**Содержание.**

**Введение………………………………………………………………………..…3**

**Глава 1. Теоретические аспекты кластерного анализа. Метод k-средних.4**

**Глава 2. Кластерный анализ регионов России.....................…………….…..7**

2.1. Евклидова Метрика…...………………………………………………….......9

2.2. Метод k-средних.............................................................................................11

**Глава 3. Регрессионный анализ в кластерах……....……………………….16**

**Глава 4. Дискриминантный анализ………....……………………………….29**

**Заключение……………………………………………………………………...35**

**Список литературы.......………………………………………………………..37**

**Приложения……………………………………………………………………..38**

**Введение.**

Необходимость развития сферы услуг требует существенной перестройки структуры потребления населения, значительного возрастания доли услуг и сокращения удельного веса материального потребления.

На современном этапе социально-экономического развития сфера сервиса становится одним из наиболее обширных секторов развития рыночных отношений, вовлечения населения в экономическую деятельность. Она теснейшим образом связана со всеми другими областями общественной жизни. Соответственно сфера сервиса является одним из важнейших источников мобилизации потенциала национального роста, повышения качества и уровня жизни.

Социальная значимость эффективности сферы услуг очевидна: она непосредственно служит увеличению доли свободного времени населения и, следовательно, создает возможности для всестороннего гармоничного развития личности.

Прогресс в отрасли определяется развитием всех видов электрической связи, среди которых наиболее распространенной является телефонная связь.

Телефонная связь приносит более половины всех доходов, получаемых от услуг связи. Одним из важнейших направлений обслуживания населения в области связи является телевидение и радиовещание, которые несет не только важную информационную и политическую нагрузку, но и являются в некоторых отдаленных районах области почти единственным средством культурного досуга.

Цель данной курсовой работы - статистическое изучение предоставления услуг связи населению в регионах России в 2008 году и влияние их на доход от услуг связи населению в расчете на одного жителя.

**Глава 1. Теоретические аспекты кластерного анализа. Метод k-средних.**

При большом количестве наблюдений иерархические методы кластерного анализа не пригодны. В таких случаях используют неиерархические методы, основанные на разделении, которые представляют собой итеративные методы дробления исходной совокупности. В процессе деления новые кластеры формируются до тех пор, пока не будет выполнено правило остановки.

Такая неиерархическая кластеризация состоит в разделении набора данных на определенное количество отдельных кластеров. Существует два подхода. Первый заключается в определении границ кластеров как наиболее плотных участков в многомерном пространстве исходных данных, т.е. определение кластера там, где имеется большое "сгущение точек". Второй подход заключается в минимизации меры различия объектов.

Наряду с иерархическими методами классификации, существует многочисленная группа так называемых итеративных методов кластерного анализа (метод *k* - средних.).

Сущность их заключается в том, что процесс классификации начинается с задания некоторых начальных условий (количество образуемых кластеров, порог завершения процесса классификации и т.д.). Название метода было предложено Дж. Мак-Куином в 1967 г. В отличие от иерархических процедур метод *k* - средних не требует вычисления и хранения матрицы расстояний или сходств между объектами. Алгоритм этого метода предполагает использование только исходных значений переменных. Для начала процедуры классификации должны быть заданы *k* выбранных объектов, которые будут служить эталонами, т.е. центрами кластеров. [[1]](#footnote-1)

Считается, что алгоритмы эталонного типа удобные и быстродействующие. В этом случае важную роль играет выбор начальных условий, которые влияют на длительность процесса классификации и на его результаты. Метод *k* - средних удобен для обработки больших статистических совокупностей.

Математическое описание алгоритма метода *k* - средних.

Пусть имеется *n* наблюдений, каждое из которых характеризуется *m* признаками *X1* , *X2* , , *Xn*. Эти наблюдения необходимо разбить на *k* кластеров.

Для начала из *n* точек исследуемой совокупности отбираются случайным образом или задаются исследователем исходя из каких-либо априорных соображений *k* точек (объектов). Эти точки принимаются за эталоны.

Каждому эталону присваивается порядковый номер, который одновременно является и номером кластера.

На первом шаге из оставшихся (*n* -*k*) объектов извлекается точка  *Xi* с координатами ( *xi*1 , *xi*2 , ... , *xim* ) и проверяется, к какому из эталонов (центров) она находится ближе всего. Для этого используется одна из метрик, например, евклидово расстояние. Проверяемый объект присоединяется к тому центру (эталону), которому соответствует минимальное из расстояний. Эталон заменяется новым, пересчитанным с учетом присоединенной точки, и вес его (количество объектов, входящих в данный кластер) увеличивается на единицу. Если встречаются два или более минимальных расстояния, то *i* -ый объект присоединяют к центру с наименьшим порядковым номером.

На следующем шаге выбираем точку *Xi+1* и для нее повторяются все процедуры. Таким образом, через (*n-k*) шагов все точки (объекты) совокупности окажутся отнесенными к одному из *k* кластеров, но на этом процесс разбиения не заканчивается. Для того чтобы добиться устойчивости разбиения по тому же правилу, все точки *X1*, *X2*,…, *Xn* опять подсоединяются к полученным кластером, при этом веса продолжают накапливаться. Новое разбиение сравнивается с предыдущим. Если они совпадают, то работа алгоритма завершается. В противном случае цикл повторяется.

Окончательное разбиение имеет центры тяжести, которые не совпадают с эталонами, их можно обозначить *C1* ,*C2* , ,*Ck*. При этом каждая точка *Xi* будет относиться к такому кластеру (классу) *l* , для которого расстояние минимально. Возможны две модификации метода *k* - средних. Первая предполагает пересчет центра тяжести кластера после каждого изменения его состава, а вторая – лишь после того, как будет завершен просмотр всех данных. В обоих случаях итеративный алгоритм этого метода минимизирует дисперсию внутри каждого кластера, хотя в явном виде такой критерий оптимизации не используется.

Достоинства алгоритма k-средних:

• простота использования;

• быстрота использования;

• понятность и прозрачность алгоритма.

Недостатки алгоритма k-средних:

• алгоритм слишком чувствителен к выбросам, которые могут искажать среднее. Возможным решением этой проблемы является использование модификации алгоритма - алгоритм k-медианы;

• алгоритм может медленно работать на больших базах данных. Возможным решением данной проблемы является использование выборки данных.

**Глава 2. Кластерный анализ регионов России.**

Нами исследуется совокупность 62 регионов, каждый из которых характеризуется по 5 замеренным на нем признакам Х. Четыре признака из них характеризуют степень оснащенности населения средствами связи и среднедушевой доход населения, а пятый – показатель дохода от услуг связи, предоставляемых населению. Данные по эти признакам приведены в *Приложении 1.* Вот эти признаки:

X1 – доходы от услуг связи населению в расчете на одного жителя (рублей);

Х2 – число квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (на конец года; штук);

Х3 – средства связи (пользовательское оборудование) для оказания услуг передачи данных и телематических служб на 1000 человек (на конец года;штук);

Х4 – число абонентских терминалов сотовой связи на 1000 человек населения (на конец года; штук);

Х5 – среднедушевые доходы населения (рублей).

Перед началом работы и анализа данных необходимо выявить наличие выбросов, и если они могут повлиять на результаты анализа, удалить их из таблицы исходных данных. Графики исследования на выбросы по признакам X1 и X2, по признакам X1 и X3, и, наконец, по признакам X4 и X5 приведены на рисунках в *Приложении 2*. Проведя анализ по этим диаграммам можно сделать следующие выводы.

Камчатская область является выбросом по трем признакам: X1, X2, X5, следовательно, наличие данных по этому региону может повлиять на результаты дальнейшего анализа. Поэтому, Камчатская область будет удалена из массива исходных данных.

Остальные выбросы определяются по одному или по двум признакам, поэтому можно сделать вывод о том, что эти выбросы обусловлены влиянием посторонних факторов и они не окажут воздействия на результаты дальнейшего анализа. Поэтому из таблицы исходных данных они удалены не будут.

В результате для дальнейшей работы будут использованы данные, указанные в таблице *Приложения 3*. . Перед началом исследования и разбиения объектов на кластеры необходимо провести проверку данных на мультиколлинеарность. Парные коэффициенты корреляции приведены в Табл. 1.1.

Табл. 1.1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *X1* | *X2* | *X3* | *X4* | *X5* |
| X1 | 1 | 0,26315 | 0,25601 | 0,26674 | 0,68608 |
| X2 | 0,26315 | 1 | 0,19883 | 0,18494 | 0,31769 |
| X3 | 0,25601 | 0,19883 | 1 | 0,05359 | 0,21421 |
| X4 | 0,26674 | 0,18494 | 0,05359 | 1 | 0,3323 |
| X5 | 0,68608 | 0,31769 | 0,21421 | 0,3323 | 1 |

Табл 1.1 Парные коэффициенты корреляции.

Как видно из таблицы (Табл. 1.1), между данными нет слишком сильных взаимосвязей, следовательно, построенные кластерные модели будут адекватно отображать действительность.

Так как наши показатели имеют разные меры измерения, то для дальнейшего кластерного анализа нам требуется простандартизировать исходные данные. Таблица стандартизированных данных представлена в *Приложении 3.*

***2.1 Евклидова Метрика***

Теперь, когда наши данные готовы к использованию, можно провести кластерный анализ. Для этого используем методы ближнего и дальнего соседа, а также методы Варда и k-средних.

В кластерном анализе, как известно более распространены иерархические процедуры. Расстояние между исследуемыми объектами будет определяться по обычному невзвешенному Евклидову расстоянию.

После проведенного кластерного анализа было получено три дендрограммы. Наиболее точное разбиение на кластеры дает метод Варда, которая приведена ниже на рис. 2.1.1.

Рис 2.1.1



Рис 2.1.1 Дендрограмма метода Варда.

Проанализировав рисунок 2.1.1, можно прийти к выводу, что здесь наблюдается три кластера. Рассмотрим более подробно какие объекты принадлежат каждому из кластеров:

К первому кластеру относятся Сахалинская область, Республика Саха (Якутия) Хабаровский край, Приморский край, Челябинская область, Самарская область, Пермский край, Мурманская область, Республика Коми, Кемеровская область, Красноярский край, Алтайский край, Ростовская область, Республика Башкортостан, Волгоградская, Новосибирская, Томская и Воронежская области.

Во второе объединение включены Амурская область, Республика Хакасия, Иркутская область, Калининградская область, Тверская область, Волгоградская область, Архангельская область, Новгородская область, Республика Карелия, Смоленская область, Республика Мордовия, Псковская, Костромская, Ульяновская, Астраханская, Ярославская, Калужская, Саратовская, Пензенская область, Тульская, Рязанская, Владимирская области.

В третий кластер вошли Республика Алтай, Республика Адыгея, Забайкальский край, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Бурятия, Курская область, Республика Калмыкия, Чувашская Республика, Республика Марий Эл, Ивановская область, Республика Северная Осетия-Алания, Липецкая, Курганская, Кировская, Орловская, Брянская, Оренбургская, Тамбовская, Белгородская области и Удмуртская Республика.

Результаты, полученные методами ближнего и дальнего соседа, приведены на рисунках *Приложения 4*. Причем, метод ближнего соседа не дал четкого разбиения данных на кластеры, а с помощью метода дальнего соседа мы получили всего 2 кластера. Очевидно, что более целесообразно использовать проанализированную выше дендрограмму метода Варда (рис.2.1.1).

***2.2. Метод k-средних.***

После того, как мы выяснили количество кластеров, можно использовать для дальнейшего анализа метод k-средних. Особенность метода k-средних состоит в том, что он строит k насколько можно различных кластеров. С помощью Евклидовой метрики и метода Варда мы получили разбиение на 3 объединения. Но я считаю, что более целесообразным будет разбить данные на 4 кластера.

Графическое изображение метода указано на рисунке 2.1.2, приведенном ниже. На нем представлены график средних значений каждого и 4х полученных кластеров.

Рис 2.2.1.



Рис 2.2.1. График средних для каждого кластера.

Определим принадлежность каждого из показателей к нужной группе и соответствующие расстояния до центра объединения. Соответствующие таблицы представлены в *Приложении* 5 в таблицах 1, 2, 3 и 4. Ниже представлена сводная таблица (табл.2.2.1) , в которой представлены все кластеры и расстояния.

