А.И. Орлов
(Москва)

ТЕОРИЯ ИЗМЕРЕНИЙ КАК ЧАСТЬ МЕТОДОВ АНАЛИЗА ДАННЫХ: РАЗМЫШЛЕНИЯ НАД ПЕРЕВОДОМ СТАТЬИ П.Ф. ВЕЛЛЕМАНА И Л. УИЛКИНСОНА[1]

Согласно современной парадигме прикладной статистики, теория измерений является неотъемлемой частью методов анализа данных. По мнению П.Ф. Веллемана и Л. Уилкинсона [5], применение теории измерений «при выборе или для рекомендации тех или иных методов статистического анализа неуместно и зачастую приводит к ошибкам». В статье приведены краткие сведения о роли шкал измерения и применении теории измерений при выборе средних величин, а затем анализируются аргументы П.Ф. Веллемана и Л. Уилкинсона против использования этой теории «при выборе или для рекомендации тех или иных методов статистического анализа». Уточняется ряд вопросов применения прикладной статистики (анализа данных): продемонстрировано влияние поставленной задачи и применяемой модели данных на определение шкал измерения этих данных; разделены области применения разведочного анализа и доказательной статистики.

Александр Иванович Орлов – профессор, доктор экономических наук, доктор технических наук, кандидат физико-математических наук, директор Института высоких статистических технологий и эконометрики МГТУ им. Н.Э. Баумана, профессор МФТИ, советник президента группы авиакомпаний «Волга-Днепр», президент Российской ассоциации статистических методов. E-mail: prof-orlov@mail.ru .

[1] Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации в рамках Постановления Правительства Российской Федерации № 218.

Ключевые слова: теория измерений, анализ данных, прикладная статистика, шкалы измерения, допустимые преобразования, инвариантность выводов.

Методы анализа данных (или прикладная статистика, статистические методы) необходимы социологу для обработки результатов массовых обследований, анализа статистических данных и результатов экспериментов, а также для подведения итогов экспертных опросов [1]. В наши дни эта научная область бурно развивается. Согласно новой парадигме прикладной статистики теория измерений – неотъемлемая часть современных методов анализа данных [2]. В наших учебниках [3; 4 и др.] уделено внимание теории измерений и ее применению при выборе адекватных методов анализа данных.

Есть и другие мнения о целесообразности использования теории измерений при анализе социологических данных. Основная идея статьи П.Ф. Веллемана и Л. Уилкинсона [5] озвучена в ее названии. По их мнению, применение теории измерений «при выборе или для рекомендации тех или иных методов статистического анализа неуместно и зачастую приводит к ошибкам» [5, с. 167].

Прежде чем обратиться к аргументам П.Ф. Веллемана и Л. Уилкинсона, целесообразно определить предмет дискуссии, в частности, используемые нами термины и сформулировать основные положения в стиле отечественной вероятностно-статистической школы, чей основоположник – А.Н. Колмогоров – превратил теорию вероятностей и математическую статистику в раздел математики. При этом мы уточняем изложение в [5, с. 168–172] и описываем способ применения теории измерений в теории средних величин, что позволило систематизировать последнюю.

Основы теории измерений

Теория измерений исходит из того, что арифметические действия с используемыми в практической работе числами не всегда имеют смысл. Например, зачем складывать или умножать номера телефонов? Далее, не всегда выполнены привычные арифметические соотношения. Например, сумма знаний двух двоечников не равна знаниям «хорошиста», т.е. для оценок знаний 2+2 не равно 4. Эти примеры призваны доказать, что практика использования чисел для описания результатов наблюдений (измерений, испытаний, анализов, опытов) заслуживает методологического анализа.

Основные шкалы измерения. Самый простой способ использования чисел – применение их для различения объектов. Например, телефонные номера нужны для того, чтобы отличать одного абонента от другого. При таком способе измерения используется только одно отношение между числами – равенство (два объекта описываются либо равными, либо различными числами). Соответствующую шкалу измерения называют шкалой наименований (при использовании термина на основе латыни – номинальной шкалой; иногда называют также классификационной шкалой). В этой шкале измерены штрих-коды товаров, номера паспортов, ИНН (индивидуальные номера налогоплательщиков) и многие иные величины, выраженные числами. С прикладной точки зрения шкала измерения – это способ приписывания чисел рассматриваемым объектам, соответствующий отношениям между объектами.

