ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИОННОЙ МОДЕЛИ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ В EXCEL

Ганиева Зулфия Самиевна - Самаркандский институт экономики и сервиса, ассистент
Садиева Самира Сухробовна - студентка Самаркандский институт экономики и сервиса
Равшанова Шахзода Равшановна - студентка Самаркандский институт экономики и сервиса

АННОТАЦИЯ: рассмотрена задача с применением регрессии. Особое внимание обращается на упрощение способа вычислений, связанного с регрессионным анализом: ошибочная оценка условий применимости метода наименьших квадратов; неправильный выбор альтернативных методов при нарушении условий применимости метода наименьших квадратов.

Ключевые слова: метод наименьших квадратов, регрессионная модель, программу Microsoft Excel, функции MS Excel.

Key words: least squares method, regression model, Microsoft Excel program, MS Excel functions

Annotation: a problem using regression is considered. Particular attention is paid to simplifying the method of calculations associated with regression analysis: erroneous assessment of the conditions for the applicability of the least squares method; incorrect choice of alternative methods when the conditions of applicability of the least squares method are violated

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 ГауссаМаркова – 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:

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

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

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

где $e_t^{\,2}$ квадраты отклонений величин. Оценки параметров $\hat{a}\,u\,\hat{b}$ должны быть подобраны таким образом, чтобы функция $F(\hat{a},\hat{b})$ была минимальной:

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

Для решения последней задачи, которая является задачей на безусловный экстремум, составляются необходимые условия экстремума (First Order Condition)

$$\begin{cases} \dfrac{\partial F}{\partial \hat{a}} = \sum_{t=1}^{n} \left(Y_t - \left(\hat{a} + \hat{b}\,\hat{X}_t\right)\right)'_{\hat{a}} = -2\sum \left(Y_t - \left(\hat{a} + \hat{b}\,\hat{X}_t\right)\right) = 0 \\ \dfrac{\partial F}{\partial \hat{b}} = \sum_{t=1}^{n} \left(Y_t - \left(\hat{a} + \hat{b}\,\hat{X}_t\right)\right)'_{\hat{b}} = -2X_t\sum \left(Y_t - \left(\hat{a} + \hat{b}\,\hat{X}_t\right)\right) = 0 \end{cases}$$

Производя некоторые преобразования систему уравнений можно записать в виде:

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

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

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:

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

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

где $\overline{X}$ и $\overline{Y}$- средние значения по выборке:

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

Подставив для а $\hat{a}$ выражения, во второе уравнение системы нормальных уравнений

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

приходим к следующей оценке параметра b

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

где $x_t = X_t - \overline{X}$, $y_t = Y_t - \overline{Y}$ – значения переменных центрированные по средним выборочным; Таким образом, МНК –оценки параметров парной регрессионной модели выражаются через выборочные данные следующим образом:

$$\hat{b} = \frac{\sum x_t y_t}{\sum x_t^{\,2}}$$

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

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 и является мощным табличным редактором, дающий высокие результаты вычислений и предоставляющий доступный сервис пользователю). Наиболее просто реализуются

вычисления коэффициентов линейной регрессионной модели (1). Для этого можно использовать следующие встроенные функций MS Excel:
ОТРЕЗОК(INTERCEPT) (диапазон_Y; диапазон_X) – определяет точку пересечения линейного тренда с осью ординат;
НАКЛОН (SLOPE)(диапазон_Y; диапазон_X) –определяет коэффициент наклона линейного тренда;
КОРРЕЛ(диапазон_Y;диапазон_X) –вычисляет 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_Y) и независимой (диапазон_X) 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 (диапазонY, [диапазонX], [константа], [статистика])
-ДиапазонY – обязательный аргумент.
Диапазон ячеек, содержащий множество значений зависимой переменной (y);
-ДиапазонX – 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 диапазонуY.
-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 возвратить 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	$v2$
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.

