

## Классификация стран по основным показателям развития алмазного рынка

© 2012 С.Ю. Айвазов

Московский государственный университет экономики,  
статистики и информатики (МЭСИ)

E-mail: ecsn@sciex.ru

В статье проведен кластерный анализ стран, играющих ведущую роль в международной торговле алмазами. В результате выделены три группы государств, формирующих в настоящее время международный алмазный рынок, определены их особенности и перспективы дальнейшего развития. Определены позиции России на мировом алмазном рынке.

*Ключевые слова:* рынок алмазов, кластерный анализ, кластер, добыча алмазов, импорт и экспорт алмазов.

При исследовании развития мирового алмазного рынка важное место отводится многомерной группировке стран, участвующих в международной торговле алмазами. Эта задача решается методами кластерного анализа и позволяет выделить группы стран со схожими параметрами развития алмазной отрасли. Кластерный анализ предоставляет возможность анализировать достаточно большой объем информации и значительно сокращать массивы данных, делать их компактными и наглядными<sup>1</sup>.

Задача кластеризации заключается в разбиении исследуемого множества объектов на однородные группы объектов, называемых кластерами. С целью исследования сложившейся в настоящее время структуры мирового алмазного рынка произведем кластеризацию стран по основным показателям его развития.

Существующая в настоящее время обобщающая статистика по добыче и торговле алмазами на международном уровне формируется на основе отчетов Кимберлийского процесса, которые находятся в открытом доступе на сайте организации<sup>2</sup>. Информационная база Кимберлийского процесса включает ежеквартальные, полугодовые и годовые отчеты. В них содержатся данные об объемах добычи, импорта и экспорта алмазов в натуральном и стоимостном выражении (карат и долларов США), а также о стоимости одного карата в долларах. Помимо этого публикуются данные сертификатов Кимберлийского процесса, выданных для ввоза или вывоза алмазов, по отдельным странам или территориальным образованиям.

При статистическом исследовании формирования и развития мирового алмазного рынка целесообразно использовать, прежде всего, данные об объемах добычи, импорта и экспорта алмазов в натуральном выражении, так как стоимостные оценки показателей в значительной степени под-

вержены влиянию инфляции<sup>3</sup>. Исходные данные для кластеризации – это совокупность показателей за 2010 г. по 32 странам, играющим наиболее заметную роль в международной торговле алмазами.

Кластеризацию будем проводить с использованием пакета STATISTICA, который предоставляет для этого значительные вычислительные возможности.

Непосредственно перед началом кластеризации исходные статистические данные стандартизируются (вычитается среднее и производится деление на среднеквадратическое отклонение). Полученные в результате переменные имеют нулевое среднее значение и единичную дисперсию<sup>4</sup>.

В современной статистической практике при решении задач кластеризации объектов активно применяются неиерархические методы, которые показывают достаточно высокую устойчивость по отношению к выбросам и неправильному выбору метрики. При этом также достигается существенная гибкость итогов кластеризации.

Самым востребованным в ряду неиерархических методов является метод  $k$ -средних. В этом случае объект относится к тому кластеру, расстояние до которого минимально. Предпочтение данного метода обусловлено простотой и быстротой его осуществления, прозрачностью алгоритма при достаточно высоком качестве результата.

В процессе реализации метода  $k$ -средних выдвигается гипотеза относительно предполагаемого количества кластеров, которая может быть сформирована на основе теоретических предположений, результатов предшествующих исследований, суждений экспертов.

На практике кластеризацию начинают, как правило, с разбиения исследуемой статистической совокупности на две группы. В дальнейшем количество кластеров постепенно увеличивают.

Производя последовательное разбиение объектов исследуемой совокупности на различное количество кластеров, можно дать оценку и сравнить качество получаемых решений.

Объективную количественную характеристику качества многомерной кластеризации объектов можно получить на основе дисперсионного анализа, по итогам которого определяются значения межгрупповых и внутригрупповых дисперсий признаков, значение  $F$ -статистики, а также уровень значимости.

