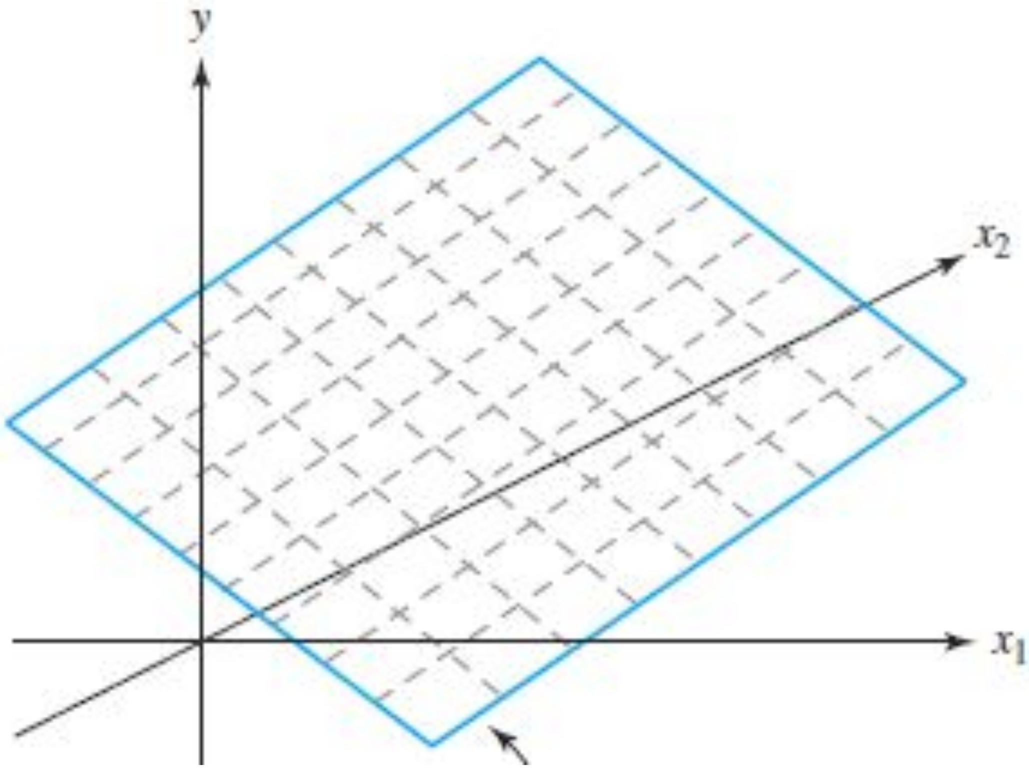
Тема 10-1

Многомерная линейная регрессия

Рябева Е.В.

2015



$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Линейная зависимость от нескольких переменных.

$$y_i = x_{i_1}\theta_1 + x_{i_2}\theta_2 + \ldots + x_{ir}\theta_r + \varepsilon_i$$

n-число проведенных экспериментов, $\vec{\theta}$

Параметров θ может быть несколько: $\theta_1, \theta_2, \ldots, \theta_r$. ***r-число параметров в общем случае может быть произвольно***

x_i – известные величины (факторы), определяющие i-тый номер эксперимента

Имеем линейную зависимость y от параметров θ

Перейдем к матричной форме написания уравнения. Введем обозначения

$$\vec{\theta} = \begin{Bmatrix} \theta_1 \\ \vdots \\ \theta_r \end{Bmatrix}^T ; \quad \vec{y} = \begin{Bmatrix} y_1 \\ \vdots \\ y_n \end{Bmatrix} ; \vec{\varepsilon} = \begin{Bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{Bmatrix}$$

- Запишем систему линейных уравнений в матричном виде

Для случая двух параметров система выглядела таким образом

$$
\begin{aligned}
y_1 &= ax_1 + b + \varepsilon_1 \\
y_2 &= ax_2 + b + \varepsilon_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
y_n &= ax_n + b + \varepsilon_n
\end{aligned}
$$

