# ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ

## МЕТОД ГЛАВНЫХ КОМПОНЕНТ

**Вариант 18.**

1) Для 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 метода главных компонент (проверить гипотезу о зависимости рассматриваемых факторов).

2) Найти собственные значения и собственные векторы выборочной матрицы ковариаций, упорядоченные по убыванию собственных значений. Оценить долю общей дисперсии объясняемой каждой главной компонентой и совокупностью главных компонент (двух, трех и т.д.)

3) Определить количество главных компонент, ориентируясь на: долю выделенной дисперсии, критерий Кайзера, 3правило сломанной трости, критерий Кэттелла (каменистой осыпи).

4) Записать выражения для главных компонент (выписать в явном виде линейные комбинации, определяющие компоненты), а также получить оценки векторов значений главных компонент по всем наблюдениям.

5) Определить какие признаки вносят наибольший вклад в каждую главную компоненту и, соответственно, долю этого вклада в компоненту.

**Задание 1.**

Импортируем исходные данные из файла с заданием с помощью функции Import[] и также вычислим в Exсel выборочную дисперсию для каждого наблюдения :

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

*s2 = {516224.2474, 152.8577908, 36344.70917, 322.7734214, 180314.4087, 439.7225455, 19.73503143};*

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.**

Нормализуем и отцентрируем исходные данный для дальнейшей работы с ними:

*XX=Standardize[X];*

Найти 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.

Оценить долю общей дисперсии объясняемой каждой главной компонентой:

Доля дисперсии совокупностью компонент (двух, трех):

**Задание 3.**

Определим количество главных компонент, ориентируясь на: долю выделенной дисперсии, критерий Кайзера, правило сломанной трости, критерий Кэттелла (каменистой осыпи).

1. По доле выделенной дисперсии

Пусть q = 0.7, тогда количество главных компонент будет равняться 2.

1. Критерий Кайзера

Так как исходные данные были нормированы, то условие отбора главных компонент выглядит следующим образом:

То есть, 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.

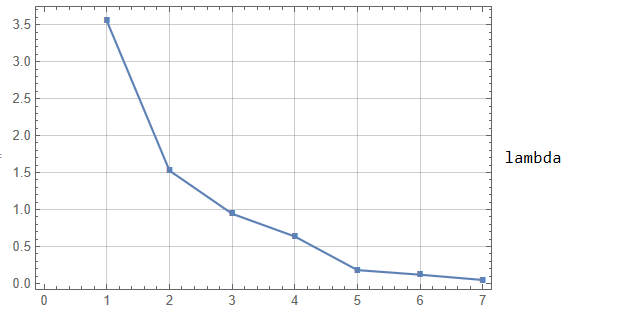


Рис. 1. Визуализация критерия каменистой осыпи.

Исходя из рисунка 1, количество главных компонент можно выбрать 1 или 4.

1. Правило сломанной трости.

Набор нормированных собственных чисел (, ) сравнивается с распределением длин обломков трости единичной длины, сломанной в случайно выбранной точке (точки разлома выбираются независимо и равномерно распределенными по длине трости). Пусть () - длины полученных кусков трости, занумерованные в порядке убывания длины: . Тогда математическое ожидание : . По правилу сломанной трости -й собственный вектор (в порядке убывания собственных чисел ) сохраняется в списке главных компонент, если .

Нормированные собственные числа имеют следующий вид:

=

Математическое ожидание:

Следовательно, согласно критерию, количество главных компонент в данном случае будет равняться 1.

**Задание 4.**

Количество главных компонент будет выбрано две. Запишем выражения для выбранных компонент:

η1 = β1(1)ξ1 +β2(1)ξ2 + β3(1)ξ3 + β4(1)ξ4 + β5(1)ξ5 + β6(1)ξ6 + β7(1)ξ7

η2 = β1(2)ξ1 +β2(2)ξ2 + β3(2)ξ3 + β4(2)ξ4 + β5(2)ξ5 + β6(2)ξ6 + β7(2)ξ7

η1 = ξ1 ξ2 ξ3 ξ4 + ξ5 )ξ6 ξ7

η2 = ξ1 ξ2 ξ3 + ξ4 + ξ5 + ξ6 ξ7

**Задание 5.**

Определим какие признаки вносят наибольший вклад в каждую главную компоненту.

Так как , то относительный вклад признака в дисперсию главной компоненты характеризует квадрат величины , то есть, квадрат соответствующей координаты вектора . Таким образом, для каждой компоненты отбираем признаки, которым соответствуют наибольшие абсолютные значения координат вектора , и определяем их суммарную долю вклада в компоненту, суммируя соответствующие квадраты координат.

Наибольший вклад в первую главную компоненту вносит признак

Наибольший вклад во вторую главную компоненту вносит признаки

Суммарная доля вклада в первую компоненту (суммируя соответствующие квадраты координат) равна

Суммарная доля вклада во вторую компоненту (суммируя соответствующие квадраты координат) равна