

Оценка параметров линейной множественной регрессии матричным методом в MS Excel

Множественная регрессия – зависимость среднего значения результативной случайной величины от значений случайных факторов

Линейная множественная регрессия имеет вид:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p + \varepsilon.$$

Линейная множественная регрессия является статистической моделью для случайных показателей.

Пример. Для изучения рынка жилья в городе по данным о 46 коттеджах было построено уравнение множественной регрессии

$$y = 21,1 - 6,2x_1 + 0,95x_2 + 3,57x_3 + \varepsilon,$$

y – цена объекта, д.е.

x_1 – расстояние до центра города, км.

x_2 – полезная площадь объекта, кв.м.

x_3 – число этажей в доме.

Уравнение регрессии позволяет получить прогнозное значение результативной переменной для конкретных значений факторов. Если расстояние до центра города равно 10 км, полезная площадь 100 кв.м., число этажей в доме равно 2, то рыночная (средняя) цена коттеджа равна

$$y = 21,1 - 6,2 \cdot 10 + 0,95 \cdot 100 + 3,57 \cdot 2 = 61,24 \text{ (д. е.)}$$

Уравнение регрессии позволяет найти ответ на многие важные вопросы.

Одной из основных задач является оценка параметров уравнения регрессии. Для этого обычно используют метод наименьших квадратов, который приводит к решению системы линейных уравнений:

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_p \sum x_p, \\ \sum yx_1 = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_2x_1 + \dots + b_p \sum x_px_1, \\ \sum yx_p = b_0 \sum x_p + b_1 \sum x_1x_p + b_2 \sum x_2x_p + \dots + b_p \sum x_p^2. \end{cases}$$

Неизвестными величинами в системе являются параметры $b_0, b_1, b_2, \dots, b_p$. Остальные значения даны по условию в виде исходных статистических данных.

Введем следующие обозначения:

$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$ – вектор-столбец значений результативной переменной

$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$ – матрица значений факторов, дополненных
единичным столбцом

$B = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_p \end{pmatrix}$ – вектор-столбец параметров уравнения множественной регрессии.

Тогда решение системы линейных уравнений в матричной форме имеет вид:

$$B = (X^T X)^{-1} \cdot (X^T Y)$$

Рассмотрим пример оценки параметров уравнения линейной множественной регрессии матричным методом в электронных таблицах MS Excel:

Пример. Имеются данные о деятельности крупнейших компаний США в 1996 году:

	A	B	C	D	E	F
1	№	y	x1	x2	x3	x4
2	1	0.9	31.3	18.9	43	40.9
3	2	1.7	13.4	13.7	64.7	40.5
4	3	0.7	4.5	18.5	24	38.9
5	4	1.7	10	4.8	50.2	38.5
6	5	2.6	20	21.8	106	37.3
7	6	1.3	15	5.8	96.6	26.5
8	7	4.1	137.1	99	347	37
9	8	1.6	17.9	20.1	85.6	36.8
10	9	6.9	165.4	60.6	745	36.3
11	10	0.4	2	1.4	4.1	35.3
12	11	1.3	6.8	8	26.8	35.3
13	12	1.9	27.1	18.9	42.7	35
14	13	1.9	13.4	13.2	61.8	26.2
15	14	1.4	9.8	12.6	121	33.1
16	15	0.4	19.5	12.2	105	32.7
17	16	0.8	6.8	3.2	33.5	32.1
18	17	1.8	27	13	142	30.5
19	18	0.9	12.4	6.9	96	29.8
20	19	1.1	17.7	15	140	25.4
21	20	1.9	12.7	11.9	59.3	29.3
22	21	-0.9	21.4	1.6	131	29.2
23	22	1.3	13.5	8.6	70.7	29.2
24	23	2	13.4	11.5	65.4	29.1
25	24	0.6	4.2	1.9	23.1	27.9
26	25	0.7	15.5	5.8	80.8	27.2

При этом:

y - чистый доход, млрд долл. США

x_1 – оборот капитала, млрд долл. США

x_2 – использованный капитал, млрд долл. США

x_3 – численность служащих, тыс. чел

x_4 – рыночная капитализация компании, млрд долл. США

Источник: Практикум по эконометрике: Учеб. пособие / И.И.Елисеева, С.В.Курьшова, Н.М.Гордеенко и др.: Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2006. – Задача 23, с.91.

Подготовим матрицу X (дополним столбцом из единиц):

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	№	y	x_1	x_2	x_3	x_4			Матрица X				
2	1	0,9	31,3	18,9	43	40,9			1	31,3	18,9	43	40,9
3	2	1,7	13,4	13,7	64,7	40,5			1	13,4	13,7	64,7	40,5
4	3	0,7	4,5	18,5	24	38,9			1	4,5	18,5	24	38,9
5	4	1,7	10	4,8	50,2	38,5			1	10	4,8	50,2	38,5
6	5	2,6	20	21,8	106	37,3			1	20	21,8	106	37,3
7	6	1,3	15	5,8	96,6	26,5			1	15	5,8	96,6	26,5
8	7	4,1	137,1	99	347	37			1	137,1	99	347	37
9	8	1,6	17,9	20,1	85,6	36,8			1	17,9	20,1	85,6	36,8
10	9	6,9	165,4	60,6	745	36,3			1	165,4	60,6	745	36,3
11	10	0,4	2	1,4	4,1	35,3			1	2	1,4	4,1	35,3
12	11	1,3	6,8	8	26,8	35,3			1	6,8	8	26,8	35,3
13	12	1,9	27,1	18,9	42,7	35			1	27,1	18,9	42,7	35
14	13	1,9	13,4	13,2	61,8	26,2			1	13,4	13,2	61,8	26,2
15	14	1,4	9,8	12,6	121	33,1			1	9,8	12,6	121	33,1
16	15	0,4	19,5	12,2	105	32,7			1	19,5	12,2	105	32,7
17	16	0,8	6,8	3,2	33,5	32,1			1	6,8	3,2	33,5	32,1
18	17	1,8	27	13	142	30,5			1	27	13	142	30,5
19	18	0,9	12,4	6,9	96	29,8			1	12,4	6,9	96	29,8
20	19	1,1	17,7	15	140	25,4			1	17,7	15	140	25,4
21	20	1,9	12,7	11,9	59,3	29,3			1	12,7	11,9	59,3	29,3
22	21	-0,9	21,4	1,6	131	29,2			1	21,4	1,6	131	29,2
23	22	1,3	13,5	8,6	70,7	29,2			1	13,5	8,6	70,7	29,2
24	23	2	13,4	11,5	65,4	29,1			1	13,4	11,5	65,4	29,1
25	24	0,6	4,2	1,9	23,1	27,9			1	4,2	1,9	23,1	27,9
26	25	0,7	15,5	5,8	80,8	27,2			1	15,5	5,8	80,8	27,2

Укажем диапазоны, в которых расположены матрицы, участвующие в расчете параметров:

Матрица X : I2:M26

Матрица Y : B2:B26

Для расчета столбца параметров B необходимо воспользоваться формулой

$$B = (X^T X)^{-1} \cdot (X^T Y)$$

Ответ:

b_0	-0,57419782
b_1	-0,01114024
b_2	0,028182637
b_3	0,007899822
b_4	0,033037721

Примечание

В электронных таблицах есть готовые инструменты для оценки параметров. После оценки параметров важно проверить модель на статистическую значимость. Все это можно сделать с помощью инструмента **Регрессия** надстройки **Анализ данных**.