



О. А. Смирнов

**ПРИМЕНЕНИЕ МОДЕЛЕЙ СОКРАЩЕНИЯ РАЗМЕРНОСТИ  
ПРИ ОЦЕНКЕ ОДНОРОДНОСТИ ОБЪЕКТОВ:  
АНАЛИЗ АЭРОПОРТОВОЙ СЕТИ**

18

*Рассматриваются особенности применения моделей сокращения размерности при анализе явлений социально-экономических систем на примере выбора направления социально-экономического развития. В частности показано, что применение алгоритмов сокращения размерности позволяет снизить вероятность ошибки при оценке апостериорных вероятностей принадлежности к классу однородных объектов, полученных в результате решения задачи дискриминантного анализа.*

*An application of dimension reduction methods in the analysis of the phenomena of socio-economic systems by analyzing the choice of direction of development of socio-economic development is considered. In particular it is shown that the use of algorithms dimension reduction methods the probability of error when estimating posterior probabilities of belonging to the class of homogeneous objects, obtained in the result of solving the problem of discriminant analysis.*

**Ключевые слова:** модели сокращения размерности, дискриминантный анализ, апостериорная вероятность, data mining.

**Key words:** dimension reduction methods, discriminant analysis, posterior probability, data mining.

Фактически все социально-экономические явления, анализируемые с помощью алгоритмов *Data mining*, описываются посредством значительного количества переменных. При этом следует учитывать, что применение ряда инструментов многомерного статистического анализа невозможно в случае, если в анализируемом массиве данных присутствует мультиколлинеарность. Исключение коррелируемых переменных не всегда подходящий метод, так как оставшиеся переменные могут содержать случайные изменения в некоторых наблюдениях, которые могут существенным образом повлиять на результат моделирования в целом. Особенно остро данная проблема стоит, когда данные представляют собой временные ряды, а задачей исследования служит выявление однородной динамики развития.

Наиболее распространенный метод снижения размерности – метод принципиальных компонент (*principal components analysis*) [1]. Его алгоритм основан на поиске в массиве данных корреляционной структуры и использовании наименьшего набора в виде линейной комбинации данных переменных. Таким образом, при использовании данного



алгоритма исходные значения переменных заменяются на их линейные комбинации. Показателем достоверности применения метода стал уровень выявленной дисперсии.

Рассмотрим применение метода принципиальных компонентов на примере выявления существующих и потенциальных хабовых аэропортов в действующей маршрутной сети гражданской авиации Российской Федерации. Согласно сложившейся терминологии *аэропорт-хаб* — это узловой аэропорт, крупный пересадочный и перегрузочный транспортный узел с необходимым набором сервисов, имеющий географическое положение, обеспечивающее концентрацию пассажиропотоков и их распределение по направлениям другого типа.

Модель хабовой организации авиаперевозок концентрирует трафик пассажиропотока в одном аэропорту и связывает его с региональными аэропортами и транспортными узлами других видов транспорта.

В современных исследованиях показано, что темп прироста пассажиропотока в российских хабовых аэропортах существенно выше аналогичных показателей в других аэропортах [2]. Однако применение показателей роста пассажиропотока в динамике как переменных признакового пространства задачи оценки однородности объектов с помощью любого метода многомерного статистического анализа может снизить адекватность результатов такой оценки, так как в данном случае будет значимо влияние случайных изменения, не связанных с основной тенденцией, например проведение в городе массовых спортивных мероприятий (Универсиада в Казани, чемпионат мира по биатлону в Ханты-Мансийске, Зимние олимпийские игры в Сочи и т. д.). Поэтому для применения данных приростов необходимо применить инструменты сокращения размерности.

По состоянию на 2013 г. из всех городов, из которых осуществляются магистральные пассажирские перевозки, согласно отчетам Министерства транспорта можно выделить группу городов, которые являются потенциальными хабами. Это Владивосток, Екатеринбург, Иркутск, Калининград, Красноярск, Нижневартовск, Новосибирск, Санкт-Петербург, Сургут, Тюмень, Хабаровск. Данные региональные аэропорты имеют магистральные маршруты, связывающие указанные центры с московским авиаузлом [3]. Применим алгоритм принципиальных компонентов, реализованный в пакете Statistica, для выявления признакового пространства для дальнейшего их применения при определении потенциальных хабовых аэропортов. В качестве эталонного были выбраны аэропорты Уфы и Казани, так как их развитие в качестве хаба было осуществлено целенаправленно с помощью специальных инструментов госрегулирования.

