УДК 519.2:303.732.4

МНОГООБРАЗИЕ ОБЪЕКТОВ НЕЧИСЛОВОЙ ПРИРОДЫ

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор

Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru

В соответствии с новой парадигмой математической статистики статистика объектов нечисловой природы (статистика нечисловых данных, нечисловая статистика) является одним из четырех основных областей математической статистики. Статистика объектов нечисловой природы состоит из центрального ядра – статистики в пространствах произвольной природы – и статистических теорий анализа конкретных видов нечисловых данных. Для выявления прикладных возможностей статистики объектов нечисловой природы целесообразно изучить многообразие объектов нечисловой природы. Этому и посвящена настоящая статья. Рассмотрены результаты измерений в шкалах, отличных от абсолютной; бинарные отношения; дихотомические (бинарные) данные; множества. Проанализированы объекты нечисловой природы как статистические данные, их значение при формировании статистической или математической модели реального явления, в качестве результата анализа данных

Ключевые слова: МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ОБЪЕКТЫ НЕЧИСЛОВОЙ ПРИРОДЫ, ТЕОРИЯ ИЗМЕРЕНИЙ, БИНАРНЫЕ ОТНОШЕНИЯ, ДИХОТОМИЧЕСКИЕ ДАННЫЕ, МНОЖЕСТВА

UDC 519.2:303.732.4

MULTIFORMITY OF OBJECTS OF NON-NUMERICAL NATURE

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci., professor
Bauman Moscow State Technical University, Moscow, Russia

In accordance with the new paradigm of mathematical statistics the statistics of objects of nonnumerical nature (statistics of nonnumerical objects, non-numerical data statistics, non-numeric statistics) is one of the four main areas of mathematical statistics. Statistics of objects of nonnumerical nature consists of a central core - statistics in spaces of arbitrary nature - and statistical theories of analysis of specific types of non-numeric data. To identify possibilities of application of statistics of objects of nonnumerical nature it is useful to explore the multiformity of objects of non-numeric nature. This is the subject of this article. We have considered the results of measurements at scales other than absolute; binary relations; dichotomous (binary) data; sets. We have also analyzed the objects of non-numerical nature as statistical data, and their importance in the formation of statistical or mathematical model of a real phenomenon, as a result of data analysis

Keywords: MATHEMATICAL STATISTICS, NON-NUMERICAL OBJECTS, MEASUREMENT THEORY, BINARY RELATIONS, DICHOTOMOUS DATA, SETS

1. Введение

В соответствии с новой парадигмой математической статистики [1] статистика объектов нечисловой природы (статистика нечисловых данных, нечисловая статистика) является одним из четырех основных областей математической статистики. Обзор исследований в этой области дан в работах [2 - 4]. Конкретные результаты приведены в ряде работ. Так, в пространствах произвольной природы введены средние величины, для которых доказаны законы больших чисел [5]. Продемонстрировано, что один из видов объектов нечисловой природы - нечеткие множества –

можно рассматривать как проекции случайных множеств, следовательно, с этой точки зрения теория нечетких множеств - часть теории вероятностей [6]. Построена развитая математико-статистическая теория для такого вида объектов нечисловой природы, как интервальные данные [7]. Предложены и изучены оценки плотности распределения вероятностей в пространствах произвольной природы, в частности, доказана их состоятельность и оценена скорость сходимости в зависимости от свойств регулярности оцениваемой плотности [8]. Развита предельная теория непараметрических статистик, прежде всего статистик интегрального типа в пространствах произвольной природы [9]. Статистика объектов нечисловой природы имеет много точек соприкосновения с системной нечеткой интервальной математикой [10 - 12].

Таким образом, статистика объектов нечисловой природы состоит из центрального ядра – статистики в пространствах произвольной природы – и статистических теорий анализа конкретных видов нечисловых данных. Для выявления прикладных возможностей статистики объектов нечисловой природы целесообразно рассмотреть многообразие объектов нечисловой природы. Этому и посвящена настоящая статья.

2. Количественные и категоризованные данные

Статистические методы – это методы анализа данных, причем обычно достаточно большого количества данных. Статистические данные могут иметь различную природу. Исторически самыми ранними были два вида данных – сведения о числе объектов, удовлетворяющих тем или иным условиям, и числовые результаты измерений.

Первый из этих видов данных до сих пор главенствует в статистических сборниках Госкомстата РФ. Такого рода данные часто называют *категоризованными*, поскольку о каждом из рассматриваемых объектов известно, в какую из нескольких заранее заданных категорий он попадает. Примером является информация Госкомстата РФ о населении страны, с разделением по возрастным категориям и полу. Часто при составлении таблиц жертвуют информацией, заменяя точное значение измеряемой величины на указание интервала группировки, в которую это значение попадает. Например, вместо точного возраста человека используют лишь один из указанных в таблице возрастных интервалов. Статистические методы анализа сгруппированных данных рассмотрены в статье [13].

Второй наиболее распространенный вид данных – количественные данные, рассматриваемые как действительные числа. Таковы результаты измерений, наблюдений, испытаний, опытов, анализов. Количественные данные обычно описываются набором чисел (выборкой), а не таблицей.

Нельзя утверждать, что категоризованные данные соответствуют первому этапу исследования, а числовые – следующему, на котором используются более совершенные методы измерения. Дело в том, что человеку свойственно давать качественные ответы на возникающие в его практической деятельности вопросы.

Иногда нецелесообразно однозначно относить данные к категоризованным или количественным. Например, в Ветхом Завете, в Четвертой книге Моисеева «Числа» указывается количество воинов в различных коленах [14]. С одной стороны, это типичные категоризованные данные, градациями служат названия колен. С другой стороны, эти данные можно рассматривать как количественные, как выборку, их вполне естественно складывать, вычислять среднее арифметическое и т.п.

Описанная ситуация типична. Существует весьма много различных видов статистических данных. Это связано, в частности, со способами их получения. Например, если испытания некоторых технических устройств продолжаются до определенного момента, то получаем т.н. *цензурированные* данные, состоящие из набора чисел – продолжительности работы ряда устройств до отказа, и информации о том, что остальные устройства продолжали работать в момент окончания испытания. Такого рода данные часто используются при оценке и контроле надежности технических устройств.

Описание вида данных и, при необходимости, механизма их порождения – начало любого статистического исследования.

3. Четыре основные области математической статистики

В простейшем случае статистические данные – это значения некоторого признака, свойственного изучаемым объектам. Значения могут быть количественными или представлять собой указание на категорию, к которой можно отнести объект. Во втором случае говорят о качественном признаке. Используют и более сложные признаки, перечень которых будет расширяться по мере развертывания изложения в учебнике.

При измерении по нескольким количественным или качественным признакам в качестве статистических данных об объекте получаем вектор. Его можно рассматривать как новый вид данных. В таком случае выборка состоит из набора векторов. Есть часть координат – числа, а часть – качественные (категоризованные) данные, то говорим о векторе разнотипных данных.

Одним элементом выборки, т.е. одним измерением, может быть и функция в целом. Например, электрокардиограмма больного или амплитуда биений вала двигателя. Или временной ряд, описывающий динамику показателей определенной фирмы. Тогда выборка состоит из набора функций.

Элементами выборки могут быть и бинарные отношения. Например, при опросах экспертов часто используют упорядочения (ранжировки) объектов экспертизы – образцов продукции, инвестиционных проектов, вариантов управленческих решений. В зависимости от регламента экспертного исследования элементами выборки могут быть различные виды бинарных отношений (упорядочения, разбиения, толерантности), множества, нечеткие множества и т.д.

