Лаб 2 ГК

в соответствии с темой 6

**6.** **СНИЖЕНИЕ РАЗМЕРНОСТИ МНОГОМЕРНОГО ПРИЗНАКА И ОТБОР**

**НАИБОЛЕЕ ИНФОРМАТИВНЫХ ПОКАЗАТЕЛЕЙ**

(см. слайды .ppt – электронный формат лекции по ММСМЭ, п. 6.2)

На основе формата данных (строки – наблюдения, объекты; столбцы – значения признаков или переменных или показателей), представленного формулой (6.3) и алгоритма вычисления главных компонент требуется получить их численные значения последовательным использованием формул (6.6), (6.7), (6.8), (6.11), (6.12), (6.14) и (6.16). Вместе с тем, поскольку в качестве самостоятельного задания предлагается освоение вычислительного алгоритма, полезно принимать во внимание ряд примечательных особенностей.

*Примечание 1.* Матрицу коэффициентов парной корреляции исходных переменных можно получить без использования формулы (6.7), например, путем запуска Excel в последовательности: Данные – анализ данных – корреляция.

*Примечание 2.* Так как алгоритмом предусмотрено применение нормированной матрицы , ее можно получить, например, в Excel-е с помощью функции . нормализуемая (стандартизируемая) переменная; среднее (арифметическое) значение переменной по наблюдаемым объектам (по столбцу); среднеквадратическое отклонение данной переменной. Если в формуле (6.7) при расчете корреляционной матрицы для усреднения используется не а (когда усреднение осуществляется по выборочной совокупности), то для корректности также следует рассчитать по формуле для выборочной совокупности:



Или же в обоих случаях формулой для генеральной совокупности (усреднение осуществляется по всему множеству объектов, т.е. через ).

Полезность формулы (6.7) состоит в том, что будет понятно как рассчитывается корреляционная матрица.

*Примечание 3.* Собственные значения (числа) корреляционной матрицы целесообразно определить в MatLab-е, где одновременно устанавливаются по одному собственному вектору, соответствующему собственному значению. Эти векторы нормированы к единице, т.е. их норма (длина) равна единице (сумма квадратов элементов вектора будет равна единице). Из этих векторов следует составить матрицу (см. формулу (6.13).

*Примечание 4.* Если исходные данные имеют формат (таблицы), то формат данных по вычисленным главным компонентам будет таким же. Исходные переменные могут иметь определенные единицы измерения, а новые переменные (главные компоненты или факторы) будут безразмерными в силу нормализации исходных переменных.

*Примечание 5.* Формулу (6.16) в силу того, что полученную матрицу следует транспонировать, можно записать так (для большей корректности):

Итак, следует найти численные значения главных компонент (их число равно числу исходных переменных) и проверить отсутствие попарной корреляции между ними по приведенным ниже трем вариантам (переменную не включать, брать только независимые переменные ).

Вариант 1

Имеются данные по консервному заводу Райсоюза одной из областей Беларуси за каждый месяц условного года. время, номера месяцев; фондовооруженность (млн. руб./ чел.); фондоотдача (млн. руб. продукции./млн. руб. основных фондов); производительность труда (туб./чел.); объем производства (туб). порядковый номер измерения; туб – тысяча условных банок.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1 | 1 | 328 | 0,054 | 0,3 | 397 |
| 2 | 2 | 329 | 0,101 | 0,6 | 670 |
| 3 | 3 | 329 | 0,099 | 1,2 | 1209 |
| 4 | 4 | 347 | 0,019 | 0,1 | 138 |
| 5 | 5 | 352 | 0,065 | 0,3 | 373 |
| 6 | 6 | 370 | 0,053 | 0,1 | 79 |
| 7 | 7 | 378 | 0,178 | 2,3 | 1883 |
| 8 | 8 | 385 | 0,174 | 2,6 | 2124 |
| 9 | 9 | 396 | 0,298 | 5,5 | 5069 |
| 10 | 10 | 399 | 0,195 | 2,4 | 2618 |
| 11 | 11 | 390 | 0,102 | 1,6 | 1265 |
| 12 | 12 | 373 | 0,138 | 0,6 | 562 |

Вариант 2

Имеются данные о продолжительности жизни по 10 странам и влияющим на этот показатель четырем показателям

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1 | 3 | 2,6 | 2,4 | 113 | 47 |
| 2 | 2,3 | 2,6 | 2,7 | 98 | 49 |
| 3 | 2,6 | 2,5 | 2,5 | 117 | 48 |
| 4 | 4,3 | 2,5 | 2,4 | 91 | 55 |
| 5 | 2,9 | 2,8 | 2,1 | 99 | 49 |
| 6 | 2,4 | 3,1 | 3,1 | 89 | 52 |
| 7 | 5,1 | 1,6 | 2,1 | 79 | 58 |
| 8 | 3,4 | 2 | 1,7 | 72 | 57 |
| 9 | 2 | 2,9 | 2,7 | 123 | 50 |
| 10 | 4,5 | 2,9 | 2,8 | 80 | 53 |

Вариант 3

Имеются данные о продолжительности жизни по 10 странам и влияющим на этот показатель четырем показателям

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1 | 7,4 | 3,1 | 4 | 46 | 68 |
| 2 | 7,4 | 2,8 | 2,7 | 73 | 59 |
| 3 | 4,9 | 3,1 | 2,8 | 124 | 47 |
| 4 | 8,3 | 2,9 | 3,3 | 90 | 60 |
| 5 | 5,7 | 2,5 | 2,7 | 96 | 51 |
| 6 | 7,5 | 2,4 | 2,2 | 55 | 57 |
| 7 | 7 | 3 | 3,8 | 45 | 67 |
| 8 | 10,8 | 1,1 | 1,1 | 34 | 69 |
| 9 | 7,8 | 2,9 | 3,1 | 56 | 57 |
| 10 | 7,6 | 2,9 | 2,6 | 90 | 51 |

В 1 – первые 6 студентов из списка, В2- вторые 6, В3 – последние 6 из списка.