УДК 519.237(075.8)

А.Ю. Трусова[*]

МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ СТРУКТУРЫ ОБЩЕСТВА

Данная статья посвящена использованию методов многомерного шкалирования для анализа социального самочувствия населения современной России. Применяя методы метрического и неметрического шкалирования, восемь типогрупп были представлены в двумерном координатном пространстве социального самочувствия.

Ключевые слова и фразы: многомерное шкалирование, многомерное пространство, двумерное пространство.

В настоящее время методы многомерного шкалирования (МШ) продолжают бурно развиваться в разделах математики и теории анализа данных и находят самое разнообразное применение в экономике и социальных науках [1]. Методы МШ являются универсальными и многофункциональными [2]. Универсальность объясняется возможностью использования самых различных отношений между изучаемыми признаками и объектами (сходства, предпочтения). Многофункциональность определяется тем, что методы МШ квалифицируются и как методы « визуализации данных», и как методы понижения размерности, и как методы поиска латентных факторов [3].

Общая схема шкалирования

Схема включает ряд последовательных этапов [4]. Первый этап – создание матрицы попарных различий или матрицы субъективных предпочтений. На втором этапе решается формальная задача построения координатного пространства и размещения в нем точек-объектов таким образом, чтобы расстояния между ними, определяемые по введенной метрике, соответствовали исходным различиям. Для построения искомого координатного пространства используется аппарат линейной или нелинейной оптимизации. Вводится критерий качества отображения, называемый стрессом и измеряющий степень рас-

---

Трусова Алла Юрьевна (trusova@ssu.samara.ru), кафедра математики, информатики и математических методов в экономике Самарского государственного университета, 443011, Россия, г. Самара, ул. Акад. Павлова, 1.

хождения между исходными различиями (близостями) и результирующими расстояниями (скалярными произведениями). Отыскивается такая конфигурация точек, которая даёт минимальное значение этого стресса. Значения координат этих точек и являются решением задачи. Далее строится геометрическое представление объектов в пространстве невысокого числа измерений. Объекты, которым в исходной матрице соответствуют большие меры различий, должны находиться далеко друг от друга, а объекты, которым соответствуют малые меры различий, – близко. Формальным критерием адекватности может служить коэффициент корреляции между исходными и результирующими данными, он должен быть достаточно высоким. Третий этап – анализ и интерпретация полученных результатов. Координатные оси теоретического пространства получают смысловое содержание и интерпретируются как факторы, определяющие расхождение между объектами.

Исходные данные, их представление и первичная обработка в МШ

В МШ основным источником данных исходной информации являются эксперты, субъективно воспринимающие и оценивающие относительное расположение объектов наблюдения в реальных условиях, или результаты прямой регистрации сведений о состоянии и поведении объектов. В МШ имеются две возможности для общего представления входной информации:

1. Матрица условных вероятностей, которая определяется относительными данными «по узнаванию стимулов». Строки такой матрицы представляют собой перечень стимулов, предъявляемых для оценки, столбцы – стимулов, распознанных экспертами.

2. Матрица мер различия профилей, в которой строки – это перечень объектов наблюдения, столбцы – характерные признаки.

На начальном этапе исследования решается задача по стандартизации исходных данных, которая снижает вероятность получения вырожденных решений.

Для измерения расстояний между стимулами в МШ используются различные метрики. Обычно в искомом пространстве вводится метрика Минковского, согласно которой расстояния вычисляются по формуле: $d_{jk} = \left| \sum_t \left| x_{jt} - x_{kt} \right|^p \right|^{\frac{1}{p}}$, где x_{jt} – t-я координата j-го объекта, а p - константа Минковского. Наиболее часто используемым случаем метрики Минковского является Евклидова метрика, соответствующая значению параметра $p=2$: $d_{jk} = \left(\sum \left(x_{jt} - x_{jk} \right)^2 \right)^{\frac{1}{2}}$. Другим частным случаем расстояния Минковского является так называемое расстояние *city-block,* соответствующее значению параметра $p=1$: $d_{jk}=\sum_{t}\left|x_{jt}-x_{kt}\right|$. Метрика *city-block* подходит для стимулов, которые имеют тенденцию быть анализируемыми вдоль некоторых выделенных направлений, причем субъективно не допускается произвольное вращение этих направлений.

