|  |
| --- |
| МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное автономное образовательное учреждение  высшего образования |
| Национальный исследовательский ядерный университет «МИФИ»  IFES-logo |

ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ

Кафедра №75 «Финансовый мониторинг»

Отчет

по дисциплине

«Макростатический анализ и прогнозирование»

по лабораторной работе № 1

Выполнил: Луканов А. В.

студент группы С18-712,

Проверила: Домашова Д. В.

Москва - 2021

Содержание

[Постановка задачи 3](#_Toc84174814)

[Выбранные показатели 4](#_Toc84174815)

[Метод «Полных связей» 5](#_Toc84174816)

[Метод Уорда: 12](#_Toc84174817)

# Постановка задачи

1. Выбрать предмет исследования, а также набор показателей, характеризующих данное явление или процесс.
2. По данным кластерного анализа:

* провести классификацию региональных образований с помощью иерархических агломеративных методов кластерного анализа;
* провести классификацию региональных образований с помощью метода К-средних.

1. Сравнить классификации, полученные с помощью агломеративных кластер-процедур и метода К-средних, обосновать выбор окончательного варианта классификации;
2. Дать экономическую интерпретацию результатов классификации.

# Выбранные показатели

Исходные данные для анализа представлены в таблице 1:

Таблица 1 – Выбранные показатели.

|  |  |
| --- | --- |
|  | **ОБЩИЕ КОЭФФИЦИЕНТЫ СМЕРТНОСТИ (число умерших на 1000 человек населения)** |
|  | **СРЕДНЕДУШЕВЫЕ ДЕНЕЖНЫЕ ДОХОДЫ НАСЕЛЕНИЯ (в месяц; рублей)** |
|  | **ОБЩАЯ ПЛОЩАДЬ ЖИЛЫХ ПОМЕЩЕНИЙ, ПРИХОДЯЩАЯСЯ В СРЕДНЕМ НА ОДНОГО ЖИТЕЛЯ (на конец года; квадратных метров)** |
|  | **Уровень занятости для людей в возрасте 15-72 лет(процентов)** |
|  | **КОЭФФИЦИЕНТЫ МИГРАЦИОННОГО ПРИРОСТА на 10 000 человек населения** |
|  | **Зарегистрировано преступлений, связанных с незаконным оборотом наркотиков, на 1000 населения** |
|  | **Зарегистрировано преступлений экономической направленности на 1000 человек** |
|  | **ЧИСЛО ПРЕДПРИЯТИЙ И ОРГАНИЗАЦИЙ (на конец года) на 1000 человек** |
|  | **РАСХОДЫ НА ОХРАНУ ОКРУЖАЮЩЕЙ СРЕДЫ(млн руб) на 1000 чел** |
|  | **ВЫБРОСЫ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В АТМОСФЕРНЫЙ ВОЗДУХ,  ОТХОДЯЩИХ ОТ СТАЦИОНАРНЫХ ИСТОЧНИКОВ тысяч тонн** |

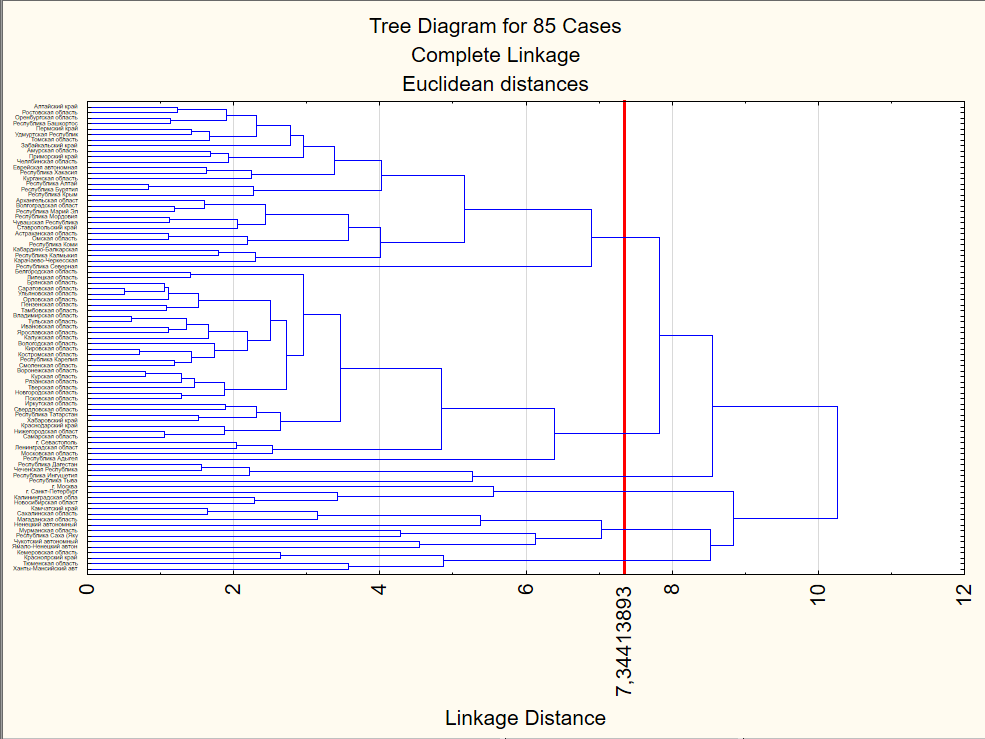
Для приведения исходных переменных к стандартизованному виду можно воспользоваться операцией центрирования и нормирования данных. Результаты преобразования данных представлены на рисунке 1:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 1 – Центрировано-нормированные значения признаков

# Метод «Полных связей»



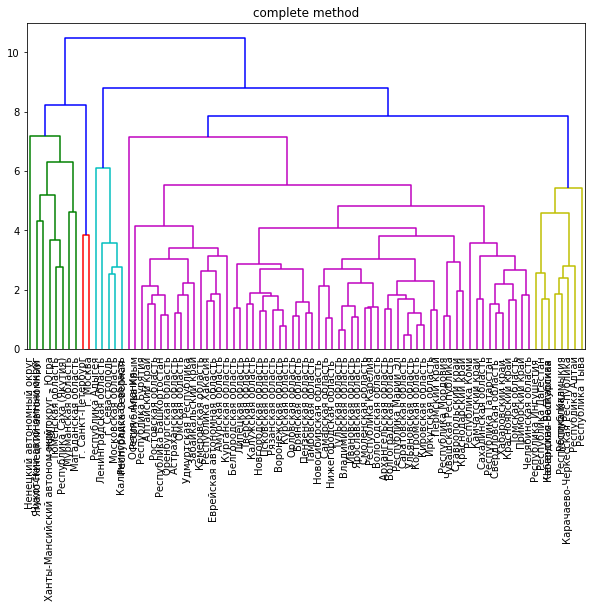


Рисунок 2 – Дендрограмма объединения классов методом «полных связей»

