

ФГОБУ
Финуниверситет при Правительстве РФ

■ **Эконометрика**

- **Семинар 5**

Уравнение множественной линейной регрессии (построение уравнения)

Множественная линейная регрессия

5.1 Классическая нормальная линейная модель множественной регрессии

Обычно на любой экономический показатель влияет не один, а несколько факторов. Например, спрос на некоторый товар определяется не только его ценой, но и ценами на замещающие и дополняющие товары, доходом потребителей и другими факторами.

В этом случае вместо функции парной регрессии

$$\hat{y}(x_1 = x) = \hat{y}(x)$$

рассматривается функция множественной регрессии:

$$\hat{y}(x_1 = x_1; x_2 = x_2; \dots; x_n = x_n) = \hat{y}(x_1; x_2; \dots; x_n). \quad (5.1)$$

Теоретическая модель множественной линейной регрессии имеет вид:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (5.2)$$

или для индивидуальных наблюдений:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i \quad (5.3)$$

$$i = 1, 2, \dots, N,$$

где N – объем генеральной совокупности.

После выбора в качестве модели линейной функции множественной регрессии необходимо оценить коэффициенты регрессии.

Пусть имеется n наблюдений вектора объясняющих переменных $X = (X_1; X_2; \dots; X_k)$ и зависимой переменной Y :

$$(X_{i1}; X_{i2}; \dots; X_{ik}; Y_i), \quad i = 1, 2, \dots, n \quad (5.4)$$

Если $k = n + 1$, то оценки коэффициентов рассчитываются *единственным* образом. Здесь n – число объясняющих переменных.

Если $k < n + 1$, то система будет иметь *бесконечное* множество решений.

Если $k > n + 1$, то *нельзя подобрать линейную функцию (5.2), точно удовлетворяющую всем наблюдениям*, и возникает необходимость оптимизации, т.е. нахождения оценок параметров модели (5.3), при которых линейная функция дает наилучшее приближение для имеющихся наблюдений.

Самым распространенным методом оценки параметров модели множественной линейной регрессии является метод наименьших квадратов (МНК).

Для применения МНК необходима выполнимость ряда предпосылок, которые позволят проводить анализ в рамках классической линейной модели множественной регрессии (КЛММР). Ранее были рассмотрены предпосылки МНК для парной линейной регрессии. Для случая множественной линейной регрессии необходимо выполнение дополнительных предпосылок. В следующем разделе перечислены все предпосылки, выполнение которых требуется для нахождения коэффициентов КЛММР.

5.1.1. Предпосылки МНК

1) Математическое ожидание случайных отклонений ε_i равно нулю:

$$E(\varepsilon_i) = 0$$

для всех наблюдений. Это означает, что случайное отклонение не должно иметь систематического смещения.

2) Дисперсия случайных отклонений постоянна для всех наблюдений:

$$E(\varepsilon_i^2) = E(\varepsilon_i^2) = \sigma^2, \quad i = 1, 2, \dots, n$$

Выполнимость данной предпосылки называется гомоскедастичностью, а невыполнимость ее называется гетероскедастичностью.

3) Случайные отклонения ε_i и ε_j ($i \neq j$) не коррелируют (отсутствует автокорреляция):

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad \varepsilon_i \neq \varepsilon_j.$$

4) Случайные отклонения должны быть статистически независимы (некоррелированы) от объясняющих переменных.

5) Отсутствие *мультиколлинеарности*. Между объясняющими переменными отсутствует строгая (сильная) линейная зависимость.

6) Отклонения ε_i , $i = 1, 2, \dots, n$ имеют нормальные распределения $\varepsilon_i \sim N(0; \sigma^2)$. Выполнимость данной предпосылки важна для проверки статистических гипотез и построения интервальных оценок.

Замечание: При построении классических линейных регрессионных моделей считается также, что объясняющие переменные не являются случайными величинами.

Теорема Гаусса – Маркова

При выполнении приведенных предпосылок МНК оценки коэффициентов b_j $j = 0; 1; 2; \dots; m$, являются несмещенными, эффективными и состоятельными.