Для построения матрицы различий необходимо рассчитать расстояния между всеми парами наблюдаемых объектов и свести результаты в матрице симметричного вида. После нахождения матрицы различий приступают к выполнению шагов алгоритма многомерного шкалирования.

Метрический и неметрический подходы к МШ

Методы неметрического МШ применяются для обработки ранговых (порядковых) данных. Решающим условием, обеспечивающим адекватность аналитических выводов, становится соответствие монотонных связей эмпирических и теоретических данных, т.е. если реально существует порядковая зависимость $\delta_{ij}<\delta_{tl}$, то в определяемом шкальном пространстве соответственно должно выполняться неравенство $d_{ij}<d_{tl}$. Вид монотонности заранее неизвестен, и методом проб подбирается функция, наилучшим образом описывающая эмпирические данные: линейная, степенная, показательная или логарифмическая.

Метрический подход к МШ

Метод Торгерсона базируется на предположении о том, что различия между объектами равны расстояниям между точками в Евклидовом пространстве. Он начал с построения матрицы Δ^{*} с двойным центрированием, элементы δ_{ij}^{*} которой посчитаны непосредственно по матрице данных. Матрица с двойным центрированием – это матрица, у которой среднее значение элементов каждой строки и каждого столбца равно нулю. Каждый элемент новой матрицы Δ^{*} определяется соотношением: $\delta_{ij}^{*}=-\frac{1}{2}\left(\delta_{ij}^{2}-\delta_{i.}^{2}-\delta_{.j}^{2}+\delta_{..}^{2}\right)$, где $\delta_{i\cdot}^{2}$ – средняя для характеристик различий в j-х столбцах i-й строки, возведенных в квадрат: $\delta_{i\cdot}^{2}=\frac{1}{j}\sum_{j}\delta_{ij}^{2}$; $\delta_{\cdot j}^{2}$ – средняя для характеристик различий в i-х строках j-го столбца, возведенных в квадрат: $\delta_{.j}^{2}=\frac{1}{i}\sum_{i}\delta_{ij}^{2}$; $\delta_{..}^{2}$ – средняя величина для квадратов характеристик различий матрицы : $\delta_{..}^2 = \frac{1}{ij}\sum_i \sum_j \delta_{ij}^2$. По Торгенсону, для исчисленных значений δ_{ij}^{*}, если $\delta_{ij} = d_{jk} = \left(\sum \left(x_{jt} - x_{jk} \right)^2 \right)^{\frac{1}{2}}$, имеет силу равенство: $\delta_{ij}^{*} = \sum_k x_{ik} x_{jk}$, или в матричном виде: $\Delta^{*} = XX'$, где X – матрица координат стимулов размерности $I * K$. Определение матрицы X, как видно, тождественно решению задачи поиска собственных векторов, при этом могут использоваться: метод главных компонент, методы факторного анализа.

Неметрический подход к МШ

Существует большое число алгоритмов для выполнения неметрического шкалирования. Основные алгоритмические шаги:

Шаг 1. Получение матрицы различий, содержащей ранговые данные – характеристики непохожести анализируемых объектов.

Шаг 2. Поиск стартовой конфигурации. Эта проблема может быть решена с использованием алгоритма Торгерсона или Краскала.

Шаг 3. Стандартизация расстояний и оценок координат, получение стартовой конфигурации. В начале каждой итерации проводится стандартизация текущих расстояний и оценок координат с целью сохранить пропорции ортонормированного стимульного пространства и избежать вырожденных решений.

