# ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ

**ФАКТОРНЫЙ АНАЛИЗ**

**МЕТОД ГЛАВНЫХ КОМПОНЕНТ**

**Вариант 18.**

1. Для относительных приращений, приведенных в варианте задания данных (цены закрытия акций на ММВБ за период с 01.01.2015 по 01.09.2015 с периодичностью 1 день: файл «Данные» лист «Котировки»), выяснить целесообразность применения факторного анализа (проверить гипотезу о зависимости рассматриваемых факторов – относительных приращений котировок ценных бумаг).
2. Провести факторный анализ относительных приращений методом главных компонент. Определить количество главных компонент, ориентируясь на долю выделенной дисперсии, а также руководствуясь критериями Кайзера и Кэттелла (каменистой осыпи).
3. Найти оценки главных компонент, упорядоченные по убыванию дисперсий (получить оценки векторов факторных нагрузок, оценки векторов значений обобщенных факторов по всем наблюдениям, записать выражения для обобщенных факторов и разложение для исходных признаков через обобщенные и характерные факторы). Оценить долю общей дисперсии, объясняемой каждым обобщенным фактором и всей совокупностью обобщенных факторов для каждого исходного признака. Оценить дисперсии характерных факторов.
4. Провести интерпретацию первых двух факторов, проведя при необходимости вращение факторов (в пространстве двух факторов), определив угол поворота приближенно графически или численно, используя метод варимах или квартимакс. Указать коэффициент информативности определяющих признаков для этих факторов.

**Задание 1.**

Посчитаем относительные приращения в Excel и выборочную дисперсию для каждого наблюдения . После импортируем полученный результат с помощью функции Import[].

*X=First[Import["import\_data.xlsx"]];*

*s2={};*

Посчитаем выборочную матрицу ковариации *S* для исходных данных.

Получим:

Воспользуемся критерием для проверки гипотезы о независимости компонент вектора . В случае получаем:

Вычислив получим: . Поправочный коэффициент в данном случае будет равен: . Таким образом, наблюдаемое значение статистики равно: . Достигнутый уровень значимости со степенями свободы в данном случае получается: Получаем, что гипотеза о независимости рассматриваемых факторов отвергается.

**Задание 2.**

Отцентрируем данные для дальнейшей работы с ними:

*XX=Standardize[X, Mean, 1&];*

Найдем собственные значения и собственные векторы выборочной матрицы ковариаций, упорядоченные по убыванию собственных значений. Для этого воспользуемся функциями Eigenvalues[], Eigenvectors[}, получим спектр собственных значений, упорядоченный по убыванию.

Нормируем собственные векторы выборочной матрицы ковариации, чтобы они удовлетворяли условию . Получим:

Определим векторы факторных нагрузок следующим образом :

**Задание 3.**

Определим количество главных компонент, которые следует оставить в факторной модели.

1. По доле выделенной дисперсии

Таблица 1. Доли дисперсии компонент

|  |  |  |
| --- | --- | --- |
| Номер компоненты | Доля дисперсии, объясняемой компонентой | Доля дисперсии совокупности компонент |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  | 1 |

Пусть q = 0.7, тогда количество главных компонент будет равняться 3.

1. Критерий Кайзера

Критерий Кайзера предполагает использование для нахождения оценок собственных значений выборочной матрицы корреляций (можно применять и в случае использования матрицы ковариаций, следует лишь в этом случае нормировать (умножить) все собственные значения на величину ). Согласно данному критерию, оставляют только те главные компоненты, дисперсия которых больше 1.

Следовательно, количество главных компонент в данном случае будет равняться 2.

1. Критерий каменистой осыпи

Критерий каменистой осыпи состоит в поиске точки, где убывание собственных значений замедляется наиболее сильно. Справа от этой точки должна находится, по-видимому, только "факторная осыпь" ("осыпь" – это геологический термин для обломков, которые скапливаются в нижней части каменистого склона). Таким образом, число выделенных факторов не должно превышать количество факторов слева от этой точки.

