

УДК 519.2: 005.521:633.1:004.8

UDC 519.2: 005.521:633.1:004.8

01.00.00 Физико-математические науки

Physics and mathematical sciences

**МЕТОДЫ СНИЖЕНИЯ РАЗМЕРНОСТИ  
ПРОСТРАНСТВА СТАТИСТИЧЕСКИХ  
ДАНЫХ****METHODS OF REDUCING SPACE  
DIMENSION OF STATISTICAL DATA**

Орлов Александр Иванович  
д.э.н., д.т.н., к.ф.-м.н., профессор  
РИНЦ SPIN-код: 4342-4994  
*Московский государственный технический  
университет им. Н.Э. Баумана, Россия, 105005,  
Москва, 2-я Бауманская ул., 5, [prof-orlov@mail.ru](mailto:prof-orlov@mail.ru)*

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

Луценко Евгений Вениаминович  
д.э.н., к.т.н., профессор  
РИНЦ SPIN-код: 9523-7101  
*Кубанский государственный аграрный  
университет, Краснодар, Россия  
[prof.lutsenko@gmail.com](mailto:prof.lutsenko@gmail.com)*

Lutsenko Eugeny Veniaminovich  
Dr.Sci.Econ., Cand.Tech.Sci., professor  
RSCI SPIN-code: 9523-7101  
*Kuban State Agrarian University, Krasnodar,  
Russia  
[prof.lutsenko@gmail.com](mailto:prof.lutsenko@gmail.com)*

Одной из «точек роста» прикладной статистики являются методы снижения размерности пространства статистических данных. Они все чаще используются при анализе данных в конкретных прикладных исследованиях, например, социологических. Рассмотрим наиболее перспективные методы снижения размерности. Метод главных компонент является одним из наиболее часто используемых методов снижения размерности. Для визуального анализа данных часто используют проекции исходных векторов на плоскость первых двух главных компонент. Обычно хорошо видна структура данных, выделяются компактные кластеры объектов и отдельно выделяющиеся вектора. Метод главных компонент является одним из методов факторного анализа. Новая идея по сравнению с методом главных компонент состоит в том, что на основе нагрузок происходит разбиение факторов на группы. В одну группу объединяются факторы, имеющие сходное влияние на элементы нового базиса. Затем из каждой группы рекомендуется оставить одного представителя. Иногда вместо выбора представителя расчетным путем формируется новый фактор, являющийся центральным для рассматриваемой группы. Снижение размерности происходит при переходе к системе факторов, являющихся представителями групп. Остальные факторы отбрасываются. На использовании расстояний (мер близости, показателей различия) между признаками и основан обширный класс методов многомерного шкалирования. Основная идея этого класса методов состоит в представлении каждого объекта точкой геометрического пространства (обычно размерности 1, 2 или 3), координатами которой служат значения скрытых (латентных) факторов, в совокупности достаточно адекватно описывающих

One of the "points of growth" of applied statistics is methods of reducing the dimension of statistical data. They are increasingly used in the analysis of data in specific applied research, such as sociology. We investigate the most promising methods to reduce the dimensionality. The principal components are one of the most commonly used methods to reduce the dimensionality. For visual analysis of data are often used the projections of original vectors on the plane of the first two principal components. Usually the data structure is clearly visible, highlighted compact clusters of objects and separately allocated vectors. The principal components are one method of factor analysis. The new idea of factor analysis in comparison with the method of principal components is that, based on loads, the factors breaks up into groups. In one group of factors, new factor is combined with a similar impact on the elements of the new basis. Then each group is recommended to leave one representative. Sometimes, instead of the choice of representative by calculation, a new factor that is central to the group in question. Reduced dimension occurs during the transition to the system factors, which are representatives of groups. Other factors are discarded. On the use of distance (proximity measures, indicators of differences) between features and extensive class are based methods of multidimensional scaling. The basic idea of this class of methods is to present each object as point of the geometric space (usually of dimension 1, 2, or 3) whose coordinates are the values of the hidden (latent) factors which combine to adequately describe the object. As an example of the application of probabilistic and statistical modeling and the results of statistics of non-numeric data, we justify the consistency of estimators of the