Нахождение коэффициентов уравнения множественной линейной регрессии

Немного теории

$$\begin{aligned}
 \frac{\partial L}{\partial x} &= \frac{\partial}{\partial x} \left(\frac{1}{2} \left(\frac{\partial x}{\partial t} - \left(\frac{\partial x}{\partial t} + \frac{\partial x}{\partial x} \frac{\partial x}{\partial x} \right) \right)^2 \right) \\
 &= \frac{\partial x}{\partial t} - \left(\frac{\partial x}{\partial t} + \frac{\partial x}{\partial x} \frac{\partial x}{\partial x} \right) = 0
 \end{aligned}$$

Получили систему нормальных уравнений:

$$\begin{aligned}
 \frac{\partial L}{\partial x} &= \frac{\partial}{\partial x} \left(\frac{1}{2} \left(\frac{\partial x}{\partial t} - \left(\frac{\partial x}{\partial t} + \frac{\partial x}{\partial x} \frac{\partial x}{\partial x} \right) \right)^2 \right) = 0 \\
 \frac{\partial L}{\partial x} &= \frac{\partial}{\partial x} \left(\frac{1}{2} \left(\frac{\partial x}{\partial t} - \left(\frac{\partial x}{\partial t} + \frac{\partial x}{\partial x} \frac{\partial x}{\partial x} \right) \right)^2 \right) = 0
 \end{aligned}$$

Решение системы нормальных уравнений возможно разными способами. Рассмотрим метод Крамера и матричный метод.

Пример. Найти коэффициенты уравнения множественной линейной регрессии по данным, приведенным в таблице:

	x1	x2	Y
1	3	5	7
2	1	7	7
3	5	2	5
4	4	1	4
5	9	2	2
6	9	1	8
7	3	3	1
8	1	7	3
9	2	4	4
10	8	8	2

$$\begin{aligned} & \beta_1 \beta_2 (\beta_{11} - (\beta_0 + \beta_1 \beta_{11})) = 0 \\ & \beta_1 = 1 \quad \beta_2 = 1 \quad \beta_1 \beta_2 = 1 \\ & \beta_1 \beta_2 (\beta_{22} - (\beta_0 + \beta_1 \beta_{22})) \beta_1 \beta_2 = 0 \\ & \beta_1 \beta_2 = 1 \quad \beta_1 \beta_2 = 1 \end{aligned}$$

Решение (система нормальных уравнений для случая двух объясняющих переменных):

$$\begin{aligned} \beta_0 &= \beta_0 + \beta_1 \beta_1 + \beta_2 \beta_2 \\ \beta_1 \beta_1 &= \beta_0 \beta_1 + \beta_1 \beta_1^2 + \beta_2 \beta_1 \beta_2 \\ \beta_1 \beta_2 \beta_2 &= \beta_0 \beta_2 + \beta_1 \beta_1 \beta_2 + \beta_2 \beta_2^2 \end{aligned}$$

	X ₁	X ₂	Y	X ₁ Y	X ₂ Y	X ₁ ²	X ₂ ²	X ₁ *X ₂
1	3	5	7	21	35	9	25	15
2	1	7	7	7	49	1	49	7
3	5	2	5	25	10	25	4	10
4	4	1	4	16	4	16	1	4
5	9	2	2	18	4	81	4	18
6	9	1	8	72	8	81	1	9
7	3	3	1	3	3	9	9	9
8	1	7	3	3	21	1	49	7
9	2	4	4	8	16	4	16	8
10	8	8	2	16	16	64	64	64
	45	40	43	189	166	291	222	151

$$\begin{aligned}
 43 &= \alpha_0 \cdot 10 + \alpha_1 \cdot 45 + \alpha_2 \cdot 40 \\
 189 &= \alpha_0 \cdot 45 + \alpha_1 \cdot 291 + \alpha_2 \cdot 151 \\
 166 &= \alpha_0 \cdot 40 + \alpha_1 \cdot 151 + \alpha_2 \cdot 222
 \end{aligned}$$

$$\begin{cases} 43 = x_0 \cdot 10 + x_1 \cdot 45 + x_2 \cdot 40 \\ 189 = x_0 \cdot 45 + x_1 \cdot 291 + x_2 \cdot 151 \\ 166 = x_0 \cdot 40 + x_1 \cdot 151 + x_2 \cdot 222 \end{cases}$$