Так, на рисунке представлен результат сжатия шести годовых значений приростов объема пассажироперевозок за 2008–2013 гг. до двух факторов с общим уровнем выявленной дисперсии более 90% (фактор 1 — 71,89%, фактор 2 — 19,03%) в виде проекции на единичную окружность, где чем ближе значение показателя к единичной окружности, тем больше характеристик учтено в заменяющих его значениях факторах.

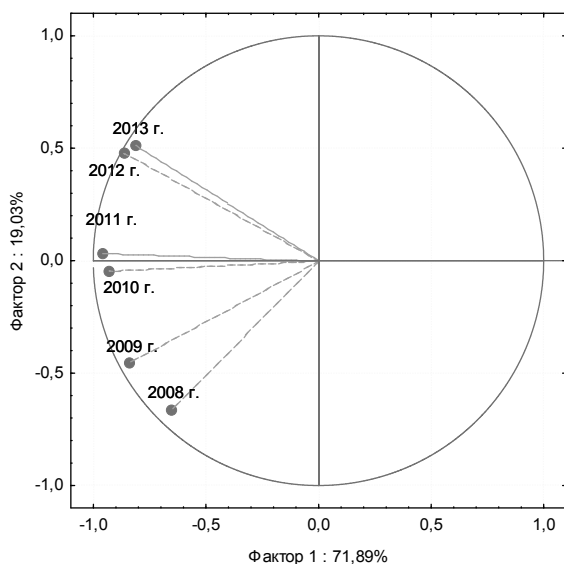


Рис. Анализ принципиальных компонент:  
проекция единичной окружности

В результате применения алгоритма дискриминантного анализа на основании значений двух факторов, являющихся линейной комбинацией шести исходных переменных на основе данных 63 федеральных аэропортов сделать вывод о том, что с апостериорной вероятностью более 98% к потенциальным хабам можно отнести 35 аэропортов, при этом в данный перечень не входит часть аэропортов, рассматриваемых как потенциальные региональные узловые аэропорты (например, Храброво, Калининград), при этом до сокращения признакового пространства таких объектов было пять.

Решение задачи дискриминантного анализа на основании сглаженных рядов динамики на три значения отнесло 58 объектов к хабовым с вероятностью более 99%.

Таким образом, применение инструментов сокращения размерности позволяет получать более объективные результаты при оценке однородности объектов, а также получить более достоверные результаты в случае, если совокупность исследуемых переменных позволяет однозначно отнести объекты наблюдения к определенному типу.

### Список литературы

1. Смирнов О.А. Мониторинг цен на авиационные билеты: разработка информационной системы // Крымский научный вестник. 2015. № 1 (1). С. 67–79.
2. Смирнов О.А. Применение хабовой модели организации авиационных перевозок в российской федерации: возможности и ограничения применения // Экономика: вчера, сегодня, завтра. 2012. № 5–6. С. 66–71.
3. Открытые данные Росавиации России. URL: <http://www.favt.ru/opendata> (дата обращения: 05.02.2015).



4. *Larose D.* Discovering knowledge in data: an introduction to Data Mining. Wiley, Hoboken, NJ, 2005.

5. *Robinson J.P., Shaver P.R., Wrightsman L.S.* Criteria for scale selection and evaluation // Measures of Personality and Social Psychological Attitudes, Academic Press, San Diego, CA, 1991.

6. *Richard A. Johnson and Dean Wichern,* Applied Multivariate Statistical Analysis, Prentice Hall, Upper Saddle River, NJ, 1998.

#### **Об авторе**

Олег Аркадьевич Смирнов – канд. физ.-мат. наук, заместитель директора Межрегионального института развития территорий, Ялта.

E-mail: smirnovoleg1952@mail.ru

21

#### **About the author**

Dr Oleg Smirnov – deputy director of Inter-regional Institute for Spatial Development, Yalta.

E-mail: smirnovoleg1952@mail.ru