|  |
| --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ |
| Федеральное государственное автономное образовательное  учреждение высшего образования |
| **НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»** |



***ИНСТИТУТ ФИНАНСОВЫХ ТЕХНОЛОГИЙ И ЭКОНОМИЧЕСКОЙ БЕЗОПАСНОСТИ***

КАФЕДРА «ФИНАНСОВЫЙ МОНИТОРИНГ»

ОТЧЕТ

по лабораторной работе №4

по курсу «Макростатистический анализ и прогнозирование»

Выполнил

студент группы С18-712

Степаненкова Маргарита

Преподаватель: Домашова Д. В.

Москва, 2021

# Постановка задачи

Исходные данные:

* X1 - валовой региональный продукт (млн руб на 1 000 человек населения)
* X2 - инвестиции в основной капитал (в фактически действовавших ценах; млн руб на 1 000 человек населения)
* X3 - число предприятий и организаций (на конец года на 1 000 человек населения)
* X4 - уровень занятости для людей в возрасте 15-72 лет (процентов)
* X5 - коэффициенты миграционного прироста на 10 000 человек населения
* X6 - зарегистрировано преступлений особой тяжести (на 1 000 человек населения)
* X7 - зарегистрировано преступлений, связанных с незаконным оборотом наркотиков (на 1 000 человек населения)
* X8 - общие коэффициенты смертности (число умерших на 1000 человек населения)
* X9 - мощность амбулаторно-поликлинических организаций всего, тыс. посещений в смену (на конец года, на 1 000 человек населения)
* X10 - численность врачей всех специальностей, всего, тыс. человек (на конец года, на 1 000 человек населения)
* X11 - численность пенсионеров, тыс. человек (на 1 000 человек населения)

1. с помощью компонентного анализа и метода главных факторов снизить размерность признакового пространства, обеспечив уровень информативности не менее 70%;
2. при необходимости провести вращение пространства новых факторов;
3. дать экономическую интерпретацию факторам;
4. найти матрицу индивидуальных значений факторов.

# Выполнение работы

Исходные данные для анализа представлены в виде матрицы. Фрагмент таблицы с исходными данными в пакете Statistica представлен на Рисунке 1.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 1 – Исходные данные для анализа

Для приведения исходных переменных к стандартизованному виду можно воспользоваться операцией центрирования и нормирования данных. Результаты преобразования данных представлены на Рисунке 2.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 2.1 – Центрировано-нормированные значения признаков (Statistica)

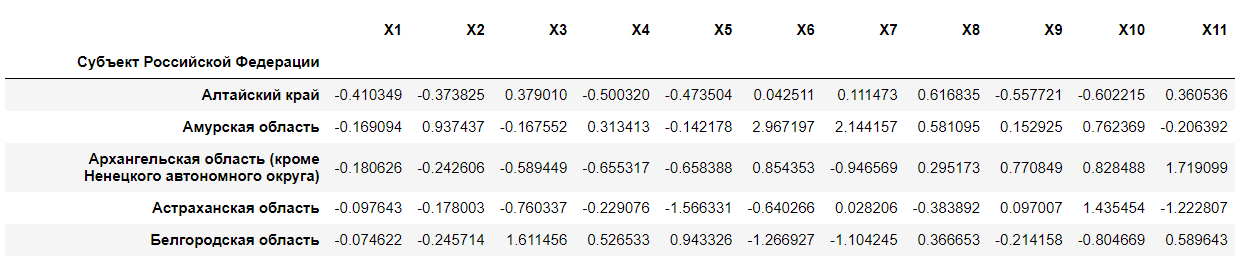


Рисунок 2.2 – Центрировано-нормированные значения признаков (Python)

Рассчитаем выборочные значения коэффициентов корреляции для каждой пары признаков (Рисунок 3).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 3.1 - Результаты расчета корреляционной матрицы (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 3.2 - Результаты расчета корреляционной матрицы (Python)

Далее согласно алгоритму, предполагая, что выборка извлечена из нормально распределенной генеральной совокупности, на уровне значимости α = 0,05 проверим гипотезу о незначимости корреляционной матрицы.

H0: Rx = Е;

H1: Rx ≠ E.