Отметим, что числа могут быть приписаны объектам разными способами. Переход от одного способа к другому мы наблюдаем, например, при замене паспортов или телефонных номеров. Каковы свойства допустимых преобразований? Для шкалы наименований естественно потребовать только взаимной однозначности. Другими словами, применив к результатам измерений взаимно однозначное преобразование, получаем новую шкалу, столь же хорошо описывающую систему исходных объектов, как и прежняя.

Шесть основных типов шкал измерения описаны в *табл.* 1.

Таблица 1

ОСНОВНЫЕ ШКАЛЫ ИЗМЕРЕНИЯ

Шкала		Примеры	Группа допустимых преобразований $\Phi = \{\varphi\}$
Тип	Определение		
Шкалы качественных признаков			
Наименований	Числа используют для различения объектов	Номера телефонов, паспортов, ИНН, штрих-коды	Все взаимно-однозначные преобразования
Порядковая	Числа используют для упорядочения объектов	Оценки экспертов, баллы ветров, отметки в школе, полезность, номера домов	Все строго возрастающие преобразования
Шкалы количественных признаков (описываются началом отсчета и единицей измерения)			
Интервалов	Начало отсчета и единица измерения произвольны	Потенциальная энергия, положение точки, температура по шкалам Цельсия и Фаренгейта	Все линейные преобразования $\varphi(x) = ax + b$, a и b произвольны, $a>0$
Отношений	Начало отсчета задано, единица измерения произвольна	Масса, длина, мощность, напряжение, сопротивление, температура по Кельвину, цены	Все подобные преобразования $\varphi(x) = ax$, a произвольно, $a>0$
Разностей	Начало отсчета произвольно, единица измерения задана	Время	Все преобразования сдвига $\varphi(x) = x + b$, b произвольно
Абсолютная	Начало отсчета и единица измерения заданы	Число людей в данном помещении	Только тождественное преобразование $\varphi(x) = x$

Кроме перечисленных в *табл.* 1, используют и иные типы шкал [6]. Отметим, что в *табл.* 1 выражение «единица измерения произвольна» означает, что эта единица может быть выбрана по соглашению специалистов, но не вытекает из каких-либо фундаментальных соотношений. При измерении времени естественная единица измерения задается периодами обращения небесных тел. Начало отсчета при измерении длины задается, скажем, длиной отрезка, у которого начало и конец совпадают.

В наши дни считается, что перед применением тех или иных алгоритмов анализа данных необходимо установить, в шкалах каких типов измерены рассматриваемые величины. При этом с течением времени тип шкалы измерения определенной величины может меняться. Например, температура сначала измерялась в порядковой шкале (теплее – холоднее), а после изобретения термометров – в шкале интервалов (по шкалам Цельсия, Фаренгейта или Реомюра). Температура C по шкале Цельсия выражается через температуру F по шкале Фаренгейта с помощью линейного преобразования

$$C = \frac{5}{9}(F - 32).$$

С открытием абсолютного нуля температур стал возможным переход к шкале отношений (шкала Кельвина).

Требование инвариантности (адекватности) выводов. Выяснить тип используемой шкалы необходимо, чтобы адекватно выбрать метод анализа данных. Основополагающее требование – выводы не должны зависеть от того, какой именно шкалой измерения (среди всех шкал, переходящих друг в друга при допустимых преобразованиях) воспользовался исследователь. Например, если речь о длинах, то выводы не должны зависеть от единиц, в которых они измерены, все равно – в метрах, аршинах, саженях, футах или дюймах. Другими словами, выводы должны быть инвариантны относительно группы допустимых преобразований шкалы измерения. Только тогда их можно назвать адекватными, т.е. избавленными от субъективизма исследователя, выбирающего определенную шкалу из множества шкал заданного типа, связанных допустимыми преобразованиями.

Требование инвариантности выводов накладывает ограничения на множество возможных алгоритмов анализа данных. Для примера рассмотрим порядковую шкалу. Одни алгоритмы анализа данных позволяют получать адекватные выводы, другие – нет. Например, в задаче проверки однородности двух независимых выборок алгоритмы ранговой статистики (т.е. использующие только ранги результатов измерений) дают адекватные выводы, а статистики Крамера-Уэлча и Стьюдента – нет. Значит, для обработки данных, измеренных в порядковой шкале, критерии Смирнова и Вилкоксона использовать можно, а критерии Крамера-Уэлча и Стьюдента – нельзя.

