Реферат

Социальные различия имущественной обеспеченности населения в российских регионах

(на основе данных Госкомстата РФ)

Москва, 2008

**Содержание**

Введение

Выбор методов

Описание использованных данных

Анализ и интерпретация

Список использованных источников

Приложение. Порядок выполнения анализа в SPSS

**Введение**

Проблема изучения различий в материальной обеспеченности населения России в региональном аспекте является весьма актуальной для нашей страны. Дело в том, что огромная территория страны подразделена на множество регионов, каждый из которых имеет своё управление, свой климат, природные богатства, свой экономический и социальный потенциал, даже своё законодательство по отдельным вопросам. При этом для обеспечения территориально-административной целостности страны выгодно было бы, чтобы между российскими регионами не существовало чрезмерных различий в экономическом развитии и уровне жизни населения. Исключительная поляризация регионов по социально-экономическим показателям приводит к нарушению различных миграционных балансов, появлению устойчиво отсталых регионов, перенаселению отдельных регионов (например, московский регион) с сопутствующим этим процессам багажом социальных проблем. Кроме прочего, сильная дифференциация между регионами способствует усилению социальной напряжённости, лишает население России понимания того, что они живут в единой стране (вспомним хотя бы многочисленные шутки и анекдоты относительно того, что современная Москва – это государство в государстве и москвичи, порой, не понимают, что за пределами Москвы есть ещё «какая-то Россия»).

Поэтому я воспользовавшись процедурами кластерного анализа решила посмотреть, можно ли разбить российские регионы на однородные группы по ряду показателей, характеризующих обеспеченность и неравенство населения по денежным доходам (например, коэффициенты неравенства доходов – Джини и коэффициент фондов, соотношения различных показателей доходов с величиной прожиточного минимума (далее – с ПМ), численности населения с доходами, ниже ПМ). Ясно, что в России есть «богатые» и «бедные» регионы. Однако, поскольку прожиточный минимум в бедных регионах, как правило ниже ПМ в богатых, можно предположить, что в этом плане уровень жизни в первых и вторых может и не слишком различаться. Ответ на этот вопрос (есть ли обособленные группы регионов или нет) и поможет нам дать проводимый далее анализ.

Естественно, что в данной работе используются данные региональной статистики, которые являются большой частью всей государственной статистики. Положительным моментом, связанным с использованием официальных статистических данных для характеристики уровня жизни населения является «повсеместный охват всей территории страны и единообразная методология и методика получения данных, что важно для межрегиональных сравнений» [2]. Л. А. Беляева отмечает и недостатки, связанные с использованием официальной статистики: недоучёт реальных доходов вследствие их сокрытия населением, условности, связанные с расчётом величины прожиточного минимума (она используется в нашей работе), а также – принципиальный недостаток, связанный с потерей связи опубликованных данных со многими параметрами, которые влияют на положение каждого отдельного индивида. Но в данном случае нас интересуют межрегиональные сопоставления, это во-первых, а во вторых – общие закономерности в развитии регионов, так что от точной оценки доходов тут, конечно, мало что зависит. В работе изучается социально-экономическое положение (прежде всего - дифференциация по доходам) в регионах РФ. Исследуются различия в социально-экономическом положении населения в регионах. На основе статистических исследований Росстата, построена разобрана кластерная модель распределения регионов России по социально-экономическому положению. Методологическая часть содержит краткое описание используемых методов анализа: теоретические аспекты, а также практическое применение для построения статистической модели - группировки регионов России по показателям обеспеченности населения.

Итак, целью данной работы являлось распределение регионов России по однородным группам и установление качественных взаимосвязей между группами регионов с близкими значениями показателей социально-экономического положения. Задача решается с помощью кластерного анализа в системе SPSS.

**Выбор методов**

При анализе социально-экономических процессов приходится довольно часто сталкиваться с многомерностью их описания. В маркетинге, например, это случается при решении задачи сегментирования рынка, в экономике – при построении типологии стран, в социологии – при анализе мнений респондентов по разным вопросам, в социальной статистике – как сейчас, при изучении регионов, каждый из которых описывается по множеству параметров. Многомерный анализ, куда входит и кластерный анализ - важный количественный инструмент исследования социально-экономических процессов, которые характеризуются большим числом показателей. Название кластерный анализ происходит от английского cluster (гроздь, скопление). Впервые определение кластерного анализа и его описание были даны в 1939 Трионом (Tryon). «Главное назначение кластерного анализа - разбиение множества исследуемых объектов и признаков на однородные в соответствующем понимании группы или кластеры. Это означает, что решается задача классификации данных и выявления соответствующей структуры в ней. Методы кластерного анализа можно применять в самых различных случаях, даже в тех случаях, когда речь идет о простой группировке, в которой все сводится к образованию групп по количественному сходству» [9].