Табл. 2.2.1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 | CASE\_NO | CLUSTER | DISTANC |
| Белгородская область | -0,8131352 | 0,004701 | -0,2307317 | -0,4534437 | 0,2923749 | 1 | 1 | 0,46 |
| Брянская область | -0,5913546 | -0,6507491 | -0,4627552 | -1,0226327 | -0,5713877 | 2 | 2 | 0,27 |
| Владимирская область | -0,1888711 | -0,8146117 | -0,2686726 | 0,2366258 | -0,7135984 | 3 | 2 | 0,51 |
| Воронежская область | 0,2590349 | 2,6265014 | 2,0493733 | -1,6559409 | -0,4880339 | 4 | 1 | 1,39 |
| Ивановская область | -0,5783203 | -1,1112028 | -0,6918601 | 0,5398779 | -1,108734 | 5 | 2 | 0,57 |
| Калужская область | 0,5040206 | 0,4618774 | -0,3066136 | 0,6423635 | -0,026406 | 6 | 3 | 0,29 |
| Костромская область | -0,2101012 | 0,6667056 | -0,6736193 | 0,4931024 | -0,7718189 | 7 | 3 | 0,57 |
| Курская область | -0,486389 | -1,4536755 | -0,5940892 | -0,4240119 | -0,1361659 | 8 | 2 | 0,37 |
| Липецкая область | -0,0039223 | 1,1206048 | -0,4284624 | -1,0073912 | 0,1383929 | 9 | 1 | 0,59 |
| Орловская область | -0,6052776 | -0,087062 | -0,5802262 | -0,5895655 | -0,6439247 | 10 | 2 | 0,44 |
| Рязанская область | -0,3097347 | -0,4819707 | -0,4160586 | 0,5377756 | -0,1679804 | 11 | 3 | 0,40 |
| Смоленская область | 0,2129211 | 1,4450526 | -0,5794965 | 1,2299473 | -0,1005337 | 12 | 3 | 0,63 |
| Тамбовская область | -0,7343369 | 0,5274224 | -0,4985072 | -0,4655317 | -0,1864327 | 13 | 1 | 0,45 |
| Тверская область | 1,6961655 | -0,694992 | -0,5532297 | 0,9734706 | -0,3127362 | 14 | 3 | 0,86 |
| Тульская область | -0,122021 | 0,3094853 | -0,1687128 | -0,3199497 | -0,1431651 | 15 | 1 | 0,29 |
| Ярославская область | 0,1330367 | 0,8059887 | -0,1169088 | 0,7874199 | 0,2379722 | 16 | 3 | 0,32 |
| Республика Карелия | -0,5992542 | 1,1074958 | -0,3927103 | 0,8615249 | 0,1240764 | 17 | 3 | 0,53 |
| Республика Коми | 0,9897458 | 1,9710513 | -0,2971283 | 1,0365387 | 2,162429 | 18 | 4 | 0,92 |
| Архангельская область | -0,4832291 | 0,5356156 | -0,3285026 | 0,9356298 | 0,9496617 | 19 | 3 | 0,46 |
| Вологодская область | -1,2570915 | 0,109573 | -0,3700917 | 1,1821207 | 0,1129413 | 20 | 3 | 0,60 |
| Калининградская область | 0,9657509 | -0,9129292 | -0,4467032 | 1,4874751 | 0,3445506 | 21 | 3 | 0,67 |
| Мурманская область | 0,0409078 | 0,9157766 | -0,2613763 | 2,5853845 | 2,2060148 | 22 | 3 | 1,22 |
| Новгородская область | -0,7687001 | 1,0009851 | -0,445244 | 1,305629 | -0,0614019 | 23 | 3 | 0,57 |
| Псковская область | -0,7074784 | 0,6404876 | -0,7108306 | 0,5598494 | -0,4924879 | 24 | 3 | 0,57 |
| Республика Адыгея | -1,989876 | -1,5011956 | -0,737827 | -3,104929 | -1,2258112 | 25 | 2 | 1,26 |
| Кабардино-Балкарская Республика | -0,7031336 | -1,5536316 | -0,7254233 | -1,7946906 | -1,03397 | 26 | 2 | 0,56 |
| Республика Калмыкия | -1,4852905 | -0,5737337 | -0,8822945 | -0,2269243 | -1,9686789 | 27 | 2 | 0,71 |
| Карачаево-Черкесская Республика | -0,6048826 | -1,1833023 | -0,8326794 | -0,4996936 | -1,0062914 | 28 | 2 | 0,17 |
| Республика Северная Осетия - Алания | 0,5064892 | 0,5225066 | -0,7268825 | -1,4220636 | -0,6366074 | 29 | 1 | 0,75 |
| Астраханская область | 0,3634081 | -0,1313049 | -0,1935204 | 1,0086836 | -0,228746 | 30 | 3 | 0,26 |
| Волгоградская область | -0,522727 | -0,1706319 | 2,167574 | -0,0072373 | -0,3095547 | 31 | 1 | 0,87 |
| Ростовская область | 0,0316258 | -0,642556 | 0,8753929 | -1,040502 | 0,1024426 | 32 | 1 | 0,57 |
| Республика Башкортостан | -0,2358736 | 0,1259592 | 1,5612484 | -0,0797656 | 0,7680011 | 33 | 1 | 0,74 |
| Республика Марий Эл | -0,9572038 | -0,563902 | -0,6816452 | 0,0805324 | -1,2713059 | 34 | 2 | 0,47 |
| Республика Мордовия | -0,406307 | 1,4253891 | -0,5794965 | -0,0550639 | -1,0991897 | 35 | 1 | 0,79 |
| Удмуртская Республика | -0,5852324 | -0,2918902 | 0,2807415 | -0,7209573 | -0,7183706 | 36 | 1 | 0,37 |
| Чувашская Республика | -1,1169727 | -0,6097835 | -0,341636 | 0,0085297 | -1,0323793 | 37 | 2 | 0,44 |
| Пермский край | 0,024615 | -0,0198784 | 0,327438 | 0,1945804 | 1,361659 | 38 | 3 | 0,69 |
| Кировская область | -0,5039655 | -0,0018535 | -0,4554588 | -0,7656304 | -0,5494358 | 39 | 1 | 0,44 |
| Оренбургская область | -0,9777427 | 0,345535 | -0,0614567 | -0,4345233 | -0,5265294 | 40 | 1 | 0,38 |
| Пензенская область | -0,2787288 | 0,0489439 | -0,2737801 | -0,1607029 | -0,5300289 | 41 | 1 | 0,39 |
| Самарская область | 0,8135852 | 0,4684319 | 0,9002004 | 1,4301883 | 1,2617616 | 42 | 4 | 0,67 |
| Саратовская область | -0,101877 | 0,3357033 | 0,4325053 | 0,0994528 | -0,8834876 | 43 | 1 | 0,42 |
| Ульяновская область | 0,4976022 | 0,2177223 | -0,4554588 | 0,3322789 | -0,6626953 | 44 | 3 | 0,51 |
| Курганская область | -0,6422081 | 0,0210872 | -0,5218555 | -0,6179461 | -0,2157021 | 45 | 1 | 0,45 |
| Челябинская область | 0,004866 | -0,4278961 | 0,9600304 | 1,1768651 | 0,7387318 | 46 | 3 | 0,67 |
| Республика Алтай | -1,7628619 | -0,8015027 | -0,7772272 | -1,5345349 | -0,5300289 | 47 | 2 | 0,63 |
| Республика Бурятия | -0,4625915 | -1,1833023 | -0,6728896 | -0,2847367 | -0,1717981 | 48 | 2 | 0,35 |
| Республика Хакасия | 0,7034849 | -1,0489351 | -0,2154094 | 0,5803466 | -0,3420055 | 49 | 3 | 0,67 |
| Алтайский край | -0,3866568 | 0,8535089 | 1,181839 | -0,9096358 | -0,6649223 | 50 | 1 | 0,49 |
| Забайкальский край | 0,2759203 | -2,0206398 | -0,7422048 | -1,4677879 | -0,2758314 | 51 | 2 | 0,75 |
| Красноярский край | 0,9006781 | -0,314831 | 2,4747497 | 0,4610428 | 1,1981327 | 52 | 4 | 0,95 |
| Иркутская область | 0,313542 | -1,5716565 | -0,1621461 | 1,0906721 | 0,3318248 | 53 | 3 | 0,78 |
| Кемеровская область | -0,0240662 | -1,4258189 | 2,2813968 | -0,3320377 | 0,827176 | 54 | 1 | 1,26 |
| Новосибирская область | 1,6268467 | 0,8567861 | 4,3973343 | 0,2087707 | 0,3178265 | 55 | 4 | 1,65 |
| Томская область | 0,8873476 | 2,339742 | 0,695903 | -0,3467536 | 0,5227116 | 56 | 4 | 0,92 |
| Республика Саха (Якутия) | 2,0954892 | 0,9567422 | -0,4802664 | -1,7920627 | 2,1958342 | 57 | 4 | 1,17 |
| Приморский край | 1,9378926 | 0,1915043 | 0,9541933 | 1,2236405 | 0,3082821 | 58 | 4 | 0,74 |
| Хабаровский край | 2,0348599 | -0,0362647 | 0,0771737 | 0,0915693 | 1,2299471 | 59 | 4 | 0,57 |
| Амурская область | 0,5284105 | -1,7240487 | -0,5641742 | -0,0119674 | 0,03086 | 60 | 2 | 0,79 |
| Сахалинская область | 3,8591686 | 1,0468666 | -0,7108306 | 0,1667253 | 4,0445726 | 61 | 4 | 1,67 |

Табл. 2.2.1 Сводная таблица метода k-средних.

Укажем какие регионы к какому кластеру относятся.

1й кластер: Белгородская область, Воронежская область, Липецкая область, Тамбовская область, Тульская область, Республика Северная Осетия – Алания, Волгоградская область, Ростовская область, Республика Башкортостан, Республика Мордовия, Удмуртская Республика, Кировская область, Оренбургская область, Пензенская область, Саратовская область, Курганская область, Алтайский край, Кемеровская область.

2й кластер: Брянская область, Владимирская область, Ивановская область, Курская область, Орловская область, Республика Адыгея, Кабардино-Балкарская Республика, Республика Калмыкия, Карачаево-Черкесская Республика, Республика Марий Эл, Чувашская Республика, Республика Алтай, Республика Бурятия, Забайкальский край, Амурская область.

3й кластер: Калужская область, Костромская область, Рязанская область, Смоленская область, Тверская область, Ярославская область, Республика Карелия, Архангельская область, Вологодская область, Калининградская область, Мурманская область, Новгородская область, Псковская область, Астраханская область, Пермский край, Ульяновская область, Челябинская область, Республика Хакасия , Иркутская область.

4й кластер: Республика Коми, Самарская область, Красноярский край, Новосибирская область, Томская область, Республика Саха (Якутия), Приморский край, Хабаровский край, Сахалинская область.

Опираясь на рисунок 2.2.2 и данные, приведенные в сводной таблице k-метода (табл. 2.2.2.) проанализируем каждый из полученных кластеров.

Самыми высокими показателями среднедушевого дохода населения, число квартирных телефонных аппаратов сети общего пользования на 1000 человек, средства связи для оказания услуг передачи данных и телематически служб на 1000 человек, и что вполне логично, доходов от услуг связи населению на одного жителя обладает кластер 4.

В регионах группы 3 наблюдается самое большое число абонентов сотовых терминалов. Также в регионах данной группы наблюдается второй по величине среднедушевые доходы населению. Оставшиеся два показателя находится на среднем уровне.

Что же касается регионов, принадлежащих первой группе, то степень оснащенности населения данных областей и Республик услугами связи, можно охарактеризовать как средний уровень.

Самым же низким уровнем оказываемых услуг связи отличается четвертый кластер. Данному объединению принадлежит и самый низкий среднедушевой доход населения. То есть, люди проживающие на данных территориях, часто просто не могут позволить такую вроде бы простую вещь как мобильный телефон у каждого члена семьи. Совокупность влияния этих факторов также влечет за собой самый низкий из 4 групп доход от услуг связи населению.

**Глава 3. Регрессионный анализ.**

Регрессионный анализ – это статистический метод исследования зависимости случайной величины Y от переменных Xj (j = 1, 2, ..., k), рассматриваемых в регрессионном анализе как неслучайные величины независимо от истинного закона распределения Xj.[[2]](#footnote-2)

Важной предпосылкой построения регрессионной модели является оценка мультиколлинеарности. Как видно из парных коэффициентов корреляции (табл. 2.1.1), приведенной выше, в наших данных мультиколлинеарности не наблюдается.

Следующим шагом моей курсовой работы является проведение регрессионного анализа по каждому из кластеров. Для начала рассмотрим группу регионов, в которых наименее всего развито оказание услуг связи населению. Это группа под номером 2. Данные этого кластера приведены в таблице 3.1.

Табл. 3.1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 |
| Брянская область | -0,59135 | -0,65075 | -0,46276 | -1,02263 | -0,57139 |
| Владимирская область | -0,18887 | -0,81461 | -0,26867 | 0,236626 | -0,7136 |
| Ивановская область | -0,57832 | -1,1112 | -0,69186 | 0,539878 | -1,10873 |
| Курская область | -0,48639 | -1,45368 | -0,59409 | -0,42401 | -0,13617 |
| Орловская область | -0,60528 | -0,08706 | -0,58023 | -0,58957 | -0,64392 |
| Республика Адыгея | -1,98988 | -1,5012 | -0,73783 | -3,10493 | -1,22581 |
| Кабардино-Балкарская Республика | -0,70313 | -1,55363 | -0,72542 | -1,79469 | -1,03397 |
| Республика Калмыкия | -1,48529 | -0,57373 | -0,88229 | -0,22692 | -1,96868 |
| Карачаево-Черкесская Республика | -0,60488 | -1,1833 | -0,83268 | -0,49969 | -1,00629 |
| Республика Марий Эл | -0,9572 | -0,5639 | -0,68165 | 0,080532 | -1,27131 |
| Чувашская Республика | -1,11697 | -0,60978 | -0,34164 | 0,00853 | -1,03238 |
| Республика Алтай | -1,76286 | -0,8015 | -0,77723 | -1,53453 | -0,53003 |
| Республика Бурятия | -0,46259 | -1,1833 | -0,67289 | -0,28474 | -0,1718 |
| Забайкальский край | 0,27592 | -2,02064 | -0,7422 | -1,46779 | -0,27583 |
| Амурская область | 0,528411 | -1,72405 | -0,56417 | -0,01197 | 0,03086 |

Табл. 3.1. Группа регионов №2.

Где:

X1 – доходы от услуг связи населению в расчете на одного жителя (рублей);

Х2 – число квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (на конец года; штук);

Х3 – средства связи (пользовательское оборудование) для оказания услуг передачи данных и телематических служб на 1000 человек (на конец года;штук);

Х4 – число абонентских терминалов сотовой связи на 1000 человек населения (на конец года; штук);

Х5 – среднедушевые доходы населения (рублей).

Далее приведена корреляционная матрица для данных показателей (таблица 3.2.):

Табл 3.2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *X1* | *X2* | *X3* | *X4* | *X5* |
| X1 | 1 | -0,40043 | 0,324542 | 0,437464 | 0,640113 |
| X2 | -0,40043 | 1 | 0,256279 | 0,360102 | -0,41142 |
| X3 | 0,324542 | 0,256279 | 1 | 0,358174 | 0,321892 |
| X4 | 0,437464 | 0,360102 | 0,358174 | 1 | 0,024324 |
| X5 | 0,640113 | -0,41142 | 0,321892 | 0,024324 | 1 |

Табл. 3.2. Корреляционная матрица для группы регионов 2.

В качестве результативного признака для регрессионного анализа возьмём показатель X1 (доходы от услуг связи населению в расчете на одного жителя), факторными же признаками будут являться все остальные признаки. Данный выбор основан на том, что довольно интересно насколько доход от услуг связи населению в каждой группе зависит от оснащенности населения средствами связи и их среднедушевого дохода.

Теперь проделаем регрессионный анализ с исключением. Все результаты представлены рисунке (Рис.3.1).

Рис. 3.1.