Для проверки гипотезы потребуются оценки собственных чисел корреляционной матрицы (Рисунок 4).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 4.1 - Результаты расчета оценок собственных чисел корреляционной матрицы в пакете Statistica

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 4.2 - Результаты расчета оценок собственных чисел корреляционной матрицы в Python

Расчет наблюдаемого значения проводится в Excel на основании полученных оценок собственных значений, которые приводятся на Рисунке 5.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 5.1 - Проверка гипотезы о незначимости корреляционной матрицы в Excel

Наблюдаемое значение рассчитывается по формуле:

χ2 = ,

где |R ̂x| - определитель матрицы R ̂\_x, равный произведению оценок собственных чисел матрицы;

k – число факторов;

n – объем выборки.

Наблюдаемое значение составило χ2 набл = 575,20. Критические значения и определяются из уравнений:



Для решения этих уравнений необходимо воспользоваться функцией ХИ2ОБР (вероятность, v) пакета Excel. Вероятность рассчитывается как 100\*()% для χкр1 и 100\*( )% для χкр2. Число степеней свободы . Критические точки принимают следующие значения:

Так как , то гипотеза отвергается, матрица парных коэффициентов корреляции значима. Это подтверждается в Python.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 5.2 - Проверка гипотезы о незначимости корреляционной матрицы в Python

# Изображение выглядит как текст Автоматически созданное описание

Рисунок 6 - Результаты факторного анализа

Изображение выглядит как текст, окно

Автоматически созданное описание

Рисунок 7.1 – Результаты расчета общностей (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 7.2 – Результаты расчета общностей (Python)

На основе оценки матрицы парных коэффициентов корреляции и оценок общностей можно составить оценку редуцированной матрицы:

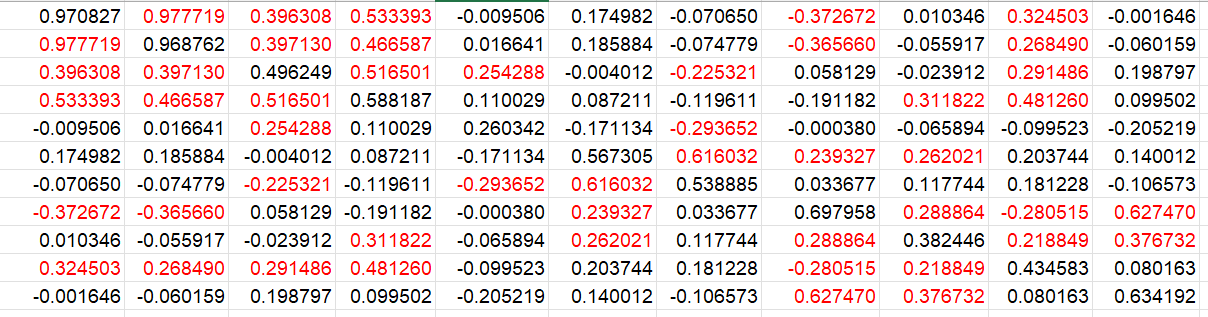


Рисунок 8.1 – Редуцированная матрица общностей (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 8.2 – Редуцированная матрица общностей (Python)

Суммарная общность составляет: 6,54. Рассчитаем собственные значения для редуцированной матрицы общностей.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 9 - Собственные значения для редуцированной матрицы общностей (Statistica)

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

Снизим размерность исходного пространства до 4 главных факторов и оценим вклад каждого фактора в суммарную дисперсию (Рисунок 10).

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 10 - вклад главных факторов в суммарную дисперсию исходных признаков

Вклад 4 главных факторов в суммарную дисперсию исходных признаков (в дисперсию процесса) составляет 60,29%.