Кластерный анализ позволяет рассматривать достаточно большой объем информации и резко сокращать, сжимать большие массивы экономической информации, делать их компактными и наглядными, то есть, в нашем случае – рассматривать не каждый регион в отдельности, а группы регионов, сравнивать их между собой.

Для проведения кластерного анализа используются различные компьютерные программы. Процесс кластерного анализа данных в системе SPSS включает в себя следующие этапы:

* Ввод данных в систему;
* Преобразование данных, адекватное методу кластерного анализа;
* Визуализацию данных с помощью различных типов графиков;
* Реализацию алгоритма метода кластерного анализа;
* Вывод результатов анализа в виде графиков и электронных таблиц с численной и текстовой информацией;
* Интерпретацию полученных результатов.

Общий принцип кластерного анализа (КА) такой. Если некая совокупность содержит набор объектов, свойства которых описываются с помощью некоторых признаков, то задача КА заключается в разбиении совокупности объектов на группы, такие, чтобы каждый объект входил только в одну группу, объекты из одной и той же группы были похожи друг на друга, а объекты из разных групп имели заметные различия. Группы сходных друг с другом объектов называют кластерами*.* Разбивку исходной совокупности на кластеры называют кластерным решением [4, 5]*.*

Как правило, общее количество кластеров и их примерный состав заранее не известны. Для отыскания и исследования кластеров применяются вычислительные алгоритмы, использующие различные способы измерения сходства объектов и групп объектов и различные схемы поиска кластерного решения. В данной работе используются алгоритмы кластерного анализа иерархический и k-средних. Их взаимодействие такое. С помощью иерархического анализа мы проводим предварительный анализ и находим, на какое число кластеров можно было бы разбить всю совокупность регионов. После этого мы проводим кластерный анализ методом k-средних, задав разбиение на выбранное число кластеров. Такая схема рекомендуется в некоторых методических публикациях [5].

Важно отметить ещё такой момент. Довольно часто признаки имеют разный масштаб и разные единицы измерения. Признак, имеющий большую изменчивость (большую дисперсию), вносит больший вклад в величину расстояния между объектами, чем другие признаки при проведении кластерного анализа. И наоборот: признаки с малой изменчивостью (малой дисперсией) фактически не влияют на величины. Поэтому при больших различиях изменчивости рекомендуется провести стандартизацию данныхи попробовать найти кластерное решение на основе стандартизированных данных. Стандартизация заключается в вычитании из значения признака его среднего значения и делении результата на стандартное отклонение признака. Стандартизированные значения иногда также называют Z-вкладами [7].

**Описание использованных данных**

В данной работе используются данные из сборника Госкомстата «Социальное положение и уровень жизни населения России. 2005» за 2004 год [1]. То есть, сборник выпущен в 2005 году, но, поскольку статистическому учёту на такой большой территории как Россия присуще некоторое запаздывание, данные относятся к 2004 году.

Данный вид информации собирается Федеральной службой государственной статистики (Росстатом). Как следует из методологических пояснений к данным, «главной задачей Росстата является удовлетворение потребностей органов власти и управления, средств массовой информации, населения, научной общественности, международных организаций в разнообразной, объективной и полной информации … Международная экспертиза признала статистические данные Федеральной службы государственной статистики надежными … Сбор статистических данных проводится органами государственной статистики в соответствии с Федеральной программой статистических работ, ежегодно утверждаемой Росстатом по согласованию с Правительством Российской Федерации. Обследование организовано во всех субъектах Российской Федерации» [1]

Таким образом, в нашем распоряжении имеются данные о социально-экономической обстановке по регионам Российской Федерации. Указанный уровень обобщения (регион) будет представлять исходные данные для построения кластерной модели в нашей работе. См. пример данных в таблице 1. Полностью данные приведены в сборнике [1].