Шаг 4. Неметрический этап. На данном этапе данные о различии и стандартизированные оценки расстояний из предыдущей итерации (или из стартовой конфигурации) используются для получения отклонений. В теоретическом пространстве шкал X_K монотонность исходных данных может нарушаться. Корректировка теоретических величин расстояний d_{ij} производится при неизменных оценках координат стимулов и таким образом, чтобы восстановить общую тенденцию к возрастанию в исходных данных о различиях. Диаграмма Шепарда позволяет наглядно отобразить несоответствие исходных и теоретических ранговых оценок.

Улучшаются оценки расстояний центрированием отклоняющихся от прямой расстояний d_{ij} посредством расчета арифметической средней. Новые центрированные значения закрепляются за двумя парами объектов, в которых возникли нарушения монотонности. С переходом от оценок d_{ij}^{c} к уточненным оценкам d_{ij}^{c+1} (*c+1* – первой итерации) неметрический этап заканчивается.

Шаг 5. Метрический этап. На данном этапе имеющимся исходным и уточненным величинам расстояний (d_{ij}^{c} и d_{ij}^{c+1}) находят уточненные оценки координат. Для расчетов используют формулу Лингоса – Роскама: $x_{ik}^{c+1} = x_{ik} - \frac{1}{j}\sum_{j}\left(1 - \frac{d_{ij}^{c+1}}{d_{ij}^{c}}\right)\left(x_{ik}^{c} - x_{jk}^{c}\right)$. Подобные расчеты проводятся для всех участвующих в анализе объектов, после этого уже по новым оценкам координат (x_{ij}^{c+1}) находят расстояния между стимулами в теоретическом пространстве (d^{c+1}).

Шаг 6. Оценка соответствий монотонных ранговых эмпирических и теоретических данных. Собственно проверке на монотонность подлежат теоретические данные d_{ij}^{c} и d_{ij}^{c+1}, рассматривается степень их улучшения на предыдущей итерации. Если улучшение существенно, итерация возобновляется после стандартизации полученных на метрическом этапе оценок координат и расстояний. Если же улучшений мало, итерации заканчиваются. Оценивание соответствий теоретических результатов эмпирическим данным осуществляется при помощи специальных стресс – формул Краскала и Юнга

$$S_1 = \left(\frac{\sum_{ij}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{ij} \hat{d}_{ij}^{\,2}}\right)^{1/2}, \quad S_2 = \left(\frac{\sum_{ij}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{ij}\left(\hat{d}_{ij} - \hat{d}_{..}\right)^2}\right)^{1/2}, \quad SS_1 = \left(\frac{\sum_{ij}\left(d_{ij}^2 - \hat{d}_{ij}^2\right)^2}{\sum_{ij} \hat{d}_{ij}^4}\right)^{\frac{1}{2}},$$

$$SS_2 = \left(\frac{\sum_{ij}\left(d_{ij}^2 - \hat{d}_{ij}^2\right)^2}{\sum_{ij}\left(\hat{d}_{ij}^2 - \hat{d}_{..}^2\right)^2}\right)^{\frac{1}{2}}$$ или коэффициента отчуждения Гутмана:

$$k = (1-\mu^2)^{\frac{1}{2}}, \text{ где } \mu = \frac{\sum d_{ij}\hat{d}_{ij}}{\left[\left(\sum_{ij} d_{ij}^2\right)\left(\sum_{ij}\hat{d}_{ij}^2\right)\right]^{\frac{1}{2}}}.$$

Применение метода МШ при анализе социального самочувствия населения

Социальное самочувствие является интегральным показателем, отражающим ощущения человека, степень его удовлетворенности своим положением в данном обществе и адаптированности к новым социально-экономическим условиям. В этой связи интересным представляется определение местонахождения социальных групп, различающихся между собой по степени наличия в них богатства и власти, в пространстве социального самочувствия и создание их образа средствами МШ. В качестве объекта исследования изучались восемь социальных групп, различающихся между собой по степени наличия в них богатства и власти: бедные-бесправные, бедные-приближенные к власти, благополучные-бесправные, благополучные-приближенные к власти, благополучные-властные, богатые-бесправные, богатые-приближенные к власти, богатые-властные (табл. 1).

