Научная статья
УДК 519.862.6
EDN XFYMSJ
DOI 10.17150/2713-1734.2023.5(2).97-114

М.П. Базилевский
Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация
А.В. Караулова
Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация

Алгоритмы и программное обеспечение численного оценивания позиномиальных регрессионных моделей с помощью метода наименьших квадратов

Аннотация. Статья посвящена построению позиномиальных регрессионных моделей с несколькими мономами для каждой входной переменной. Такие модели обобщают известные полиномиальные регрессии. Исследован позином с одной переменной и двумя мономами. Предложен алгоритм «А» приближенного оценивания позиномиальных регрессий с помощью метода наименьших квадратов. Главным недостатком алгоритма «А» является то, что из-за межмономной мультиколлинеарности оцененная позиномиальная регрессия может оказаться не согласованной с содержательным смыслом входящих в нее факторов. Поэтому был разработан алгоритм «Б», гарантирующий не только сохранность содержательного смысла факторов, но и значимость по t-критерию Стьюдента всех коэффициентов позиномиальной регрессии. Алгоритм «Б» был реализован в виде программного комплекса ПОЗИТОН-1. Продемонстрирована работа программного комплекса на примере решения конкретной прикладной задачи.

Ключевые слова. Позином, нелинейность, математическая модель, позиномиальная регрессия, метод наименьших квадратов, мультиколлинеарность, алгоритм, программный комплекс.

Информация о статье. Дата поступления: 3 марта 2023 г.; дата принятия к публикации: 14 апреля 2023 г.; дата онлайн-размещения: 14 мая 2023 г.

Original article

M. P. Bazilevskiy
Irkutsk State Transport University, Irkutsk, Russian Federation
A.V. Karaulova
Irkutsk State Transport University, Irkutsk, Russian Federation

Algorithms and Software for Numerical Estimation of Posinomial Regression Models Using the Least Squares Method

Abstract. The article is devoted to the construction of posinomial regression models with several monomes for each input variable. Such models generalize wellknown polynomial regressions. Investigated by a posin with one variable and two monomes. The algorithm «A» of approximate estimation of posinomial regressions using the least squares method is proposed. The main disadvantage of the algorithm «A» is that due to the inter-monomic multicollinearity, the estimated posinomial regression may not be consistent with the meaningful meaning of the factors included in it. Therefore, algorithm «B» was developed, which guarantees not only the preservation of the meaningful meaning of the factors, but also the significance of all posinomial regression coefficients according to the Student's t-criterion. Algorithm «B» was implemented in the form of the POSITON-1 software package. The work of the software package is demonstrated by the example of solving a specific applied problem.

Keywords. Posinom, nonlinearity, mathematical model, posinomial regression, least squares method, multicollinearity, algorithm, software package.

Article info. Received 3 March, 2023; Accepted 14 April, 2023; Available online 14 May, 2023.

Введение

Машинное обучение [1–3] стало актуальным еще с самого начала развития искусственного интеллекта. Как отмечено в [1], «машинное обучение, наверное, самая горячая и быстроразвивающаяся дисциплина в современной информатике, если не в современной науке вообще». К одному из методов машинного обучения относится регрессионный анализ [4; 5]. Параметрическая регрессионная модель представляет собой зависимость выходной (объясняемой) переменной от одной или нескольких входных (объясняющих) переменных. Регрессии бывают линейные и нелинейные. Различных спецификаций нелинейных регрессионных моделей существует огромное количество. Среди них хотелось бы выделить степенные [6], кусочно-линейные [7], неэлементарные [8; 9], логистические [10], степенно-показательные [11], линейно-логарифмические [11] регрессии. Многие нелинейные спецификации, применяемые в экономике, изложены в [12].

Весьма эффективными на практике среди нелинейных спецификаций являются полиномиальные регрессии, которые находят широкое применение при решении различных прикладных задач. Так, в [13] полиномиальная регрессия была применена для моделирования расхода топлива для тяжелых транспортных средств, в [14] — для численного моделирования зарядного баланса легкового автомобиля, в [15] —для моделирования электромагнитных процессов при работе силовых трансформаторов под нагрузкой и в режиме холостого хода, в [16] — для моделирования оптимальных кредитных лимитов в микрофинансовых организациях, в [17] — для моделирования процесса спасения маломобильных групп населения при пожаре, в [18] — для прогнозирования мощности утечки на этапе планировки физического проектирования интегральных схем.

В [19] предложено обобщение полиномиальной регрессии — позиномиальная регрессия (ПозРМ) следующего вида:

$$y_i = \alpha_0 + \sum_{k=1}^{q} \alpha_k \prod_{j=1}^{m} x_{ij}^{b_{kj}} + \varepsilon_i, \quad i = \overline{1, n}, \tag{1}$$

где n — объем выборки; m — число объясняющих переменных; q — число мономов; y_i — i-е значение объясняемой переменной y; x_{ij} — i-е значение j-й объясняющей переменной; a_j, $j = \overline{0, q}$, b_{kj}, $k = \overline{1, q}$, $j = \overline{1, m}$,– неизвестные параметры; ε_i — i-я ошибка аппроксимации.

ПозРМ (1) представляют собой более гибкий инструмент математического моделирования, нежели полиномиальные регрессии.

ПозРМ (1) является нелинейной по параметрам, поэтому задачу ее оценивания методом наименьших квадратов (МНК) можно решить только численно, используя специальные методы. К сожалению, на сегодняшний день нет программных продуктов, предназначенных для нахождения МНК-оценок модели (1). Целью данной работы является разработка алгоритмов и программного обеспечения численного оценивания упрощенной формы ПозРМ (1) с помощью МНК.

1. Исследование позинома с одной переменной и двумя мономами

Рассмотрим позином с одной переменной и двумя мономами:

$$g(x) = a_1 x^{b_1} + a_2 x^{b_2}, \tag{2}$$

где переменная $x > 0$, коэффициенты a_1, a_2 и степени b_1, $b_2 (b_1 < b_2)$ — любые отличные от нуля действительные числа.

Первая производная функции (2) имеет вид:

$$g'(x) = a_1 b_1 x^{b_1 - 1} + a_2 b_2 x^{b_2 - 1}.$$

Тогда нетрудно определить, что функция (2) при x > 0 будет иметь только один экстремум в точке

$$x = \sqrt[b_2 - b_1]{\frac{-a_1 b_1}{a_2 b_2}}, \tag{3}$$

при условии $\frac{a_1 b_1}{a_2 b_2} < 0$.

