

Научная статья | Original paper

УДК 519.6

Идентификация параметров линейной регрессии и математического правила для формирования значений входящей в модель фиктивной переменной

М.П. Базилевский

Иркутский государственный университет путей сообщения,
Иркутск, Российская Федерация

✉ mik2178@yandex.ru

Резюме

Контекст и актуальность. Задача построения регрессионных моделей в ситуации, когда объясняющие переменные носят категориальный характер, актуальна в настоящее время. Категории могут отражать, например, пол человека (мужчина и женщина), уровень высшего образования (бакалавриат, специалитет и магистратура) или времена года (зима, весна, лето и осень). Для внедрения дихотомических категорий в модель используются фиктивные переменные, принимающие при наличии какого-либо признака значение 1, а при его отсутствии — 0. Причём, значения фиктивных переменных зачастую формируют по определенным заранее правилам. **Цель.** Цель работы состоит в исследовании задачи одновременной идентификации параметров модели множественной линейной регрессии и значений входящей в неё фиктивной переменной. **Гипотеза.** Гипотеза состоит в том, что сформулированная задача может быть формализована в терминах аппарата математического программирования. **Методы и материалы.** Для оценки неизвестных параметров линейной регрессии применён метод наименьших модулей. **Результаты.** Задача идентификации указанных характеристик формализована в виде задачи частично булевого линейного программирования. Её решение даёт не просто оценки линейной регрессии и оптимальные значения фиктивной переменной, но и правило, по которому эти значения идентифицируются. Правило основано на преобразовании линейной комбинации объясняющих переменных с помощью целочисленной функции «пол». По известным данным о работе выпарного аппарата на большом промышленном предприятии построено 4 разновидности линейной регрессии с фиктивной переменной. **Выводы.** Поскольку в процессе



построения линейных регрессий нашим способом идентифицируется и правило формирования фиктивной переменной, то это позволяет использовать их для решения задач прогнозирования. Идентифицировав правило, при необходимости можно переоценивать полученные линейные регрессии, например, с помощью метода наименьших квадратов, имеющего более ёмкую совокупность критериев верификации моделей.

Ключевые слова: регрессионный анализ, фиктивная переменная, целочисленная функция «пол», метод наименьших модулей, задача частично булевого линейного программирования

Для цитирования: Базилевский, М.П. (2025). Идентификация параметров линейной регрессии и математического правила для формирования значений входящей в модель фиктивной переменной. *Моделирование и анализ данных*, 15(2), 139—151. <https://doi.org/10.17759/mda.2025150208>

Identification of linear regression parameters and a mathematical rule for generating values of a dummy variable included in the model

M.P. Bazilevskiy

Irkutsk State Transport University, Irkutsk, Russian Federation

✉ mik2178@yandex.ru

Abstract

Context and relevance. The task of constructing regression models in a situation where explanatory variables are categorical is currently relevant. Categories can reflect, for example, gender (male and female), level of higher education (bachelor's degree, specialist degree, and master's degrees), or seasons (winter, spring, summer, and autumn). To introduce dichotomous categories into the model, dummy variables are used. These take the value 1 if a certain feature is present, and 0 if it is absent. The values of these dummy variables are often determined by predefined rules. **Objective.** The aim of the work is to formulate a problem of simultaneous identification of parameters of multiple linear regression models and values of dummy variables included in them. **Hypothesis.** The hypothesis is that the formulated problem can be formalized in terms of the apparatus of mathematical programming. **Methods and materials.** The least absolute deviations method is used to estimate the unknown regression parameters. **Results.** The task of identifying these features is formalized as a mixed 0—1 integer linear programming problem. Its solution gives not only estimates of the linear regression and optimal values of the dummy variable, but also a rule by which these values are identified. The rule is based on the transformation of a linear combination of explanatory variables using the integer function floor. Based on known data on the operation of an evaporator at a large industrial enterprise,



4 types of linear regressions with a dummy variable were constructed. **Conclusions.** Since the process of constructing linear regressions using our method also identifies a rule for forming a dummy variable, it allows them to be used for solving forecasting problems. After identifying the rule, it is possible, if necessary, to re-estimate the obtained linear regressions, for example, using the ordinary least squares method, which has a more extensive set of model validation criteria.