В качестве исходных признаков рассматривались: X_1 – наличие качеств, необходимых современной России, X_2 –самооценка профессионализма, X_3 –опасения потерять работу, X_4 –самооценка общественного признания, X_5 – удовлетворенность материальным положением, X_6 –удовлетворенность жизнью, X_7 – беспокойство о « завтрашнем дне», X_8 –способность самостоятельно изменить жизнь к лучшему, X_9 – здоровье. Исходные данные были опубликованы в центральной печати [5].

Таблица 1

Частные ранги социального самочувствия типогрупп.
(1-низшие ранги, 8-высшие ранги)

	X1	X2	X3	X4	X5	X6	X7	X8	X9
Бедные – бесправные	1	1	1	1	1	1	1	1	1
Бедные-приближенные к власти	4	3	3	3	2	2	2	4	3
Благополучные – бесправные	3	4	4	2	3	3	3	3	2
Благополучные – приближенные к власти	5	5	5	5	4	4	4	5	5
Благополучные – властные	7	6	8	7	5	6	6	8	8
Богатые- бесправные	2	2	2	4	6	5	5	2	4
Богатые-приближенные к власти	8	7	7	6	7	8	8	6	7
Богатые – властные	6	8	6	8	8	7	7	7	6

Для построения матрицы различий Δ были рассчитаны расстояния между всеми парами наблюдаемых объектов. В качестве меры расстояний была выбрана тривиальная метрика Евклида. Для нахождения матрицы координат стимулов использовался метод главных компонент, были вычислены собственные значения и собственные векторы матрицы Δ^*. В табл. 2 представлены положительные собственные значения, полученные по данным матрицы Δ^*.

Таблица 2

Собственные значения матрицы Δ^*

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
326,746	32,941	8,317	7,637	1,938	0,305	0,112

Каждое собственное значение связано с одной координатной осью. Оно равно сумме квадратов шкальных значений стимулов по этой оси. Для определения размерности теоретического пространства стимулов проводился анализ графиков зависимостей собственных значений от числа координатных осей (рис. 1). С учетом уровня информативности в анализе остаются 2 координатные оси (табл. 3).

