МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ. ПРИМЕР РАСЧЕТА

Пусть существует некоторая сложная система, которая имеет три входа и один выход, или три входных фактора и одну выходную характеристику. Зависимость, которая существует между ними, имеет вид $Y = f(X_1, X_2, X_3)$, в нашем случае предполагаем ее линейной, то есть $Y = a_1X_1 + a_2X_2 + a_3X_3$. Коэффициенты $a_{i,}\ i = \overline{1,3}$ необходимо определить. Известна табл. 1 данных эксперимента (или статистических данных).

Таблица 1

X1	X2	X3	Y
1	9	12	23
3	8	23	43
5	3	34	12
7	2	29	26
9	5	38	76
12	6	45	43
15	7	54	23
18	11	56	76
21	1	67	18
23	5	78	44

Исследуем исходные данные на мультиколлинеарность по критерию Фаррара-Глобера. На первом шаге нормируем исходные данные и получим табл.2. При нормализации учитываем, что $n = 10$, $\overline{X_1} = 11.4$, $\overline{X_2} = 5.7$, $\overline{X_3} = 43.6$, $\sigma_1 = 7{,}2691$, $\sigma_2 = 3{,}002$, $\sigma_3 = 19{,}38$. Транспонируем матрицу табл. 2 и умножим транспонированную матрицу на исходную из табл.2. Получим корреляционную матрицу $r = \begin{pmatrix} 1 & -0.146 & 0.983 \\ -0.146 & 1 & -0.21 \\ 0.983 & -0.21 & 1 \end{pmatrix}$. Следующим шагом есть определение критерия $\chi^2 = -(10 - 1 - \frac{1}{6} \times$

$$\times (2 \cdot 3 + 5)) \cdot \ln|r| = -(9 - \frac{11}{6}) \cdot \ln(0.028434) = 25.51463.$$

Таблица 2

X1n	X2n	X3n
-0,45243	0,347657	-0,51572
-0,36542	0,242307	-0,3362
-0,27842	-0,28445	-0,15668
-0,19141	-0,3898	-0,23828
-0,10441	-0,07375	-0,09139
0,026102	0,031605	0,022849
0,156611	0,136956	0,169732
0,287119	0,558359	0,202372
0,417628	-0,49515	0,381896
0,504634	-0,07375	0,56142

Сравним вычисленное значение с табличным при 3 степенях свободы и уровне значимости $\alpha = 0{,}05$ $\chi^2_{табл.} = 7.8$. Поскольку вычисленное значение больше табличного, то в массиве факторов существует мультиколлинеарность.

Определим мультиколлинеарность каждого фактора с остальными. Для этого найдем обратную матрицу $C = r^{-1} = \begin{pmatrix} 33.614 & -2.141 & -33.5 \\ -2.138 & 1.182 & 2.351 \\ 33.49 & 2.353 & 34.42 \end{pmatrix}$ и вычислим F-критерии. Так, $F_1 = 114.1498$, $F_2 = 0.63827$, $F_3 = 119.979$. Поскольку табличное значение критерия при 7 и 2 степенях свободы равно $F_{табл.} = 19.36$, то сравнивая вычисленные значения и табличное, получим, что первый и третий фактор мультиколлинеарен с другими.

Для выяснения мультиколлинеарности каждой пары переменных находим частичные коэффициенты корреляции: $r_{12} = 0.339543$, $r_{13} = 0.984722$, $r_{23} = -0.36845$ и вычисляем значения t-критерия: $t_{12} = 0.9551$, $t_{13} = 14.961$, $t_{23} = -1.049$. Вычисленные значения сравниваем с табличным при 7 степенях свободы и уровне значимости $\alpha = 0{,}05$, $t_{табл.} = 2.45$. Мультиколлинеарность существует между первым и третьим факторами.

Далее для поиска коэффициентов линейной модели используем метод главных компонент. Вначале нормализуем вектор факторов (начальный), получим табл. 3. Вычислим корреляционную матрицу $r = \begin{pmatrix} 1 & -0{,}14 & 0{,}983 \\ -0{,}14 & 1 & -0{,}21 \\ 0{,}983 & -0{,}21 & 1 \end{pmatrix}$. Далее находим собственные (характеристические) числа матрицы r. Получим $\text{eigenvals}(a) = \begin{pmatrix} 2.045 \\ 0.012 \\ 0.943 \end{pmatrix}$. Вычисляем собственные векторы $\text{eigenvecs}(a) = \begin{pmatrix} 0.682 & -0.701 & 0.189 \\ -0.237 & 0.045 & 0.974 \\ 0.692 & 0.711 & 0.128 \end{pmatrix}$. Упорядочим собственные числа, получим массив (2.045, 0.943, 0.012). Соответственно массив собственных векторов будет таким $\begin{pmatrix} 0.682 & 0.189 & -0.701 \\ -0.237 & 0.974 & 0.045 \\ 0.692 & 0.128 & 0.711 \end{pmatrix}$. Вычислим главные компоненты:

$Z_1 = (-2.323, -1.762, -0.747, -0.678, -0.365, 0.079, 0.613, 0.644, 2.117, 2.37)$,

$Z_2 = (0.602, 0.382, -1.111, -1.385, -0.29, 0.121, 0.582, 1.978, -1.154, 0.335)$,

$Z_3 = (-0.107, 0.094, 0.235, -0.202, -0.012, -0.002, 0.053, -0.103, -0.137, 0.135)$

Таблица 3

X1	X2	X3
-1,431	1,099	-1,631
-1,156	0,766	-1,063
-0,88	-0,9	-0,495
-0,605	-1,23	-0,753
-0,33	-0,23	-0,289
0,083	0,1	0,0723
0,495	0,433	0,5367
0,908	1,766	0,64
1,321	-1,57	1,2077
1,596	-0,23	1,7754

Следующим шагом есть определение параметров модели $Y = Zb$ по формуле $b = (Z^T Z)^{-1} Z^T Y$. Полученный результат: $b = (1{,}431,\ 12{,}992,\ -32.11)$. И остается вычислить параметры модели $Y = X\beta$ по формуле $\beta = a \cdot b$, где a - массив собственных векторов. Результат $\beta = (25{,}94,\ 10{,}87,\ -20{,}17)$. Таким образом, искомая модель имеет вид $Y = 25{,}94 X_1 + 10{,}87 X_2 - 20{,}17 X_3$. Она изображена на рис.1. желтым цветом. То, что линия исходных данных размещена ниже, чем линия модели $Y = Zb$, объясняется отсутствием коэффициента смещения в модели главных компонент.

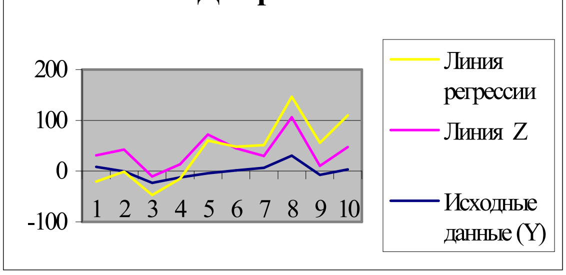

Рис. 1- Диаграмма значений