



КЛАССИФИКАЦИЯ РЕГИОНОВ РФ ПО УРОВНЮ РАЗВИТИЯ СИСТЕМЫ ЗДРАВООХРАНЕНИЯ

Болдина Кристина Сергеевна

Магистрант специальности прикладной информатики в аналитической экономике
Института управления и региональной экономики,
Волгоградский государственный университет
kristi_b92@mail.ru, manag@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Аннотация. В данной статье проведен анализ развития системы здравоохранения в регионах РФ с использованием факторного и кластерного анализа. Задачей анализа являлось конструирование обобщенных характеристик, описывающих различия в сфере здравоохранения в регионах РФ.

Ключевые слова: здравоохранение, основные показатели, факторный анализ, кластерный анализ, метод главных компонент, матрица факторных нагрузок.

1. Показатели, характеризующие уровень развития здравоохранения в регионах РФ

Для того чтобы узнать, в каких регионах наиболее и наименее развито здравоохранение, необходимо было собрать данные региональной статистики по группе показателей. В качестве источника данных были использованы сайты Федеральной службы государственной статистики [3]. Все данные были собраны на конец 2010, 2011 и 2012 гг., по всем регионам Российской Федерации.

Для определения уровня развития здравоохранения регионов РФ были взяты следующие данные:

1. Смертность населения по основным классам причин смерти по субъектам РФ (умершие от всех причин смерти) (число умерших на 100 000 человек населения).
2. Число больничных коек по субъектам РФ (тысяч коек).
3. Общая численность инвалидов по субъектам РФ (на 1 000).

4. Среднесписочная численность работников здравоохранения (тысяч человек).
5. Число больничных организаций по субъектам РФ (на 1 000 человек).
6. Заболеваемость населения (диагноз первый раз в жизни, на 1 000 человек).
7. Численность врачей всех специальностей (тысяч человек).

Задачей анализа является конструирование обобщенных характеристик, описывающих различия в сфере здравоохранения в регионах РФ.

2. Факторный анализ показателей, характеризующих развитие здравоохранения

Факторный анализ представляет собой многомерный статистический метод, который применяется для изучения взаимосвязей между значениями переменных. Главная цель данного анализа состоит в том, чтобы уменьшить размерность исходных данных для их экономного описания при условии минимальных потерь исходной информации [1]. Результатом факторного анализа является переход от множества исходных переменных к существенно меньшему числу новых переменных – факторов. При этом фактор можно интерпретировать, как причину совместной изменчивости нескольких исходных переменных.

В исследовании посредством метода анализа главных компонент определено минимальное число факторов, которые вносят наибольший вклад в дисперсию данных. Эти факторы называются главными компонентами.

Таблица 1

Таблица корреляционных значений признаков

Variable	Смертность	число больничных коек	Числ. инвалидов	числ. работников	Больш. организации	Заболеваемость	Врачи всех спец.
Смертность	1,000	-0,009	0,360	-0,026	0,060	-0,031	-0,076
число больничных коек	-0,009	1,000	-0,232	0,988	-0,486	-0,061	0,954
Численность инвалидов	0,360	-0,232	1,000	-0,239	-0,022	-0,069	-0,251
численность работников	-0,026	0,988	-0,239	1,000	-0,503	-0,057	0,943
Больничные организации	0,060	-0,486	-0,022	-0,503	1,000	-0,011	-0,448
Заболеваемость	-0,031	-0,061	-0,069	-0,057	-0,011	1,000	-0,053
Врачи всех спец.	-0,076	0,954	-0,251	0,943	-0,448	-0,053	1,000

В таблице 1 изображены значения корреляций признаков. По ней можно заметить, что между значениями коэффициента «число больничных коек» и коэффициентом «среднесписочная численность работников здравоохранения» имеется сильная зависимость ($r = 0,988$). Очень сильная связь также наблюдается между коэффициентами «число больничных коек» и «численность врачей всех специальностей» ($r = 0,954$). По данной таблице можно сказать, что имеется достаточное количество сильно коррелирующих значений между показателями для того, чтобы применять метод главных компонент.