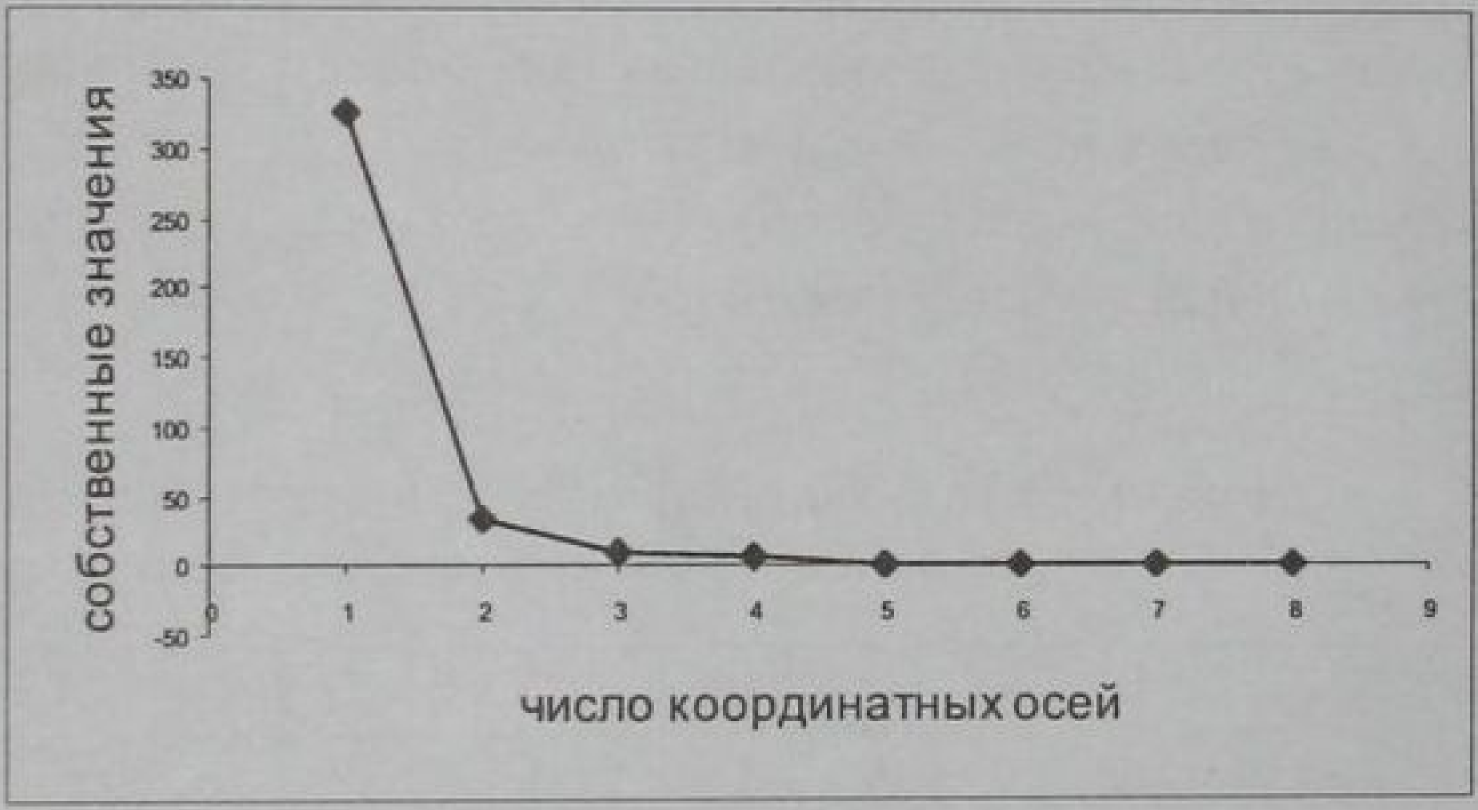


Рис. 1. График зависимости собственных значений от числа координатных осей

Далее, реализуя алгоритм неметрического МШ, была проведена стандартизация оценок координат и расстояний: $Z_{ij} = \dfrac{x_{ij} - \overline{x_j}}{\sigma_j}$. Матрица стандартизированных оценок расстояний представлена в таблице 3.

Таблица 3

Значения координат стимулов

	Нестандартизированные оценки		Стандартизированные оценки		Новые координаты стимула после 1 итерации		Новые координаты стимула после 2 итерации	
Бедные – бесправные	-0,580	-0,021	-4,002	-0,147	-2,839	-0,734	-2,962	-0,640
Бедные-приближенные к власти	-0,270	-0,356	-1,871	-2,493	-2,154	-6,948	-1,980	-7,479
Благополучные бесправные	-0,250	-0,093	-1,733	-0,654	-1,529	-3,310	-1,224	-3,538
Благополучные – приближенные к власти	0,029	-0,220	0,183	-1,538	0,221	-6,121	-0,369	-8,125
Благополучные – властные	0,380	-0,431	2,601	-3,016	1,531	-5,523	1,592	-4,405
Богатые – бесправные	-0,159	0,745	-1,108	5,216	-1,571	17,431	-1,532	19,273
Богатые-приближенные к власти	0,433	0,143	2,965	1,002	2,671	2,273	2,470	1,798
Богатые- властные	0,433	0,233	2,965	1,631	3,671	2,931	4,005	3,117

На неметрическом этапе, предназначенном для упорядочения оценок расстояний между стимулами, строилась диаграмма Шепарда, которая отражала несоответствие между исходными и теоретическими ранговыми оценками.

