

*T.S. Kagramanyan, L.O. Babeshko*

## ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИОННОЙ МОДЕЛИ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ В EXCEL

*Аннотация:* в данной статье рассмотрены проблемы, связанные с применением регрессии. Особое внимание обращается на трудности, связанные с регрессионным анализом: ошибочная оценка условий применимости метода наименьших квадратов; неправильный выбор альтернативных методов при нарушении условий применимости метода наименьших квадратов; применение регрессионного анализа без глубоких знаний о предмете исследования.

*Ключевые слова:* статистика, диапазон, метод наименьших квадратов, регрессионный анализ, линейный метод наименьших квадратов, константа.

*T.S. Kagramanyan, L.O. Babeshko*

## ESTIMATION PARAMETER REGRESSION MODEL BY LEAST SQUARES IN EXCEL

*Abstract:* the article deals with the problems associated with the use of regression. Much attention is given to the difficulties associated with the regression analysis. Report assessment of the conditions of applicability of the method of least squares; wrong choice of alternative methods in violation of the conditions of applicability of the method of least squares; the use of regression analysis without in-depth knowledge about the subject.

*Keywords:* statistics, Ordinary Least Squares, regression analysis, linear least squares, range, constant.

Метод наименьших квадратов (МНК, англ. *Ordinary Least Squares, OLS*) – один из методов оценки параметров регрессионных моделей. Достоинством метода являются – статистические свойства МНК-оценок (при выполнении предпосылок Гаусса-Маркова – несмещенность и эффективность), простота математических выводов и практической реализации.

МНК позволяет решить задачу «наилучшего» приближения выборочных данных  $X_t, Y_t, t = 1, \dots, n$  линейной функцией:

$$f(X) = a + bX \quad (1)$$

– для парной регрессии.

Смысл «наилучшего» приближения определяется выбором критерия. В методе наименьших квадратов – это сумма квадратов отклонений (остатков)

$$F(\hat{a}, \hat{b}) = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - (\hat{a} + \hat{b}X_t))^2,$$

и оценки параметров  $\hat{a}$  и  $\hat{b}$  должны быть подобраны таким образом, чтобы функция  $F(\hat{a}, \hat{b})$  была минимальна:

$$F(\hat{a}, \hat{b}) = \sum_{t=1}^n e_t^2 \rightarrow \min. \quad (2)$$

Для решения задачи (2) (задача на безусловный экстремум) составляются *необходимые условия экстремума (First Order Condition)*:

$$\begin{cases} \frac{\partial F}{\partial \hat{a}} = -2 \sum (Y_t - \hat{a} - \hat{b}X_t) = 0 \\ \frac{\partial F}{\partial \hat{b}} = -2 \sum X_t (Y_t - \hat{a} - \hat{b}X_t) = 0 \end{cases},$$

которые, можно также записать следующим образом:

$$\begin{cases} \sum (Y_t - \hat{a} - \hat{b}X_t) = 0 \\ \sum X_t (Y_t - \hat{a} - \hat{b}X_t) = 0 \end{cases}, \quad (3)$$

или в компактной форме:

$$\begin{cases} \sum e_t = 0 \\ \sum X_t e_t = 0 \end{cases}.$$

Система (3) называется системой нормальных уравнений. В (3) столько уравнений, сколько параметров требуется оценить по выборочным данным. Из решения системы нормальных уравнений находятся МНК-оценки параметров:

$$\begin{cases} \sum Y_t - \hat{a}n - \hat{b} \sum X_t = 0 \\ \sum X_t Y_t - \hat{a} \sum X_t - \hat{b} \sum X_t^2 = 0 \end{cases}$$

$$\hat{a} = \frac{1}{n} \sum Y_t - \hat{b} \frac{1}{n} \sum X_t = \bar{Y} - \hat{b} \bar{X},$$

где  $\bar{X}$  и  $\bar{Y}$  – средние значения по выборке:

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t, \quad \bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t.$$

Подстановка, полученного для  $\hat{a}$  выражения, во второе уравнение системы нормальных уравнений

$$\sum X_t Y_t - \frac{1}{n} (\sum Y_t)(\sum X_t) + \hat{b} \frac{1}{n} (\sum X_t)(\sum X_t) - \hat{b} \sum X_t^2 = 0,$$

