

УДК 338.001.36

МЕТОД ГЛАВНЫХ ФАКТОРОВ КАК ИНСТРУМЕНТ ФОРМИРОВАНИЯ СТРУКТУРНО-ОДНОРОДНЫХ ОБЪЕКТОВ

И. С. Максимова

В работе изучены показатели экономической и транспортной сфер Приволжского Федерального округа. В соответствии с алгоритмом Г. Хотгелинга было выделено два фактора развития дорожно-транспортной сферы, с целью улучшения их интерпретируемости была проведена процедура «варимакс» вращения. Предложена методика кластеризации субъектов, основанная на их расположении в пространстве латентных факторов. Методика включает политетический и монотетический подходы. Получены средние значения значимых показателей в кластерах, и проведено их сравнение со средними значениями в регионе. Для каждого подхода рассчитан критерий качества кластеризации, и, согласно полученным результатам, выбран оптимальный подход.

Ключевые слова: факторный анализ, латентные факторы, кластеризация, ПФО, дорожно-транспортная сфера.

Транспортная система России представляет собой совокупность транспортных средств инфраструктуры и управления, функционирующих на территории Российской Федерации. Целью функционирования системы является удовлетворение транспортных нужд человека, она включает в себя объекты транспортировки, средства транспортировки и окружающую среду. В настоящее время удовлетворение спроса на высококачественные транспортно-логистические услуги осуществляется с помощью преобразования характеристик системы. В частности, рассматриваются географическая и технологическая доступность транспорта, его стоимостные характеристики. Доступность транспортных услуг определяет эффективность развития производства, бизнеса и социальной сферы. Роль транспорта как системообразующего фактора постоянно растет, и в этой связи проводится множество исследований с целью выявления направлений его модернизации. Анализ объемов торговли и производства товаров позволяет определить их роль в общем количестве грузоперевозок. Цены на транспортные услуги и объемы инвестированных средств в управление и обслуживание транспорта также определяют функционирование системы.

Сфера транспорта, как отдельного сектора экономики, широко изучается многими авторами. Выделяются следующие факторы, определяющие специфику этого сектора в современном мире:

1) система управления развитием экономики РФ включает в себя деятельность дорожно-транспортного комплекса;

2) деятельность транспорта как сферы экономики по базовым направлениям обеспечивает реализацию стратегии развития экономики РФ [1].

Самым распространенным видом транспорта в РФ является железнодорожный транспорт. В статьях также проводится изучение вопроса регулирования данного вида транспорта. Главным документом, регламентирующим деятельность управления железнодорожным транспортом, является Устав железнодорожного транспорта.

При сравнении железнодорожного и автомобильного видов транспорта выделяются факторы конкурентоспособности, а также проводится их качественная и количественная оценка. На основе анализа особенностей каждого из рассматриваемых видов транспорта были выделены следующие конкурентные преимущества:

© Максимова И. С., 2020.

Максимова Ирина Сергеевна (irina.maximova@yandex.ru),
студент IV курса Института экономики и управления Самарского университета,
443086, Россия, г. Самара, Московское шоссе, 34.

– автомобильный транспорт отличаются маневренностью, высокая проходимость, оперативность, низкая стоимость инфраструктуры;

– железнодорожный транспорт независим от погодных условий, является надёжным средством доставки, при низких энергозатратах [2].

По территории Приволжского федерального округа (ПФО) проходят железнодорожные линии, связывающие Центральный и Северо-Западный федеральные округа с Уральским, Сибирским, Дальневосточным федеральными округами, а также с Республикой Казахстан. Вторым по популярности является автомобильный транспорт. Однако, уровень развития железнодорожного и автомобильного транспорта у субъектов ПФО отличаются, и в условиях комплексного развития субъектов это отражается на их экономическом состоянии.

Таким образом, проблема определения зависимости транспортных и экономических показатели в субъектах регионов представлена недостаточно полно. Поэтому важно и актуально исследовать состояние и развитие данных показателей.

Средствами многомерных методов решаются разнообразные задачи [3]. Однако, специфической особенностью многомерных данных является отсутствие возможности их графического представления. Поэтому далее в работе визуализация и кластеризация субъектов проводилась с помощью факторного анализа: на основе их расположения в пространстве латентных факторов.