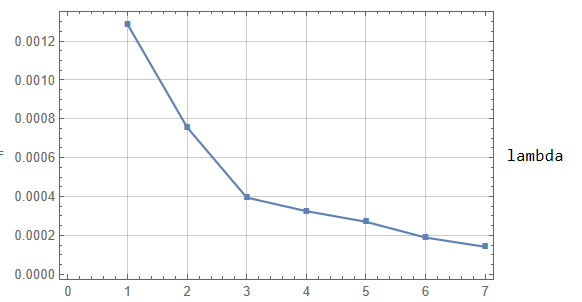


Рисунок 1 – Визуализация критерия каменистой осыпи.

Исходя из рисунка 1, количество главных компонент можно выбрать равное 2.

**Задание 4.**

Количество главных компонент будет выбрано три. Запишем выражения для обобщенных факторов:

f1 = β1(1)ξ1 +β2(1)ξ2 + β3(1)ξ3 + β4(1)ξ4 + β5(1)ξ5 + β6(1)ξ6 + β7(1)ξ7

f2 = β1(2)ξ1 +β2(2)ξ2 + β3(2)ξ3 + β4(2)ξ4 + β5(2)ξ5 + β6(2)ξ6 + β7(2)ξ7

f3 = β1(3)ξ1 +β2(3)ξ2 + β3(3)ξ3 + β4(3)ξ4 + β5(3)ξ5 + β6(3)ξ6 + β7(3)ξ7

f1 = ξ1 +ξ2 + ξ3 + ξ4 + ξ5  ξ6 + ξ7

f2 = ξ1 ξ2 + ξ3 + ξ4 + ξ5 + ξ6 + 2ξ7

f3 = ξ1 ξ2 + ξ3 ξ4 ξ5 + ξ6 ξ7

Разложим исходные признаки через обобщенные факторы:

ξ1 = α(1)1 f(1) + α(2)1 f(2) + α(3)1 f(3) + ε1

ξ2 = α(1)2 f(1) + α(2)2 f(2) + α(3)2 f(3) + ε2

ξ3 = α(1)3 f(1) + α(2)3 f(2) + α(3)3 f(3) + ε3

ξ4 = α(1)4 f(1) + α(2)4 f(2) + α(3)4 f(3) + ε4

ξ5 = α(1)5 f(1) + α(2)5 f(2) + α(3)5 f(3) + ε5

ξ6 = α(1)6 f(1) + α(2)6 f(2) + α(3)6 f(3) + ε6

ξ7 = α(1)7 f(1) + α(2)7 f(2) + α(3)7 f(3) + ε7

**Задание 5.**

Оценим долю дисперсии , объясняемая факторами.

Таблица 1. Доля дисперсии , объясняемая факторами.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Исходный признак | Доля дисперсии, объясняемая фактором f1 | Доля дисперсии, объясняемая фактором f2 | Доля дисперсии, объясняемая фактором f3 | Доля дисперсии, объясняемая факторами |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Оценим долю дисперсии характерных факторов εi.

Таблица 2. Долю дисперсии εi

|  |  |
| --- | --- |
| Исходный признак | Доля дисперсии εi, объясняемая факторами |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Исходя из результатов, полученных в таблице 1, фактор f2 объясняет долю дисперсии только 5-го признака. Следовательно, не коррелирует с остальными признаками. Таким образом, при вращении факторов коэффициенты информативности улучшить не получится. Изобразим исходные признаки в системе координат первых двух факторов на рисунке 1.

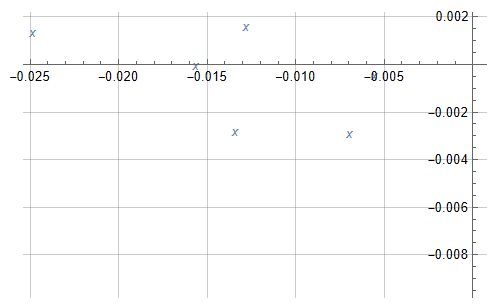


Рисунок 1. Изображение исходных признаков в системе координат первых двух факторов.

Рассчитаем коэффициенты информативности определяющих признаков для этих факторов.

Коэффициент информативности первого фактора для признаков , ,, равен .

Коэффициент информативности второго фактора для признаков , ,, равен .

Коэффициент информативности третьего фактора для признаков , ,, равен