Выбор средних величин в соответствии со шкалами измерения

Требование инвариантности считается достаточно сильным. Из многих алгоритмов анализа статистических данных ему удовлетворяют лишь некоторые. Покажем это на примере сравнения средних величин.

Средние по Коши. Среди всех методов анализа данных важное место занимают алгоритмы усреднения. В 1970-х гг. удалось выяснить, какими видами средних можно пользоваться при анализе данных, измеренных в тех или иных шкалах.

Пусть $X_1, X_2, \ldots, X_n$ – выборка объема n. Наиболее общее понятие средней величины введено французским математиком первой половины XIX в. О. Коши. Средняя величина (по Коши) – это любая функция $f(X_1, X_2, ..., X_n)$, такая, что при всех возможных значениях аргументов значение этой функции не меньше чем минимальное из чисел $X_1, X_2, ..., X_n$ и не больше чем максимальное из этих чисел. Среднее по Коши есть среднее арифметическое, медиана, мода, среднее геометрическое, среднее гармоническое, среднее квадратическое.

Средние величины используются обычно для того, чтобы заменить совокупность чисел (выборку) одним числом, а затем сравнивать совокупности с помощью средних. Пусть, например, $Y_1, Y_2, ..., Y_n$ – совокупность оценок экспертов (или респондентов), «выставленных» одному объекту экспертизы, $Z_1, Z_2, ..., Z_n$ – второму. Как сравнивать эти совокупности? Самый простой способ – по средним значениям.

При допустимом преобразовании шкалы значение средней величины, очевидно, меняется. Но выводы о том, для какой совокупности среднее больше, а для какой – меньше, не должны меняться (в соответствии с требованием инвариантности выводов, принятом как основное требование в теории измерений). Сформулируем соответствующую математическую задачу поиска вида средних величин, результат сравнения которых устойчив относительно допустимых преобразований шкалы.

Пусть $f(X_1, X_2, ..., X_n)$ – среднее по Коши. Пусть среднее по первой совокупности меньше среднего по второй совокупности:

$$f(Y_1, Y_2, ..., Y_n) < f(Z_1, Z_2, ..., Z_n).$$

Тогда, согласно теории измерений, для устойчивости результата сравнения средних необходимо, чтобы для любого допустимого преобразования (из группы допустимых преобразований в соответствующей шкале) было справедливо также неравенство

$$f(g(Y_1), g(Y_2), ..., g(Y_n)) < f(g(Z_1), g(Z_2), ..., g(Z_n)),$$

т.е. среднее преобразованных значений из первой совокупности было меньше среднего преобразованных значений для второй совокупности. Причем сформулированное условие должно быть выполнено для любых двух совокупностей $Y_1, Y_2, ..., Y_n$ и $Z_1, Z_2, ..., Z_n$. И, напомним, для любого допустимого преобразования. Средние величины, удовлетворяющие сформулированному условию, назовем *допустимыми* (в соответствующей шкале). Согласно теории измерений при анализе мнений экспертов и иных данных, измеренных в рассматриваемой шкале, можно пользоваться, только допустимыми средними величинами.

С помощью математической теории, развитой в [7], удается описать вид допустимых средних величин в основных шкалах.

Средние величины в порядковой шкале. Для определенности рассмотрим обработку мнений экспертов, измеренных в порядковой шкале. Справедливо следующее утверждение.

Теорема 1. Из всех средних по Коши допустимыми средними в порядковой шкале являются только члены вариационного ряда (порядковые статистики).

Теорема 1, впервые полученная в статье [8], справедлива при условии, что среднее $f(X_1, X_2, ..., X_n)$ является непрерывной (по совокупности переменных) и симметрической функцией. Последнее означает, что при перестановке аргументов значение функции $f(X_1, X_2, ..., X_n)$ не меняется. Это условие вполне естественно, ибо среднюю величину находим для *совокупности (множества)* чисел, а не для *последовательности*. Множество не меняется в зависимости от того, в какой последовательности мы перечисляем его элементы.

Согласно теореме 1 в качестве среднего для данных, измеренных в порядковой шкале, можно использовать в частности медиану (при нечетном объеме выборки). При четном же объеме следует применять один из двух центральных членов вариационного ряда – как их иногда называют, левую медиану или правую медиану. Моду тоже можно использовать – она всегда остается членом вариационного ряда. Можно применять, например, выборочные квартили, минимум и максимум, децили. Но никогда нельзя рассчитывать среднее арифметическое, среднее геометрическое и т.д.