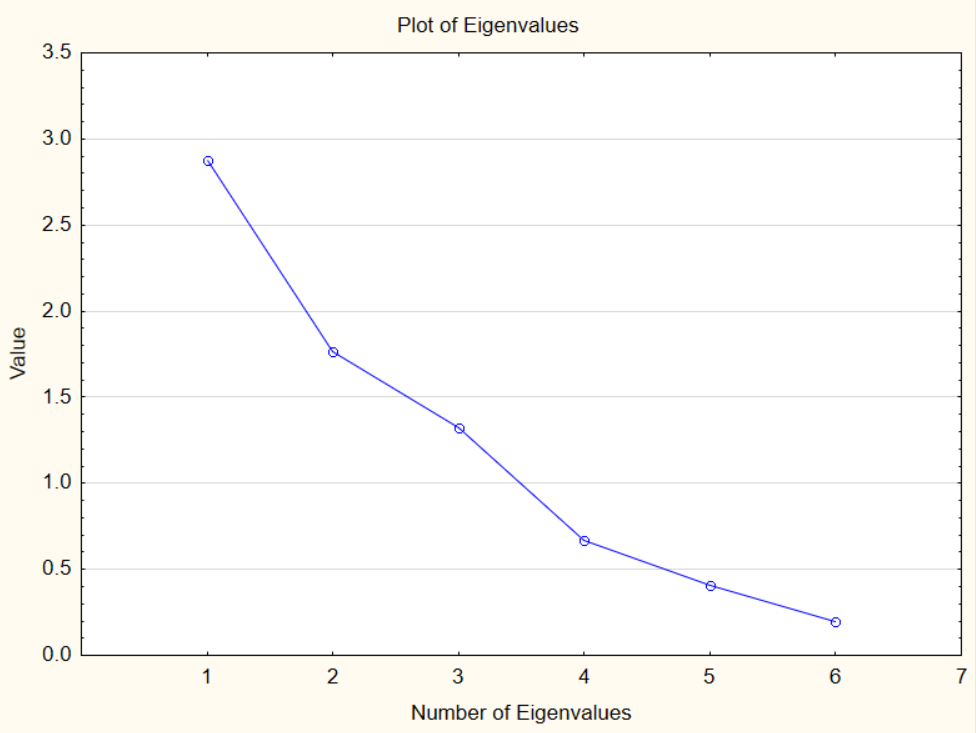


Рисунок 11.1 - График собственных значений (Statistica)

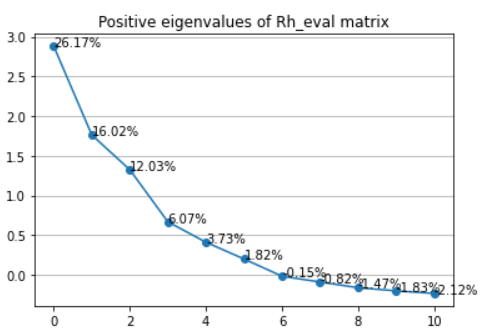


Рисунок 11.2 - График собственных значений (Python)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 12.1 - Весовые коэффициенты при общих факторах (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 12.2 - Весовые коэффициенты при общих факторах (Python)