Методом «полных связей» при пороговом значении расстояния ρ=7,34 все регионы РФ разбиваются на шесть классов , состав которых приведен в таблице 2:

Таблица 2 – Результаты классификации регионов РФ методом «полных связей»

|  |  |  |
| --- | --- | --- |
| Номер  кластера | Количество  объектов в кластере | Состав класса |
| кластер 1 | 4 | Кемеровская область  Красноярский край  Тюменская область  Ханты-Мансийский автономный округ – Югра |
| кластер 2 | 4 | Республика Дагестан Республика Ингушетия Республика Тыва Чеченская Республика |
| кластер 3 | 4 | г. Москва  г. Санкт-Петербург  Калининградская область  Новосибирская область |
| кластер 4 | 35 | Белгородская область  Брянская область  Владимирская область  Вологодская область  Воронежская область  г. Севастополь  Ивановская область  Иркутская область  Калужская область  Кировская область  Костромская область  Краснодарский край  Курская область  Ленинградская область  Липецкая область  Московская область  Нижегородская область  Новгородская область  Орловская область  Пензенская область  Псковская область  Республика Адыгея  Республика Карелия  Республика Татарстан  Рязанская область  Самарская область  Саратовская область  Свердловская область  Смоленская область  Тамбовская область  Тверская область  Тульская область  Ульяновская область  Хабаровский край  Ярославская область |
| кластер 5 | 30 | Алтайский край  Амурская область  Архангельская область  Астраханская область  Волгоградская область  Еврейская автономная область  Забайкальский край  "Кабардино-Балкарская  Республика"  Карачаево-Черкесская Республика  Курганская область  Омская область  Оренбургская область  Пермский край  Приморский край  Республика Алтай  Республика Башкортостан  Республика Бурятия  Республика Калмыкия  Республика Коми  Республика Крым  Республика Марий Эл  Республика Мордовия  "Республика Северная  Осетия – Алания"  Республика Хакасия  Ростовская область  Ставропольский край  Томская область  Удмуртская Республика  Челябинская область  Чувашская Республика |
| кластер 6 | 8 | Камчатский край  Магаданская область  Мурманская область  Ненецкий автономный округ  Республика Саха (Якутия)  Сахалинская область  Чукотский автономный округ  "Ямало-Ненецкий автономный  округ" |

В данном случае уровень порогового значения выбирается из тех соображений, чтобы получить небольшое количество кластеров.

Далее рассчитываются средние значения показателей в каждом классе. Графическое изображение информации о средних значениях признаков в классах представлено на рисунке 3.

Таблица 3 – Средние значения признаков в классах

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер кластера | 1 | 2 | 3 | 4 | 5 | 6 |
| Среднее по полю X1 | -0,812326876 | -2,560035153 | -0,37761771 | 0,675316328 | -0,029850377 | -0,967580149 |
| Среднее по полю X2 | 0,513272642 | -0,7890829 | 0,912253718 | -0,185677296 | -0,473179424 | 2,268539284 |
| Среднее по полю X3 | -0,607878408 | -2,337731023 | -0,349691451 | 0,779507647 | -0,374219213 | -0,359373463 |
| Среднее по полю X4 | 0,565556249 | -1,779077883 | 1,235451718 | 0,097184202 | -0,450290064 | 1,252441817 |
| Среднее по полю X5 | -0,037076809 | -0,060297622 | 0,938197342 | 0,350352506 | -0,431087565 | -0,336625297 |
| Среднее по полю X6 | 0,058224214 | -0,216206646 | -0,388307354 | -0,384586258 | 0,492352127 | 0,109389289 |
| Среднее по полю X7 | -1,013729646 | -0,974997092 | 0,135336097 | -0,199977148 | 0,294569927 | 0,696958118 |
| Среднее по полю X8 | -0,246344565 | -1,535774045 | 3,101686028 | 0,139652045 | -0,439618798 | 0,377809087 |
| Среднее по полю X9 | 1,187180335 | -0,703212177 | -0,477353909 | -0,263920252 | -0,296217824 | 2,262160817 |
| Среднее по полю X10 | 3,977224078 | -0,510668865 | -0,358526945 | -0,171274059 | -0,191033767 | -0,088313495 |

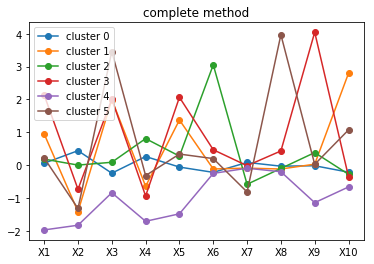


Рисунок 3 – График средних значений признаков в каждом кластере

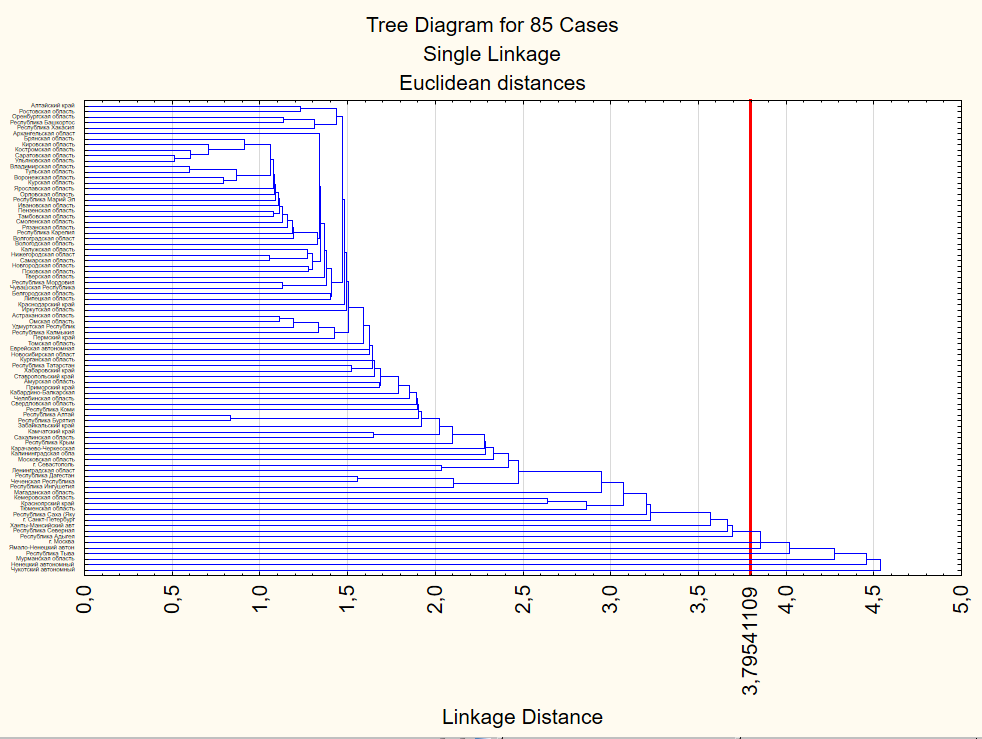
**Выводы:**

Анализируя график средних значений в классах, можно сделать следующие выводы:

* Первый класс довольно сильно отличается от остальных классов только показателем выбросы загрязняющих веществ в атмосферу, отходящих от стационарных источников.
* Второй класс характеризуется тем, что по сравнению с остальными классами, низкими значениями всех признаков, с небольшим повышением к коэффициенту миграционного прироста.
* Третий класс, где находятся г. Москва, г. Санкт-Петербург, Калининградская область, Новосибирская область характеризуется, по сравнению с остальными классами, повышенными значениями числа предприятий и организаций на 1000 человек.
* Четвертый класс, имеет самые высокие показатели коэффициента смертности и общей площади жилых помещений, приходящихся в среднем на одного человека.
* Пятый класс, включающий в себя 30 регионов, имеет высокие показатели зарегистрированных преступлений, связанных с незаконным оборотом наркотиков и экономических преступлений.
* Шестой класс выделяется самыми высокими показателями средне-душевого дохода населения, уровня занятости населения, числа экономических преступлений и расходов на охрану окружающей среды.

С помощью метода «полных связей» получено достаточно неравномерное распределение объектов по классам, так в первый, второй и третий кластер вошло по 4 объекта.

# Метод «Одиночной связи»



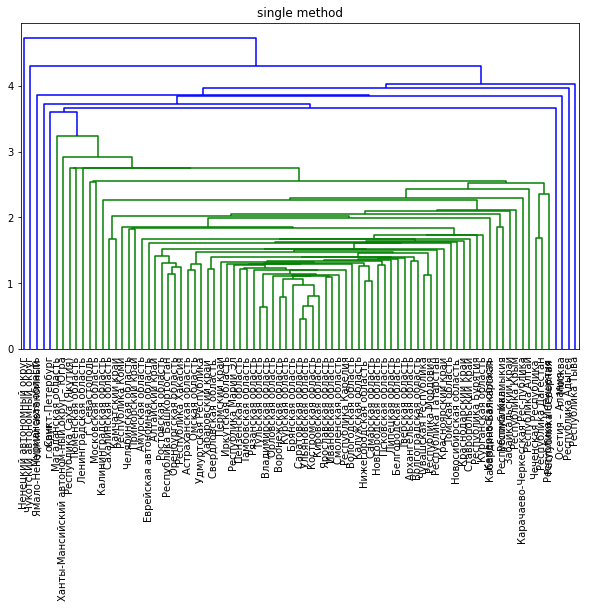


Рисунок 2 – Дендрограмма объединения классов методом «полных связей»

Методом «полных связей» при пороговом значении расстояния ρ=3,80 все регионы РФ разбиваются на семь классов , состав которых приведен в таблице 2:

Таблица 2 – Результаты классификации регионов РФ методом «полных связей»

|  |  |  |
| --- | --- | --- |
| Номер  кластера | Количество  объектов в кластере | Состав класса |
| кластер 1 | 1 | г. Москва |
| кластер 2 | 1 | Мурманская область |
| кластер 3 | 1 | Ненецкий автономный округ |
| кластер 4 | 1 | Республика Тыва |
| кластер 5 | 1 | Чукотский автономный округ |
| кластер 6 | 1 | Ямало-Ненецкий автономный  округ |
| кластер 7 | 79 | Алтайский край  Амурская область  Архангельская область  Астраханская область  Белгородская область  Брянская область  Владимирская область  Волгоградская область  Вологодская область  Воронежская область  г. Санкт-Петербург  г. Севастополь  Еврейская автономная область  Забайкальский край  Ивановская область  Иркутская область  "Кабардино-Балкарская  Республика"  Калининградская область  Калужская область  Камчатский край  Карачаево-Черкесская Республика  Кемеровская область  Кировская область  Костромская область  Краснодарский край  Красноярский край  Курганская область  Курская область  Ленинградская область  Липецкая область  Магаданская область  Московская область  Нижегородская область  Новгородская область  Новосибирская область  Омская область  Оренбургская область  Орловская область  Пензенская область  Пермский край  Приморский край  Псковская область  Республика Адыгея  Республика Алтай  Республика Башкортостан  Республика Бурятия  Республика Дагестан  Республика Ингушетия  Республика Калмыкия  Республика Карелия  Республика Коми  Республика Крым  Республика Марий Эл  Республика Мордовия  Республика Саха (Якутия)  "Республика Северная  Осетия – Алания"  Республика Татарстан  Республика Хакасия  Ростовская область  Рязанская область  Самарская область  Саратовская область  Сахалинская область  Свердловская область  Смоленская область  Ставропольский край  Тамбовская область  Тверская область  Томская область  Тульская область  Тюменская область  Удмуртская Республика  Ульяновская область  Хабаровский край  Ханты-Мансийский автономный округ – Югра  Челябинская область  Чеченская Республика  Чувашская Республика  Ярославская область |

В данном случае уровень порогового значения выбирается из тех соображений, чтобы получить небольшое количество кластеров.

Далее рассчитываются средние значения показателей в каждом классе. Графическое изображение информации о средних значениях признаков в классах представлено на рисунке 3.

Таблица 3 – Средние значения признаков в классах

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер кластера | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Среднее по полю X1 | -0,998630803 | -0,324388016 | -1,31800897 | -1,42446835 | -1,14057665 | -2,701981 | 0,100101947 |
| Среднее по полю X2 | 2,96178974 | 0,846377675 | 3,45758058 | -1,1156429 | 3,62388478 | 3,60281296 | -0,169326618 |
| Среднее по полю X3 | -1,81490243 | -0,239961995 | -0,317418082 | -3,15747461 | -0,704698518 | -1,53089678 | 0,0982956 |
| Среднее по полю X4 | 2,2014604 | 0,900503982 | 0,337403445 | -2,74994088 | 2,93931627 | 2,33738121 | -0,075520562 |
| Среднее по полю X5 | 0,571308494 | -1,3420865 | 0,199775485 | -0,357524029 | 1,92740398 | -0,580443835 | -0,005296628 |
| Среднее по полю X6 | -0,499940246 | 1,23036958 | -0,81623344 | 2,90486296 | -1,52324176 | -0,4441238 | -0,010780928 |
| Среднее по полю X7 | 1,69754907 | -0,678047519 | 3,96985886 | -1,45269858 | 1,02618482 | -0,832977731 | -0,047213531 |
| Среднее по полю X8 | 4,34625495 | -0,250706008 | 0,024687967 | -1,35726641 | -0,002726818 | -0,602113704 | -0,027318101 |
| Среднее по полю X9 | -0,305057385 | 4,67056657 | 1,0733148 | -0,55226544 | 2,84272519 | 4,51624882 | -0,155006741 |
| Среднее по полю X10 | -0,359114365 | 0,007436053 | -0,377911823 | -0,523592117 | -0,493046249 | 1,25746697 | 0,006186855 |