Средние по Колмогорову. Естественная система аксиом (требований к средним величинам) приводит к так называемым ассоциативным средним. Их общий вид нашел в 1930 г. А.Н. Колмогоров [9]. Теперь их называют «средними по Колмогорову».

Для чисел $X_1, X_2, ..., X_n$ среднее по Колмогорову – это

$$G\{(F(X_1) + F(X_2) + ... + F(X_n))/n\},$$

где F – строго монотонная функция (т.е. строго возрастающая или строго убывающая), G – функция, обратная к F. Среди средних по Колмогорову – много хорошо известных персонажей. Так, если $F(x) = x$, то среднее по Колмогорову – это среднее арифметическое, если $F(x) = \ln x$, то среднее геометрическое, если $F(x) = 1/x$, то среднее гармоническое, если $F(x) = x^2$, то среднее квадратическое, и т.д. (в последних трех случаях усредняются положительные величины).

С одной стороны, среднее по Колмогорову – частный случай среднего по Коши. С другой – такие популярные средние, как медиана и мода, нельзя представить в виде средних по Колмогорову. В статье [10] впервые доказаны следующие утверждения.

Теорема 2. В шкале интервалов из всех средних по Колмогорову допустимо только среднее арифметическое.

Таким образом, среднее геометрическое или среднее квадратическое температур (в шкале Цельсия), потенциальных энергий или координат точек не имеют смысла. В качестве среднего надо применять среднее арифметическое. А также можно использовать медиану или моду.

Теорема 3. В шкале отношений из всех средних по Колмогорову допустимыми являются только степенные средние с $F(x) = x^c$, $c \neq 0$, и среднее геометрическое.

Есть ли средние по Колмогорову, которыми нельзя пользоваться в шкале отношений? Конечно, есть. Например, с $F(x) = e^x$.

Замечание 1. Среднее геометрическое является пределом степенных средних при $c \to 0$.

Замечание 2. Теоремы 1 и 2 справедливы при выполнении некоторых внутриматематических условий регулярности. Доказательства теорем 1–3 приведены в [7]. Перенос на случай взвешенных средних дан в [11].

Аналогично средним величинам могут быть изучены и другие статистические характеристики – показатели разброса, связи, расстояния и др. (см., например: [6; 7]). Не трудно показать, например, что коэффициент корреляции при любом допустимом преобразовании в шкале интервалов не меняется, как и отношение дисперсий. Дисперсия не меняется в шкале разностей, коэффициент вариации – в шкале отношений, и т.д. В [12] рассмотрены дальнейшие результаты о средних величинах.

В русле рассматриваемого подхода сначала надо установить, в каких шкалах измерены социологические данные, а затем использовать только алгоритмы обработки данных, инвариантные относительно этих шкал.

В [5] теория измерений обозначена как «ограничения Стивенса», порядковая шкала названа ординальной, шкала отношений – относительной, нет понятия «группа допустимых преобразований». Будем пользоваться устоявшимися в прикладной статистике терминами [4]. В целом же позиция сторонников использования теории измерений при анализе данных верно описана в [5].

На русском языке вышло в свет достаточно много публикаций, посвященных теории измерений, написанных строго, квалифицированными авторами. Поскольку мы не ставим целью дать здесь обзор по теории измерений, отошлем читателей к [6; 7; 12] и имеющимся там ссылкам на литературные источники.

Первые размышления над переводом статьи П.Ф. Веллемана и Л. Уилкинсона

Эта статья написана в виде обзора различных публикаций, изложение идет на словесном уровне, строгие определения, формулы, таблицы, примеры почти отсутствуют. Поэтому приходится додумывать за авторов. Однако не всегда удается придать точный смысл их высказываниям.

В статье [5, с. 173] выделено три направления критики.

1. Требование инвариантности выводов относительно допустимых преобразований шкал измерения «представляется опасным для анализа данных».

2. Подход на основе теории измерений «слишком строг, чтобы его можно было применять для реальных данных».

3. Этот подход «часто ведет к понижению уровня данных через их преобразования в ранги и последующее ненужное обращение к непараметрическим методам».

Начнем с разбора в общих терминах этих трех направлений критики.

1. Для получения обоснованных выводов, наоборот, опасен отказ от требования инвариантности. Разве можно опираться на выводы, которые меняются при допустимом преобразовании шкалы?