Для того чтобы определить, сколько первых главных компонент следует оставить для дальнейшего анализа, необходимо определить собственные значения, построить график для наглядного представления, а также определить дисперсии главных компонент и их относительный вклад в суммарную дисперсию признаков.

Таблица собственных значений и дисперсий компонент				
Переменная	Собственные значения	% общей дисперсии	Кумулят. собств. значения	Кумулят. % (накопленные частоты)
Смертность	3,307042	47,24346	3,307042	47,2435
Койки	1,337682	19,10975	4,644724	66,3532
Инвалиды	0,990225	14,14607	5,634949	80,4993
Численность работников	0,794342	11,34774	6,429291	91,8470
Больничные организации	0,499402	7,13432	6,928693	98,9813
Заболеваемость	0,060711	0,86730	6,989404	99,8486
Врачи	0,010596	0,15137	7,000000	100,0000

Простейшим методом отбора главных компонент является правило Кайзера. Оно гласит, что значимы только те главные компоненты, для которых собственное значение будет больше единицы. В нашем случае по таблице 2 видно, что для дальнейшего рассмотрения будут взяты первые два фактора, так как их собственные значения больше 1. Таким образом, будет рассматриваться только 66,3532 % от общей дисперсии.

Для того чтобы удостовериться в правильности числа выборов факторов, можно воспользоваться графиком собственных значений.

На рисунке 1 изображен график собственных значений. Опираясь на данный график, для определения числа факторов можно воспользоваться критерием каменистой осыпи, который впервые был предложен Кэттелем в 1966 году. Им было предложено найти такое место на графике, где убывание собственных значений слева направо максимально замедляется. Опираясь на этот критерий, можно выделить первые два фактора, так как дальше идет заметное замедление. Таким образом, для дальнейшего анализа оставляем первые 2 главные компоненты.

Далее необходимо построить матрицу факторных нагрузок, которая будет показывать значения парных коэффициентов корреляции исходных признаков.

В таблице 3 представлены значения матрицы факторных нагрузок без вращения. Полученные данные трудно проинтерпретировать, поэтому, без сомнения, нужно воспользоваться вращением матрицы факторных нагрузок.

В качестве метода вращения была выбрана стратегия «варимакс», с помощью которой происходит вращение, которое максимизирует дисперсии исходного пространства переменных. Другими словами, целью вращения является максимизация дисперсии новой переменной (фактора) и минимизация разброса вокруг нее.