приводит к следующей оценке параметра  $b$

$$\hat{b} = \frac{n \sum X_t Y_t - (\sum X_t)(\sum Y_t)}{n \sum X_t^2 - (\sum X_t)^2} = \frac{\sum x_t \cdot y_t}{\sum x_t^2},$$

где  $x_t = X_t - \bar{X}$ ,  $y_t = Y_t - \bar{Y}$  – значения переменных центрированные по средним выборочным;

Таким образом, МНК – оценки параметров парной регрессионной модели выражаются через выборочные данные следующим образом:

$$\left. \begin{aligned} \hat{b} &= \frac{\sum x_t y_t}{\sum x_t^2} \\ \hat{a} &= \frac{1}{n} \sum Y_t - \frac{1}{n} \sum X_t \hat{b} = \bar{Y} - \hat{b} \bar{X} \end{aligned} \right\} \quad (4)$$

#### *Реализация регрессионного анализа в программе MS Excel*

Для проведения расчетов по линейному методу МНК можно использовать программу *Microsoft Excel* (входит в программный пакет *Microsoft Office*).

Наиболее просто реализуются вычисления коэффициентов линейной регрессионной модели (1).

Для этого можно использовать следующие встроенные функций *MS Excel*:

ОТРЕЗОК(диапазон\_Y;диапазон\_X)

НАКЛОН(диапазон\_Y;диапазон\_X)

КОРРЕЛ(диапазон\_Y;диапазон\_X)

Первая функция вычисляет свободный член уравнения регрессии ( $a$  в выражении (1), вторая – наклон прямой ( $b$  в выражении (1)). Третья функция позволяет вычислить коэффициент корреляции.

Каждая из функций принимает два аргумента, разделяемых знаком точка с запятой «;». Каждый из аргументов определяет диапазон ячеек, в котором находятся значения зависимой (диапазон\_У) и независимой (диапазон\_Х) переменных. Диапазоны должны быть одинаковой формы (вектор-строка или вектор-столбец одинаковой длины).

В более общем виде линейный МНК может быть реализован с помощью встроенной функции ЛИНЕЙН, которая производит вычисления коэффициентов линейной регрессии и дополнительно рассчитывает ряд статистических показателей. Вычисленные коэффициенты регрессии и статистики возвращаются в виде массива чисел. Поскольку возвращается массив значений, функция должна задаваться в виде формулы массива.

Функция ЛИНЕЙН может принимать от одного до четырех аргументов. Обязателен только первый аргумент, остальные – необязательные:

ЛИНЕЙН(диапазонУ, [диапазонХ], [константа], [статистика])

*ДиапазонУ* – обязательный аргумент. Диапазон ячеек, содержащий множество значений зависимой переменной ( $y$ );

*ДиапазонХ* – диапазон ячеек, содержащий множество значений независимых переменных. Если переменных несколько, то они должны располагаться в смежных ячейках. Каждый диапазон значений независимой переменной должен иметь форму, аналогичную диапазонуУ.

*Константа*. Необязательный аргумент. Логическое значение, которое указывает, требуется ли, чтобы константа  $a$  была равна 0. Если аргумент константа имеет значение ИСТИНА или опущен, то свободный член вычисляется обычным образом.

Если аргумент константа имеет значение ЛОЖЬ, то значение  $a$  полагается равным 0 и значения коэффициентов регрессии подбираются с этим условием.

*Статистика*. Необязательный аргумент. Логическое значение, которое указывает, требуется ли вернуть дополнительную регрессионную статистику. Если аргумент *статистика* имеет значение ИСТИНА, функция ЛИНЕЙН возвращает дополнительную регрессионную статистику. Возвращаемый массив чисел будет иметь следующий вид:

$\hat{b}$	$\hat{a}$
$S_{\hat{b}}$	$S_{\hat{a}}$
$R^2$	$s = \hat{\sigma}$
$F$	$\nu_2$
$RSS$	$ESS$

Если аргумент *статистика* имеет значение ЛОЖЬ или опущен, функция ЛИНЕЙН возвращает только коэффициенты (то есть, вектор-строку). Размер диапазона ячеек, в которые будет записан результат выполнения функции ЛИНЕЙН следующий:

1. Если *статистика*=ЛОЖЬ, то 1 строка и  $n$  столбцов ( $n$ -число определяемых параметров).
2. Если *статистика*=ИСТИНА, то 5 строк и  $k$  столбцов (число столбцов равно числу оцениваемых параметров, для парной регрессии – 2).