Keywords: regression analysis, dummy variable, integer function floor, least absolute deviations method, mixed 0—1 integer linear programming problem

For citation: Bazilevskiy, M.P. (2025). Identification of linear regression parameters and a mathematical rule for generating values of a dummy variable included in the model. *Modelling and Data Analysis*, 15(2), 139—151. (In Russ.). <https://doi.org/10.17759/mda.2025150208>

Введение

В настоящее время регрессионный анализ активно используется для обработки статистических данных в различных сферах человеческой деятельности (см., например, (Khan et al., 2024; Wang et al., 2024)). Для оценки неизвестных параметров моделей множественной линейной регрессии существует широкий арсенал методов. В данной работе используется метод наименьших модулей (МНМ), состоящий в минимизации суммы модулей отклонений фактических от расчетных значений выходной переменной. Ещё в 1959 году задача поиска минимума такой суммы в работе (Wagner, 1959) была сведена к задаче линейного программирования. Благодаря этому, возникло целое научное направление, связанное с использованием в регрессионном анализе аппарата математического программирования. Важные результаты в этом направлении были получены в работах (Носков, 1996; Konno, Yamamoto, 2009), в которых для МНМ задача отбора наиболее существенных регрессоров в линейной регрессии описана в терминах аппарата частично булевого линейного программирования (ЧБЛП). А повышение эффективности методов и аппаратных средств решения задач ЧБЛП за последние годы, как следует из (Koch et al., 2022), привело к появлению новых подходов к регрессионному моделированию. Так, в (Базилевский, Носков, 2017) сформулирована задача ЧБЛП для отбора наиболее существенных регрессоров в линейной регрессии с контролем величины критерия «согласованности поведения». В (Базилевский, Ойдопова, 2023) рассмотрена задача ЧБЛП для построения с помощью МНМ модульных регрессионных моделей. Причём, аппарат ЧБЛП стал применяться не только при использовании МНМ, но и метода наименьших квадратов (Базилевский, 2024a).

На написание данной статьи сподвигла работа (Базилевский, 2024б), посвященная оценке с помощью МНМ параметров нелинейных регрессий с целочисленными функциями «пол» и «потолок». Если ограничить значения этих целочисленных функций диапазоном от 0 до 1, то можно получить правило формирования значений



фиктивной переменной. Фиктивные переменные (dummy-переменные) (Елисеева, Курышева, 2010), которым присваивают значения 0 и 1, довольно часто применяются в процессе построения регрессионных моделей. Например, при построении зависимости заработной платы работника от его стажа, введение в модель фиктивной переменной «пол» позволяет определить среднюю разницу в зарплатах мужчин и женщин. Традиционно dummy-переменные используются при моделировании сезонности во временных рядах. Например, в (Крылова, Ефимова, 2019) с помощью фиктивных переменных измеряется сезонность числа зарегистрированных преступлений в Республике Саха, в (Wang, He, Zhao, 2021) осуществляется прогнозирование сезонных потреблений природного газа в США, в (Zhou et al., 2021) исследуется квартальное потребление электроэнергии, ежемесячная пропускная способность грузов, ежемесячная добыча сырой нефти и ежемесячное производство бензина в Китае. С помощью dummy-переменных решаются и другие задачи анализа данных. Так, в (Кривошеева, Подкур, 2021) строится зависимость цены квартиры в городе Кемерово от районного расположения, в (Горидько, 2016) исследуется связь между индексом потребительских цен Камчатского края, безработицей и фиктивной переменной, отражающей максимальные отклонения отдельных наблюдений от полученного тренда, в (Шумилина, Цвиль, 2020) с помощью dummy-переменных улучшено качество нелинейного тренда в местах его сильных колебаний. В (Cui et al., 2024) обсуждается возможность применения фиктивных переменных для изучения структурных изменений в экономике Китая, в (Singh et al., 2023) исследуется успеваемость выпускников сельскохозяйственных вузов в Индии в зависимости от пола и касты, а в (Sahoo, 2021) строится модель влияния COVID-19 и дня недели на фондовый рынок в Индии.

Проанализированные работы объединяет то, что фиктивным переменным присваивают значения 0 и 1 по определенным заранее правилам. Цель настоящей статьи состоит в формализации задачи одновременной идентификации и неизвестных параметров линейной регрессии с помощью МНМ, и определенного правила, выраженного с использованием целочисленной функции «пол», по которому формируются значения входящей в модель фиктивной переменной.