Классическая модель факторного анализа определяет следующее матричное уравнение, для определения коэффициентов при общих факторах F_r :

$$Z_j = AF + a_j D_j, \quad (1)$$

где D_j – характерный фактор.

Решение уравнения при условии максимизации сумм:

$\sum_{j=1} a_{j1}^2 = \lambda_1$ – первый максимум в части описанной дисперсии ($D(Z_j)$);

$\sum_{j=1} a_{j2}^2 = \lambda_2$ – второй максимум, относительно оставшийся после λ_1 дисперсии и т. д.

Далее работа сводится к определению

собственных значений и собственных векторов симметричной матрицы R из равенства $(R - \lambda E) U = 0$. В соответствии с методом главных компонент, имея известные значения λ_j и U_j , коэффициенты a_{jr} можно вычислить согласно формуле:

$$A = V * A^{\frac{1}{2}}, \quad (2)$$

где V – матрица нормированных векторов,

$$U_j: V_j = \frac{U_j}{|U_j|}.$$

Тогда общая формула вычисления a_{jr} имеет вид:

$$a_{jr} = \frac{u_{jr} \sqrt{\lambda_r}}{\sqrt{u_{1r}^2 + u_{2r}^2 + \dots + u_{nr}^2}} \quad (3)$$

В реальной практике применяются различные приёмы, способы нахождения параметров модели главных факторов λ , U_j , a_{jr} . В частности, в работе применен метод, разработанный Г. Хоттелином. Данный метод предусматривает итеративное решение. Первая итерация включает многократное возведение матрицы парных корреляций в квадрат, с целью получения сходимости по первому собственному значению λ_1 . После этого проводится вычисление значений собственного вектора U_1 и факторных нагрузок a_{1r} . В завершении этапа находят произведение векторов $A_1 * \hat{A}_1 = R^+$, где R^+ – воспроизведенная матрица корреляций. При этом остаточная матрица корреляций рассчитывается как $R - R^+$. Если при проверке разность $R - R^+$ существенна, переходят ко второму этапу и описанная выше итерация повторяется, но относительно второго собственного числа λ_2 , вычисляемого по данным матрицы остаточных коэффициентов корреляции. Итерации повторяют до тех пор, пока разность $R - R^+$ не станет достаточно малой и тогда алгоритм завершается [4]. С целью повышения определенности расположения факторной координатной системы в пространстве применяется процедура вращения общих факторов. Методика реализации ортогонального вращения предполагает получение произведения матрицы факторных нагрузок и матрицы поворота T . Матрица T определяет угол поворота пространства и ее размерность эквивалента числу общих факторов. Поворот пространства может производиться в как по часовой стрелке, так и против. При числе общих факторов равного двум, матрица T будет иметь размерность 4×4 и примет следующий вид:

$$T = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix},$$

Вращение по часовой стрелке

$$T = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}. \quad (4)$$

Вращение против часовой стрелки

При проведении процедуры факторного анализа матрица исходных данных имела следующий вид, представленный в таблице 1.

Исследуемые показатели:

– добыча полезных ископаемых (объём отгруженных товаров собственного производства, млн. руб.) (X_1);

– обрабатывающее производство (объём отгруженных товаров собственного производства в млн. руб.) (X_2);

– индексы тарифов на грузовые перевозки, % (X_3);

– инвестиции в основной капитал по

виду деятельности – машины, оборудование, транспортные средства, % от общего объёма инвестиций (X_4);

– торговля со странами СНГ (млн. долларов США):

– экспорт (X_5);

– импорт (X_6).

По данным таблицы была построена исходная матрица корреляции (табл. 2). Для дальнейшего осуществления операций в работе использовалась редуцированная матрица корреляций, полученная методом максимальной корреляции (табл. 3).

Редуцированная матрица корреляции возводилась в степени. Итерации, включали возведение редуцированной матрицы корреляции во вторую степень, на каждом шаге производился расчёт погрешности α , а также разность $d = R - R^+$.