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}$$

$$\vec{\theta} = \begin{pmatrix} b \\ a \end{pmatrix}^T$$

$$\vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ .. & .. \\ 1 & x_n \end{pmatrix}$$

$$\vec{y} = \mathbf{X}\vec{\theta} + \vec{\varepsilon}$$

Для случая r параметров система

$$\begin{cases} y_1 = x_{11}\theta_1 + x_{12}\,\theta_2 + \ldots + x_{1r}\theta_r + \varepsilon_1 \\ y_2 = x_{21}\theta_1 + x_{22}\theta_2 + \ldots + x_{2r}\theta_r + \varepsilon_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = x_{n1}\theta_1 + x_{n2}\theta_2 + \ldots + x_{nr}\theta_r + \varepsilon_n \end{cases}$$

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}$$

$$\vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}$$

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_r \end{pmatrix}^T$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & \ldots & x_{1r} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nr} \end{pmatrix}$$

$$\vec{y} = \mathbf{X}\vec{\theta} + \vec{\varepsilon}$$

$$\boxed{y - \varepsilon - \mathbf{X}\theta = 0}$$

Эту систему уравнений надо решить относительно *θ*, используя метод максимального правдоподобия и тот факт, что погрешности измерений *yi с дисперсиями σi* подчиняются нормальному распределению, т.е.:

$$\boxed{f(y_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[\frac{(y_i - \eta_i)^2}{2\sigma_i^2}\right] = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\varepsilon_i^2}{\sigma_i^2}\right)}$$

Функция правдоподобия для n измерений имеет вид:

$$L=\prod_{i=1}^{n} f(y_i)=(2\pi)^{-\frac{n}{2}}\left(\prod_{i=1}^{n}\sigma_i^{-1}\right)\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma_i^2}\right\}$$

Логарифмическая функция правдоподобия

$$l=\ln L=-\frac{n}{2}\ln 2\pi+\ln\left(\prod_{i=1}^{n}\sigma_i^{-1}\right)-\frac{1}{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma_i^2}$$

Это выражение достигает максимума, когда последнее слагаемое будет минимально.Условие МНК

$$Q=\sum_{i=1}^{n}\frac{\varepsilon_i^2}{\sigma_i^2}=\sum_{i=1}^{n}\frac{(y_i-\vec{x}_i^{\,T}\vec{\theta})^2}{\sigma_i^2}\to\min$$

Как записать *εi*, используя введённые матричные обозначения

$$\vec{y}-\vec{\varepsilon}-\vec{x}\vec{\theta}=\vec{0}\;\Rightarrow\;\vec{\varepsilon}=\vec{y}-\vec{x}\vec{\theta}\;\Rightarrow\;\varepsilon_{\mathrm{i}}=y_i-\vec{x}_i^{\,T}\vec{\theta}$$

Введём вектор-столбец $\vec{x}_i=(x_{i_1}, x_{i_2}\ldots x_{i_r})$

матрица ошибок.

$$\mathbf{C_y} = \mathbf{C} = \begin{Bmatrix} \sigma_1^2 & \cdots & 0 \\ 0 & \sigma_2^2 & \\ \vdots & \ddots & \\ 0 & \cdots & \sigma_n^2 \end{Bmatrix}$$

Обратная ей матрица называется весовой:

$$g_i = \frac{1}{\sigma_i^2}$$

$$\mathbf{G_y} = \mathbf{G} = \begin{Bmatrix} g_1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & g_n \end{Bmatrix} = \begin{Bmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{Bmatrix}$$

С учётом введённых обозначений для *Gy* можно переписать

$$Q = \sum_{i=1}^{n} \frac{\varepsilon_i^2}{\sigma_i^2} = \vec{\varepsilon}^{\,T} \mathbf{G_y} \vec{\varepsilon} \to \min$$

или

$$Q = (\vec{y} - \mathbf{X}\vec{\theta})^T \mathbf{G_y} (\vec{y} - \mathbf{X}\vec{\theta}) \to \min$$

Для получения min необходимо образовать частные производные Q по θi и приравнять их к 0.

