлять внимательность, чтобы избежать необязательных ограничений, которые могут быть обусловлены удобством программиста, а не фундаментальным пониманием данных и их анализа.

Заклочение

Теория измерений важна для интерпретации статистического анализа. Однако применение типологии Стивенса в статистике обнажает множество непростых проблем. Статистические программы, основанные на типологии Стивенса, подразумевают, что статистика – это просто вопрос декларирования типа шкалы и выбора модели. Более того, допускается, что тип шкалы как бы самоочевиден и *не зависит от того, какой вопрос ставит исследователь перед своими данными*. Таким образом исключается возможность поставить по отношению к данным ряд вопросов. Подобные ограничения ведут к некачественному анализу и негодной науке.

Недавние попытки разработки основанного на «искусственном интеллекте» программного обеспечения поддержали использование этой терминологии в статистике и проигнорировали тонкости творческого анализа данных. Разумеется, аналитик должен нести ответственность за свой выбор методов анализа, соответствующих имеющимся данным и поставленным вопросам. Статистическое программное обеспечение, способствующее любому анализу для любых данных допускает и безответственный анализ. Рассмотрение вопроса: *убедительны ли предположения о типах шкал после проведения анализа?* – может помочь выявить бессмыслицу. Но программное обеспечение, налагающее произвольные ограничения, в равной мере способно вести к ошибочным выводам.

Благодарности

Авторы выражают благодарность Сергею Адамову, Ингверу Боргу, Ласло Энгельману, Пету Флери, Давиду Хоаглину, Джону Тьюки за ценные комментарии.

Перевод В.И. Хоменко.

ЛИТЕРАТУРА

1. *Stevens S.S.* On the Theory of Measurement Scales // *Science*. 1946. No. 103. P. 677–680.
2. *Сунпес П., Зинес Дж.* Основы теории измерений // Психологические измерения / Под ред. Л.Д. Мешалкина. М.: МИР, 1967.
3. *Стивенс С.С.* Математика, измерение и психофизика // Экспериментальная психология / Под ред. С. Стивенса. М.: Иностран. лит., 1960. С. 19–99.
4. *Luce R.D.* On the Possible Psychophysical Laws // *Psychological Review*. 1956. No. 66. P. 81–95.
5. *Blalock H.M.Jr.* Social Statistics. N.Y.: McGraw-Hill, 1960.
6. *Sigel S.* Nonparametric Statistics for the Behavioral Sciences. N.Y.: McGraw-Hill, 1956.
7. *Andrews F.M., Klem L., Davidson T.N., O'Malley P.M., Rodgers W.L.* A Guide for Selecting Statistical Techniques for Analyzing Social Science Data. Ann Arbor: Univ. of Michigan; Institute for social research, 1981.
8. *Lord F.* On the Statistical Treatment of Football Numbers // *American Psychologist*. 1953. No. 6. P. 750–751.
9. *Baker B.O., Hardyck C.D., Petrinovich L.F.* Weak Measurements vs. Strong Statistics: an Empirical Critique of S.S. Stevens's Proscriptions on Statistics // *Educational and Psychological Measurement*. 1966. No. 26. P. 291–309.
10. *Borgatta E.F., Bohrnstedt G.F.* Level of Measurement – Once Over Again // *Sociological Methods and Research*. 1980. No. 9. P. 147–160.
11. *Guttman L.* What Is Now What in Statistics // *The Statistician*. 1977. No. 26. P. 81–107.
12. *Tukey J.W.* Data Analysis and Behavioral Science of Learning to Bear the Quantitative Man's Burden by Shunning Badmandments // *The Collected Works of John W. Tukey* / Ed. by L.V. Jones. Belmont, CA: Wadsworth, 1961. Vol. 3. P. 391–484.
13. *Krantz D.H., Luce R.D., Suppes P., Tversky A.* Foundations of Measurement. N.Y.: Academic press, 1971. Vol. 1.
14. *Luce R.D., Krantz D.H., Suppes P., Tversky A.* Foundations of Measurement. N.Y.: Academic press, 1990. Vol. 3.
15. *Narens L., Luce R.D.* Measurement: the Theory of Numerical Assignments. Reading, MA: Addison-Wesley, 1986.
16. *Roberts F.S.* Measurement Theory. Reading, MA: Addison-Wesley, 1979.
17. *Zumbo B. D., Zimmerman D. W.* Levels of Measurement and the Relation Between Parametric and Nonparametric Tests. Working Paper 91-1. Edometrics Research Group, University of Ottawa.
18. *Gaito J.* Measurement Scales and Statistics: Resurgence of an Old Misconception // *Psychological Bulletin*. 1980. No. 87. P. 564–567.