Заметим, что частным случаем позинома (2) является полином второй степени $g(x) = a_1 x + a_2 x^2$, график которого представляет собой параболу, имеющую абсциссу вершины в точке $x_0 = -\frac{a_1}{2a_2}$. То же самое получается, если в выражении (3) взять $b_1 = 1$, $b_2 = 2$.

Вторая производная функции (2) имеет вид:

$$g''(x) = a_1 b_1 (b_1 - 1) x^{b_1 - 2} + a_2 b_2 (b_2 - 1)\, x^{b_2 - 2}.$$

Из этого следует, что функция (2) при $x > 0$ будет иметь только один перегиб в точке

$$x = \sqrt[b_2 - b_1]{\frac{-a_1 b_1 (b_1 - 1)}{a_2 b_2 (b_2 - 1)}}, \tag{4}$$

при условии $\dfrac{a_1 b_1 (b_1 - 1)}{a_2 b_2 (b_2 - 1)} < 0.$

Обобщая исследование, сделаем вывод, что функция (2) при $x > 0$ может иметь либо экстремум в точке (3), либо перегиб в точке (4), либо и экстремум, и перегиб, либо быть монотонной.

2. Алгоритмы численного МНК-оценивания ПозРМ

Рассмотрим частный случай ПозРМ (1) с двумя мономами для каждой входной переменной:

$$y_i = \alpha_0 + \sum_{j=1}^{m} \left(\alpha_{1j} x_{ij}^{b_{1j}} + \alpha_{2j} x_{ij}^{b_{2j}} \right) + \varepsilon_i, \; i = \overline{1, n}. \tag{5}$$

Если степени b_{1j}, b_{2j}, $j = \overline{1, m}$, известны, то модель (5) линейна по параметрам a_{1j}, a_{2j}, $j = \overline{1, m}$, поэтому для ее оценивания можно воспользоваться традиционным МНК. Если степени b_{1j}, b_{2j}, $j = \overline{1, m}$, неизвестны, то найти приближенные оценки ПозРМ (5) можно по следующему алгоритму «А».

1. Назначить для каждой входной переменной x_j нижнюю $b_j^{\text{нижн}}$ и верхнюю $b_j^{\text{верхн}}$ границу области возможных значений ее степеней. Выбор можно сделать либо произвольно, либо, например, принимая во внимание масштабы значений объясняющих переменных. В последнем случае $b_j^{\text{нижн}}$ и $b_j^{\text{верхн}}$ определяются из решения неравенств

$$\left(x_j^{\max} \right)^{b_j^{\text{верхн}}} \le \text{Big}, \; \left(x_j^{\max} \right)^{b_j^{\text{нижн}}} \ge \text{Small}, \tag{6}$$

где $x_j^{\max} = \max\left\{ x_{1j}, x_{2j}, ..., x_{nj} \right\}$, Big, Small — большое и малое положительные числа соответственно.

Если хотя бы одна из степеней ПозРМ (5) равна нулю, то из-за совершенной мультиколлинеарности невозможно найти МНК-оценки этой модели. При близких к нулю значениях будет возникать частичная мультиколлинеарность. Поэтому на данном этапе требуется исключить из каждого интервала $\left[b_j^{\text{нижн}}, b_j^{\text{верхн}} \right]$ промежутки [–Small, Small]. В результате для каждой переменной x_j будет выбрана область

$$[b_j^{\text{нижн}},\ -\text{Small}] \cup \left[\text{Small},\ b_j^{\text{верхн}}\right]. \tag{7}$$

2. Равномерно разбить каждый из объединенных интервалов области (7) p точками.

3. Оценив регрессии со всеми возможными комбинациями степеней, общее число которых составляет $\left(C_{2(p+2)}^2\right)^m$, выбрать модель с наименьшей величиной суммы квадратов остатков, либо с наибольшим значением коэффициента детерминации R^2. При необходимости в процессе перебора можно исключать значительно нелинейные зависимости, используя критерий нелинейности $NC_{\max}$ [19,20]. Для этого для каждой функции $\tilde{\alpha}_{1j}x_j^{\tilde{b}_{1j}} + \tilde{\alpha}_{2j}x_j^{\tilde{b}_{2j}}$ оцененной ПозРМ (5) сначала определяются точки экстремума и перегиба по формулам (3) и (4), попадающие на интервал $\left[x_j^{\min}, x_j^{\max}\right]$. Если ни одна точка не попала в интервал, то для j-й переменной строится вектор критериев нелинейности $VNC(x_j)$ [19] из одной компоненты, если попала одна точка — из двух компонент, если две точки — из трех компонент. Затем по векторам вычисляется критерий нелинейности ПозРМ в целом $NC_{\max}$ [19]:

$$NC_{\max} = \max\left\{\max\left(VNC\left(x_1\right)\right),\ \max\left(VNC\left(x_2\right)\right),\ \ldots,\ \max\left(VNC\left(x_m\right)\right)\right\}. \tag{8}$$

Критерий (8) принимает значения из промежутка [0,1). Если его значение равно 0, то все функции объясняющих переменных ПозРМ представляют собой либо прямые линии, либо ломаные, поэтому их просто интерпретировать. Чем ближе значение критерия $NC_{\max}$ к 1, тем выше степень нелинейности ПозРМ. Допустимо исключать модели, у которых $NC_{\max}$ превышает величину 0,1–0,3.

Алгоритм «А» может применяться для приближенного оценивания ПозРМ (5) с помощью МНК. Но ему свойственны следующие недостатки.

1. Высокая вычислительная сложность. Например, даже для восьми разбиений ($p = 8$) для оценки ПозРМ с тремя переменными ($m = 8$) потребуется осуществить перебор $\left(C_{20}^2\right)^3 = 6\ 859\ 000$ шестифакторных регрессий.

2. Оцененная ПозРМ может оказаться не согласованной с содержательным смыслом входящих в модель факторов. Причиной этому может служить сильная мультиколлинеарность, возникающая не только между переменными, но и между мономами одной переменной. Особенно сильно она будет выражена, когда степени мономов одной переменной будут значительно близки.

Поэтому для приближенного оценивания ПозРМ был предложен следующий алгоритм «Б».

ISSN 2713-1734

1. Сформировать все ПозРМ, включающие в себя только одну объясняющую переменную:

$$y_i = \theta_{01} + \theta_{11} x_1^{b_{11}} + \theta_{21} x_1^{b_{21}} + \varepsilon_{i1},\ i = \overline{1, n},$$
$$y_i = \theta_{02} + \theta_{12} x_2^{b_{12}} + \theta_{22} x_2^{b_{22}} + \varepsilon_{i2},\ i = \overline{1, n},$$
$$\ldots$$
$$y_i = \theta_{0m} + \theta_{1m} x_m^{b_{1m}} + \theta_{2m} x_m^{b_{2m}} + \varepsilon_{im},\ i = \overline{1, n}, \tag{9}$$

где $\theta_{0j},\ \theta_{1j},\ \theta_{2j},\ j = \overline{1, m}$ — неизвестные параметры.