*Таблица 1 «Показатели социально-экономической дифференциации регионов РФ за 2004 г» (фрагмент)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Регион | Коэффициент Джини | Коэффициент фондов | Соотношение среднедушевых денежных доходов с величиной ПМ, % | Соотношение среднемесячной начисленной зарплаты с величиной ПМ, % | Соотношение среднего размера назначенных месячных пенсий с величиной ПМ, % | Численность населения с денежными доходами ниже величины ПМ |
| Белгородская область | 0,352 | 10,2 | 244,2 | 263,o | 125,2 | 21,2 |
| Брянская обл. область | 0,355 | 10,3 | 223,2 | 209,2 | 125,2 | 25,3 |
| … | … | … | … | … | … | … |
| Санкт-Петербург | 0,410 | 15,3 | 381,3 | 267,3 | 106,2 | 13,5 |

Единицей наблюдения является регион, а показателями – статистика по этому региону, полученная на основе исследований. Хотя данные присутствовали почти по всем российским регионам, в анализе была использована информация только по 70 из них, включая Москву и Санкт-Петербург (о причинах этого см. ниже).

Регион характеризуется 6 показателями. По результатам кластерного анализа можно ожидать появление «богатых» и «бедных» регионов (или же регионов с высоким и низким уровнем жизни). В данном случае нас интересует типовые группы регионов рассматриваемых по схожим социально-экономическим показателям.

Для работы с базой данных и статистического анализа используется статистический пакет SPSS 13.0 для Windows. Для обработки в статистическом пакете информация должна быть организована в особом виде. Традиционным представлением является прямоугольная таблица, матрица данных. В исходных данных представлена статистика по регионам, а также – информация по группам регионов (федеральным округам) и России в целом. Для того чтобы обработать данные в статистическом пакете, нужно привести их к нужной структуре т.е. оставить только информацию по регионам.

В файле данных информация по показателям социально-экономического положения представлена переменными (информация об одном и том же показателе записывается в один столбец, а регион формирует строку файла данных). Список переменных с их краткой характеристикой из [1] представлен в таблице 2.

*Таблица 2 «Список переменных»*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Имя | Тип | Описание | Смысл показателя |
| 1 | Region | Номинальный | Регион | Регион |
| 2 | Gini | Числовой | Коэффициент Джини  | (индекс концентрации доходов / заработной платы) Характеризует степень отклонения линии фактического распределения общего объема доходов /заработной платы от линии их равномерного распределения. |
| 3 | Fond | Числовой | Коэффициент Фондов | Коэффициент дифференциации доходов / заработной платы. Характеризует степень соц. расслоения и определяется как соотношение между средними уровнями денежных доходов / заработной платы 10% процентов населения (работников) с самыми высокими доходами и 10 % процентов населения (работников) с самыми низкими доходами / заработной платой |
| 4 | Sdohod | Числовой | Соотношение среднедушевых денежных доходов с величиной ПМ, % | Характеризует общий уровень денежных доходов населения относительно установленного прожиточного минимума |
| 5 | Szarplata | Числовой | Соотношение среднемесячной начисленной зарплаты с величиной ПМ, % | Характеризует общий уровень заработной платы населения относительно установленного прожиточного минимума |
| 6 | Spensii | Числовой | Соотношение среднего размера назначенных месячных пенсий с величиной ПМ, % | Характеризует общий уровень пенсий населения относительно установленного прожиточного минимума. |
| **7** | ChislMin | Числовой | Численность населения с денежными доходами ниже величины ПМ | Определяется на основе данных о распределении населения по величине среднедушевых денежных доходов и является результатом их соизмерения с величиной прожиточного минимума |

На практике большую проблему представляют пропущенные значения (пункты, по которым отсутствует информация). Связано это с тем, что нельзя отнести регион к какому либо кластеру, не имея полной информации о нем. В данной таблице пропусков довольно мало. Информация полностью отсутствует по Чеченской Республике. Статистические исследования в этом регионе не проводились в связи с проходившими на территории Чечни военными действиями. Также, вне зоны нашего внимания останутся такие регионы как Архангельская область, Пермская область, Тюменская область, Красноярский край, Иркутская область и Читинская область из-за частичного или полного отсутствия статистических данных. По остальным регионам, включая Москву и Петербург, все данные находятся в нашем распоряжении. Поэтому в дальнейшем исключим выше перечисленные субъекты из рассмотрения, и модель будем строить на базе информации о 70 регионах РФ.

**Анализ и интерпретация**

В задачи работы входит построение кластерной модели социально-экономического положения по регионам РФ. Требуется выделить группы регионов, имеющих схожую, однородную социально-экономическою обстановку. Таким образом, исходными данными должна являться статистика показателей социально-экономического положения на региональном уровне (по всем регионам РФ).