На метрическом этапе были найдены уточненные оценки координат (новые координаты стимула). Для расчетов использовалась формула Лингоса – Роскама. После этого уже по новым оценкам координат (x_{ik}^{c+1}) были найдены расстояния между стимулами в теоретическом пространстве (d^{c+1}), и первая итерация на этом была закончена. Оценивание соответствия теоретических результатов эмпирическим данным было произведено по стресс-формулам Краскала.

После проведения оценки соответствия монотонных ранговых эмпирических и теоретических данных была проведена еще одна итерация. На метрическом этапе использовались оценки расстояний на предыдущей итерации (d_{ij}^{c}) и оценки координат на предыдущей итерации (x_{ik}) для получения новых оценок координат(x_{ik}^{c+1}), по которым рассчитывались новые оценки расстояний(d_{ij}^{c+1}) (табл. 3).

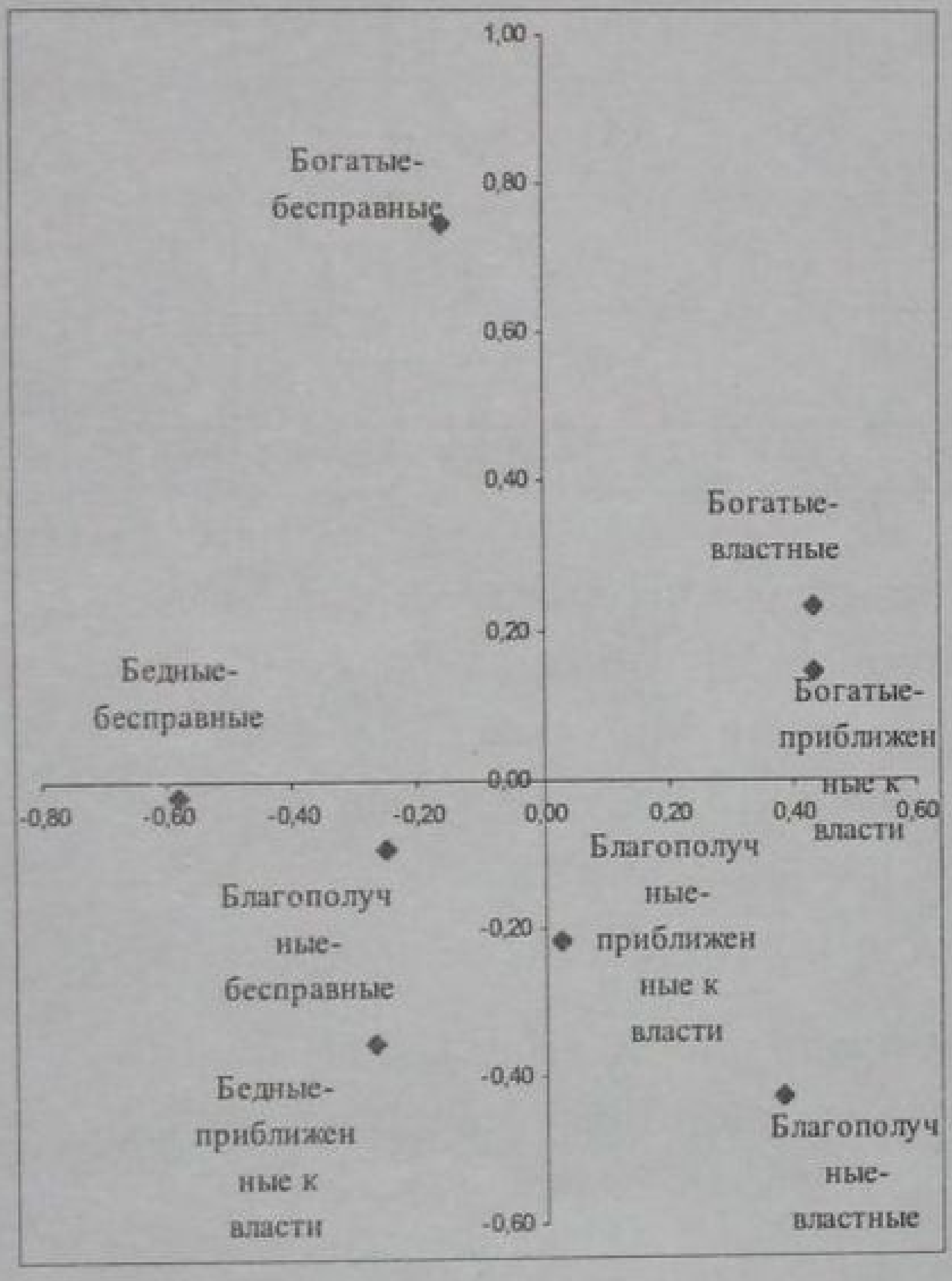
После проведения второй итерации существенных улучшений обнаружено не было (таблица 4), поэтому итерации были закончены.