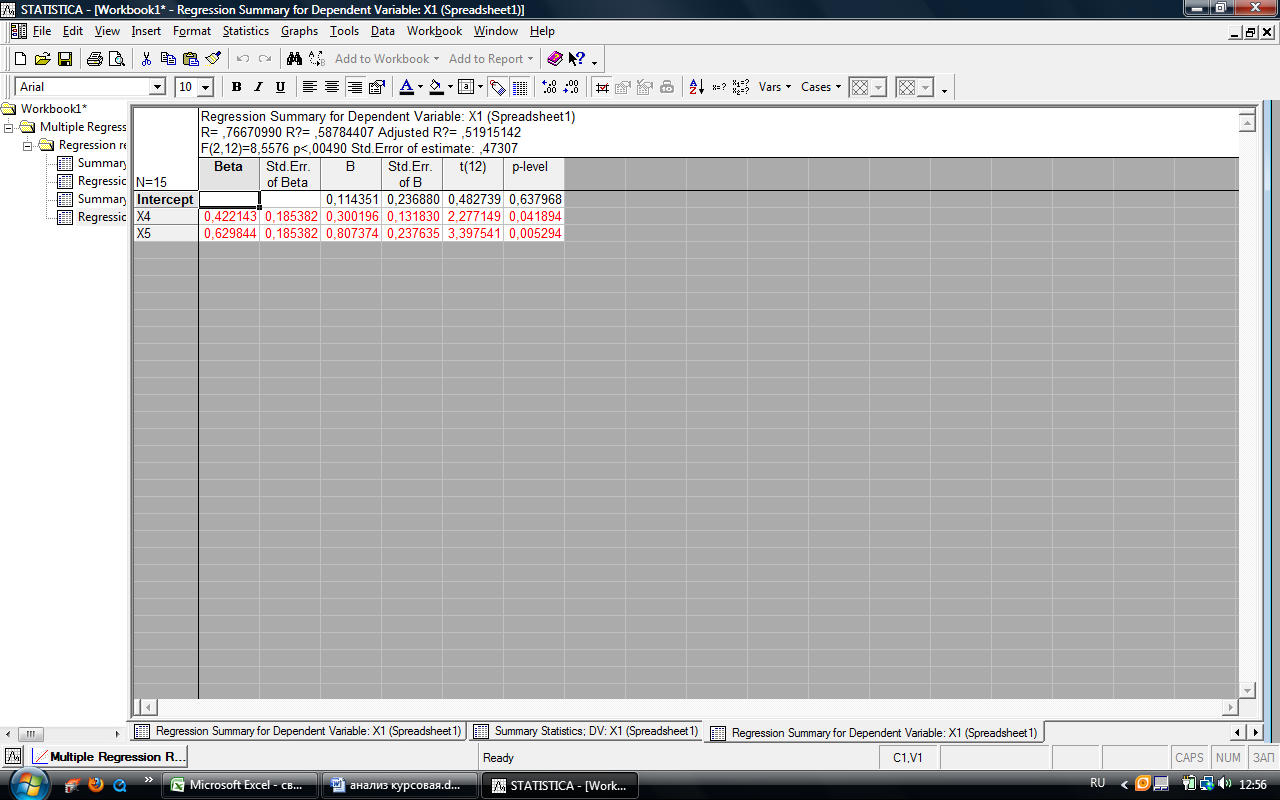


Рис. 3.1. Результаты регрессионного анализа для кластера 2.

Исходя из рисунка 3.1. можно построить следующее уравнение регрессии:

X1=0,114351+0,300196\*X4+0,807374\*X5

Необходимо проверить значимость уравнения регрессии. Для этого находим наблюдаемое значение статистики F. И получаем, что F=8,5576. Теперь найдем критическое значение статистики F на уровне значимости 0,1, оно равно 2,807. Так как наблюдаемое значение статистики F превосходит его критическое, то на уровне значимости 0,1 можно утверждать, что полученное уравнение регрессии значимое.

Далее необходимо проверить значимость коэффициентов уравнения. С вероятностью 0,1 можно утверждать, что коэффициенты при X4 и Х5 значимы. Коэффициент детерминации составил 58,8%. Следовательно, на долю вариации факторных признаков приходится большая часть по сравнению с остальными неучтенными в модели факторами, влияющими на изменение результативного показателя. А значит данная регрессионная модель имеет высокое практическое значение.

Увеличение числа абонентских терминалов сотовой связи в регионах страны с самым маленьким среднедушевым доходом на единицу ведет к увеличению дохода от услуг связи населению на одного жителя на 0,3002 единиц в этих регионах. Это обусловлено тем, что в современной ситуации от услуг связи населению, основная доля дохода приходится именно на доход от пользователей аппаратами сотовой связи. А увеличение среднедушевого дохода населения единицы своего измерения приводит к увеличению дохода от услуг связи населению на 0,8074 единицы.

Далее рассмотрим 3 кластер, в котором собраны регионы с самым высоким числом зарегистрированных сотовых терминалов на 1000 человек. Они представлены в таблице 3.3.

Табл. 3.3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 |
| Калужская область | 0,504021 | 0,461877 | -0,30661 | 0,642363 | -0,02641 |
| Костромская область | -0,2101 | 0,666706 | -0,67362 | 0,493102 | -0,77182 |
| Рязанская область | -0,30973 | -0,48197 | -0,41606 | 0,537776 | -0,16798 |
| Смоленская область | 0,212921 | 1,445053 | -0,5795 | 1,229947 | -0,10053 |
| Тверская область | 1,696166 | -0,69499 | -0,55323 | 0,973471 | -0,31274 |
| Ярославская область | 0,133037 | 0,805989 | -0,11691 | 0,78742 | 0,237972 |
| Республика Карелия | -0,59925 | 1,107496 | -0,39271 | 0,861525 | 0,124076 |
| Архангельская область | -0,48323 | 0,535616 | -0,3285 | 0,93563 | 0,949662 |
| Вологодская область | -1,25709 | 0,109573 | -0,37009 | 1,182121 | 0,112941 |
| Калининградская область | 0,965751 | -0,91293 | -0,4467 | 1,487475 | 0,344551 |
| Мурманская область | 0,040908 | 0,915777 | -0,26138 | 2,585385 | 2,206015 |
| Новгородская область | -0,7687 | 1,000985 | -0,44524 | 1,305629 | -0,0614 |
| Псковская область | -0,70748 | 0,640488 | -0,71083 | 0,559849 | -0,49249 |
| Астраханская область | 0,363408 | -0,1313 | -0,19352 | 1,008684 | -0,22875 |
| Пермский край | 0,024615 | -0,01988 | 0,327438 | 0,19458 | 1,361659 |
| Ульяновская область | 0,497602 | 0,217722 | -0,45546 | 0,332279 | -0,6627 |
| Челябинская область | 0,004866 | -0,4279 | 0,96003 | 1,176865 | 0,738732 |
| Республика Хакасия | 0,703485 | -1,04894 | -0,21541 | 0,580347 | -0,34201 |
| Иркутская область | 0,313542 | -1,57166 | -0,16215 | 1,090672 | 0,331825 |

Табл. 3.3. Группа регионов №3.

Где:

X1 – доходы от услуг связи населению в расчете на одного жителя (рублей);

Х2 – число квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (на конец года; штук);

Х3 – средства связи (пользовательское оборудование) для оказания услуг передачи данных и телематических служб на 1000 человек (на конец года; штук);

Х4 – число абонентских терминалов сотовой связи на 1000 человек населения (на конец года; штук);

Х5 – среднедушевые доходы населения (рублей).

Далее приведена корреляционная матрица для данных показателей (Табл. 3.4.):

Табл. 3.4.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *X1* | *X2* | *X3* | *X4* | *X5* |
| X1 | 1 | -0,49087 | 0,027889 | -0,00173 | -0,0982 |
| X2 | -0,49087 | 1 | -0,28401 | 0,127065 | 0,06861 |
| X3 | 0,027889 | -0,28401 | 1 | 0,02649 | 0,498841 |
| X4 | -0,00173 | 0,127065 | 0,02649 | 1 | 0,583459 |
| X5 | -0,0982 | 0,06861 | 0,498841 | 0,583459 | 1 |

Табл.3.4. Корреляционная матрица для группы регионов 3.

В качестве результативного признака для регрессионного анализа опять возьмём показатель X1 (доходы от услуг связи населению в расчете на одного жителя), факторными же признаками будут являться все остальные признаки. Все результаты представлены в таблице (Рис 3.2).

Рис. 3.2.

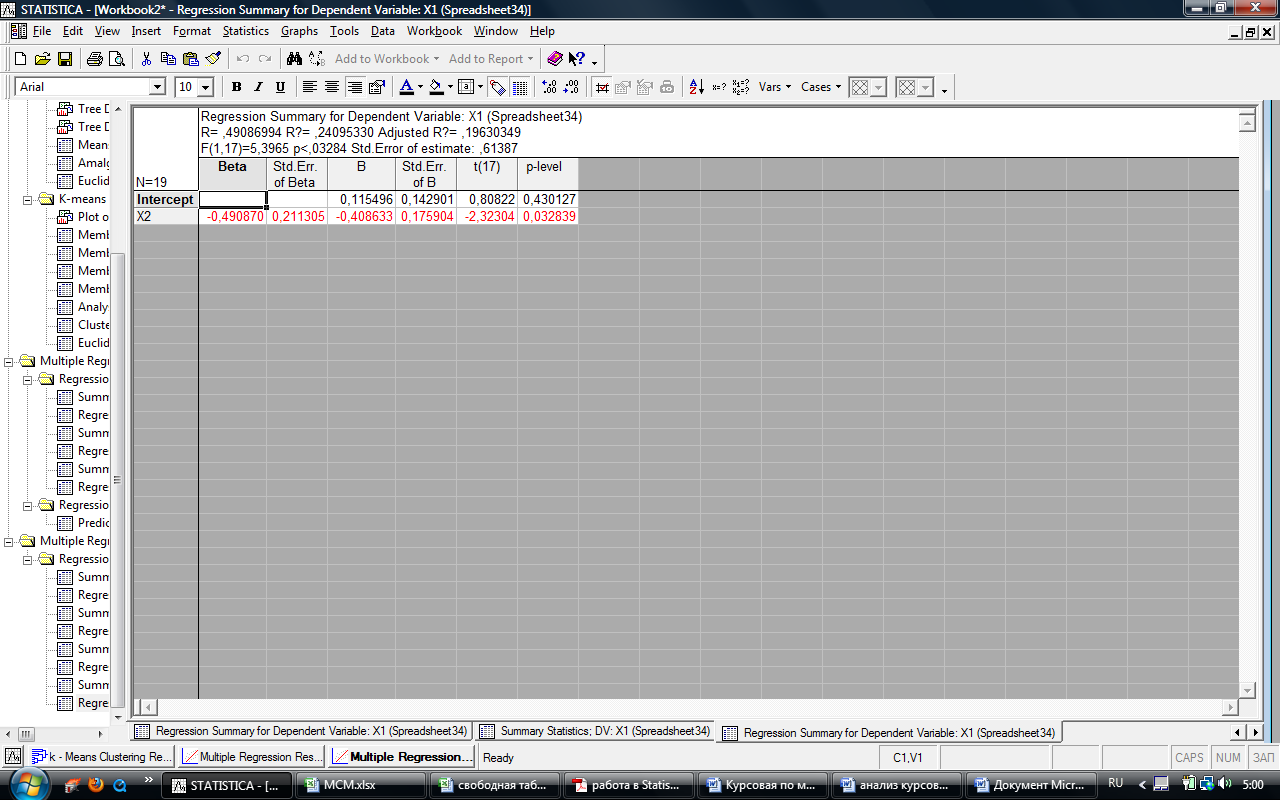


Рис. 3.2. Результаты регрессионного анализа для кластера 3.

Исходя из таблицы (Рис. 3.2) можно построить следующее уравнение регрессии:

X1=0,115496-0,408633\*X2

Необходимо проверить значимость уравнения регрессии. Для этого находим наблюдаемое значение статистики F. И получаем, что F=5,3965. Теперь найдем критическое значение статистики F на уровне значимости 0,1, оно равно 3,026. Так как наблюдаемое значение статистики F превосходит его критическое, то на уровне значимости 0,1 можно утверждать, что полученное уравнение регрессии значимое.

Далее необходимо проверить значимость коэффициентов уравнения. С вероятностью 0,1 можно утверждать, что коэффициент при X2 значим. Коэффициент детерминации составил 24,1%.  Следовательно, на долю вариации факторных признаков приходится меньшая часть по сравнению с остальными неучтенными в модели факторами, влияющими на изменение результативного показателя. А значит данная регрессионная модель имеет низкое практическое значение.

Из уравнения регрессии можно сделать следующий вывод. При увеличении числа квартирных телефонных аппаратов на единицу своего измерения, доход от услуг связи населению уменьшается на 0, 4086 единиц.

Далее рассмотрим 4 кластер, в котором собраны регионы с самым высоким уровнем среднедушевого дохода. Они представлены в таблице 3.5.

Табл.3.5.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 |
| Республика Коми | 0,989746 | 1,971051 | -0,29713 | 1,036539 | 2,162429 |
| Новосибирская область | 1,626847 | 0,856786 | 4,397334 | 0,208771 | 0,317826 |
| Томская область | 0,887348 | 2,339742 | 0,695903 | -0,34675 | 0,522712 |
| Республика Саха (Якутия) | 2,095489 | 0,956742 | -0,48027 | -1,79206 | 2,195834 |
| Приморский край | 1,937893 | 0,191504 | 0,954193 | 1,223641 | 0,308282 |
| Хабаровский край | 2,03486 | -0,03626 | 0,077174 | 0,091569 | 1,229947 |
| Сахалинская область | 3,859169 | 1,046867 | -0,71083 | 0,166725 | 4,044573 |

Табл. 3.5. Группа регионов №4.

Где:

X1 – доходы от услуг связи населению в расчете на одного жителя (рублей);

Х2 – число квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (на конец года; штук);

Х3 – средства связи (пользовательское оборудование) для оказания услуг передачи данных и телематических служб на 1000 человек (на конец года;штук);

Х4 – число абонентских терминалов сотовой связи на 1000 человек населения (на конец года; штук);

Х5 – среднедушевые доходы населения (рублей).

Далее приведена корреляционная матрица для данных показателей (Табл. 3.6.):

Табл. 3.6.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *X1* | *X2* | *X3* | *X4* | *X5* |
| X1 | 1 | -0,44955 | -0,30478 | -0,1064 | 0,710283 |
| X2 | -0,44955 | 1 | -0,11721 | -0,0998 | 0,116032 |
| X3 | -0,30478 | -0,11721 | 1 | 0,18099 | -0,66355 |
| X4 | -0,1064 | -0,0998 | 0,18099 | 1 | -0,18994 |
| X5 | 0,710283 | 0,116032 | -0,66355 | -0,18994 | 1 |

Табл. 3.6. Корреляционная матрица для группы 4.