**Описательная статистика.** Для начала работы стоит провести разведочный анализ с целью определения перспектив кластеризации. Прибегнем к возможностям SPSS и получим описательные статистики показателей социально-экономического положения.

*Таблица 3 «Описательная статистика показателей»*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   | Количество | Минимум | Максимум | Среднее | Стандартное отклонение |
| Gini | 70 | ,314 | ,578 | ,36346 | ,036434 |
| Fond | 70 | 7,8 | 44,0 | 11,430 | 4,4475 |
| Sdohod | 70 | 105,1 | 674,0 | 231,597 | 73,1439 |
| Szarplata | 70 | 168,1 | 309,4 | 234,240 | 36,9582 |
| Spensii | 70 | 80,7 | 132,6 | 111,811 | 13,0245 |
| ChislMin | 70 | 13,5 | 73,0 | 26,624 | 10,0924 |

Наиболее важным показателем для нас является стандартное отклонение. Чем выше стандартное отклонение величины (чем больше ее изменчивость), тем больше эта переменная будет оказывать влияние на результаты кластеризации. Мы видим, например что переменные Sdohod или Szarplata имеют наибольшее стандартное отклонение и возможно разделение регионов на группы именно по этим признакам. Практически большее стандартное отклонение означает, что между регионами существует большая дифференциация по данным показателям: в какой-то части регионов показатели малы, а в какой-то части регионов наоборот велики. Таким образом, справедливо ожидать, что показатели с большей «дифференцирующей способностью» (большей дисперсией) окажут большее влияние на результат кластеризации [4, 5]

Видно, что, как и для многих экономических показателей, чем меньше среднее показателя, тем меньше стандартное отклонение этого показателя. Связано это с невозможностью переменных принимать отрицательные значения. Это значит, что наибольшее влияние на результаты кластеризации окажут переменные с большей средней величиной.

Так же большое подспорье в оценке перспектив кластеризации окажут гистограммы абсолютных или стандартизированных значений по тем переменным по которым она проводиться. Далее мы будем рассматривать только стандартизированные значения, в связи с тем, что стандартные отклонения слишком различны. Стандартизация показателей проводилась с помощью меню SPSS Analyze – Descriptive Statistics – Descriptives с установкой флажка стандартизации. Изучим гистограммы наших показателей.

При анализе гистограммы Коэффициента Джини мы видим, что все регионы отчетливо делятся на две группы. Причем одна из этих групп крайне велика и туда входит большая часть всех регионов РФ.

Анализ гистограмм для коэффициента фондов, численности населения с доходами, ниже ПМ и соотношения денежных доходов с ПМ выявил аналогичные распределение регионов как и для коэффициента Джини.

А вот анализ гистограмм Szarplata и Spensii не выявил очевидного деления на какие-либо группы регионов.

Стоит также рассмотреть и корреляции всех показателей.

*Таблица 4 «Матрица корреляций показателей»*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|   | Gini | Fond | Sdohod | Szarplata | Spensii | ChislMin |
| Gini | 1 | ,953(\*\*) | ,872(\*\*) | ,455(\*\*) | -,225 | -,384(\*\*) |
| Fond | ,953(\*\*) | 1 | ,881(\*\*) | ,370(\*\*) | -,257(\*) | -,301(\*) |
| Sdohod | ,872(\*\*) | ,881(\*\*) | 1 | ,599(\*\*) | -,046 | -,658(\*\*) |
| Szarplata | ,455(\*\*) | ,370(\*\*) | ,599(\*\*) | 1 | -,069 | -,668(\*\*) |
| Spensii | -,225 | -,257(\*) | -,046 | -,069 | 1 | -,250(\*) |
| ChislMin | -,384(\*\*) | -,301(\*) | -,658(\*\*) | -,668(\*\*) | -,250(\*) | 1 |

Как следует из справочной информации по пакету SPSS, звёздочками отмечены значимые корреляции, то есть, те, на которые имеет смысл обращать внимание. Исследовав таблицу корреляций мы можем заметить, что Коэффициент фондов и Коэффициент Джини дают нам крайне близкую информацию (корреляция 0,953) что в свою очередь позволяет нам вместо обоих показателей воспользоваться одним. В данной ситуации более предпочтительным для работы является коэффициент фондов так как он в свою очередь обладает большим относительным разбросом. Об этом нам может сказать коэффициент вариации (отношение стандартного отклонения переменной к среднему значению этой переменной). У коэффициента Джини он составляет порядка 0,1 а у коэффициентов фондов около 0,389 (в таблицах не показано).