Таблица 4

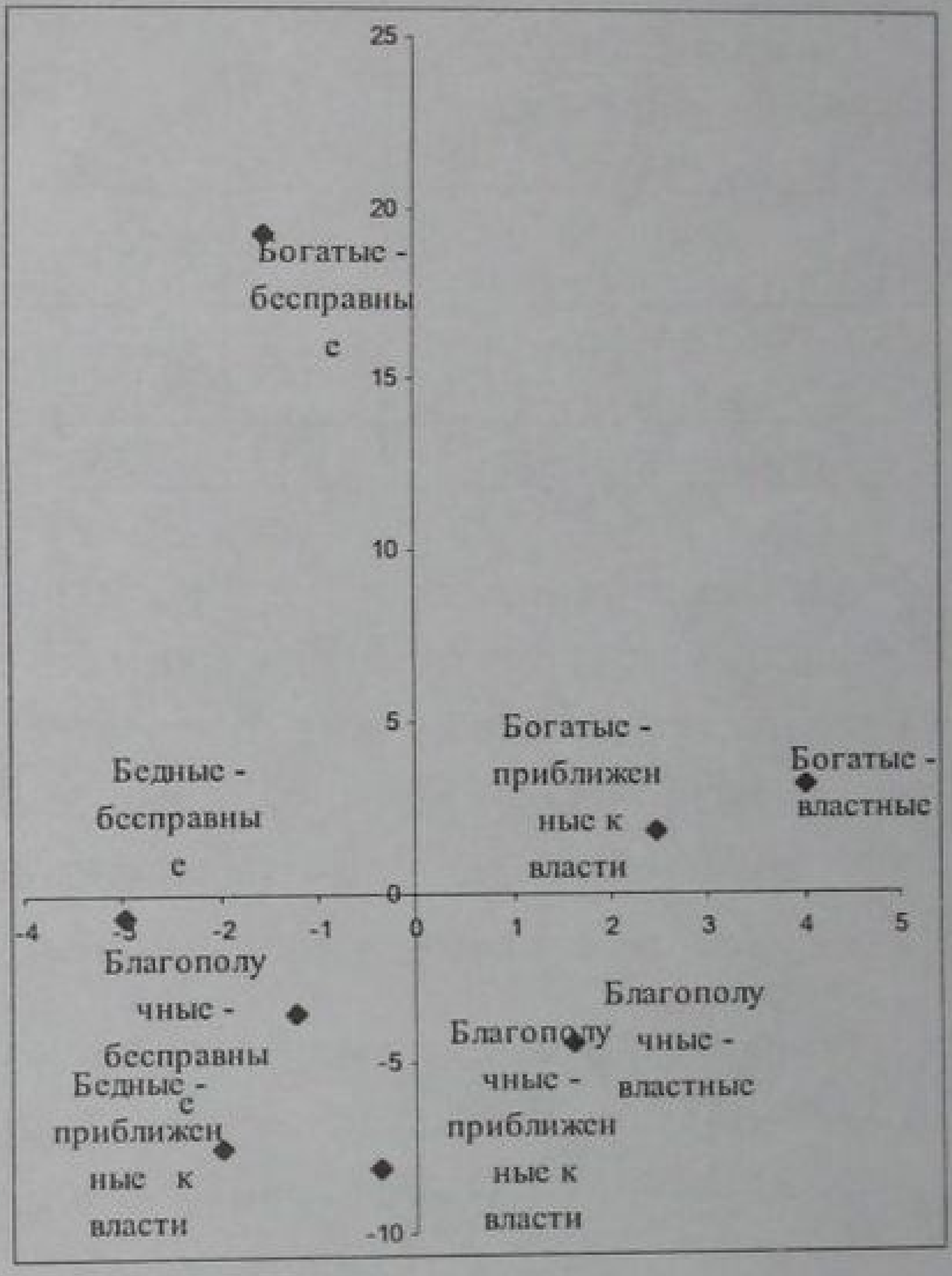
Уточненные оценки координат стимулов

	После 1 итерации	После 2 итерации
S_1	0,955113654	0,964428904
S_2	1,208835567	1,209120159

Выбранные координатные оси соответствуют субъективным и объективным факторам, влияющим на социальное самочувствие (рис. 2). Полученные результаты имеют адекватную интерпретируемость.



а



б

Рис. 3. Графическое представление координат стимулов:
а – метрическое МШ, б – неметрическое МШ.

Первая шкала (ось абсцисс) соответствует объективному фактору, влияющему на социальное самочувствие. К нему можно отнести образование, вид трудовой деятельности, материальный достаток и т.д. Вторая шкала (ось ординат) соответствует субъективному фактору, к которому относятся психологический склад личности, его самооценка, удовлетворенность жизнью, характеристики ценностно-мотивационной сферы, личностные амбиции.

Библиографический список:

1. Трусова, А.Ю. Математическое моделирование социальных процессов / А.Ю. Трусова, И.С. Макарова // Образовательные технологии: межвуз. сб. науч. тр. – Воронеж, 2003. – Вып. 10. – С. 87-91
2. Дейвисон, М. Многомерное шкалирование: Методы наглядного представления данных / М. Дейвисон; пер. с англ.- М.: Финансы и статистика, 1988. – 254 с.