Материалы и методы

Пусть исследуется влияние l объясняющих (входных) переменных $x_1, x_2, \dots, x_l$ на объясняемую (выходную) переменную y по выборочной совокупности объема n . Предположим, что все эти переменные принимают значения $y_i, x_{i1}, x_{i2}, \dots, x_{il}$, $i = \overline{1, n}$. Рассмотрим модель множественной линейной регрессии с неизвестными параметрами $\alpha_0, \alpha_1, \dots, \alpha_l$ и ошибками ε_i , $i = \overline{1, n}$:

$$y_i = \sum_{j=0}^l \alpha_j x_{i,j} + \varepsilon_i, \quad i = \overline{1, n}, \quad (1)$$

где x_0 — объясняющая переменная, принимающая единичные значения.



Введем в линейную регрессию (1) фиктивную переменную:

$$y_i = \sum_{j=0}^l \alpha_j x_{i,j} + \lambda \cdot d_i \cdot x_{i,s} + \varepsilon_i, \quad i = \overline{1, n}, \quad (2)$$

где s — назначенный исследователем номер объясняющей переменной из множества $\{0, 1, \dots, l-1, l\}$; d — фиктивная переменная с неизвестными булевыми значениями d_i , $i = \overline{1, n}$; λ — неизвестный параметр. Поскольку значения dummy-переменной неизвестны, то модель (2) относится к нелинейным по параметрам регрессиям.

Пусть, например, $l = 3$. Тогда при $s = 0, 1, 2$ и 3 совокупность возможных регрессионных моделей (2) будет иметь вид

$$y_i = \alpha_0 + \alpha_1 x_{i,1} + \alpha_2 x_{i,2} + \alpha_3 x_{i,3} + \lambda \cdot d_i + \varepsilon_i, \quad i = \overline{1, n}, \quad (3)$$

$$y_i = \alpha_0 + \alpha_1 x_{i,1} + \alpha_2 x_{i,2} + \alpha_3 x_{i,3} + \lambda \cdot d_i \cdot x_{i,1} + \varepsilon_i, \quad i = \overline{1, n}, \quad (4)$$

$$y_i = \alpha_0 + \alpha_1 x_{i,1} + \alpha_2 x_{i,2} + \alpha_3 x_{i,3} + \lambda \cdot d_i \cdot x_{i,2} + \varepsilon_i, \quad i = \overline{1, n}, \quad (5)$$

$$y_i = \alpha_0 + \alpha_1 x_{i,1} + \alpha_2 x_{i,2} + \alpha_3 x_{i,3} + \lambda \cdot d_i \cdot x_{i,3} + \varepsilon_i, \quad i = \overline{1, n}. \quad (6)$$

Как видно, если $d_i = 0$, то в регрессии (3) свободный член равен α_0 , в (4) коэффициент при переменной x_1 равен α_1 , в (5) коэффициент при переменной x_2 равен α_2 , в (6) коэффициент при переменной x_3 равен α_3 . Если $d_i = 1$, то указанные коэффициенты в этих моделях переключаются на $(\alpha_0 + \lambda)$, $(\alpha_1 + \lambda)$, $(\alpha_2 + \lambda)$ и $(\alpha_3 + \lambda)$ соответственно.

Будем считать, что значения фиктивной переменной d в регрессии (2) связаны со значениями объясняющих переменных правилом:

$$d_i = \left\lfloor \beta_0 + \sum_{j=1}^l \beta_j x_{i,j} \right\rfloor, \quad i = \overline{1, n}, \quad (7)$$

где $\beta_0, \beta_1, \dots, \beta_l$ — неизвестные параметры; $[x]$ — целочисленная функция «пол» (Грэхем, Кнут, Паташник, 1998), округляющая число x до ближайшего целого в меньшую сторону. Например, $[6,95] = 6$.

