РАЗДЕЛ IX. МАТЕМАТИКА

Базилевский М.П.

Интерпретация коэффициентов линейно-мультипликативных регрессионных моделей*Иркутский государственный университет путей сообщения
(Россия, Иркутск)*

doi: 10.18411/sr-10-04-2018-41

idsp: 000001:sr-10-04-2018-41

Аннотация

В статье рассмотрена линейно-мультипликативная регрессионная модель, являющаяся расширенной формой линейной регрессии. Предложен способ интерпретации коэффициентов линейно-мультипликативной регрессии. Разработанный способ продемонстрирован на примере решения задачи моделирования перевозок крупнотоннажных контейнеров на железнодорожном пункте пропуска Забайкальск – Маньчжурия.

Ключевые слова: линейная регрессия, линейно-мультипликативная регрессия, интерпретация, объемы импорта крупнотоннажных контейнеров.

Abstract

The paper considers a linear-multiplicative regression model, which is an extended form of linear regression. A method for interpreting coefficients of linear-multiplicative regression is proposed. The developed method is demonstrated on the example of solution of problem of modeling transportation of large-capacity containers at the Zabaikalsk-Manchuria railway checkpoint.

Keywords: linear regression, linear-multiplicative regression, interpretation, import volumes of large-capacity containers.

Наиболее популярная на практике линейная регрессионная модель имеет вид:

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} + \varepsilon_i, \quad i = \overline{1, n}, \quad (1)$$

где n – объем выборки; l – количество независимых переменных; y_i и x_{ij} , $i = \overline{1, n}$, $j = \overline{1, l}$ – известные значения зависимой и независимых переменных соответственно; α_i , $i = \overline{0, l}$ – подлежащие оцениванию параметры; ε_i , $i = \overline{1, n}$ – ошибки аппроксимации.

В линейной регрессии (1) оценка параметра α_j , $j = \overline{1, l}$, интерпретируется как число, которое показывает, на какую величину в среднем изменится значение переменной y , если значение переменной x_j изменить на одну единицу (при неизменных значениях остальных объясняющих переменных). Такая трактовка естественна при взятии частной производной функции (1) по переменной x_j :

$$\frac{\partial y}{\partial x_j} = \alpha_j,$$

которая показывает скорость изменения функции относительно изменения аргумента x_j , если остальные аргументы остаются неизменными. Как видно, в случае линейной регрессии эта скорость постоянна.

Используя все возможные комбинации произведений независимых переменных, от уравнения (1) можно перейти к линейно-мультипликативной регрессии (ЛМР) [1,2]:

$$y_i = \alpha_0 + \sum_{j=1}^p \alpha_j \prod_{k=1}^l x_{ik}^{\lambda_{jk}} + \varepsilon_i, \quad i = \overline{1, n}, \quad (2)$$

где p – количество слагаемых, причем, $p = 2^l - 1$;

λ_{jk} – булева переменная, заданная по правилу:

$$\lambda_{jk} = \begin{cases} 1, & \text{если } k\text{-я независимая переменная входит в } j\text{-е слагаемое,} \\ 0, & \text{в противном случае.} \end{cases}$$

В выражении (2) величина $\prod_{k=1}^l x_{ik}^{\lambda_{jk}}$ называется j -й мультипликацией

(произведением) регрессоров порядка g , если для неё справедливо равенство $\sum_{k=1}^l \lambda_{jk} = g$

. Каждая мультипликация состоит мультипликаторов (множителей, регрессоров).

Удобно представлять числа λ_{jk} как элементы бинарной матрицы Λ , в строках которой размещены все возможные комбинации «0» и «1» по l , кроме комбинации из всех нулей.

По сути своей, регрессия (2) является расширенной версией линейной модели (1). Тогда каким же образом интерпретировать её коэффициенты?