Рисунок 3 – График средних значений признаков в каждом кластере

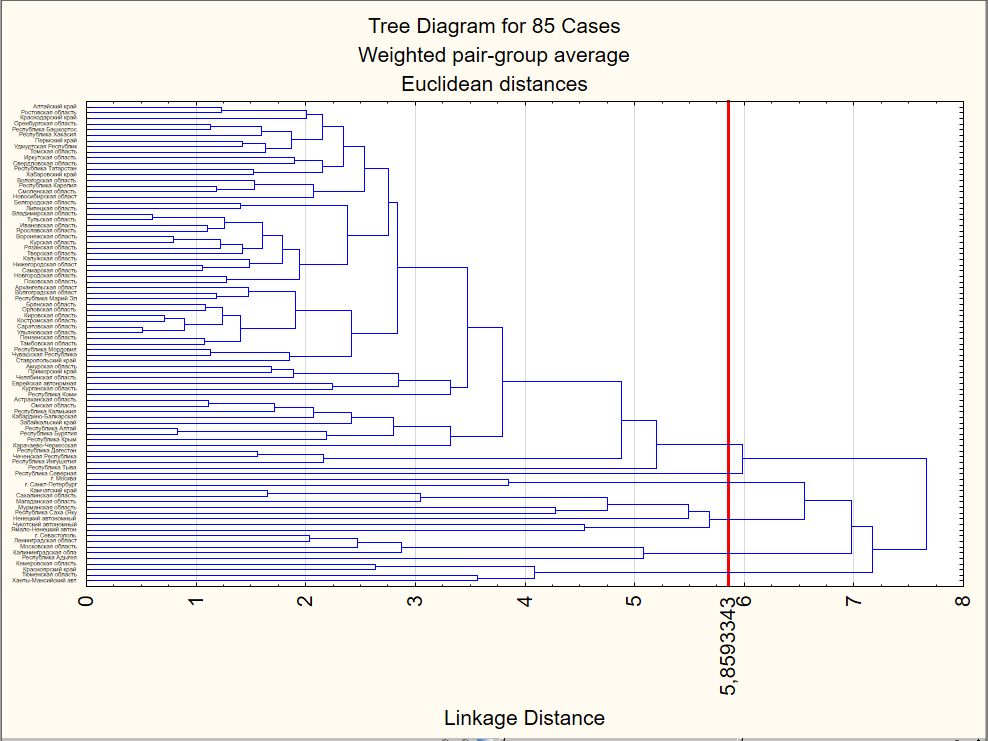
**Выводы:**

Анализируя график средних значений в классах, можно сделать следующие выводы:

* Первый класс характеризуется, самым высоким показателем числа предприятий и организаций.
* Второй класс обладает самым низким коэффициентом миграционного прироста.
* Третий класс, характеризуется самым высоким числом зарегистрированных экономических преступлений.
* Четвёртый класс, выделяется в высоких значениях преступлений, связанных с незаконным оборотом наркотиков, на 1000 населения, а также обладает самыми низкими показателями уровня занятости населения и общей жило площади, приходящейся на одного человека.
* Пятый класс имеет высокие показатели миграционного прироста и уровня занятости населения.
* Шестой класс отличается высокими показателями среднедушевых доходов населения и расходов на охрану окружающей среды.
* Седьмой класс, включающий в себя 79 объектов, имеет средние показатели относительно других классов.

С помощью метода «Одиночной связи» получено неравномерное распределение объектов по классам, так в первом, втором, третьем, четвертом, пятом и шестом классе вошло только по одному объекту, а в седьмом 79.

**Метод «Взвешенного попарного среднего»**



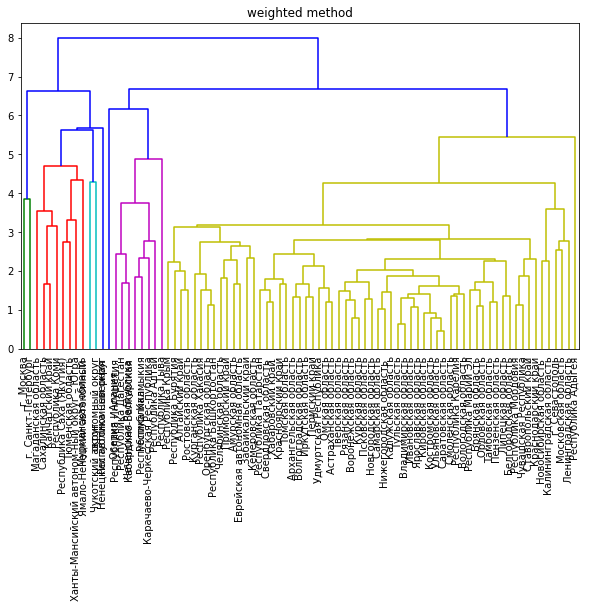


Рисунок 4 – Дендрограмма объединения классов методом «взвешенного попарного среднего»

Методом «взвешенного попарного среднего» при пороговом значении расстояния ρ=5,86 все регионы РФ разбиваются на шесть классов , состав которых приведен в таблице 4.

Таблица 4 – Результаты классификации регионов РФ методом «средней связи»

|  |  |  |
| --- | --- | --- |
| Номер  кластера | Количество  объектов в кластере | Состав класса |
| кластер 1 | 1 | Республика Северная  Осетия – Алания |
| кластер 2 | 2 | г. Москва  г. Санкт-Петербург |
| кластер 3 | 4 | Кемеровская область  Красноярский край  Тюменская область  Ханты-Мансийский автономный округ – Югра |
| кластер 4 | 5 | г. Севастополь  Калининградская область  Ленинградская область  Московская область  Республика Адыгея |
| кластер 5 | 65 | Алтайский край  Амурская область  Архангельская область  Астраханская область  Белгородская область  Брянская область  Владимирская область  Волгоградская область  Вологодская область  Воронежская область  Еврейская автономная область  Забайкальский край  Ивановская область  Иркутская область  "Кабардино-Балкарская  Республика"  Калужская область  Карачаево-Черкесская Республика  Кировская область  Костромская область  Краснодарский край  Курганская область  Курская область  Липецкая область  Нижегородская область  Новгородская область  Новосибирская область  Омская область  Оренбургская область  Орловская область  Пензенская область  Пермский край  Приморский край  Псковская область  Республика Алтай  Республика Башкортостан  Республика Бурятия  Республика Дагестан  Республика Ингушетия  Республика Калмыкия  Республика Карелия  Республика Коми  Республика Крым  Республика Марий Эл  Республика Мордовия  Республика Татарстан  Республика Тыва  Республика Хакасия  Ростовская область  Рязанская область  Самарская область  Саратовская область  Свердловская область  Смоленская область  Ставропольский край  Тамбовская область  Тверская область  Томская область  Тульская область  Удмуртская Республика  Ульяновская область  Хабаровский край  Челябинская область  Чеченская Республика  Чувашская Республика  Ярославская область |
| кластер 6 | 8 | Камчатский край  Магаданская область  Мурманская область  Ненецкий автономный округ  Республика Саха (Якутия)  Сахалинская область  Чукотский автономный округ  "Ямало-Ненецкий автономный  округ" |