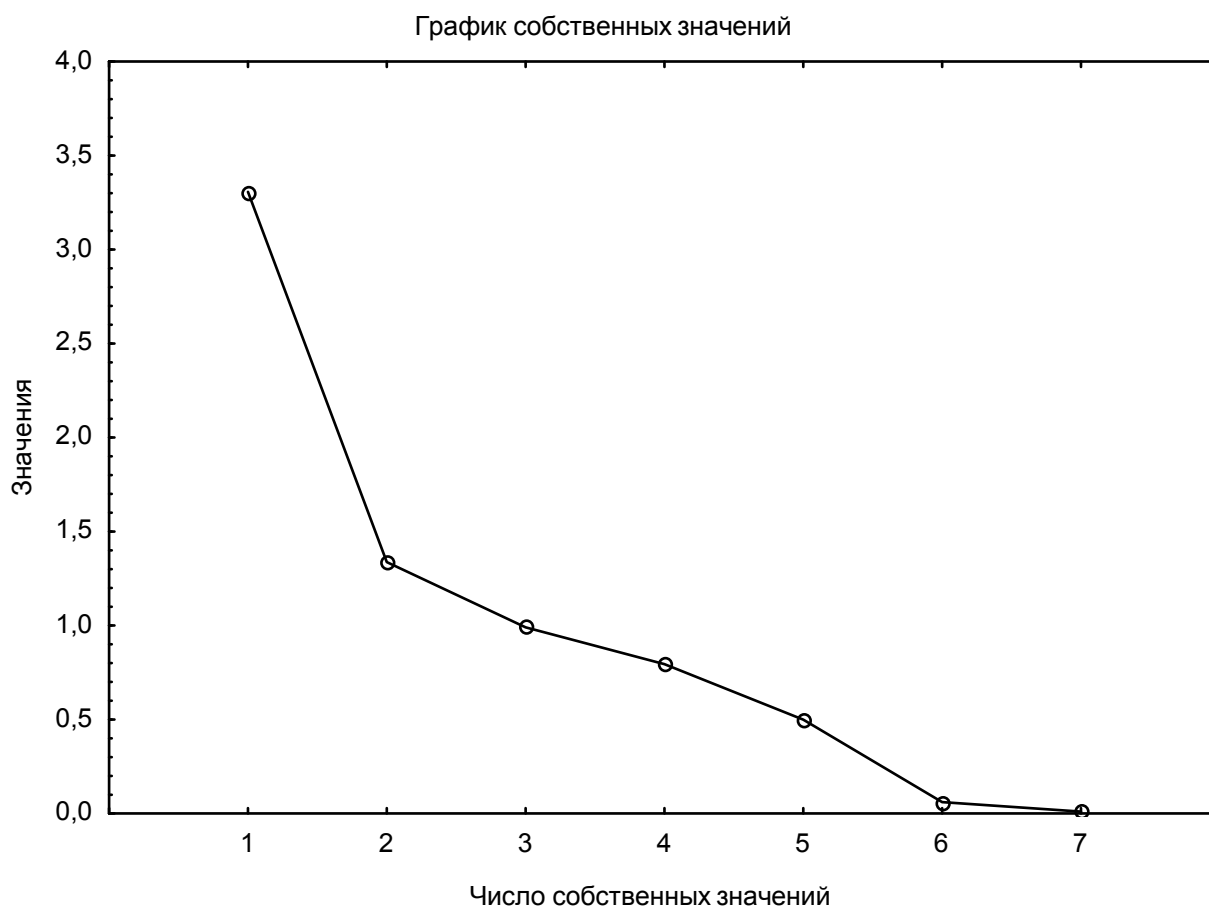


Рис. 1. График собственных значений

Таблица 3

Матрица факторных нагрузок без вращения

Переменные	Фактор 1	Фактор 2
Смертность	0,110	0,793
Больничные койки	-0,976	0,092
Инвалиды	0,313	0,768
Численность работников	-0,977	0,079
Больничные организации	0,604	-0,192
Заболеваемость	0,058	-0,258
Врачи всех специализаций	-0,960	0,030

Таблица 4

Матрица факторных нагрузок после вращения методом «варимакс»

Переменные	Фактор 1	Фактор 2
Смертность	0,004	-0,800
Больничные койки	0,979	0,048
Инвалиды	-0,200	-0,805
Численность работников	0,978	0,061
Больничные организации	-0,626	0,104
Заболеваемость	-0,095	0,247
Врачи всех специализаций	0,954	0,107

В таблице 4 изображена матрица факторных нагрузок после вращения. Сравнивая данную таблицу с таблицей 3, можно заметить, что интерпретация показателей стала намного проще. Таким образом, можно объяснить данную матрицу следующим образом:

- первый фактор связан с числом больничных коек, среднесписочной численностью работников здравоохранения и численностью врачей всех специальностей;
- второй фактор связан со смертностью и с общей численностью инвалидов.

Последним шагом данного анализа является построение выражений для главных компонент через исходные признаки. Для этого нужно рассчитать матрицу факторных коэффициентов.

В таблице 5 представлены значения коэффициентов факторов, на основе которых можно записать выражения для каждой главной компоненты через исходные признаки.

Таблица 5

Матрица коэффициентов факторов

Переменные	Фактор 1	Фактор 2
Смертность	0,052	-0,591
Больничные койки	0,302	-0,026
Инвалиды	-0,012	-0,582
Численность работников	0,301	-0,016
Больничные организации	-0,201	0,116
Заболеваемость	-0,045	0,188
Врачи всех специализаций	0,290	0,019

Для первой главной компоненты уравнение будет иметь следующий вид:

Фактор 1 = 0,052 * Смертность + 0,302 * Койки – 0,012 * Инвалиды + 0,301 * Численность работников – 0,201 * Бол. орган – 0,045 * Заболеваемость + 0,290 * Врачи.

