Лабораторная работа №1

Тема: Определение однородности данных. Метод шаров.

Цель работы: Получить навыки определения однородности данных путем разбиения исходного набора на однородные изотропные подмножества.

Методические указания к выполнению работы

Реализации многомерных случайных величин являются объектами таксономических исследований. Потому их рассматривают как векторы или точки, расположенные во многомерном пространстве, которое задается рассмотренной совокупностью признаков. Эти объекты различаются как за уровнем значений их описывающихся признаков, так и за структурой значений этих признаков, т.е. по пропорциям их значений. Чаще всего используются такие таксономические методы, с помощью которых выделяются изотропные, изотонические или изоморфические подмножества. Пусть в результате измерений получена совокупность значений x_{ij}, $i = \overline{1, N}$, $j = \overline{1, n}$, N – количество объектов, n – количество измеряемых признаков.

Подмножества, однородные в содержании изотропности, состоят из объектов, которые отличаются друг от друга за уровнем и структурой значений признаков. При этом избавиться от единиц измерения признаков можно путем так называемой стандартизации:

$$y_{ij} = \frac{x_{ij} - \overline{x_{ij}}}{\sigma_j}, \tag{1.1}$$

где $\overline{x_{ij}}$ – математическое ожидание величины X, σ_j – среднеквадратическое отклонение.

Элементы матрицы расстояний D определяются по формулам:

$$d_{ik} = \sqrt{\sum_{j=1}^{n} \left(y_{ij} - y_{kj} \right)^2} \,. \tag{1.2}$$

Совокупность операций, приводящих к получению подмножеств данных, которые состоят из элементов, однородных за уровнем значений признаков, будем называть изотоническим анализом, а полученные подмножества данных – изотоническими подмножествами.

Избавиться от структурного компонента можно с помощью превращения:

$$v_{ij} = \frac{x_{ij}}{\sum\limits_{i=1}^{N} x_{ij}} \,. \tag{1.3}$$

Заметим, что длины векторов $\mathbf{v}_i$ характеризуются тем, что их нормы равны единице. Благодаря проведенному превращению в компонентах полученных векторов остался лишь признак, характеризирующий уровень значений свойств данного объекта в рассмотренной совокупности. Далее по полученным векторам $\mathbf{v}_i$ определяются изотонические показатели:

$$w_i = \sum_{j=1}^{n} v_{ij} . \quad (1.4)$$

По этим показателям определяется матрица расстояний D между объектами с компонентами:

$$d_{ij} = \left| w_i - w_k \right| . \quad (1.5)$$

Чем высшее значение d_{ij}, тем больше отличаются между собой рассмотренные объекты за уровнем значений свойств. Низкие значения d_{ij} свидетельствуют о малых расхождениях объектов за уровнем значений свойств.

Если объекты, входящие в состав подмножеств, будут иметь близкие структуры значений признаков (пропорции значений соответствующих признаков будут мало отличаться друг от друга), то такие подмножества называются изоморфическими. Изоморфическое превращение осуществляется путем нормирования векторов $\mathbf{v}_i$, полученных в предыдущем методе, на их длину:

$$z_{ij} = \frac{\dfrac{x_{ij}}{\sum_{i=1}^{N} x_{ij}}}{\sum_{j=1}^{n} \dfrac{x_{ij}}{\sum_{i=1}^{n} x_{ij}}} = \frac{v_{ij}}{\sum_{j=1}^{n} v_{ij}} . \quad (1.6)$$

Нормы векторов $\mathbf{z}_i$ равны единице. Для этих векторов вычисляется матрица расстояний D, компоненты которой определяются по формулам:

$$d_{ik} = \sqrt{\sum_{j=1}^{n} \left(z_{ij} - z_{kj} \right)^2} . \quad (1.7)$$

Можно сказать, что наименьшее значение расстояния между объектами, координаты которых преобразованы по формуле (1.7), наблюдается в том случае, если векторы $\mathbf{z}_i$, представляющие данные объекты, коллинеарные. Другими словами, объекты $\mathbf{z}_i$ имеют идентичную структуру значений признаков. Наибольшее же значение расстояния между объектами достигается при наличии перпендикулярных векторов (объектов с максимально разной структурой значений признаков). Чем меньше различаются объекты за структурой значений признаков, тем меньше расстояние между объектами, рассчитанное по формуле (1.7). И, наоборот, большая величина расстояния между объектами свидетельствует о большем расхождении пропорций значений признаков, описывающих данные объекты. Одним из методов получения однородных подмножеств есть метод шаров. Суть его заключается в следующем:

1. Выполняются предыдущие превращения исходных значений x_{ij} по одному из описанных способов.
2. Строится соответствующая матрица расстояний.
3. Для каждого столбца матрицы расстояний находится минимальный элемент (элементы, расположенные на диагонали, равняется 0, и на данном шаге в расчеты не включаются). Дальше среди этих элементов находится наибольший элемент. Он и задает радиус шаров, которыми совокупность объектов будет разделяться на подмножества.
4. Для каждого столбца матрицы расстояний подсчитывается количество значений, меньших радиуса шара. Среди этих чисел выбирается наибольшие. Тогда соответствующий номер столбца определяет центр шара, а номера значений – точки, которые вошли в этот шар (в данное подмножество).
5. Для значений, которые остались, строится новая матрица расстояний путем вычеркивания строк и столбцов исходной матрицы расстояний для точек, которые вошли в первое подмножество.
6. Пункты 4 и 5 продолжаются, пока не будут получены все подмножества.

Задание для самостоятельного выполнения

В результате проведения эксперимента по изучению 2-х свойств было обследовано N объектов. Результаты измерений свойств представлены в таблице 1.1. Необходимо провести

1) изотропное;
2) изотоническое;
3) изоморфическое

превращение полученной совокупности данных. По методу шаров разбить исследуемую совокупность объектов на подмножества.

Как результат представить:

1) матрицу расстояний, полученную для каждого из видов превращения;
2) критическое значение радиуса шара и совокупность полученных подмножеств.

Выходные даны приведенные в таблице 1.1.

Таблица 1.1

Вариант 1		**Вариант 2**		**Вариант 3**		**Вариант 4**		**Вариант 5**		**Вариант 6**	
X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
1,2	3,2	1,4	2,2	1,2	4,2	1,4	4,2	1,2	3,2	1,2	3,7
1,5	3,8	1,5	2,3	1,3	4,7	1,5	4,4	1,5	3,8	1,3	3,8
1,6	4	1,6	2,9	1,4	4	1,6	4,8	1,6	4	1,6	4
1,8	2,1	1,8	2,3	1,5	3,1	1,8	5,1	1,8	2,1	1,8	2,1
2,1	2,3	2,1	2,6	1,7	3,3	2,1	6,8	2,1	2,3	2,1	2,3
2,2	4,6	2,2	2,7	2,2	4,6	2,2	6,6	2,2	4,6	2,2	4,6

2,3	4,5	2,3	4,7	2,3	7,5	3,3	7,8	2,3	4,5	2,3	4,5
2,6	8,2	2,6	5,9	2,6	8,2	3,6	4,9	2,6	8,2	2,6	8,2
2,7	5,2	2,7	5,8	2,7	9,3	3,8	5,2	2,7	5,2	2,7	5,2
2,9	5,3	2,9	5,3	2,9	10,5	4,2	5,3	2,9	5,3	2,9	6
Вариант 7		**Вариант 8**		**Вариант 9**		**Вариант 10**		**Вариант 11**		**Вариант 12**	
X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
2,1	4,1	2	4	3	4	4,2	8,1	2,4	5,3	1	2,4
2,3	4,8	2,1	4,1	3,2	3,8	4,5	8,8	2,6	5,8	1,2	2,5
2,4	4,9	2,5	5	3,5	3,8	1,6	4	2,9	6	1,3	2,6
2,8	6,2	2,8	5,5	3,6	4	1,8	2,1	1,8	2,1	1,4	2,6
4,1	8,1	4,2	7,1	3,7	4,4	2,1	2,3	2,1	2,3	2,1	3,9
4,4	8,6	4,6	7,7	5,6	6,7	2,2	4,6	2,2	4,6	2,2	4,1
5,2	9,1	4,7	7,6	5,7	7,2	2,3	4,5	2,3	4,5	2,3	4,5
5,6	10	4,9	8,2	5,9	7	2,6	8,2	2,6	8,2	2,6	4,6
5,7	10,2	6,1	9	6	7,2	2,7	5,2	5,7	7,2	2,7	5,2
6,9	13,1	6,6	10,2	6,2	7,7	2,9	5,3	5,9	7,4	2,9	5,5
Вариант 13		**Вариант 14**		**Вариант 15**		**Вариант 16**		**Вариант 17**		**Вариант 18**	
X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
3,2	3,1	1,1	2,7	1,4	2,2	2,1	4	2	4	3	4
3,3	3,3	1,2	2,7	1,5	2,6	2,3	3,8	2,1	4,1	3,2	4,1
3,6	3,5	1,5	3,1	1,6	2,9	2,4	3,8	2,5	5	3,5	5
3,8	3,9	1,6	3,2	1,8	3,2	2,8	4	2,8	5,5	3,6	5,5
3,7	3,8	6,2	6,7	2,1	3,8	4,1	4,4	4,2	7,1	3,7	7,1
4,4	4,1	6,5	8,1	2,2	4,9	4,4	6,7	4,6	7,7	5,6	7,7
4,5	4,2	6,7	8	2,3	6,1	5,2	7,2	4,7	7,6	5,7	7,6
7,1	6,4	6,9	8,1	2,5	9	5,6	7	4,9	8,2	5,9	8,2
7,2	6,6	7,2	8,4	2,7	10,2	5,7	7,2	6,1	9	6	9
7,9	7	7,7	8,6	2,9	16,1	6,9	7,7	6,6	10,2	6,2	10,2
Вариант 19		**Вариант 20**		**Вариант 21**		**Вариант 22**		**Вариант 23**		**Вариант 24**	
X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
1,1	3,2	1,2	2,2	1,4	4,2	1,2	4,2	1,4	3,2	1,2	3,7
1,2	3,8	1,5	2,4	1,5	4,7	1,3	4,4	1,5	3,9	1,5	3,8
1,5	4,4	1,6	2,9	1,6	4	1,4	4,8	1,6	4	1,6	4,3
1,6	2,1	1,8	2,3	1,8	3,1	1,5	5,1	1,8	2,1	1,8	2,1
6,2	2,3	2,1	2,8	2,1	3,3	1,7	6,8	2,1	2,3	2,1	2,3