Далее рассчитываются средние значения показателей в каждом классе. Графическое изображение информации о средних значениях признаков в классах представлено на рисунке 5.

Таблица 5 – Средние значения признаков в классах

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номер кластера | 1 | 2 | 3 | 4 | 5 | 6 |
| Среднее по полю X1 | -0,714739103 | -0,732482335 | -0,812326876 | 0,002087439 | 0,202449466 | -0,967580149 |
| Среднее по полю X2 | -0,554295289 | 2,00809478 | 0,513272642 | 0,087891483 | -0,370811639 | 2,268539284 |
| Среднее по полю X3 | 0,715329746 | -0,937066778 | -0,607878408 | 0,80827705 | 0,037291076 | -0,359373463 |
| Среднее по полю X4 | -1,9149987 | 2,06553958 | 0,565556249 | 0,275268215 | -0,24421817 | 1,252441817 |
| Среднее по полю X5 | -1,13774335 | 0,469136917 | -0,037076809 | 3,053148996 | -0,188076397 | -0,336625297 |
| Среднее по полю X6 | 3,7979261 | -0,016197714 | 0,058224214 | -0,581804368 | -0,030223385 | 0,109389289 |
| Среднее по полю X7 | 2,47220014 | 0,43228567 | -1,013729646 | -0,099641392 | -0,067066321 | 0,696958118 |
| Среднее по полю X8 | -1,06318054 | 4,04531538 | -0,246344565 | 0,38257552 | -0,168883419 | 0,377809087 |
| Среднее по полю X9 | -0,448887526 | -0,353000766 | 1,187180335 | -0,27838888 | -0,312294991 | 2,262160817 |
| Среднее по полю X10 | -0,518892753 | -0,368513094 | 3,977224078 | -0,340316908 | -0,188382844 | -0,088313495 |

Рисунок 5 – График средних значений признаков в каждом кластере

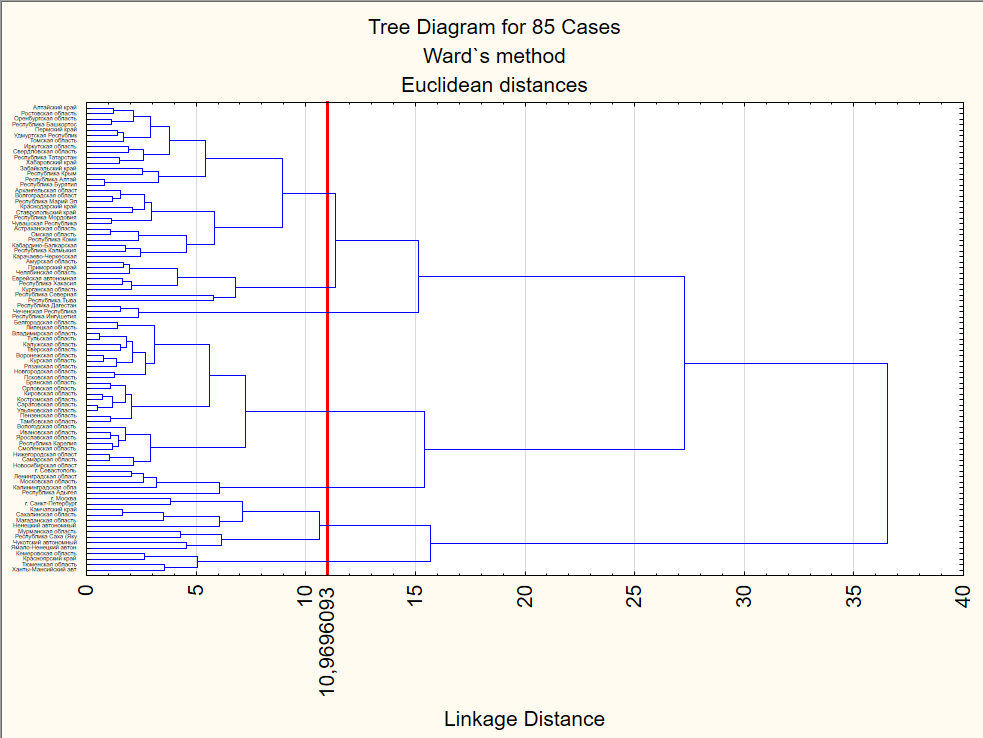
**Выводы:**

Анализируя график средних значений в классах, можно сделать следующие выводы:

* Первый класс, куда вошла только республика Северная Осетия - Алания, по сравнению с остальными классами, имеет существенные различия в показателях уровень занятости населения и коэффициент миграционного прироста (самые низкие показатели среди всех классов), а также совершенные преступления как и экономического характера, так и связанные с незаконным оборотом наркотиков(самые высокие показатели среди других классов)
* Второй класс, где находятся г. Москва и г. Санкт-Петербург, характеризуется, по сравнению с остальными классами, повышенными значениями числа предприятий и организаций и уровнем занятости населения.
* Третий класс, имеет усредненные значения по всем признакам кроме выбросов загрязняющих веществ в атмосферу. По этому признаку третий класс имеет наибольший показатель.
* Четвертый класс, обладает самым высоким коэффициентом миграционного прироста.
* Пятый класс, включающий в себя 65 регионов, имеет усреднённые средние значения по всем признакам.
* Шестой класс, включающий в себя дальневосточные и северные регионы, выделяется самыми высокими показателями среднедушевых доходов населения и расходов на охрану окружающей среды.

С помощью метода «взвешенного попарного среднего» получено неравномерное распределение объектов по классам, так в первый класс вошёл 1 объект, в третий – 2 объекта, а в то время, как в пятый вошло 65 объектов. Данный недостаток можно устранить методом Уорда.