Поскольку фиктивная переменная d должна принимать только булевы значения, то справедливы будут следующие ограничения:

$$d_i \in \{0, 1\}, \quad i = \overline{1, n}. \quad (8)$$

Используя (7), представим регрессию (2) в виде:

$$y_i = \sum_{j=0}^l \alpha_j x_{i,j} + \lambda \cdot \left\lfloor \beta_0 + \sum_{j=1}^l \beta_j x_{i,j} \right\rfloor \cdot x_{i,s} + \varepsilon_i, \quad i = \overline{1, n}. \quad (9)$$

Сформулируем следующую задачу: для выбранного номера s с помощью МНМ требуется оценить неизвестные параметры α_j , β_j , $j = \overline{0, l}$, λ и $d_i \in \{0, 1\}$, $i = \overline{1, n}$, нелинейной регрессии (9). Покажем далее, что эта задача сводится к задаче ЧБЛП.



Так, как это сделано в (Базилевский, 2024б), заменим целочисленную функцию

$$\left[\beta_0 + \sum_{j=1}^l \beta_j x_{i,j} \right] \text{ линейными ограничениями}$$

$$d_i \leq \beta_0 + \sum_{j=1}^l \beta_j x_{ij} \leq d_i + 1 - \Delta, \quad i = \overline{1, n}, \quad (10)$$

где Δ — близкое к нулю положительное число. Из (10) следует, что при $d_i = 1$ линейная комбинация $\beta_0 + \sum_{j=1}^l \beta_j x_{ij} \in [1, 2 - \Delta]$, а при $d_i = 0$ — $\beta_0 + \sum_{j=1}^l \beta_j x_{ij} \in [0, 1 - \Delta]$.

Как того требует МНМ, введём неотрицательные переменные u_i и v_i , $i = \overline{1, n}$, удовлетворяющие условиям:

$$u_i = \begin{cases} y_i - \sum_{j=0}^l \alpha_j x_{i,j} - \lambda \cdot d_i \cdot x_{i,s}, & \text{если } y_i - \sum_{j=0}^l \alpha_j x_{i,j} - \lambda \cdot d_i \cdot x_{i,s} > 0, \\ 0, & \text{в противном случае,} \end{cases}$$

$$v_i = \begin{cases} -y_i + \sum_{j=0}^l \alpha_j x_{i,j} + \lambda \cdot d_i \cdot x_{i,s}, & \text{если } y_i - \sum_{j=0}^l \alpha_j x_{i,j} - \lambda \cdot d_i \cdot x_{i,s} < 0, \\ 0, & \text{в противном случае.} \end{cases}$$

Подчеркнём ограничения неотрицательности введенных переменных:

$$u_i \geq 0, \quad v_i \geq 0, \quad i = \overline{1, n}. \quad (11)$$

Учитывая введенные неотрицательные переменные, перепишем регрессию (2) в следующем виде:

$$y_i = \sum_{j=0}^l \alpha_j x_{i,j} + \lambda \cdot d_i \cdot x_{i,s} + u_i - v_i, \quad i = \overline{1, n}. \quad (12)$$

Как известно (Wagner, 1959), для получения МНМ-оценок необходимо минимизировать следующую функцию:

$$\sum_{i=1}^n (u_i + v_i) \rightarrow \min. \quad (13)$$

Перейдём к линеаризации ограничений (12). Для этого введём вспомогательную переменную w , значения которой находятся по формулам $w_i = \lambda \cdot d_i$, $i = \overline{1, n}$. Это означает, что

$$w_i = \begin{cases} \lambda, & \text{если } d_i = 1, \\ 0, & \text{если } d_i = 0. \end{cases}$$



Тогда нелинейные ограничения (12) можно представить в виде уже линейных ограничений

$$y_i = \sum_{j=0}^l \alpha_j x_{i,j} + w_i \cdot x_{i,s} + u_i - v_i, \quad i = \overline{1, n}. \quad (14)$$

Из вышеизложенного следует, что значения переменной w должны удовлетворять неравенствам:

$$-M \cdot (1 - d_i) \leq w_i - \lambda \leq M \cdot (1 - d_i), \quad i = \overline{1, n}, \quad (15)$$

$$-M \cdot d_i \leq w_i \leq M \cdot d_i, \quad i = \overline{1, n}, \quad (16)$$

где M — заранее выбранное большое положительное число.