В качестве результативного признака для регрессионного анализа опять возьмём показатель X1 (доходы от услуг связи населению в расчете на одного жителя), факторными же признаками будут являться все остальные признаки. Все результаты представлены таблице (Рис. 3.3).

Рис. 3.3.

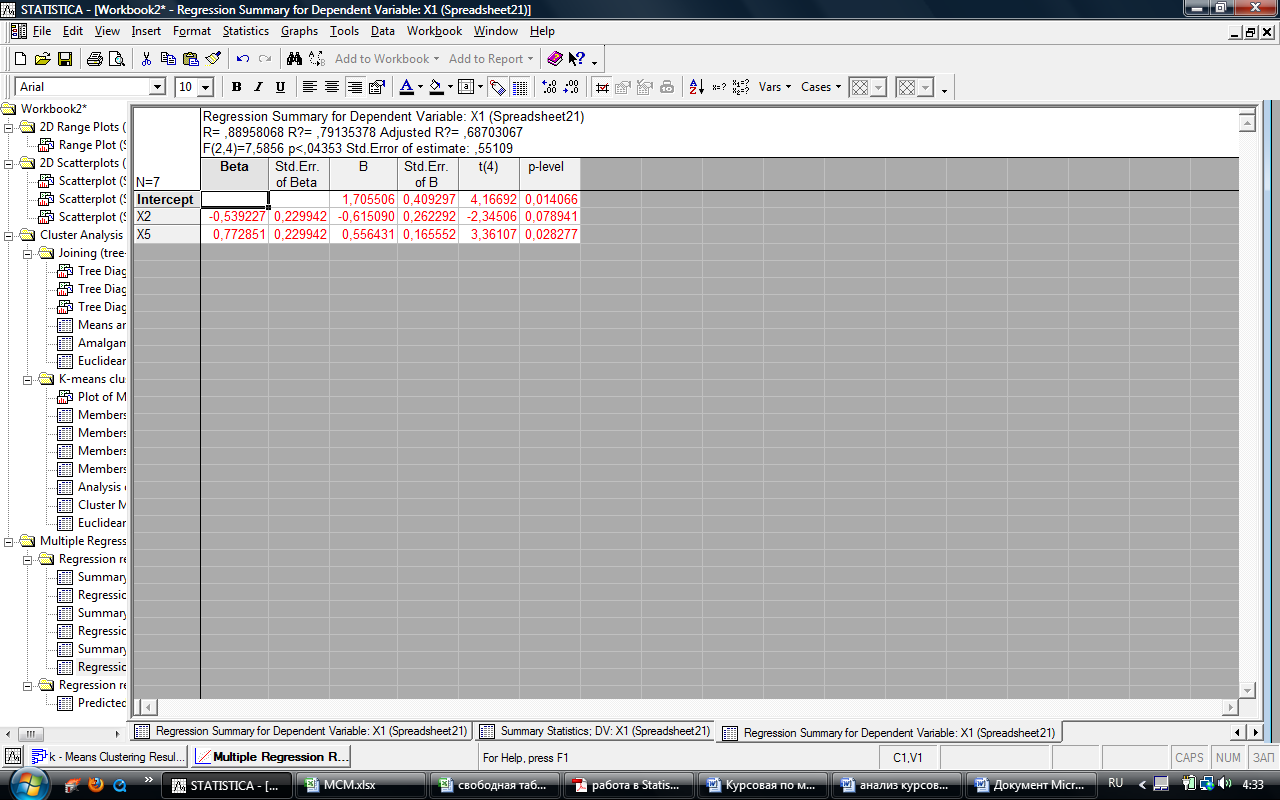


Рис.3.3. Результаты регрессионного анализа для группы 4.

Исходя из таблицы (рис. 3.3) можно построить следующее уравнение регрессии:

X1=1,705506-0,615090\*X2+0,556431\*X5

Необходимо проверить значимость уравнения регрессии. Для этого находим наблюдаемое значение статистики F. И получаем, что F=7,5856. Теперь найдем критическое значение статистики F на уровне значимости 0,1, оно равно 4,325. Так как наблюдаемое значение статистики F превосходит его критическое, то на уровне значимости 0,1 можно утверждать, что полученное уравнение регрессии значимое.

Далее необходимо проверить значимость коэффициентов уравнения. С вероятностью 0,1 можно утверждать, что коэффициенты при X2 и Х5 значимы. Коэффициент детерминации составил 79,1%. Следовательно, на долю вариации факторных признаков приходится большая часть по сравнению с остальными неучтенными в модели факторами, влияющими на изменение результативного показателя. А значит данная регрессионная модель имеет высокое практическое значение.

Увеличение числа квартирных телефонных аппаратов в Дальневосточном округе страны на единицу ведет к уменьшению дохода от услуг связи населению на одного жителя на 0,6151 единиц в этих регионах. Скорее всего это обусловлено тем, что чем больше квартирных телефонных аппаратов у населения, тем меньше они заинтересованы в использовании более современных и дорогостоящих средствах связи, что и вызывает уменьшение дохода. А увеличение среднедушевого дохода населения единицы своего измерения приводит к увеличению дохода от услуг связи населению на 0,5564 единицы.

И, наконец, последней будет рассмотрена группа регионов под номером 1. К ней отнесены регионы с примерно усреднёнными показателями оказываемых услуг связи и среднедушевым доходом населения. Они представлены в таблице 3.7.

Табл.3.7.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 |
| Белгородская область | -0,81314 | 0,004701 | -0,23073 | -0,45344 | 0,292375 |
| Воронежская область | 0,259035 | 2,626501 | 2,049373 | -1,65594 | -0,48803 |
| Липецкая область | -0,00392 | 1,120605 | -0,42846 | -1,00739 | 0,138393 |
| Тамбовская область | -0,73434 | 0,527422 | -0,49851 | -0,46553 | -0,18643 |
| Тульская область | -0,12202 | 0,309485 | -0,16871 | -0,31995 | -0,14317 |
| Республика Северная Осетия - Алания | 0,506489 | 0,522507 | -0,72688 | -1,42206 | -0,63661 |
| Волгоградская область | -0,52273 | -0,17063 | 2,167574 | -0,00724 | -0,30955 |
| Ростовская область | 0,031626 | -0,64256 | 0,875393 | -1,0405 | 0,102443 |
| Республика Башкортостан | -0,23587 | 0,125959 | 1,561248 | -0,07977 | 0,768001 |
| Республика Мордовия | -0,40631 | 1,425389 | -0,5795 | -0,05506 | -1,09919 |
| Удмуртская Республика | -0,58523 | -0,29189 | 0,280741 | -0,72096 | -0,71837 |
| Кировская область | -0,50397 | -0,00185 | -0,45546 | -0,76563 | -0,54944 |
| Оренбургская область | -0,97774 | 0,345535 | -0,06146 | -0,43452 | -0,52653 |
| Пензенская область | -0,27873 | 0,048944 | -0,27378 | -0,1607 | -0,53003 |
| Саратовская область | -0,10188 | 0,335703 | 0,432505 | 0,099453 | -0,88349 |
| Курганская область | -0,64221 | 0,021087 | -0,52186 | -0,61795 | -0,2157 |
| Алтайский край | -0,38666 | 0,853509 | 1,181839 | -0,90964 | -0,66492 |
| Кемеровская область | -0,02407 | -1,42582 | 2,281397 | -0,33204 | 0,827176 |

Табл. 3.7. Группа регионов 1.

Где:

X1 – доходы от услуг связи населению в расчете на одного жителя (рублей);

Х2 – число квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (на конец года; штук);

Х3 – средства связи (пользовательское оборудование) для оказания услуг передачи данных и телематических служб на 1000 человек (на конец года;штук);

Х4 – число абонентских терминалов сотовой связи на 1000 человек населения (на конец года; штук);

Х5 – среднедушевые доходы населения (рублей).

Далее приведена корреляционная матрица для данных показателей (Табл. 3.8.):

Табл. 3.8.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 |
| X1 | 1 | 0,215458 | 0,235416 | -0,48011 | 0,068809 |
| X2 | 0,215458 | 1 | -0,11719 | -0,40711 | -0,46561 |
| X3 | 0,235416 | -0,11719 | 1 | 0,017502 | 0,387714 |
| X4 | -0,48011 | -0,40711 | 0,017502 | 1 | 0,07353 |
| X5 | 0,068809 | -0,46561 | 0,387714 | 0,07353 | 1 |

Табл. 3.8. Корреляционная матрица для группы 1.

В качестве результативного признака для регрессионного анализа опять возьмём показатель X1 (доходы от услуг связи населению в расчете на одного жителя), факторными же признаками будут являться все остальные признаки. Все результаты представлены таблице (Рис. 3.4).

Рис. 3.4.

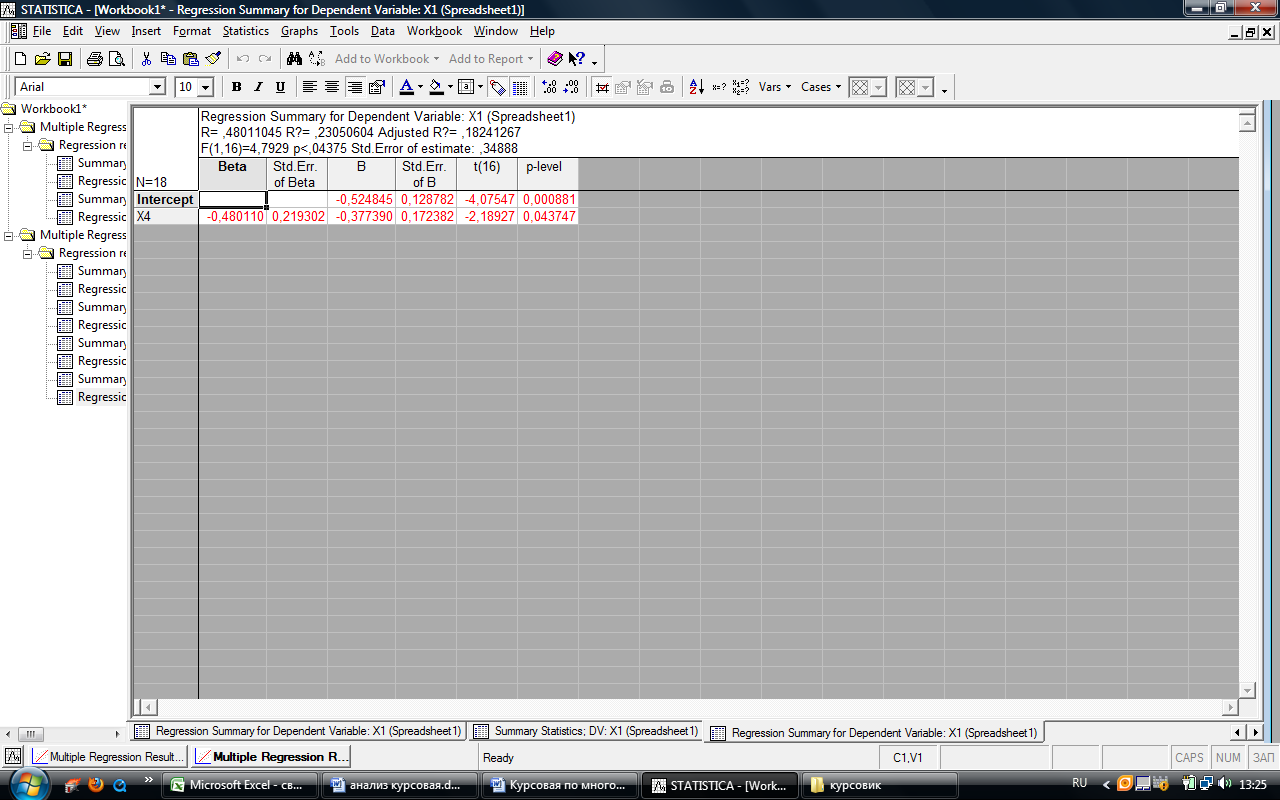


Рис.3.4. Результаты регрессионного анализа для группы 1.

Исходя из таблицы (Рис. 3.4) можно построить следующее уравнение регрессии:

X1=-0,524845-0,377390\*X4

Необходимо проверить значимость уравнения регрессии. Для этого находим наблюдаемое значение статистики F. И получаем, что F=4,7929. Теперь найдем критическое значение статистики F на уровне значимости 0,1, оно равно 3,136. Так как наблюдаемое значение статистики F превосходит его критическое, то на уровне значимости 0,1 можно утверждать, что полученное уравнение регрессии значимое.

Далее необходимо проверить значимость коэффициентов уравнения. С вероятностью 0,1 можно утверждать, что коэффициент при X4 значим. Коэффициент детерминации составил 23,1%. Следовательно, на долю вариации факторных признаков приходится меньшая часть по сравнению с остальными неучтенными в модели факторами, влияющими на изменение результативного показателя. А значит данная регрессионная модель имеет низкое практическое значение.

Увеличение числа абонентских терминалов сотовой связи в рассматриваемых регионах страны на единицу ведет к уменьшению дохода от услуг связи населению на одного жителя на 0,3774 единиц в этих регионах.