объект. В качестве примера применения вероятностно-статистического моделирования и результатов статистики нечисловых данных обоснуем состоятельность оценки размерности пространства данных в многомерном шкалировании, ранее предложенной Краскалом из эвристических соображений. Рассмотрен ряд работ по оцениванию размерностей моделей (в регрессионном анализе и в теории классификации). Дана информация об алгоритмах снижения размерности в автоматизированном системно-когнитивный анализе

Ключевые слова: МАТЕМАТИКА, ПРИКЛАДНАЯ СТАТИСТИКА, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ТОЧКИ РОСТА, МЕТОД ГЛАВНЫХ КОМПОНЕНТ, ФАКТОРНЫЙ АНАЛИЗ, МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ, ОЦЕНИВАНИЕ РАЗМЕРНОСТИ ДАННЫХ, ОЦЕНИВАНИЕ РАЗМЕРНОСТИ МОДЕЛИ

dimension of the data in multidimensional scaling, which are proposed previously by Kruskal from heuristic considerations. We have considered a number of consistent estimations of dimension of models (in regression analysis and in theory of classification). We also give some information about the algorithms for reduce the dimensionality in the automated system-cognitive analysis

Keywords: MATHEMATICS, APPLIED STATISTICS, MATHEMATICAL STATISTICS, , GROWTH POINTS, THE PRINCIPAL COMPONENT ANALYSIS, FACTOR ANALYSIS, MULTIDIMENSIONAL SCALING, ESTIMATION OF DATA DIMENSION, ESTIMATION OF MODEL DIMENSION

## 1. Введение

Как уже отмечалось, одной из «точек роста» [1] прикладной статистики являются методы снижения размерности пространства статистических данных. Они все чаще используются при анализе данных в конкретных прикладных исследованиях, например, социологических. Рассмотрим наиболее перспективные методы снижения размерности. В качестве примера применения вероятностно-статистического моделирования и результатов статистики нечисловых данных обоснуем состоятельность оценки размерности пространства, ранее предложенной Краскалом из эвристических соображений [2, 3].

В многомерном статистическом анализе каждый объект описывается вектором, размерность которого произвольна (но одна и та же для всех объектов). Однако человек может непосредственно воспринимать лишь числовые данные или точки на плоскости. Анализировать скопления точек в трехмерном пространстве уже гораздо труднее. Непосредственное восприятие данных более высокой размерности невозможно. Поэтому вполне естественным является желание перейти от многомерной выборки к данным небольшой размерности, чтобы «на них можно было

посмотреть». Например, маркетолог может наглядно увидеть, сколько имеется различных типов поведения потребителей (т.е. сколько целесообразно выделять сегментов рынка) и какие именно (с какими свойствами) потребители в них входят.

Кроме стремления к наглядности, есть и другие мотивы для снижения размерности. Те факторы, от которых интересующая исследователя переменная не зависит, лишь мешают статистическому анализу. Во-первых, на сбор информации о них расходуются финансовые, временные, кадровые ресурсы. Во-вторых, как можно доказать, их включение в анализ ухудшает свойства статистических процедур (в частности, увеличивает дисперсию оценок параметров и характеристик распределений). Поэтому желательно избавиться от таких факторов.

При анализе многомерных данных обычно рассматривают не одну, а множество задач, в частности, по-разному выбирая независимые и зависимые переменные. Поэтому рассмотрим задачу снижения размерности в следующей формулировке. Дана многомерная выборка. Требуется перейти от нее к совокупности векторов меньшей размерности, максимально сохранив структуру исходных данных, по возможности не теряя информации, содержащихся в данных. Задача конкретизируется в рамках каждого конкретного метода снижения размерности.