Итак, математическая природа элементов выборки в различных задачах прикладной статистики может быть самой разной. Однако можно выделить два класса статистических данных – числовые и нечисловые. Соответственно прикладная статистика разбивается на две части – числовую статистику и нечисловую статистику (ее называют также статистикой нечисловых данных или статистикой объектов нечисловой природы).

Числовые статистические данные – это числа, вектора, функции. Их можно складывать, умножать на коэффициенты. Поэтому в числовой статистике большое значение имеют разнообразные суммы. Математический аппарат анализа сумм случайных элементов выборки – это (классические) законы больших чисел и центральные предельные теоремы.

Нечисловые статистические данные – это категоризованные данные, вектора разнотипных признаков, бинарные отношения, множества, нечеткие множества и др. Их нельзя складывать и умножать на коэффициенты. Поэтому не имеет смысла говорить о суммах нечисловых статистических данных. Они являются элементами нечисловых математических пространств (множеств). Математический аппарат анализа нечисловых статистических данных основан на использовании расстояний между элементами (а также мер близости, показателей различия) в таких пространствах. С помощью расстояний определяются эмпирические и теоретические средние, доказываются законы больших чисел, строятся непараметрические оценки плотности распределения вероятностей, решаются задачи диагностики и кластерного анализа, и т.д.

Сведем информацию об основных областях математической статистики в табл. 1. Отметим, что модели порождения цензурированных данных входят в состав каждой из рассматриваемых областей.

Таблица 1. Области математической статистики

№ п/п	Вид статистических данных	Область прикладной статистики
1	Числа	Статистика (случайных) величин
2	Конечномерные вектора	Многомерный статистический анализ
3	Функции	Статистика случайных процессов и временных рядов
4	Объекты нечисловой природы	Нечисловая статистика

Статистика объектов нечисловой природы - это направление в математической статистике, в котором в качестве исходных статистических данных (результатов наблюдений) рассматриваются объекты нечисловой природы. Так принято называть объекты, которые нецелесообразно описывать числами, в частности элементы различных нелинейных пространств. Примерами являются бинарные отношения (ранжировки, разбиения, толерантности и др.), результаты парных и множественных сравнений, множества, нечеткие множества, измерение в шкалах, отличных от абсолютных. Этот перечень примеров не претендует на законченность. Он складывался постепенно, по мере того, как развивались теоретические исследования в области нечисловой статистики (статистики нечисловых данных) и расширялся опыт применений этого направления прикладной математической статистики.

Объекты нечисловой природы широко используются в теоретических и прикладных исследованиях по экономике, менеджменту и другим проблемам управления, в частности управления качеством продукции, в технических науках, социологии, психологии, медицине и т.д., а также практически во всех отраслях народного хозяйства.

4. Результаты измерений в шкалах, отличных от абсолютной.

Рассмотрим основные виды объектов нечисловой природы.

Теория измерений [15 - 18] является одной из составных частей математической статистики. Она входит в состав статистики объектов нечисловой природы. В соответствии с теорией измерений при математическом моделировании реального явления или процесса следует прежде всего установить типы шкал, в которых измерены те или иные переменные. Тип шкалы задает группу допустимых *преобразований шкалы*. Допустимые преобразования не меняют соотношений между объектами измерения. Например, при измерении длины переход от аршин к метрам не меняет соотношений между длинами рассматриваемых объектов - если первый объект длиннее второго, то это будет установлено и при измерении в аршинах, и при измерении в метрах. При этом численное значение длины в аршинах отличается от численного значения длины в метрах - не меняется лишь результат сравнения длин двух объектов.

Выделяют шесть основных видов шкал – наименований, порядковые, интервалов, отношений, разностей, абсолютные [19]. Статистические выводы должны быть инвариантны относительно допустимых преобразований тех шкал, в которых измерены обрабатываемые статистические данные. Это условие позволяет, например, указать вид средних величин, которыми можно пользоваться при анализе данных, измеренных в порядковой шкале, а также шкалах интервалов и отношений.

Рассмотрим конкретное исследование в области маркетинга образовательных услуг, послужившее поводом к развитию одного из направлений отечественных исследований по теории измерений [20]. При изучении привлекательности различных профессий для выпускников новосибирских школ был составлен список из 30 профессий. Опрашиваемых просили оценить каждую из этих профессий одним из баллов 1,2,...,10 по правилу: чем больше нравится, тем выше балл. Для получения социологических выводов необходимо было дать единую оценку привлекательности определенной профессии для совокупности выпускников школ. В качестве такой оценки в работе [21] использовалось среднее арифметическое баллов, выставленных профессии опрошенными школьниками. В частности, физика получила средний балл 7,69, а математика - 7,50. Поскольку 7,69 больше, чем 7,50, был сделан вывод, что физика более предпочтительна для школьников, чем математика.

Однако этот вывод противоречит данным работы [22], согласно которым ленинградские школьники средних классов больше любят математику, чем физику. Как объяснить это противоречие? Есть много подходов к выяснению причин различия выводов новосибирских и ленинградских исследователей. Здесь обсудим одно из возможных объяснений этого противоречия, основанное на идеях статистики объектов нечисловой природы. Оно сводится к указанию на неадекватность (с точки зрения теории измерений) методики обработки статистических данных о предпочтениях выпускников школ, примененной в работе [21].

Дело в том, что баллы 1, 2, ..., 10 введены конкретными исследователями, т.е. субъективно. Если одна профессия оценена в 10 баллов, а вторая - в 2, то из этого нельзя заключить, что первая ровно в 5 раз привлекательней второй. Другой коллектив социологов мог бы принять иную систему баллов, например 1, 4, 9, 16, ..., 100. Естественно предположить, что упорядочивание профессий по привлекательности, присущее школьникам, не зависит от того, какой системой баллов им предложит пользоваться социолог. Раз так, то распределение профессий по градациям десятибалльной системы не изменится, если перейти к другой системе баллов с помощью любого допустимого преобразования в порядковой шкале, т.е. с помощью строго возрастающей функции $g: R^1 \to R^1$. Если Y_1, Y_2, ... , Y_n - ответы n выпускников школ, касающиеся математики, а Z_1, Z_2, ... , Z_n - физики, то после перехода к новой системе баллов ответы относительно математики будут иметь вид $g(Y_1)$, $g(Y_2)$, ... , $g(Y_n)$, а относительно физики - $g(Z_1)$, $g(Z_2)$, ... , $g(Z_n)$.

Пусть единая оценка привлекательности профессии вычисляется с помощью функции $f(X_1, X_2, \dots, X_n)$. Какие требования естественно наложить на функцию $f: R^n \to R^1$, чтобы полученные с ее помощью выводы не зависели от того, какой именно системой баллов пользовался социолог (в рассматриваемом исследовании он выступал как специалист по маркетингу образовательных услуг)?

Замечание. Обсуждение можно вести в терминах экспертных оценок [23]. Тогда вместо сравнения математики и физики n экспертов (а не выпускников школ) оценивают по конкурентоспособности на мировом рынке, например, две марки стали. Однако в настоящее время маркетинговые и социологические исследования более привычны, чем экспертные.

Единая оценка вычислялась для того, чтобы сравнивать профессии по привлекательности. Пусть $f(X_1, X_2,...,X_n)$ - среднее по Коши[1]. Пусть среднее по первой совокупности меньше среднего по второй совокупности:

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

Тогда согласно теории измерений необходимо потребовать, чтобы для любого допустимого преобразования g из группы допустимых преобразований в порядковой шкале было справедливо также неравенство

[1] Среднее по Коши - любая функция $f(X_1, X_2,...,X_n)$ такая, что

$$\min(X_1, X_2,...,X_n) \le f(X_1, X_2,...,X_n) \le \max(X_1, X_2,...,X_n)$$

при всех $X_1, X_2,...,X_n$.