Что же касательно общего уравнения регрессии по всей совокупности данных то оно будет выглядеть так (рис. 3.5):

Рис. 3.5

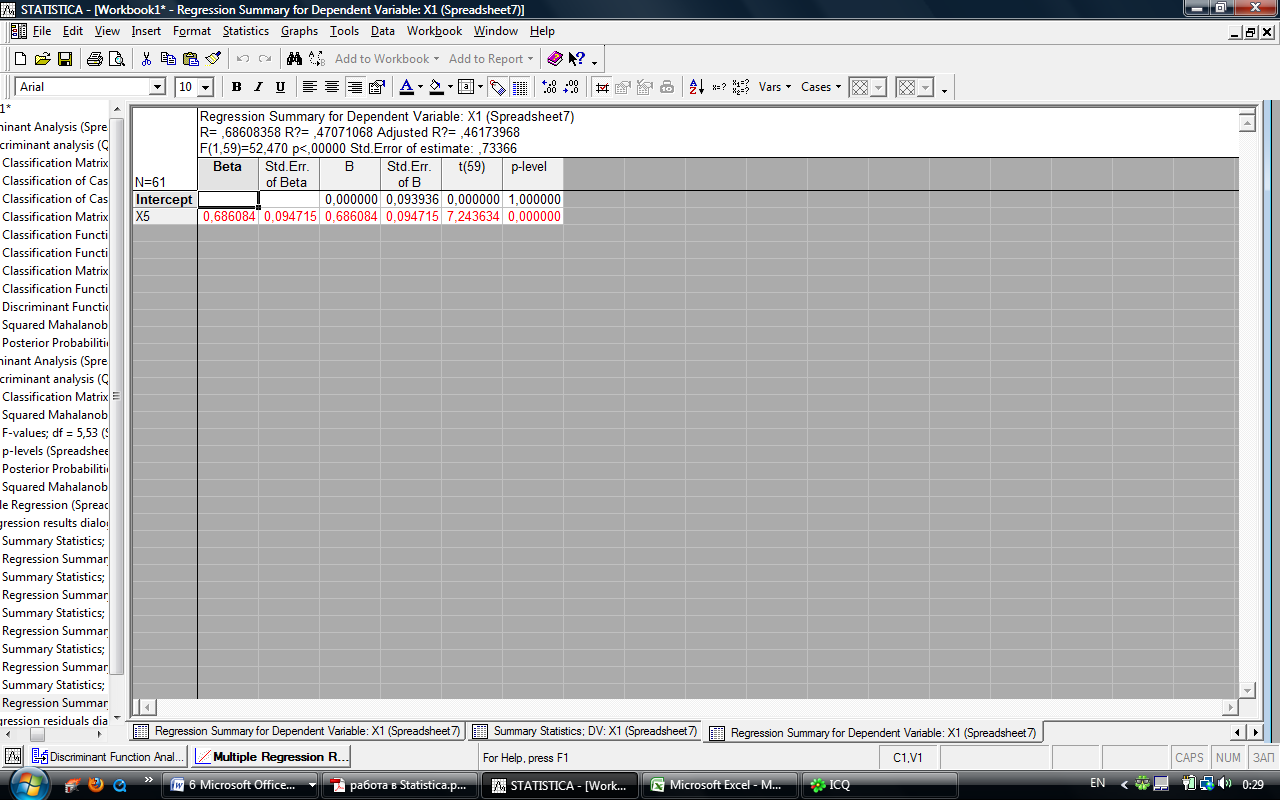


Рис. 3.5 Результаты регрессионного анализа для всей совокупности регионов.

X1=0,686084\*X5

Необходимо проверить значимость уравнения регрессии. Для этого находим наблюдаемое значение статистики F. И получаем, что F=52,470. Так как наблюдаемое значение статистики F превосходит его критическое, то на уровне значимости 0,1 можно утверждать, что полученное уравнение регрессии значимое.

Далее необходимо проверить значимость коэффициентов уравнения. С вероятностью 0,1 можно утверждать, что коэффициент при X5 значим. Коэффициент детерминации составил 47,1%. Следовательно, на долю вариации факторных признаков приходится меньшая часть по сравнению с остальными неучтенными в модели факторами, влияющими на изменение результативного показателя. А значит данная регрессионная модель имеет низкое практическое значение.

Мы провели регрессионный анализ в каждом из кластеров, которые были получены в ходе кластерного анализа. В каждой из групп влияние на доход от услуг связи населению. Далее представлена сводная таблица (табл. 3.9).

Табл.3.9.

|  |  |  |  |
| --- | --- | --- | --- |
| Кластеры | Уравнение регрессии | R^2 | Fнабл |
| 1 | X1=-0,5248-0,3774\*X4 | 23,1 % | 4,7929 |
| 2 | X1=0,1144+0,3002\*X4+0,8074\*X5 | 58,8 % | 8,5576 |
| 3 | X1=0,1155-0,4086\*X2 | 24,1 % | 5,3965 |
| 4 | X1=1,7055-0,6151\*X2+0,5564\*X5 | 79,1 % | 7,5856 |

Табл. 3.9. Сводная таблица регрессионного анализа по кластерам.

Далее рассчитаем коэффициенты эластичности для каждого показатели в каждом кластере. Коэффициент эластичности рассчитывается по следующей формуле:



Коэффициент эластичности показывает влияние каждого из факторов регрессионный модели на зависимый признак.

Ниже представлена сводная таблица, в которой рассчитаны коэффициенты эластичности по каждому из кластеров (табл. 3.10). Расчет коэффициентов эластичности представлен в таблицах расчета коэффициента эластичности по кластерам *Приложения 6.*

Табл. 3.10

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | X2 | X3 | X4 | X5 |
| 1 | - | - | -0,03916 | - |
| 2 | - | - | |  | | --- | | 0,017657 | | 0,05484 |
| 3 | -0,05004 | - | - | - |
| 4 | -0,05971 | - | 0,004338 | - |

Табл. 3.10. Сводная таблица коэффициентов эластичности.

Проанализировав таблицу 3.10 можно сделать следующие выводы. Влияние факторов на доход от услуг связи населению, перечисленных в данных регрессионных моделях, невелик. Причем, число квартирных телефонных аппаратов влияет на результативный признак только в кластерах 3 и 4, то есть в регионах, где среднедушевой доход населения либо самый большой, либо усреднен. Как видно из сводной таблицы показатель средства связи (пользовательское оборудование) для оказания услуг передачи данных и телематических служб на 1000 человек( на конец года; штук) вовсе не влияет на исследуемый признак. Это обусловлено тем, что данная услуга связи устаревает и потеряла свою популярность в современном. Среднедушевой доход влияет на исследуемый признак только лишь во втором кластере, где сосредоточены регионы с его низким уровнем. Зато, число зарегистрированных абонентских терминалов сотовой связи влияет на доход от услуг связи населению в трёх кластерах: в первом, втором и четвертом. Причем, в первом кластере это единственный показатель, который влияет на результативный признак. Данная тенденция вызвана тем, что в современном обществе очень большую роль играет мобильная связь.

**Глава 4. Дискриминантный анализ.**

Дискриминантный анализ является разделом многомерного статистического анализа, который позволяет изучать различия между двумя и более группами объектов по нескольким переменным одновременно. Дискриминантный анализ – это общий термин, относящийся к нескольким тесно связанным статистическим процедурам. Эти процедуры можно разделить на методы *интерпретации межгрупповых различий* – *дискриминации* и методы *классификации наблюдений* по группам.[[3]](#footnote-3)

Цель дискриминантного анализа состоит в том, чтобы на основе некоторой «зависимой переменной» определить линейные классификационные модели, позволяющие «предсказать» поведение новых элементов (или исключенных элементов) на основании измерения ряда независимых факторов, которыми они характеризуются. Дискрминантный анализ используется как метод разведочного анализа.

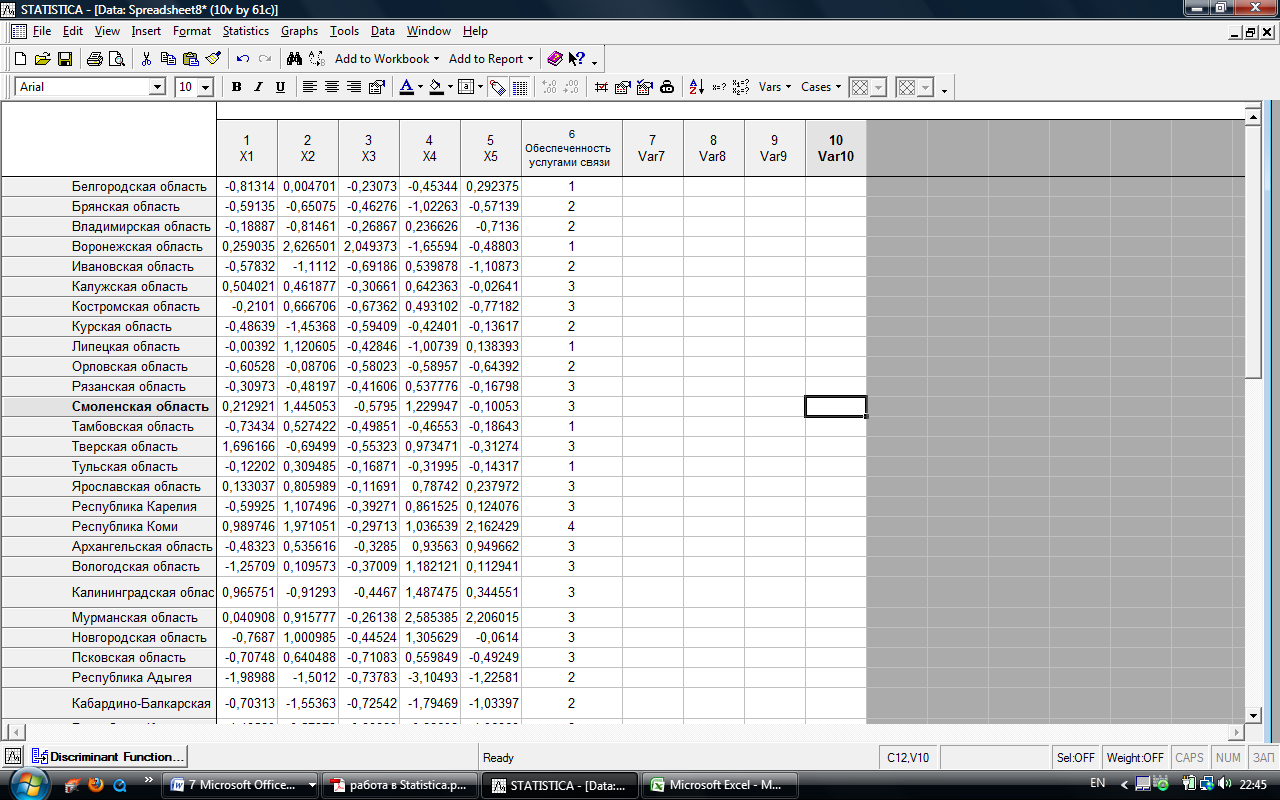


Рис 4.1. Исходные данные.

Проведем дискриминантный анализ на основе рейтинга регионов России по обеспеченности населения услугами связи и среднедушевому доходу. Исходный массив данных составляет 61 регион России (исходные данные приведены на рис. 4.1), обследованных по следующим пяти признакам:

X1 – доходы от услуг связи населению в расчете на одного жителя (рублей);

Х2 – число квартирных телефонных аппаратов сети общего пользования на 1000 человек населения (на конец года; штук);

Х3 – средства связи (пользовательское оборудование) для оказания услуг передачи данных и телематических служб на 1000 человек (на конец года;штук);

Х4 – число абонентских терминалов сотовой связи на 1000 человек населения (на конец года; штук);

Х5 – среднедушевые доходы населения (рублей).

Данные показатели в ходе анализа будут являться дискриминантными.

После проведенного кластерного анализа было выделено 4 группы регионов России:

* Кластер 1: Регионы со средним уровнем оказываемых услуг связи и среднедушевого дохода.
* Кластер 2: Регионы с низким уровнем оказываемых услуг связи и среднедушевого дохода.
* Кластер 3: Регионы высоким уровнем оказываемых услуг связи и срднедушевого дохода.
* Кластер 4: Регионы с самым высоким уровнем оказываемых услуг связи и среднедушевым доходом населения.

При этом к первому классу отнесено 18 регионов России, ко второму – 15 регионов, к третьему 19 регионов и к четвертому классу было отнесено 7 регионов России.

В качестве проверки корректности обучающих выборок посмотрим результаты классификационной матрицы (табл. 4.1).

Табл. 4.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Percent | G\_1:1 | G\_2:2 | G\_3:3 | G\_4:4 |
| cases | Correct | p=,29508 | p=,24590 | p=,31148 | p=,14754 |
| G\_1:1 | 100,0000 | 18 | 0 | 0 | 0 |
| G\_2:2 | 100,0000 | 0 | 15 | 0 | 0 |
| G\_3:3 | 100,0000 | 0 | 0 | 19 | 0 |
| G\_4:4 | 88,8889 | 0 | 0 | 1 | 8 |
| Total | 98,3607 | 18 | 15 | 20 | 8 |

Табл. 4.1. Классификационная матрица.

Как можно заметить из классификационной матрицы почти все объекты были распределены верно по кластерам, но как видно из рис. 4.2. не все объекты попали в в верную группу. Ошибочно распределенные объекты помечены знаком «\*».

Рис 4.2.

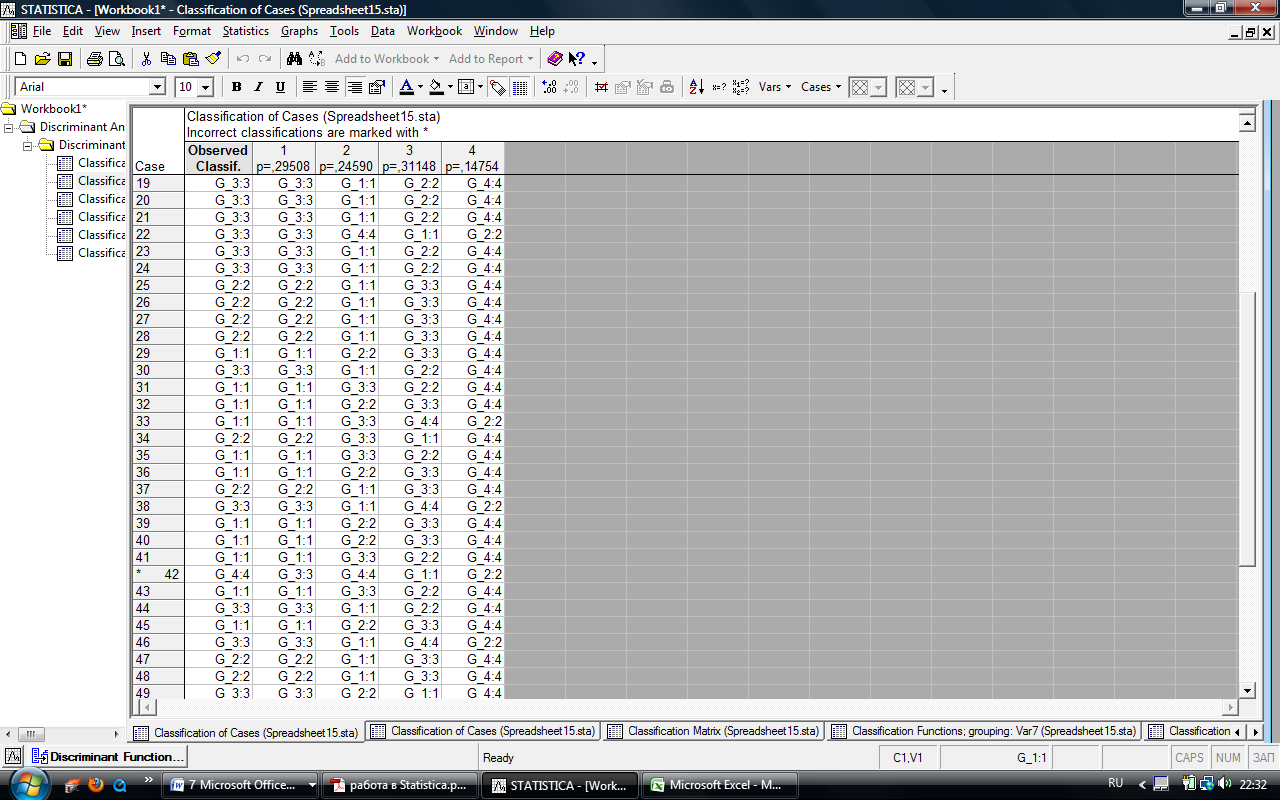


Рис 4.1. Классификация случаев.

