

ОП «Политология», 2022-23**Введение в МСА****Иерархический кластерный анализ (13 января)***Д. В. Сальникова, А. А. Тамбовцева*

Задача 1. Перед вами небольшая таблица, содержащая данные по четырём респондентам (x – число килограммов, набранных за новогодние праздники, y – среднее число часов, проведённых на свежем воздухе):

id	x	y
1	0	4
2	7	0
3	1	2
4	4	2

- (a) Постройте диаграмму рассеивания для представленного набора данных.
- (b) Реализуйте иерархический кластерный анализ по следующему алгоритму.
 - i. Вычислите средние значения \bar{x} и \bar{y} , стандартные отклонения s_x и s_y .
 - ii. Выполните нормировку предложенных данных, вычитая из каждого значения среднее значение по столбцу и деля результат на стандартное отклонение столбца¹.
 - iii. Используя в качестве метрики манхэттенское расстояние, запишите матрицу расстояний для преобразованного набора данных.
 - iv. Используя в качестве метода агрегирования метод дальнего соседа, реализуйте иерархический кластерный анализ и постройте дендрограмму.

Задача 2. Какие из перечисленных ниже матриц могут быть матрицами расстояний? Обоснуйте свой ответ.

$$A = \begin{pmatrix} 0 & 2 & 3 \\ 3 & 0 & 5 \\ 3 & 5 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 1 & 5 \\ 4 & 5 & 0 \end{pmatrix} \quad C = \begin{pmatrix} 0 & 1.5 & 6 \\ 1.5 & 0 & 5 \\ 6 & 5 & 0 \end{pmatrix} \quad D = \begin{pmatrix} 0 & -1.5 & 6 \\ -1.5 & 0 & 5 \\ 6 & 5 & 0 \end{pmatrix}$$

¹для удобства стандартное отклонение можно округлить до целых