Эконометрика

Уравнение множественной регрессии

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} \dots + a_n x_{nt} + u_t = \sum_{i=0}^{n} a_i x_{it} + u_t \qquad (7.1)$$

Наилучшая линейная процедура получения оценок параметров уравнения (7.1) и условия, при которых эта процедура дает несмещенные и эффективные оценки, сформулирована в теореме Гаусса-Маркова

Карл Фридрих Гаусс
Время жизни
30.04.1777 - 23.02.1855
Научная сфера – математика, физика, астрономия

Андрей Андреевич Марков
Время жизни
14.06.1856 - 20.07.1922
Научная сфера - математика

Теорема Гаусса - Маркова

<u>Постановка задачи:</u>
Имеем случайную выборку наблюдений за поведением экономического объекта объемом **n**

$$\begin{pmatrix} y_1 & x_{11} & x_{21} & \cdots & x_{k1} \\ y_2 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_n & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

Выборка наблюдений за переменными модели (7.1)
Первый индекс – номер регрессора
Второй индекс – номер наблюдения

$$\begin{cases} y_1 = a_0 + a_1 x_{11} + a_2 x_{21} \ldots + a_n x_{k1} + u_1 \\ y_2 = a_0 + a_1 x_{12} + a_2 x_{22} \ldots + a_n x_{k2} + u_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_1 = a_0 + a_1 x_{1n} + a_2 x_{2n} \ldots + a_n x_{kn} + u_n \end{cases} \qquad (7.2)$$

(7.2) - Система уравнений наблюдений, связывающая наблюдения в выборке

Теорема Гаусса - Маркова

Сформируем вектора и матрицу коэффициентов на основе системы (7.2)

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

Y – вектор выборочных значений эндогенной переменной

U – вектор выборочных значений случайного возмущения

A - вектор неизвестных параметров модели

x – вектор регрессоров

X – матрица коэффициентов при неизвестных параметрах

$$\vec{Y} = \vec{A}\vec{x} + \vec{U}$$

Теорема Гаусса - Маркова

По данным выборки найти: Ã, Cov(ÃÃ), σ_u, $\sigma(\tilde{y}(z))$

Теорема (Гаусса – Маркова)

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

1. $M(u_i) = 0$ — Математическое ожидание всех случайных возмущений равно нулю

2. $\sigma^2(u_i) = \sigma_u^2$ — Дисперсия случайных возмущений постоянна во всех наблюдениях (условие **ГОМОСКЕДАСТИЧНОСТИ**)

3. $Cov(u_i, u_j) = 0$ при $i \neq j$ — Случайные возмущения в разных наблюдениях не зависимы

4. $Cov(x_i, u_i) = 0$ — Случайные возмущения и регрессоры не зависимы

Теорема Гаусса - Маркова

Тогда наилучшей линейной процедурой оценки параметров модели (7.1) является:

$$\tilde{A} = \left(X^T X\right)^{-1} X^T \vec{Y} \qquad (7.3)$$

которая удовлетворяет методу наименьших квадратов

При этом:

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

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

$$\tilde{Y}\left(\vec{z}\right) = \tilde{a}_0 + \tilde{a}_1 z_1 + \ldots + \tilde{a}_k z_k$$

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

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

Теорема Гаусса - Маркова

Доказательство

Воспользуемся методом наименьших квадратов

$$S = \sum_{i=1}^{n} u_i^2 = \vec{u}^T \vec{u} \Rightarrow \min \qquad (7.4)$$

где

$$\vec{u} = \vec{Y} - \tilde{\vec{Y}} = \vec{Y} - X\tilde{\vec{a}} \qquad (7.5)$$

Подставив (7.5) в (7.4) получим

$$S = \vec{u}^T \vec{u} = \left(\vec{Y} - X\tilde{\vec{a}}\right)^T \left(\vec{Y} - X\tilde{\vec{a}}\right) = \left(Y^T - \tilde{\vec{a}}^T X^T\right)\left(Y - X\tilde{\vec{a}}\right) =$$

$$= \vec{Y}^T \vec{Y} - 2\tilde{\vec{a}}^T X^T \vec{Y} + \tilde{a}^T X^T X\tilde{\vec{a}} \qquad (7.6)$$

Теорема Гаусса - Маркова

Для получения необходимого условия экстремума дифференцируем (7.6) по вектору параметров

$$\frac{\partial S}{\partial \vec{\tilde{a}}} = -2X^{T}\vec{Y} + 2X^{T}X\vec{\tilde{a}} = 0$$