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

т.е. среднее преобразованных значений из первой совокупности также было меньше среднего преобразованных значений для второй совокупности. Более того, два рассматриваемых неравенства должны быть равносильны. Причем сформулированное условие должно быть верно для любых двух совокупностей $Y_1, Y_2, \dots, Y_n$ и $Z_1, Z_2, \dots, Z_n$ и, напомним, любого допустимого преобразования. Средние величины, удовлетворяющие сформулированному условию, называют допустимыми (в порядковой шкале). Согласно теории измерений только такими средними можно пользоваться при анализе мнений выпускников школ или экспертов, обработке иных данных, измеренных в порядковой шкале.

Какие единые оценки привлекательности профессий $f(X_1, X_2, \dots, X_n)$ устойчивы относительно сравнения? Ответ на этот вопрос был получен в [20]. В частности, оказалось, что средним арифметическим, как в работе [21] новосибирских социологов (специалистов по маркетингу образовательных услуг), пользоваться нельзя. А порядковыми статистиками, т.е. членами вариационного ряда (и только ими) - можно.

Методы анализа конкретных статистических данных, измеренных в шкалах, отличных от абсолютной, являются предметом изучения в статистике объектов нечисловой природы. Основные шкалы измерения делятся на качественные (шкалы наименований и порядка) и количественные (шкалы интервалов, отношений, разностей, абсолютная). Методы анализа статистических данных в количественных шкалах сравнительно мало отличаются от таковых в абсолютной шкале. Добавляется только требование инвариантности относительно преобразований сдвига и/или масштаба. Методы анализа качественных данных - принципиально иные.

Исходным понятием теории измерений является совокупность $\Phi = \{\varphi\}$ допустимых преобразований шкалы (обычно Ф - группа),

$\varphi: R^1 \to R^1$. Алгоритм обработки данных *W*, т.е. функция $W: R^n \to A$ (здесь *A*-множество возможных результатов работы алгоритма) называется адекватным в шкале с совокупностью допустимых преобразований Ф, если

$$W(x_1, x_2, ..., x_n) = W(\varphi(x_1), \varphi(x_2), ..., \varphi(x_n)) \qquad (1)$$

для всех $x_i \in R^1$, $i = 1,2,...,n$, и всех $\varphi \in \Phi$. Таким образом, теорию измерений рассматриваем как теорию инвариантов относительно различных совокупностей допустимых преобразований Ф. Интерес вызывают две задачи:

а) дана группа допустимых преобразований Ф (т.е. задана шкала измерения); какие алгоритмы анализа данных *W* из определенного класса являются адекватными (т.е удовлетворяют тождеству (1))?

б) дан алгоритм анализа данных *W*; для каких шкал (т.е. групп допустимых преобразований Ф) он является адекватным?

В [19, разд. 3.1] первая задача рассматривается для алгоритмов расчета средних величин. Информацию о других результатах решения задач указанных типов можно найти в работах [2, 20, 24].

5. Бинарные отношения

Пусть $W: R^n \to A$ - адекватный алгоритм в шкале наименований. Можно показать, что этот алгоритм задается некоторой функцией от матрицы $B = \|b_{ij}\| = B(x_1, x_2, ..., x_n)$ порядка $n \times n$, где

$$b_{ij} = \begin{cases} 1, & x_i = x_j, i, j = 1,2,...,n, \\ 0, & x_i \neq x_j, i, j = 1,2,...,n. \end{cases}$$

Если $W: R^n \to A$ - адекватный алгоритм в шкале порядка, то этот алгоритм задается некоторой функцией от матрицы $C = \|c_{ij}\| = C(x_1, x_2, ..., x_n)$ порядка $n \times n$, где

$$c_{ij} = \begin{cases} 1, & x_i \le x_j, i, j = 1,2,...,n, \\ 0, & x_i > x_j, i, j = 1,2,...,n. \end{cases}$$

Матрицы B и C можно проинтерпретировать в терминах бинарных отношений. Пусть некоторая характеристика измеряется у n объектов $q_1,q_2,\ldots,q_n$, причем x_i - результат ее измерения у объекта q_i. Тогда матрицы B и C задают бинарные отношения на множестве объектов $Q =\{q_1,q_2,\ldots,q_n\}$. Поскольку бинарное отношение можно рассматривать как подмножество декартова квадрата $Q \times Q$, то любой матрице $D = ||d_{ij}||$ порядка $n \times n$ из 0 и 1 соответствует бинарное отношение $R(D)$, определяемое следующим образом: $(q_i, q_j) \in R(D)$ тогда и только тогда, когда $d_{ij} = 1$.

Бинарное отношение $R(B)$ - отношение эквивалентности, т.е. симметричное рефлексивное транзитивное отношение. Оно задает разбиение Q на классы эквивалентности. Два объекта q_i и q_j входят в один класс эквивалентности тогда и только тогда, когда $x_i = x_j$, $b_{ij} = 1$.

Выше показано, как разбиения возникают в результате измерений в шкале наименований. Разбиения могут появляться и непосредственно. Так, при оценке качества промышленной продукции эксперты дают разбиение показателей качества на группы [23]. Для изучения психологического состояния людей их просят разбить предъявленные рисунки на группы сходных между собой. Аналогичная методика применяется и в иных экспериментальных психологических исследованиях, необходимых для оптимизации управления персоналом.

Во многих эконометрических задачах разбиения получаются «на выходе» (например, в кластерном анализе) или же используются на промежуточных этапах анализа данных (например, сначала проводят классификацию с целью выделения однородных групп, а затем в каждой группе строят регрессионную зависимость).

Бинарное отношение $R(C)$ задает разбиение Q на классы эквивалентности, между которыми введено отношение строгого порядка. Два объекта q_i и q_j входят в один класс тогда и только тогда, когда c_{ij}= 1 и c_{ji}= 1, т.е. $x_i = x_j$. Класс эквивалентности Q_1 предшествует классу эквивалентности Q_2 тогда и только тогда, когда для любых $q_i \in Q_1$, $q_j \in Q_2$ имеем $c_{ij} = 1$, c_{ji}= 0, т.е. $x_i < x_j$. Такое бинарное отношение в статистике часто называют ранжировкой со связями; связанными считаются объекты, входящие в один класс эквивалентности. В литературе встречаются и другие названия: линейный квазипорядок, упорядочение, квазисерия, ранжирование. Если каждый из классов эквивалентности состоит только из одного элемента, то имеем обычную ранжировку (другими словами, строгий линейный порядок).

Как известно, ранжировки возникают в результате измерений в порядковой шкале. Так, при описанном выше опросе ответ выпускника школы - это ранжировка (со связями) профессий по привлекательности. Ранжировки часто возникают и непосредственно, без промежуточного этапа - приписывания объектам квазичисловых оценок - баллов. Многочисленные примеры тому даны английским статистиком М. Кендэлом [25]. При оценке качества промышленной продукции широко применяемые нормативные и методические документы предусматривают использование ранжировок.

Для прикладных областей, кроме ранжировок и разбиений, представляют интерес толерантности, т.е. рефлексивные симметричные отношения. Толерантность - математическая модель для выражения представлений о сходстве (похожести, близости). Разбиения - частный вид толерантностей. Толерантность, обладающая свойством транзитивности - это разбиение. Однако в общем случае толерантность не обязана быть транзитивной. Толерантности появляются во многих постановках теории экспертных оценок, например, как результат парных сравнений (см. ниже).