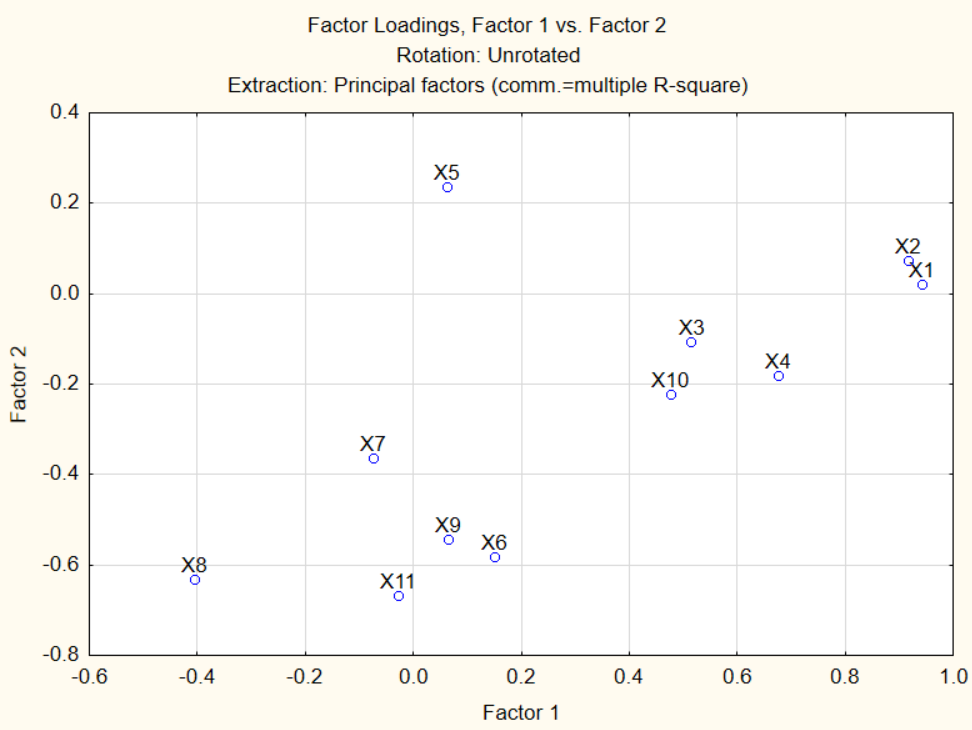


Рисунок 13.1 - Расположение исходных признаков на плоскости, образованной главными факторами (Statistica)

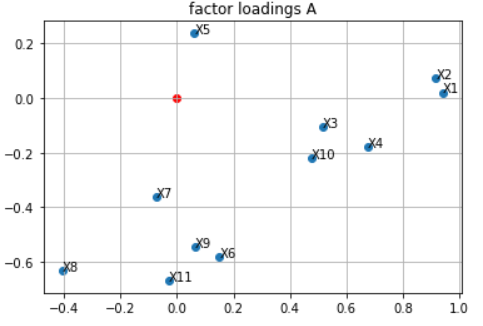


Рисунок 13.2 - Расположение исходных признаков на плоскости, образованной главными факторами (Python)

Попробуем упростить структуру главных факторов с помощью ортогонального вращения.

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 14 - Весовые коэффициенты при общих факторах после вращений

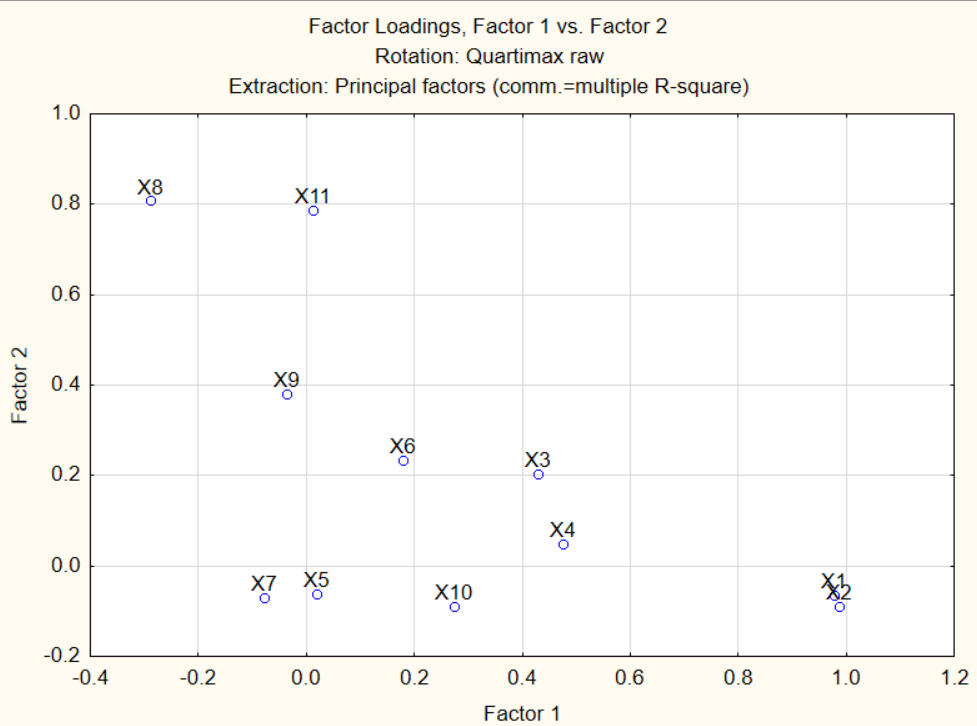


Рисунок 15.1 - Расположение исходных признаков на плоскости, образованной обобщенными факторами (Statistica)

