Лекция № 2. Модель множественной линейной регрессии

- Вопросы

- 1. Задача построения множественной линейной регрессии.
- 2. Оценка качества модели множественной линейной регрессии.
- 3. Прогнозирование по модели множественной линейной регрессии.

- 1. В общем случае зависимая переменная y может быть функцией нескольких переменных: x_1, x_2, …, x_m.
- В каждом наблюдении (i) получают совокупность значений независимой переменной: x_{i1}, x_{i2} ,…, x_{im} и совокупность значений зависимой переменной y_i. Итак, допустим,
- $y_i = \alpha_1 x_{i1} + \alpha_2 x_{i2} + \ldots + \alpha_m x_{im} + \varepsilon_i$. (1)

- **Введем матричные *обозначения*.**
- **Пусть $\{\alpha_j\}$, $j = 1,\ldots, m$ – вектор неизвестных параметров; $Y = \{y_i\}$, $i = 1,\ldots,n$**
- **- вектор зависимой переменной; $X = (x_{ij})$ – матрица независимых переменных размером nm; $\varepsilon = \{\varepsilon_i\}$- вектор ошибок.**
- **Тогда линейная модель (1) перепишется в виде**

  - $$Y = X\alpha + \varepsilon \qquad (2)$$

- **Относительно ошибок *ε* сделаем следующие *предположения*:**
- **1) возмущение *ε* является случайной величиной;**
- **2) *M(ε) = 0*;**
- **3) *Д(ε)* = const;**
- **4) последовательные значения *ε* не зависят друг от друга;**
- **5) матрица *X* состоит из линейно-независимых векторов-столбцов.**

- **Последнее обстоятельство эквивалентно тому, что ранг матрицы *X* равен *m*, а это, в свою очередь, означает, что $|X'X| \neq 0$, т.е. матрица *X'X* обратима (*X'* - транспонированная для *X*).**
- **Матрица *X* не содержит ошибок.**

- **Оценку выражения (1), полученную по выборочным данным, запишем в виде**
  - $y_i = a_1x_{i1} + a_2x_{i2} + \ldots + a_mx_{im} + e_i$ **(3)**
- **или в матричном виде** $Y = Xa + e$.
- **Сумму квадратов отклонений теперь можно определить как**
- $Q = \Sigma e_i^2 = e'e = (Y - Xa)' (Y - Xa) =$
- $= Y'Y - a'X'Y - Y'Xa + a'X'Xa$.
- **Так как** $a'X'Y = Y'Xa$, **то**
- $Q = Y'Y - 2a'X'Y + X'Xa^2$ **(4).**

- **Продифференцируем *Q* по *a*, получим**

$$\frac{\partial Q}{\partial a} = -2X'Y + 2\left(X'X\right)a = 0,$$

$$X'Y = X'Xa,$$

$$a = \left(X'X\right)^{-1} X'Y \qquad (5).$$

- **Оценку *a*, найденную по формуле (5), будем называть оценкой МНК.**

- **Таким образом, для определения вектора *a* необходимо по данным наблюдений найти матрицу, обратную к матрице *X'X*, и вектор *X'Y*:**

$$X'X = \begin{pmatrix} \sum_i x_{i1}^2 & \sum_i x_{i1}x_{i2} & \dots & \sum_i x_{i1}x_{im} \\ \sum_i x_{i2}x_{i1} & \sum_i x_{i2}^2 & \dots & \sum_i x_{i2}x_{im} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ \sum_i x_{im}x_{i1} & \sum_i x_{im}x_{i2} & \dots & \sum_i x_{im}^2 \end{pmatrix} \qquad (6)$$

$$X'Y = \begin{pmatrix} \sum_{i=1}^{n} y_i x_{i1} \\ \sum_{i=1}^{n} y_i x_{i2} \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} y_i x_{im} \end{pmatrix} \qquad (7)$$

- **Обычно предполагается, что уравнение регрессии имеет свободный член, т.е. a_0. Чтобы получить оценку этого параметра, расширим матрицу (6), введя в нее переменную $X_{i0} = 1$.**
-

- **Тогда матрицу X в развернутом виде можно записать так:**

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix},$$

- Тогда

$$X'X = \begin{pmatrix} n & \sum_i x_{i1} & \dots & \sum_i x_{im} \\ \sum_i x_{i1} & \sum_i x_{i1}^2 & \dots & \sum_i x_{i1} x_{im} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \sum_i x_{im} & \sum_i x_{i1} x_{im} & \dots & \sum_i x_{im}^2 \end{pmatrix} \qquad (8)$$

