Уравнение множественной регрессии

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + \ldots + a_k x_{kt} + U_t \qquad (8.1)$$

Наилучшая линейная процедура получения оценок параметров уравнения (8.1) и условия, при которых эта процедура дает несмещенные и эффективные оценки, сформулирована в теореме Гаусса-Маркова.

Постановка задачи:

Имеем случайную выборку наблюдений за поведением экономического объекта объемом n.

Выборка

$y_1 \quad x_{11} \quad x_{21} \quad x_{31} \ldots x_{k1}$

$y_2 \quad x_{12} \quad x_{22} \quad x_{32} \ldots x_{k2}$

$y_3 \quad x_{13} \quad x_{23} \quad x_{33} \ldots x_{k3}$

……………………

$y_n \quad x_{1n} \quad x_{2n} \quad x_{3n} \ldots x_{kn}$

Система уравнений наблюдений

$$y_1 = a_0 + a_1 x_{11} + a_2 x_{21} + a_3 x_{31} + \ldots + a_k x_{k1} + u_1$$

$$y_2 = a_0 + a_1 x_{12} + a_2 x_{22} + a_3 x_{32} + \ldots + a_k x_{k2} + u \qquad (8.2)$$

$$y_3 = a_0 + a_1 x_{13} + a_2 x_{23} + a_3 x_{33} + \ldots + a_k x_{k3} + u_3$$

……………………

$$y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + a_3 x_{3n} + \ldots + a_k x_{kn} + u_n$$

Уравнение множественной регрессии

Вводятся обозначения:

$$X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \qquad \vec{A} = \begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_k \end{pmatrix} \qquad \vec{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} \qquad \vec{U} = \begin{pmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{pmatrix}$$

X – матрица коэффициентов при неизвестных параметрах;

Y – вектор выборочных значений эндогенной переменной;

U – вектор выборочных значений случайного возмущения;

A - вектор неизвестных параметров модели.

Найти: $\tilde{A}$, $Cov(\tilde{A}\tilde{A})$, σ_u, $\sigma(\tilde{y}(z))$

Уравнение множественной регрессии

Теорема Гаусса-Маркова.

Если матрица X неколлинеарна и вектор случайных возмущений удовлетворяет следующим требованиям:

1. **$M(u) = 0$**
2. **$\sigma^2(u) = \sigma^2_u$**
3. **$Cov(u_i, u_j) = 0$ при $i \neq j$**
4. **$Cov(x_i, u_i) = 0$**

Уравнение множественной регрессии

Теорема Гаусса-Маркова (Продолжение).

Тогда наилучшей линейной процедурой оценки параметров модели (8.1) является:

$$\hat{A} = \left(X^T X\right)^{-1} X^T \vec{Y}$$

При этом:

$$Cov(\hat{A}, \hat{A}) = \tilde{\sigma}_u^2 \left(X^T X\right)^{-1}$$

$$\tilde{\sigma}_u^2 = \frac{1}{n-k} \sum \left(y_i - \tilde{y}_i\right)^2 = \sum u_i^2$$

$$\tilde{Y}(Z) = \tilde{a}_0 + a_1 z_1 + \ldots + \tilde{a}_k z_k$$

$$\sigma^2(\tilde{y}(Z)) = \tilde{\sigma}_u^2 \left(1 + q_0\right)$$

$$q_0 = \vec{Z}^T \left(X^T X\right)^{-1} \vec{Z}$$

Теорема Гаусса-Маркова

Примеры применения теоремы Гаусса-Маркова

Пример 1. Пусть имеем выборку из n наблюдений за случайной величиной Y.

Найти наилучшие оценки среднего значения и дисперсии этой переменной.

В терминах теоремы Гаусса –Маркова задача формулируется так: необходимо построить модель типа $\mathbf{Y = a_0 + u}$, при этом имеем:

$$X = \begin{pmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} \qquad X^T = \begin{pmatrix} 1 & 1 & \ldots & 1 \end{pmatrix}$$

Примеры применения теоремы Гаусса-Маркова

Пример 1. (Продолжение)

Решение.

$$\left(X^T X\right) = \left(1\ 1\ \dots\ 1\right)\begin{pmatrix}1\\1\\\dots\\1\end{pmatrix} = n \qquad \left(X^T X\right)^{-1} = \frac{1}{n}$$

$$\left(X^T \vec{Y}\right) = \left(1\ 1\ \dots\ 1\right)\begin{pmatrix}y_1\\y_2\\\dots\\y_n\end{pmatrix} = \sum y_i \qquad \tilde{a}_0 = \left(X^T X\right)^{-1}\left(X^T \vec{Y}\right) = \frac{\sum y_i}{n}$$

$$\boldsymbol{\sigma}^2(Y) = \boldsymbol{\sigma}_u^2 = \frac{1}{n-1}\sum\left(y_i - \tilde{a}_0\right)^2 \qquad \boldsymbol{\sigma}^2\left(a_0\right) = \boldsymbol{\sigma}_u^2\left(X^T X\right)^{-1} = \frac{\boldsymbol{\sigma}_u^2}{n}$$

Примеры применения теоремы Гаусса-Маркова

Пример 2. Уравнение парной регрессии.