Таблица 1

Массив исходных данных для проведения факторного анализа (объяснения в тексте)

Субъекты ПФО	X_1	X_2	X_3	X_4	X_5	X_6
Республика Башкортостан	233703,0	1082923	103,8	31,5	1415,8	254,6
Республика Марий Эл	656,0	152086	97,8	31,7	53,6	10,1
Республика Мордовия	77,3	167180	10,3	27,5	155,0	30,1
Республика Татарстан	547515,0	1596330	101,9	26,0	1510,9	607,3
Удмуртская Республика	191064,0	321066	102,4	50,4	130,4	453,1
Чувашская Республика	706,0	172565	100,1	32,9	141,2	56,1
Пермский край	294130,0	933960	132,0	31,1	697,0	161,6
Кировская область	1107,0	199822	101,4	41,0	179,7	19,9
Нижегородская область	1168,0	1237833	130,9	39,0	1151,3	371,8
Оренбургская область	389692,0	304238	100,0	39,8	1625,1	547,7
Пензенская область	890,0	195823	101,4	33,5	117,2	46,3
Самарская область	273744,0	969528	104,6	36,1	2983,2	390,1
Саратовская область	30492,0	352459	108,2	38,4	575,4	103,8
Ульяновская область	7096,0	256480	108,0	33,7	144,0	80,0

Таблица 2

Исходная матрица корреляции (объяснения в тексте)

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1,00	0,65	0,20	-0,14	0,65	0,82
X_2	0,65	1,00	0,39	-0,28	0,66	0,64
X_3	0,20	0,39	1,00	0,31	0,22	0,26
X_4	-0,14	-0,28	0,31	1,00	-0,07	0,23
X_5	0,65	0,66	0,22	-0,07	1,00	0,67
X_6	0,82	0,64	0,26	0,23	0,67	1,00

Таблица 3

Редуцированная матрица корреляции, полученная методом наибольшей корреляции (объяснения в тексте)

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	0,823	0,645	0,197	-0,139	0,653	0,823
X_2	0,645	0,645	0,388	-0,282	0,664	0,639
X_3	0,197	0,388	0,388	0,313	0,224	0,259
X_4	-0,139	-0,282	0,313	0,313	-0,065	0,226
X_5	0,653	0,664	0,224	-0,065	0,664	0,673
X_6	0,823	0,639	0,259	0,226	0,673	0,823

Таблица 4

Матрица компонент до и после вращения (объяснения в тексте)

	Перед вращением		После вращения	
	F1	F2	F1	F2
X_1	0,87	-0,14	0,83	0,32
X_2	0,80	-0,18	0,78	0,24
X_3	0,35	0,44	0,08	0,55
X_4	-0,03	0,65	-0,35	0,54
X_5	0,79	-0,08	0,72	0,33
X_6	0,88	0,22	0,65	0,63

Таким образом, всего было выполнено четыре цикла итераций и достигнута точность $d(5) \leq 0,045$. Далее в работе осуществлялся расчёт нагрузок первого главного фактора. Была получена воспроизведенная, а затем остаточная матрица корреляции. Аналогично остаточная матрица возводилась в степень и были вычислены нагрузки второго главного фактора. Таким образом были получены значения векторов главных факторов, представленных в таблице 4 (перед вращением).

В работе проводилось варимакс вращение пространства латентных факторов по часовой стрелке на 30 градусов. В таблице 4

представлена матрица факторного отображения до и после варимаксного вращения по двум выделенным латентным факторам.

Система линейных комбинаций латентных факторов примет вид:

$$\begin{cases} F1 = 0,83X_1 + 0,78X_2 + 0,72X_5 \\ F2 = 0,55X_3 + 0,54X_4 \end{cases}$$

Таким образом, первый латентный фактор представляет собой множество: X_1, X_2, X_5 , второй латентный фактор: X_3, X_4 . Первый фактор аккумулирует в себе компоненты добычи, переработки и реализации товаров

собственного производства, второй представляет собой внешнюю экономическую среду. При вращении показатель импорта имеет одинаковое влияние и в первом и во втором факторе. Координаты латентных факторов для субъектов ПФО в новом пространстве представлены в таблице 5.

Расположение субъектов ПФО в пространстве латентных факторов представлено на рисунке 1.

На основе полученного двумерного расположения субъектов ПФО в пространстве латентных факторов были выделены три кластера субъектов, представленных в таблице 6.