2. Для каждой ПозРМ из набора (9) реализовать алгоритм «А», выбирая не одну лучшую модель, а несколько, например, k штук. В результате будет сформирован следующий список оцененных регрессий:

$$\tilde{y} = \tilde{\theta}_{01}^{(1)} + \tilde{\theta}_{11}^{(1)} x_1^{\tilde{b}_{11}^{(1)}} + \tilde{\theta}_{21}^{(1)} x_1^{\tilde{b}_{21}^{(1)}},\ \tilde{y} = \tilde{\theta}_{01}^{(2)} + \tilde{\theta}_{11}^{(2)} x_1^{\tilde{b}_{11}^{(2)}} + \tilde{\theta}_{21}^{(2)} x_1^{\tilde{b}_{21}^{(2)}},\ \ldots,$$
$$\tilde{y} = \tilde{\theta}_{01}^{(k)} + \tilde{\theta}_{11}^{(k)} x_1^{\tilde{b}_{11}^{(k)}} + \tilde{\theta}_{21}^{(k)} x_1^{\tilde{b}_{21}^{(k)}},$$
$$\tilde{y} = \tilde{\theta}_{01}^{(2)} + \tilde{\theta}_{11}^{(2)} x_1^{\tilde{b}_{11}^{(2)}} + \tilde{\theta}_{21}^{(2)} x_1^{\tilde{b}_{21}^{(2)}},\ \ldots,\ \tilde{y} = \tilde{\theta}_{01}^{(k)} + \tilde{\theta}_{11}^{(k)} x_1^{\tilde{b}_{11}^{(k)}} + \tilde{\theta}_{21}^{(k)} x_1^{\tilde{b}_{21}^{(k)}},$$
$$\tilde{y} = \tilde{\theta}_{02}^{(1)} + \tilde{\theta}_{12}^{(1)} x_2^{\tilde{b}_{12}^{(1)}} + \tilde{\theta}_{22}^{(1)} x_2^{\tilde{b}_{22}^{(1)}},\ \tilde{y} = \tilde{\theta}_{02}^{(2)} + \tilde{\theta}_{12}^{(2)} x_2^{\tilde{b}_{12}^{(2)}} + \tilde{\theta}_{22}^{(2)} x_2^{\tilde{b}_{22}^{(2)}},\ \ldots,$$
$$\tilde{y} = \tilde{\theta}_{02}^{(k)} + \tilde{\theta}_{12}^{(k)} x_2^{\tilde{b}_{12}^{(k)}} + \tilde{\theta}_{22}^{(k)} x_2^{\tilde{b}_{22}^{(k)}},$$
$$\ldots$$
$$\tilde{y} = \tilde{\theta}_{0m}^{(1)} + \tilde{\theta}_{1m}^{(1)} x_m^{\tilde{b}_{1m}^{(1)}} + \tilde{\theta}_{2m}^{(1)} x_m^{\tilde{b}_{2m}^{(1)}},\ \tilde{y} = \tilde{\theta}_{0m}^{(2)} + \tilde{\theta}_{1m}^{(2)} x_m^{\tilde{b}_{1m}^{(2)}} + \tilde{\theta}_{2m}^{(2)} x_m^{\tilde{b}_{2m}^{(2)}},\ \ldots,$$
$$\tilde{y} = \tilde{\theta}_{0m}^{(k)} + \tilde{\theta}_{1m}^{(k)} x_m^{\tilde{b}_{1m}^{(k)}} + \tilde{\theta}_{2m}^{(k)} x_m^{\tilde{b}_{2m}^{(k)}},$$

где $\tilde{\theta}_{0j}^{q},\ \tilde{\theta}_{1j}^{q},\ \tilde{\theta}_{1j}^{q}$ — МНК-оценки занявшей q-е место по величине коэффициента детерминации R^2 ПозРМ с j-й объясняющей переменной.

Удостовериться, что все полученные зависимости удовлетворяют содержательному смыслу решаемой задачи. Для этого достаточно просмотреть на плоскости графики всех полученных функций. Если окажется, что зависимость y от x_j противоречит смыслу решаемой задачи, то такие регрессии нужно исключить из списка.

На основе полученных регрессий введем новые переменные:

$$z_1^{(1)} = \tilde{\theta}_{11}^{(1)} x_1^{\tilde{b}_{11}^{(1)}} + \tilde{\theta}_{21}^{(1)} x_1^{\tilde{b}_{21}^{(1)}},\ z_1^{(2)} = \tilde{\theta}_{11}^{(2)} x_1^{\tilde{b}_{11}^{(2)}} + \tilde{\theta}_{21}^{(2)} x_1^{\tilde{b}_{21}^{(2)}},\ \ldots,$$
$$z_1^{(k)} = \tilde{\theta}_{11}^{(k)} x_1^{\tilde{b}_{11}^{(k)}} + \tilde{\theta}_{21}^{(k)} x_1^{\tilde{b}_{21}^{(k)}},$$
$$z_2^{(1)} = \tilde{\theta}_{12}^{(1)} x_2^{\tilde{b}_{12}^{(1)}} + \tilde{\theta}_{22}^{(1)} x_2^{\tilde{b}_{22}^{(1)}},\ z_2^{(2)} = \tilde{\theta}_{12}^{(2)} x_2^{\tilde{b}_{12}^{(2)}} + \tilde{\theta}_{22}^{(2)} x_2^{\tilde{b}_{22}^{(2)}},$$

$$z_2^{(k)} = \tilde{\theta}_{12}^{(k)} x_2^{\tilde{b}_{12}^{(k)}} + \tilde{\theta}_{22}^{(k)} x_2^{\tilde{b}_{22}^{(k)}},$$

$$\ldots$$

$$z_m^{(1)} = \tilde{\theta}_{1m}^{(1)} x_m^{\tilde{b}_{1m}^{(1)}} + \tilde{\theta}_{2m}^{(1)} x_m^{\tilde{b}_{2m}^{(1)}},\ z_m^{(2)} = \tilde{\theta}_{1m}^{(2)} x_m^{\tilde{b}_{1m}^{(2)}} + \tilde{\theta}_{2m}^{(2)} x_m^{\tilde{b}_{2m}^{(2)}},$$

$$z_m^{(k)} = \tilde{\theta}_{1m}^{(k)} x_m^{\tilde{b}_{1m}^{(k)}} + \tilde{\theta}_{2m}^{(k)} x_m^{\tilde{b}_{2m}^{(k)}}.$$