Напомним, что любое бинарное отношение на конечном множестве может быть описано матрицей из 0 и 1.

6. Дихотомические (бинарные) данные

Это данные, которые могут принимать одно из двух значений (0 или 1), т.е. результаты измерений значений альтернативного признака. Как уже было показано, измерения в шкале наименований и порядковой шкале приводят к бинарным отношениям, а те могут быть выражены как результаты измерений по нескольким альтернативным признакам, соответствующим элементам матриц, описывающих отношения. Дихотомические данные возникают в прикладных исследованиях и многими иными путями.

В настоящее время в большинстве технических регламентов, стандартов, технических условий, договоров на поставку конкретной продукции предусмотрен контроль по альтернативному признаку. Это означает, что единица продукции относится к одной из двух категорий - «годных» или «дефектных», т.е. соответствующих или не соответствующих требованиям стандарта. Отечественными специалистами проведены обширные теоретические исследования проблем статистического приемочного контроля по альтернативному признаку. Основополагающими в этой области являются работы академика А.Н. Колмогорова. Подход отечественной вероятностно-статистической научной школы к проблемам контроля качества продукции отражен в монографиях [26, 27] (см. также главу 10 учебника [28]).

Дихотомические данные – давний объект математической статистики. Особенно большое применение они имеют в экономических и социологических исследованиях, в которых большинство переменных, интересующих специалистов, измеряется по качественным шкалам. При этом дихотомические данные зачастую являются более адекватными, чем результаты измерений по методикам, использующим большее число градаций. В частности, психологические тесты типа MMPI (расшифровывается как Миннесотское многофакторное личностное исследование) используют только дихотомические данные. На них опираются и популярные в технико-экономическом анализе методы парных сравнений [29].

Элементарным актом в методе парных сравнений является предъявление эксперту для сравнения двух объектов (сравнение может проводиться также прибором). В одних постановках эксперт должен выбрать из двух объектов лучший по качеству, в других - ответить, похожи объекты или нет. В обоих случаях ответ эксперта можно выразить одной из двух цифр (меток) - 0 или 1. В первой постановке: 0, если лучшим объявлен первый объект; 1 - если второй. Во второй постановке: 0, если объекты похожи, схожи, близки; 1 - в противном случае.

Подводя итоги, можно сказать, что рассмотренные выше виды данных могут быть представлены в виде векторов из 0 и 1 (при обосновании этого утверждения используется тот очевидный факт, что матрицы могут быть записаны в виде векторов). Более того, поскольку все мыслимые результаты наблюдений имеют лишь несколько значащих цифр, то, используя двоичную систему счисления, любые виды анализируемых статистическими методами данных можно записать в виде векторов конечной длины (размерности) из 0 и 1. Представляется, однако, что эта возможность в большинстве случаев имеет лишь академический интерес. Но во всяком случае можно констатировать, что анализ дихотомических данных необходим во многих прикладных постановках.

7. Множества

Совокупность X^n векторов $X = (x_1, x_2,\dots,x_n)$ из 0 и 1 размерности n находится во взаимно-однозначном соответствии с совокупностью 2^n всех подмножеств множества $N = \{1, 2, ..., n\}$. При этом вектору $X = (x_1, x_2,\ldots,x_n)$ соответствует подмножество $N(X)\subseteq N$, состоящее из тех и только из тех i, для которых $x_i = 1$. Это объясняет, почему изложение вероятностных и статистических результатов, относящихся к анализу данных, являющихся объектами нечисловой природы перечисленных выше видов, можно вести на языке конечных случайных множеств, как это было сделано в монографии [20].

Множества как исходные данные появляются и в иных постановках. Из геологических задач исходил Ж. Матерон [30], из электротехнических - Н.Н. Ляшенко [31], и др. Случайные множества применялись для описания процесса случайного распространения, например распространения информации, слухов, эпидемии или пожара, а также в математической экономике. В монографии [20] рассмотрены приложения случайных множеств в теории экспертных оценок и в теории управления запасами и ресурсами (логистике).

Отметим, что с точки зрения математики реальные объекты можно моделировать случайными множествами как из конечного числа элементов, так и из бесконечного, однако при расчетах на ЭВМ неизбежна дискретизация, т.е. переход к первой из названных возможностей.

8. Объекты нечисловой природы как статистические данные

В теории и практике статистических методов наиболее распространенный объект изучения и применения - выборка $x_1, x_2,\ldots,x_n$, т.е. совокупность результатов n наблюдений (измерений, испытаний. анализов, опытов). В различных областях статистики результат наблюдения - это или число, или конечномерный вектор, или функция... Соответственно проводится, как уже отмечалось, деление прикладной статистики: одномерная статистика, многомерный статистический анализ, статистика временных рядов и случайных процессов... В статистике объектов нечисловой природы (нечисловой статистике, статистике нечисловых данных) в качестве результатов наблюдений рассматриваются объекты нечисловой природы. В частности, математические объекты перечисленных выше видов - измерения в шкалах, отличных от абсолютной, бинарные отношения, вектора из 0 и 1, множества. А также нечеткие множества [6] или интервальные данные [7], и др. Выборка может состоять из *n* ранжировок, или *n* толерантностей, или *n* множеств, или *n* нечетких множеств и т.д.

Отметим необходимость развития методов статистической обработки «разнотипных данных», обусловленную большой ролью в прикладных исследованиях «признаков смешанной природы». Речь идет о том, что результат наблюдения состояния объекта зачастую представляет собой вектор, у которого часть координат измерена по шкале наименований, часть – по порядковой шкале, часть – по шкале интервалов и т.д. Классические статистические методы ориентированы обычно либо на абсолютную шкалу, либо на шкалу наименований (анализ таблиц сопряженности), а потому зачастую непригодны для обработки разнотипных данных. Есть и более сложные модели разнотипных данных, например, когда некоторые координаты вектора наблюдений описываются нечеткими множествами.

Для обозначения подобных неклассических результатов наблюдений в 1979 г. в монографии [20] предложен собирательный термин - объекты нечисловой природы. Термин «нечисловой» означает, что структура пространства, в котором лежат результаты наблюдений, не является структурой действительных чисел, векторов или функций, она вообще не является структурой линейного (векторного) пространства. В памяти компьютеров и при расчетах объекты числовой природы, разумеется, изображаются с помощью чисел, но эти числа нельзя складывать и умножать.

С целью «стандартизации математических орудий» (выражение группы французских математиков, действовавшей в середине XX в. под псевдонимом Н. Бурбаки) целесообразно разрабатывать методы статистического анализа данных, пригодные одновременно для всех перечисленных выше видов результатов наблюдений. Кроме того, в процессе развития теоретических и прикладных исследований выявляется необходимость использования новых видов объектов нечисловой природы, отличных от рассмотренных выше, например, в связи с развитием статистических методов обработки текстовой информации. Поэтому целесообразно ввести еще один вид объектов нечисловой природы - объекты произвольной природы, т.е. элементы множеств, на которые не наложено никаких условий (кроме «условий регулярности», необходимых для справедливости доказываемых теорем). Другими словами, в этом случае предполагается, что результаты наблюдений (элементы выборки) лежат в произвольном пространстве X. Для получения теорем необходимо потребовать, чтобы X удовлетворяло некоторым внутриматематическим условиям, например, было топологическим пространством [32]. Как известно, ряд результатов классической математической статистики получен именно в такой постановке. Так, при изучении оценок максимального правдоподобия элементы выборки могут лежать в пространстве произвольной природы. Это не влияет на рассуждения, поскольку в них рассматривается лишь зависимость плотности вероятности от параметра. Методы классификации, использующие лишь расстояние между классифицируемыми объектами, могут применяться к совокупностям объектов произвольной природы, лишь бы в пространстве, где они лежат, была задана метрика [33]. Цель статистики объектов нечисловой природы состоит в том, чтобы систематически рассматривать методы статистической обработки данных как произвольной природы, так и относящихся к указанным выше конкретным видам объектов нечисловой природы, т.е. методы описания данных, оценивания и проверки гипотез. Взгляд с общей точки зрения позволяет получить новые результаты и в других областях прикладной статистики.