Конечно, при первоначальном разведочном анализе данных можно их «прогнать» через весь арсенал методов обработки, имеющихся в программном продукте, – вдруг удастся что-нибудь интересное заметить? Полученные нестрогими методами «находки» необходимо затем проверить с помощью обоснованных процедур анализа данных [3].

Практика зачастую вынуждает использовать соображения теории измерений. Так, при проведении опросов летного состава авиакомпании «Волга-Днепр» выяснилось, что пилотам легче сказать, какое событие встречается чаще, а какое реже, чем оценить число событий на 1000 полетов. Оценивать в абсолютной шкале (вероятность событий) пилоты не берутся, в то время как задача сравнения встречаемости событий по частоте или оценка встречаемости условными баллами (значениями качественных признаков) не вызывают трудностей. Таким образом, полученные при опросах пилотов оценки измерены в порядковых шкалах.

2. При практической работе обычно вполне ясно, в каких шкалах измерены данные. Если попытаться навязать респондентам «неправильную» шкалу, их ответы станут произвольными, не отражающими истинных мнений, или же они могут попросту отказаться давать ответы, как это было при опросах летного состава авиакомпании «Волга-Днепр».

Можно признать, что только в редких случаях определение типа шкалы измерения данных требует специальных исследований.

3. Уже ко времени появления статьи П.Ф. Веллемана и Л. Уилкинсона (1993) с помощью непараметрических методов можно было решать все те задачи анализа данных, для которых все еще порой используются параметрические методы. Современная парадигма прикладной статистики [2] вместо параметрических методов, характерных для устаревшей парадигмы середины XX в., предлагает применять непараметрические методы.

Согласно современным взглядам, параметрические методы основаны на вероятностно-статистических моделях, в которых распределения случайных величин принадлежат тому или иному из параметрических семейств – семейству нормальных, логарифмически-нормальных, гамма-распределений или иных, входящих в 4-параметрическое семейство К. Пирсона, введенное им в начале XX в. Непараметрические методы исходят из распределений произвольного вида. «Преобразование в ранги» [5] не обязательно при применении непараметрических методов. Оно соответствует случаю, когда данные измерены в порядковой шкале.

Как показали многочисленные исследования, почти все распределения реальных данных не принадлежат ни одному из известных параметрических семейств [3]. Боязни непараметрических методов невозможно найти рационального обоснования, она порождена предрассудками устаревшей парадигмы прикладной статистики середины XX в.

От общих возражений против применения теории измерений при анализе социологических данных перейдем к рассмотрению конкретных примеров, приведенных П.Ф. Веллеманом и Л. Уилкинсоном. Не станем здесь повторять формулировки примеров, предполагая, что читатели могут познакомиться с переводом их исходной статьи [5].

В критике Лорда [5, с. 173] выделим несколько составляющих. Во-первых, выбор типа шкалы может быть связан с решаемой задачей. Так, номера договоров предприятия служат прежде всего для того, чтобы различать эти договора (и связанные с ними действия), т.е. естественно принять, что они измерены в шкале наименований. Однако эти номера возрастают с течением времени (в соответствии с датами заключения договоров), поэтому в некоторых задачах принятия управленческих решений естественно считать, что они измерены в порядковой шкале. Во-вторых, при обработке порядковых данных с помощью алгоритмов, не являющихся инвариантными в порядковой шкале, может создаться впечатление, что получены обоснованные выводы. Лорд рассказывает о применении неравенства Чебышева (можно было использовать критерий Крамера-Уэлча [3]). Однако при применении той же процедуры анализа к данным, подвергнутым некоторому допустимому преобразованию в порядковой шкале, выводы будут прямо противоположными. Для обнаружения различия между двумя независимыми выборками следовало применить непараметрические критерии однородности, например критерий Вилкоксона [3].

Бейкер, Хардик и Петринович, Боргатта и Борштейн [5, с. 173–174] не хотят применять непараметрические методы, объяснений нет. Веллеман и Уилкинсон напрасно критикуют их за нежелание «связываться с проблемой робастности» [5, с. 174]. Робастные, т.е. устойчивые к малым отклонениям функций распределения данных, методы не позволяют справиться с произвольными допустимыми преобразованиями. Если же от робастности перейти к более общей системе понятий – к общей схеме устойчивости, то оказывается, что устойчивые к допустимым преобразованиям шкал методы анализа данных – это ранговые методы как частный случай непараметрических [4; 7].