## **2. Метод главных компонент**

Он является одним из наиболее часто используемых методов снижения размерности. Основная его идея состоит в последовательном выявлении направлений, в которых данные имеют наибольший разброс. Пусть выборка состоит из векторов, одинаково распределенных с вектором  $X = (x(1), x(2), \dots, x(n))$ . Рассмотрим линейные комбинации

$$Y(\lambda(1), \lambda(2), \dots, \lambda(n)) = \lambda(1)x(1) + \lambda(2)x(2) + \dots + \lambda(n)x(n),$$

где

$$\lambda^2(1) + \lambda^2(2) + \dots + \lambda^2(n) = 1.$$

Здесь вектор  $\lambda = (\lambda(1), \lambda(2), \dots, \lambda(n))$  лежит на единичной сфере в  $n$ -мерном пространстве.

В методе главных компонент прежде всего находят направление максимального разброса, т.е. такое  $\lambda$ , при котором достигает максимума дисперсия случайной величины  $Y(\lambda) = Y(\lambda(1), \lambda(2), \dots, \lambda(n))$ . Тогда вектор  $\lambda$  задает первую главную компоненту, а величина  $Y(\lambda)$  является проекцией случайного вектора  $X$  на ось первой главной компоненты.

Затем, выражаясь терминами линейной алгебры, рассматривают гиперплоскость в  $n$ -мерном пространстве, перпендикулярную первой главной компоненте, и проектируют на эту гиперплоскость все элементы выборки. Размерность гиперплоскости на 1 меньше, чем размерность исходного пространства.

В рассматриваемой гиперплоскости процедура повторяется. В ней находят направление наибольшего разброса, т.е. вторую главную компоненту. Затем выделяют гиперплоскость, перпендикулярную первым двум главным компонентам. Ее размерность на 2 меньше, чем размерность исходного пространства. Далее – следующая итерация.

С точки зрения линейной алгебры речь идет о построении нового базиса в  $n$ -мерном пространстве, ортами которого служат главные компоненты.

Дисперсия, соответствующая каждой новой главной компоненте, меньше, чем для предыдущей. Обычно останавливаются, когда она меньше заданного порога. Если отобрано  $k$  главных компонент, то это означает, что от  $n$ -мерного пространства удалось перейти к  $k$ -мерному, т.е. сократить размерность с  $n$ -до  $k$ , практически не исказив структуру исходных данных.

Для визуального анализа данных часто используют проекции исходных векторов на плоскость первых двух главных компонент. Обычно

хорошо видна структура данных, выделяются компактные кластеры объектов и отдельно выделяющиеся вектора.

### 3. Факторный анализ

Метод главных компонент является одним из методов *факторного анализа* [4]. Различные алгоритмы факторного анализа объединены тем, что во всех них происходит переход к новому базису в исходном  $n$ -мерном пространстве. Важным является понятие «нагрузка фактора», применяемое для описания роли исходного фактора (переменной) в формировании определенного вектора из нового базиса.

Новая идея по сравнению с методом главных компонент состоит в том, что на основе нагрузок происходит разбиение факторов на группы. В одну группу объединяются факторы, имеющие сходное влияние на элементы нового базиса. Затем из каждой группы рекомендуется оставить одного представителя. Иногда вместо выбора представителя расчетным путем формируется новый фактор, являющийся центральным для рассматриваемой группы. Снижение размерности происходит при переходе к системе факторов, являющихся представителями групп. Остальные факторы отбрасываются.

Описанная процедура может быть осуществлена не только с помощью факторного анализа. Речь идет о кластер-анализе признаков (факторов, переменных). Для разбиения признаков на группы можно применять различные алгоритмы кластер-анализа [5 - 7]. Достаточно ввести расстояние (меру близости, показатель различия) между признаками. Пусть  $X$  и  $Y$  – два признака. Различие  $d(X, Y)$  между ними можно измерять с помощью выборочных коэффициентов корреляции:

$$d_1(X, Y) = 1 - |r_n(X, Y)|, \quad d_2(X, Y) = 1 - |\rho_n(X, Y)|,$$

где  $r_n(X, Y)$  – выборочный линейный коэффициент корреляции Пирсона,  $\rho_n(X, Y)$  – выборочный коэффициент ранговой корреляции Спирмена.