Решаем систему по методу Крамера

$$\Delta = \begin{vmatrix} 10 & 45 & 40 \\ 45 & 291 & 151 \\ 40 & 151 & 222 \end{vmatrix} = 46460$$

$$\Delta_0 = \begin{vmatrix} 43 & 45 & 40 \\ 189 & 291 & 151 \\ 166 & 151 & 222 \end{vmatrix} = 246623$$

$$\Delta_1 = \begin{vmatrix} 10 & 43 & 40 \\ 45 & 189 & 151 \\ 40 & 166 & 222 \end{vmatrix} = -4530$$

$$\Delta_2 = \begin{vmatrix} 10 & 45 & 43 \\ 45 & 291 & 189 \\ 40 & 151 & 166 \end{vmatrix} = -6615$$

$$\Delta_0 = \frac{\Delta_0}{\Delta} = 5,308287$$

$$\Delta_1 = \frac{\Delta_1}{\Delta} = -0,0975$$

$$\Delta_2 = -0,14238$$

$$\Delta = 5,3 - 0,097\Delta_1 - 0,143\Delta_2 \blacktriangleleft$$

Задача 1. Найти коэффициенты уравнения множественной линейной регрессии по данным, приведенным в таблице:

	x1	x2	Y
1	3	5	7
2	2	7	7
3	5	2	5
4	4	4	4
5	9	2	7
6	9	1	8
7	3	3	1
8	2	7	3
9	2	4	4
10	8	8	6

Самостоятельно, используя систему нормальных уравнений (метод Крамера)!

Решение (система нормальных уравнений для случая двух объясняющих переменных):

$$\begin{aligned}
 \sum_{i=1}^n Y_i &= \sum_{i=1}^n \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \\
 \sum_{i=1}^n Y_i X_{i1} &= \sum_{i=1}^n \beta_0 X_{i1} + \beta_1 \sum_{i=1}^n X_{i1}^2 + \beta_2 \sum_{i=1}^n X_{i1} X_{i2} \\
 \sum_{i=1}^n Y_i X_{i2} &= \sum_{i=1}^n \beta_0 X_{i2} + \beta_1 \sum_{i=1}^n X_{i1} X_{i2} + \beta_2 \sum_{i=1}^n X_{i2}^2
 \end{aligned}$$

	x_1	x_2	y	x_1y	x_2y	x_1^2	x_2^2	$x_1 * x_2$
1	3	5	7	21	35	9	25	15
2	2	7	7	14	49	4	49	14
3	5	2	5	25	10	25	4	10
4	4	4	4	16	16	16	16	16
5	9	2	7	63	14	81	4	18
6	9	1	8	72	8	81	1	9
7	3	3	1	3	3	9	9	9
8	2	7	3	6	21	4	49	14
9	2	4	4	8	16	4	16	8
10	8	8	6	48	48	64	64	64
	47	43	52	276	220	297	237	177

$$52 = \alpha_0 \cdot 10 + \alpha_1 \cdot 47 + \alpha_2 \cdot 43$$

$$276 = \alpha_0 \cdot 47 + \alpha_1 \cdot 297 + \alpha_2 \cdot 177$$

$$220 = \alpha_0 \cdot 43 + \alpha_1 \cdot 177 + \alpha_2 \cdot 237$$

$$\Delta = 33348$$

$$\Delta_0 = 77952$$

$$\Delta_1 = 15560$$

$$\Delta_2 = 5192$$

$$\text{Ответ: } \alpha_0 = 2,337 + 0,466\alpha_1 + 0,156\alpha_2 \leftarrow$$

Задача 2. Найти коэффициенты уравнения множественной линейной регрессии по данным, приведенным в таблице:

	x1	x2	y
1	5	5	7
2	8	4	9
3	4	5	8
4	5	3	5
5	3	6	9
6	9	7	5
7	6	5	7
8	4	8	8
9	8	6	4
10	2	4	3
11	3	5	7
12	5	4	8

Самостоятельно, используя метод определителей.

$$80 = \alpha_0 \cdot 12 + \alpha_1 \cdot 62 + \alpha_2 \cdot 62$$

$$409 = \alpha_0 \cdot 62 + \alpha_1 \cdot 374 + \alpha_2 \cdot 326$$

$$417 = \alpha_0 \cdot 62 + \alpha_1 \cdot 326 + \alpha_2 \cdot 342$$

$$\Delta = 13568$$

$$\Delta_0 = 83840$$

$$\Delta_1 = -1376$$

$$\Delta_2 = 2656$$

Ответ: $\alpha_0 = 6,179 - 0,101\alpha_1 + 0,195\alpha_2 \blacktriangleleft$

Решим эту же задачу матричным методом

$$\mathbb{X} = (\mathbb{X}^{\mathbb{X}} \mathbb{X} \mathbb{X})^{-1} \mathbb{X} \mathbb{X}^{\mathbb{X}} \mathbb{X} \mathbb{X}$$