3. Выбрать оптимальный состав регрессоров с точки зрения минимизации суммы квадратов остатков (максимизации R^2) для регрессионной модели

$$y_i = \alpha_0 + \alpha_1 \begin{Bmatrix} z_{i1}^{(1)} \\ z_{i1}^{(2)} \\ \ldots \\ z_{i1}^{(k)} \end{Bmatrix} + \alpha_2 \begin{Bmatrix} z_{i2}^{(1)} \\ z_{i2}^{(2)} \\ \ldots \\ z_{i2}^{(k)} \end{Bmatrix} + \ldots + \alpha_m \begin{Bmatrix} z_{im}^{(1)} \\ z_{im}^{(2)} \\ \ldots \\ z_{im}^{(k)} \end{Bmatrix} + \varepsilon_i,\ i = \overline{1, n}.$$

Для этого нужно воспользоваться методом полного перебора моделей, общее количество которых равно k^m. Для того чтобы гарантировать получение ПозРМ, удовлетворяющей содержательному смыслу решаемой задачи, необходимо исключать в процессе перебора те модели, у которых хотя бы одна из оценок параметров $a_1, a_2, \ldots, a_m$ отрицательна. Для того чтобы гарантировать получение ПозРМ со всеми значимыми по t-критерию Стьюдента коэффициентами, необходимо исключать в процессе перебора те модели, у которых хотя бы один коэффициент при новых переменных не значим по t-критерию Стьюдента.

Перечислим достоинства алгоритма «Б».

1. Вычислительная сложность значительно ниже, чем у алгоритма «А». Например, при $p = 8$ и $m = 3$ на первом этапе потребуется оценить всего $m \cdot C_{2(p+2)}^2 = 3C_{20}^2 = 570$ двухфакторных моделей, а на втором при $k = 5$ — 125 трехфакторных моделей.

2. Если итоговая ПозРМ существует, то она гарантированно удовлетворяет содержательному смыслу факторов, все ее коэффициенты при переменных z значимы по t-критерию Стьюдента, и если исключены значительно нелинейные уравнения, то ее еще и можно интерпретировать.

Заметим, что ПозРМ, построенная по алгоритму «Б», может проигрывать модели, построенной по алгоритму «А», по прогностическим способностям.

3. Программный комплекс построения позиномиальных регрессионных моделей (ПК ПОЗИТОН-1)

Для построения ПозРМ (5) с помощью описанного выше алгоритма «Б» в среде программирования Delphi был разработан

ISSN 2713-1734

специализированный программный комплекс ПК ПОЗИТОН-1. Причем, ПК ПОЗИТОН-1 позволяет оценивать ПозРМ не только с двумя, но и с тремя мономами для каждой переменной. Однако в последнем случае пока недоступна функция контроля степени нелинейности модели.

Блок-схема алгоритма работы ПК ПОЗИТОН-1 представлена на рис. 1.

Для начала работы с программой пользователь должен выбрать исходные статистические данные, хранящиеся в текстовом файле с расширением *.txt. Данные должны располагаться в этом файле в виде таблицы. В первом столбце должны быть указаны

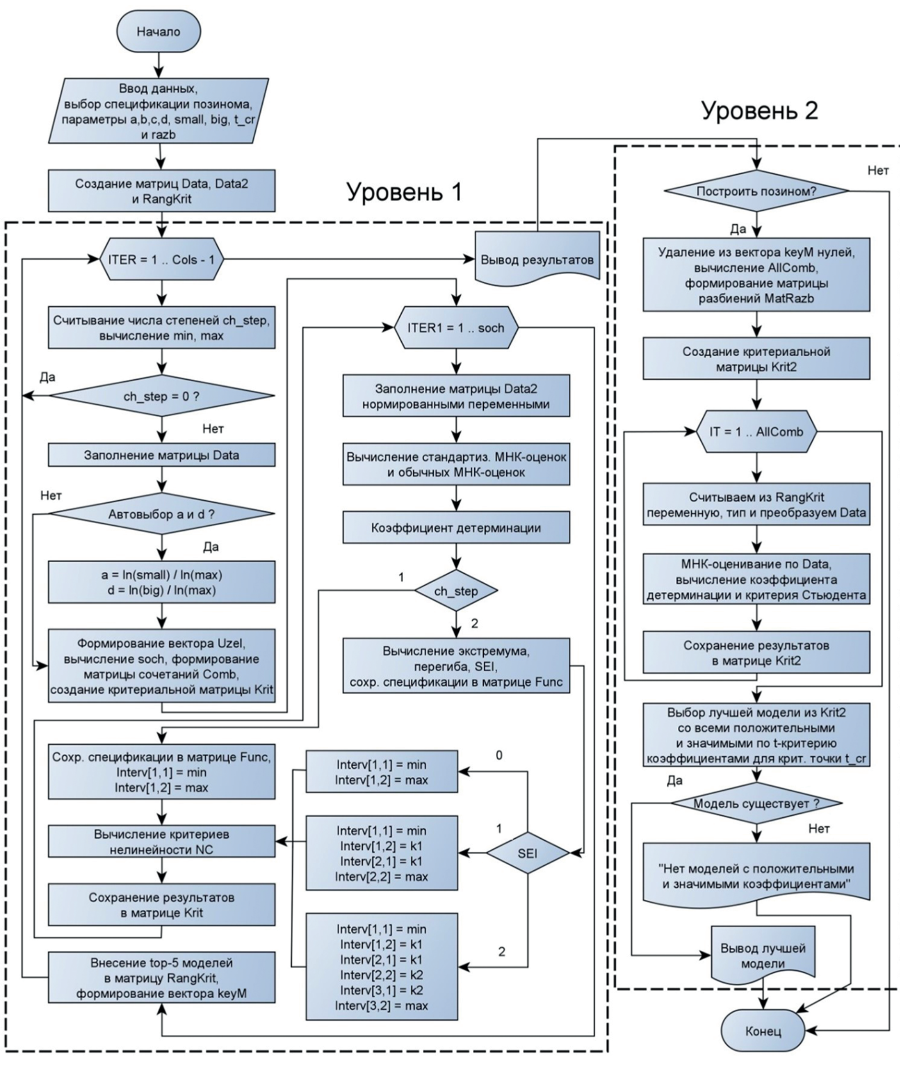

Рис. 1. Блок-схема алгоритма работы ПК ПОЗИТОН-1

значения объясняемой переменной y, во втором — первой объясняющей переменной x_1, в третьем — переменной x_2 и т.д. Столбцы отделяются друг от друга символом «Tab». В действительных числах целые части от дробных отделяются символом «,». Для выбора данных в главном меню ПК ПОЗИТОН-1 нужно воспользоваться командой Файл => Загрузка.