Данный подход к кластеризации – политетический. Он предполагает выделение кластеров на основе их расположения в пространстве двух латентных факторов.

Таблица 5

Координаты латентных факторов для субъектов ПФО (объяснения в тексте)

Субъекты	W1	W2
1	476413,11	-91886,95
2	55358,48	-12255,80
3	60675,53	-13510,84
4	774583,82	-139183,58
5	184615,48	-29278,25
6	62830,48	-13891,91
7	443739,13	-80981,08
8	72859,45	-16116,10
9	449547,23	-100080,05
10	249822,07	-31500,00
11	71320,36	-15785,14
12	450047,10	-83306,49
13	138807,37	-28983,10
14	95536,84	-20799,18

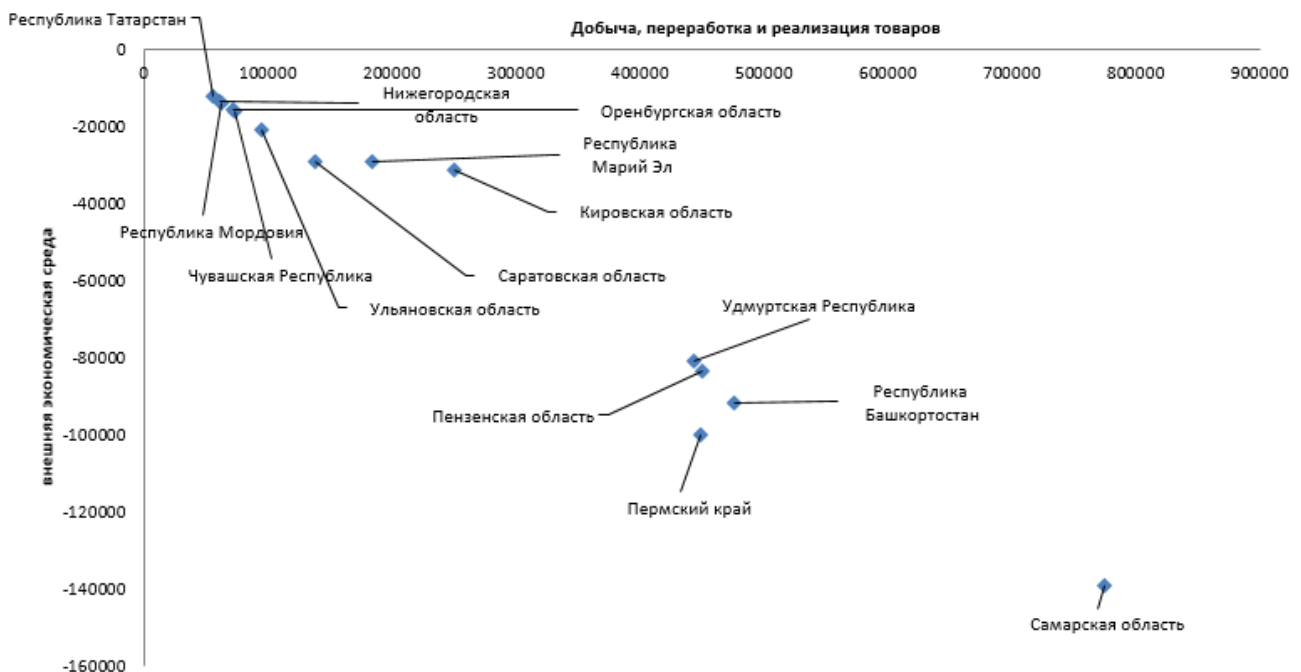


Рис. 1. Субъекты ПФО в пространстве латентных факторов

Таблица 6

Кластеры субъектов ПФО

1 кластер	2 кластер	3 кластер
Татарстан, Республика Мордовия, Нижегородская область, Оренбургская область, Кировская область, Республика Марий Эл, Саратовская область, Ульяновская область, Чувашская Республика	Удмуртская Республика, Пензенская область, Пермский Край, Республика Башкортостан	Самарская область

Таблица 7

Средние значения значимых показателей у субъектов первого кластера

Название показателя	Среднее значение в первом кластере	Среднее значение в ПФО
Добыча полезных ископаемых (объем отгруженных товаров собственного производства), млн. руб.	69086,7	140860,0214
Обрабатывающее производство (объем отгруженных товаров собственного производства), млн. руб.	235746,5556	567306,6429
Инвестиции в основной капитал по виду деятельности – машины, оборудование, транспортные средства, %	36,5	35,2

Во всех кластерах максимальная корреляция по первому фактору выделяется у показателей «добыча полезных ископаемых», и «обрабатывающее производство». В двумерном пространстве латентных факторов второй фактор превалирует на уровне инвестиций.