6,5	4,6	2,2	2,7	2,2	4,6	2,2	6,6	2,2	4,6	2,2	4,7
6,7	4,5	2,3	4,7	2,3	7,5	2,3	7,8	3,3	4,5	2,3	4,5
6,9	8,2	2,6	5,9	2,6	8,2	2,6	4,9	3,6	8,2	2,6	8,2
7,2	5,1	2,7	5,8	2,7	9,3	2,7	5,2	3,8	5,2	2,7	5,2
7,7	5,3	2,9	5,3	2,9	10,7	2,9	5,3	4,2	5,3	2,9	6
Вариант 25		**Вариант 26**		**Вариант 27**		**Вариант 28**		**Вариант 29**		**Вариант 30**	
X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
2,1	4,2	2,4	4	3	4	4,2	8,1	2,4	5,3	1	2,4
2,3	4,4	2,1	4,1	3,2	3,6	4,5	8,6	2,3	5,6	1,2	2,5
2,4	4,9	2,5	5	3,5	3,9	1,6	3,7	2,9	6	1,3	2,6
2,8	6,2	2,8	5,5	3,5	4	1,8	2,2	1,8	2,1	1,5	2,4
4,2	8,1	4,2	7,3	3,7	4,5	2,1	2,3	2,1	2,3	2,1	3,9
4,4	8,4	4,6	7,5	5,6	6,6	2,2	4,6	2,2	4,6	2,2	4,1
5,2	9,1	4,7	7,6	5,7	7,4	2,3	4,5	2,3	6,5	2,4	4,5
5,5	10	4,8	8,2	5,9	7,2	2,6	8,2	2,6	8,2	2,8	4,6
5,7	10,2	6,1	8,6	6	7,5	2,7	5,2	5,7	8,1	2,8	5,2
6,9	12,1	6,6	10,3	6,2	7,3	2,9	5,3	5,9	8,4	2,9	5,5

Рекомендуемая литература

1. Вознесенский В.А. Статистические методы планирования эксперимента в технико-экономических исследованиях / В.А. Вознесенский. – М. : Статистика, 1974. – 192 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. Уч. пособие для втузов. – М.: Высш. школа, 2002. – 479 с.
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. Уч. пособие для втузов. – М.: Высш. Школа, 2002. –ы 400 с.
4. Жалдак М.И. Теория вероятностей с элементами информатики. Практикум: Уч. Пособие / М.И. Жалдак, А.Н. Квитко / Под общ. ред. М.И. Ядренко. – К.: Вища шк., 1989. – 263 с.
5. Советов Б.Я. Моделирование систем : учеб. для вузов / Б.Я. Советов, С.А. Яковлев. – М. : Высш. шк., 2001. – 343 с.
6. Советов Б.Я. Моделирование систем: практикум : учеб. пособие для студентов вузов. Изд. 2-е, перераб. и доп. / Б. Я. Советов, С. А. Яковлев. – М.: Высшая школа, 2003. – 295 с.
7. Тарасова П.В. Введение в математическое моделирование: учеб. пособие для вузов / под ред. П.В. Тарасова. – М.: Интермет Инжиниринг, 2000. – 200 с.