#### **4. Многомерное шкалирование.**

На использовании расстояний (мер близости, показателей различия)  $d(X, Y)$  между признаками  $X$  и  $Y$  основан обширный класс методов многомерного шкалирования [8, 9]. Основная идея этого класса методов состоит в представлении каждого объекта точкой геометрического пространства (обычно размерности 1, 2 или 3), координатами которой служат значения скрытых (латентных) факторов, в совокупности достаточно адекватно описывающих объект. При этом отношения между объектами заменяются отношениями между точками – их представителями. Так, данные о сходстве объектов – расстояниями между точками, данные о превосходстве – взаимным расположением точек [10].

#### **5. Проблема оценки истинной размерности факторного пространства**

В практике анализа социологических данных используется ряд различных моделей многомерного шкалирования. Во всех них встает проблема оценки истинной размерности факторного пространства. Рассмотрим эту проблему на примере обработки данных о сходстве объектов с помощью метрического шкалирования.

Пусть имеется  $n$  объектов  $O(1), O(2), \dots, O(n)$ , для каждой пары объектов  $O(i), O(j)$  задана мера их сходства  $s(i, j)$ . Считаем, что всегда  $s(i, j) = s(j, i)$ . Происхождение чисел  $s(i, j)$  не имеет значения для описания работы алгоритма. Они могли быть получены либо непосредственным измерением, либо с использованием экспертов, либо путем вычисления по совокупности описательных характеристик, либо как-то иначе.

В евклидовом пространстве рассматриваемые  $n$  объектов должны быть представлены конфигурацией  $n$  точек, причем в качестве меры близости точек-представителей выступает евклидово расстояние  $d(i, j)$

между соответствующими точками. Степень соответствия между совокупностью объектов и совокупностью представляющих их точек определяется путем сопоставления матриц сходства  $\|s(i,j)\|$  и расстояний  $\|d(i,j)\|$ . Метрический функционал сходства имеет вид

$$S = \sum_{i < j} |s(i, j) - d(i, j)|^2$$

Геометрическую конфигурацию надо выбирать так, чтобы функционал  $S$  достигал своего наименьшего значения [8, 9].

*Замечание.* В неметрическом шкалировании вместо близости самих мер близости и расстояний рассматривается близость упорядочений на множестве мер близости и множестве соответствующих расстояний. Вместо функционала  $S$  используются аналоги ранговых коэффициентов корреляции Спирмена и Кендалла. Другими словами, неметрическое шкалирование исходит из предположения, что меры близости измерены в порядковой шкале.

Пусть евклидово пространство имеет размерность  $m$ . Рассмотрим минимум среднего квадрата ошибки

$$\alpha_m = \frac{2}{n(n-1)} \min S$$

где минимум берется по всем возможным конфигурациям  $n$  точек в  $m$ -мерном евклидовом пространстве. Можно показать, что рассматриваемый минимум достигается на некоторой конфигурации. Ясно, что при росте  $m$  величина  $\alpha_m$  монотонно убывает (точнее, не возрастает). Можно показать, что при  $m \geq n - 1$  она равна 0 (если  $s(i,j)$  – метрика). Для увеличения возможностей содержательной интерпретации желательно действовать в пространстве возможно меньшей размерности. При этом, однако, размерность необходимо выбрать так, чтобы точки представляли объекты без больших искажений. Возникает вопрос: *как рационально выбирать размерность пространства*, т.е. натуральное число  $m$ ?

## 6. Модели и методы оценивания размерности пространства данных

В рамках детерминированного анализа данных обоснованного ответа на этот вопрос, видимо, нет. Следовательно, необходимо изучить поведение  $\alpha_m$  в тех или иных вероятностных моделях. Если меры близости  $s(i,j)$  являются случайными величинами, распределение которых зависит от «истинной размерности»  $m_0$  (и, возможно, от каких-либо еще параметров), то можно в классическом математико-статистическом стиле ставить задачу оценки  $m_0$ , искать состоятельные оценки и т.д.