После выбора исходных данных следует определиться с настройками, воспользовавшись командой Настройки в главном меню программы. В открывшемся окне «Настройки» следует выбрать границы области (7), обозначенные в системе как [a, b] и [c, d]. Внутренние границы области b и c всегда задаются вручную и по умолчанию равны –0,25 и 0,25. Внешние границы a и d также можно менять и по умолчанию они равны –5 и 5. Однако для них предусмотрен и режим автоматического выбора. Для этого следует задать маленькое и большое положительные числа small и big. В результате решения неравенств (6) будут автоматически определены границы a и d.

В этом же окне нужно выбрать число точек razb, разбивающих каждый из промежутков [a, b] и [c, d]. Завершив ввод настроек, нужно нажать кнопку «Применить».

Затем нужно выбрать структуру ПозРМ. Это делается в специальной таблице на панели «Этап 1. Выбор переменных» (рис. 2).

Эта таблица состоит из двух столбцов — «Перем.» и «Число степ.». Первый столбец заполняется автоматически после загрузки исходных данных и содержит имена объясняющих переменных.

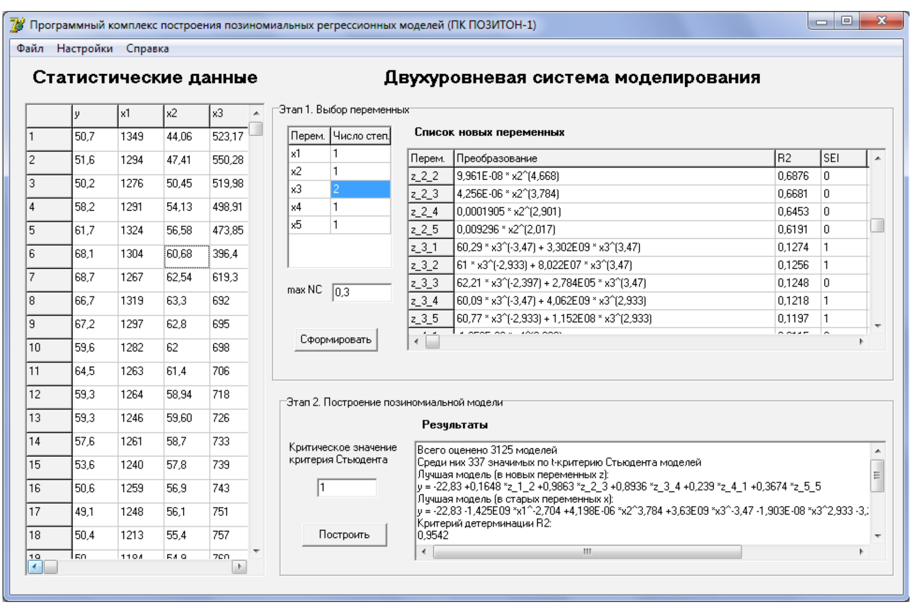

Рис. 2. Главная форма ПК ПОЗИТОН-1

ISSN 2713-1734

Во втором столбце для каждой переменной пользователю нужно задать число их степеней (число мономов). Пока предусмотрено, что это число не должно превосходить 3. Если число мономов выбрано равным 0, то это будет означать, что данная переменная не должна входить в модель. Например, если на рис. 2 во второй столбец таблицы со структурой ПозРМ внести числа 3, 0, 2, 1, 1, то это будет означать выбор следующей спецификации регрессии:

$$y_i = \alpha_0 + \alpha_{11}x_{i1}^{b_{11}} + \alpha_{21}x_{i1}^{b_{21}} + \alpha_{31}x_{i1}^{b_{31}} + \alpha_{13}x_{i3}^{b_{13}} + \alpha_{23}x_{i3}^{b_{23}} +$$

$$+ \alpha_{23}x_{i3}^{b_{23}} + \alpha_{14}x_{i4}^{b_{14}} + \alpha_{15}x_{i5}^{b_{15}} + \varepsilon_i,\ i = \overline{1,\ n}.$$

При необходимости в поле «max NC» нужно выбрать пороговое значение критерия нелинейности (8). Если нет необходимости исключать в процессе построения слишком нелинейные зависимости, то в этом поле следует выбрать значение 1.

Завершив ввод необходимой информации, нужно нажать кнопку «Сформировать» на панели «Этап 1. Выбор переменных». В результате в программе активируется главный цикл по параметру ITER — номеру выбранной объясняющей переменной. Если переменная под номером ITER не выбрана пользователем, то происходит переход на следующую итерацию. Если переменная под номером ITER выбрана, то для нее из таблицы структуры ПозРМ считывается количество мономов ch_step, а также вычисляется ее минимальное min и максимальное значение max. Потом в зависимости от числа мономов заполняется матрица данных Data, в которой первый столбец включает значения объясняемой переменной y, а остальные ch_step столбцов одинаковы и включают значения объясняющей переменной под номером ITER. После чего в зависимости от выбранного режима считываются границы области изменения степеней [a, b] и [c, d] и формируется вектор Uzel из 2*(razb+2) элементов. Этот вектор содержит значения всех точек, равномерно разбивших область [a, b] и [c, d]. Затем вычисляется число перебираемых регрессий soch, как число сочетаний из 2*(razb+2) элементов по ch_step, и в лексикографическом порядке формируется матрица сочетаний Comb размера soch * ch_step. Далее активируется вложенный цикл по параметру ITER1 — номер комбинации из матрицы Comb.

Сначала для комбинации под номером ITER1 с помощью матрицы Data формируется матрица данных Data2, в которой первый столбец включает значения объясняемой переменной y, а остальные ch_step столбцов преобразуются в зависимости от комбинации. Затем происходит нормирование всех столбцов матрицы Data2 по известному правилу $\frac{x_i - \bar{x}}{\sigma_x}$, где $\bar{x}$ — среднее значение переменной, σ_x — ее среднеквадратическое отклонение. Нормирование выполняется для повышения точности МНК-оценок ПозРМ в условиях частичной мультиколлинеарности. После чего по матрице Data2 определяются стандартизованные МНК-оценок и снова происходит переход к обычным МНК-оценкам. По последним определяется значение коэффициента детерминации. Далее с помощью разработанной функции NC вычисляется критерий нелинейности по формуле (8). Входными параметрами в этой функции выступают:

– матрица Interv, содержащая промежутки, на которые экстремумы и перегибы, найденные по формулам (3) и (4), разбили отрезок [min, max];

– число мономов ch_step;

– вектор коэффициентов позинома.