Гутман предлагает использовать «функцию потерь, выбранную для проверки качества модели» [5, с. 174]. Действительно, если задана функция потерь, то нет необходимости привлекать теорию измерений. Проблема в том, чтобы выбрать эту функцию, причем обоснованно. Не с одним таким практиком за более чем 40 лет консультирования в области анализа данных мне встретиться не довелось. Тот, кто сможет выбрать функцию потерь, уже не практик, а квалифицированный специалист в области математической статистики.

Тьюки задает вопрос: «Какое знание не основано на некоторой приблизительности?» [5, с. 175] Действительно, при первоначальном разведочном анализе специалисту достаточно одного взгляда для формулировки вывода. Однако и практики, и теоретики настаивают на том, чтобы интуитивные выводы были обоснованы строгими рассуждениями.

Дискуссия о статистиках и шкальных типах

Названный так раздел [5, с. 175–177] начинается словами: «Статистики отвергли запрет на методы, основанный на ограничениях, связанных с допустимыми преобразованиями». Это совершенно неверно. Статистики приняли этот запрет (см. обсуждения в: [1–8; 10–12]). Особенно ясно это сегодня, спустя 20 лет после написания статьи [5]. В настоящее время сомнения остаются у некоторых из тех, кто не является профессионалом в области анализа данных, к тому же склонен к принятию простых решений и не хочет утруждать себя изучением теории измерений и непараметрической статистики. Такой настрой практиков вполне естествен и разумен, но не плодотворен. Современную прикладную статистику [1; 3; 4] трудно признать простой, для ее усвоения нужно приложить усилия и затратить время.

Приходится констатировать, что в статью [5] включено большое количество категоричных утверждений, не подтвержденных аргументами и противоречащих практике анализа данных. Например, «ключевой аргумент против использования предписания статистик на основе шкального типа гласит: это не работает!» [5, с. 176]. Еще как работает – и на практике, и при развитии теории (в начальных разделах настоящей статьи показано, что теория измерений позволила придать теории средних законченный вид). Или: «опыт показывает, что применение запрещенных статистик к данным приводит к научно значимым результатам, важным при принятии решений и ценным для дальнейших исследований» [5, с. 177]. Примеров нет. Видимо, потому, что это утверждение неверно.

В [5] часто используются термины без определений. Отечественного читателя может поразить заявление о «фундаментальной разнице между математикой и наукой» [5, с. 176][1]. В нашей стране согласно традиции и нормативным документам Минобразования и ВАК математика – одна из наук. Мы считаем, что статистические методы и анализ данных – это одно и то же. Именно поэтому наша недавняя книга называется «Статистические методы анализа данных» [13]. Конечно, можно определить термины так, что математика не будет наукой, а анализ данных станет отличаться от математической статистики. Дискуссия о терминах – увлекательное занятие. Только в одной брошюре [14] приведено около 200 определений термина «статистика». Однако ясно, что использование терминов без определений, как это сделано в [5, с. 176], может только запутать читателя.

Различные виды данных

Нельзя не согласиться с Веллеманом и Уилкинсоном, что данные – это не всегда числа [5, с. 177–179]. Элементами выборок мо-

[1] В соответствующем фрагменте речь идет о различии между математикой и эмпирической наукой, и цитируется работа Тьюки: «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”».

гут быть вектора, функции, различные виды объектов нечисловой природы – бинарные отношения, множества, нечеткие множества, интервалы и др. [3, 4] Тем более это касается результатов расчетов, таких как доли или набор точек на плоскости, полученных в результате многомерного шкалирования. Обратите внимание: при рассказе о применении теории измерений при анализе данных в начале этой статьи шла речь об инвариантности выводов, сделанных на основе обработки наборов чисел. Следовательно, теория измерений используется не во всех разделах прикладной статистики, а лишь при статистическом анализе числовых величин [3, гл. 8]. Это замечание понадобится при дальнейшем разборе статьи [5].

Необходимо всегда различать разведочный статистический анализ, нацеленный на «интуитивное проникновение в закономерности массива данных» [5, с. 179], и доказательную статистику, основанную на строгих рассуждениях. Именно к разведочному анализу относятся методы преобразования данных [5, с. 178] и многомерного шкалирования [5, с. 179]. При разведочном анализе соблюдать требования теории измерений не обязательно, а в доказательной статистике – наоборот.