3. Дубров, А.М. Многомерные статистические методы: учебник / А.М. Дубров, В.С. Мхитарян, Л.И. Трошин. – М.: Финансы и статистика, 2000. – 352 с.
4. Сошникова, Л.А. Многомерный статистический анализ в экономике: учеб. пособие для вузов / Л.А. Сошникова, В.Н. Тимашевич, Г. Уебе [и др.]; под ред. проф. В.Н. Тимашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.
5. Душацкий, Л.Е. Материально-властные ресурсы россиян в самооценке и социальном самочувствии / Л.Е. Душацкий // СОЦИС. – 2004. – №4 – С. 64-71.

A.Yu. Trusova

MULTI-DIMENSIONAL SCALING OF SOCIAL STRUCTURE

The paper is devoted to the studying of multidimensional scaling and to the interpretation of the objects in the multidimensional space. In this article, methods of multi-dimensional scaling are applied to analyze the social mood and conditions of contemporary population of Russia. Applying the methods of metrical and non-metrical scaling eight typical groups were presented in two-dimensional plane. These groups differ from each other with the level of financial position and authority. Social health of definite social group is satisfaction of her position in existent community.

Keywords and phrases: *multidimensional scaling, multidimensional space, two-dimensional plane*

Статья принята в печать в окончательном варианте 04.12.08 г.