Вспомним:

1) Умножение матриц

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} = \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \quad b = c, a = e, d = f$$

Выделяем на листе *Excel* прямоугольник ячеек $a*d$,

вызываем **МУМНОЖ** и вводим массивы матриц A и B , умножения двух матриц A и B , получаем матрицу C размерностью $a*d$. (**ctrl+shift** вызываются на клавиатуре слева).

2) Определитель матрицы: вызываем **МОПРЕД**, вводим массив *квадратной* матрицы и нажимаем **enter**. Результат в ячейке, где находился курсор.

3) Обратная матрица: выделяем на листе *Excel* *квадрат* ячеек, равный исходной матрице, вызываем **МОБР**, вводим массив исходной матрицы и нажимаем **ctrl+shift+enter**. Получаем результат в выделенном квадрате ячеек.

Вспомним исходные данные:

	x1	x2	y
1	5	5	7
2	8	4	9
3	4	5	8
4	5	3	5
5	3	6	9
6	9	7	5
7	6	5	7
8	4	8	8
9	8	6	4
10	2	4	3
11	3	5	7
12	5	4	8

Матрица $X_{12 \times 3}$ (1 – единица в первом столбце соответствует переменной при коэффициенте b_0)

	x1	x2
1	5	5
1	8	4
1	4	5
1	5	3
1	3	6
1	9	7
1	6	5
1	4	8
1	8	6
1	2	4
1	3	5
1	5	4

Матрица (вектор столбец) $Y_{12 \times 1}$

y
7
9
8
5
9
5
7
8
4
3
7
8

Матрица $X^T_{3 \times 12}$

	1	1	1	1	1	1	1	1	1	1	1	1
x1	5	8	4	5	3	9	6	4	8	2	3	5
x2	5	4	5	3	6	7	5	8	6	4	5	4

Произведение матриц (получено с использованием *Excel*):

$$(X^T * X)_{3 \times 3} = \begin{vmatrix} 12 & 62 & 62 \\ 62 & 374 & 326 \\ 62 & 326 & 342 \end{vmatrix}$$

В нашем случае $X^T_{3 \times 12}$; $X_{12 \times 3}$

Обратная матрица (получено с использованием Excel):

$(X^T * X)^{-1}_{3 \times 3} =$	1,59434	-0,07311	-0,21934
	-0,07311	0,019163	-0,00501
	-0,21934	-0,00501	0,047465

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

$X^T * Y =$	80	
	409	
	417	
$B_{3 \times 1} =$	6,179245	
	-0,10142	
	0,195755	

Задача 2-а. Найти коэффициенты уравнения множественной линейной регрессии по данным, приведенным в таблице:

№	y	x1	x2
1	8	8	8
2	9	9	9
3	10	10	10
4	11	11	11
5	8	8	8
6	11	10	10
7	8	9	8
8	9	10	10
9	11	10	11
10	9	8	7

Самостоятельно, используя метод Крамера и матричный метод!

№	y	x1	x2	x1y	x2y	x1 ²	x2 ²	x1*x2
1	8	8	8	64	64	64	64	64
2	9	9	9	81	81	81	81	81
3	10	10	10	100	100	100	100	100
4	11	11	11	121	121	121	121	121
5	8	8	8	64	64	64	64	64
6	11	10	10	110	110	100	100	100
7	8	9	8	72	64	81	64	72
8	9	10	10	90	90	100	100	100
9	11	10	11	110	121	100	121	110
10	9	8	7	72	63	64	49	56
	94	93	92	884	878	875	864	868

$$\Delta = 240 \quad \Delta_0 = 352 \quad \Delta_1 = 88 \quad \Delta_2 = 118$$

$$\hat{y} = 1,466 + 0,366x_1 + 0,492x_2$$

	10	93	92
$X^T * X =$	93	875	868
	92	868	864
	10,73333	-2,06667	0,933333
$(X^T * X)^{-1} =$	-2,06667	0,733333	-0,51667
	0,933333	-0,51667	0,420833
	94		1,466667
$X^T * Y =$	884	B =	0,366667
	878		0,491667

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

Задача 3. Найти коэффициенты уравнения множественной линейной регрессии по данным, приведенным в таблице:

	x1	x2	y
1	8	6	7
2	9	8	9
3	7	3	8
4	3	2	5
5	8	8	9
6	7	6	5
7	8	7	7
8	2	9	8
9	4,5	4	4
10	3	6	3

Самостоятельно, используя метод Крамера и матричный метод!