В разделе «Хороший анализ данных не основан на допущениях о типе данных» [5, с. 179–180] Веллеман и Уилкинсон справедливо обращают внимание на важность правильного выбора статистической модели. В следующем разделе «Стивенсовские категории не описывают фиксированных свойств данных» [5, с. 180–182] речь фактически идет о том же: в ряде ситуаций «шкальный тип зависит от интерпретации данных или от наличия дополнительной информации» [5, с. 180]. Это утверждение совершенно верно, набор чисел сам по себе не дает возможности обосновать тип шкалы. Результат измерения равен 2911397 – какая шкала? Если это число из бухгалтерского отчета, то шкала отношений (переход от одной валюты к другой – подобное преобразование). Если же это число из телефонного справочника, то номер телефона измерен в шкале наименований. На эту тему мы говорили ранее в связи с разбором работы Лорда [5, с. 173]. Итак, весьма важен выбор статистической модели, им определяются шкалы измерения данных.

В разделе «Категории Стивенса недостаточны для описания шкал данных» [5, с. 182–183] рассматриваются «многомерные шкалы». Что это такое – неясно, так как определений нет. Однако квазипрактический пример, заданный *табл.* 1 в [5], достаточно понятен. Поскольку я пять лет проработал в медицинских учреждениях (в «Кремлевской больнице» и в НИИ профессиональных заболеваний и гигиены труда АМН СССР), то отмечу, что число наблюдающихся у пациента симптомов нельзя рассматривать как показатель тяжести заболевания, поскольку подобное рассмотрение предполагает, что все симптомы равноценны по вкладу в тяжесть заболевания. Такого в медицине не бывает.

О чем идет речь в абзаце, посвященном работе Андерсона [5, с. 183], остается для меня неясным, поскольку определений используемых понятий нет.

Робастность, шкалы и анализ данных

В разделе «Статистические процедуры не могут классифицироваться по критериям Стивенса» [5, с. 183–184] Веллеман и Уилкинсон обсуждают обратную задачу (в терминологии [7; 11; 12]), где требуется установить, в каких шкалах заданная процедура позволяет получить инвариантные выводы. Действительно, нами доказано, что вывод о сравнении рассчитанных по двум выборкам значений линейной функции от порядковых статистик, заданной формулой (5) на с. 185 [5], инвариантен в порядковой шкале, если только один весовой коэффициент отличен от 0 (см.: [8] и теорему 1 в этой статье), и в шкале интервалов (и в шкалах с более узкими группами преобразований – отношений, разностей, абсолютной), если по крайней мере два весовых коэффициента отличны от 0 (см. [11]). Остальной текст этого раздела статьи [5] не поддается интерпретации в строгих терминах. Отметим только, что рассматривается иная задача чем раньше – увязка процедур расчетов со шкалами измерения, а не установление типа шкалы измерения исходных данных.

В разделе «Шкальные типы – не точные категории» [5, с. 184–185] в очередной раз бездоказательно утверждается, что «реальные данные не удовлетворяют требованиям шкальных типов». Вместе с тем правильно отмечено, что при сомнениях «следует осуществить понижение уровня» шкалы, например с интервальной до порядковой. В задаче, рассмотренной Тьюки в 1961 г., была бы полезна статистика интервальных данных, развиваемая с начала 1980-х гг. [3; 4].

В разделе «Шкалы и анализ данных» [5, с. 185–186] рассуждения построены на смешении разведочного статистического анализа, когда можно не обращать внимания на шкалы, в которых измерены данные, и анализа данных на стадии получения строгих выводов, немыслимых без обращения к теории измерений. Странно, что Веллеман и Уилкинсон считают «хорошим» только разведочный анализ. Утверждение: «Хороший анализ данных редко следует формальной парадигме проверки гипотезы» [5, с. 186] демонстрирует их нигилизм по отношению к математической статистике, который никак нельзя оправдать.

В разделе «Осмысленность» [5, с. 186–188] термин, давший название разделу, так и остался без определения. Как справедливо отмечают Веллеман и Уилкинсон, согласно теории измерений осмысленность – это то, что сохраняется при допустимых преобразованиях. Такое определение им не нравится, но дать другое они не могут, занимаясь общими рассуждениями о праве на ошибку. Странно читать такое: «Если бы наука была ограничена доказуемо осмысленными суждениями, она не смогла бы развиваться». Математика же успешно развивается!