**Иерархическая кластеризация.** После изучения полученного результата описательной статистики показателей социально-экономического положения было установлено, что ряд переменных не стоят нашего внимания, а именно – коэффициент Джини.

Было решено провести серию пробных разбиений наблюдений на 2, 3 и так далее кластеров, чтобы установить переменные либо данные, которые играют малозаметную роль в разбиении регионов на кластеры. При этом воспользуемся методом k-средних.

При первом же разбиении на 2 кластера, мы можем судить о том, что город Москва явно превзошел все остальные регионы по уровню социально-экономического развития. И при любом количестве кластеров Москва всегда будет отделяться в отдельный кластер. Поэтому целесообразнее всего будет исключить Москву из рассмотрения в работе, ибо дальнейшее ее рассмотрение не поможет нам объективно взглянуть на общую социально-экономическую обстановку в целом по России. То есть далее мы будем рассматривать только 69 регионов.

Кластерный анализ с разбиением на 3 кластера создал у нас две достаточно большие группы: 24 и 38 регионов и одну маленькую (7 регионов). В самую малочисленную группу попали самые худшие по показателям регионы с достаточно малыми доходами, зарплатами и большой прослойкой населения, доходы которых явно меньше прожиточного минимума. Такие регионы как Ингушетия, Калмыкия, республика Тыва. В основном это регионы с крайне низким уровнем жизни и не развитой экономикой.

Две другие группы оказались более подкованными в этом плане. Во вторую группу попали такие регионы как Тамбовская, Тульская области республика Саха. Лучший результат же показали регионы первого кластера. Самые доходные и социально обеспеченные. Такие как Московская область, Мурманская и Вологодская области.

Далее приводится дендрограмма (график объединения) для иерархического кластерного анализа с оставшимися переменными. Из нее будет видно на каких расстояниях объекты объединяются в кластеры, из этого можно будет сделать вывод на сколько кластеров разбить всю совокупность.

Красной линией на дендрограмме мы отметили один из вариантов кластерного решения, который предусматривает разбиение на 6 кластеров. Это решение даёт следующие центры кластеров:

*Таблица 6 «Кластерные центры по итогам иерархического анализа»*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Кластер** | **1** | **2** | **3** | **4** | **5** | **6** |
| zFond | ,18 | -,32 | -,15 | -,03 | 1,06 | -,70 |
| zSdohod | ,58 | -,25 | -,89 | ,01 | 1,37 | -1,73 |
| zSzarplata | 1,18 | -,35 | -1,23  | ,68  | 1,14 | -1,29 |
| zSpensii. | ,86 | ,36  | -,53 | -1,83  | -,55 | -1,41 |
| zChislMin | -,80 | -,07 | 1,54 | -,21 | -,99 | 4,60 |

**Кластеризация методом k-средних**. Повторим разбиение на 6 групп с помощью метода k-средних. Таблица 7 показывает, как распределились в итоге регионы по кластерам. Последняя графа показывает расстояние от региона до центра его кластера, то есть, как бы говорит, далеко ли регион находится от типичного для данного кластера региона.

Таблица 7 «Принадлежность к кластерам»