Для ЛМР частная производная функции (2) по переменной x_k имеет вид:

$$\frac{\partial y}{\partial x_k} = \sum_{j=1}^p \alpha_j \lambda_{jk} x_k^{\lambda_{jk}-1} \prod_{K \in \{1, 2, \dots, l\} \setminus \{k\}} x_K^{\lambda_{jK}}. \quad (3)$$

Как видно, $\frac{\partial y}{\partial x_k}$ уже не является постоянной величиной, как в случае линейной

модели, а представляет собой линейную комбинацию произведений регрессоров. При этом производная (3) не зависит от переменной x_k : если $\lambda_{jk} = 0$, то в сумме (3) множитель $\lambda_{jk} x_k^{\lambda_{jk}-1} = 0$, а если $\lambda_{jk} = 1$, то $\lambda_{jk} x_k^{\lambda_{jk}-1} = 1$. Поэтому справедлива следующая интерпретация ЛМР. Если зафиксировать значения независимых переменных $x_K = x_K^*$, $K \in \{1, 2, \dots, l\} \setminus \{k\}$, то:

- если $\frac{\partial y}{\partial x_k} \Big|_{x_K^*} = 0$, то изменение переменной x_k никак не влияет на зависимую переменную y ;
- если $\frac{\partial y}{\partial x_k} \Big|_{x_K^*} > 0$, то рост (падение) переменной x_k на 1 единицу увеличивает (уменьшает) значение зависимой переменной на величину $\frac{\partial y}{\partial x_k} \Big|_{x_K^*}$;

- если $\frac{\partial y}{\partial x_k} \Big|_{x_k^*} < 0$, то рост (падение) переменной x_k на 1 единицу уменьшает (увеличивает) значение зависимой переменной на величину $\frac{\partial y}{\partial x_k} \Big|_{x_k^*}$.

Таким образом, варьируя в ЛМР значения независимых переменных x_K , $K \in \{1, 2, \dots, l\} \setminus \{k\}$, можно контролировать степень и знак влияния одной конкретной переменной x_k на зависимую переменную y , что невозможно в линейной регрессионной модели (1). При этом в ЛМР, используя выборку, можно проследивать и динамику скорости изменения функции y относительно изменения аргумента x_k :

$\frac{\partial y}{\partial x_k} \Big|_{x_{ik}}$, $i = \overline{1, n}$. Еще одной характеристикой ЛМР является средняя по выборке скорость v_k изменения функции y относительно изменения аргумента x_k :

$$v_k = \frac{1}{n} \sum_{i=1}^n \frac{\partial y}{\partial x_k} \Big|_{x_{ik}}, \quad (4)$$

которая, в случае линейной регрессии (1), совпадает с оценкой параметра α_k .

Продemonстрируем предложенный способ интерпретации ЛМР на примере решения задачи моделирования перевозок крупнотоннажных контейнеров на железнодорожном пункте пропуска Забайкальск – Маньчжурия. Отметим, что подобная задача решалась в работах [3,4]. При этом построение математических моделей объема пропуска производилось на основании анализа статистических данных торгово-экономических показателей России и Китая за период с 1990 по 2015 гг.

В работе [3] с помощью метода наименьших квадратов была построена линейная регрессия:

$$y_1 = -28636,1 + 14,910x_1 + 7,468x_2 - 7,330x_3 + 33,738z_1 + 15,508z_3, \quad (5)$$

где y_1 – объем импорта крупнотоннажных контейнеров, ДФЭ (в 20-футовом эквиваленте); x_1 – валовой внутренний продукт (ВВП) РФ по паритету покупательной способности (ППС), млрд. долл.; x_2 – инвестиции в основной капитал РФ, млрд. руб.; x_3 – оборот розничной торговли в РФ, млрд. руб.; z_1 – объем импорта КНР, млрд. долл.; z_3 – грузооборот железнодорожного транспорта КНР, млрд. т-км.

В работе [4] с помощью метода наименьших модулей была построена ЛМР:

$$y_1 = 5383,68 + 23,8257x_2 - 9,313x_3 - 0,000857x_2x_3 + 2,003 \cdot 10^{-7}x_2x_3z_1 + 5,126 \cdot 10^{-7}x_1x_3z_3. \quad (6)$$

Займемся теперь вопросом содержательной интерпретации ЛМР (6). Частные производные функции (6) по переменным x_1 , x_2 , x_3 , z_1 , z_3 имеют вид:

$$\frac{\partial y_1}{\partial x_1} = 5,126 \cdot 10^{-7}x_3z_3, \quad (7)$$

$$\frac{\partial y_1}{\partial x_2} = 23,8257 - 0,000857x_3 + 2,003 \cdot 10^{-7}x_3z_1, \quad (8)$$

$$\frac{\partial y_1}{\partial x_3} = -9,313 - 0,000857 x_2 + 2,003 \cdot 10^{-7} x_2 z_1 + 5,126 \cdot 10^{-7} x_1 z_3, \quad (9)$$