9. Объекты нечисловой природы при формировании статистической или математической модели реального явления

Использование объектов нечисловой природы часто порождено желанием обрабатывать более объективную, более освобожденную от погрешностей информацию. Как показали многочисленные опыты, человек более правильно (и с меньшими затруднениями) отвечает на вопросы качественного, например, сравнительного, характера, чем количественного. Так, ему легче сказать, какая из двух гирь тяжелее, чем указать их примерный вес в граммах. Другими словами, использование объектов нечисловой природы – средство повысить устойчивость эконометрических и экономико-математических моделей реальных явлений. Сначала конкретные области статистики объектов нечисловой природы (а именно, прикладная теория измерений, нечеткие и случайные множества) были рассмотрены в монографии [20] при анализе частных постановок проблемы устойчивости математических моделей социально-экономических явлений и процессов к допустимым отклонениям исходных данных и предпосылок модели. А затем была понята необходимость проведения работ по развитию статистики объектов нечисловой природы как самостоятельного научного направления.

Обсуждение начнем со шкал измерения. Науку о единстве мер и точности измерений называют метрологией. Таким образом, теория измерений - часть метрологии. Методы обработки данных должны быть адекватны относительно допустимых преобразований шкал измерения в смысле репрезентативной теории измерений. Однако установление типа шкалы, т.е. задание группы преобразований Ф - дело специалиста соответствующей прикладной области. Так, оценки привлекательности профессий мы считали измеренными в порядковой шкале [20]. Однако отдельные социологи не соглашались с этим, считая, что выпускники школ пользуются шкалой с более узкой группой допустимых преобразований, например, интервальной шкалой. Очевидно, эта проблема относится не к математике, а к наукам о человеке. Для ее решения может быть поставлен достаточно трудоемкий эксперимент. Пока же он не поставлен, целесообразно принимать порядковую шкалу, так как это гарантирует от возможных ошибок.

Как уже отмечалось, номинальные и порядковые шкалы широко распространены не только в социально-экономических исследованиях. Они применяются в медицине, минералогии, географии и т.д. Так, по шкале интервалов измеряют величину потенциальной энергии или координату точки на прямой, на которой не отмечены ни начало, ни единица измерения; по шкале отношений - большинство физических единиц: массу тела, длину, заряд, а также цены и стоимостные величины в экономике. Время измеряется по шкале разностей, если год принимаем естественной единицей измерения, и по шкале интервалов в общем случае. В процессе развития соответствующей области знания тип шкалы может меняться. Так, сначала температура измерялась по порядковой шкале (холоднее - теплее), затем - по интервальной (шкалы Цельсия, Фаренгейта, Реомюра) и, наконец, после открытия абсолютного нуля температур - по шкале отношений (шкала Кельвина). Следует отметить, что среди специалистов иногда имеются разногласия по поводу того, по каким шкалам следует считать измеренными те или иные реальные величины.

Отметим, что термин «репрезентативная теория измерений» использовался, чтобы отличить рассматриваемый подход к теории измерений от классической метрологии, а также от работ А.Н. Колмогорова и А. Лебега, связанных с измерением геометрических величин, от «алгоритмической теории измерения», и от других научных направлений.

Необходимость использования в математических моделях реальных явлений таких объектов нечисловой природы, как бинарные отношения, множества, нечеткие множества, кратко была показана выше. Здесь же обратим внимание, что анализируемые в классической статистике результаты наблюдений также «не совсем числа». А именно, любая величина X измеряется всегда с некоторой погрешностью ΔX и результатом наблюдения является

$$Y = X + \Delta X.$$

Как уже отмечалось, погрешностями измерений занимается метрология. Отметим справедливость следующих фактов:

а) для большинства реальных измерений невозможно полностью исключить систематическую ошибку, т.е. $M(\Delta X) \neq 0$;

б) распределение ΔX в подавляющем большинстве случаев не является нормальным [28, 34];

в) измеряемую величину X и погрешность ее измерения ΔX обычно нельзя считать независимыми случайными величинами;

г) распределение погрешностей оценивается по результатам специально проведенных измерений, следовательно, полностью известным считать его нельзя; зачастую исследователь располагает лишь границами для систематической погрешности и оценками таких характеристик случайной погрешности, как дисперсия или размах.

Приведенные факты показывают ограниченность области применимости распространенной модели погрешностей, в которой X и ΔX рассматриваются как независимые случайные величины, причем ΔX имеет нормальное распределение с нулевым математическим ожиданием.

Строго говоря, результаты наблюдения всегда имеют дискретное распределение, поскольку описываются числами, у которых немного значащих цифр (обычно от 1 до 5). Возникает дилемма: либо признать, что непрерывные распределения - внутриматематическая фикция, и прекратить ими пользоваться, либо считать, что непрерывные распределения имеют «реальные» величины X, которые наблюдаются с принципиально неустранимой погрешностью ΔX. Первый выход в настоящее время нецелесообразен, так как он требует отказа от большей части разработанного математического аппарата. Из второго следует необходимость изучения влияния неустранимых погрешностей на статистические выводы.

Погрешности ΔX можно учитывать либо с помощью вероятностной модели (ΔX - случайная величина, имеющая некоторую функцию распределения, вообще говоря, зависящую от X), либо с помощью нечетких множеств. Во втором случае приходим к теории нечетких чисел [6] и к ее частному случаю - статистике интервальных данных [7], при более продвинутом подходе – к системной нечеткой интервальной математике [10 – 12].

Другой источник появления погрешности ΔX связан с принятой в конструкторской и технологической документации системой допусков на контролируемые параметры изделий и деталей, с использованием шаблонов при проверке контроля качества продукции [35]. В этих случаях характеристики ΔX определяются не свойствами средств измерения, а применяемой технологией проектирования и производства. В терминах прикладной статистики сказанному соответствует группировка данных, при которой мы знаем, какому из заданных интервалов принадлежит наблюдение, но не знаем точного значения результата наблюдения. Применение группировки может дать экономический эффект, поскольку зачастую легче (в среднем) установить, к какому интервалу относится результат наблюдения, чем точно измерить его.

10. Объекты нечисловой природы как результат статистической обработки данных

Объекты нечисловой природы появляются не только на «входе» статистической процедуры, но и в процессе обработки данных, и на «выходе» в качестве итога статистического анализа.

Рассмотрим простейшую прикладную постановку задачи регрессии (см. также [28]). Исходные данные имеют вид $(x_i, y_i) \in R^2,\ i = 1,2,...,n$. Цель состоит в том, чтобы с достаточной точностью описать y как многочлен (полином) от x, т.е. модель имеет вид

$$y_i = \sum_{k=0}^{m} a_k x_i^k + \varepsilon_i, \quad i = 1,2,...,n, \qquad (2)$$

где m - неизвестная степень полинома; $a_0, a_1, a_2, ..., a_m$ - неизвестные коэффициенты многочлена; $\varepsilon_i,\ i = 1,2,...,n$, - погрешности, которые для простоты примем независимыми и имеющими одно и то же нормальное распределение.