Очевидно, что если $d_i = 1$, то ограничения (15), (16) преобразуются к виду

$$w_i = \lambda,$$

$$-M \leq w_i \leq M;$$

а если $d_i = 0$, то к виду

$$-M \leq -\lambda \leq M,$$

$$w_i = 0.$$

Таким образом, решение задачи ЧБЛП с целевой функцией (13), с линейными ограничениями (8), (10), (11), (14) — (16) приводит для выбранного номера s к идентификации с помощью МНМ как параметров линейной регрессии, так и правила (7), с помощью которого формируются значения входящей в модель фиктивной переменной d .

Результаты

Для демонстрации предложенного нами способа использовалась известная выборка данных (Дрейпер, Смит, 1986; Дрейпер, Смит, 1987), включающая наблюдения за работой выпарного аппарата на большом промышленном предприятии. Объем выборочной совокупности равен 25. Были выбраны следующие переменные:

y — количество используемого пара в фунтах ежемесячно;

x_1 — количество активной жирной кислоты в фунтах, накопленное за месяц;

x_2 — число рабочих дней в месяце;

x_3 — средняя температура воздуха (°F).

Значения этих переменных представлены во 2-м, 3-м, 4-м и 5-м столбцах табл.

Сначала по этим данным с помощью МНМ была построена модель множественной линейной регрессии без фиктивной переменной:

$$\tilde{y} = 8,8543 + 0,5001x_1 + 0,1013x_2 - 0,07785x_3. \quad (17)$$



Таблица / Table

Исходные данные и идентифицированные значения фиктивных переменных
Initial data and identified values of dummy variables

№	y	x ₁	x ₂	x ₃	d ⁽⁰⁾	d ⁽¹⁾	d ⁽²⁾	d ⁽³⁾
1	10,98	5,2	20	35,3	1	1	1	1
2	11,13	5,12	20	29,7	1	1	1	1
3	12,51	6,19	23	30,8	1	1	1	1
4	8,4	3,89	20	58,8	1	1	1	0
5	9,27	6,28	21	61,4	0	0	0	0
6	8,73	5,76	22	71,3	1	1	1	0
7	6,36	3,45	11	74,4	0	0	0	0
8	8,5	6,57	23	76,7	1	1	1	0
9	7,82	5,69	21	70,7	0	0	0	0
10	9,14	6,14	20	57,5	0	0	0	0
11	8,24	4,84	20	46,4	0	0	0	0
12	12,19	4,88	21	28,9	1	1	1	1
13	11,88	6,03	21	28,1	1	1	1	1
14	9,57	4,55	19	39,1	0	0	0	0
15	10,94	5,71	23	46,8	1	1	1	1
16	9,58	5,67	20	48,5	0	0	0	0
17	10,09	6,72	22	59,3	1	1	1	0
18	8,11	4,95	22	70	1	1	1	0
19	6,83	4,62	11	70	0	0	0	0
20	8,88	6,6	23	74,5	1	1	1	0
21	7,68	5,01	20	72,1	0	0	0	0
22	8,47	5,68	21	58,1	0	0	0	0
23	8,86	5,28	20	44,6	0	0	0	0
24	10,36	5,36	20	33,4	0	0	0	0
25	11,08	5,87	22	28,6	1	1	1	1

Для (17) сумма модулей остатков *SAE* составила 10,475. Под остатком понимается разность между фактическим и расчетным по модели значением выходной переменной *y*.

Затем строились модели (3), (4), (5) и (6) с фиктивными переменными. Для этого в каждом случае решалась задача ЧБЛП (13), (8), (10), (11), (14) — (16) с параметрами $M=100$, $\Delta=0,00001$. Для решения использовался пакет LPSolve.

При $s = 0$ была получена линейная регрессия

$$\tilde{y} = 9,888 + 0,625x_1 - 0,02112x_2 - 0,07328x_3 + 0,8119d^{(0)}, \quad (18)$$

в которой фиктивная переменная $d^{(0)}$ находится по правилу

$$d^{(0)} = \left\lfloor 0,996 - 0,0001848x_1 + 0,0002642x_2 - 1,03068 \cdot 10^{-5}x_3 \right\rfloor;$$



при $s = 1$ регрессия

$$\tilde{y} = 10,0135 + 0,4397x_1 + 0,02065x_2 - 0,07255x_3 + 0,1359d^{(1)}x_1, \quad (19)$$