Построить модель типа **$Y=a_0+a_1x+u$,** по данным выборки наблюдений за переменными Y и x объемом n.

В схеме Гаусса-Маркова имеем:

$$X=\begin{pmatrix}1 & x_1\\ 1 & x_2\\ \dots & \dots\\ 1 & x_n\end{pmatrix}\quad X^T=\begin{pmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{pmatrix}\quad \vec{Y}=\begin{pmatrix}y_1\\ y_2\\ \dots\\ y_n\end{pmatrix}\quad \vec{A}=\begin{pmatrix}a_0\\ a_1\end{pmatrix}\quad \vec{U}=\begin{pmatrix}u_1\\ u_2\\ \dots\\ u_n\end{pmatrix}$$

$$\left(X^T X\right)=\begin{pmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{pmatrix}\begin{pmatrix}1 & x_1\\ 1 & x_2\\ \dots & \dots\\ 1 & x_n\end{pmatrix}=\begin{pmatrix}n & \sum x_i\\ \sum x_i & \sum x_i^2\end{pmatrix}\quad \left(X^T X\right)^{-1}=\frac{1}{n\sum x_i^2-\sum x_i\sum x_i}\begin{pmatrix}\sum x_i^2 & -\sum x_i\\ -\sum x_i & n\end{pmatrix}$$

Примеры применения теоремы Гаусса-Маркова

Пример 2. Уравнение парной регрессии.(Продолжение).

Последовательно вычисляем X^TY и оценку вектора A.

$$\left(X^T \vec{Y}\right) = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

Система нормальных уравнений имеет вид:

$$\left(X^T X\right)\vec{A} = \left(X^T \vec{Y}\right)$$

$$\tilde{\vec{A}} = \frac{1}{n\sum x_i^2 - \sum x_i \sum x_i} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$\tilde{\vec{A}} = \frac{1}{n\sum x_i^2 - \sum x_i \sum x_i} \begin{pmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ n\sum x_i y_i - \sum x_i \sum y_i \end{pmatrix} \qquad (8.3)$$

Примеры применения теоремы Гаусса-Маркова

Пример 2. Уравнение парной регрессии.(Продолжение).

Вычислим дисперсии (ковариационную матрицу) параметров модели.

$$\mathrm{Cov}\left(\tilde{A}, \tilde{A}\right) = \boldsymbol{\sigma}_u^2 \left(X^T X\right)^{-1} = \boldsymbol{\sigma}_u^2 \frac{1}{n\sum x_i^2 - \sum x_i \sum x_i} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}$$

Следовательно:

$$\boldsymbol{\sigma}^2\left(\tilde{a}_0\right) = \boldsymbol{\sigma}_u^2 \frac{\sum x_i^2}{n\sum x_i^2 - \sum x_i \sum x_i}$$

$$\boldsymbol{\sigma}^2\left(\tilde{a}_1\right) = \boldsymbol{\sigma}_u^2 \frac{n}{n\sum x_i^2 - \sum x_i \sum x_i}$$

$$\mathrm{Cov}\left(\tilde{a}_0, \tilde{a}_1\right) = \boldsymbol{\sigma}_u^2 \frac{-\sum x_i}{n\sum x_i^2 - \sum x_i \sum x_i}$$

Примеры применения теоремы Гаусса-Маркова

Пример 2. Уравнение парной регрессии.(Продолжение).

Расчет дисперсии прогнозирования.

Прогноз осуществляется в точке Z={1,z}

$$\boldsymbol{\sigma}^2(Y(z)) = \boldsymbol{\sigma}_u^2(1+q_0)$$

$$q_0 = Z^T (X^T X)^{-1} Z = (1\ z)\frac{1}{n\sum x_i^2 - \sum x_i \sum x_i}\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}\begin{pmatrix} 1 \\ z \end{pmatrix} =$$

$$= \frac{1}{n\sum x_i^2 - \sum x_i \sum x_i}\begin{pmatrix} \sum x_i^2 & -z\sum x_i \\ -\sum x_i & nz \end{pmatrix}\begin{pmatrix} 1 \\ z \end{pmatrix} = \frac{\sum x_i^2 - 2z\sum x_i + n z^2}{n\sum x_i^2 - \sum x_i \sum x_i} = \frac{\sum x_i^2 - \frac{1}{n^2}\sum x_i^2 \sum x_i^2 + (z-\bar{x})^2}{n\sum x_i^2 - \sum x_i \sum x_i}$$

$$\boldsymbol{\sigma}^2(Y(z)) = \boldsymbol{\sigma}_u^2\left(1 + \frac{1}{n} + \frac{(z-\bar{x})^2}{n\sum((x_i - \bar{x})^2)}\right)$$

Примеры применения теоремы Гаусса-Маркова

Пример 2. Уравнение парной регрессии.(Продолжение).