$$\frac{\partial y_1}{\partial z_1} = 2,003 \cdot 10^{-7} x_2 x_3, \quad (10)$$

$$\frac{\partial y_1}{\partial z_3} = 5,126 \cdot 10^{-7} x_1 x_3. \quad (11)$$

Выражения (7) – (11) указывают на скорости изменения объема импорта крупнотоннажных контейнеров y_1 относительно изменения соответствующих аргументов x_1 , x_2 , x_3 , z_1 , z_3 . Как видно, все эти скорости не являются постоянными, как в случае линейной регрессии. Например, соотношение (7) показывает, что на скорость изменения объема импорта крупнотоннажных контейнеров y_1 относительно изменения ВВП РФ x_1 влияют две переменные – оборот розничной торговли РФ x_3 и грузооборот железнодорожного транспорта КНР z_3 . График зависимости скорости изменения y_1 относительно x_1 представлен на рисунке 1.

Рисунок 1. Зависимость скорости изменения y_1 относительно x_1

График на рисунке 1 отражает динамику скорости изменения y_1 относительно x_1 . Как видно, с 1990 по 1998 годы эта скорость примерно была равна нулю, т.е. изменение ВВП РФ x_1 никак не влияло на объемы импорта y_1 . Начиная же с 1999 г. и до 2014 г. включительно наблюдается значительный положительный рост скорости изменения y_1 относительно x_1 с низких значений и до 36,8 (за исключением некоторого спада в кризисном 2008 г.), т.е. усиливается влияние переменной x_1 на y_1 . Такое усиление, в силу формулы (7), объясняется ростом оборота розничной торговли РФ x_3 и грузооборота железнодорожного транспорта КНР z_3 . В 2015 г. эта скорость упала с 36,8 до 28,5, т.е. влияние переменной x_1 на y_1 несколько ослабло. При этом на всем временном интервале скорость изменения y_1 относительно x_1 всегда положительна, что говорит о том, что с ростом x_1 значения y_1 всегда возрастало, но разными темпами.

В соответствии с формулой (4) средняя по выборке скорость изменения y_1 относительно x_1 равна $v_1 = 10,477$. Эта величина отличается от оценки параметра $\alpha_1 = 14,910$ линейной регрессии (5), и это при том, что обе они характеризуют среднее изменение объема импорта крупнотоннажных контейнеров y_1 с изменением ВВП РФ x_1 на 1 единицу.

Аналогично интерпретируются соотношения (8) – (11).

Средние по выборке скорости изменения y_1 относительно переменных x_2 , x_3 , z_1 и z_3 равны соответственно $v_2 = 19,048$, $v_3 = -10,188$, $v_4 = 16,018$ и $v_5 = 10,147$. Имеющие аналогичный смысл оценки параметров линейной регрессии (5) равны $\alpha_2 = 7,468$, $\alpha_3 = -7,330$, $\alpha_4 = 33,738$ и $\alpha_5 = 15,508$.

Таким образом, предложенный способ интерпретации коэффициентов линейно-мультипликативной регрессии может быть использован на практике при решении конкретных прикладных задач.

1. Базилевский М.П., Носков С.И. Алгоритм построения линейно-мультипликативной регрессии // Современные технологии. Системный анализ. Моделирование. – Иркутск, 2011. – №1(29). – С.88-92.
2. Базилевский М.П., Носков С.И. Формализация задачи построения линейно-мультипликативной регрессии в виде задачи частично-булевого линейного программирования // Современные технологии. Системный анализ. Моделирование. – Иркутск, 2017. – №3(55). – С.101-105.
3. Банщикова А.А., Базилевский М.П., Тихомиров В.А. Прогнозирование объема пропуска перевозимых на нетяговом подвижном составе крупнотоннажных контейнеров в экспортно-импортном сообщении в направлении РФ – КНР // Современные технологии. Системный анализ. Моделирование. – Иркутск, 2017. – №2(54). – С.185-190.
4. Базилевский М.П., Банщикова А.А., Тихомиров В.А. Построение линейно-мультипликативных регрессионных моделей объема пропуска крупнотоннажных контейнеров на железнодорожном пункте пропуска Забайкальск-Маньчжурия в направлении Россия – Китай // Современное развитие науки и техники сборник научных трудов Всероссийской национальной научно-практической конференции. – Ростов-на-Дону, 2017. – С. 90-94.