где

$$d^{(1)} = \lfloor -0,03611x_1 + 0,06316x_2 - 0,002089x_3 \rfloor;$$

при $s = 2$ модель

$$\tilde{y} = 10,1192 + 0,5922x_1 - 0,01472x_2 - 0,07581x_3 + 0,03645d^{(2)}x_2, \quad (20)$$

где

$$d^{(2)} = \lfloor -0,03611x_1 + 0,06316x_2 - 0,002089x_3 \rfloor;$$

при $s = 3$ зависимость

$$\tilde{y} = 8,0794 + 0,4372x_1 + 0,1032x_2 - 0,06293x_3 + 0,02219d^{(3)}x_3, \quad (21)$$

где

$$d^{(3)} = \lfloor -0,11802x_1 + 0,093036x_2 - 0,00683x_3 \rfloor.$$

Значения идентифицированных фиктивных переменных $d^{(0)}$, $d^{(1)}$, $d^{(2)}$ и $d^{(3)}$ представлены в 6-м, 7-м, 8-м и 9-м столбцах таблицы соответственно. Как видно по табл., $d^{(0)} = d^{(1)} = d^{(2)} \neq d^{(3)}$. Значения критерия SAE для моделей (18) — (21) составили 7,098, 7,083, 7,063 и 7,554 соответственно. Это означает, что все четыре построенные модели с фиктивными переменными оказались лучше по качеству аппроксимации линейной регрессии (17) без фиктивных переменных. Лучшей из моделей (18) — (21) следует признать зависимость (20).

Обсуждение результатов

Подчеркнём следующие важные на наш взгляд моменты.

Во-первых, после идентификации dummy-переменной можно переоценить найденное уравнение линейной регрессии другими методами. Это можно сделать, например, с помощью метода наименьших квадратов, который имеет более ёмкую совокупность критериев верификации модели — коэффициент детерминации R^2 , критерий Фишера, Стюдента, Дарбина-Уотсона и пр. В результате проведенной переоценки значения R^2 для перестроенных моделей (18) — (21) составили 0,9387, 0,9357, 0,9393 и 0,9285. Для перестроенной линейной регрессии (17) $R^2=0,8795$.



Во-вторых, поскольку в процессе оценки идентифицируется правило формирования фиктивной переменной, то построенную модель можно использовать для прогнозирования. Причём, поскольку при решении задачи ЧБЛП выполняются условия (8) и (10), то фиктивная переменная для любого наблюдения выборки обязательно принимает либо значение 0, либо 1. Но для новых наблюдений это свойство может нарушаться. В таком случае достаточно округлять полученное по найденному правилу целое число до ближайшего числа из множества $\{0, 1\}$.

Заключение

В статье с использованием целочисленной функции «пол» сформулирована задача ЧБЛП, решение которой приводит к идентификации с помощью МНМ как параметров линейной регрессии, так и правила, с помощью которого формируются значения входящей в модель фиктивной переменной. Предложенный математический аппарат продемонстрирован при примере моделирования количество используемого пара при работе выпарного аппарата на большом промышленном предприятии. Показано, что благодаря идентификации правила формирования фиктивной переменной можно использовать построенную линейную регрессию для прогнозирования значений выходной переменной. Открытым пока остаётся вопрос с выбором в задаче ЧБЛП чисел M и Δ . Тем не менее, разработанный метод идентификации dummy-переменной может успешно применяться для улучшения аппроксимационных качеств линейных по параметрам регрессионных моделей.

Список источников / References

1. Базилевский, М.П. (2024). Интеграция ограничений на коэффициенты интеркорреляций в оптимизационную задачу и условия построения вполне интерпретируемых неэлементарных линейных регрессий. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*, (69), 31—41.
Bazilevskiy, M.P. (2024). Integration of constraints on intercorrelation coefficients into the optimization problem and conditions for constructing quite interpretable non-elementary linear regressions. *Tomsk State University Journal of Control and Computer Science*, (69), 31—41. (In Russ.).
2. Базилевский, М.П. (2024). Оценивание с помощью метода наименьших модулей регрессионных моделей с целочисленными функциями пол и потолок. *International Journal of Open Information Technologies*, 12(10), 56—61.
Bazilevskiy, M.P. (2024). Estimation using least absolute deviations method of regression models with integer floor and ceiling functions. *International Journal of Open Information Technologies*, 12(10), 56—61. (In Russ.).
3. Базилевский, М.П., Носков, С.И. (2017). Программный комплекс построения линейной регрессионной модели с учётом критерия согласованности поведения фактической и расчетной траекторий изменения значений объясняемой переменной. *Вестник Иркутского государственного технического университета*, 21(9), 37—44.