Начнем строить вероятностные модели. Примем, что объекты представляют собой точки в евклидовом пространстве размерности  $k$ , где  $k$  достаточно велико. То, что «истинная размерность» равна  $m_0$ , означает, что все эти точки лежат на гиперплоскости размерности  $m_0$ . Примем для определенности, что совокупность рассматриваемых точек представляет собой выборку из кругового нормального распределения с дисперсией  $\sigma^2(0)$ . Это означает, что объекты  $O(1), O(2), \dots, O(n)$  являются независимыми в совокупности случайными векторами, каждый из которых строится как

$$\zeta(1)e(1) + \zeta(2)e(2) + \dots + \zeta(m_0)e(m_0),$$

где  $e(1), e(2), \dots, e(m_0)$  – ортонормальный базис в подпространстве размерности  $m_0$ , в котором лежат рассматриваемые точки, а  $\zeta(1), \zeta(2), \dots, \zeta(m_0)$  – независимые в совокупности одномерные нормальные случайные величины с математическим ожиданием 0 и дисперсией  $\sigma^2(0)$ .

Рассмотрим две модели получения мер близости  $s(i,j)$ . В первой из них  $s(i,j)$  отличаются от евклидова расстояния между соответствующими точками из-за того, что точки известны с искажениями. Пусть  $c(1), c(2), \dots, c(n)$  – рассматриваемые точки. Тогда

$$s(i,j) = d(c(i) + \varepsilon(i), c(j) + \varepsilon(j)), \quad i, j = 1, 2, \dots, n,$$

где  $d$  – евклидово расстояние между точками в  $k$ -мерном пространстве, вектора  $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(n)$  представляют собой выборку из кругового нормального распределения в  $k$ -мерном пространстве с нулевым математическим ожиданием и ковариационной матрицей  $\sigma^2(1)I$ , где  $I$  – единичная матрица. Другими словами,

$$\varepsilon(i) = \eta(1)e(1) + \eta(2)e(2) + \dots + \eta(k)e(k),$$

где  $e(1), e(2), \dots, e(k)$  – ортонормальный базис в  $k$ -мерном пространстве, а  $\{\eta(i,t), i = 1, 2, \dots, n, t = 1, 2, \dots, k\}$  – совокупность независимых в совокупности одномерных случайных величин с нулевым математическим ожиданием и дисперсией  $\sigma^2(1)$ .

Во второй модели искажения наложены непосредственно на сами расстояния:

$$s(i,j) = d(c(i), c(j)) + \varepsilon(i,j), \quad i,j = 1, 2, \dots, n, i \neq j,$$

где  $\{\varepsilon(i,j), i,j = 1, 2, \dots, n\}$  – независимые в совокупности нормальные случайные величины с математическим ожиданием 0 и дисперсией  $\sigma^2(1)$ .

В работе [11] показано, что для обеих сформулированных моделей минимум среднего квадрата ошибки  $\alpha_m$  при  $n \rightarrow \infty$  сходится по вероятности к

$$f(m) = f_1(m) + \sigma^2(1)(k - m), \quad m = 1, 2, \dots, k,$$

где

$$f_1(m) = \begin{cases} \sigma^2(1)(m_0 - m), & m < m_0, \\ 0, & m \geq m_0. \end{cases}$$

Таким образом, функция  $f(m)$  линейна на интервалах  $[1, m_0]$  и  $[m_0, k]$ , причем на первом интервале она убывает быстрее, чем на втором. Отсюда следует, что статистика

$$m^* = \underset{m}{\text{Arg min}} \{ \alpha_{m+1} - 2\alpha_m + \alpha_{m-1} \}$$

является состоятельной оценкой истинной размерности  $m_0$ .

Итак, из вероятностной теории вытекает рекомендация – в качестве оценки размерности факторного пространства использовать  $m^*$ . Отметим, что подобная рекомендация была сформулировано как эвристическая одним из основателей многомерного шкалирования Дж. Краскалом [2, 3, 8]. Он исходил из опыта практического использования многомерного шкалирования и вычислительных экспериментов. Вероятностная теория позволила обосновать эту эвристическую рекомендацию.

## **7. Оценивание размерности модели**

Если возможные подмножества признаков образуют расширяющееся семейство, например, оценивается степень полинома, то естественно ввести термин «размерность модели» (это понятие во многом аналогично используемому в многомерном шкалировании понятию размерности пространства данных). Автору настоящей статьи принадлежит ряд работ по оцениванию размерности модели, которые целесообразно сопоставить с работами по оцениванию размерности пространства данных, рассмотренными выше.