|  |  |  |
| --- | --- | --- |
| **Кластер** | **Регион** | **Расстояние**  |
| 1 | Белгородская область | ,986 |
| Московская область | 1,071 |
| Липецкая область | ,968 |
| Ярославская область | ,404 |
| Вологодская область | ,633 |
| Республика Башкортостан | ,933 |
| Ростовская область | 1,217 |
| Волгоградская область | 1,092 |
| Астраханская область | ,862 |
| Республика Татарстан | 1,185 |
| Челябинская область | ,745 |
| Кемеровская область | ,841 |
| Омская область | ,834 |
| Томская область | 1,230 |
| Магаданская область | 1,251 |
| 2 | Брянская область | ,756 |
| Владимирская область | 1,281 |
| Воронежская область | 1,03 |
| Калужская область | ,837 |
| Костромская область | ,529 |
| Орловская область | 1,390 |
| Рязанская область | ,588 |
| Смоленская область | ,579 |
| Тамбовская область | 1,312 |
| Тверская область | 1,190 |
| Тульская область | 1,054 |
| Республика Карелия | 1,161 |
| Ленинградская область | 1,937 |
| Калининградская область | 1,439 |
| Новгородская область | ,633 |
| Псковская область | 1,124 |
| Кабардино-Балкарская Республика | ,793 |
| Карачаево-Черкесская Республика | 1,135 |
| Республика Северная Осетия-Алания | 1,696 |
| Краснодарский край | 1,478 |
| Ставропольский край | 1,052 |
| Удмуртская Республика | ,934 |
| Чувашская Республика | ,867 |
| Кировская область | 1,027 |
| Нижегородская область | ,672 |
| Оренбургская область | ,608 |
| Пензенская область | ,972 |
| Саратовская область | ,931 |
| Республика Хакасия | 1,135 |
| Алтайский край | 1,647 |
| Новосибирская область | 1,273 |
| 3 | Ивановская область | 1,734 |
| Республика Калмыкия | 1,684 |
| Республика Адыгея | ,908 |
| Республика Дагестан | 1,051 |
| Республика Мордовия | 1,155 |
| Республика Марий Эл | 1,317 |
| Ульяновская область | ,843 |
| Курганская область | 1,007 |
| Республика Алтай | 1,197 |
| Республика Бурятия | 2,107 |
| Республика Тыва | 1,163 |
| 4 | Курская область | 1,489 |
| Мурманская область | 1,090 |
| Республика Саха | 1,124 |
| Приморский край | 1,385 |
| Хабаровский край | ,563 |
| Амурская область | ,764 |
| Сахалинская область | 1,198 |
| 5 | Санкт-Петербург | ,794 |
| Свердловская область | ,897 |
| Республика Коми | ,873 |
| Самарская область | 1,381 |
| 6 | Республика Ингушетия | ,000 |

Чем меньше показатель расстояния у данного кластера тем он более ярко выраженный представитель этого кластера.

*Таблица 8. «Количество регионов, попавших в каждый из кластеров»*

|  |  |  |
| --- | --- | --- |
| кластер | 1 | 15,000 |
| 2 | 31,000 |
| 3 | 11,000 |
| 4 | 7,000 |
| 5 | 4,000 |
| 6 | 1,000 |
| верно |  69,000 |
| пропущено | ,000 |

По результатам кластеризации мы получили один крупный кластер (2) , два средних кластера (1), (3), два маленьких кластера (4),(5) и один очень маленький полученный из одного региона РФ (6).

Для получения более ясного представления о кластерах, стоит упомянуть ярких представителей своих групп. В первый кластер попали такие регионы как: Московская область, Омская область, Ярославская область. Во второй попали регионы такие как: Костромская область, Нижегородская и Смоленская области. У третьего кластера яркими представителями оказались Ульяновская область и Адыгея. У четвертого кластера выделились Хабаровский край, Амурская область. А яркими представителями пятого кластера стали Санкт-Петербург, Свердловская область. А вот шестой кластер состоит лишь из одного региона России- Республики Ингушетии.

Для создания качественного представления о социально-экономическом положении (различиях в имущественном обеспечении и неравенстве в доходах) очень полезно будет рассмотреть таблицу окончательных кластерных центров.

*Таблица 9 «Окончательные кластерные центры»*

|  |  |
| --- | --- |
|  | **кластер** |
| 1 | 2 | 3 | 4 | 5 | 6 |
| zFond | ,15515 | -,34257 | -,21524 | -,02923 | 1,06128 | -,70377 |
| zSdohod | ,53023 | -,26863 | -,86786 | ,01469 | 1,37473 | -1,72943 |
| zSzarplata | 1,01466 | -,36758 | -1,20269 | ,67806 | 1,14142 | -1,28902 |
| zSpensii | ,72391 | ,40307 | -,44828 | -1,83259 | -,54984 | -1,40592 |
| zChislMin | -,76469 | -,05674 | 1,37576 | -,21190 | -,99077 | 4,59510 |

В данной таблице отображены окончательные кластерные центры. Координаты кластерного центра – это средние значения по каждой переменной кластеризации по всем регионам, входящим в этот кластер. Учитывая, что переменные стандартизированы, средние значения даны здесь в стандартизированном виде. То есть, 0 означает средний уровень по России, положительное значение – выше среднероссийского уровня, отрицательное – ниже. Таким образом, кластерные центры показывают, чем характеризуется каждый кластер, по каким переменным есть отличия. В данном случае они характеризуют материальное расслоение общества.