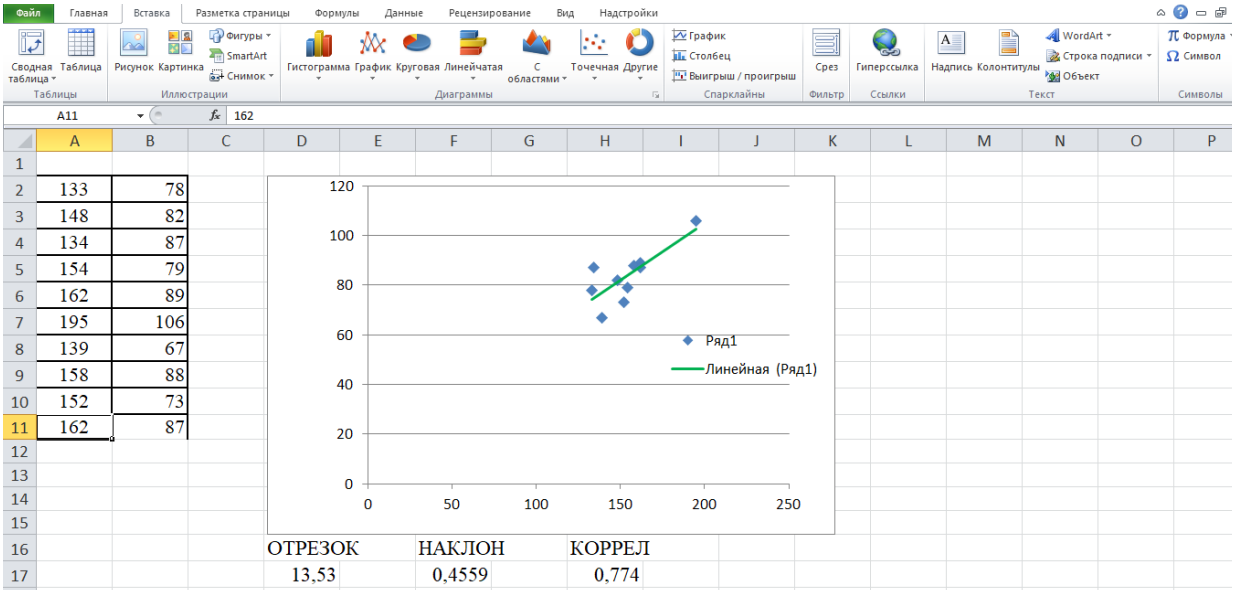
Величина	Описание
$\hat{b}, \hat{a}$	МНК-оценки параметров.
$S_{\hat{b}}$, $S_{\hat{a}}$ и т. д.	Стандартные значения ошибок для коэффициентов *b*; *a*; ...
R^2	*Коэффициент детерминации*. Он характеризует тесноту связи между результативным показателем и набором факторных показателей. Принимает только 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 возмущения.
F	*F*-статистика или *F*-наблюдаемое значение. *F*-статистика используется для определения того, является ли случайной наблюдаемая взаимосвязь между зависимой и независимой переменными.
v	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 *ЛИНЕЙН*.
RSS	Регрессионная сумма квадратов.
ESS	Остаточная сумма квадратов, равна сумме квадратов разностей для каждой точки между прогнозируемым значением y и фактическим значением y.

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:

b	0,920431	76,97649	a
$s_{\hat{b}}$	0,279716	24,21156	$s_{\hat{a}}$
R^2	0,519877	12,54959	$s = \sigma$
F	10,82801	10	v_2
RSS	1705,328	1574,922	ESS

Тогда уравнение модели будет записано в виде:

$$Y = 76{,}98 + 0{,}92\, X_t + e_t$$

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

Литература:

1. Бабешко Л.О. Основы эконометрического моделирования. – М.: КомКнига, 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.
5. Гарнаев А.Ю., Использование MS EXCEL и VBA в экономике и финансах. – СПб.: БХВ - Петербург, 1999.–332с.