МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ»

(НИЯУ МИФИ)

Институт Финансовых Технологий и Экономической Безопасности

Кафедра Финансового Мониторинга

Лабораторная работа №4:

По курсу «Макростатический анализ и прогнозирование»

Работу выполнил: студент группы С18-712: Луканов А.В.

Проверила: Домашова Д. В.

Москва 2021

# 1. Постановка задачи

Субъект РФ характеризуются следующими признаками:

X1 - ОБЩИЕ КОЭФФИЦИЕНТЫ СМЕРТНОСТИ (число умерших на 1000 человек населения);

Х2 – СРЕДНЕДУШЕВЫЕ ДЕНЕЖНЫЕ ДОХОДЫ НАСЕЛЕНИЯ (в месяц; рублей);

Х3 – ОБЩАЯ ПЛОЩАДЬ ЖИЛЫХ ПОМЕЩЕНИЙ, ПРИХОДЯЩАЯСЯ В СРЕДНЕМ НА ОДНОГО ЖИТЕЛЯ (на конец года; квадратных метров);

Х4 – Уровень занятости для людей в возрасте 15-72 лет(процентов)

Х5 – коэффициенты миграционного прироста на 10 000 человек населения;

Х6 – зарегистрировано преступлений, связанных с незаконным оборотом наркотиков (на 1000 человек);

Х7 – Зарегистрировано преступлений экономической направленности на 1000 человек;

Х8 – ЧИСЛО ПРЕДПРИЯТИЙ И ОРГАНИЗАЦИЙ (на конец года) на 1000 человек;

Х9 – РАСХОДЫ НА ОХРАНУ ОКРУЖАЮЩЕЙ СРЕДЫ(млн руб) на 1000 чел;

Х10 - ВЫБРОСЫ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В АТМОСФЕРНЫЙ ВОЗДУХ, ОТХОДЯЩИХ ОТ СТАЦИОНАРНЫХ ИСТОЧНИКОВ тысяч тонн;

Ставится задача на основании статистических данных по показателям, соответствующим нужному варианту, снизить размерность признакового пространства методом главных компонент, обеспечив уровень информативности новой системы признаков не ниже 70%.

# 2. Ход работы.

Первым делом стандартизируем исходные данные, чтобы привести их к одному масштабу.

Если исходные признаки, по которым производится классификация объектов, имеют разные единицы измерения, то необходимо перейти к стандартизованным переменным одним из следующих способов:

; ; ; ; ; ,

где  - исходное значение j-го признака на *i*-ом объекте наблюдения;

 - нормированное значение исходного *j*-го признака на *i*-ом объекте наблюдения;

 - среднее значение *j*-го признака;

 - выборочное среднеквадратическое отклонение *j*-го признака;

 - максимальное значение *j*-го признака;

 - минимальное значение *j*-го признака.

Была рассчитана матрица корреляций для исходных данных. Матрицы представлены на Рисунках 1.1 и 1.2.

Изображение выглядит как текст, кроссворд, табло

Автоматически созданное описание

Рисунок 1.1 – Результат расчета корреляционной матрицы (Statistica).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 1.2 – Результат расчета корреляционной матрицы (Python).

Теперь проверим гипотезу о незначимости матрицы корреляций. Для этого посчитаем собственные числа для данной матрицы. Результаты представлены на Рисунках 2.1, 2.2.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2.1 – Результаты расчета оценок собственных чисел корреляционной матрицы (Statistica)

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 3.1 – Результаты расчета оценок собственных чисел корреляционной матрицы (Python)

Затем были посчитаны наблюдаемое и критическое значение статистики Хи-квадрат, в результате гипотеза о незначимости корреляционной матрицы была отвергнута. Результаты представлены на Рисунках 3.1 и 3.2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Hi2набл | Hi2крит1 | Hi2крит2 | alpha | k | n | det(R) = |
| 310,9174 | 65,64662 | 118,1359 | 0,05 | 10 | 85 | 0,021356 |

Рисунок 3.1– Проверка гипотезы о незначимости корреляционной матрицы в Excel



Рисунок 3.2– Проверка гипотезы о незначимости корреляционной матрицы в Python

Рассчитаем оценки общностей для 5 главных факторов. (Рисунки 4.1 и 4.2)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 4.1 - Результаты расчета общностей (Statistica)