Раздел «Роль типов данных» [5, с. 188–189] начинается неожиданно – с признания важности теории измерений: «Были бы ошибкой полагать, что типы данных не имеют значения…

Понятие типа шкалы важно, а терминология Стивенса [т.е. теории измерений. – *А.О.*] зачастую бывает удобна». Дальнейшие рассуждения Веллемана и Уилкинсона снова посвящены констатации: тип шкалы определяется не самими данными, а моделью, соответствующей решаемой задаче. Вторая идея, которая также уже встречалась, – упор на разведочный анализ и умаление роли доказательной статистики.

Заключение

Раздел «Заключение» статьи [5, с. 190] написан взвешенно, высказанные в нем положения в целом справедливы. Как уже говорилось, нельзя считать, «что тип шкалы как бы самоочевиден и не зависит от того, какой вопрос ставит исследователь перед своими данными». За 20 лет, прошедших после написания статьи [5], стало ясно, что поставив вопрос, исследователь должен описать модель анализа данных, обычно вероятностно-статистическую, включающую выбор типа шкал измерения данных, а затем в рамках этой модели разработать метод решения задачи или выбрать его из уже имеющихся [3].

Совершенно верно, что «статистическое программное обеспечение, способствующее любому анализу для любых данных, допускает и безответственный анализ». Об этом предупреждал В.В. Налимов более 40 лет назад [15]. Он имел в виду прежде всего склонность к проведению расчетов без знакомства с сутью применяемых методов.

Итак, необходимо констатировать пользу от сопоставления подходов теории измерений и критических замечаний по ее поводу, собранных в статье Веллемана и Уилкинсона [5]. Дискуссия позволила уточнить ряд вопросов применения прикладной статистики (анализа данных). Прежде всего выявлена роль решаемой задачи и применяемой модели данных для установления типов шкал измерения этих данных, разделены области применения разведочного анализа и доказательной статистики. Подтвердилась справедливость пословицы: «В споре рождается истина».

ЛИТЕРАТУРА

1. *Орлов А.И.* Статистические методы в российской социологии: тридцать лет спустя // Социология: методология, методы, математические модели. 2005. № 20. С.32–53.

2. *Орлов А.И.* Новая парадигма прикладной статистики // Заводская лаборатория. 2012. Т. 78. №1. Ч. I. С. 87–93.

3. Орлов А.И. Прикладная статистика: учебник. М.: Экзамен, 2006.

4. *Орлов А.И.* Организационно-экономическое моделирование: учебник: В 3 ч. Ч. 1: Нечисловая статистика. М.: Изд-во МГТУ им. Н.Э. Баумана, 2009.

5. *Веллеман П.Ф., Уилкинсон Л.* Типология номинальных, ординальных, интервальных и относительных шкал вводит в заблуждение // Социология: методология, методы, математическое моделирование. 2011. № 33. С. 166–193.

6. *Толстова Ю.Н.* Измерения в социологии. М.: Инфра-М, 1998.

7. *Орлов А.И.* Устойчивость в социально-экономических моделях. М.: Наука, 1979.

8. *Орлов А.И.* Допустимые средние в некоторых задачах экспертных оценок и агрегирования показателей качества // Многомерный статистический анализ в социально-экономических исследованиях. М.: Наука, 1974. С. 388–393.

9. *Колмогоров А.Н.* Об определении среднего // Избр. труды: математика и механика. М.: Наука, 1985. С. 136–138.

10. *Орлов А.И.* Допустимые преобразования в задаче сравнения средних: пси-постоянные статистики // Алгоритмы многомерного статистического анализа и их применения. М.: Изд-во ЦЭМИ АН СССР, 1975. С. 121–127.

11. *Орлов А.И.* Связь между средними величинами и допустимыми преобразованиями шкалы // Математические заметки. 1981. Т. 30. № 4. С. 561–568.

12. *Барский Б.В., Соколов М.В.* Средние величины, инвариантные относительно допустимых преобразований шкалы измерения // Заводская лаборатория. 2006. Т. 72. № 1. С.59–66.

13. *Орлов А.И.* Организационно-экономическое моделирование: учебник: В 3 ч. Ч. 3: Статистические методы анализа данных. М.: Изд-во МГТУ им. Н.Э. Баумана, 2012.

14. *Никитина Е.П., Фрейдлина В.Д., Ярхо А.В.* Коллекция определений термина «статистика». М.: МГУ, 1972.

15. *Налимов В.В.* О преподавании математики экспериментаторам // О преподавании математической статистики экспериментаторам. Препринт межфакультетской лаборатории статистических методов №17. М.: Изд-во МГУ им. М.В. Ломоносова, 1971. С. 5–39.