

*Г.Л. ВОРОНИН*

## **ЕЩЕ РАЗ О «КЛАСТЕРАХ НА ФАКТОРАХ»**

Статья посвящена вопросу типологизации респондентов в ходе анализа социологических данных. Рассматриваемый алгоритм выделения типов респондентов предполагает проведение на первом этапе факторного, а затем кластерного анализов. Анализ осуществляется с помощью компьютерной программы IBM SPSS Statistics 18.

*Ключевые слова:* факторный анализ методом главных компонент, корреляционный анализ, кластерный анализ методом k-средних, анализ пригодности.

### **Постановка исследовательской задачи**

Дискуссии о правомерности применения кластерного анализа к значениям факторов возникают с определенной регулярностью. Периодическое возвращение этой темы в поле зрения социологов указывает на важность обсуждаемой проблемы. Вопрос выделения «кластеров на факторах» нередко актуализируется как на страницах социологических журналов, так и в ходе защит диссертационных исследований. Необходимо подчеркнуть, что это не просто «профессиональные игры», а действительно важная проблема. Об этой проблеме заявил А.О. Крыштановский в статье «“Кластеры на факторах” — об одном распространенном заблуждении» [5, с. 172–187]. Уже название статьи показывает отрицательное отношение автора к такому алгоритму анализа социологических данных. Главный аргумент, на котором базируется отрицание «кластеров на факторах», сводится к неадекватности или низкой информативности отдельных факторов и, как следствие, низкой информационной результативности кластерного анализа методом k-средних, проведенного на «плохих» факторных значениях.

В продолжение этой темы публикуется статья Е.Г. Галицкой и Е.Б. Галицкого «“Кластеры на факторах”: как избежать распространенных ошибок?» [4]. Авторы показывают, что «неудовлетворительный результат кластерного анализа [приведенного в работе [5]. — *Г.В.*] вполне закономерен. Он является результатом двух обстоятельств: во-первых, выбора излишнего числа факторов и, во-вторых, использования этих факторов без всяких преобразований» [4, с. 157].

---

**Воронин Геннадий Леонидович** — доктор социологических наук, ведущий научный сотрудник Института социологии РАН. Адрес: 117218, Москва, ул. Кржижановского, д. 24/35, корп. 5.  
Электронная почта: VoroninG@isras.ru

В нашей статье предложен алгоритм анализа социологических данных, описывающий последовательность действий социолога по выделению типов респондентов в зависимости от их ответов на вопросы анкеты. В качестве методологического основания анализа социологических данных выступает определение социологии как социальной топологии, данное П. Бурдые в его «Социологии политики» [2, с. 55]. Это определение как нельзя лучше стимулирует социолога-исследователя к построению модели социального пространства и поиску однотипных объектов внутри этого пространства. Типологизация объектов исследования (в нашем случае респондентов) предполагает поиск гомогенных групп. При этом группы не только должны быть внутренне однородными, но и «качественно отличными друг от друга... характеризующимися типобразующими признаками, природа которых различна, и интерпретируемыми как носители различных типов изучаемого социального феномена» [7, с. 204–205].

Первый методический вопрос, возникающий в ходе решения поставленной задачи в многомерном пространстве, связан с правоммерностью применения порядковых шкал в процедуре факторного анализа. Ответ на этот вопрос, на наш взгляд, лежит как в области математики, так и в области социологии. Прежде всего, необходимо обратить внимание на то, чтобы количество объектов было на порядок больше количества подлежащих факторизации переменных. Что касается социологии, то вопрос верификации результатов факторного анализа решается с позиции соответствия полученной модели существующей социальной реальности. В статье также представлены результаты двух кластеризаций методом *k*-средних, где в качестве переменных выступают значения факторов двухфакторной модели, а затем первичные переменные (вопросы анкеты), включенные в процедуру факторного анализа.

Для реализации алгоритма «факторный анализ — кластерный анализ» воспользуемся файлом данных Российского мониторинга экономического положения и здоровья населения (РМЭЗ) 2002 года<sup>1</sup>. В ходе исследования респонденту предлагается оценить степень согласия – несогласия с приведенными в анкете семнадцатью утверждениями, касающимися его социально-психологического самочувствия. Для ответа предлагалась четырехбалльная шкала, где 1 означает полное

<sup>1</sup> РМЭЗ представляет собой серию общенациональных репрезентативных опросов, проводящихся в Российской Федерации с 1992 г. по вероятностной, стратифицированной, многоступенчатой территориальной выборке, репрезентативной на федеральном уровне. Мониторинг проводят Институт социологии РАН, Исследовательский центр «Демоскоп», Государственный университет — Высшая школа экономики и Университет Северной Каролины в Чепел Хилле (США).

