

## Лабораторная работа №4

**Тема:** Построение синтетических и индивидуальных диагностических признаков.

**Цель работы:** Получить навыки определения таксономического показателя и выделения диагностических признаков методом потенциалов и методом центра тяжести.

### Методические указания к выполнению работы

Таксономический показатель является синтетической величиной, образованной из всех признаков, характеризующих исследуемое экономическое явление. При построении таксономического показателя применяется матрица данных:  $X = [x_{sn}]$ , где  $x_{sn}$  – стандартизированная реализация  $n$ -го свойства на  $s$ -м объекте.

Как известно, в выделенной совокупности могут встречаться стимуляторы, дестимуляторы и номинаторы. Их стоит привести к одному типу, например, превратить в стимуляторы. Дестимуляторы меняются на стимуляторы путем одного из следующих превращений:

$$\begin{aligned}x_{ij} &= 1 - y_{ij}, \\x_{ij} &= -y_{ij}, \\x_{ij} &= \frac{1}{y_{ij}},\end{aligned}\tag{4.1}$$

где  $y_{ij}$  –  $i$ -ая реализация  $j$ -го признака дестимулятора.

Номинаторы меняются признаками, имеющих близкую содержательную интерпретацию, но в то же время являющихся или стимуляторами, или дестимуляторами.

После выполнения указанных операций рассчитывается матрица  $D$  расстояний между точками-данными и точкой верхнего полюса:

$$D = [d_{i0}],\tag{4.2}$$

$$\text{где } d_{i0} = \left[ \sum_{j=1}^n (x_{ij} - x_{0j})^2 \right]^{1/2}, \quad i = \overline{1, t}.$$

Полученные расстояния служат исходными элементами при расчете таксономического показателя:

$$M^* = \frac{D}{\|D\|},\tag{4.3}$$

где  $\|D\|$  – норма  $D$ .

Значения этого показателя имеют такой вид:

$$\begin{aligned}
 m_{i0}^* &= \frac{d_{i0}}{c_0}, \\
 c_0 &= \bar{x} + \alpha s_d, \\
 \bar{x} &= \frac{1}{t} \sum_{i=1}^t d_{i0}, \\
 s_d &= \left[ \frac{1}{t} \sum_{i=1}^t (d_{i0} - \bar{x})^2 \right]^{1/2}.
 \end{aligned} \tag{4.4}$$

Величина  $\alpha$  – некоторое положительное число, избранное таким образом, чтобы значения показателя  $M^*$  изменялись в интервале от 0 до 1; чаще всего принимается, что  $\alpha = 2$  или  $\alpha = 3$ .

Интерпретация показателя  $M^*$  не сопоставляется с интуитивными представлениями (показатель растет с удалением значений стимуляторов от верхнего полюса и убывает с приближением к нему), потому его приводят к виду:  $M = i - M^*$ , где  $i$  – единичный вектор.

### Группирование исходных признаков и выбор репрезентантов

Редукция признаков, характеризующих исследуемое явление, должна отвечать некоторым требованиям. Главное заключается в сохранении достаточно полного описания явления.

С помощью соответствующих таксономических методов совокупность признаков разделяется на однородные подмножества сильно коррелируемых признаков (например, по методу шаров). Элементы матрицы расстояний рассчитываются по формуле:

$$M^* = \frac{D}{\|D\|}, \tag{4.5}$$

где  $c_{ij}$  – расстояние от  $i$ -го признака к  $j$ -ому признаку;  $r_{ij}$  – коэффициент корреляции между признаками  $i$  и  $j$ . Если далее определенным образом подобрать репрезентанты в каждой из избранных однородных подмножеств, то получим некоррелируемые или слабо коррелируемые признаки.

### Метод потенциалов

Первая операция заключается в нахождении наименьших чисел в каждой строке матрицы расстояний. С помощью найденных чисел  $c_{ip}$  определяют строку и столбец. Их номера записывают в таком же порядке. Первый номер указывает вершину, из которой дуга выходит, а второй – вершину, в которую дуга заходит. Так получают ориентированные дуги.

После выделения ориентированных скоплений определяют так званый потенциал вершины: число стрелок, которые заходят в вершину. Потенциал определяет ранг вершины. В каждом скоплении есть по крайней мере одна вершина с наибольшим потенциалом. Если же таких элементов немного, то из них выбирают один из наименьшего значения величины:

$$\bar{c}_i = \frac{1}{\omega} \sum_{p=1}^{\omega} c_{ip}, \quad (4.6)$$

где  $c_{ip}$  – расстояние  $i$ -го элемента с наибольшим потенциалом от элемента  $p$  в множестве;  $\omega$  – количество таких элементов.

Таким образом, в совокупность диагностических признаков войдет по одному элементу из каждого скопления.

### Метод центра тяжести

В методе центра тяжести способ выбора репрезентантов зависит от числа элементов, входящих в однородное подмножество. Различается три типа групп: одноэлементные, двухэлементные и другие.

Одноэлементные группы содержат такие признаки, значения которых резко отличаются от значений других признаков, поэтому их включают в диагностическую совокупность.