Чем больше величина межгрупповой дисперсии признака и меньше величина внутригрупповой дисперсии, тем лучше признак характеризует принадлежность объектов к определенному кластеру и, следовательно, тем качественнее кластеризация<sup>5</sup>.

Наилучшей кластеризации также соответствуют большие значения  $F$ -критерия и меньшие значения уровня значимости. Признаки с большим уровнем значимости (обычно больше 0,05) часто исключаются из процедуры кластеризации.

Результаты дисперсионного анализа в случае разбиения исследуемой совокупности государств, участвующих в международной торговле алмазами, на два кластера оказались неудовлетворитель-

ными. Величины межгрупповых дисперсий при такой кластеризации неоднократно превышают значения внутригрупповых дисперсий исследуемых признаков.

Уровень значимости во всех случаях существенно ниже величины 0,05 и свидетельствует, что вклад всех признаков в процесс многомерной кластеризации является существенным.

Результаты дисперсионного анализа, представленные в табл. 2, отчетливо показывают, что дальнейшее увеличение числа кластеров не позволяет добиться более качественной кластеризации стран, участвующих в международной торговле алмазами. Полученные величины изменяются разнонаправленно, не обнаруживая явного улучшения качества многомерной кластеризации изучаемых объектов.

Все вышесказанное позволяет сделать вывод, что многомерную кластеризацию, при которой происходит дробление исследуемой совокупности государств на три кластера, в целом можно считать оптимальной.

Первый кластер, являющийся самым малочисленным, содержит только Европейский союз и Индию.

Второй кластер включает в себя Анголу, Австралию, Ботсвану, Канаду, Демократическую

Таблица 1. Дисперсионный анализ (3 кластера)

| Показатель      | Межгрупповая дисперсия | Внутригрупповая дисперсия | $F$ -статистика | Уровень значимости |
|-----------------|------------------------|---------------------------|-----------------|--------------------|
| Добыча алмазов  | 21,47815               | 9,52185                   | 32,7072         | 0,000000           |
| Импорт алмазов  | 29,06734               | 1,93266                   | 218,0809        | 0,000000           |
| Экспорт алмазов | 17,73945               | 13,26055                  | 19,3975         | 0,000004           |

Республику Конго, Россию, Южно-Африканскую Республику и Зимбабве.

Третий кластер является самым многочисленным по своему составу и охватывает остальные 22 страны, участвующие в международной торговле алмазами.

Для выделенных кластеров рассчитаем Евклидово расстояние между центрами кластеров (табл. 3). Над диагональю таблицы приведены квадраты этих расстояний.

Данные табл. 3 свидетельствуют о том, что максимальное расстояние отмечается между первым и вторым кластерами, минимальное - между вторым и третьим кластерами.

Итоги дисперсионного анализа, приведенные в табл. 1, позволяют сделать заключение, что разбиение изучаемой совокупности стран на

Республику Конго, Россию, Южно-Африканскую Республику и Зимбабве.

Третий кластер является самым многочисленным по своему составу и охватывает остальные 22 страны, участвующие в международной торговле алмазами.

Таблица 2. Дисперсионный анализ (4 кластера)

| Показатель      | Межгрупповая дисперсия | Внутригрупповая дисперсия | $F$ -статистика | Уровень значимости |
|-----------------|------------------------|---------------------------|-----------------|--------------------|
| Добыча алмазов  | 28,89769               | 2,10231                   | 128,2928        | 0,000000           |
| Импорт алмазов  | 29,06734               | 1,93266                   | 140,3739        | 0,000000           |
| Экспорт алмазов | 18,52892               | 12,47108                  | 13,8670         | 0,000010           |