При этом если ch_step = 1, то

$$\mathrm{Interv}[1,1] = \min,\ \mathrm{Interv}[1,2] = \max;$$

если ch_step = 2, то

при SEI = 0

$$\mathrm{Interv}[1,1] = \min,\ \mathrm{Interv}[1,2] = \max,$$

при SEI = 1

$$\begin{gathered}\mathrm{Interv}[1,1] = \min,\ \mathrm{Interv}[1,2] = \mathrm{k1},\\ \mathrm{Interv}[2,1] = \mathrm{k1},\ \mathrm{Interv}[2,2] = \max,\end{gathered}$$

при SEI = 2

$$\begin{gathered}\mathrm{Interv}[1,1] = \min,\ \mathrm{Interv}[1,2] = \mathrm{k1},\\ \mathrm{Interv}[2,1] = \mathrm{k1},\ \mathrm{Interv}[2,2] = \mathrm{k2},\\ \mathrm{Interv}[3,1] = \mathrm{k2},\ \mathrm{Interv}[3,2] = \max,\end{gathered}$$

где SEI — количество экстремумов и перегибов, k1, k2 — значения экстремума и перегиба на интервале [min, max].

После определения критерия нелинейности результаты моделирования для комбинации под номером ITER1 сохраняются в критериальной матрице Krit. После окончания работы вложенного цикла с параметром ITER1 по матрице Krit отбирается максимум top-5 лучших моделей по величине коэффициента детерминации, у которых критерии нелинейности не превосходят заданной пользователем величины. Эти модели сохраняются в матрицу RangKrit, а их количество заносится в вектор keyM. После окончания работы главного цикла с параметром ITER результаты моделирования из матрицы RangKrit выводятся в таблицу «Список новых переменных» (рис. 2), расположенную на панели «Этап 1. Выбор переменных» главной формы ПК ПОЗИТОН-1. В первом столбце этой

ISSN 2713-1734

таблицы содержатся имена новых объясняющих переменных z. Например, имя z_2_1 означает, что переменная z сформирована из позинома с объясняющей переменной x_2, занявшего 1-е место по величине коэффициента детерминации. Во втором столбце содержатся формулы, устанавливающие связь между новыми переменными z и старыми переменными x. В третьем столбце содержатся значения коэффициентов детерминации построенных позиномов, в четвертом — число экстремумов и перегибов SEI, в оставшихся столбцах указаны значения критериев нелинейности на промежутках, полученных разбиением отрезка [min, max] точками экстремума и перегиба. Если нажать левой кнопкой мыши по любой характеристике переменной z, то в отдельном окне будет построен график ее позинома на диаграмме рассеивания y от x. Эти графики следует использовать для контроля степени переобучения позинома. Если по графику видно, что позином недообучен, то следует либо увеличить в нем число мономов, либо совсем его исключить, а если переобучен, то число мономов лучше сократить.

Затем пользователь может прекратить построение ПозРМ, ограничившись моделями, содержащими только по одной объясняющей переменной, либо перейти к построению ПозРМ с несколькими объясняющими переменными. В последнем случае в поле «Критическое значение критерия Стьюдента» (рис. 2) панели «Этап 2. Построение позиномиальной модели» главной формы ПК ПОЗИТОН-1 предварительно требуется ввести критическое значение t-критерия Стьюдента для исключения моделей с незначимыми коэффициентами. Далее нужно нажать кнопку «Построить».

При нажатии на кнопку «Построить» из вектора keyM исключаются нулевые элементы, означающие, что для данной переменной не нашлось ни одного позинома, удовлетворяющего заданным условиям. После чего вычисляется произведение элементов AllComb вектора keyM. Число AllComb означает сколько моделей потребуется перебрать программе, чтобы построить многофакторную ПозРМ. Затем из вектора keyM в лексикографическом порядке формируется матрица разбиений MatRazb размера AllComb * w, где w — число элементов вектора keyM. Например, если вектор keyM = (2,3,1), то матрица разбиений будет иметь вид:

$$\text{MatRazb} = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}^T.$$

Далее в программе активируется главный цикл по параметру IT — номеру выбранной комбинации из матрицы MatRazb. Для комбинации номер IT из матрицы RangKrit считывается переменная, ее тип и формируется матрица данных Data, в которой в первом столбце содержатся значения объясняемой переменной y, а в остальных столбцах — соответствующие значения новых переменных z. Затем по матрице Data определяются МНК-оценки ПозРМ, ее коэффициент детерминации и t-критерии Стьюдента. После чего результаты моделирования сохраняются в критериальной матрице Krit2. После окончания работы главного цикла с параметром IT по матрице Krit2 выбирается лучшая по величине коэффициента детерминации ПозРМ со всеми положительными и значимыми по t-критерию Стьюдента коэффициентами. Если такая модель не существует, то в поле «Результаты» (рис. 2) панели «Этап 2. Построение позиномиальной модели» выводится сообщение «Нет моделей с положительными и значимыми коэффициентами». Если ПозРМ существует, то в поле «Результаты» выводится следующая информация: общее количество оцененных моделей, количество значимых по t-критерию Стьюдента моделей с положительными коэффициентами, лучшая ПозРМ в новых переменных z и в старых переменных x, ее коэффициент детерминации и t-критерий Стьюдента.

4. Пример

Для демонстрации работы ПК ПОЗИТОН-1 решалась задача моделирования грузооборота на Дальневосточной железной дороге в зависимости от среднего веса поездов. В работе [21] показано, что средний вес поездов действительно влияет на грузооборот. Для построения модели были использованы находящиеся в открытом доступе статистические данные [22] за период с 2000 по 2018 гг. по следующим переменным:y — эксплуатационный грузооборот, млн т-км; x — средний вес грузового поезда, т.

При работе с ПК ПОЗИТОН-1 в настройках были установлены следующие параметры: границы области изменения степеней a = –5, b = –0,25, c = 0,25, d = 5; число разбиений razb = 20; max NC = 1.