Рассчитаем редуцированную матрицу общностей (Рисунки 5.1 и 5.2).Изображение выглядит как стол

Автоматически созданное описание

Рисунок 5.1 – Редуцированная матрица общностей (Excel)

Суммарная общность равна 4,668078. Рассчитаем собственные значения для редуцированной матрицы общностей (Рисунки 6.1 и 6.2).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 6.1 - Собственные значения для редуцированной матрицы общностей (Statistica)

Так как сумма оценок первых двух собственных чисел редуцированной матрицы больше суммарной общности (), то размерность признакового пространства можно снизить до двух главных факторов.

Снизим размерность исходного пространства до 2 главных факторов и ценим вклад каждого фактора в суммарную дисперсию (Рисунки 7.1 и 7.2)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 7.1 – Вклад главных факторов в суммарную дисперсию исходных признаков (Statistica)

Вклад двух главных факторов в суммарную дисперсию исходных признаков (в дисперсию процесса) составляет 39,68%. Был построен график собственных чисел, представленный на рисунке 8.1, 8.2.

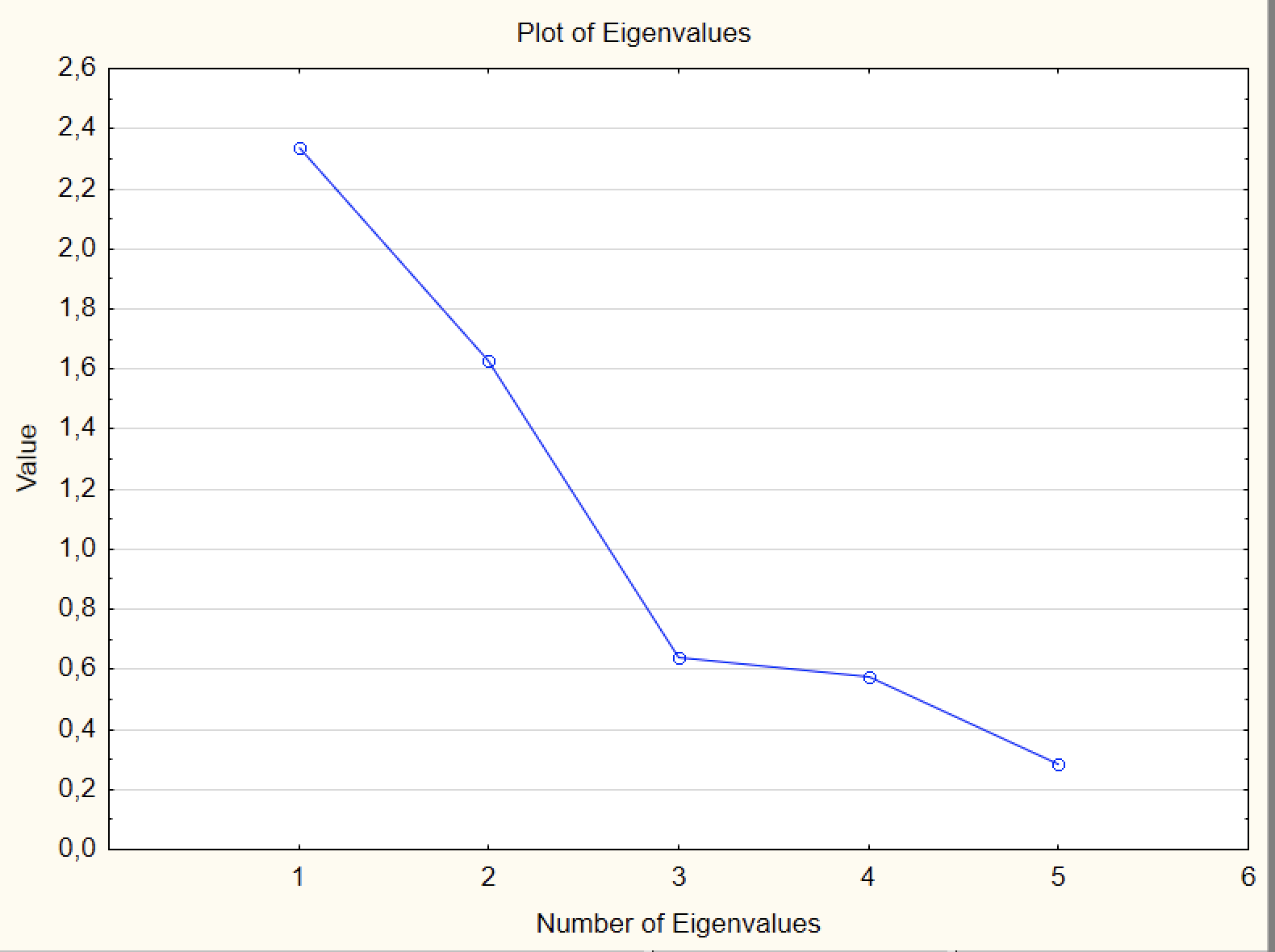


Рисунок 8.1 – График собственных чисел (Statistica)

Рассчитаем таблицу весовых коэффициентов при главных факторах (Рисунок 9.1 и 9.2)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 9.1 - Таблица весовых коэффициентов при главных факторах (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 9.2 - Таблица весовых коэффициентов при главных факторах (Python)

Проверка гипотезы о достаточности 2 главных факторов (Рисунки 10.1 и 10.2)

Получим расположение исходных признаков на плоскости, образованной главными факторами (Рисунок 11.1 и 11.2)