Показатели субъектов первого кластера находятся в диапазоне от 55300 до 249822 по шкале показателей добычи, переработки и реализации и в диапазоне от –12255 до –31499 по шкале «внешняя экономическая среда». На основе таблицы исходных данных для факторного анализа, был рассчитаны средние значения показателей по кластерам, а также получено значение критерия качества кластеризации: общая сумма дисперсий равна 49003701633. Сравнение средних значений значимых показателей у субъектов кластера со средними значениями в регионе представлено в таблице 7.

Среднее значение добычи полезных ископаемых (объем отгруженных товаров собственного производства) в кластере равно 69086,7 млн. руб., что в практически два раза меньше, чем среднее значение в регионе. Среднее значение обрабатывающего производства в регионе превышает значение в

кластере в 2,4 раза. Среднее значение инвестиций в основной капитал по виду деятельности – машины, оборудование, транспортные средства в кластере превышает значение по округу на 1,36 %.

Показатели субъектов второго и третьего кластеров располагаются значительно дальше по каждому из факторов. Для второго кластера характерны следующие средние значения значимых показателей первого фактора:

– добыча полезных ископаемых – 200686,25 млн. руб.;

– обрабатывающее производство – 1056061 млн. руб.

Среднее значение инвестиций внутри кластера равно 34,425 %. Таким образом, во втором кластере значение добычи полезных ископаемых больше среднего значения в регионе в 1,42 раза, обрабатывающего производства – в 1,8 раза. Однако значение инвестиций у данных субъектов ниже среднего значения по округу на 0,76 %.

При монотетическом подходе кластеризация субъектов производится по каждому отдельному фактору. По первому фактору были выделены следующие кластеры субъектов, представленные в таблице 8.

Первый кластер субъектов располагается в диапазоне от 55358 до 95536,7 по фактору добычи переработки и реализации товаров. Второй кластер субъектов имеет диапазон от 138807 до 249822. Третий и четвертый кластеры по шкале первого фактора расположены аналогично второму и третьему кластеру в политетическом подходе. Результат критерия качества кластеризации по первому фактору равен 50088852689,2. Субъекты второго, третьего и четвертого кластеров имеют больший объём добычи полезных ископаемых, чем в среднем по региону. Однако среднее значение данного показателя в первом кластере значительно меньше, чем в округе – в 80 раз. Наибольший объём отгруженных товаров собственного производства по добыче полезных ископаемых наблюдается в Самарской области.

Результат кластеризации субъектов на уровне второго фактора представлен в таблице 9. Диапазон значений первого кластера по фактору «внешняя экономическая среда» равен от –20799 до –12225. Самым значимым в данном факторе является показатель инвестиций, и его среднее значение в кластере равно 33,38%. Среднее значение индекса тарифов на грузоперевозки равно 86,5 % (декабрь к декабрю предыдущего года).

Второй кластер расположен в области от –31499 до –28983. Среднее значение инвестиций в факторе равно 42,87 %, показателя индекса тарифов на грузоперевозки – 103,53 %.

Третий кластер имеет диапазон от –83306 до –80981. Среднее значение инвестиций в основной капитал по виду деятельности – машины, оборудование, транспортные средства равно 33,6 %. Средняя индексация тарифов на грузоперевозки в кластере равна 118,3 %.

Область расположения четвертого кластера субъектов представляет диапазон значений от –100080 до –91886. Показатель инвестиций имеет среднее значение в кластере равное 35,25%, а показатель индекса тарифов на грузоперевозки – 117,35 %. Общая сумма дисперсий в данной кластеризации равна 42245509804.