Метод Уорда:



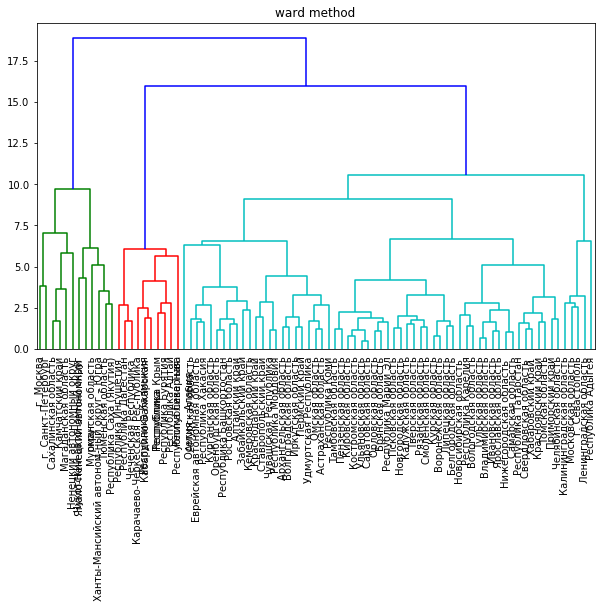


Рисунок 6 – Дендрограмма объединения классов методом Уорда

Методом Уорда при пороговом значении расстояния ρ=10,97 все регионы РФ разбиваются на 7 классов , состав которых приведен в таблице 6:

Таблица 6 – Результаты классификации регионов РФ методом Уорда

|  |  |  |
| --- | --- | --- |
| Номер  кластера | Количество  объектов в кластере | Состав класса |
| кластер 1 | 3 | Республика Дагестан  Республика Ингушетия Чеченская Республика |
| кластер 2 | 4 | Кемеровская область  Красноярский край  Тюменская область  Ханты-Мансийский автономный округ – Югра |
| кластер 3 | 5 | г. Севастополь  Калининградская область  Ленинградская область  Московская область  Республика Адыгея |
| кластер 4 | 8 | Амурская область  Еврейская автономная область  Курганская область  Приморский край  "Республика Северная  Осетия – Алания"  Республика Тыва  Республика Хакасия  Челябинская область |
| кластер 5 | 27 | Белгородская область  Брянская область  Владимирская область  Вологодская область  Воронежская область  Ивановская область  Калужская область  Кировская область  Костромская область  Курская область  Липецкая область  Нижегородская область  Новгородская область  Новосибирская область  Орловская область  Пензенская область  Псковская область  Республика Карелия  Рязанская область  Самарская область  Саратовская область  Смоленская область  Тамбовская область  Тверская область  Тульская область  Ульяновская область  Ярославская область |
| кластер 6 | 28 | Алтайский край  Архангельская область  Астраханская область  Волгоградская область  Забайкальский край  Иркутская область  "Кабардино-Балкарская  Республика"  Карачаево-Черкесская Республика  Краснодарский край  Омская область  Оренбургская область  Пермский край  Республика Алтай  Республика Башкортостан  Республика Бурятия  Республика Калмыкия  Республика Коми  Республика Крым  Республика Марий Эл  Республика Мордовия  Республика Татарстан  Ростовская область  Свердловская область  Ставропольский край  Томская область  Удмуртская Республика  Хабаровский край  Чувашская Республика |
| кластер 7 | 10 | г. Москва  г. Санкт-Петербург  Камчатский край  Магаданская область  Мурманская область  Ненецкий автономный округ  Республика Саха (Якутия)  Сахалинская область  Чукотский автономный округ  "Ямало-Ненецкий автономный  округ" |

Далее рассчитываются средние значения показателей в каждом классе. Графическое изображение информации о средних значениях признаков в классах представлено на рисунке 7.

Таблица 7 – Средние значения признаков в классах

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер кластера | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Среднее по полю X1 | -2,93855742 | -0,812326876 | 0,002087439 | 0,119192766 | 0,853236814 | -0,097528131 | -0,920560586 |
| Среднее по полю X2 | -0,680229566 | 0,513272642 | 0,087891483 | -0,453858268 | -0,314453884 | -0,374830075 | 2,216450383 |
| Среднее по полю X3 | -2,06448316 | -0,607878408 | 0,80827705 | -0,536876996 | 0,895104368 | -0,376432244 | -0,474912126 |
| Среднее по полю X4 | -1,455456883 | 0,565556249 | 0,275268215 | -0,842195093 | 0,060527702 | -0,297124869 | 1,41506137 |
| Среднее по полю X5 | 0,038777847 | -0,037076809 | 3,053148996 | -0,521927386 | -0,019910608 | -0,313072758 | -0,175472855 |
| Среднее по полю X6 | -1,256563181 | 0,058224214 | -0,581804368 | 2,2048317 | -0,406912836 | -0,037461122 | 0,084271888 |
| Среднее по полю X7 | -0,815763262 | -1,013729646 | -0,099641392 | 0,019138439 | -0,182653382 | 0,190668316 | 0,644023629 |
| Среднее по полю X8 | -1,59527659 | -0,246344565 | 0,38257552 | -0,489806546 | 0,177315397 | -0,290137013 | 1,111310345 |
| Среднее по полю X9 | -0,753527756 | 1,187180335 | -0,27838888 | -0,333898325 | -0,357107027 | -0,22051437 | 1,739128501 |
| Среднее по полю X10 | -0,506361114 | 3,977224078 | -0,340316908 | -0,249560434 | -0,292453012 | -0,048284981 | -0,144353415 |

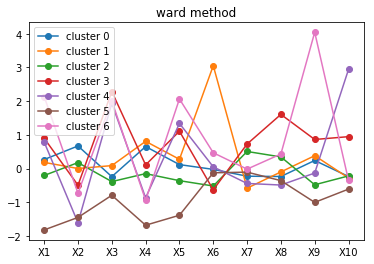


Рисунок 7 – График средних значений признаков в каждом кластере

**Выводы:**

* Первый класс, где находятся кавказские регионы, характеризуется усредненными значениями, по сравнению с остальными классами, но при этом самым высоким значением коэффициента смертности.
* Второй класс отличается от остальных регионов высокими выбросами загрязняющих веществ в атмосферу.
* Третий класс характеризуется, по сравнению с остальными классами, средними показателями. При этом обладает повышенными значениями коэффициента миграционного прироста.
* Четвертый класс характеризуется высокими значениями преступлений, связанных с незаконным оборотом наркотиков, на 1000 населения.
* Пятый класс характеризуется, по сравнению с остальными классами, самыми низкими значениями общего коэффициента смертности, общей пощади жилых помещений, приходящейся на одного человека, уровня занятости населения, числа предприятий и организаций, числа преступлений, связанных с незаконным оборотом наркотиков, на 1000 населения и расходами на охрану окружающей среды.
* Шестой класс обладает, по сравнению с остальными классами, средними показателями всех признаков.
* Седьмой класс характеризуется самыми высокими среднедушевыми доходами и уровнем занятости населения.