Если число элементов в группе больше двух, для каждого из них рассчитывают расстояние от других элементов той же группы:

$$d_{ik} = \sum_{\substack{j=1 \\ j \neq i}}^{n_k} c_{ij}, \quad (4.7)$$

где  $c_{ij}$  – расстояние к  $i$ -ому элементу от  $j$ -го элемента  $k$ -ой группы;  $n_k$  – число признаков в  $k$ -ой группе. В диагностическую совокупность включается признак, находящийся в “середине” данной группы, т.е. признак с наименьшей суммой расстояний:

$$d_{mk} = \min_i \{d_{ik}\}, \quad (4.8)$$

Процедура завершается выбором репрезентантов двухэлементных групп. Для этого рассчитываются расстояния каждого из двух элементов группы от диагностических признаков, избранных на предыдущих этапах:

$$d_i = \sum_{j=1}^s c_{ij}, \quad (4.9)$$

где  $i = \overline{1, 2k}$ ;  $k$  – число двухэлементных групп;  $2k$  – общее число признаков в двухэлементных группах;  $s$  – число одиночных признаков и признаков, избранных из многоэлементных групп.

Репрезентантом каждой из двухэлементных групп будет тот признак, расстояние которого от одиночных элементов и элементов, взятых из многоэлементных групп, окажется наибольшим:

$$d_{mk} = \min_i \{d_{ik}\}, \quad (4.10)$$

где  $i = \overline{1, 2}$ .

Избранные элементы владеют тем свойством, что находятся вблизи “центров тяжести” групп, и потому удовлетворяют основным требованиям, предъявленным к диагностическим признакам.

### Задание для индивидуального выполнения

Для заданного набора исходных данных необходимо построить синтетические и индивидуальные показатели, используя приведенные в указаниях методы. Для расчетов необходимо взять 20 строк таблицы 4.2, руководствуясь таблицей 4.1.

Таблица 4.1

Номер варианта	Диапазон рядков	
N	Первый	Последний
	N	19+N

Таблица 4.2

	Номера признаков						
	1	2	3	4	5	6	7
1	1,36	26,35	28,08	87,72	1270	25,48	200,85
2	1,45	28,56	32,94	144,48	1680	31,98	351
3	1,54	29,41	37,8	167,7	1900	31,2	351
4	1,47	27,88	42,12	168,99	1900	35,36	237,9
5	0,91	14,79	34,02	23,22	100	10,4	154,05
6	0,94	14,62	35,1	25,8	170	20,54	132,6
7	0,94	14,79	35,64	29,67	200	18,98	175,5
8	0,94	14,79	37,26	34,83	220	19,5	212,55
9	3,96	64,6	170,1	387	2180	50,7	813,15
10	3,95	62,39	179,28	434,73	2320	65,26	850,2
11	3,57	60,69	198,18	501,81	4400	106,08	542,1
12	3,83	60,18	234,36	553,41	4610	91,52	791,7
13	4,1	53,89	155,52	570,18	2980	83,46	1027,65
14	0,19	31,7	104,32	46,45	107	22,86	115,6
15	0,4	5,1	27	64,5	500	10,4	117
16	0,19	2,55	6,48	30,96	230	4,42	40,95

17	1,77	36,38	19,98	69,66	670	13	132,6
18	1,77	35,87	21,06	78,69	720	12,48	179,4
19	1,85	36,38	25,92	86,43	870	16,12	234
20	1,9	36,72	31,86	98,04	940	15,86	200,85
21	1,85	36,72	40,5	45,15	490	22,36	185,25
22	1,95	35,53	54,54	55,47	580	24,96	191,1
23	1,92	38,42	64,8	72,24	710	27,04	253,5
24	1,9	37,4	67,5	86,43	740	38,22	308,1
25	0,35	38,08	3,78	47,73	450	4,68	74,1
26	0,35	7,65	4,32	55,47	500	5,98	95,55
27	0,55	6,97	6,48	76,11	680	9,36	95,55
28	0,57	10,88	6,48	87,72	760	10,92	120,9
29	0,86	12,41	28,62	33,54	480	16,12	136,5
30	0,86	12,75	29,7	38,7	590	32,5	163,8
31	0,84	12,41	34,02	42,57	600	54,34	154,05
32	0,82	11,73	37,26	51,6	640	106,86	154,05
33	0,82	16,49	8,64	64,5	340	12,74	169,65
34	0,82	15,13	9,18	78,69	410	14,82	216,45
35	0,81	15,3	10,26	94,17	450	16,64	212,55
36	0,8	14,96	11,88	104,49	500	60,84	366,6
37	1,19	19,04	32,94	45,15	580	18,46	136,5
38	1,2	20,23	34,56	55,47	660	19,76	169,65
39	1,29	22,1	43,2	68,37	780	23,66	222,3
40	1,27	21,25	45,9	77,4	830	26	200,85

### 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 вузов. – М.: Высш. школа, 2002. – 479 с.
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. Уч. пособие для вузов. – М.: Высш. Школа, 2002. – 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. Ивченко Г.И. Математическая статистика: учеб. пособие для втузов / Г.И. Ивченко, Ю.И. Медведев. – М.: Высш. шк., 1984. – 248 с.

6. Методичні рекомендації до виконання лабораторних робіт з навчальної дисципліни "Моделювання систем" для студентів напрямку підготовки 0804 "Комп'ютерні науки" всіх форм навчання / укл. В. М. Задачин, І. Г. Конюшенко. – Харків : Вид. ХНЕУ, 2007. – 96 с.