Ответ:

$$\bar{x}_1 = 2,072 + 0,387x_{11} + 0,360x_{21}$$

	x ₁	x ₂	y	x ₁ y	x ₂ y	x ₁ ²	x ₂ ²	x ₁ *x ₂
1	8	6	7	56	42	64	36	48
2	9	8	9	81	72	81	64	72
3	7	3	8	56	24	49	9	21
4	3	2	5	15	10	9	4	6
5	8	8	9	72	72	64	64	64
6	7	6	5	35	30	49	36	42
7	8	7	7	56	49	64	49	56
8	2	9	8	16	72	4	81	18
9	4,5	4	4	18	16	20,25	16	18
10	3	6	3	9	18	9	36	18
	59,5	59	65	414	405	413,25	395	363

$$65 = \beta_0 + \beta_1 \cdot 59,5 + \beta_2 \cdot 59$$

$$414 = \beta_0 + \beta_1 \cdot 413,25 + \beta_2 \cdot 363$$

$$405 = \beta_0 + \beta_1 \cdot 363 + \beta_2 \cdot 395$$

$$\Delta = 26348,5$$

$$\Delta_0 = 54595,5$$

$$\Delta_1 = 10211$$

$$\Delta_2 = 9477$$

Эта же задача, решенная матричным способом:

$$X^T * X = \begin{vmatrix} 10 & 59,5 & 59 \\ 59,5 & 413,25 & 363 \\ 59 & 363 & 395 \end{vmatrix}$$

$$(X^T * X)^{-1} = \begin{vmatrix} 1,194176 & -0,07915 & -0,10563 \\ -0,07915 & 0,0178 & -0,00454 \\ -0,10563 & -0,00454 & 0,022478 \end{vmatrix}$$

$$X^T * Y = \begin{vmatrix} 65 \\ 414 \\ 405 \end{vmatrix}$$

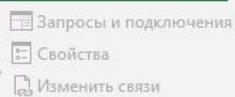
$$B = \begin{vmatrix} 2,072053 \\ 0,387536 \\ 0,359679 \end{vmatrix}$$



То же, используя программу *Регрессия Excel*:



Получить и преобразовать данные



Запросы и подключения



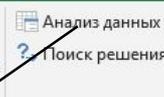
Сортировка и фильтр



Работа с данными

Прогноз

Структура



Анализ

E3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	x1	x2	y																
2	8	6	7																
3	9	8	9																
4	7	3	8																
5	3	2	5																
6	8	8	9																
7	7	6	5																
8	8	7	7																
9	2	9	8																
10	4,5	4	4																
11	3	6	3																

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

Метки Константа - ноль

Уровень надежности: %

Параметры вывода

Выходной интервал:

Новый рабочий дист.

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

OK Отмена Справка

Лист1

Готово

100%

	x1	x2	y	ВЫВОД ИТОГОВ								
1	8	6	7									
2	9	8	9	<i>Регрессионная статистика</i>								
3	7	3	8	Множест	0,672079							
4	3	2	5	R-квадрат	0,45169							
5	8	8	9	Нормиро	0,29503							
6	7	6	5	Стандарт	1,781113							
7	8	7	7	Наблюде	10							
8	2	9	8									
9	4,5	4	4	<i>Дисперсионный анализ</i>								
10	3	6	3		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>ачимость F</i>			
				Регрессия	2	18,29346	9,14673	2,883255	0,122065			
				Остаток	7	22,20654	3,172363					
				Итого	9	40,5						
				<i>Коэффициентная статистика - Значения ниже 95% верхние 95% жние 95,0%</i>								
				Y-пересе	2,072053	1,946371	1,064573	0,322407	-2,53038	6,674489	-2,53038	6,674489
				x1	0,387536	0,237629	1,630844	0,146944	-0,17437	0,94944	-0,17437	0,94944
				x2	0,359679	0,267034	1,346942	0,219972	-0,27176	0,991113	-0,27176	0,991113

Спасибо за внимание!