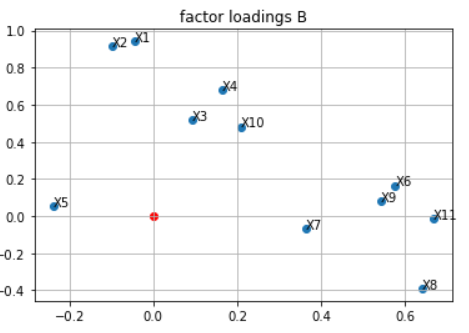


Рисунок 15.2 - Расположение исходных признаков на плоскости, образованной обобщенными факторами (Python)

Матрица факторных нагрузок после вращения имеет вид:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | F1 | F2 | F3 | F4 |
| X1: ВРП (млн руб на 1 000 человек населения) | 0.977854 | -0.064582 | 0.028509 | 0.096365 |
| X2: инвестиции в основной капитал (в фактически действовавших ценах; млн руб на 1 000 человек населения) | 0.986000 | -0.088152 | 0.016323 | -0.000940 |
| X3: число предприятий и организаций (на конец года на 1 000 человек населения) | 0.429130 | 0.205206 | -0.301341 | 0.372098 |
| X4: уровень занятости для людей в возрасте 15-72 лет (процентов) | 0.476239 | 0.048743 | -0.084391 | 0.647298 |
| X5: коэффициенты миграционного прироста на 10 000 человек населения | 0.019969 | -0.062667 | -0.391731 | 0.080679 |
| X6: зарегистрировано преступлений особой тяжести (на 1 000 человек населения) | 0.179069 | 0.234044 | 0.703923 | 0.064065 |
| X7: зарегистрировано преступлений, связанных с незаконным оборотом наркотиков (на 1 000 человек населения) | -0.077390 | -0.069626 | 0.770746 | 0.001061 |
| X8: общие коэффициенты смертности (число умерших на 1000 человек населения) | -0.289131 | 0.807958 | 0.050745 | -0.120416 |
| X9: мощность амбулаторно-поликлинических организаций всего, тыс. посещений в смену (на конец года, на 1 000 человек населения) | -0.036474 | 0.381241 | 0.235280 | 0.388375 |
| X10: численность врачей всех специальностей, всего, тыс. человек (на конец года, на 1 000 человек населения) | 0.274460 | -0.089650 | 0.226545 | 0.578515 |
| X11: численность пенсионеров, тыс. человек (на 1 000 человек населения) | 0.012594 | 0.787305 | 0.012377 | 0.138887 |

Первый главный фактор тесно связан с X1: ВРП (млн руб на 1 000 человек населения) и X2: инвестиции в основной капитал (в фактически действовавших ценах; млн руб на 1 000 человек населения). Поэтому первый фактор можно интерпретировать как «Показатель экономической активности».

Второй главный фактор тесно связан с X8: общие коэффициенты смертности (число умерших на 1000 человек населения) и X11: численность пенсионеров, тыс. человек (на 1 000 человек населения). Его можно интерпретировать как «Влияние возрастного состава населения на уровень смертности».

Третий главный фактор тесно связан с X6: зарегистрировано преступлений особой тяжести (на 1 000 человек населения) и X7: зарегистрировано преступлений, связанных с незаконным оборотом наркотиков (на 1 000 человек населения). Его можно интерпретировать как «Преступность».

Четвертый главный фактор связан с X4: уровень занятости для людей в возрасте 15-72 лет (процентов) и X10: численность врачей всех специальностей, всего, тыс. человек (на конец года, на 1 000 человек населения). Его можно интерпретировать как «Обеспеченность врачами».

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 15 - Вклады главных факторов в дисперсию признаков

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16.1 – оценка редуцированной матрицы (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 16.2 – оценка редуцированной матрицы (Python)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 17.1 - Оценка остаточной матрицы парных коэффициентов корреляции (Statistica)

Изображение выглядит как текст, окно

Автоматически созданное описание

Рисунок 17.2 - Оценка остаточной матрицы парных коэффициентов корреляции (Python)

Была рассчитана матрица индивидуальных значений обобщенных факторов (Рисунок 18).

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 18.1 - Индивидуальные значения обобщенных факторов (Statistica)

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 18.2 - Индивидуальные значения обобщенных факторов (Python)