19. *Townsend J. T., Ashby F. G.* Measurement Scales and Statistics: the Misconception Misconceived // *Psychological Bulletin*. 1984. No. 96. P. 394–401.
20. *Michel J.* Measurement scales and Statistics: a Clash of Paradigms // *Psychological Bulletin*. 1986. No. 100. P. 398–407.
21. *Tukey J.W.* The Future of Data Analysis // *The Collected Works of John W. Tukey* / Ed. by L.V. Jones. Belmont, CA: Wadsworth, 1961. Vol. 3. P. 187–389.
22. *Мостеллер Ф., Тьюки Дж.* Анализ данных и регрессия. М.: Финансы и статистика, 1982.
23. *Bartlett M.S.* The Use of Transformation // *Biometrics*. 1947. No. 3. P. 39–52.
24. *Tukey J.W.* On the Comparative Anatomy of Transformations // *Annals of Mathematical Statistics*. 1957. No. 28. P. 602–632.
25. *Box G.E.P., Cox D.R.* An Analysis of Transformations // *Journal of the Royal Statistical Society*. 1963. No. 26. P. 211–252.
26. *Hoaglin D.C.* Transformations in Everyday Experience // *Chance*. 1988. No. 1. P. 40–45.
27. *Abelson R.P., Tukey J.W.* Efficient Utilization of Non-numerical Information in Quantitative Analysis: General Theory and the Case of Simple Order // *Annals of Mathematical Statistics*. 1963. No. 34. P. 1347–1369.
28. *Shepard R.N.* The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function // *Psychometrika*. 1962. No. 27. P. 125–139.
29. *Kruskal J.B.* Nonmetric Multidimensional Scaling: a Numerical Method // *Psychometrika*. 1964. No. 29. P. 115–129.
30. *Guttman L.* A General Nonmetric for Finding the Smallest Coordinate for a Configuration of Points // *Psychometrika*. 1968. No. 33. P. 469–506.
31. *Joiner B.F.* Lurking Variables: Some Examples // *The American Statistician*. 1981. No. 35. P. 227–233.
32. *Rozenboom W.W.* Scaling Theory and the Nature of Measurement // *Synthese*. 1966. No. 16. P. 170–233.
33. *Goodman L.A.* A New Model for Scaling Response Patterns: an Application of Quasi-independent Concept // *Journal of the American Statistical Association*. 1975. No. 70. P. 755–768.
34. *Shye S.* Partial Order Scalogram Analysis // *Theory Construction and Data Analysis in the Behavioral Sciences* / Ed. by S. Shye. San Francisco: Jossey-Bass, 1978.
35. *Green P.E., Rao V.R.* Conjoint Measures for Quantifying Judgmental Data // *Journal of Marketing Research*. 1971. No. 8. P. 355–363.
36. *Anderson N.H.* Scales and Statistics: Parametric and Nonparametric // *Psychological Bulletin*. 1961. No. 58. P. 305–316.
37. *Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В.* Робастность в статистике: подход на основе функций влияния. М.: Мир, 1989.
38. *Hoaglin D.C., Mosteller F., Tukey J.W.* Understanding Robust and Exploratory Data Analysis. N.Y.: John Wiley, 1983.

39. Тьюки Дж. Анализ данных наблюдений: разведочный анализ / Под ред. В.М. Писаренко. М.: Мир, 1981.
40. *Velleman P.W., Hoaglin D.C.* Applications, Basics and Computing of Exploratory Data Analysis. Boston: Duxbury press, 1981.
41. *Savage I.R.* Nonparametric Statistics // Journal of the American Statistical Association. 1957. No. 52. P. 331–334.
42. *Sigel S.* Nonparametric statistics for the behavioral sciences. N.Y.: McGraw-Hill, 1956.
43. *Bacon F.* Novum Organum // The works of Francis Bacon / Ed. by J. Spedding, R.L. Ellis, D.D. Heath. Belmont; N.Y., 1868. Vol. 8.
44. *Кун Т.* Структура научных революций. М.: Прогресс, 1977.