Таблица 3. Евклидово расстояние между кластерами

|           | Кластер 1 | Кластер 2 | Кластер 3 |
|-----------|-----------|-----------|-----------|
| Кластер 1 | 0,000000  | 9,091042  | 8,282887  |
| Кластер 2 | 3,015136  | 0,000000  | 1,241301  |
| Кластер 3 | 2,878000  | 1,114137  | 0,000000  |

Таблица 4. Средние значения нормированных показателей по кластерам

| Показатель      | Кластер 1 | Кластер 2 | Кластер 3 |
|-----------------|-----------|-----------|-----------|
| Добыча алмазов  | -0,491508 | 1,418981  | -0,471311 |
| Импорт алмазов  | 3,686060  | -0,332253 | -0,214277 |
| Экспорт алмазов | 2,817660  | 0,083375  | -0,286469 |

В целях более полной интерпретации итогов кластерного анализа для каждого выделенного кластера определим средние значения нормированных показателей (усреднение производится внутри кластера).

Данные табл. 4 позволяют сделать вывод, что по средним значениям показателей импорта и экспорта алмазов Европейский союз и Индия, образующие первый кластер, являются безусловными лидерами мирового алмазного рынка. Вместе с тем по добыче алмазов они явно уступают странам, входящим во второй кластер.

Рисунок 1 позволяет просмотреть средние значения для каждого кластера на линейном графике.

В то же время исчисление характеристик по нормированным величинам затрудняет интерпре-

тацию результатов кластерного анализа. Поэтому определим средние значения для каждого кластера по исходным данным.

Анализ рис. 1 и табл. 5 позволяет выявить отличительные особенности каждого кластера.

Первый кластер составляют Европейский союз и Индия. Они являются безусловными лидерами по объемам импорта и экспорта алмазов.

Можно с уверенностью утверждать, что они играют доминирующую роль на мировом алмазном рынке. Вместе с тем доля этого кластера в мировой добыче алмазов минимальна, а в Европейском союзе добыча отсутствует совсем.

Второй кластер образуют Ангола, Австралия, Ботсвана, Канада, Демократическая Республика Конго, Россия, Южно-Африканская Республика