Субъекты второго кластера: Саратовская область, Кировская область, Республика Марий Эл имеют наибольшее среднее значение инвестиций в основной капитал по виду деятельности – машины, оборудование, транспортные средства, превышающее среднее значение в регионе на 7,68 %.

Таблица 8

Кластеры субъектов ПФО в первом факторе

1 кластер	2 кластер	3 кластер	4 кластер
Татарстан, республика Мордовия, Нижегородская область, Оренбургская область, Ульяновская область, Чувашская Республика	Саратовская область, Кировская область, Республика Марий Эл	Удмуртская Республика, Пензенская область, Пермский Край, Республика Башкортостан	Самарская область

Таблица 9

Кластеры субъектов ПФО в втором факторе

1 кластер	2 кластер	3 кластер	4 кластер	5 кластер
Татарстан, Республика Мордовия, Нижегородская область, Оренбургская область, Ульяновская область, Чувашская Республика	Саратовская область, Кировская область, Республика Марий Эл	Удмуртская Республика, Пензенская область	Пермский Край, республика Башкортостан	Самарская область

Самарская область обладает самым низким значением данного показателя и отличается от среднего по округу на 9,18 %. Средняя индексация тарифов на грузоперевозки в регионе равно 100,2 %, наиболее низким средним значением данного показателя обладают субъекты первого кластера, наиболее высоким – субъекты третьего кластера – Удмуртская Республика, Пензенская область.

Таким образом, средствами факторного анализа данные визуализированы и классифицированы с использованием политетического подхода по двум латентным факторам, и монотетического – по каждому фактору в отдельности. Рассчитаны средние значения и критерий качества кластеризации, проведено сравнение средних значений ключевых показателей со средними показателями в Приволжском Федеральном округе. При обоих методах Самарская область выделялась в качестве отдельного кластера. Предложенная методика кластеризации предполагает исследователю самостоятельно выбрать подход к кластеризации. Согласно критерию качества кластеризации в качестве оптимального выбран монотетический подход. Используемая в работе методика кластериза-

ции на латентных факторах позволяет достаточно быстро сформировать структурно-однородные объекты, с использованием стандартного программного продукта Microsoft Excel.

Литература

1. Чередник М. А. Транспорт как сфера экономики. Государственное регулирование железнодорожного транспорта // Вектор экономики. 2018. № 4 (электронный журнал). URL: <http://www.vectoreconomy.ru/images/publications/2018/4/world-economy/Cherednik.pdf> (дата обращения: 10.06.2020).
2. Давыдова Д. В. Анализ и оценка конкурентных преимуществ железнодорожного транспорта во взаимодействии с другими видами транспорта // Новое слово в науке: стратегии развития: сб. матер. V Междунар. научно-практ. конф. 2018. С. 228–230.
3. Трусова А. Ю. Анализ показателей инновационного потенциала многомерными статистическими методами // Вестник Самарского университета. Серия Экономика и управление. 2018. Т. 9. № 4 С. 82–87.
4. Сошникова Л. А., Тамашевич В. Н. Многомерный статистический анализ в экономике. М.: ЮНИТИ-ДАНА, 1999. 201 с.

THE METHOD OF MAJOR FACTORS AS A TOOL FOR FORMING STRUCTURAL-HOMOGENEOUS OBJECTS

I. S. Maksimova

The paper studies the indicators of the economic and transport sectors of the Volga Federal District. Based on the indicators, factor analysis tools identified two main factors in the development of the road transport sector and obtained the location of the Volga Federal District subjects in the space of latent factors. Based on the location obtained, clustering of the constituent entities by two approaches was carried out. A method for clustering subjects based on their location in the space of latent factors is proposed. The technique includes polythetic and monothetic approaches. The average values of significant indicators in the clusters are obtained, and they are compared with the average values in the region. For each approach, a clustering quality criterion was calculated, and, according to the results obtained, the optimal approach was chosen.

Key words: factor analysis, latent factors, clustering, VFD, road transport sphere.

Статья поступила в редакцию 01.10.2020 г.

© Maksimova I. S., 2020.

Maksimova Irina Sergeevna (irina.maximova@yandex.ru), student IV course of the Institute of Economics and Management of the of the Samara University, 443086, Russia, Samara, Moskovskoye Shosse, 34.