На рисунке классификации случаев (рис 4.2) некорректно отнесённым объектом оказался регион под номером 42. Это Самарская область. Изначально, она была отнесена к самой оснащённой услугами связи группе регионов.

Таким образом, задача получения корректных обучающих выборок состоит в том, чтобы исключить из обучающих выборок те объекты, которые по своим показателям не соответствуют большинству объектов, образующих однородную группу.

Для этого с помощью метрики Махаланобиса определятся расстояние от всех *n* объектов до центра тяжести каждой группы (вектор средних), определяемых по обучающей выборке. Отнесение экспертом *i*-го объекта в *j*-ю группу считается ошибочным, если расстояние Махаланобиса от объекта до центра его группы значительно выше, чем от него до центра других групп, а апостериорная вероятность попадания в свою группу ниже критического значения. В этом случае объект считается некорректно отнесенным и должен быть исключен из выборки.[[4]](#footnote-4)

Процедура исключения объекта из обучающих выборок состоит в том, что в таблице исходных данных (рис 4.1) у объекта, который должен быть исключен из выборки (он помечен "\*"), убирается номер принадлежности к этой группе, после чего процесс тестирования повторяется.

В результате дальнейшего анализа, получаем следующие таблицы (табл 4.2.; рис 4.3). Из Табл. 4.2. видно, что дискриминантный анализ выполняется корректно. Такие выводы, я делаю исходя их того факта, что лямбда Уилкса стремится к нулю, а также F-критерий значим.

Табл 4.2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Discriminant Function Analysis Summary (Spreadsheet15.sta) | | | | | |  |
| No. of vars in model: 5; Grouping: обеспеченность услугами связи (4 grps) | | | | | | |
| Wilks' Lambda: ,08337 approx. F (15,143)=14,007 p<0,0000 | | | | | |  |
|  | Wilks' | Partial | F-remove | p-level | Toler. | 1-Toler. |
|  | Lambda | Lambda |  |  |  | (R-Sqr.) |
| X1 | 0,125544 | 0,664058 | 8,76882 | 0,000084 | 0,845137 | 0,154863 |
| X2 | 0,138232 | 0,603108 | 11,40667 | 0,000007 | 0,908463 | 0,091537 |
| X3 | 0,113222 | 0,736330 | 6,20683 | 0,001100 | 0,938229 | 0,061771 |
| X4 | 0,157078 | 0,530746 | 15,32512 | 0,000000 | 0,975267 | 0,024733 |
| X5 | 0,093371 | 0,892876 | 2,07960 | 0,114184 | 0,879408 | 0,120592 |
|  |  |  |  |  |  |  |

Табл 4.2. Результат дискриминационного анализа.

Рис. 4.3.

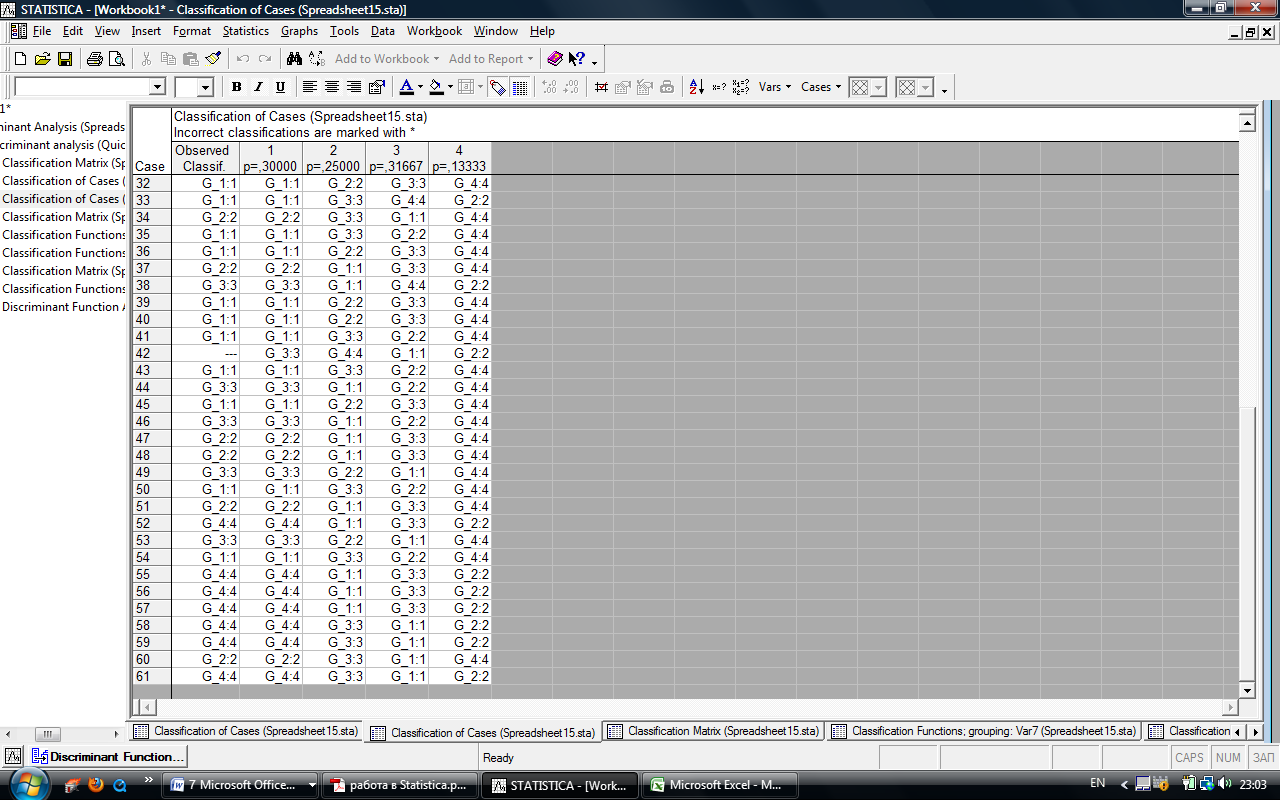


Рис. 4.3. Классификация случаев. (после корректировки исходных данных)

Для того, чтобы определить, к какой же в итоге группе отнести оставшиёся регион, воспользуемся классификационными функциями (Рис. 4.4.):

Рис. 4.4

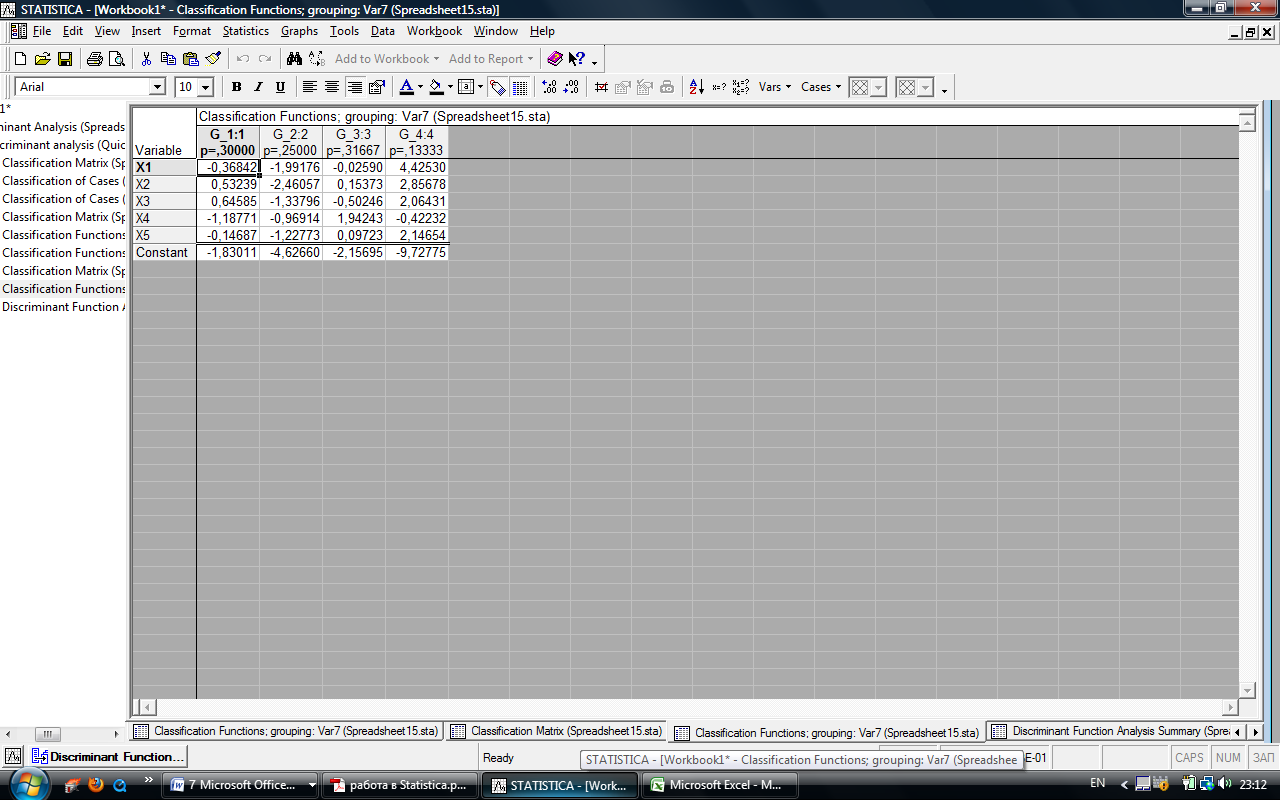


Рис. 4.4. Классификационные функции.

Низкая обеспеченность услугами связи (2):

Y=-4,6266-1,99176\*X1-2,46057\*X2-1,33796\*X3-0,96914\*X4-1,22773\*X5

Средняя обеспеченность услугами связи (1):

Y=-1,83011-0,36842\*X1+0,53239\*X2-0,64585\*X3-1,18771\*X4-0,14687\*X5

Высокая обеспеченность услугами связи (3):

Y=-2,15695-0,0259\*X1+0,15373\*X2-0,50246\*X3+1,94243\*X4+0,09723\*X5

Самая высокая обеспеченность услугами связи (4):

Y=-9,72775+4,42530\*X1+2,85678\*X2+2,06431\*X3-0,42232\*X4+2,13654\*X5

Подставив, соответствующие значения в эти функции получим, что Самарская область изначально была определена у нас в группу с самым высокой степенью обеспеченностью услугами связи и среднедушевым доходом, тогда как анализ классификационных функций показал, что на самом деле по рассматриваемым показателям ее следует отнести к субъектам РФ с высоким уровнем обеспеченности услугами связи и среднедушевым доходом.

Наши выводы подтверждают таблицы (*Приложение 6*). Расстояния квадратов Махаланобиса и Апостериорные вероятности классификации.

Расстояние Махалонобиса у рассматриваемого субъекта до третьей группы действительно минимально, а апостериорная вероятность принадлежности объекта к второй группе максимальна.

**Заключение.**

Влияние таких факторов, как число квартирных телефонных аппаратов на 1000 человек, число абонентских терминалов сотовой связи, средства связи для оказания услуг передачи данных и телематических служб на 1000 человек в совокупности со среднедушевым доходом населения оказывают различное влияние на доход от услуг связи населению в различных группах регионов России.

Так же было замечено, что в тех регионах, где среднедушевой доход выше, увеличивается и степень обеспеченности населения услугами связи. В особенности средствами мобильной связи.

C помощью методов кластерного и дискриминантного анализа регионы России были разбиты на четыре кластера: с очень высокой степенью обеспеченности услугами связи, с высокой степенью обеспеченности услугами связи, со средней и соответственно низкой степенями обеспеченности услугами связи.

**Список литературы:**

1. Кошелева В.А.«Анализ методов автоматического извлечения знаний из реляционных баз данных»
2. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы и основы эконометрики. / Учебное пособие./ Московский государственный университет экономики, статистики и информатики. М.: МЭСИ, 2002г.
3. Буреева Н.Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. Учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007
4. www.gks.ru

Приложение 1.