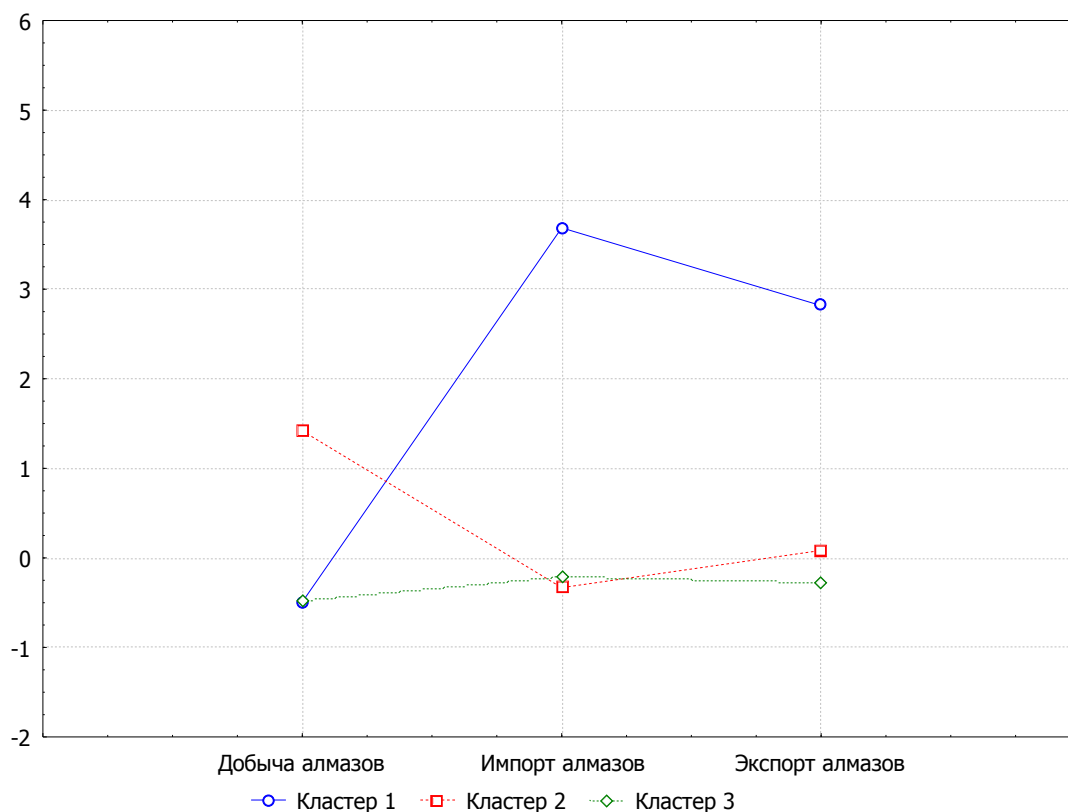


Рис. 1. График средних значений нормированных показателей по кластерам

Таблица 5. Средние значения показателей по кластерам

| Показатель             | Кластер 1      | Кластер 2     | Кластер 3    |
|------------------------|----------------|---------------|--------------|
| Добыча алмазов, карат  | 9042,20        | 15 560 168,09 | 173 447,15   |
| Импорт алмазов, карат  | 153 491 681,54 | 237 797,80    | 4 737 245,93 |
| Экспорт алмазов, карат | 94 329 257,88  | 15 319 273,48 | 4 632 254,82 |

и Зимбабве. На долю этих стран в настоящее время приходится более 90,0 % мировой добычи алмазов. По итогам 2010 г. Россия уступила лидерство по стоимости добываемых алмазов Ботсване, однако сохранила первенство по объему их добычи.

Третий кластер формируют 22 страны: Армения, Беларусь, Бразилия, Центрально-Африканская Республика, Китай, Республика Конго, Гана, Гвинея, Гайана, Израиль, Япония, Ливан, Лесото, Либерия, Намибия, Сьерра-Леоне, Шри-Ланка, Швейцария, Танзания, Таиланд, Объединенные Арабские Эмираты, Соединенные Штаты Америки.

Являясь самым многочисленным кластером по количеству стран, по показателям участия в международной торговле алмазами данный кластер не выделяется.

Далее проведем кластеризацию стран, формирующих мировой алмазный рынок, по основным показателям его развития на основе агломеративного иерархического метода.

Предварительное исследование позволило сделать заключение, что наиболее отчетливые кластеры образуются при использовании в качестве метрики так называемого манхэттенского расстояния, или “расстояния городских кварталов”. При этом расстояние между кластерами целесообразно определять методом полной связи.

Рисунок 2 позволяет сделать вывод, что в исследуемой совокупности стран, участвующих в международной торговле алмазами, явно выделяются три кластера.

Необходимо отметить, что иерархический и неиерархический методы разбиения на кластеры исследуемой совокупности государств, функционирующих на международном рынке алмазов, в данном случае приводят к аналогичным результатам.

Таким образом, проведенный кластерный анализ позволил отчетливо выделить три группы государств, формирующих в настоящее время международный алмазный рынок.