В результате в ПК ПОЗИТОН-1 из 13244 регрессий была выбрана следующая ПозРМ:

$$\tilde{y} = 7{,}167 \cdot 10^5 - 4{,}382 \cdot 10^9 \cdot x^{-0{,}9286} + 1{,}849 \cdot 10^9 \cdot x^{-0{,}7024} - 2{,}176 \cdot 10^8 \cdot x^{-0{,}4762}. \quad (10)$$

Абсолютно все коэффициенты модели (10) значимы по t-критерию Стьюдента на уровне значимости 0,01. Коэффициент детерминации ПозРМ (10) равен 0,9286, что говорит о ее высоком качестве.

На рис. 3 представлены диаграмма рассеивания и график полученной зависимости (10).

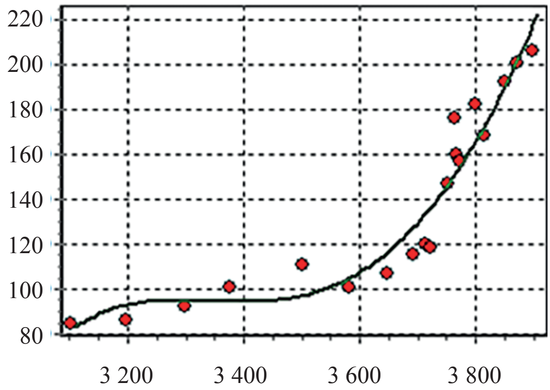

Рис. 3. График зависимости

Из графика видно, что полученная модель вполне удовлетворяет содержательному смыслу задачи. Так, при увеличении среднего веса грузового поезда увеличивается грузооборот, что, в свою очередь, приводит к снижению загруженности железнодорожной сети.

Заключение

В работе проведено исследование позинома с одной переменной и двумя мономами, представлены алгоритмы численного МНК-оценивания ПозРМ. Приводится подробное описание ПК ПОЗИТОН-1, предназначенного для численного оценивания ПозРМ с двумя и тремя мономами с помощью МНК. С помощью ПК ПОЗИТОН-1 решена задача моделирования грузооборота на Дальневосточной железной дороге.

В дальнейшем в программе планируется доработать функцию контроля степени нелинейности ПозРМ с различным числом мономов.

Разработанный программный комплекс может быть использован для решения прикладных задач из самых разных предметных областей.

Список использованной литературы

1. Бурков А. Машинное обучение без лишних слов / А. Бурков. — Санкт-Петербург : Питер, 2020. — 192 с.

2. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных/ пер. с англ. А.А. Слинкина. —Москва : ДМК Пресс, 2022. — 402 с.

3. Луис П.К. Построение систем машинного обучения на языке Python / П.К. Луис, В. Ричарт. — 2-е изд. — Москва : ДМК Пресс, 2022. — 302 с.

4. Montgomery D.C. Introduction to linear Regression Analysis / D.C. Montgomery, E.A. Peck, G.G. Vining. — New York : John Wiley & Sons, 2021. — 680 p.

5. Gelman A. Regression and Other Stories / A. Gelman, J. Hill, A. Vehtari. — Cambridge University Press, 2020. — 548 p.

6. Применение производственной функции Кобба — Дугласа для анализа промышленного комплекса региона / Н.В. Суворов, Р.Р. Ахунов, Р.В. Губарев [и др.]. — DOI 10.17059/2020-l-14. — EDN ZNLTJO // Экономика региона. — 2020. Т. 16, № 1. — С. 187–200.

7. Носков С.И. Программный комплекс построения некоторых типов кусочно-линейных регрессий / С.И. Носков, А.А. Хоняков. — EDN UTFPOD // Информационные технологии и математическое моделирование в управлении сложными системами. — 2019. — № 3. — С. 47–55.

8. Базилевский М.П. Оценивание линейно-неэлементарных регрессионных моделей с помощью метода наименьших квадратов / М.П. Базилевский. — DOI 10.26102/2310-6018/2020.31.4.026. — EDN QNQQIE // Моделирование, оптимизация и информационные технологии. — 2020. — Т. 8, № 4. — С. 26.

9. Базилевский М.П. Метод построения неэлементарных линейных регрессий на основе аппарата математического программирования / М.П. Базилевский. — DOI 10.25728/pu.2022.4.1. — EDN PHHVHL // Проблемы управления. — 2022. — № 4. — С. 3–14.

10. Мармыш Д.Е. Применение логистической регрессии к вычислению повреждаемости твердого деформируемого тела / Д.Е. Мармыш. — DOI 10.46864/1995-0470-2020-1-54-46-53. — EDN NDRLLT // Механика машин, механизмов и материалов. — 2021. — № 1. — С. 46–53.

11. Базилевский М.П. Построение степенно-показательных и линейно-логарифмических регрессионных моделей / М.П. Базилевский. — DOI 10.25728/pu.2021.3.3. — EDN BHFIAT // Проблемы управления. — 2021. — № 3. — С. 25–32.

12. Клейнер Г.Б. Производственные функции: теория, методы, применение / Г.Б. Клейнер. — Москва : Финансы и статистика, 1986. — 239 с.

13. Дурницын О.А. Моделирование расхода топлива для тяжелых транспортных средств с использованием искусственных нейронных сетей / О.А. Дурницын. — DOI 10.24412/2413-2527-2021-428-9-15 // Интеллектуальные технологии на транспорте. — 2021. — № 4. С. 9–15.

14. Брачунова У.В. Численное моделирование зарядного баланса легкового автомобиля / У.В. Брачунова. — DOI 10.24412/2071-6168-2022-9-453-459 // Известия Тульского государственного университета. Технические науки. — 2022. — № 9. — С. 453–458.

15. Моделирование электромагнитных процессов при работе силовых трансформаторов под нагрузкой и в режиме холостого хода / Д.С. Ярымбаш, М.И. Коцур, С.Т. Ярымбаш, И.М. Килимник. — DOI 10.5281/zenodo.3713396. — EDN GWYNHT // Проблемы региональной энергетики. — 2020. — № 1. — С. 1–13.

16. Сорокин А.С. Моделирование оптимальных кредитных лимитов в микрофинансовых организациях / А.С. Сорокин. — DOI 10.17323/1813-8691-2022-26-2-285-306. — EDN AZJWAV // Экономический журнал Высшей школы экономики. — 2022. — Т. 26, № 2. — С. 285–306.

17. Вытовтов А.В. Математическое моделирование процесса спасения маломобильных групп населения при пожаре / А.В. Вытовтов, Д.С. Королев, А.В. Федоров. — EDN HONNTV // Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России. — 2020. — № 4. — С. 126–131.

