Множественная линейная регрессия

Множественной называют линейную регрессию, в модели которой число независимых переменных две и более.

Уравнение множественной линейной регрессии имеет вид:

$$y = k_0 + k_1x_1 + k_2x_2 + \cdots + k_nx_n$$

Как и в простой линейной регрессии, параметры модели k_n вычисляются при помощи метода наименьших квадратов.

Представим данные наблюдений и коэффициенты модели в матричной форме.

$$Y=\begin{bmatrix} y_1 \\ \ldots \\ y_n \end{bmatrix}; X=\begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nm} \end{bmatrix}; K=\begin{bmatrix} k_1 \\ \ldots \\ k_n \end{bmatrix}$$

Здесь Y — n-мерный вектор-столбец наблюдений зависимой переменной; X — матрица размерности n x (m+1), в которой i-я строка i = 1, 2,…, n представляет i-е наблюдение вектора значений независимых переменных X_1, X_2, …, X_m , единица соответствует переменной при свободном члене k_0. K — вектор-столбец размерности (m+1) параметров уравнения множественной регрессии.

$$y = k_0 + k_1x_1 + k_2x_2 + \cdots + k_nx_n$$

В матричном виде соотношение примет вид:

$$e = Y – XK$$

Согласно методу наименьших квадратов:

$$\sum_{i=1}^{n} e_i^2 = e^T e = (Y - XK)^T (Y - XK) \rightarrow min$$

Можно показать, что предыдущее условие выполняется, если вектор-столбец коэффициентов K найти по формуле:

$$K = (X^T X)^{-1} X^T Y$$

$(X^T X)^{-1}$ – матрица, обратная к $(X^T X)$. Соотношение справедливо для уравнений регрессии с произвольным количеством m объясняющих переменных.

X^T – транспонированная матрица X.

X^{-1} – обратная матрица X.