Откуда система нормальных уравнений для определения искомых параметров получает вид

$$X^{T}X\vec{\tilde{a}} = X^{T}\vec{Y} \qquad (7.7)$$

Решение системы (7.7) в матричном виде есть

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

Выражение (7.3) доказано

Теорема Гаусса - Маркова

Докажем несмещенность оценок (7.3)

$$M\left(\tilde{a}\right) = M\left[\left(X^{T} X\right)^{-1} X^{T} Y\right] = M\left[\left(X^{T} X\right)^{-1} X^{T}\left(X a + u\right)\right] =$$

$$= M\left[\left(X^{T} X\right)^{-1}\left(X^{T} X\right) a + u\right] = a$$

Несмещенность оценки (7.3) доказана

Вычислим ковариационную матрицу оценок (7.3)

$$\mathrm{Cov}\left(\tilde{a}, \tilde{a}\right) = \mathrm{Cov}\left[\left(X^{T} X\right)^{-1} X^{T} Y, \left(X^{T} X\right)^{-1} X^{T} Y\right] = \sigma_u^2 \left(X^{T} X\right)^{-1}$$

В результате получено выражение (7.4)

Теорема 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 **Y = a_0 +u,** при этом имеем:

$$X = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} \qquad Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \qquad X^T = (1\ 1\ \dots\ 1)$$

Теорема Гаусса - Маркова

Решение

1. Вычисляем $(X^T X)^{-1}$

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

2. Вычисляем $(X^T Y)$

$$\left(X^T Y\right) = \left(1\ 1\ \ldots\ 1\right)\begin{pmatrix}y_1\\y_2\\\ldots\\y_n\end{pmatrix} = \sum y_i$$

3. Вычисляем оценку параметра a_0

$$\tilde{a}_0 = \left(X^T X\right)^{-1} X^T Y = \frac{1}{n}\sum_{i=1}^{n} y_i$$

4. Находим дисперсию среднего

$$\sigma^2\left(\tilde{a}_0\right) = \sigma_u^2\left(X^T X\right)^{-1} = \frac{\sigma_u^2}{n-1}$$

Теорема Гаусса - Маркова

<u>Пример 2.</u> Уравнение парной регрессии

Построить модель типа $\mathbf{Y=a_0+a_1x+u}$, по данным выборки наблюдений за переменными Y и x объемом n

В схеме Гаусса-Маркова имеем:

$$X=\begin{pmatrix}1 & x_1\\ 1 & x_2\\ \dots & \dots\\ 1 & x_n\end{pmatrix}\quad X^T=\begin{pmatrix}1 & 1 & \dots & 1\\ x_1 & x_2 & \dots & x_n\end{pmatrix}\quad \vec{Y}=\begin{pmatrix}y_1\\ y_2\\ \dots\\ y_n\end{pmatrix}\quad \vec{A}=\begin{pmatrix}a_0\\ a_1\end{pmatrix}\quad \vec{U}=\begin{pmatrix}u_1\\ u2\\ \dots\\ u_n\end{pmatrix}$$

1. Вычисляем матрицы (X^TX) и $(X^TX)^{-1}$

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

Теорема Гаусса - Маркова

2. Вычисляем $X^T Y$

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

3. Вычисляем оценку вектора параметров a

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

$$\tilde{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \frac{1}{n\sum x_i^2 - \sum x_i \sum x_i} \begin{pmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ n\sum x_i y_i - \sum x_i \sum y_i \end{pmatrix}$$

Теорема Гаусса - Маркова

Вычислим дисперсии (ковариационную матрицу) параметров модели

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

Следовательно:

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

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

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

Теорема Гаусса - Маркова

Расчет дисперсии прогнозирования

Прогноз осуществляется в точке $Z=\{1,z\}^T$

$$\sigma^2(Y(z)) = \sigma_u^2(1+q_0)$$

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

$$= \frac{1}{n\sum x_i^2 - \sum x_i \sum x_i}\begin{pmatrix} \sum x_i^2 & -z\sum x_i \\ -\sum x_i & nz \end{pmatrix}\begin{pmatrix} 1 \\ z \end{pmatrix} =$$

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

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

Оценка уравнений регрессии с помощью EXCEL

Процедура «ЛИНЕЙН» в приложении EXCEL

Алгоритм использования процедуры:

1. Подготовка таблицы исходных данных
2. Вызов процедуры «ЛИНЕЙН»
3. Ввод исходных данных в процедуру
4. Анализ результата

Рассмотрим алгоритм на примере

Теорема Гаусса - Маркова

Выводы:

1. Теорема Гаусса-Маркова формулирует наилучшую линейную процедуру расчета оценок параметров линейной модели множественной регрессии

2. Линейная процедура соответствует методу наименьших квадратов

3. Предпосылки теоремы обеспечивают получение оценок, обладающих свойствами несмещенности и эффективности

4. При выполнении предпосылок свойства эффективности и несмещенности достигаются при любом законе распределения случайного возмущения