Таким образом, с помощью метода Уорда получилось добиться примерно равномерного распределения объектов на классы.

Протокол объединения классов методом Уорда представлен на рисунке 8.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 8 – Протокол объединения кластеров методом Уорда

Матрица обычных евклидовых расстояний между объектами представлена на рисунке 9.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 9 – Матрица расстояний

# Метод К-средних

Метод К-средних позволяет получить более устойчивое разбиение, но требует задания некоторых начальных условий (количество образуемых кластеров, порог завершения процесса классификации и т. д.).

Информация о составе кластеров, а также евклидово расстояние от объектов до центра соответствующего класса представлены на рисунках ниже.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 1 - Класс 1

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2 - Класс 2

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 3 - Класс 3

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 4 - Класс 4

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 5 - Класс 5

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 6 - Класс 6

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 7 - Класс 7

В первой таблице, представленной на рисунке 8, указаны средние значения признаков в каждом классе. Во второй таблице, представленной на рисунке 9, приведены расстояния между классами. Причем, ниже главной диагонали указаны расстояния между классами, рассчитанные по метрике обычного евклидового расстояния, а выше главной диагонали – расстояния между классами, рассчитанные по метрике квадратичного евклидового расстояния.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 8 - Средние значения признаков в классах

Изображение выглядит как текст, кроссворд, внутренний

Автоматически созданное описание

Рисунок 9 - Расстояния между классами

Как видно из рисунка 8 наименьшее расстояние наблюдается между пятым и третьим классами (0,455465).

Информация о значениях сумм квадратов при расчете межгрупповой дисперсии (Between) и внутригрупповой дисперсии (Within) по каждому признаку, а также соответствующие им степени свободы представлена на рисунке 11.

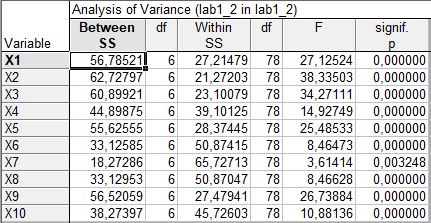
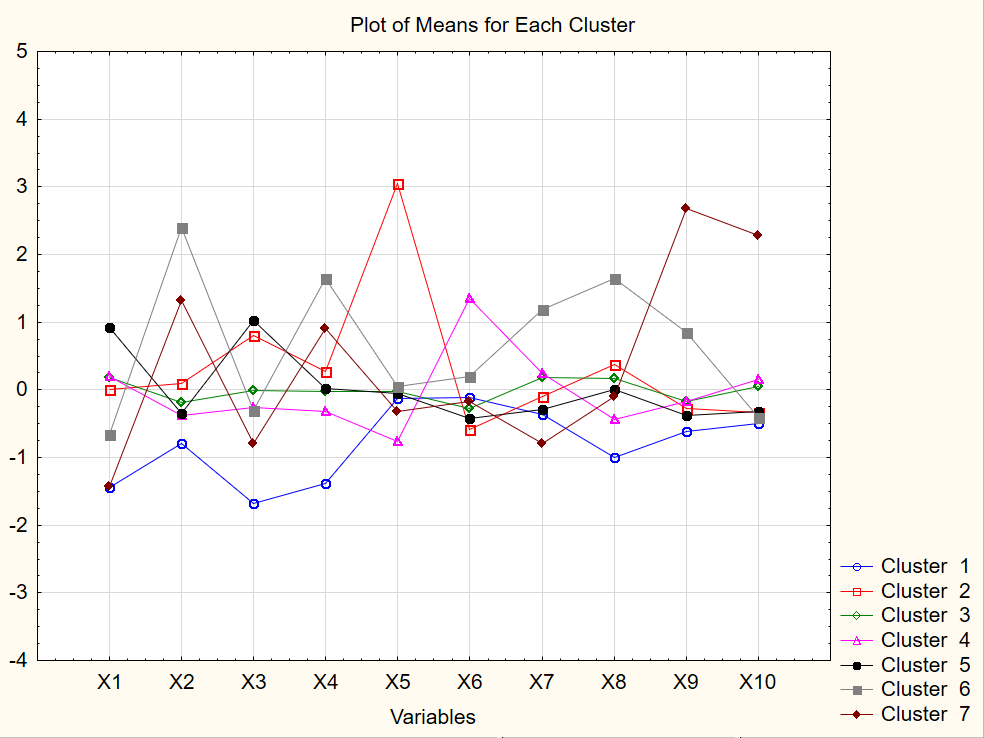


Рисунок 10 – Анализ дисперсий

|  |  |  |
| --- | --- | --- |
| Признаки | Межгрупповая дисперсия | Внутригрупповая дисперсия |
| X1 | 9,464201667 | 0,348907564 |
| X2 | 10,45466167 | 0,272718333 |
| X3 | 10,14986833 | 0,296163974 |
| X4 | 7,483125 | 0,501298077 |
| X5 | 9,270925 | 0,363775 |
| X6 | 5,520975 | 0,652232692 |
| X7 | 3,045476667 | 0,842655513 |
| X8 | 5,521588333 | 0,652185513 |
| X9 | 9,420098333 | 0,352300128 |
| X10 | 6,378995 | 0,586231154 |

Таблица 11 – Значения межгрупповых и внутригрупповых дисперсий

Таблица, представленная на рисунке 10, содержит также наблюденное значение F-критерия, а также значимость нулевой гипотезы о равенстве межгрупповой и внутригрупповой дисперсий. На уровне значимости 0,05 по всем признакам нулевая гипотеза отвергается. Это означает, что каждый из признаков вносит существенный вклад в разделение объектов на классы.



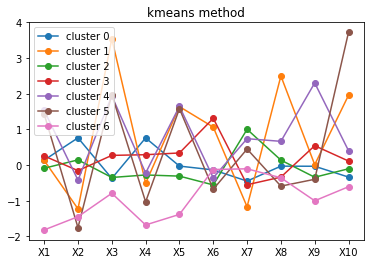


Рисунок 12 – График средних значений признаков в каждом кластере

На рисунках ниже представлены результаты расчетов описательных статистик для каждого кластера: среднего арифметического, оценку среднего квадратичного отклонения, несмещенную оценку дисперсии по каждому признаку.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 13 – Результаты расчета описательных статистик первого кластера.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 14 – Результаты расчета описательных статистик второго кластера.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 15 – Результаты расчета описательных статистик третьего кластера.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16 – Результаты расчета описательных статистик четвёртого кластера.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 17 – Результаты расчета описательных статистик пятого кластера.

