УДК 517.8

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ РАССТОЯНИЙ (МНКР) ДЛЯ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

А. Е. Ефремов, студент гр.ПИб – 151, 1 курс
Научный руководитель: В. А. Гоголин, д.т.н., профессор
Кузбасский государственный технический университет им. Т. Ф. Горбачева
г. Кемерово

В принятом методе наименьших квадратов рассматривается минимизация суммы квадратов отклонений экспериментальных значений от соответствующих значений искомой линейной функции регрессии [1]. Для вертикальных отклонений Δ_y (Рис.1) находятся коэффициенты прямой регрессии $y = \alpha x + \beta$ по минимуму следующей функции

$$\sum \Delta_y^2 = \sum(\alpha x_i + \beta - y_i)^2 = G(\alpha, \beta).$$

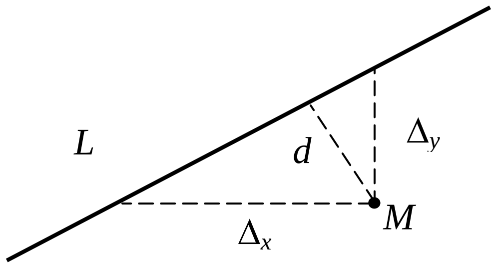

Рис.1 Отклонения экспериментальной точки *M* от линии регрессии *L*

Для горизонтальных отклонений Δ_x (Рис.1) находятся коэффициенты обратной регрессии $x = \gamma y + \delta$ по минимуму следующей функции

$$\sum \Delta_x^2 = \sum(\gamma y_i + \delta - x_i)^2 = H(\gamma, \delta).$$

Из этих условий получают следующие системы двух линейных уравнений на коэффициенты регрессий:

$$\begin{cases} \alpha \sum x_i^2 + \beta \sum x_i = \sum x_i y_i \\ \alpha \sum x_i + \beta n = \sum y_i \end{cases} \text{ и } \begin{cases} \gamma \sum y_i^2 + \delta \sum y_i = \sum y_i x_i \\ \gamma \sum y_i + \delta n = \sum x_i \end{cases}, \qquad (1)$$

где n – число экспериментальных точек.

Решение этих систем дает две различные регрессионные прямые.

В предлагаемом методе минимизируется сумма квадратов расстояний d от экспериментальных точек $(x_i; y_i)$ до прямой регрессии $y = ax + b$ (Рис.1). В соответствии с [2], эта сумма имеет следующий вид:

$$\sum d^2 = \sum \frac{(ax_i - y_i + b)^2}{a^2 + 1} = F(a, b).$$

Минимум полученной функции реализуется для значений a и b , при которых частные производные этой функции обращаются в ноль

$$\begin{cases} F_a'(a,b)=0, \\ F_b'(a,b)=0. \end{cases}$$

$$F_a' = \frac{2}{(a^2+1)^2}\cdot\sum[-a\cdot(ax_i-y_i+b)^2+x_i\cdot(ax_i-y_i+b)]=0,$$

$$F_b' = \sum 2\cdot(a^2+1)^{-1}\cdot\sum(ax_i-y_i+b)=0.$$

Выражаем из второго уравнения $b=\bar{y}-a\cdot\bar{x}$, подставляем это выражение в первое уравнение и после преобразований получим следующее квадратное уравнение на угловой коэффициент линейной регрессии a:

$$(\bar{x}\cdot\bar{y}-\overline{xy})\cdot a^2-(\overline{x^2}-\overline{y^2}+\bar{y}^2-\bar{x}^2)\cdot a+(\overline{xy}-\bar{x}\cdot\bar{y})=0. \qquad (2)$$

Нетрудно проверить, что это уравнение имеет два действительных корня разного знака, которые дают две взаимно перпендикулярные прямые [2]. Одно решение соответствует минимуму суммы квадратов расстояний от экспериментальных точек до прямой регрессии, а второе – максимуму. Знак углового коэффициента соответствует знаку свободного члена уравнения, определяющему знак коэффициента корреляции.

Если прямая проходит через начало координат, то есть имеет вид $y=a\cdot x$, то ее угловой коэффициент определяется по существующему методу как $a=\overline{xy}/\overline{x^2}$. В предлагаемом методе этот коэффициент в рассматриваемом случае находится из уравнения, аналогичному уравнению (2):

$$\overline{xy}\cdot a^2+(\overline{x^2}-\overline{y^2})\cdot a-\overline{xy}=0. \qquad (3)$$

Рассмотрим пример на определение уравнений линейной регрессии по существующему и предлагаемому методам двух факторов: посещаемость X и успеваемость $У$ по следующим данным (результаты экзамена по математике):

X_i	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
$У_i$	2,0	2,2	2,5	2,3	2,6	3,3	4,2	3,5	4,1

Из решения систем уравнений (1) находим коэффициенты прямой и регрессии по существующему методу (МНК). Из уравнения (2) находим угловой коэффициент линейной регрессии по предлагаемому методу (МНКР). Данные проведенных таким образом расчетов представлены на рис.2.

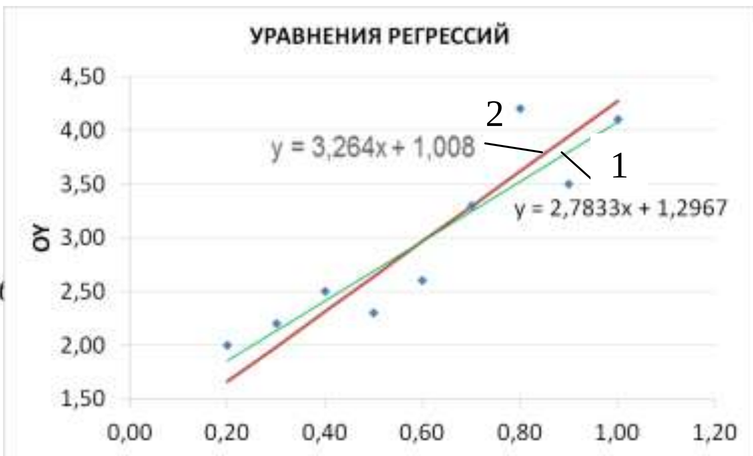

Рис.2. Уравнения линейной регрессии по МНК (1) и МНКР (2)

Как видно из рис.2, использование критерия минимизации расстояний дает заметное различие в уравнениях регрессии. Поэтому при нахождении прогнозных значений встает вопрос о выборе наиболее адекватной регрессионной модели, который пока остается открытым. Следует только заметить, что принятый МНК, дает на самом деле два уравнения регрессии, а предлагаемый МНКР - только одно, что говорит в его пользу.

Список литературы:

1. Горелова Г.В. Теория вероятностей и математическая статистика в примерах и задачах с применением Excel : учеб. пособие для вузов / Г. В. Горелова, И. А. Кацко //Ростов н/Д: Феникс , 2005.– 480 с.

2. Беклемишев Д.Б. Курс аналитической геометрии и линейной алгебры. М.: Наука, 1987. – 320 с.