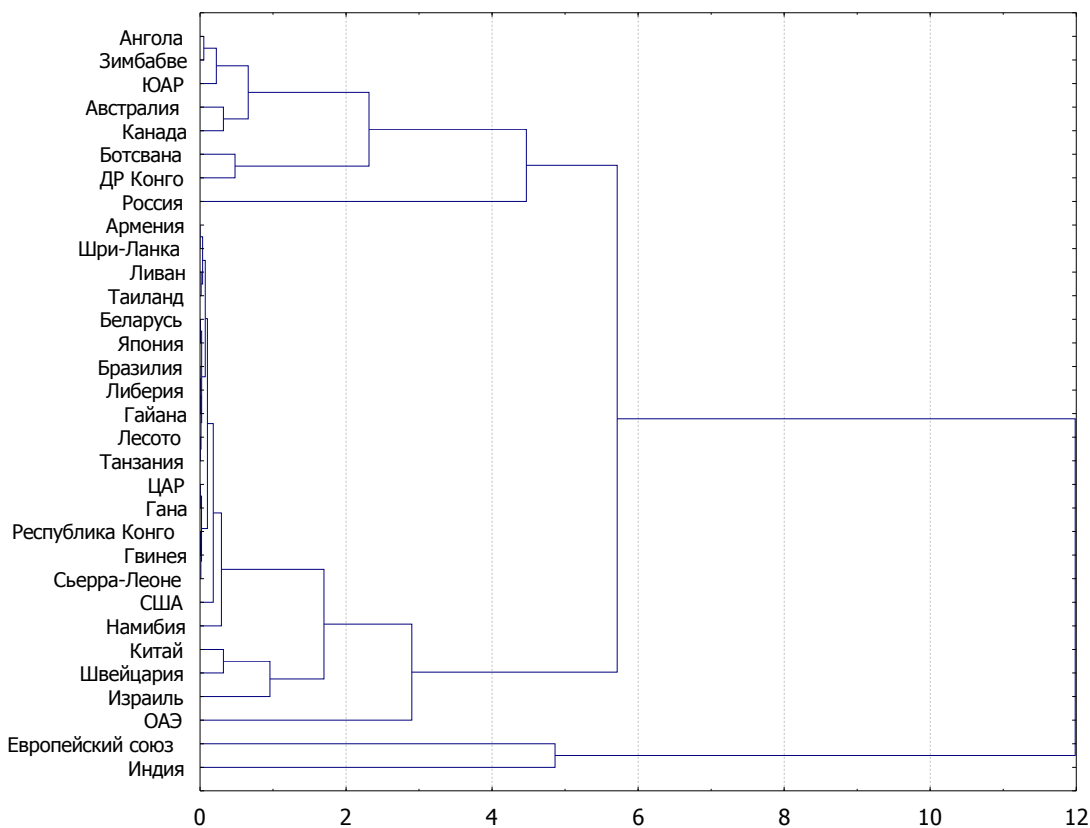


Рис. 2. Дендрограмма классификации стран

В первый кластер входят участники рынка, являющиеся безусловными лидерами по объемам импорта и экспорта алмазов. Они являются центрами международной торговли алмазами.

Второй кластер образует группа алмазодобывающих стран, куда входит и Россия. Можно утверждать, что именно они формируют предложение на мировом алмазном рынке, так как на долю этих стран в настоящее время приходится более 90,0 % добычи алмазов в мире. При этом Россия и Ботсвана на протяжении многих лет сохраняют первенство по объемам добычи алмазов.

Третий кластер составляют страны, которые не выделяются по показателям участия в международной торговле алмазами. Тем не менее данная группа является самой многочисленной по количеству образующих ее государств и в пер-

спективе может оказывать значительное влияние на процесс формирования спроса и предложения на мировом алмазном рынке.

<sup>1</sup> См.: Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян [и др.]. СПб., 2004; Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы : учебник. М., 2003.

<sup>2</sup> URL: [https://kimberleyprocessstatistics.org/public\\_statistics](https://kimberleyprocessstatistics.org/public_statistics).

<sup>3</sup> Айвазов С.Ю., Минашкин В.Г. Эволюция и структура мирового рынка алмазов // Вестн. Ростов. гос. экон. ун-та (РИНХ). 2012. Июнь (□ 2 (38)).

<sup>4</sup> Боровиков В.П. Программа STATISTICA для студентов и инженеров. 2-е изд. М., 2001. С. 186.

<sup>5</sup> Халафян А.А. STATISTICA 6. Статистический анализ данных : учебник. 3-е изд. М., 2007. С. 250.

*Поступила в редакцию 04.04.2012 г.*