Для второй главной компоненты уравнение будет иметь следующий вид:

Фактор 2 = -0,591 * Смертность – 0,026 * Койки – 0,582 * Инвалиды – 0,016 * Численность работников + 0,116 * Больничные организации + 0,188 * Заболеваемость + 0,019 * Врачи.

Эти результаты могут означать следующее: первый фактор показывает развитие инфраструктуры здравоохранения, а второй фактор – общее состояние заболеваемости населения.

3. Кластерный анализ регионов РФ по уровню развития системы здравоохранения

Кластерный анализ является многомерной статистической процедурой. Суть данного анализа состоит в том, чтобы объединить некоторые объекты в классы (кластеры) так, чтобы в один кластер попадали максимально схожие, а объекты различных классов максимально отличались друг от друга [2]. Метод k -средних является одним из методов кластерного анализа. В его основу положено то, что вся совокупность признаков разделяется на k кластеров, заданное самостоятельно. Причем каждое наблюдение относится к тому кластеру, к центру которого оно ближе всего. В данной работе будет производиться кластерный анализ по двум главным компонентам методом k -средних, которые были выявлены в ходе факторного анализа в пункте 2.

Для проведения кластерного анализа по регионам необходимо было сохранить значения двух факторов: первый фактор, показывающий развитие инфраструктуры, и второй фактор, который показывает общее состояние заболеваемости населения.

Таблица 6

Значения двух факторов, выделенных с помощью факторного анализа, по всем регионам

Регион	Значение фактора 1	Значение фактора 2
Белгородская область	0,3344	-2,0806
Брянская область	0,3437	-0,6202
Владимирская область	0,4855	-0,7022
...
Чукотский автономный округ	0,3665	2,1681

В таблице 6 представлены значения факторов по нескольким регионам в виде краткой иллюстрации. Далее необходимо на основе этой таблицы методом k -средних провести разбиение регионов по двум факторам на несколько кластеров. Количество кластеров было выбрано самостоятельно (2 кластера), а правильность выбора была проверена с помощью дисперсионного анализа. После проведения кластерного анализа получились следующие данные.

Таблица 7

Классификация регионов после проведения кластерного анализа

Номер кластера	Регионы, входящие в кластер
Кластер 1	Области: Белгородская, Брянская, Владимирская, Воронежская, Ивановская, Калужская, Костромская, Курская, Липецкая, Орловская, Рязанская, Смоленская, Тамбовская, Тверская, Тульская, Ярославская, Вологодская, Калининградская, Ленинградская, Новгородская, Псковская, Кировская, Нижегородская, Оренбургская, Пензенская, Ульяновская, Курганская, Иркутская, Кемеровская, Омская, Амурская, Еврейская автономная область.

	Республики: Карелия, Адыгея, Карачаево-Черкесская, Северная Осетия – Алания, Чеченская, Марий Эл, Мордовия, Алтай, Бурятия; Ставропольский край, Пермский край, Забайкальский край
Кластер 2	Города: Москва и Санкт-Петербург. Области: Московская, Архангельская, Мурманская, Астраханская, Вологодская, Ростовская, Самарская, Саратовская, Свердловская, Тюменская, Челябинская, Новосибирская, Томская, Магаданская, Сахалинская. Республики: Коми, Калмыкия, Дагестан, Ингушетия, Кабардино-Балкарская, Башкортостан, Татарстан, Удмуртская, Чувашская, Тыва, Хакасия. Краснодарский край, Алтайский край, Красноярский край. Ханты-Мансийский автономный округ – Югра, Ямало-Немецкий автономный округ, Чукотский автономный округ

В таблице 7 представлена классификация регионов по двум главным компонентам. В первый кластер входит 44 региона. Во второй кластер вошло 38 регионов, где один из них – г. Москва. При этом расстояние данного объекта до других кластеров получилось наибольшее. Это значит, что при выборе большего количества кластеров, возможно, г. Москва стала бы, вероятнее всего, отдельным кластером.