несогласие, 2 — несогласие, 3 — согласие и 4 — полное согласие респондента с оцениваемым суждением. В нашем случае факторному анализу подлежит матрица 17×9664, где 17 — это количество вопросов, имеющих четырехбалльную шкалу, 9664 — количество респондентов.

### Предварительные процедуры

Одним из требований к переменным, подлежащим факторному анализу, является нормальность распределения, где среднее значение совпадает с наиболее часто встречающимся значением (модой) и с медианой — значением, которое делит распределение на две равные части, а 68% наблюдений находится в пределах одного стандартного отклонения от среднего значения. Нормальность распределения проверяется при помощи значений статистики критерия Колмогорова-Смирнова.

Таблица 1

### Средние значения переменных, N=9664

(1 — совсем не согласны, 4 — полностью согласны)

Суждения	Среднее	Станд. откл.
1. Я не могу справиться со своими проблемами	2,14	0.72
2. Иногда я чувствую, что мной помыкают в жизни	2,13	0.70
3. Я мало могу влиять на то, что со мной происходит	2,26	0.72
4. Я всегда могу выполнить задуманное	2,45	0.69
5. Я часто чувствую себя беспомощным перед проблемами, возникающими в моей жизни	2,32	0.71
6. То, что со мной произойдет в будущем, во многом зависит от меня	2,73	0.69
7. То, что я могу сделать, мало что изменит в моей жизни	2,39	0.69
8. Я думаю, что я ничем не хуже других	3,08	0.49
9. Я считаю, что у меня есть много хороших качеств	3,06	0.48
10. В целом, мне кажется, что я неудачник (неудачница)	1,91	0.65
11. Я могу все делать не хуже других	2,99	0.51
12. Я думаю, что мне особенно нечем гордиться	2,18	0.66
13. Я хорошо отношусь к самому (самой) себе	2,91	0.54
14. В целом, я удовлетворен (удовлетворена) собой	2,78	0.60
15. Иногда я чувствую себя бесполезным (бесполезной)	2,20	0.68
16. Я хотел (хотела) бы относиться к себе с большим уважением	2,76	0.61
17. Иногда мне кажется, что я нехороший человек	1,95	0.67

Учитывая, что нормальное распределение переменной в социологическом исследовании встречается крайне редко, исследователю необходимо проанализировать средние значения и стандартное отклонение (мера разброса вокруг среднего) в переменных, подлежащих факторизации, обратив при этом внимание на те из них, для которых средние значения не совпадают со смысловым средним значением по шкале. Зачем это надо делать? Это необходимо по той причине, что в

ходе интерпретации факторов респонденты, согласные и полностью согласные, с одной стороны, и несогласные и полностью несогласные, с другой, должны иметь противоположные (положительный и отрицательный) знаки в переменной «значение фактора» и, следовательно, относиться к разным типологическим группам. В случае несовпадения среднего значения переменной со смысловым средним значением шкалы респонденты, полностью согласные и частично согласные (или полностью несогласные и частично несогласные) с тем или иным суждением, попадают в разные группы. Это делает практически невозможным корректную интерпретацию факторов и, как следствие, ставит под сомнение возможность адекватного понимания позиций респондента в пространстве факторов. Учитывая, что при замере позиции респондента используется четырехбалльная шкала, переменные со средними значениями менее 2 и более 3 баллов не подлежат факторизации (см. табл. 1).

#### **Факторный анализ**

Суть факторного анализа можно определить как процедуру сокращения размерности и выделение латентных переменных, включающих первичные переменные [1, с. 717–746]. Иными словами, факторный анализ позволяет измерить объект всесторонне и в то же время компактно. При проведении факторного анализа для социолога, как правило, возникают две главных проблемы — сколько факторов выделять и как их интерпретировать.