**Первый кластер.** Регионы попавшие в данный кластер это по большей части довольно стабильные и высокоразвитые в плане промышленности и социального обеспечения. Достаточно крепкие середняки если можно так сказать. Населения живущее ниже прожиточного минимума не много, небольшое расслоение общества соответствует неплохой социальной подкованности регионов. Уверенные показатели доходов, пенсий и заработной платы говорят о благополучии данных регионов.

**Кластер №2** - самый многочисленный кластер. В него входит основная часть регионов России. Надо сказать достаточно посредственные, ничем особо не выделяющиеся, регионы с довольно сносными показателями. И хотя показатели по большей части не плохи смотрятся они явно хуже чем регионы первого кластера. Небольшое расслоение общества, показатели доходов и заработной платы находятся в отрицательном соотношении с прожиточным минимумом. Приятно удивляет показатель по пенсиям. В общем если данные регионы и можно отнести к середнякам то уж точно к их худшей части.

Третий кластер показал достаточно плохие показатели по численности населения живущего бедно, имеющие достаток явно ниже прожиточного минимума. Низкие доходы, пенсии и особенно зарплаты. Расслоение общества невелико. Связано это прежде всего с отсутствием серьезных возможностей для развития регионов.

Четвертый кластер показал абсолютно близкие к средним по стране показателям по расслоению общества и среднедушевым доходам. Зато заработная плата по большей части оказалась выше среднего, что приятно удивило. Но регионы не без изъянов, а именно уровень пенсий крайне низок. В свою очередь показатель населения живущего на доходы ниже прожиточного минимума не велико.

А вот у пятого кластера социальное расслоение достаточно большое. Очень хорошие, по отношению к другим регионам, показатели заработной платы и доходов говорят о хорошем экономическом развитии регионов. Плохо, по-прежнему, живут пенсионеры чьи пенсии оказались, в среднем, ниже чем в по остальной России. Основная часть населения живет хорошо, об этом говорит показатель численности населения живущего с доходами ниже чем прожиточный минимум. Вполне можно сказать что данные регионы в социально-экономическом плане одни из самых подкованных. Определенно лучшие и самые богатые регионы с большими возможностями и потенциалами.

Шестой кластер явно худший регион страны. Республика Ингушетия является беднейшим. Расслоение не велико. Но это, прежде всего, связано с плохим уровнем жизни абсолютно всего населения региона. Маленькие зарплаты, пенсии, доходы. Скорее всего очень плохо влияет географическое расположение. Близость к Чеченской республике, обилие беженцев. Отсутствие должного количества рабочих мест. Никакой толковой социальной обеспеченности. Привели к тому что регион ярко выразился в худшую сторону по отношению ко всем остальным рассматриваемым регионам.

*Таблица 10*

*«Расстояние между окончательными кластерными центрами»*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Кластер | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 |   | 3,014 | 1,627 | 3,529 | 2,584 | 2,607 |
| 2 | 3,014 |   | 1,774 | 1,804 | 2,578 | 4,928 |
| 3 | 1,627 | 1,774 |   | 2,990 | 2,850 | 3,470 |
| 4 | 3,529 | 1,804 | 2,990 |   | 2,120 | 5,665 |
| 5 | 2,584 | 2,578 | 2,850 | 2,120 |   | 4,737 |
| 6 | 2,607 | 4,928 | 3,470 | 5,665 | 4,737 |   |

В этой таблице показаны расстояния между кластерными центрами. На основании сопоставления данной таблицы со средними расстояниями в каждом из кластеров можно сделать вывод, насколько кластеры обособленны.

Таблица 11 «Значимость переменных при разбиении на 6 кластеров»

|  |  |
| --- | --- |
|  | Sig. |
| zFond | ,000 |
| zSdohod | ,000 |
| zSzarplata | ,000 |
| zSpensii | ,000 |
| zChislMin | ,000 |

Из таблицы дисперсионного анализа можно проанализировать результаты кластеризации, оценив значимость всех переменных. Значимость всех переменных получилась удовлетворительной по уровню 0.05, это говорит о том, что все переменные оказывают достаточно сильное влияние на результаты кластеризации.

Вероятнее всего, явных кластеров в пространстве наших переменных не существует. Учитывая, что многие из исходных переменных имели распределение, близкое к нормальному, скорее всего, рабочая модель распределения регионов в социально-экономическом пространстве выглядит как общероссийский центр с основной массой регионов с характеристиками, близкими к среднероссийскому уровню и небольшое количество «периферийных» регионов, чья структура в силу особенностей конкретного региона заметно отличается от среднероссийской.