Здесь наглядно проявляется одна из причин живучести статистических моделей на основе нормального распределения. Такие модели, хотя и, как правило, неадекватны реальной ситуации [28, 34], с математической точки зрения позволяет проникнуть глубже в суть изучаемого явления. Поэтому они пригодны для первоначального анализа ситуации, как и в рассматриваемом случае. Дальнейшие научные исследования должны быть направлены на снятие нереалистического предположения нормальности и переход к непараметрическим моделям погрешности.

Распространенная процедура восстановления зависимости с помощью многочлена такова: сначала пытаются применить модель (2) для линейной функции (m = 1), при неудаче (неадекватности модели) переходят к многочлену второго порядка (m = 2), если снова неудача, то берут модель (2) с m= 3 и т.д. (адекватность модели проверяют по F-критерию Фишера).

Обсудим свойства этой процедуры в терминах прикладной статистики. Если степень полинома задана ($m = m_0$), то его коэффициенты оценивают методом наименьших квадратов, свойства этих оценок хорошо известны (см., например, учебник [28] или монографию [36, гл.26]). Однако в описанной выше реальной постановке m тоже является неизвестным параметром и подлежит оценке. Таким образом, требуется оценить объект (m, a_0, a_1, a_2, …, a_m), множество значений которого можно описать как $R^1 \cup R^2 \cup R^3 \cup \ldots$. Это - объект нечисловой природы, обычные методы оценивания для него неприменимы, так как m - дискретный параметр. В рассматриваемой постановке разработанные к настоящему времени методы оценивания степени полинома носят в основном эвристический характер (см., например, гл. 12 монографии [37]). Свойства описанной выше распространенной процедуры рассмотрены в [28]. Показано, что обычно используемыми методами степень полинома m оценивается несостоятельно, и найдено предельное распределение оценок этого параметра, оказавшееся геометрическим. Отметим, что для степени многочлена давно предложены состоятельные оценки [38, 39].

В более общем случае линейной регрессии данные имеют вид $(y_i, X_i),\ i = 1,2,...,n$, где $X_i = (x_{i1}, x_{i2},..., x_{iN}) \in R^N$ - вектор предикторов (факторов, объясняющих переменных), а модель такова:

$$y_i = \sum_{j \in K} a_j x_{ij} + \varepsilon_i, \quad i = 1,2,...,n \qquad (3)$$

(здесь K - некоторое подмножество множества $\{1, 2, …, n\}$; ε_i - те же, что и в модели (2); a_j - неизвестные коэффициенты при предикторах с номерами из K). Модель (2) сводится к модели (3), если

$$x_{i1} = 1,\ x_{i2} = x_i,\ x_{i3} = x_i^2,\ x_{i4} = x_i^3,..., x_{ij} = x_i^{j-1},...$$

В модели (2) есть естественный порядок ввода предикторов в рассмотрение - в соответствии с возрастанием степени независимой переменной, а в модели (3) естественного порядка нет, поэтому здесь стоит произвольное подмножество множества предикторов. Есть только частичный порядок - чем мощность подмножества меньше, тем лучше. Модель (3) особенно актуальна в технических исследованиях (см. многочисленные примеры в журнале «Заводская лаборатория. Диагностика материалов»). Она применяется в задачах управления качеством продукции и других технико-экономических исследованиях, в медицине, экономике, маркетинге и социологии, когда из большого числа факторов, предположительно влияющих на изучаемую переменную, надо отобрать по возможности наименьшее число значимых факторов и с их помощью сконструировать прогнозирующую формулу (3).

Задача оценивания модели (3) разбивается на две последовательные задачи: оценивание множества K - подмножества множества всех предикторов, а затем - неизвестных параметров a_j. Методы решения второй задачи хорошо известны и подробно изучены (обычно используют метод наименьших квадратов). Гораздо хуже обстоит дело с оцениванием объекта нечисловой природы K. Как уже отмечалось, существующие методы - в основном эвристические, они зачастую не являются даже состоятельными. Даже само понятие состоятельности в данном случае требует специального определения. Пусть K_0 - истинное подмножество предикторов, т.е. подмножество, для которого справедлива модель (3), а подмножество предикторов K_n - его оценка. Оценка K_n называется состоятельной, если

$$\lim_{n\to\infty} Card(K_n \Delta K_0) = 0\,,$$

где Δ - символ симметрической разности множеств; $Card(K)$ означает число элементов множества K, а предел понимается в смысле сходимости по вероятности.

Задача оценивания в моделях регрессии, таким образом, разбивается на две - оценивание структуры модели и оценивание параметров при заданной структуре. В модели (2) структура описывается неотрицательным целым числом m, в модели (3) – множеством K. Структура – объект нечисловой природы. Задача ее оценивания сложна, в то время как задача оценивания численных параметров при заданной структуре хорошо изучена, разработаны эффективные (в смысле прикладной математической статистики) методы.

Такова же ситуация и в других методах многомерного статистического анализа - в факторном анализе (включая метод главных компонент) и в многомерном шкалировании, в иных оптимизационных постановках проблем прикладного многомерного статистического анализа [40].

Перейдем к объектам нечисловой природы на «выходе» статистической процедуры. Примеры многочисленны. Разбиения - итог работы многих алгоритмов классификации, в частности, алгоритмов кластер-анализа. Ранжировки - результат упорядочения профессий по привлекательности или автоматизированной обработки мнений экспертов - членов комиссии по подведению итогов конкурса научных работ. (В последнем случае используются ранжировки со связями; так, в одну группу, наиболее многочисленную, попадают работы, не получившие наград.) Из всех объектов нечисловой природы, видимо, наиболее часты на «выходе» дихотомические данные - принять или не принять гипотезу, в частности, принять или забраковать партию продукции. Результатом статистической обработка данных может быть множество, например зона наибольшего поражения при аварии, или последовательность множеств, например, «среднемерное» описание распространения пожара (см. главу 4 в монографии [20]). Нечетким множеством Э. Борель [41] еще в начале XX в. предлагал описывать представление людей о числе зерен, образующем «кучу». С помощью нечетких множеств формализуются значения лингвистических переменных, выступающих как итоговая оценка качества систем автоматизированного проектирования, сельскохозяйственных машин, бытовых газовых плит, надежности программного обеспечения или систем управления. Можно констатировать, что все виды объектов нечисловой природы могут появляться «на выходе» статистического исследования.

К многообразию объектов нечисловой природы, кроме рассмотренных в статье, относятся графы [42], текстовые документы [43], чертежи (блок-схемы), изображения, нотные записи и многие другие.

Литература

1. Орлов А.И. Основные черты новой парадигмы математической статистики / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №06(090). С. 187 – 213. – IDA [article ID]: 0901306013. – Режим доступа: http://ej.kubagro.ru/2013/06/pdf/13.pdf

2. Орлов А.И. Статистика объектов нечисловой природы (Обзор) // Заводская лаборатория. Диагностика материалов. 1990. Т.56. №3. С. 76 – 83.

3. Орлов А.И. Тридцать лет статистики объектов нечисловой природы (обзор) // Заводская лаборатория. Диагностика материалов. 2009. Т.75. №5. С. 55 – 64.