В нашем случае в качестве первичных переменных выступают вопросы анкеты, диагностирующие уровень социально-психологической адаптации респондента. Необходимо отметить, что идеология факторного анализа сконцентрирована в двух базовых положениях, а именно: «а) сущность вещей заключена в их простых и вместе с тем многообразных проявлениях, которые могут быть объяснены с помощью комбинации нескольких основных факторов; б) общую сущность наблюдаемых вещей мы постигаем, совершая бесконечные приближения к ней» [3, с. 54]. Именно эта интерпретация основных идей факторного анализа как нельзя лучше согласуется с социологическим анализом данных.

Прежде чем применять факторный анализ, необходимо сформировать исследовательскую цель. В самом общем случае мы можем говорить о поиске и выработке теоретических положений или проверке гипотез. Это может быть разведочный (эксплораторный) факторный анализ, когда исследователь не знает ни количества, ни структуры факторов, или проверочный (конфирматорный) факторный анализ, когда проверяется исследовательская гипотеза [3, с. 31–32]. В нашем случае проверяется гипотеза о существовании двух независимых латентных факторов: социальной и психологической адаптации респондентов.

Из 17 переменных, отобранных для факторного анализа, у нас остается 13 переменных (см. табл. 1). Приступаем непосредственно к

процедуре факторного анализа при помощи компьютерной программы IBM SPSS Statistics. Воспользуемся следующими установками: метод — главные компоненты, фиксированное количество факторов — 2 (*проверяется гипотеза о существовании ортогональных факторов* — социальной и психологической адаптации); вращение факторов методом Варимакс, который является ортогональным методом вращения, минимизирующим число переменных с высокими нагрузками на каждый фактор, что позволяет упростить интерпретацию факторов. Затем сохраняем значения факторов и устанавливаем формат вывода коэффициентов на уровне не менее 0,3. (Вопрос о границах последнего показателя был решен в статье А.О. Крыштановского [5] и здесь не обсуждается.)

После получения результатов обращаем внимание на таблицу «Общности», где представлена информативность первичных переменных в факторной модели, понимаемая как квадрат множественной корреляции для переменной при использовании факторов в качестве предикторов. Никаких указаний на величину этих показателей в литературе нет, но из практики известно, что первичные переменные с показателями менее 0,2 желательно в факторизацию не включать. В нашем случае исключается из анализа переменная под номером 16 «Я хотел(-а) бы относиться к себе с большим уважением», так как показатель «Извлечение» составляет для этой переменной 0.095. Окончательно определившись с количеством переменных, используемых в факторном анализе, — в нашем случае 12 вопросов, — мы повторяем процедуру факторизации в полном объеме еще раз.

Наиболее ответственным и творческим этапом анализа является интерпретация полученных факторов. Мы используем двухфакторную модель, где доля объясненной дисперсии составляет 41,8%. Здесь, как правило, встает вопрос о нижней границе этого показателя. Решение этого вопроса находится за рамками математического анализа. Допустимая нижняя граница доли объясненной дисперсии определяется, прежде всего, областью применения результатов исследования: социология, экономика, медицина, маркетинг и др. Полагаем, что для социологических и маркетинговых данных этот показатель не должен опускаться ниже 30%.

Аналитик должен обратить внимание еще на два важных показателя качества факторной модели. Это мера Кайзера-Мейера-Олкина и критерий сферичности Бартлетта.

Мера Кайзера-Мейера-Олкина (КМО) является важным показателем проверки адекватности применения факторного анализа. Высокие значения этого показателя — от 0,5 до 1,0 — указывают на возможность проведения факторного анализа, тогда как низкие — менее 0,5 — указывают, что применение факторного анализа недопустимо. КМО является мерой выборочной адекватности, используемой для

проверки гипотезы о том, что частные корреляции между переменными малы [1, с. 720]. Критерий сферичности Бартлетта проверяет гипотезу о том, что корреляционная матрица является единичной матрицей, то есть то, что выборка взята из многомерного нормального распределения. Если нулевую гипотезу отвергнуть нельзя, то факторная модель непригодна (нулевая гипотеза отвергается при значимости критерия менее 0,01) [1, с. 720].

В результате факторизации 12 первичных переменных и вращения факторов получили хорошо структурированную матрицу (табл. 2).