Bazilevskiy, M.P., Noskov, S.I. (2017). Program complex for linear regression model construction considering behavior consistency criterion of actual and calculated trajectories of explained variable value change. *Proceedings of Irkutsk State Technical University*, 21(9), 37–44. (In Russ.).

4. Базилевский, М.П., Ойдопова, А.Б. (2023). Оценивание модульных линейных регрессионных моделей с помощью метода наименьших модулей. *Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления*, (45), 130–146.
Bazilevskiy, M.P., Oydopova, A.B. (2023). Estimation of modular linear regression models using the least absolute deviations. *Bulletin of Perm National Research Polytechnic University. Electrotechnics, Informational Technologies, Control Systems*, (45), 130–146. (In Russ.).
5. Горидько, Н.П. (2016). Методика сценарного прогнозирования изменения экономических показателей региона с использованием фиктивных переменных. *Известия Волгоградского государственного технического университета*, (7), 12–17.
Gorid'ko, N.P. (2016). Methods of scenario forecasting for changes of regional economic indicators using dummy variables. *Izvestia VSTU*, (7), 12–17. (In Russ.).
6. Грэхем, Р., Кнут, Д., Паташник, О. (1998). *Конкретная математика. Основание информатики*. М.: Мир.
Grekhem, R., Knut, D., Patashnik, O. (1998). *Concrete Mathematics. A Foundation for Computer Science*. Moscow: Mir. (In Russ.).
7. Дрейпер, Н., Смит, Г. (1986). *Прикладной регрессионный анализ. Книга 1*. М.: Финансы и статистика.
Dreyper, N., Smit, G. (1986). *Applied Regression Analysis. Part 1*. Moscow: Finance & Statistics. (In Russ.).
8. Дрейпер, Н., Смит, Г. (1987). *Прикладной регрессионный анализ. Книга 2*. М.: Финансы и статистика.
Dreyper, N., Smit, G. (1987). *Applied Regression Analysis. Part 2*. Moscow: Finance & Statistics. (In Russ.).
9. Елисеева, И.И., Курышева, С.В. (2010). Фиктивные переменные в анализе данных. *Социология: методология, методы, математическое моделирование*, (30), 43–63.
Eliseeva, I.I., Kuryшева, S.V. (2010). Dummy variables in data analysis. *Sociology: Methodology, Methods, Mathematical Modeling*, (30), 43–63. (In Russ.).
10. Кривошеева, М.В., Подкур, П.Н. (2021). Исследование цен на однокомнатные квартиры с помощью фиктивных переменных. *Россия молодая*, 95020.1-95020.4.
Krivosheeva, M.V., Podkur, P.N. (2021). A Study of one-room apartment prices using dummy variables. *Young Russia*, 95020.1-95020.4. (In Russ.).
11. Крылова, Е.А., Ефимова, Н.Р. (2019). Использование фиктивных переменных при моделировании числа зарегистрированных преступлений по Республике Саха (Якутия). *Южно-Сибирский научный вестник*, (2), 111–116.
Krylova, E.A., Efimova, N.R. (2019). Application of indicator variables in modeling of number of recorded crimes in the Republic of Sakha (Yakutia). *South-Siberian Scientific Bulletin*, (2), 111–116. (In Russ.).
12. Носков, С.И. (1996). *Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных*. Иркутск: РИЦ ГП «Облформпечать».



