Методы регрессии и получаемые на их основе инварианты

к.э.н. Ивашнев Л.И.
Университет машиностроения
8(985) 284-26-98

Аннотация. Статья содержит изложение инвариантов, полученных на основе трех базовых и трех взвешенных методов линейной множественной регрессии: метода наименьших квадратов, метода получения уравнений регрессии без свободного члена и метода получения уравнений регрессии общего вида. Статья содержит формулы, которые используются для расчета указанных инвариантов. В статье дан пример расчета значений инвариантов по данным представленной выборки.

Ключевые слова: *формулы, инварианты, базовые и взвешенные методы, метод наименьших квадратов, метод получения уравнений регрессии без свободного члена, метод получения уравнений регрессии общего вида.*

В учебном пособии [3] изложены шесть методов линейной множественной регрессии, которые можно разделить на:

три *базовых метода*:

- базовый метод наименьших квадратов;
- базовый метод получения вырожденных уравнений регрессии;
- базовый метод получения уравнений регрессии общего вида,

и три *метода взвешенной линейной множественной регрессии*:

- взвешенный метод наименьших квадратов;
- взвешенный метод получения вырожденных уравнений регрессии;
- взвешенный метод получения уравнений регрессии общего вида.

Также в этом пособии дана матричная форма каждого из этих методов.

На основе этих методов можно определить несколько важных инвариантов, то есть констант, характеризующих исследуемую выборку. Рассмотрим их.

Поскольку при отсутствии факторов, то есть только при наличии значений показателя Y базовое уравнение регрессии принимает вид:

$$Y = a_0, \tag{1}$$

то система нормальных уравнений для метода наименьших квадратов превращается в одно уравнение:

$$na_0 = \sum_{i=1}^{n} y_i , \tag{2}$$

откуда

$$a_0 = \frac{\sum_{i=1}^{n} y_i}{n} . \tag{3}$$

Из выражения (3) видно, что коэффициент a_0 уравнения (2) представляет собой **среднее арифметическое значений показателя *Y***.

В случае вырожденного метода регрессии для показателя Y и фактора X уравнение регрессии принимает вид:

$$Y = a_1 X , \tag{4}$$

а система нормальных уравнений превращается в уравнение:

$$a_1 \sum_{i=1}^{n} x_{1i} x_{1i} = \sum_{i=1}^{n} x_{1i} y_i , \tag{5}$$

откуда получаем выражение для a_1:

$$a_1 = \frac{\sum_{i=1}^{n} x_{1i} y_i}{\sum_{i=1}^{n} x_{1i} x_{1i}}. \qquad (6)$$

Вполне возможно, что выражение (6) может использоваться для оценки **уровня корреляции случайных величин *X* и *Y***.

Уравнение регрессии общего вида в случае одного фактора имеет вид:

$$a_1 X_1 = 1, \qquad (7)$$

поэтому система нормальных уравнений для получения таких уравнений регрессии состоит из одного уравнения и принимает следующий вид:

$$a_1 \sum_{i=1}^{n} x_{1i} x_{1i} = \sum_{i=1}^{n} x_{1i}, \qquad (8)$$

откуда получаем выражение для коэффициента a_1:

$$a_1 = \frac{\sum_{i=1}^{n} x_{1i}}{\sum_{i=1}^{n} \left(x_{1i}\right)^2}. \qquad (9)$$

Из выражения (9) получаем:

$$a_1 = \frac{\dfrac{\sum_{i=1}^{n} x_{1i}}{n}}{\dfrac{\sum_{i=1}^{n} \left(x_{1i}\right)^2}{n}},$$

то есть коэффициент a_1 представляет собой отношение первого начального момента случайной величины X ко второму начальному моменту этой же величины, вычисленным по заданной выборке:

$$a_1 = \frac{M_1}{M_2}. \qquad (10)$$

Отмечаем, что этот инвариант зависит только от значений рассматриваемой случайной величины.

Если требуется получить взвешенное уравнение регрессии в случае одного показателя Y, то соответствующее уравнение принимает вид:

$$Y = a_0, \qquad (11)$$

Система нормальных уравнений, предназначенная для получения таких взвешенных уравнений регрессии, имеет вид:

$$a_0 \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} p_i y_i, \qquad (12)$$

откуда получаем:

$$a_0 = \frac{\sum_{i=1}^{n} p_i y_i}{\sum_{i=1}^{n} p_i}. \qquad (13)$$

Легко видеть, что выражение (13) представляет собой **взвешенную среднюю значений показателя *Y***.