Первая такая работа выполнена автором настоящей статьи во время командировки во Францию в 1976 г. В ней была изучена одна оценка размерности модели в регрессии, а именно, оценка степени полинома в предположении, что зависимость описывается полиномом. Эта оценка была известна в литературе, но позже ее стали ошибочно приписывать автору настоящей статьи, который лишь изучил ее свойства, в частности, установил, что она не является состоятельной, и нашел ее предельное геометрическое распределение [12]. Другие, уже состоятельные оценки размерности регрессионной модели были предложены и изучены в статье [13]. Этот цикл завершила содержащая ряд уточнений работа [14].

Крайняя публикация на эту тему включает в себя обсуждение результатов изучения скорости сходимости в полученных мною предельных теоремах методом Монте-Карло [15].

Аналогичные по методологии оценки размерности модели в задаче расщепления смесей (часть теории классификации [5 - 7]) рассмотрены в статье [16].

Рассмотренные выше оценки размерности модели в многомерном шкалировании изучаются в работах [11, 17, 18]. В этих же работах установлено предельное поведение характеристик метода главных компонент (с помощью асимптотической теории поведения решений экстремальных статистических задач [19, 20]).

## **8. Алгоритмы снижения размерности в автоматизированном системно-когнитивный анализе**

В автоматизированном системно-когнитивный анализе (АСК-анализе) предложен и в системе "Эйдос" реализован еще один метод снижения размерности. Он описан в работе [21] в разделах 4.2 "Описание алгоритмов базовых когнитивных операций системного анализа (БКОСА)" и 4.3 "Детальные алгоритмы БКОСА (АСК-анализа)". Приведем краткое описание двух алгоритмов - БКОСА-4.1 и БКОСА-4.2.

*БКОСА-4.1. "Абстрагирование факторов (снижение размерности семантического пространства факторов)"*

С помощью метода последовательных приближений (итерационный алгоритм) при заданных граничных условиях снижается размерность пространства атрибутов без существенного уменьшения его объема. Критерий остановки итерационного процесса – достижение одного из граничных условий.

*БКОСА-4.2. "Абстрагирование классов (снижение размерности семантического пространства классов)"*

С помощью метода последовательных приближений (итерационный алгоритм) при заданных граничных условиях снижается размерность пространства классов без существенного уменьшения его объема. Критерий остановки итерационного процесса – достижение одного из граничных условий.

Здесь приведены все реальные алгоритмы, реализованные в системе "Эйдос" той версии, которая была реализована на момент подготовки работы [21] (2002 год) : <http://lc.kubagro.ru/aidos/aidos02/4.3.htm>

#### Суть алгоритмов такова.

1. Рассчитывается количество информации в значениях факторов о переходе объекта в состояния, соответствующие классам.

2. Рассчитывается ценность значения фактора для дифференциации объекта по классам. Эта ценность - это просто вариабельность информативностей значений факторов (количественных мер вариабельности много: среднее отклонение от среднего, среднее квадратическое отклонение, и др.). Иначе говоря, если в значении фактора в среднем содержится мало информации о принадлежности и не принадлежности объекта к классу, то это значение не очень ценное, а если много - то ценное.

3. Рассчитывается ценность описательных шкал для дифференциации объектов по классам. В работах Е.В. Луценко сейчас это делается как среднее от ценностей градаций данной шкалы.

4. Потом проводится Парето-оптимизация значений факторов и описательных шкал:

- значения факторов (градации описательных шкал) ранжируются в порядке убывания ценности и удаляются из модели те наименее ценные, которые идут правее касательной к Парето-кривой  $45^\circ$ ;

- факторы (описательные шкалы) ранжируются в порядке убывания ценности и удаляются из модели те наименее ценные, которые идут правее касательной к Парето-кривой  $45^\circ$ .

В результате размерность пространства, построенного на описательных шкалах, существенно снижается за счет удаления коррелирующих между собой шкал, т.е. по сути это ортонормирование пространства в информационной метрике.

