

Задачи по теме:
«Дискриминантный анализ»
Вариант 3

По представленным данным с помощью дискриминантного анализа классифицировать объекты.

Задача 1

Таблица 1. Анализируемые данные

| № п/п | Группы регионов | Объем реализованной продукции растениеводства, млн. руб. | Объем реализованной продукции животноводства, млн. руб. |
|-------|----------------------------|---|--|
| 1 | Высокий уровень (X) | 25 | 21 |
| 2 | | 31 | 37 |
| 3 | | 27 | 22 |
| 4 | | 33 | 36 |
| 1 | Низкий уровень (Y) | 47 | 38 |
| 2 | | 50 | 67 |
| 3 | | 52 | 45 |
| 4 | | 39 | 49 |
| 5 | | 66 | 33 |
| 1 | Подлежат дискриминации (Z) | 32 | 42 |
| 2 | | 67 | 33 |
| 3 | | 46 | 56 |

Решение:

Определим векторы средних и оценки ковариационных X, Y матриц Sx и Sy соответственно для групп регионов высокого и низкого уровня.

Векторы средних: X_{ср. трансп.} = (29 29); Y_{ср. трансп.} = (50,8 46,4)

| | |
|----------------------|------|
| Матрица ковариации X | |
| 10 | 22,5 |
| 22,5 | 56,5 |

| | |
|----------------------|--------|
| Матрица ковариации Y | |
| 77,36 | -44,12 |
| -44,12 | 136,64 |

Для подсчета значений векторов и матриц, используются формулы:

$$s_{jl(x)} = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i^{(j)} - \bar{x}^{(j)})(x_i^{(l)} - \bar{x}^{(l)}),$$

$$\bar{x}^{(j)} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(j)}$$

Вектор (X_{ср.}-Y_{ср.}) трансп. = (-21,8 -17,4)

Далее найдем несмещенную оценку суммарной ковариационной матрицы по формуле:

$$\hat{S} = \frac{1}{n_1 + n_2 - 2} [n_1 \cdot S_x + n_2 \cdot S_y]$$

| Несмещенная оценка суммарной ковариационной матрицы | |
|---|--------------|
| 60,97142857 | -18,65714286 |
| -18,65714286 | 129,8857143 |

| Обратная матрица | |
|------------------|----------------|
| 0,01715517014 | 0,002464216036 |
| 0,002464216036 | 0,008053042911 |

Вычислим вектор оценок коэффициентов дискриминантной функции:

$$\hat{a} = \hat{S}^{-1} (\bar{X} - \bar{Y})$$

| | |
|----|---------------|
| a^ | -0,416860068 |
| | -0,1938428562 |

Теперь найдем оценки дискриминантной функции для обучающих выборок:

| U _x =X*a^ | U _y =Y*a^ |
|----------------------|----------------------|
| -14,49220168 | -26,95845173 |
| -20,09484779 | -33,83047477 |
| -15,51976467 | -30,39965207 |
| -20,73472507 | -25,75584261 |
| | -33,90957874 |

Получим средние значения оценок дискриминантной функции, а затем оценку константы дискриминации:

U_{хср} = -17,71; U_{уср} = -30,17

$$\hat{c} = \frac{1}{2} (\bar{U}_x + \bar{U}_y)$$

$$c^{\wedge} = \frac{1}{2} * (-17,71 + (-30,17)) = -23,94$$

Чтобы определить, к какой группе относятся регионы, подлежащие дискриминации, рассчитаем для них дискриминантную функцию:

U(z1) = 32*a1+42*a2=-21,48 > c^=-23,94 => Данный регион можно отнести к региону высокого уровня

U(z2) = -34,32 < c^=-23,94 => Данный регион можно отнести к региону низкого уровня

$U(z_3) = -30,03 < c^* = -23,94 \Rightarrow$ Данный регион можно отнести к региону низкого уровня

Задача 2

Имеются 12 фирм, характеризующиеся тремя показателями: производительность труда (млн. руб./чел.), удельный вес потерь заказов (%) и годовой рост заявок на техобслуживание сети (%). Из этих компаний выделены две обучающие выборки, первая из которых включает $n_1 = 4$ группы А, а вторая - $n_2 = 5$ группы В. Используя данные таблицы, с помощью дискриминантного анализа провести классификацию трех последних фирм (Z).