Табл. 1. Исходные данные.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **X1** | **X2** | **X3** | **X4** | **X5** |
| Белгородская область | 2450,5 | 396,1 | 95,6 | 1211,9 | 12758 |
| Брянская область | 2675,1 | 356,1 | 63,8 | 1103,6 | 10043 |
| Владимирская область | 3082,7 | 346,1 | 90,4 | 1343,2 | 9596 |
| Воронежская область | 3536,3 | 556,1 | 408,1 | 983,1 | 10305 |
| Ивановская область | 2688,3 | 328 | 32,4 | 1400,9 | 8354 |
| Калужская область | 3784,4 | 424 | 85,2 | 1420,4 | 11756 |
| Костромская область | 3061,2 | 436,5 | 34,9 | 1392 | 9413 |
| Курская область | 2781,4 | 307,1 | 45,8 | 1217,5 | 11411 |
| Липецкая область | 3270 | 464,2 | 68,5 | 1106,5 | 12274 |
| Орловская область | 2661 | 390,5 | 47,7 | 1186 | 9815 |
| Рязанская область | 2960,3 | 366,4 | 70,2 | 1400,5 | 11311 |
| Смоленская область | 3489,6 | 484 | 47,8 | 1532,2 | 11523 |
| Тамбовская область | 2530,3 | 428 | 58,9 | 1209,6 | 11253 |
| Тверская область | 4991,7 | 353,4 | 51,4 | 1483,4 | 10856 |
| Тульская область | 3150,4 | 414,7 | 104,1 | 1237,3 | 11389 |
| Ярославская область | 3408,7 | 445 | 111,2 | 1448 | 12587 |
| Республика Карелия | 2667,1 | 463,4 | 73,4 | 1462,1 | 12229 |
| Республика Коми | 4276,3 | 516,1 | 86,5 | 1495,4 | 18636 |
| Архангельская область | 2784,6 | 428,5 | 82,2 | 1476,2 | 14824 |
| Вологодская область | 2000,9 | 402,5 | 76,5 | 1523,1 | 12194 |
| Калининградская область | 4252 | 340,1 | 66 | 1581,2 | 12922 |
| Мурманская область | 3315,4 | 451,7 | 91,4 | 1790,1 | 18773 |
| Новгородская область | 2495,5 | 456,9 | 66,2 | 1546,6 | 11646 |
| Псковская область | 2557,5 | 434,9 | 29,8 | 1404,7 | 10291 |
| Республика Адыгея | 1258,8 | 304,2 | 26,1 | 707,4 | 7986 |
| Кабардино-Балкарская Республика | 2561,9 | 301 | 27,8 | 956,7 | 8589 |
| Республика Калмыкия | 1769,8 | 360,8 | 6,3 | 1255 | 5651 |
| Карачаево-Черкесская Республика | 2661,4 | 323,6 | 13,1 | 1203,1 | 8676 |
| Республика Северная Осетия - Алания | 3786,9 | 427,7 | 27,6 | 1027,6 | 9838 |
| Астраханская область | 3642 | 387,8 | 100,7 | 1490,1 | 11120 |
| Волгоградская область | 2744,6 | 385,4 | 424,3 | 1296,8 | 10866 |
| Ростовская область | 3306 | 356,6 | 247,2 | 1100,2 | 12161 |
| Республика Башкортостан | 3035,1 | 403,5 | 341,2 | 1283 | 14253 |
| Республика Марий Эл | 2304,6 | 361,4 | 33,8 | 1313,5 | 7843 |
| Республика Мордовия | 2862,5 | 482,8 | 47,8 | 1287,7 | 8384 |
| Удмуртская Республика | 2681,3 | 378 | 165,7 | 1161 | 9581 |
| Чувашская Республика | 2142,8 | 358,6 | 80,4 | 1299,8 | 8594 |
| Пермский край | 3298,9 | 394,6 | 172,1 | 1335,2 | 16119 |
| Кировская область | 2763,6 | 395,7 | 64,8 | 1152,5 | 10112 |
| Оренбургская область | 2283,8 | 416,9 | 118,8 | 1215,5 | 10184 |
| Пензенская область | 2991,7 | 398,8 | 89,7 | 1267,6 | 10173 |
| Самарская область | 4097,9 | 424,4 | 250,6 | 1570,3 | 15805 |
| Саратовская область | 3170,8 | 416,3 | 186,5 | 1317,1 | 9062 |
| Ульяновская область | 3777,9 | 409,1 | 64,8 | 1361,4 | 9756 |
| Курганская область | 2623,6 | 397,1 | 55,7 | 1180,6 | 11161 |
| Челябинская область | 3278,9 | 369,7 | 258,8 | 1522,1 | 14161 |
| Республика Алтай | 1488,7 | 346,9 | 20,7 | 1006,2 | 10173 |
| Республика Бурятия | 2805,5 | 323,6 | 35 | 1244 | 11299 |
| Республика Хакасия | 3986,4 | 331,8 | 97,7 | 1408,6 | 10764 |
| Алтайский край | 2882,4 | 447,9 | 289,2 | 1125,1 | 9749 |
| Забайкальский край | 3553,4 | 272,5 | 25,5 | 1018,9 | 10972 |
| Красноярский край | 4186,1 | 376,6 | 466,4 | 1385,9 | 15605 |
| Иркутская область | 3591,5 | 299,9 | 105 | 1505,7 | 12882 |
| Кемеровская область | 3249,6 | 308,8 | 439,9 | 1235 | 14439 |
| Новосибирская область | 4921,5 | 448,1 | 729,9 | 1337,9 | 12838 |
| Томская область | 4172,6 | 538,6 | 222,6 | 1232,2 | 13482 |
| Республика Саха (Якутия) | 5396,1 | 454,2 | 61,4 | 957,2 | 18741 |
| Камчатский край | 6735 | 531,2 | 43,5 | 1421,1 | 19063 |
| Приморский край | 5236,5 | 407,5 | 258 | 1531 | 12808 |
| Хабаровский край | 5334,7 | 393,6 | 137,8 | 1315,6 | 15705 |
| Амурская область | 3809,1 | 290,6 | 49,9 | 1295,9 | 11936 |
| Сахалинская область | 7182,2 | 459,7 | 29,8 | 1329,9 | 24552 |

Приложение 2



Рис 1. График распределения показателей по регионам X1 и X2.



Рис 2. График распределения показателей по регионам X1 и X3.



Рис 3. График распределения показателей по регионам X1 и X2.

Приложение 3

Табл. 1 Стандартизированный данные

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **X1** | **X2** | **X3** | **X4** | **X5** |
| Белгородская область | -0,813135 | 0,004700974 | -0,23073167 | -0,45344367 | 0,29237491 |
| Брянская область | -0,591355 | -0,65074913 | -0,46275516 | -1,02263269 | -0,57138775 |
| Владимирская область | -0,188871 | -0,81461165 | -0,26867262 | 0,236625756 | -0,71359839 |
| Воронежская область | 0,2590349 | 2,626501386 | 2,04937333 | -1,65594088 | -0,48803385 |
| Ивановская область | -0,57832 | -1,11120283 | -0,69186011 | 0,539877894 | -1,10873402 |
| Калужская область | 0,5040206 | 0,461877421 | -0,30661357 | 0,642363452 | -0,026406 |
| Костромская область | -0,210101 | 0,666705578 | -0,67361927 | 0,493102435 | -0,77181886 |
| Курская область | -0,486389 | -1,4536755 | -0,59408921 | -0,42401192 | -0,1361659 |
| Липецкая область | -0,003922 | 1,120604774 | -0,42846238 | -1,00739125 | 0,13839291 |
| Орловская область | -0,605278 | -0,08706204 | -0,58022617 | -0,58956551 | -0,64392472 |
| Рязанская область | -0,309735 | -0,48197073 | -0,41605861 | 0,537775627 | -0,16798036 |
| Смоленская область | 0,2129211 | 1,445052575 | -0,57949654 | 1,229947319 | -0,1005337 |
| Тамбовская область | -0,734337 | 0,527422431 | -0,49850721 | -0,46553171 | -0,18643275 |
| Тверская область | 1,6961655 | -0,69499201 | -0,55322973 | 0,97347064 | -0,31273617 |
| Тульская область | -0,122021 | 0,309485272 | -0,16871282 | -0,31994966 | -0,14316508 |
| Ярославская область | 0,1330367 | 0,805988725 | -0,11690883 | 0,787419935 | 0,23797218 |
| Республика Карелия | -0,599254 | 1,107495772 | -0,39271033 | 0,861524877 | 0,1240764 |
| Республика Коми | 0,9897458 | 1,971051283 | -0,29712833 | 1,036538676 | 2,16242901 |
| Архангельская область | -0,483229 | 0,535615557 | -0,32850258 | 0,935629819 | 0,9496617 |
| Вологодская область | -1,257091 | 0,109572991 | -0,37009169 | 1,182120725 | 0,11294134 |
| Калининградская область | 0,9657509 | -0,91292917 | -0,44670322 | 1,487475131 | 0,34455063 |
| Мурманская область | 0,0409078 | 0,915776617 | -0,26137629 | 2,58538452 | 2,20601482 |
| Новгородская область | -0,7687 | 1,00098513 | -0,44524395 | 1,305628962 | -0,06140191 |
| Псковская область | -0,707478 | 0,640487574 | -0,71083058 | 0,559849439 | -0,49248788 |
| Республика Адыгея | -1,989876 | -1,50119564 | -0,73782703 | -3,104929 | -1,22581124 |
| Кабардино-Балкарская Республика | -0,703134 | -1,55363165 | -0,72542326 | -1,79469056 | -1,03397003 |
| Республика Калмыкия | -1,485291 | -0,57373374 | -0,88229448 | -0,22692431 | -1,96867894 |
| Карачаево-Черкесская Республика | -0,604883 | -1,18330234 | -0,8326794 | -0,49969356 | -1,00629145 |
| Республика Северная Осетия - Алания | 0,5064892 | 0,522506555 | -0,72688252 | -1,42206358 | -0,63660739 |
| Астраханская область | 0,3634081 | -0,13130492 | -0,19352036 | 1,008683627 | -0,22874598 |
| Волгоградская область | -0,522727 | -0,17063193 | 2,16757397 | -0,00723732 | -0,30955472 |
| Ростовская область | 0,0316258 | -0,642556 | 0,87539286 | -1,04050197 | 0,10244257 |
| Республика Башкортостан | -0,235874 | 0,125959243 | 1,56124845 | -0,07976556 | 0,76800112 |
| Республика Марий Эл | -0,957204 | -0,56390199 | -0,68164524 | 0,080532367 | -1,27130592 |
| Республика Мордовия | -0,406307 | 1,425389072 | -0,57949654 | -0,05506391 | -1,09918968 |
| Удмуртская Республика | -0,585232 | -0,2918902 | 0,28074148 | -0,72095725 | -0,71837056 |
| Чувашская Республика | -1,116973 | -0,6097835 | -0,34163598 | 0,008529693 | -1,03237931 |
| Пермский край | 0,024615 | -0,0198784 | 0,32743803 | 0,194580399 | 1,36165899 |
| Кировская область | -0,503966 | -0,00185353 | -0,45545882 | -0,76563045 | -0,54943577 |
| Оренбургская область | -0,977743 | 0,345535028 | -0,06145668 | -0,43452326 | -0,52652935 |
| Пензенская область | -0,278729 | 0,048943856 | -0,27378006 | -0,16070287 | -0,53002894 |
| Самарская область | 0,8135852 | 0,468431922 | 0,90020041 | 1,430188332 | 1,26176158 |
| Саратовская область | -0,101877 | 0,335703276 | 0,43250527 | 0,099452778 | -0,88348762 |
| Ульяновская область | 0,4976022 | 0,217722258 | -0,45545882 | 0,332278943 | -0,66269525 |
| Курганская область | -0,642208 | 0,021087227 | -0,52185548 | -0,61794613 | -0,21570206 |
| Челябинская область | 0,004866 | -0,42789609 | 0,96003036 | 1,176865055 | 0,73873182 |
| Республика Алтай | -1,762862 | -0,80150265 | -0,77722724 | -1,53453491 | -0,53002894 |
| Республика Бурятия | -0,462591 | -1,18330234 | -0,67288964 | -0,28473667 | -0,1717981 |
| Республика Хакасия | 0,7034849 | -1,04893507 | -0,21540937 | 0,580346551 | -0,34200547 |
| Алтайский край | -0,386657 | 0,853508857 | 1,18183898 | -0,90963579 | -0,66492226 |
| Забайкальский край | 0,2759203 | -2,02063984 | -0,74220483 | -1,46778791 | -0,27583139 |
| Красноярский край | 0,9006781 | -0,31483095 | 2,47474972 | 0,46104285 | 1,19813266 |
| Иркутская область | 0,313542 | -1,57165652 | -0,16214612 | 1,090672073 | 0,33182484 |
| Кемеровская область | -0,024066 | -1,42581888 | 2,28139681 | -0,3320377 | 0,82717602 |
| Новосибирская область | 1,6268467 | 0,856786108 | 4,39733426 | 0,208770707 | 0,31782648 |
| Томская область | 0,8873476 | 2,339741966 | 0,695903 | -0,34675357 | 0,52271162 |
| Республика Саха (Якутия) | 2,0954892 | 0,956742249 | -0,48026637 | -1,79206273 | 2,1958342 |
| Приморский край | 1,9378926 | 0,191504253 | 0,95419329 | 1,223640515 | 0,30828214 |
| Хабаровский край | 2,0348599 | -0,03626466 | 0,0771737 | 0,091569274 | 1,22994712 |
| Амурская область | 0,5284105 | -1,72404867 | -0,56417423 | -0,01196742 | 0,03086003 |
| Сахалинская область | 3,8591686 | 1,046866638 | -0,71083058 | 0,16672535 | 4,04457261 |

Приложение 4.



Рис 1. Евклидова метрика. Метод ближнего соседа.



Рис 1. Евклидова метрика. Метод дальнего соседа.

Приложение 5

Табл. 1 Члены кластера №1 с соответствующими расстояниями.



Табл. 2 Члены кластера №2 с соответствующими расстояниями



Табл. 3 Члены кластера № 3 с соответствующими расстояниями.



Табл. 4 Члены кластера №4 с соответствующими расстояниями.