Таблица 2

**Матрица факторных нагрузок (после вращения)**

Суждения	Компонента	
	1	2
5. Я часто чувствую себя беспомощным перед проблемами, возникающими в моей жизни	0,739	
3. Я мало могу влиять на то, что со мной происходит	0,733	
1. Я не могу справиться со своими проблемами	0,700	
7. То, что я могу сделать, мало что изменит в моей жизни	0,619	
2. Иногда я чувствую, что мной помыкают в жизни	0,602	
15. Иногда я чувствую себя бесполезным (бесполезной)	0,540	
12. Я думаю, что мне особенно нечем гордиться	0,434	
13. Я хорошо отношусь к самому (самой) себе		0,760
14. В целом, я удовлетворен (удовлетворена) собой		0,741
11. Я могу все делать не хуже других		0,600
4. Я всегда могу выполнить задуманное		0,471
6. То, что со мной произойдет в будущем, во многом зависит от меня		0,399

Первый фактор включает переменные, диагностирующие уровень социальной адаптации респондента: в зоне положительных значений фактора сосредоточены респонденты, заявляющие, что они не могут справиться с возникающими на их жизненном пути проблемами, в зоне отрицательных значений фактора сосредоточены респонденты, хорошо приспособившиеся к современным условиям. Определяемся с названием первого фактора — «Уровень социальной адаптации».

Второй фактор группирует переменные, где респондент оценивает свое психологическое самочувствие. Так как максимальное значение по шкале (4 балла) означает согласие респондента с тезисом, в зоне положительных значений фактора сосредоточены респонденты, заявившие о комфортном психологическом самочувствии. В зоне отрицательных значений фактора находятся респонденты, заявившие о наличии психологических проблем. По аналогии с первым второй фактор определяем как «Уровень психологической адаптации».

Процедура факторного анализа позволяет нам «выявить и осуществить конструирование и анализ внутренних факторов по информации

об их “внешних” проявлениях» [3, с. 5]. В нашем примере осуществлен переход от матрицы данных  $12 \times 9664$  к матрице  $2 \times 9664$  при одновременном снижении информативности со 100% до 41,8% (доля объясненной дисперсии).

Выполним проверку качества полученной модели. Это можно сделать разными способами. В качестве примера воспользуемся процедурой «анализ пригодности». Это модель внутренней согласованности, основанная на средней межпунктовой корреляции [8, с. 409–416]. Проверим согласованность первичных переменных, образующих первый и второй факторы. Для этого достаточно подсчитать коэффициент Альфа Кронбаха (Cronbach's Alpha), показывающий уровень внутренней согласованности модели (0 — модель не согласована, 1 — модель согласована). Для первого фактора «Уровень социальной адаптации», включающего семь первичных переменных, этот показатель составил 0,776 (при максимальном значении 1), а для второго фактора «Уровень психологической адаптации», включающего пять первичных переменных, — 0,609. Эта проверка не является обязательной, но, тем не менее, она подтверждает гипотезу о существовании двух латентных переменных, фиксирующих уровень социально-психологической адаптации.

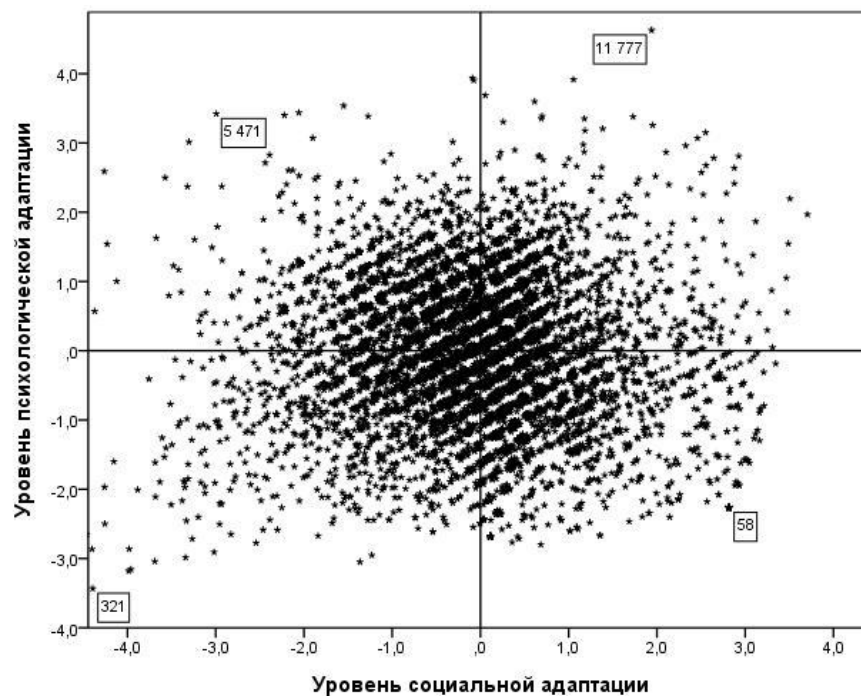


Рис. 1. Модель расположения респондентов в пространстве социально-психологической адаптации (двухфакторное пространство)