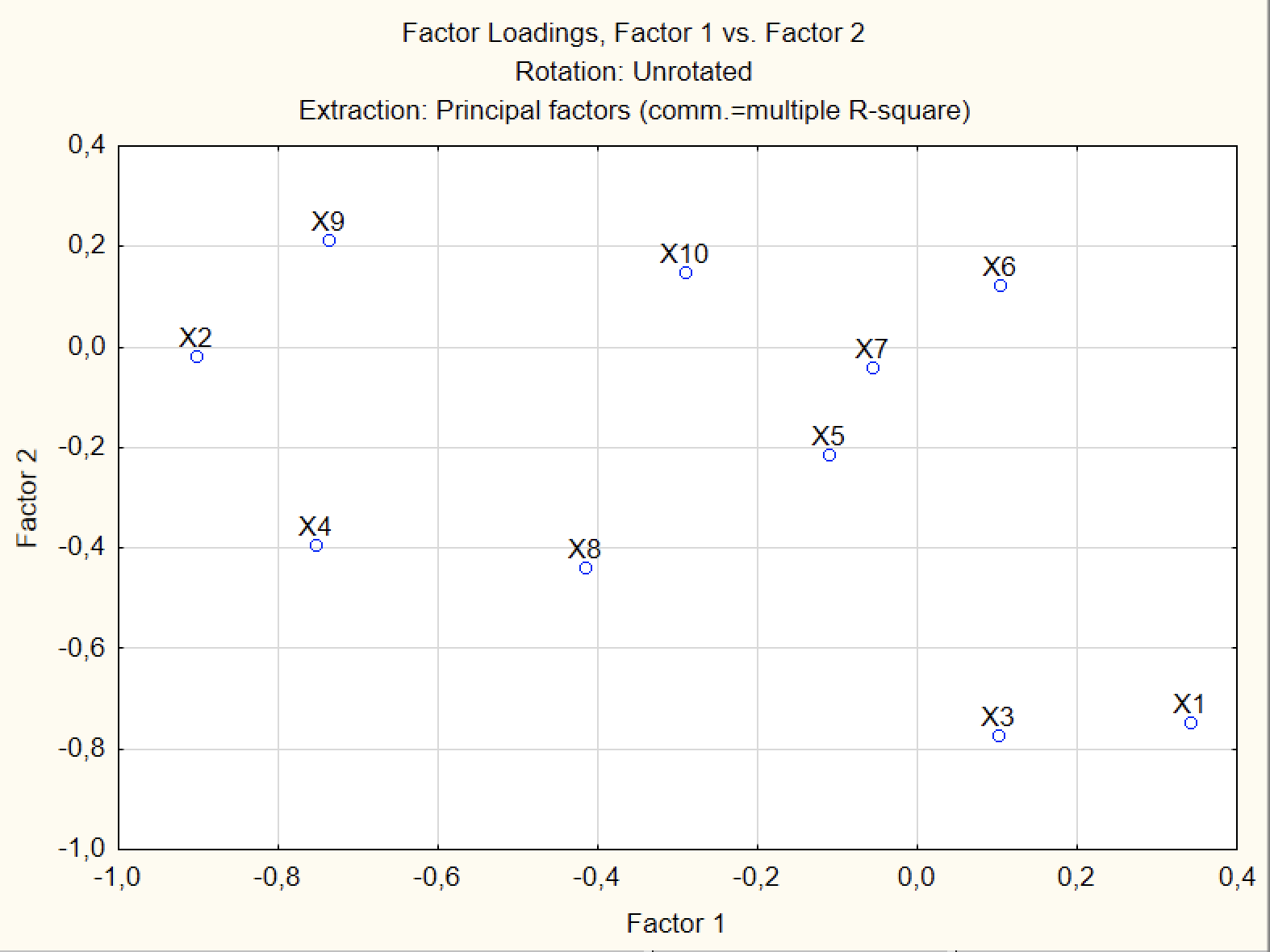


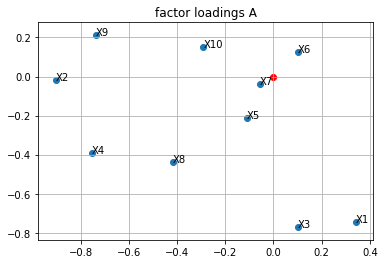
Рисунок 11.1 - Расположение исходных признаков на плоскости, образованной главными факторами (Statistica) 

Рисунок 11.2 - Расположение исходных признаков на плоскости, образованной главными факторами (Python)

Упростим структуру главных факторов с помощью ортогонального вращения (Рисунок 12.1 и 12.2).Изображение выглядит как стол

Автоматически созданное описание

Рисунок 12.1 - Весовые коэффициенты факторов после вращения (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 12.2 - Весовые коэффициенты факторов после вращения (Python)