Этот процесс может повторяться, т.е. быть итерационным, при этом в новой версии системе "Эйдос" итерации запускаются вручную.

Аналогично ортонормируется информационное пространство классов.

Шкалы и их градации могут быть числовыми (тогда обрабатываются интервальные значения), а также могут быть текстовыми (порядковыми или даже номинальными).

Таким образом, с помощью алгоритмов БКОСА (АСК-анализа) размерность пространства максимально снижается с минимальной потерей информации.

Для анализа статистических данных в прикладной статистике разработан ряд других алгоритмов снижения размерности. В задачи настоящей статьи не входит описание всего многообразия таких алгоритмов.

## **Литература**

1. Орлов А.И. Точки роста статистических методов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 103. С. 136–162.
2. Краскал Дж. Взаимосвязь между многомерным шкалированием и кластер-анализом // Классификация и кластер. М.: Мир, 1980. С.20-41.

3. Kruskal J.B., Wish M. Multidimensional scaling // Sage University paper series: Qualitative applications in the social sciences. 1978. №11.
4. Харман Г. Современный факторный анализ. М.: Статистика, 1972. 489 с.
5. Орлов А.И. Заметки по теории классификации. / Социология: методология, методы, математические модели. 1991. № 2. С.28-50.
6. Орлов А.И. Базовые результаты математической теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 110. С. 219–239.
7. Орлов А.И. Математические методы теории классификации // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2014. № 95. С. 23 – 45.
8. Терехина А.Ю. Анализ данных методами многомерного шкалирования. - М.: Наука, 1986. 168 с.
9. Перекрест В.Т. Нелинейный типологический анализ социально-экономической информации: Математические и вычислительные методы. - Л.: Наука, 1983. 176 с.
10. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации. М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика", 1981. - 80 с.
11. Орлов А.И. Общий взгляд на статистику объектов нечисловой природы // Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. С.58-92.
12. Орлов А.И. Предельное распределение одной оценки числа базисных функций в регрессии // Прикладной многомерный статистический анализ. Ученые записки по статистике, т.33. - М.: Наука, 1978. С.380-381.
13. Орлов А.И. Оценка размерности модели в регрессии // Алгоритмическое и программное обеспечение прикладного статистического анализа. Ученые записки по статистике, т.36. - М.: Наука, 1980. С.92-99.
14. Орлов А.И. Асимптотика некоторых оценок размерности модели в регрессии // Прикладная статистика. Ученые записки по статистике, т.45. - М.: Наука, 1983. С.260-265.
15. Орлов А.И. Об оценивании регрессионного полинома // Заводская лаборатория. Диагностика материалов. 1994. Т.60. № 5. С.43-47.
16. Орлов А.И. Некоторые вероятностные вопросы теории классификации // Прикладная статистика. Ученые записки по статистике, т.45. - М.: Наука, 1983. С.166-179.
17. Orlov A.I. On the Development of the Statistics of Nonnumerical Objects // Design of Experiments and Data Analysis: New Trends and Results. - М.: ANTAL, 1993. P.52-90.
18. Орлов А.И. Методы снижения размерности // Приложение 1 к книге: Толстова Ю.Н. Основы многомерного шкалирования: Учебное пособие для вузов. – М.: Издательство КДУ, 2006. - 160 с.
19. Орлов А.И. Асимптотика решений экстремальных статистических задач // Анализ нечисловых данных в системных исследованиях. Сборник трудов. Вып.10. - М.: Всесоюзный научно-исследовательский институт системных исследований, 1982. С. 4-12.
20. Орлов А.И. Организационно-экономическое моделирование: учебник : в 3 ч. Часть 1: Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана. – 2009. – 541 с.

21. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. <http://elibrary.ru/item.asp?id=18632909>

## References

1. Orlov A.I. Točki rosta statističeskijh metodov // Politematičeskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 103. S. 136–162.
2. Kraskal Dzh. Vzaimosvjaz' mezhdu mnogomernym shkalirovanijem i klaster-analizom // Klassifikacija i klaster. M.: Mir, 1980. S.20-41.
3. Kruskal J.B., Wish M. Multidimensional scaling // Sage University paper series: Qualitative applications in the social sciences. 1978. №11.
4. Harman G. Sovremennyj faktornyj analiz. M.: Statistika, 1972. 489 s.
5. Orlov A.I. Zametki po teorii klassifikacii. / Sociologija: metodologija, metody, matematičeskie modeli. 1991. № 2. S.28-50.
6. Orlov A.I. Bazovye rezul'taty matematičeskoj teorii klassifikacii // Politematičeskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2015. № 110. S. 219–239.
7. Orlov A.I. Matematičeskie metody teorii klassifikacii // Politematičeskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. 2014. № 95. S. 23 – 45.
8. Terehina A.Ju. Analiz dannyh metodami mnogomernogo shkalirovanija. - M.: Nauka, 1986. 168 s.
9. Perekrest V.T. Nelinejnyj tipologičeskij analiz social'no-jekonomičeskoj informacii: Matematičeskie i vychislitel'nye metody. - L.: Nauka, 1983. 176 s.
10. Tjurin Ju.N., Litvak B.G., Orlov A.I., Satarov G.A., Shmerling D.S. Analiz nečislovoj informacii. M.: Nauchnyj Sovet AN SSSR po kompleksnoj probleme "Kibernetika", 1981. - 80 s.
11. Orlov A.I. Obshhij vzgljad na statistiku ob#ektov nečislovoj prirody // Analiz nečislovoj informacii v sociologičeskijh issledovanijah. - M.: Nauka, 1985. S.58-92.
12. Orlov A.I. Predel'noe raspredelenie odnoj ocenki čisla bazisnyh funkcij v regressii // Prikladnoj mnogomernyj statističeskij analiz. Učenyje zapiski po statistike, t.33. - M.: Nauka, 1978. S.380-381.
13. Orlov A.I. Ocenka razmernosti modeli v regressii // Algoritmičeskoe i programmnoe obespečenie prikladnogo statističeskogo analiza. Učenyje zapiski po statistike, t.36. - M.: Nauka, 1980. S.92-99.
14. Orlov A.I. Asimptotika nekotoryh ocenok razmernosti modeli v regressii // Prikladnaja statistika. Učenyje zapiski po statistike, t.45. - M.: Nauka, 1983. S.260-265.
15. Orlov A.I. Ob ocenivanii regressionnogo polinoma // Zavodskaja laboratorija. Diagnostika materialov. 1994. T.60. № 5. S.43-47.
16. Orlov A.I. Nekotorye verojatnostnye voprosy teorii klassifikacii // Prikladnaja statistika. Učenyje zapiski po statistike, t.45. - M.: Nauka, 1983. S.166-179.
17. Orlov A.I. On the Development of the Statistics of Nonnumerical Objects // Design of Experiments and Data Analysis: New Trends and Results. - M.: ANTAL, 1993. R.52-90.

18. Orlov A.I. Metody snizhenija razmernosti // Prilozhenie 1 k knige: Tolstova Ju.N. Osnovy mnogomernogo shkalirovanija: Uchebnoe posobie dlja vuzov. – M.: Izdatel'stvo KDU, 2006. - 160 s.

19. Orlov A.I. Asimptotika reshenij jekstremal'nyh statisticheskikh zadach // Analiz nechislovyh dannyh v sistemnyh issledovanijah. Sbornik trudov. Vyp.10. - M.: Vsesojuznyj nauchno-issledovatel'skij institut sistemnyh issledovanij, 1982. S. 4-12.

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

21. Lucenko E.V. Avtomatizirovannyj sistemno-kognitivnyj analiz v upravlenii aktivnymi ob#ektami (sistemnaja teorija informacii i ee primenenie v issledovanii jekonomicheskikh, social'no-psihologicheskikh, tehnologicheskikh i organizacionno-tehnicheskikh sistem): Monografija (nauchnoe izdanie). – Krasnodar: KubGAU. 2002. – 605 s. <http://elibrary.ru/item.asp?id=18632909>