4. Орлов А.И. О развитии статистики объектов нечисловой природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №09(093). С. 273 – 309. – IDA [article ID]: 0931309019. – Режим доступа: http://ej.kubagro.ru/2013/09/pdf/19.pdf

5. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №05(089). С. 556 – 586. – IDA [article ID]: 0891305038. – Режим доступа: http://ej.kubagro.ru/2013/05/pdf/38.pdf

6. Орлов А.И. Теория нечетких множеств – часть теории вероятностей / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 589 – 617. – IDA [article ID]: 0921308039. – Режим доступа: http://ej.kubagro.ru/2013/08/pdf/39.pdf

7. Орлов А.И. Основные идеи статистики интервальных данных / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный

ресурс]. – Краснодар: КубГАУ, 2013. – №10(094). С. 867 – 892. – IDA [article ID]: 0941310060. – Режим доступа: http://ej.kubagro.ru/2013/10/pdf/60.pdf

8. Орлов А.И. Оценки плотности распределения вероятностей в пространствах произвольной природы / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №05(099). С. 33 – 49. – IDA [article ID]: 0991405003. – Режим доступа: http://ej.kubagro.ru/2014/05/pdf/03.pdf

9. Орлов А.И. Предельная теория непараметрических статистик / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 226 – 244. – IDA [article ID]: 1001406011. – Режим доступа: http://ej.kubagro.ru/2014/06/pdf/11.pdf

10. Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики / А.И. Орлов, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: http://ej.kubagro.ru/2013/07/pdf/15.pdf

11. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: http://ej.kubagro.ru/2014/01/pdf/07.pdf

12. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

13. Орлов А.И. Статистическое оценивание для сгруппированных данных / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №04(098). С. 1097 – 1117. – IDA [article ID]: 0981404080. – Режим доступа: http://ej.kubagro.ru/2014/04/pdf/80.pdf

14. Орлов А.И. Основные этапы становления статистических методов / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). – IDA [article ID]: 0971401086. – Режим доступа: http://ej.kubagro.ru/2014/03/pdf/86.pdf

15. Орлов А.И. Репрезентативная теория измерений и ее применения // Заводская лаборатория. Диагностика материалов. 1999. Т.65. №3. С. 57 – 62.

16. Орлов А.И. Математические методы исследования и теория измерений // Заводская лаборатория. Диагностика материалов. 2006. Т.72. №1. С. 67 – 70.

17. Орлов А.И. Теория измерений как часть методов анализа данных: размышления над переводом статьи П.Ф. Веллемана и Л. Уилкинсона // Социология: методология, методы, математическое моделирование. 2012. № 35. С. 155-174.

18. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар:

КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: http://ej.kubagro.ru/2013/08/pdf/58.pdf

19. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. – 541 с.
20. Орлов А.И. Устойчивость в социально-экономических моделях. – М.: Наука, 1979. – 296 с.
21. Шубкин В.П. Социологические опыты. – М.: Мысль,1970. – 256 с.
22. Щукина Г.И. Проблема познавательного интереса в педагогике. – М.: Педагогика, 1971. – 352 с.
23. Орлов А.И. Организационно-экономическое моделирование: учеб. Ч.2. Экспертные оценки. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. – 486 с.
24. Орлов А.И. Объекты нечисловой природы // Заводская лаборатория. Диагностика материалов. 1995. Т.61. №3. С. 43 – 52.
25. Кендэл М. Ранговые корреляции. – М.: Статистика, 1975. – 216 с.
26. Беляев Ю.К. Вероятностные методы выборочного контроля. – М.: Наука, 1975. – 408 с.
27. Лумельский Я.П. Статистические оценки результатов контроля качества. – М.: Изд-во стандартов, 1979. – 200 с.
28. 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. – 624 с.
29. Дэвид Г. Метод парных сравнений. – М.: Статистика, 1978. – 144 с.
30. Матерон Ж. Случайные множества и интегральная геометрия. – М.: Мир, 1978. – 320 с.
31. Ляшенко Н.Н. Статистика случайных множеств // Прикладная статистика. Ученые записки по статистике, т.45. – М.: Наука, 1983. – С. 40 – 59.
32. Келли Дж. Общая топология. – М.: Наука, 1968. – 384 с.
33. Орлов А.И. Математические методы теории классификации / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 423 – 459. – IDA [article ID]: 0951401023. – Режим доступа: http://ej.kubagro.ru/2014/01/pdf/23.pdf
34. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. Диагностика материалов. 1991. Т.57. №7. С.64-66.
35. Организация и планирование машиностроительного производства (производственный менеджмент): Учебник / К.А. Грачева, М.К. Захарова, Л.А. Одинцова и др. Под ред. Ю.В. Скворцова, Л.А. Некрасова. – М.: Высшая школа, 2003. – 470 с.
36. Кендалл М.Дж., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973. – 900 с.
37. Себер Дж. Линейный регрессионный анализ. – М.: Мир, 1980. – 456 с.
38. Орлов А.И. Асимптотика некоторых оценок размерности модели в регрессии // Прикладная статистика. Ученые записки по статистике, т.45. – М.: Наука, 1983. – С. 260 – 265.
39. Орлов А.И. Об оценивании регрессионного полинома // Заводская лаборатория. Диагностика материалов. 1994. Т. 60. № 5. С. 43 – 47.
40. Орлов А.И. Асимптотическое поведение решений экстремальных статистических задач // Заводская лаборатория. Диагностика материалов. 1996. Т.62. №10. С. 45 – 46.

41. Борель Э. Вероятность и достоверность. – М.: ГИФМЛ, 1961. – 120 с.
42. Орлов А. И. Графы при моделировании процессов управления промышленными предприятиями / Управление большими системами. Специальный выпуск 30.1 «Сетевые модели в управлении». М.: ИПУ РАН, 2010. С. 62 – 75.
43. Бородкин А. А., Толчеев В. О. Разработка и исследование методов взвешивания ближайших соседей (на примере классификации библиографических текстовых документов) // Заводская лаборатория. Диагностика материалов. 2013. Т. 79. № 7. С. 70 – 74.

References

1. Orlov A.I. Osnovnye cherty novoj paradigmy matematicheskoj statistiki / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №06(090). S. 187 – 213. – IDA [article ID]: 0901306013. – Rezhim dostupa: http://ej.kubagro.ru/2013/06/pdf/13.pdf
2. Orlov A.I. Statistika ob#ektov nechislovoj prirody (Obzor) // Zavodskaja laboratorija. Diagnostika materialov. 1990. T.56. №3. S. 76 – 83.
3. Orlov A.I. Tridcat' let statistiki ob#ektov nechislovoj prirody (obzor) // Zavodskaja laboratorija. Diagnostika materialov. 2009. T.75. №5. S. 55 – 64.
4. Orlov A.I. O razvitii statistiki ob#ektov nechislovoj prirody / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №09(093). S. 273 – 309. – IDA [article ID]: 0931309019. – Rezhim dostupa: http://ej.kubagro.ru/2013/09/pdf/19.pdf
5. Orlov A.I. Srednie velichiny i zakony bol'shih chisel v prostranstvah proizvol'noj prirody / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №05(089). S. 556 – 586. – IDA [article ID]: 0891305038. – Rezhim dostupa: http://ej.kubagro.ru/2013/05/pdf/38.pdf
6. Orlov A.I. Teorija nechetkih mnozhestv – chast' teorii verojatnostej / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №08(092). S. 589 – 617. – IDA [article ID]: 0921308039. – Rezhim dostupa: http://ej.kubagro.ru/2013/08/pdf/39.pdf
7. Orlov A.I. Osnovnye idei statistiki interval'nyh dannyh / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №10(094). S. 867 – 892. – IDA [article ID]: 0941310060. – Rezhim dostupa: http://ej.kubagro.ru/2013/10/pdf/60.pdf
8. Orlov A.I. Ocenki plotnosti raspredelenija verojatnostej v prostranstvah proizvol'noj prirody / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №05(099). S. 33 – 49. – IDA [article ID]: 0991405003. – Rezhim dostupa: http://ej.kubagro.ru/2014/05/pdf/03.pdf
9. Orlov A.I. Predel'naja teorija neparametricheskih statistik / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №06(100). S. 226 – 244. – IDA [article ID]: 1001406011. – Rezhim dostupa: http://ej.kubagro.ru/2014/06/pdf/11.pdf