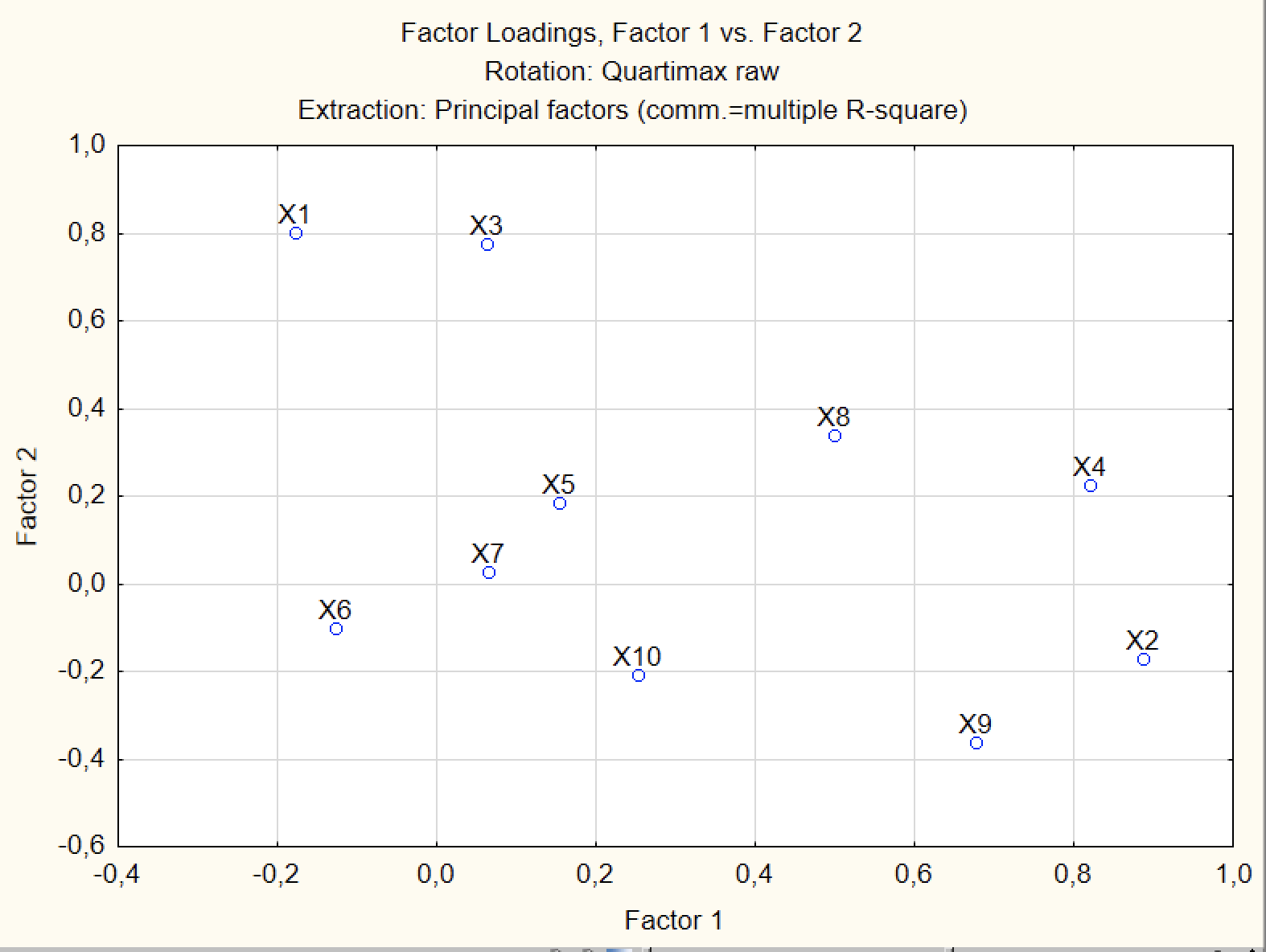


Рисунок 13.1 - Расположение исходных признаков на плоскости, образованной обобщенными факторами (Statistica)

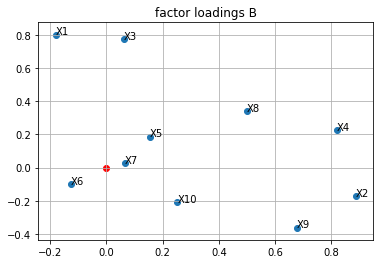


Рисунок 13.2 - Расположение исходных признаков на плоскости, образованной обобщенными факторами (Python)

X1 - ОБЩИЕ КОЭФФИЦИЕНТЫ СМЕРТНОСТИ (число умерших на 1000 человек населения);

Х2 – СРЕДНЕДУШЕВЫЕ ДЕНЕЖНЫЕ ДОХОДЫ НАСЕЛЕНИЯ (в месяц; рублей);

Х3 – ОБЩАЯ ПЛОЩАДЬ ЖИЛЫХ ПОМЕЩЕНИЙ, ПРИХОДЯЩАЯСЯ В СРЕДНЕМ НА ОДНОГО ЖИТЕЛЯ (на конец года; квадратных метров);

Х4 – Уровень занятости для людей в возрасте 15-72 лет(процентов)

Х5 – коэффициенты миграционного прироста на 10 000 человек населения;

Х6 – зарегистрировано преступлений, связанных с незаконным оборотом наркотиков (на 1000 человек);

Х7 – Зарегистрировано преступлений экономической направленности на 1000 человек;

Х8 – ЧИСЛО ПРЕДПРИЯТИЙ И ОРГАНИЗАЦИЙ (на конец года) на 1000 человек;

Х9 – РАСХОДЫ НА ОХРАНУ ОКРУЖАЮЩЕЙ СРЕДЫ(млн руб) на 1000 чел;

Х10 - ВЫБРОСЫ ЗАГРЯЗНЯЮЩИХ ВЕЩЕСТВ В АТМОСФЕРНЫЙ ВОЗДУХ, ОТХОДЯЩИХ ОТ СТАЦИОНАРНЫХ ИСТОЧНИКОВ тысяч тонн;

Проведем анализ матрицы нагрузок.

Первая главный фактор тесно связан (модуль коэффициента корреляции больше 0,7) с четырьмя исходными признаками: x2 (среднедушевые доходы населения), x4 (уровень занятости населения), x9 (расходы на охрану окружающей среды). Поэтому первый главный фактор можно интерпретировать как «Экономический фактор».