Если требуется взвешенное вырожденное уравнение регрессии для одного показателя Y

и одного фактора *X*, то система нормальных уравнений превращается в одно уравнение:

$$a_1 \sum_{i=1}^{n} p_i x_i x_i = \sum_{i=1}^{n} p_i x_i y_i \,. \qquad (14)$$

Из этого уравнения получаем еще один инвариант:

$$a_1 = \frac{\sum_{i=1}^{n} p_i x_i y_i}{\sum_{i=1}^{n} p_i x_i x_i}\,, \qquad (15)$$

который видимо можно использовать для взвешенной оценки уровня корреляции случайных величин *X* и *Y*.

И наконец, если требуется взвешенное уравнение регрессии общего вида для одного фактора X, то система нормальных уравнений превращается в одно уравнение

$$a_1 \sum_{i=1}^{n} p_i x_i x_i = \sum_{i=1}^{n} p_i x_i \,. \qquad (16)$$

Из этого уравнения получаем еще один инвариант:

$$a_1 = \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i x_i x_i}\,, \qquad (17)$$

разделив числитель и знаменатель которого на $\sum_{i=1}^{n} p_i$ получаем:

$$a_1 = \frac{\dfrac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} p_i}}{\dfrac{\sum_{i=1}^{n} p_i x_i x_i}{\sum_{i=1}^{n} p_i}} = \frac{M_1}{M_2}\,,$$

т.е. коэффициент a_1 взвешенного уравнения регрессии общего вида от выборки, состоящей из значений одной величины *X*, равен **отношению первого взвешенного момента величины *X* к ее второму взвешенному моменту**

$$a_1 = \frac{M_1}{M_2}\,. \qquad (18)$$

Отмечаем, что этот инвариант измеряется отношением средневзвешенной величины X к средневзвешенной величине X^2 и зависит только от значений случайной величины *X*.

В качестве примера рассмотрим следующую задачу.

По данным следующей выборки рассчитать основные инварианты, характеризующие следующие случайные величины (таблица 1):

Таблица 1

X_1	2,5797	0,3678	1,0706	2,9002	0,2893	2,9431	1,861
X_2	1,505637	3,367738	3,76242	3,112084	1,145446	0,835847	1,739339
P	1	2	3	4	5	6	7

По данным выборки рассчитаем указанные инварианты. Получаем следующие оценки (таблица 2).

Таблица 2

Показатель	X_1	X_2
Среднее арифметическое $a_0 = \sum_{i=1}^{n} y_i / n$	1,715957	2,209787
$a_1 = \sum_{i=1}^{n} x_{1i} y_i / \sum_{i=1}^{n} x_{1i} x_{1i}$	0,847613	0,569718
Отношение моментов $a_1 = \sum_{i=1}^{n} x_{1i} / \sum_{i=1}^{n} x_{1i} x_{1i}$	0,420632	0,36409
Среднее взвешенное $a_0 = \sum_{i=1}^{n} p_i y_i / \sum_{i=1}^{n} p_i$	2,015096	1,564774
$a_1 = \sum_{i=1}^{n} p_i x_i y_i / \sum_{i=1}^{n} p_i x_i x_i$	0,776047	0,678002
Отношение моментов $a_1 = M_1 / M_2 = \sum_{i=1}^{n} p_i x_i / \sum_{i=1}^{n} p_i x_i x_i$	0,416607	0,397535

Отмечаем, что из полученных инвариантов широко известными являются только среднее арифметическое и среднее взвешенное. Смысл и возможное применение остальных инвариантов еще предстоит выяснить.

Литература

1. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов. – М.: ЮНИТИ-ДАНА, 2003. –543 с.
2. Ивашнев Л.И. Методы регрессии в экономической математике: Монография. – М.: Изд-во МГОУ, 2005. – 200 с.
3. Ивашнев Л.И. Методы и модели в экономике: Учеб. пособие. – М.: Издательский дом «Лидер-М», 2011. – 328 с.