- Noskov, S.I. (1996). *Technology for modeling objects with unstable functioning and uncertainty in data*. Irkutsk: RITZ GP Oblinformpechat'. (In Russ.).
13. Шумилина, В.Е., Цвиль, М.М. (2020). Построение модели регрессии по временным рядам с целью прогнозирования индекса производительности труда в Российской Федерации. *Вестник евразийской науки*, 12(1), 73.
- Shumilina, V.E., Tsvil', M.M. (2020). Building a time series regression model with the aim of predicting the labor productivity index in the Russian Federation. *The Eurasian Scientific Journal*, 12(1), 73. (In Russ.).
14. Cui, J., Zhang, J., Liu, C., Liu, Y., Guo, S. (2024). Dummy variable threshold effect model and its economic applications. *Journal of Nonlinear and Convex Analysis*, 25(4), 779–788.
15. Khan, F., Muhammadullah, S., Sharif, A., Lee, C.C. (2024). The role of green energy stock market in forecasting China's crude oil market: An application of IIS approach and sparse regression models. *Energy Economics*, 130, 107269. <https://doi.org/10.1016/j.eneco.2023.107269>
16. Koch, T., Berthold, T., Pedersen, J., Vanaret, C. (2022). Progress in mathematical programming solvers from 2001 to 2020. *EURO Journal on Computational Optimization*, 10, 100031. <https://doi.org/10.1016/j.ejco.2022.100031>
17. Konno, H., Yamamoto, R. (2009). Choosing the best set of variables in regression analysis using integer programming. *Journal of global optimization*, 44, 273–282. <https://doi.org/10.1007/s10898-008-9323-9>
18. Sahoo, M. (2021). COVID-19 impact on stock market: Evidence from the Indian stock market. *Journal of Public Affairs*, 21(4), e2621. <https://doi.org/10.1002/pa.2621>
19. Singh, H., Das, A., Dey, S., Narsimhaiah, L., Pandit, P., Sinha, K., Sahu, P.K., Mishra, P. (2023). A study on academic attainment of agriculture students and its correlates: a dummy regression approach. *Annals of Data Science*, 10(1), 129–152. <https://doi.org/10.1007/s40745-020-00275-z>
20. Wagner, H.M. (1959). Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, 54(285), 206–212. <https://doi.org/10.1080/01621459.1959.10501506>
21. Wang, S., Chen, Y., Cui, Z., Lin, L., Zong, Y. (2024). Diabetes risk analysis based on machine learning LASSO regression model. *Journal of Theory and Practice of Engineering Science*, 4(01), 58–64. [https://doi.org/10.53469/jtpes.2024.04\(01\).08](https://doi.org/10.53469/jtpes.2024.04(01).08)
22. Wang, Z.X., He, L.Y., Zhao, Y.F. (2021). Forecasting the seasonal natural gas consumption in the US using a gray model with dummy variables. *Applied Soft Computing*, 113, 108002. <https://doi.org/10.1016/j.asoc.2021.108002>
23. Zhou, W., Cheng, Y., Ding, S., Chen, L., Li, R. (2021). A grey seasonal least square support vector regression model for time series forecasting. *ISA transactions*, 114, 82–98. <https://doi.org/10.1016/j.isatra.2020.12.024>

Информация об авторе

Михаил Павлович Базилевский, кандидат технических наук, доцент кафедры математики, факультет «Управление на транспорте и информационные технологии», Иркутский государственный университет путей сообщения (ФГБОУ ВО ИрГУПС), Иркутск, Российская Федерация, ORCID: <https://orcid.org/0000-0002-3253-5697>, e-mail: mik2178@yandex.ru



Information about the author

Mikhail P. Bazilevskiy, Candidate of Science (Technical), Associate Professor, Department of Mathematics, Irkutsk State Transport University, Irkutsk, Russian Federation, ORCID: <https://orcid.org/0000-0002-3253-5697>, e-mail: mik2178@yandex.ru

Вклад автора

Базилевский М.П. — идеи исследования; аннотирование, написание и оформление рукописи; планирование исследования; контроль за проведением исследования; применение статистических, математических или других методов для анализа данных; проведение эксперимента; сбор и анализ данных; визуализация результатов исследования.

Contribution of the Author

Mikhail P. Bazilevskiy — ideas; annotation, writing and design of the manuscript; planning of the research; control over the research; application of statistical, mathematical or other methods for data analysis; conducting the experiment; data collection and analysis; visualization of research results.

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Conflict of Interest

The author declare no conflict of interest.

Поступила в редакцию 10.04.2025

Поступила после рецензирования 15.04.2025

Принята к публикации 15.05.2025

Опубликована 30.06.2025

Received 2025.04.10

Revised 2025.04.15

Accepted 2025.05.15

Published 2025.06.30