Изображение выглядит как стол

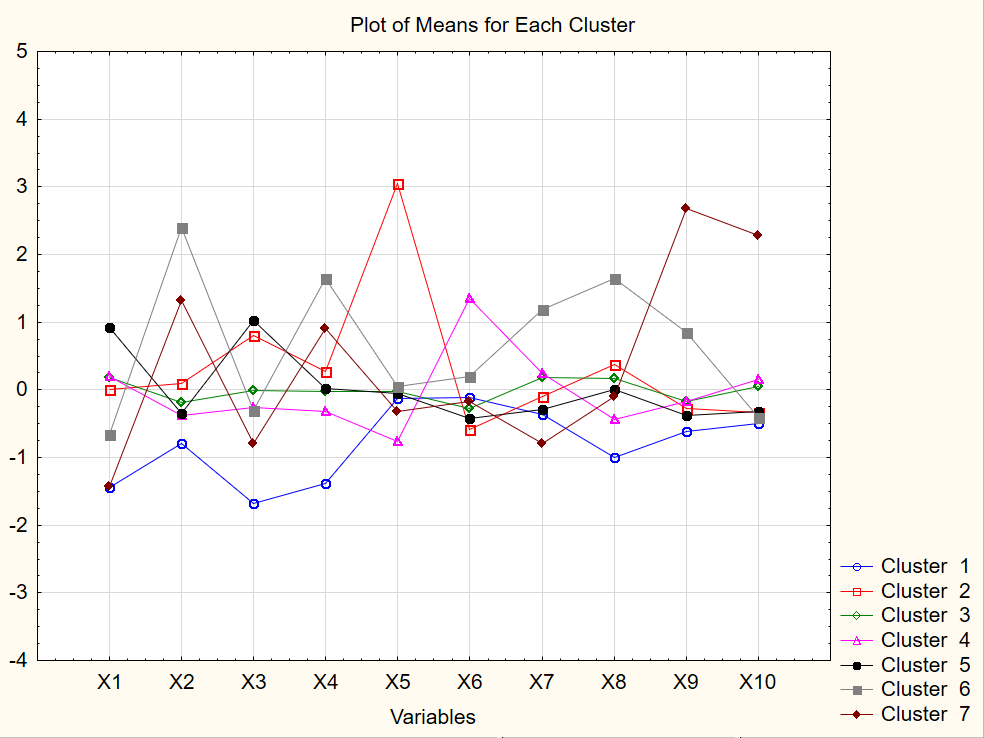
Автоматически созданное описание

Рисунок 18 – Результаты расчета описательных статистик шестого кластера.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 19 – Результаты расчета описательных статистик седьмого кластера.

**Выводы:** 

Первый класс содержит 10 объектов. Типичные представители кластера -Кавказские и Сибирские регионы: Республика Ингушетия, Республика Дагестан, Чеченская Республика, республика Алтай, республика Тыва. Кластер характеризуется самыми низкими значениями общего коэффициента смертности(Х1), среднедушевых доходов населения(Х2), общей пощади жилых помещений, приходящейся на одного человека(Х3), уровня занятости населения(Х4), числа предприятий и организаций(Х8), числа преступлений, связанных с незаконным оборотом наркотиков(Х9). Это может быть связано с культурными особенностями данных регионов.

Второй класс содержит 5 объектов. Представители кластера: г. Севастополь, Калининградская область, Московская область, Ленинградская область, Республика Адыгея характеризуется, по сравнению с остальными классами, средними показателями в диапазоне от -0.5 до 1. При этом обладает самым высоким значением коэффициента миграционного прироста(Х5). Данные регионы активно развиваются, в них строятся много новостроек и тем самым являются привлекательными для переезда из другого региона.

Третий класс состоит из 20 объектов. Типичные представители - регионы Поволжья: Самарская область, Республика Татарстан, Нижегородская область, Волгоградская область. Кластер характеризуется, по сравнению с остальными классами, средними показателями. Кластер обладает высоким показателем числа предприятий и организаций на 1000 человек и достаточно высоким коэффициентом смертности. Данные регионы расположены в центральной части России, имеют хорошие климатические условия, а также имеют достаточно развитую инфраструктуру.

Четвертый содержит 14 регионов. Типичные представители: Омская область, Челябинская область, Приморский край, Забайкальский край. Регионы характеризуются самыми высоким количеством преступлений, связанных с незаконным оборотом наркотиков, на 1000 населения и самым низким коэффициентом миграционного прироста. Регионы данного класса расположены далеко от центральной части России, и с этим связано низкий коэффициент миграционного прироста.

Пятый класс состоит из 23 объектов. Типичные представители: Костромская область, Кировская область, Ульяновская область, Воронежская область. Класс характеризуется, по сравнению с остальными классами, средними показателями. При этом обладает самыми высокими значениями коэффициента смертности (Х1) и общей площадью жилых помещений, приходящихся на одного человека(Х3). Показатели всех остальных признаков для данного кластера находятся примерно на одном уровне в диапазоне от -0.5 до 0.5. В данном кластере много регионов со средними для России показателями, находятся они достаточно далеко от центра, поэтому в них низкий миграционный прирост.

Шестой класс состоит из 7 объектов: Сахалинская область, Камчатский край, г. Санкт-Петербург, Магаданская область, Ненецкий автономный округ, Чукотский автономный округ, г. Москва. Кластер обладает самыми высокими среднедушевыми доходами населения(Х2), уровнем занятости(Х4), количеством экономических преступлений(Х7) и числом организаций и предприятий(Х8). Данный кластер включает в себя регионы с самыми высокими доходами в РФ.

Седьмой класс состоит из 6 объектов. Представители данного кластера – регионы Сибири и Северной части РФ: Республика Саха (Якутия), Ханты-Мансийский автономный округ – Югра, Мурманская область, Красноярский край, Тюменская область, Ямало-Ненецкий автономный округ. Эти регионы расположены далеко от центральной части России. Кластер характеризуется самыми высокими расходами на охрану окружающей среды(Х9) и выбросами загрязняющих веществ в атмосферу(Х10).

# Сравнение классификаций

С помощью метода «полных связей», метода Уорда и метода К-средних были получены различные классификации. Для выбора лучшей классификации необходимо воспользоваться функционалом качества разбиения сумма квадратов расстояний от каждого объекта до центра кластера:

Тогда значение функционала качества для классификации, полученной методом «полной связи», рассчитывается следующим образом:

420,9190596

Аналогичным образом можно рассчитать функционал качества и для других классификаций. Функционал качества классификации, полученной методом «взвешенной попарно средней»:

486,6195323

Функционал качества классификации, полученной методом Уорда:

387,6516522

Функционал качества классификации, полученной методом К-средних:

379,74050

По выбранному функционалу качества наилучшей является классификация , полученная методом К-средних.