Для того чтобы проверить правильность количества выбранных кластеров, был проведен дисперсионный анализ (табл. 8).

Таблица 8

Дисперсионный анализ факторов

Variable	Analysis of Variance (Spreadsheet14)					
	Between SS	df	Within SS	df	F	signif. p
FACTOR1	10,15384	1	70,84616	80	11,46579	0,001101
FACTOR2	39,52582	1	41,47418	80	76,24178	0,000000

При рассмотрении уровня значимости в таблице 8 можно сделать вывод, что регионы существенно различаются по значениям всех факторов. Следовательно, выбранное количество кластеров можно оставить.

Для того чтобы дать характеристику и выявить отличительные особенности каждого кластера, необходимо построить график средних значений главных факторов по кластерам (см. рис. 2).

Опираясь на рисунок 2, можно дать характеристику каждому кластеру. Как видно, первый кластер отличается высоким значением первого фактора, который показывает развитие инфраструктуры, и низким значением второго фактора, описывающего общее состояние заболеваемости населения. Это означает, что в этих регионах большое количество среднесписочной численности работников здравоохранения и врачей всех специальностей, а также число больничных коек. При этом в данных регионах низкая смертность и общее число инвалидов.

Второй кластер отличается низким значением первого фактора, характеризующим общее состояние заболеваемости населения и высоким значением второго фактора, показывающим развитие инфраструктуры здравоохранения.

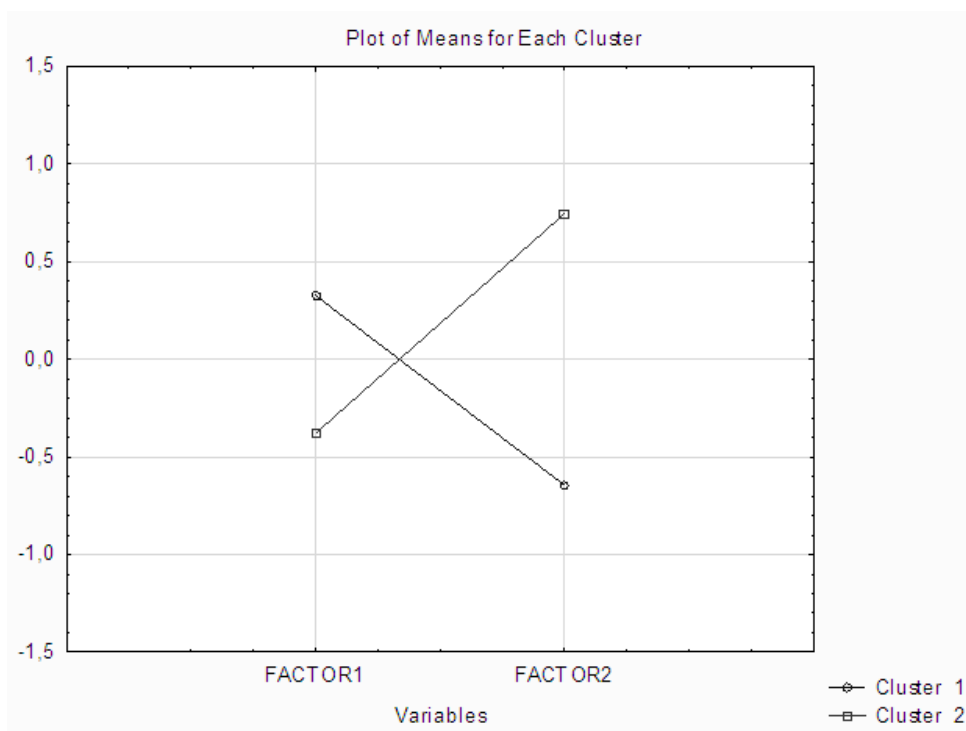


Рис. 2. График средних значений главных компонент

Для улучшения развития инфраструктуры в регионах второй группы необходимо сосредоточиться на снижении смертности и числа инвалидов. Для этого необходимо улучшить инфраструктуру здравоохранения за счет увеличения числа работников здравоохранения и числа больничных коек.