10. Orlov A.I. Sistemnaja nechetkaja interval'naja matematika (SNIM) – perspektivnoe napravlenie teoreticheskoj i vychislitel'noj matematiki / A.I. Orlov, E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 255 – 308. – IDA [article ID]: 0911307015. – Rezhim dostupa: http://ej.kubagro.ru/2013/07/pdf/15.pdf

11. Lucenko E.V. Kognitivnye funkcii kak obobshhenie klassicheskogo ponjatija funkcional'noj zavisimosti na osnove teorii informacii v sistemnoj nechetkoj interval'noj matematike / E.V. Lucenko, A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №01(095). S. 122 – 183. – IDA [article ID]: 0951401007. – Rezhim dostupa: http://ej.kubagro.ru/2014/01/pdf/07.pdf

12. Orlov A.I., Lucenko E.V. Sistemnaja nechetkaja interval'naja matematika. Monografija (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s.

13. Orlov A.I. Statisticheskoe ocenivanie dlja sgruppirovannyh dannyh / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №04(098). S. 1097 – 1117. – IDA [article ID]: 0981404080. – Rezhim dostupa: http://ej.kubagro.ru/2014/04/pdf/80.pdf

14. Orlov A.I. Osnovnye jetapy stanovlenija statisticheskih metodov / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №03(097). – IDA [article ID]: 0971401086. – Rezhim dostupa: http://ej.kubagro.ru/2014/03/pdf/86.pdf

15. Orlov A.I. Reprezentativnaja teorija izmerenij i ee primenenija // Zavodskaja laboratorija. Diagnostika materialov. 1999. T.65. №3. S. 57 – 62.

16. Orlov A.I. Matematicheskie metody issledovanija i teorija izmerenij // Zavodskaja laboratorija. Diagnostika materialov. 2006. T.72. №1. S. 67 – 70.

17. Orlov A.I. Teorija izmerenij kak chast' metodov analiza dannyh: razmyshlenija nad perevodom stat'i P.F. Vellemana i L. Uilkinsona // Sociologija: metodologija, metody, matematicheskoe modelirovanie. 2012. № 35. S. 155-174.

18. Lucenko E.V. Metrizacija izmeritel'nyh shkal razlichnyh tipov i sovmestnaja sopostavimaja kolichestvennaja obrabotka raznorodnyh faktorov v sistemno-kognitivnom analize i sisteme «Jejdos» / E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №08(092). S. 859 – 883. – IDA [article ID]: 0921308058. – Rezhim dostupa: http://ej.kubagro.ru/2013/08/pdf/58.pdf

19. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch. 1. Nechislovaja statistika. – M.: Izd-vo MGTU im. N.Je. Baumana, 2009. – 541 s.

20. Orlov A.I. Ustojchivost' v social'no-jekonomicheskih modeljah. – M.: Nauka, 1979. – 296 s.

21. Shubkin V.P. Sociologicheskie opyty. – M.: Mysl',1970. – 256 s.

22. Shhukina G.I. Problema poznavatel'nogo interesa v pedagogike. – M.: Pedagogika, 1971. – 352 s.

23. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie: ucheb. Ch.2. Jekspertnye ocenki. – M.: Izd-vo MGTU im. N.Je. Baumana, 2011. – 486 s.

24. Orlov A.I. Ob#ekty nechislovoj prirody // Zavodskaja laboratorija. Diagnostika materialov. 1995. T.61. №3. S. 43 – 52.

25. Kendjel M. Rangovye korreljacii. – M.: Statistika, 1975. – 216 s.

26. Beljaev Ju.K. Verojatnostnye metody vyborochnogo kontrolja. – M.: Nauka, 1975. – 408 s.
27. Lumel'skij Ja.P. Statisticheskie ocenki rezul'tatov kontrolja kachestva. – M.: Izd-vo standartov, 1979. – 200 s.
28. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch.3. Statisticheskie metody analiza dannyh. – M.: Izd-vo MGTU im. N.Je. Baumana, 2012. – 624 s.
29. Djevid G. Metod parnyh sravnenij. – M.: Statistika, 1978. – 144 s.
30. Materon Zh. Sluchajnye mnozhestva i integral'naja geometrija. – M.: Mir, 1978. – 320 s.
31. Ljashenko N.N. Statistika sluchajnyh mnozhestv // Prikladnaja statistika. Uchenye zapiski po statistike, t.45. – M.: Nauka, 1983. – S. 40 – 59.
32. Kelli Dzh. Obshhaja topologija. – M.: Nauka, 1968. – 384 s.
33. Orlov A.I. Matematicheskie metody teorii klassifikacii / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №01(095). S. 423 – 459. – IDA [article ID]: 0951401023. – Rezhim dostupa: http://ej.kubagro.ru/2014/01/pdf/23.pdf
34. Orlov A.I. Chasto li raspredelenie rezul'tatov nabljudenij javljaetsja normal'nym? // Zavodskaja laboratorija. Diagnostika materialov. 1991. T.57. №7. S.64-66.
35. Organizacija i planirovanie mashinostroitel'nogo proizvodstva (proizvodstvennyj menedzhment): Uchebnik / K.A. Gracheva, M.K. Zaharova, L.A. Odincova i dr. Pod red. Ju.V. Skvorcova, L.A. Nekrasova. – M.: Vysshaja shkola, 2003. – 470 s.
36. Kendall M.Dzh., St'juart A. Statisticheskie vyvody i svjazi. – M.: Nauka, 1973. – 900 s.
37. Seber Dzh. Linejnyj regressionnyj analiz. – M.: Mir, 1980. – 456 s.
38. Orlov A.I. Asimptotika nekotoryh ocenok razmernosti modeli v regressii // Prikladnaja statistika. Uchenye zapiski po statistike, t.45. – M.: Nauka, 1983. – S. 260 – 265.
39. Orlov A.I. Ob ocenivanii regressionnogo polinoma // Zavodskaja laboratorija. Diagnostika materialov. 1994. T. 60. № 5. S. 43 – 47.
40. Orlov A.I. Asimptoticheskoe povedenie reshenij jekstremal'nyh statisticheskih zadach // Zavodskaja laboratorija. Diagnostika materialov. 1996. T.62. №10. S. 45 – 46.
41. Borel' Je. Verojatnost' i dostovernost'. – M.: GIFML, 1961. – 120 s.
42. Orlov A. I. Grafy pri modelirovanii processov upravlenija promyshlennymi predprijatijami / Upravlenie bol'shimi sistemami. Special'nyj vypusk 30.1 «Setevye modeli v upravlenii». M.: IPU RAN, 2010. S. 62 – 75.
43. Borodkin A. A., Tolcheev V. O. Razrabotka i issledovanie metodov vzveshivanija blizhajshih sosedej (na primere klassifikacii bibliograficheskih tekstovyh dokumentov) // Zavodskaja laboratorija. Diagnostika materialov. 2013. T. 79. № 7. S. 70 – 74.