Вторая главный фактор тесно отрицательно связан с признаком х1 (общий коэффициент смертности) и x3 (общая площадь жилых помещений, приходящихся на одного жителя), поэтому второй главный фактор можно интерпретировать «Смертность».

Рассчитаем вклады факторов в дисперсию признаков (Рисунок 14.1 и 14.2)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 14.1 - Вклады одного и двух главных факторов в дисперсию признаков (Statistica)

Оценка редуцированной матрицы парных коэффициентов корреляции, рассчитанная по матрице нагрузок, и оценка остаточной матрицы парных коэффициентов представлены на рисунках 15.1, 15.2, 16.1 и 16.2.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 15.1 – Оценка редуцированной матрицы (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 15.2 – Оценка редуцированной матрицы (Python)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16.1 – Оценка остаточной матрицы парных коэффициентов корреляции (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16.2 – Оценка остаточной матрицы парных коэффициентов корреляции (Python)

Была рассчитана матрица индивидуальных значений обобщенных факторов (Рисунок 17.1 и 17.2).

Изображение выглядит как стол

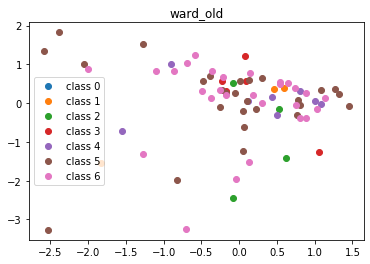
Автоматически созданное описание

Рисунок 17.1 – Индивидуальные значения общих факторов (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 17.2 – Индивидуальные значения общих факторов (Statistica)



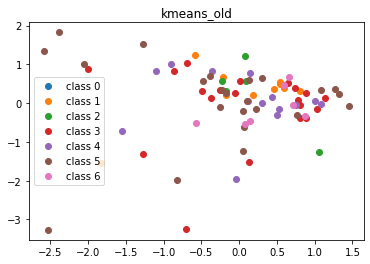
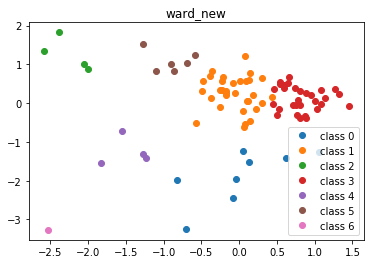


Рисунок 13.1, 13.2 – точечные графики старых кластеров в новой системе координат (2 главных фактора)



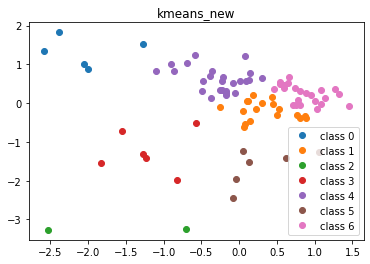
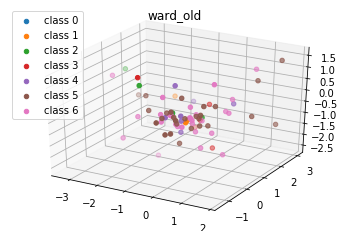


Рисунок 14.1, 14.2 – точечные графики новых кластеров в новой системе координат (2 главных фактора)



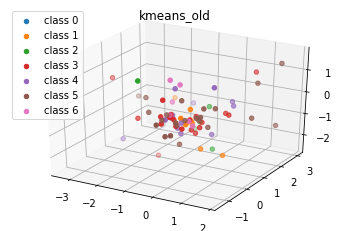
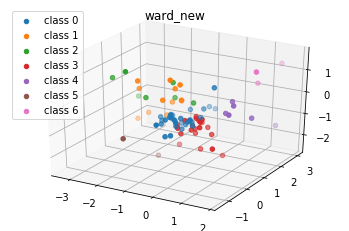


Рисунок 15.1, 15.2 – точечные графики старых кластеров в новой системе координат (3 главных фактора)



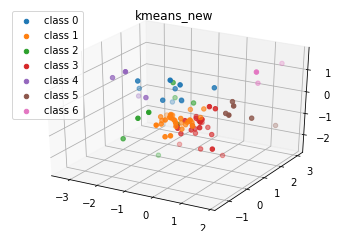


Рисунок 16.1, 16.2 – точечные графики новых кластеров в новой системе координат (3 главных фактора)