Еще одним важным аргументом в пользу применения факторного анализа к социологическим данным можно считать тот, что в этом случае социологическая анкета может выступать в качестве социологического теста (по аналогии с психологическим тестом), где респондент не только оценивает ту или иную социальную ситуацию, но и презентует себя в социальном пространстве. Результаты самопрезентации респондентов в пространстве социально-психологической адаптации визуализированы и представлены на рис. 1.

Респонденты, занимающие в двухфакторном пространстве социально-психологической адаптации строго определенное место, обозначены звездочками. Положение в пространстве предоставляет исследователю как качественную, так и количественную информацию о респонденте. Так, респондент под номером 11777, отвечая на вопросы анкеты, согласился (4 пункт шкалы) со всеми предложенными к оценке тезисами, тогда как респондент номер 321 не согласился с этими тезисами (1 пункт шкалы). Респондент номер 5471 не согласен с вопросами, образующими первый фактор, но соглашается с вопросами, образующими второй фактор. Респондент номер 58 занимает противоположную позицию по отношению к респонденту номер 5471 как в модели пространства (см. рис. 1), так и в ответах на первичные вопросы анкеты.

#### **Типологизация респондентов в пространстве социально-психологической адаптации**

Полученная в ходе анализа двухфакторная модель изучаемого явления предполагает выделение пяти типологических групп респондентов. Выделение этих групп диктуется следующими обстоятельствами: четыре группы по количеству квадрантов (первый квадрант — положительные значения первого и второго факторов; второй квадрант — отрицательные значения первого и положительные значения второго факторов; третий квадрант — отрицательные значения первого и второго факторов; четвертый квадрант — положительное значение первого и отрицательное значение второго факторов). Пятая группа формируется вокруг нулевых значений обоих факторов. Решить задачу выделения типов (групп, кластеров) можно различными путями, но в настоящей статье рассмотрим наиболее типичный способ, связанный с методом кластерного анализа.

#### ***Метод k-средних***

Кластерный анализ предполагает поиск и выделение однотипных объектов в пространстве заданных характеристик. В основе кластерного анализа лежит гипотеза о том, что вычленение однородных групп зависит от близости объектов друг к другу. То есть объекты, находящиеся в одном кластере, должны быть максимально похожими; но не менее важная задача кластерного анализа состоит в том,

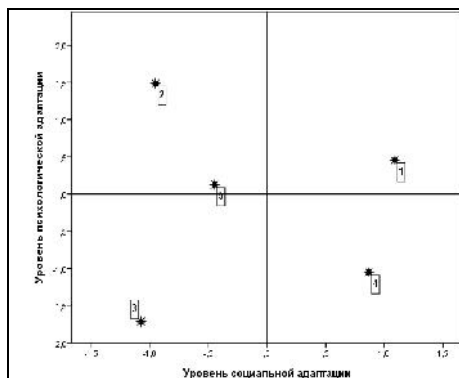


чтобы сделать максимально непохожими друг на друга выделенные кластеры. Одним из популярных в социологии методов кластеризации данных является метод *k*-средних. Особенность этого метода — в том, что вначале определяются центры кластеров, а затем идет формирование групп (кластеров) в пределах заданного от центра порогового значения. Используя метод *k*-средних, исследователь сам определяет количество кластеров. Как пишет М.Ф. Черныш, «Огрубляя детали этой процедуры, можно сказать, что в данном случае выполняется одномерный дисперсионный анализ, в рамках которого размер групп неизвестен, а наблюдения включаются в кластеры так, чтобы *F*-величина достигла максимальных значений» [8, с. 136].