- И

$$X'Y = \begin{pmatrix} \sum_i y_i \\ \sum_i y_i x_{i1} \\ \cdot \\ \cdot \\ \cdot \\ \sum_i y_i x_{im} \end{pmatrix} \qquad (9).$$

- В частном случае, когда m=2, имеем

$$X'X = \begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{pmatrix} \quad (10).$$

Условный пример

i	y_i	x_{i1}	x_{i2}
1	10	2	1
2	12	2	2
3	17	8	10
4	13	2	4
5	15	6	8
6	10	3	4
7	14	5	7
8	12	3	3
9	16	9	10
10	18	10	11

- Таким образом,

$$Y = \begin{pmatrix} 10 \\ 12 \\ 17 \\ 13 \\ 15 \\ 10 \\ 14 \\ 12 \\ 16 \\ 18 \end{pmatrix}, X = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 8 & 10 \\ 1 & 2 & 4 \\ 1 & 6 & 8 \\ 1 & 3 & 4 \\ 1 & 5 & 7 \\ 1 & 3 & 3 \\ 1 & 9 & 10 \\ 1 & 10 & 11 \end{pmatrix}, тогда$$

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 8 & 2 & 6 & 3 & 5 & 3 & 9 & 10 \\ 1 & 2 & 10 & 4 & 8 & 4 & 7 & 3 & 10 & 11 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 8 & 10 \\ 1 & 2 & 4 \\ 1 & 6 & 8 \\ 1 & 3 & 4 \\ 1 & 5 & 7 \\ 1 & 3 & 3 \\ 1 & 9 & 10 \\ 1 & 10 & 11 \end{pmatrix} =$$

$$\begin{pmatrix} 10 & 50 & 60 \\ 50 & 336 & 398 \\ 60 & 398 & 480 \end{pmatrix}.$$

- Далее

$$X'Y=\begin{pmatrix}1&1&1&1&1&1&1&1&1&1\\2&2&8&2&6&3&5&3&9&10\\1&2&10&4&8&4&7&3&10&11\end{pmatrix}\cdot\begin{pmatrix}10\\12\\17\\13\\15\\10\\14\\12\\16\\18\end{pmatrix}=$$

$$=\begin{pmatrix}137\\756\\908\end{pmatrix}.$$

Замечание

- **Матрицу X′X и вектор X′Y можно получить и по формулам (8) и (9), предварительно подсчитав необходимые суммы:**

$$\sum y_i = 137; \quad \sum x_{i1} = 50; \quad \sum x_{i2} = 60;$$

$$\sum y_i^2 = 1947; \quad \sum x_{i1}^2 = 336; \quad \sum x_{i2}^2 = 480;$$

$$\sum y_i x_{i1} = 756; \quad \sum x_{i1} x_{i2} = 398; \quad \sum y_i x_{i2} = 908.$$

Результат, естественно, будет одинаковым.

- **Определив *X′X* и *X′Y*, находим *a*:**
- **a = (X′X)$^{-1}$X′Y =**

$$=\frac{1}{7160}\begin{pmatrix} 2876 & -120 & -260 \\ -120 & 1200 & -980 \\ -260 & -980 & 860 \end{pmatrix}\cdot\begin{pmatrix} 137 \\ 756 \\ 908 \end{pmatrix}=\begin{pmatrix} 9,3871 \\ 0,1285 \\ 0,61174 \end{pmatrix}.$$

- **Таким образом, искомое уравнение**

$$\hat{y}=9,3871+0,1285x_1+0,61174x_2 .$$

- **Приведем рассчитанные по данному уравнению регрессии значения независимой переменной (y_i) и соответствующие ошибки (e_i):**

- Имеем:

i	y_i	$\hat{y}_i$	e_i
1	10	10,2559	−0,2559
2	12	10,8676	1,1324
3	17	16,5324	0,4676
4	13	12,0910	0,9090
5	15	15,0520	−0,0520
6	10	12,2195	−2,2195
7	14	14,3117	−0,3117
8	12	11,6078	0,3922
9	16	16,6609	−0,6609
10	18	17,4012	0,5988.

- **Здесь $\Sigma y_i = \Sigma \hat{y}_i = 137$; $\Sigma e_i = 0$.**

- **Как можно заключить из примера, наиболее трудоемкой операцией здесь является обращение матрицы.**
- **Уравнение (5) позволяет в принципе решить задачу оценивания для любого конечного числа независимых переменных (на ЭВМ).**