Выводы

Рассмотрены основные показатели региональной статистики, отвечающие за развитие здравоохранения в регионах РФ. Проведен факторный и кластерный анализ для классификации регионов по уровню развития системы здравоохранения. На основе данных видов анализов получились следующие выводы:

1. Факторный анализ выявил два основных фактора, по которым можно получить представление об уровне развития здравоохранения в регионах РФ. Первый фактор характеризует уровень развития здравоохранения посредством числа больничных коек, среднесписочной численностью работников здравоохранения и численностью врачей всех специальностей, тем самым показывает ресурсную обеспеченность. Вторым фактором показывает общее состояние заболеваемости, о котором можно судить по динамике таких показателей, как смертность и общая численность инвалидов.

2. При проведении кластерного анализа методом k -средних объекты наблюдения были разбиты на два кластера.

Для регионов первого кластера характерно большое количество среднесписочной численности работников здравоохранения и врачей всех специальностей, а также число больничных коек. При этом в данных регионах низкая смертность и общее число инвалидов.

Второй кластер отличается высокой смертностью и общим числом инвалидов при низкой среднесписочной численности работников здравоохранения, врачей всех специальностей и низким числом больничных коек, что должно лежать в основе комплекса корректирующих мероприятий в программах развития здравоохранения в регионах.

СПИСОК ЛИТЕРАТУРЫ

1. Калинина, В. Н. Введение в многомерный статистический анализ / В. Н. Калинина, В. И. Соловьев. – М. : ГУУ, 2003.
2. Харман, Г. Современный факторный анализ / Г. Харман. – М. : Статистика, 1972. – 488 с.
3. <http://www.gks.ru/>

REFERENCES

1. Kalinina V.N., Solovyev V.I. Vvedenie v mnogomernyy staticheskiy analiz [Introduction to the Multi-dimensional Statistical Analysis]. Moscow, GUU Publ., 2003.
2. Kharman G. Sovremennyy faktornyy analiz [Modern Factorial Analysis]. Moscow, Statistika Publ., 1972. 488 p.
3. <http://www.gks.ru/>

**CLASSIFICATION OF RUSSIAN REGIONS BY THE LEVEL
OF THE HEALTHCARE SYSTEM DEVELOPMENT**

Boldina Kristina Sergeevna

Master Student in Applied Informatics in Analytical Economy, Institute of Management and Regional Economy, Volgograd State University
 kristi_b92@mail.ru, manag@volsu.ru
 Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. This article emphasizes the classification of regions by the level of the healthcare system development in the Russian Federation. The main goal of factorial and cluster analysis consists in classifying variables by homogenous groups (clusters). First of all, with the help of the regional statistics data on the indicators groups it is possible to carry out the factorial analysis on the basis of the main components. It is aimed at sorting out those factors, which have the biggest impact on the healthcare system development in these regions. The result of this factorial analysis is the transition from a set of initial variables to a significantly smaller number of new variables – factors. A possible interpretation of the received results can be made after finding a matrix of factorial capacity, in which the first factor shows the development of infrastructure and healthcare. The second factor shows the general condition of the population's incidence. Subsequently, a cluster analysis was carried out with the help of the method that deals with the k-averages on the two main components. These components were revealed during the factorial analysis to unite some objects in classes (clusters) in order to get the most similar objects in one cluster and ensure that those objects belonging to different classes can be distinguished from each other on the highest level possible. After carrying out the cluster analysis the following data was obtained: the first cluster includes 44 regions, the second cluster includes 38 regions. After that, a schedule of average values of the main factors on clusters was constructed and each cluster was characterized. The results showed that the first cluster emphasizes a high value of the first factor and a low value of the second one, while the second cluster shows a high value of the second one.

Key words: healthcare, main indicators, factorial analysis, cluster analysis, method of main components, matrix of factorial loadings.