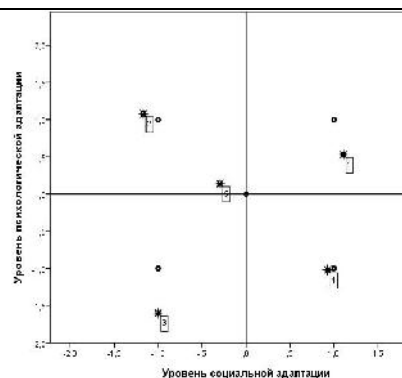
Необходимо помнить, что метод кластерного анализа является разведочным, а это предполагает проведение нескольких процедур для поиска оптимальной модели. Верификация результатов исследования и определение оптимальности полученной модели осуществляются на уровне смысловых характеристик выделенных групп, а именно с точки зрения того, будут ли полезны эти кластеры в проводимом социологическом исследовании.

Для типологизации респондентов в пространстве двух факторов методом *k*-средних установим в окне «число кластеров» программы IBM SPSS Statistics значение «5». В качестве переменных кластеризации задаем значение факторов «Уровень социальной адаптации» и «Уровень психологической адаптации». Сохраняем переменную с кластерным номером наблюдения. На рисунке 2 показаны центры выделенных кластеров. Такое расположение выглядит вполне логичным, то есть представлены четыре группы по числу квадрантов и пятая группа находится в центре, но несколько сдвинута в зону отрицательных значений по первому фактору. Практика проведения и подачи результатов социологического исследования с применением факторного анализа убеждает в необходимости давать названия выделенным кластерам и отдельным частям социального пространства в пределах квадрантов. Респондентов, образующих второй кластер и расположенных во втором квадранте, назовем «комфортными». Эта группа заявляет о своем высоком уровне социальной и психологической адаптации. Следуя логике факторного анализа, респонденты, вошедшие в четвертый кластер, являются «дискомфортными»: есть сложности во взаимоотношении с социальным окружением и плохое психологическое самочувствие.

Проведем повторный кластерный анализ методом *k*-средних, но зададим начальные центры кластеров (на рисунке 3 они обозначены кружочком). Эта процедура предусмотрена в программе IBM SPSS Statistics. Кластерные центры практически совпадают, что в какой-то мере указывает на устойчивость полученной модели (рис. 2, 3). Но произошла некоторая корректировка пространственного расположения пятого кластера, что, на наш взгляд, несколько улучшает модель.



**Рис. 2. Расположение кластерных центров (метод k-средних без установки начальных центров)**



**Рис. 3. Расположение кластерных центров (метод k-средних с установкой начальных центров)**

В таблице 3 представлены результаты двух видов кластеризации методом k-средних. Несовпадение результатов составило 7,5%. Процедура установки кластерных центров в процессе кластеризации методом k-средних необходима для выделения кластеров, совпадающих по своим содержательным характеристикам с содержательной интерпретацией факторного пространства, в нашем случае — двухфакторной модели.

*Таблица 3*

**Таблица сопряженности двух кластеризаций методом k-средних переменных «Уровень социальной адаптации» и «Уровень психологической адаптации»,  $N=9664$**

	Кластеризации					Итого
	без установки начальных центров кластеров 1-5	с установкой начальных центров кластеров 1-5				
	1	2	3	4	5	
1	2046	0	0	45	84	2175
2	43	901	0	0	30	974
3	0	0	812	0	0	812
4	0	0	44	1438	23	1505
5	0	354	96	0	3748	4198
Итого	2089	1255	952	1483	3885	9664

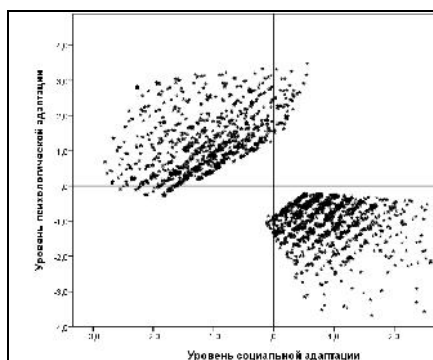
В то же время, несовпадение результатов кластерного анализа методом k-средних с выделением пяти кластеров в пространстве двух факторов (с установкой начальных кластерных центров) и в пространстве 12 первичных переменных составляет 41% (табл. 4). Это говорит о серьезной рассогласованности в моделях и сложностях, возникающих в процессе выбора путей и методов типологизации данных и интерпретации полученных результатов.