Приложение 6

1 кластер:

X1=-0,5248-0,3774\*X4

Табл 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y(X1 расчитанное) | X2 | X3 | X4 | X5 |
| Белгородская область | -0,3537 | 0,004701 | -0,23073 | -0,45344 | 0,292375 |
| Воронежская область | 0,1002 | 2,626501 | 2,049373 | -1,65594 | -0,48803 |
| Липецкая область | -0,1446 | 1,120605 | -0,42846 | -1,00739 | 0,138393 |
| Тамбовская область | -0,3491 | 0,527422 | -0,49851 | -0,46553 | -0,18643 |
| Тульская область | -0,4041 | 0,309485 | -0,16871 | -0,31995 | -0,14317 |
| Республика Северная Осетия - Алания | 0,0119 | 0,522507 | -0,72688 | -1,42206 | -0,63661 |
| Волгоградская область | -0,5221 | -0,17063 | 2,167574 | -0,00724 | -0,30955 |
| Ростовская область | -0,1321 | -0,64256 | 0,875393 | -1,0405 | 0,102443 |
| Республика Башкортостан | -0,4947 | 0,125959 | 1,561248 | -0,07977 | 0,768001 |
| Республика Мордовия | -0,5040 | 1,425389 | -0,5795 | -0,05506 | -1,09919 |
| Удмуртская Республика | -0,2527 | -0,29189 | 0,280741 | -0,72096 | -0,71837 |
| Кировская область | -0,2359 | -0,00185 | -0,45546 | -0,76563 | -0,54944 |
| Оренбургская область | -0,3608 | 0,345535 | -0,06146 | -0,43452 | -0,52653 |
| Пензенская область | -0,4642 | 0,048944 | -0,27378 | -0,1607 | -0,53003 |
| Саратовская область | -0,5623 | 0,335703 | 0,432505 | 0,099453 | -0,88349 |
| Курганская область | -0,2916 | 0,021087 | -0,52186 | -0,61795 | -0,2157 |
| Алтайский край | -0,1815 | 0,853509 | 1,181839 | -0,90964 | -0,66492 |
| Кемеровская область | -0,3995 | -1,42582 | 2,281397 | -0,33204 | 0,827176 |
| сумма | Y=-5,5407 |  |  | -10,3489 |  |
|  |  |  |  | -0,57494 |  |
|  |  |  | Э4= | -0,03916 |  |

2 кластер:

X1=0,1144+0,3002\*X4+0,8074\*X5

Табл 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y (X1 рассчитанное) | X2 | X3 | X4 | X5 |
| Брянская область | -0,65393 | -0,65075 | -0,46276 | -1,02263 | -0,57139 |
| Владимирская область | -0,39072 | -0,81461 | -0,26867 | 0,236626 | -0,7136 |
| Ивановская область | -0,61872 | -1,1112 | -0,69186 | 0,539878 | -1,10873 |
| Курская область | -0,12283 | -1,45368 | -0,59409 | -0,42401 | -0,13617 |
| Орловская область | -0,58249 | -0,08706 | -0,58023 | -0,58957 | -0,64392 |
| Республика Адыгея | -1,80742 | -1,5012 | -0,73783 | -3,10493 | -1,22581 |
| Кабардино-Балкарская Республика | -1,25919 | -1,55363 | -0,72542 | -1,79469 | -1,03397 |
| Республика Калмыкия | -1,54323 | -0,57373 | -0,88229 | -0,22692 | -1,96868 |
| Карачаево-Черкесская Республика | -0,84809 | -1,1833 | -0,83268 | -0,49969 | -1,00629 |
| Республика Марий Эл | -0,88788 | -0,5639 | -0,68165 | 0,080532 | -1,27131 |
| Чувашская Республика | -0,71658 | -0,60978 | -0,34164 | 0,00853 | -1,03238 |
| Республика Алтай | -0,77421 | -0,8015 | -0,77723 | -1,53453 | -0,53003 |
| Республика Бурятия | -0,10979 | -1,1833 | -0,67289 | -0,28474 | -0,1718 |
| Забайкальский край | -0,54894 | -2,02064 | -0,7422 | -1,46779 | -0,27583 |
| Амурская область | 0,135724 | -1,72405 | -0,56417 | -0,01197 | 0,03086 |
| сумма | Y=-10,7283 |  |  | -10,0959 | -11,659 |
|  |  |  |  | -0,63099 | -0,72869 |
|  |  |  | Э4 и Э5 соответственно: | 0,017657 | 0,05484 |

3 Кластер:

X1=0,1155-0,4086\*X2

Табл 3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y (X1 рассчитанное) | X2 | X3 | X4 | X5 |
| Калужская область | -0,0732231 | 0,4618774 | -0,3066136 | 0,6423635 | -0,026406 |
| Костромская область | -0,1569159 | 0,6667056 | -0,6736193 | 0,4931024 | -0,7718189 |
| Рязанская область | 0,3124332 | -0,4819707 | -0,4160586 | 0,5377756 | -0,1679804 |
| Смоленская область | -0,4749485 | 1,4450526 | -0,5794965 | 1,2299473 | -0,1005337 |
| Тверская область | 0,3994737 | -0,694992 | -0,5532297 | 0,9734706 | -0,3127362 |
| Ярославская область | -0,213827 | 0,8059887 | -0,1169088 | 0,7874199 | 0,2379722 |
| Республика Карелия | -0,3370228 | 1,1074958 | -0,3927103 | 0,8615249 | 0,1240764 |
| Архангельская область | -0,1033525 | 0,5356156 | -0,3285026 | 0,9356298 | 0,9496617 |
| Вологодская область | 0,0707285 | 0,109573 | -0,3700917 | 1,1821207 | 0,1129413 |
| Калининградская область | 0,4885229 | -0,9129292 | -0,4467032 | 1,4874751 | 0,3445506 |
| Мурманская область | -0,2586863 | 0,9157766 | -0,2613763 | 2,5853845 | 2,2060148 |
| Новгородская область | -0,2935025 | 1,0009851 | -0,445244 | 1,305629 | -0,0614019 |
| Псковская область | -0,1462032 | 0,6404876 | -0,7108306 | 0,5598494 | -0,4924879 |
| Астраханская область | 0,1691512 | -0,1313049 | -0,1935204 | 1,0086836 | -0,228746 |
| Пермский край | 0,1236223 | -0,0198784 | 0,327438 | 0,1945804 | 1,361659 |
| Ульяновская область | 0,0265387 | 0,2177223 | -0,4554588 | 0,3322789 | -0,6626953 |
| Челябинская область | 0,2903383 | -0,4278961 | 0,9600304 | 1,1768651 | 0,7387318 |
| Республика Хакасия | 0,5440949 | -1,0489351 | -0,2154094 | 0,5803466 | -0,3420055 |
| Иркутская область | 0,7576789 | -1,5716565 | -0,1621461 | 1,0906721 | 0,3318248 |
| сумма | Y=1,1249007 | 2,6177173 |  |  |  |
|  |  | 0,1377746 |  |  |  |
|  |  | Э2=-0,0500441 |  |  |  |

4 Кластер:

X1=1,7055-0,6151\*X2+0,5564\*X5

Табл 4.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Y (X1 рассчитанное) | X2 | X3 | X4 | X5 |
| Республика Коми | 1,696282 | 1,971051 | -0,29713 | 1,036539 | 2,162429 |
| Новосибирская область | 1,35533 | 0,856786 | 4,397334 | 0,208771 | 0,317826 |
| Томская область | 0,557161 | 2,339742 | 0,695903 | -0,34675 | 0,522712 |
| Республика Саха (Якутия) | 2,33877 | 0,956742 | -0,48027 | -1,79206 | 2,195834 |
| Приморский край | 1,759234 | 0,191504 | 0,954193 | 1,223641 | 0,308282 |
| Хабаровский край | 2,412149 | -0,03626 | 0,077174 | 0,091569 | 1,229947 |
| Сахалинская область | 3,311973 | 1,046867 | -0,71083 | 0,166725 | 4,044573 |
| сумма: | Y=13,4309 | 7,326428 |  |  | 10,7816 |
|  |  | 1,046633 |  |  | 1,540229 |
|  |  | Э2=-0,04793 |  | Э5= | 0,063807 |

Приложение 7.

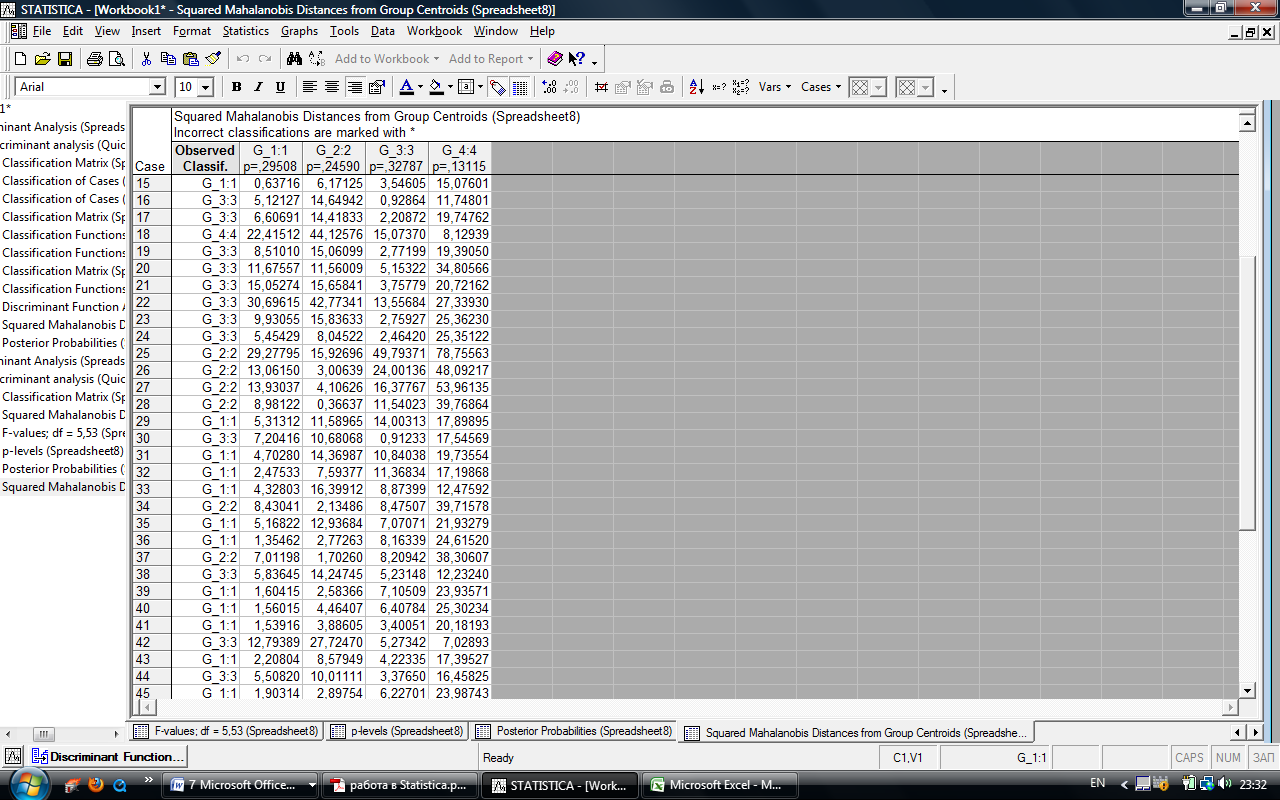


Рис. 1. Квадрат Мааланобисных расстояний

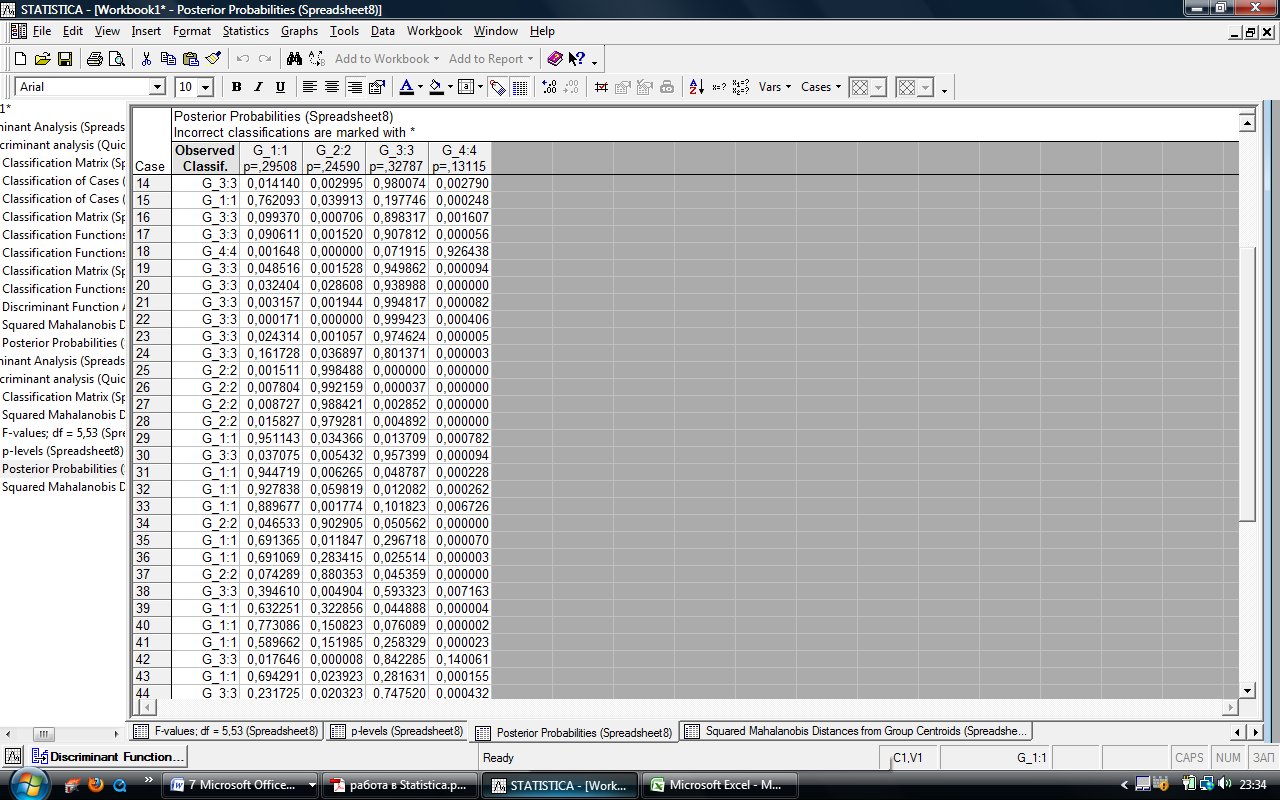


Рис. 2 Апостерирорные вероятности.

1. Кошелева В.А.«Анализ методов автоматического извлечения знаний из реляционных баз данных» [↑](#footnote-ref-1)
2. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы и основы эконометрики. / Учебное пособие./ Московский государственный университет экономики, статистики и информатики. М.: МЭСИ, 2002г. [↑](#footnote-ref-2)
3. Буреева Н.Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. Учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007 [↑](#footnote-ref-3)
4. Буреева Н.Н. Многомерный статистический анализ с использованием ППП “STATISTICA”. Учебно-методический материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007 [↑](#footnote-ref-4)