Таблица 2. Анализируемые данные

| № п/п | Группы компаний | Производительность труда, млн. руб./чел. | Удельный вес потерь заказов, % | Годовой рост заявок на техобслуживание сети, % |
|-------|----------------------------|--|--------------------------------|--|
| 1 | Группа А (X_1) | 9,4 | 0,15 | 1,91 |
| 2 | | 9,9 | 0,34 | 1,68 |
| 3 | | 9,1 | 0,09 | 1,89 |
| 4 | | 9,4 | 0,21 | 2,30 |
| 1 | Группа В (X_2) | 6,6 | 0,48 | 0,88 |
| 2 | | 4,3 | 0,41 | 0,62 |
| 3 | | 7,4 | 0,62 | 1,09 |
| 4 | | 6,6 | 0,50 | 1,32 |
| 5 | | 5,5 | 1,20 | 0,68 |
| 1 | Подлежат дискриминации (Z) | 5,5 | 0,05 | 1,02 |
| 2 | | 10 | 0,32 | 2,62 |

Решение:

Определим векторы средних и оценки ковариационных X_1 , X_2 матриц S_{x1} и S_{x2} соответственно для групп компаний, относящихся к группам компаний А и В.

Векторы средних: $X_{1ср. трансп.} = (9,45 \quad 0,1975 \quad 1,945)$;

$X_{2ср. трансп.} = (6,08 \quad 0,642 \quad 0,918)$

| Матрица ковариации X_1 | | |
|--------------------------|------------|------------|
| 0,0825 | 0,025875 | -0,029 |
| 0,025875 | 0,00856875 | -0,0064375 |
| -0,029 | -0,0064375 | 0,050125 |

| Матрица ковариаций X_2 | | |
|--------------------------|-----------|-----------|
| 1,1576 | -0,01956 | 0,21696 |
| -0,01956 | 0,082416 | -0,023676 |
| 0,21696 | -0,023676 | 0,067616 |

Для подсчета значений векторов и матриц, используются формулы:

$$s_{jl(x)} = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i^{(j)} - \bar{x}^{(j)})(x_i^{(l)} - \bar{x}^{(l)}),$$

$$\bar{x}^{(j)} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i^{(j)}$$

Вектор (X1ср.-X2ср.) трансп. = (3,37 -0,4445 1,027)

Далее найдем несмещенную оценку суммарной ковариационной матрицы по формуле:

$$\hat{S} = \frac{1}{n_1 + n_2 - 2} [n_1 \cdot S_x + n_2 \cdot S_y]$$

| Несмещенная оценка суммарной ковариационной матрицы | | |
|---|-----------------|----------|
| 0,874 | 0,0008142857143 | 0,1384 |
| 0,0008142857143 | 0,063765 | -0,02059 |
| 0,1384 | -0,02059 | 0,07694 |

| Обратная матрица | | |
|------------------|--------------|--------------|
| 1,665414799 | -1,082119067 | -3,285342341 |
| -1,082119067 | 17,86905896 | 6,72847937 |
| -3,285342341 | 6,72847937 | 20,70744438 |

Вычислим вектор оценок коэффициентов дискриминантной функции:

$$\hat{a} = \hat{S}^{-1} (\bar{X} - \bar{Y})$$

| | |
|----|--------------|
| a^ | 2,719403213 |
| | -4,679389648 |
| | 7,204132604 |

Теперь найдем оценки дискриминантной функции для обучающих выборок:

| Ux1=X1*a^ | Ux2=X2*a^ |
|-------------|-------------|
| 38,62037502 | 22,04159086 |
| 37,4340421 | 14,24144627 |
| 37,94123479 | 25,07486673 |
| 41,14922336 | 25,11782142 |
| | 14,24026026 |

Получим средние значения оценок дискриминантной функции, а затем оценку константы дискриминации:

$$U_{x1cp} = 38,78; U_{x2cp} = 20,14$$

$$\hat{c} = \frac{1}{2}(\bar{\hat{U}}_x + \bar{\hat{U}}_y)$$

$$c^{\wedge} = \frac{1}{2} * (38,78 + 20,14) = 29,46$$

Чтобы определить, к какой группе относятся компании, подлежащие дискриминации, рассчитаем для них дискриминантную функцию:

$$U(z1) = 5,5 * a1 + 0,05 * a2 + 1,02 * a3 = 22,07 < c^{\wedge} = 29,46 \Rightarrow \text{Данную компанию можно}$$

отнести к группе В. $U(z2) = 44,57 > c^{\wedge} = 29,46 \Rightarrow \text{Данную компанию можно отнести к}$
группе А.