Это значит, что можно предложить другие, равнозначные варианты группировки. Разделить на другое количество кластеров, использовать другие показатели или их комбинации и т.д.

**Выводы**

В результате проделанной работы удалось обобщить и проанализировать исходные данные по социально-экономическому положению регионов РФ по состоянию на конец 2004 года. Был выбран метод проведения исследования и построения статистической модели. На основе иерархического кластерного анализа были сделаны предварительные выводы о возможном разбиении на кластеры. Окончательная кластеризация проведена с помощью метода k-средних.

В полученной 6-кластерной модели обнаружены значительные различия в социальном и экономическом развитии регионов, попавших в разные кластеры. Особенности каждого кластера были рассмотрены, также были предложены возможные причины данных особенностей. Был получен один большой кластер, отражающий характерное в среднем положение регионов по России, а также 5 более мелких, менее характерных для России кластера.

В работе удалось обнаружить, что большинство регионов находится в положении близком к среднему по всей России, и лишь небольшая часть регионов сильно отличается от средних показателей. В основном это было заметно на бедных регионах Кавказа и богатых регионов центральной части РФ. Примером может послужить Москва постоянно отделявшаяся в обособленный кластер.

Тем не менее, полученная группировка регионов может быть использована в дальнейшем. Ее можно верифицировать на более поздних данных. Если изменений мало, значит полученная классификация хотя и являлась одной из многих возможных, но всё-таки не случайна.

В работе использовались данные Росстата за 2004 год, которые к настоящему моменту уже несколько устарели. На веб-сайте Госкомстата [3] уже имеются данные о начисленных пенсиях, зарплатах и прожиточном минимуме даже за 2006 год, однако не все исследованы показатели даны в разрезе по регионам (например, индекс Джини или коэффициент фондов даны только в целом по России). Но если запастись свежим статистическим сборником (имеется аналогичный сборник 2006 года), например, то можно построить аналогичную модель по новым данным и сравнить. Возможно, ситуация осталась такой же, возможно появились какие-то изменения, и тогда можно будет поставить вопрос о причине переходов регионов из одного кластера в другой, о смене типологий.

**Список использованных источников**

1. Социальное положение и уровень жизни населения России. 2005: Стат. сб. / Росстат. -М., 2005
2. Беляева, Л.А. Материальное неравенство в России. Реальность и тенденции // Социологические исследования, 2007, №11.
3. Федеральная служба государственной статистики, http//www.gks.ru
4. Бююль, А., Цёфель, П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. – Диасофт, 2005
5. Решение задач в программе SPSS, http://www.spsstools.ru
6. SPSS 13.0. Справочная система.
7. Гайдышев, И. Анализ и обработка данных. Специальный справочник – С.-Пб., 2001
8. Кластерный анализ в задачах социально-экономического прогнозирования, http://www.ref.by/refs/49/28133/1.html
9. Интернет-университет информационных технологий. 13. Лекция: Методы кластерного анализа. Иерархические методы, http://www.intuit.ru/department/database/datamining/13/datamining\_13.html
10. Кунец Н.Л. Кластерный анализ в портфельном инвестировании. Курсовая работа. http://www.5ballov.ru/referats/preview/71794/1

**Приложение. Порядок выполнения анализа в SPSS**

1. Перенесём табличную информацию (исходные данные) из статистического сборника в файл Excel, подписав вверху названия переменных (Gini, Fond и т.д., чтобы затем их SPSS прочёл как переменные);
2. Загрузим SPSS и импортируем туда сохранённые данные из Excel (File – Open – Data, указать тип .xls);
3. При необходимости можно подписать метки переменных;
4. Получим описательную статистику показателей (Analyze – Descriptive Statistics – Descriptives);
5. Удалим из файла данных те регионы, где отсутствует полная информация по всем показателям (осталось 70);
6. Стандартизируем показатели (повторим Analyze – Descriptive Statistics – Descriptives с флажком Save standardized values as variables;
7. Построим гистограммы показателей (Graph – Histogram);
8. Найдём корреляции показателей (Analyze – Correlate – Bivariate);
9. Удалим из файла данных Москву;
10. Проведём пробный иерархический анализ (Analyze – Classify – Hierarchical Cluster);
11. Найдём координаты кластерных центров для сохранённого решения с 6 кластерами (Analyze – Compare Means – Means);
12. Проведём кластеризацию на 6 кластеров с методом k-средних (Analyze – Classify – K-Means Cluster).