$$\frac{\partial Q}{\partial \theta_i} = 0 \text{ ; i=1...r} \qquad \frac{\partial \varepsilon^{\mathrm{T}}}{\partial \theta} = x^T$$

$$\frac{\partial Q}{\partial \theta} = \frac{\partial \varepsilon^T}{\partial \theta}\mathbf{G_y}\varepsilon + \varepsilon^T\frac{\partial}{\partial \theta}\left(\mathbf{G_y}\varepsilon\right) = \mathbf{x}^T\mathbf{G_y}\varepsilon + \varepsilon^T\mathbf{G_y}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\mathbf{G_y}\varepsilon + \mathbf{x}^{\mathrm{T}}\mathbf{G_y}\varepsilon = 2\mathbf{x}^{\mathrm{T}}\mathbf{G_y}\varepsilon =$$

$$= 2\mathbf{x}^{\mathrm{T}}\mathbf{G_y}(y - x\theta) = 0 \ \Rightarrow$$

$$\mathbf{x}^{\mathrm{T}}\mathbf{G_y}y = \mathbf{x}^{\mathrm{T}}\mathbf{G_y}\mathbf{x}\theta$$

Для получения искомой оценки параметра $\hat{\theta}$

необходимо подействовать оператором $(\mathbf{x}^{\mathrm{T}}\mathbf{G_y}\mathbf{x})^{-1}$

$$\boxed{\hat{\theta} = (\mathbf{x}^{\mathrm{T}}\mathbf{G_y}\mathbf{x})^{-1}\mathbf{x}^{\mathrm{T}}\mathbf{G_y}\,y}$$

Оценка параметров θ МНК линейный случай

Оценка дисперсии оценок параметров

Дисперсия элементов вектора оценок (ковариационная матрица оценок параметров) то же может быть записана в векторном виде

$$\mathrm{var}(\tilde{\theta}) = \mathbf{C}_{\tilde{\theta}} = \sigma_y^{\,2}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

Дисперсию y можно оценить как и двухмерном случае суммой квадратов отклонений от выровненной (рассчитанной) поверхности

$$\tilde{\sigma}_y^2 = S_y^2 = \frac{1}{n-r-1}\sum_{i=1}^{n}\left(y_i - \left(\sum_{j=1}^{r} x_{ij}\theta_j\right)\right)^2$$

диагональные элементы этой матрицы- дисперсии оценок

$$S_{\theta i}^{\;2} = \mathbf{C}_{\tilde{\theta}_{i,i}}$$

случайная величина

$$\boxed{\frac{S_y^2}{\sigma_y^2} \in \chi^2_{n-r-1}}$$

а величины

$$\boxed{\frac{\tilde{\theta}_i}{S_{\theta_i}} \in t_{n-r-1}}$$

могут быть использованы для проверки гипотезы о значимости (отличие от 0) θ

Пример 10-4

на многомерный МНК-потребление –производство –градус-дней

$$\theta = (x^T G_y x)^{-1} x^T G_y y$$

Рассмотрим применение этой оценка к простым случаям

Случай простой линейной регрессии

Зависимость от двух параметров θ0 и θ1 и одной переменной x

$y = \theta_0 + \theta_1 x$

Пусть произвели только 2 измерения

$$\begin{cases} y_1 = x_{11}\theta_1 + \theta_0 \\ y_2 = x_{12}\theta_1 + \theta_0 \end{cases}$$

Можно представить задание в матричной форме:

$$x_{10} = 1$$
$$x_{20} = 1$$

$$x = \begin{pmatrix} x_{11} & 1 \\ x_{12} & 1 \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \theta = (\theta_0, \theta_1)$$

Решение для простой линейной регрессии в матричном виде

$$\boxed{\theta = \left(x^T x\right)^{-1} x^T y}$$

Оценить θ для прямых неравноточных измерений.
В этом случае: $y_i = \theta + \varepsilon_i \quad i = 1, \ldots n.$

$$\hat{\theta} = (x^T G_y x)^{-1} x^T G_y y$$

$$\hat{\theta} = \theta \ ; \ x_0 = 0 \ ;$$

$$x = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \qquad G_y = \begin{pmatrix} g_1 & \\ & g_n \end{pmatrix} \qquad g_i = \frac{1}{\sigma_i^2}$$

$$\hat{\theta} = \left[(1,1\ldots1) \begin{pmatrix} g_1 & \\ & g_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right] \times (1,1\ldots1) \times \begin{pmatrix} g_1 & \\ & g_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \left(\sum_{i=1}^{n} g_i \right)^{-1} \sum_{i=1}^{n} g_i y_i$$