Описание значений, вычисляемых функцией, приведены в таблице ниже.

Таблица 1

Величина	Описание
$\hat{b}, \hat{a}$	МНК-оценки параметров.
$S_{\hat{b}}, S_{\hat{a}}$ и т. д.	Стандартные значения ошибок для коэффициентов $b; a; \dots$
$R^2$	<i>Коэффициент детерминации</i> . Он характеризует тесноту связи между результативным показателем и набором факторных показателей. Принимает только положительные значения в пределах от 0 до 1. Чем ближе значение коэффициента к 1, тем больше теснота связи. И, наоборот, чем ближе к 0, тем зависимость меньше.
$\hat{\sigma}$	Оценка ско возмущения.
$F$	$F$ -статистика или $F$ -наблюдаемое значение. $F$ -статистика используется для определения того, является ли случайной наблюдаемая взаимосвязь между зависимой и независимой переменными.

$\nu$	Степени свободы. Степени свободы полезны для нахождения $F$ -критических значений в статистической таблице. Для определения уровня надежности модели необходимо сравнить значения в таблице с $F$ -статистикой, возвращаемой функцией <i>ЛИНЕЙН</i> .
RSS	Регрессионная сумма квадратов.
ESS	Остаточная сумма квадратов, равна сумме квадратов разностей для каждой точки между прогнозируемым значением $y$ и фактическим значением $y$ .

### Практическая часть

*Пример.* По территориям региона приводятся данные за 20XX г.

Таблица 2

Номер региона	Среднедневная заработная плата, руб., $y$	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., $x$
1	133	78
2	148	82
3	134	87
4	154	79
5	162	89
6	195	106
7	139	67
8	158	88
9	152	73
10	162	87
11	159	76
12	173	115

Используя функцию *ЛИНЕЙН*, оценим регрессионную модель зависимости размера средней заработной платы в регионе от среднедушевого прожиточного минимума:

$\hat{b}$	0,920431	76,97649	$\hat{a}$
$S_{\hat{b}}$	0,279716	24,21156	$S_{\hat{a}}$
$R^2$	0,519877	12,54959	$s = \hat{\sigma}$
$F$	10,82801	10	$\nu_2$
RSS	1705,328	1574,922	ESS

Можем записать стандартную форму записи оцененной модели:

$$Y = 76.98 + 0.92 \times X_t + e_t$$

(24.21)    (0.28)    (12.55)

Таким образом, в нашем случае коэффициент детерминации равен 0,52, следовательно, оцененная модель среднего качества.

С увеличением среднедневной зарплаты среднедушевой прожиточный минимум увеличивается на 0,92 процентных пунктов.

### *Список литературы*

1. Бабешко Л.О. Основы эконометрического моделирования. – М.: Ком-Книга, 2010. – 432 с.
2. Бородич С.А. Эконометрика. – Минск: Новое знание, 2001. – 408 с.
3. Орлов А.И. Эконометрика. – М.: Экзамен, 2002.
4. Кремер Н.Ш. Эконометрика: Учебник для вузов / Н.Ш. Кремер, Б.А. Путко; под ред. проф. Н.Ш. Кремера. – М.: ЮНИТА-ДАНА, 2005. – 311 с.

---

**Каграманян Татевик Самвеловна** – студентка ФГОБУ ВО «Финансовый университет при Правительстве РФ», Россия, Москва.

**Kagramanyan Tatevik Samvelovna** – student of Financial university under the government of the Russian Federation, Russia, Moscow.

**Бабешко Людмила Олеговна** – д-р экон. наук, профессор ФГОБУ ВО «Финансовый университет при Правительстве РФ», Россия, Москва.

**Babeshko Ludmila Olegovna** – professor of Financial university under the government of the Russian Federation, Russia, Moscow.

---