Таблица 4

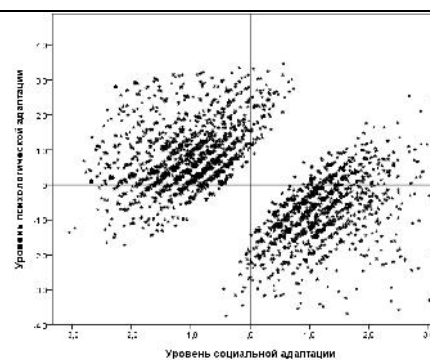
**Таблица сопряженности двух кластеризаций методом k-средних значений факторов с установкой кластерных центров и 12 первичных переменных,  $N=9664$**

	Кластеризации					Итого
	без установки начальных центров кластеров 1-5	по 12 первичным переменным				
	1	2	3	4	5	
1	1107	18	1	600	363	2089
2	1	1237	3	0	14	1255
3	14	89	624	48	177	952
4	226	0	30	1077	150	1483
5	446	1382	411	0	1646	3885
Итого	1794	2726	1069	1725	2350	9664

Для выбора пути типологизации объектов исследования будет полезным прием визуализации данных. Построим диаграммы рассеяния для двух кластеризаций методом k-средних, проведенных для двух факторов и для 12 первичных переменных.



**Рис. 4. Расположение респондентов второго и четвертого кластеров в пространстве факторов (метод k-средних с установкой начальных центров по факторам)**



**Рис. 5. Расположение респондентов второго и четвертого кластеров в пространстве факторов (метод k-средних по 12 первичным переменным)**

На рис. 4 и 5 показано расположение респондентов в пространстве двухфакторной модели. Можно видеть, что более удобной для интерпретации является типологизация, выполненная при помощи кластерного анализа методом k-средних с установкой кластерных центров.

Другой алгоритм верификации выделенных кластеров с позиции их интерпретации и однородности предоставляют таблицы сопряженности. Рассмотрим позицию респондентов второго и четвертого кластеров, выделенных методом k-средних с установкой начальных кластерных центров, по отношению к первым трем первичным переменным, образующим первый фактор «уровень социальной адаптации».

Респонденты второго и четвертого кластеров занимают по отношению к предложенным для оценки утверждениям прямо противоположную позицию (табл. 5). Респонденты, образующие второй кластер (комфортные), высказывают несогласие, тогда как респонденты, отнесенные к четвертому кластеру (дискомфортные), заявляют о своем согласии.

Распределение ответов респондентов из второго и четвертого кластеров на первичные вопросы анкеты, образующие второй фактор (уровень психологической адаптации), также будут диаметрально противоположными. Верифицируя полученные результаты, мы вправе утвердительно ответить на вопрос: «Имеют ли полученные в ходе анализа кластеры какой-либо содержательный смысл?». Следовательно, цели, которые мы ставили перед собой, применяя описанные выше методы, достигнуты.

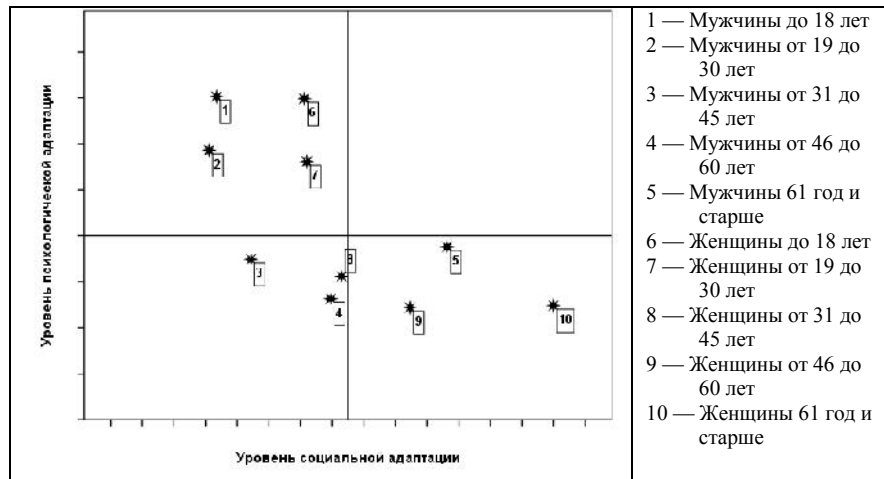
Таблица 5

**Распределение респондентов, отнесенных ко второму и четвертому кластерам (метод k-средних с установкой начальных кластерных центров), %**

Суждения	Комфортные: 2 кластер		Дискомфортные: 4 кластер	
	Варианты ответов			
	совсем не согласны	не со- гласны	соглас- ны	полностью согласны
5. Я часто чувствую себя беспомощным перед проблемами, возникающими в моей жизни	39,9	54,1	74,9	9,4
3. Я мало могу влиять на то, что со мной происходит	45,7	48,1	69,3	5,7
1. Я не могу справиться со своими проблемами	49,0	47,3	57,8	6,7

***Соединение пространств***

Дальнейший путь социологического анализа данных предполагает соединение пространства социально-психологической адаптации, полученного в процессе факторизации первичных переменных, с социально-демографическими характеристиками респондентов. Ограничим социально-демографические характеристики респондентов выделением десяти половозрастных групп. Понятно, что при проведении исследования аналитик не ограничивается двумя параметрами, и групп может быть достаточно много. Например, пол (2 группы), возраст (4 группы), образование (4 группы), место жительства (4 группы), занятость (5 групп):  $2 \times 4 \times 4 \times 4 \times 5 = 640$ . Дробность деления на группы определяется исследовательскими задачами и объемом выборки.



**Рис. 6. Топология половозрастных групп респондентов в пространстве социально-психологической адаптации.**

Приступая к этому этапу работы, исследователю надо иметь ясное понимание типов респондентов и, соответственно, модели рассматриваемого социального (в нашем случае — социально-психологического) пространства. На рисунке 6 представлено расположение половозрастных групп в пространстве социально-психологической адаптации. Во втором квадранте сосредоточены группы молодых мужчин и женщин до 18 лет. Именно этой группе свойственна самоидентификация, которую вполне правомерно выразить фразой «я хороший». В четвертом квадранте расположены группа мужчин старше 60 лет и две группы женщин: 46–60 лет и от 61 года. Этим респондентам не свойственно восприятие себя и своего социального окружения «сквозь розовые очки». Мужчины и женщины среднего возраста расположились в третьем квадранте. Позицию этой части респондентов можем интерпретировать как наличие определенной уверенности в происходящем и заниженной самооценки. В первом квадранте не представлено ни одной половозрастной группы, таким образом, правомерно сделать вывод, что пол и возраст не являются значимыми социально-демографическими характеристиками, связанными с этим типом социально-психологической адаптации — неуверенность в социальном окружении и удовлетворенность собой.

Таким образом, поиск типов респондентов, базирующийся на методах факторного и кластерного анализов, вполне правомерен. Но, напоминая о необходимости строгого соблюдения установленных правил их применения. Прежде всего, социологу-аналитику следует обратить внимание на: соотношение количества факторизуемых переменных и количество случаев (case); средние значения и стандартные отклонения переменных; таблицу «обшности» в процедуре факторного

анализа; меру КМО (адекватность применения факторного анализа); показатель доли объясненной дисперсии и подбор оптимального количества кластеров. В противном случае полученные результаты способны направить исследователя по ложному пути.

#### ЛИТЕРАТУРА

1. *Благуш П.* Факторный анализ с обобщениями. М.: Финансы и статистика. 1989.
2. *Бурдые П.* Социология политики. М.: Socio-Logos, 1993.
3. *Бююль А.* SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: [Версия 10]: [Пер. с нем.] / А. Бююль, П. Цёфель; Под ред. В.Е. Момота. М. [и др.]: DiaSoft(DS), 2002.
4. *Галицкая Е.Г.* Кластеры на факторах: как избежать распространенных ошибок? / Е.К. Галицкая и Е.Б. Галицкий // Социология: методология, методы, математическое моделирование. 2006. № 22. С. 145–161.
5. *Крыштановский А.О.* «Кластеры на факторах» — об одном распространенном заблуждении // Социология: методология, методы, математическое моделирование. 2005. № 21. С. 172–187.
6. *Малхорта Н.К.* Маркетинговые исследования. Практическое руководство. 3-е изд.: Пер. с англ. М.: Издательский дом «Вильямс», 2002.
7. *Татарова Г.Г.* Методология анализа данных в социологии: Учеб. для вузов / Г.Г. Татарова. 2-е изд., испр. М.: NOTA BENE, 1999.
8. *Черныш М.Ф.* Опыт применения кластерного анализа // Социология: методология, методы, математическое моделирование. 2000. № 12. С. 129–141.