18. Джанполадов В.А. Прогнозирование мощности утечки на основе машинного обучения на этапе планировки физического проектирования ИС / В.А. Джанполадов, С.В. Гаврилов. — DOI 10.24151/1561-5405-2022-27-6-763-773. — EDN YQEXUG // Известия высших учебных заведений. Электроника. — 2022. — Т. 27, № 6. — С. 763–773.

19. Базилевский М.П., Караулова А.В. Способ измерения степени нелинейности многофакторных полиномиальных и позиномиальных регрессионных моделей / М.П. Базилевский, А.В. Караулова. — DOI 10.26731/2658-3704.2022.4(16).1-9. — EDN WFVGGR // Информационные технологии и математическое моделирование в управлении сложными системами. — 2022. — № 4. — С. 1–9.

20. Базилевский М.П. Критерии нелинейности квазилинейных регрессионных моделей / М.П. Базилевский. — DOI 10.26102/2310-6018/2018.23.4.015. — EDN YZSOFF // Моделирование, оптимизация и информационные технологии. — 2018. — Т. 6, № 4. — С. 185–195.

21. Колышев А.С. Масса поезда как фактор повышения эффективности грузовых перевозок / А.С. Колышев. — EDN XAAOCF // Наука и бизнес: пути развития. — 2016. — № 10. — С. 68–71.

22. Краковский Ю.М. Многофакторное прогнозирование грузооборота по данным Дальневосточной железной дороги / Ю.М. Краковский, Н.Н. Попова. — EDN ANMFCD // Транспортная инфраструктура Сибирского региона. — 2019. — Т. 1. — С. 337–341.

References

1. Burkov A. *Machine learning without further ado.* Saint Petersburg, Piter Publ., 2020. 192 p.

2. Flach P. *Machine Learning. The Art and Science of Algorithms that Make Sense of Data*. New York, 2012. 383 p. (Russ. ed.: Flakh P. *Machine Learning. The Art and Science of Algorithms that Make Sense of Data.* Moscow, DMK Press Publ., 2022. 402 p.).

3. Luis P.C., Richart V. *Building Machine Learning Systems with Python*. Birmingham; Mumbai, 2015. 295 p. (Russ. ed.: Luis P.C., Richart V. *Building Machine Learning Systems with Python*. 2nd ed. Moscow, DMK Press Publ., 2022. 302 p.).

4. Montgomery D.C., Peck E.A., Vining, G.G. *Introduction to linear Regression Analysi.* New York, John Wiley & Sons, 2021. 680 p.

5. Gelman A., Hill J., Vehtari A. *Regression and Other Stories*. Cambridge University Press, 2020. 548 p.

6. Suvorov N.V., Akhunov R.R., Gubarev R.V., Dzyuba E.I., Faizullin F.S. Applying the Cobb-Douglas Production Function for Analysing the Region's Industry. *Ekonomika regiona = Economy of Region*, 2020, vol. 16, no. 1, pp. 178–200. (In Russian). EDN: ZNLTJO. DOI: 10.17059/2020-l-14.

7. Noskov S.I., Khonyakov A.A. Software Complex for Building Some Types Pieces of Linear Regressions. *Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami = Information Technology and Mathematical Modeling in the Management of Complex Systems*, 2019, no. 3, pp. 47–55. (In Russian). EDN: UTFPOD.

8. Bazilevskiy M.P. Estimation Linear Non-Elementary Regression Models Using Ordinary Least Squares. *Modelirovanie, optimizacija i informacionnye tehnologii = Modeling, Optimization and Information Technology*, 2020, vol. 8, no. 4, pp. 26. (In Russian). EDN: QNQQIE. DOI: 10.26102/2310-6018/2020.31.4.026.

9. Bazilevskiy M.P. A Method for Constructing Nonelementary Linear Regressions Based on Mathematical Programming. *Problemy upravleniya = Control Sciences*, 2022, no. 4, pp. 3–14. (In Russian). EDN: PHHVHL. DOI: 10.25728/pu.2022.4.1.

10. Marmysh D.E. Application of Logistic Regression in Calculation of Damageability of Solid Deformable Body. *Mekhanika mashin, mekhanizmov i materialov = Mechanics of machines, mechanisms and materials*, 2021, no. 1, pp. 46–53. (In Russian). EDN: NDRLLT. DOI: 10.46864/1995-0470-2020-1-54-46-53.

11. Bazilevskii M.P. Constructing Power-Exponential and Linear-Logarithmic Regression Models. *Problemy upravleniya = Control Sciences*, 2021, no. 3, pp. 25–32. (In Russian). EDN: BHFIAT. DOI: 10.25728/pu.2021.3.3.

12. Kleiner G.B. *Production functions: theory, methods, application.* Moscow, Finansy i statistika Publ., 1986. 239 p.

13. Durnitsyn O.A. Simulation of Fuel Consumption for Heavy Vehicles Using Artificial Neural Networks. *Intellektual'nye tekhnologii na transporte = Intellectual Technologies on Transpor*, 2021, no. 4, pp. 9–15. (In Russian). DOI: 10.24412/2413-2527-2021-428-9-15.

14. Brachunova U.V. Numerical Simulation of the Charging Balance of a Passenger Car. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = Izvestiya Tula State University. Technical sciences*, 2022, no. 9, pp. 453–458. (In Russian). DOI: 10.24412/2071-6168-2022-9-453-459.

15. Yarymbash D.S., Kotsur M.I., Yarymbash S.T., Kilimnik I.M. Electromagnetic Processes Simulation of Power Transformers in Operation and in No-Load Modes. *Problemy regional'noi energetiki = Problems of regional energy*, 2020, no. 1, pp. 1–13. (In Russian). EDN: GWYNHT. DOI: 10.5281/zenodo.3713396.

16. Sorokin A.S. Modeling of Optimal Credit Limits in Microfinance Organizations. *Ekonomicheskii zhurnal Vysshei shkoly ekonomiki = Higher School of Economics Economic Journal*, 2022, vol. 26, no. 2, pp. 285–306. (In Russian). EDN: AZJWAV. DOI: 10.17323/1813-8691-2022-26-2-285-306.

17. Vytovtov A.V., Korolev D.S., Fedorov A.V. Mathematical Modeling of the Rescue Process Small-Mobile Population in the Fire. *Vestnik Sankt-Peterburgskogo universiteta Gosudarstvennoi protivopozharnoi sluzhby MChS Rossii = Bulletin of the St. Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia*, 2020, no. 4, pp. 126–131. (In Russian). EDN: HONNTV.

18. Dzhanpoladov V.A., Gavrilov S.V. A Machine Learning-Based Leakage Power Prediction at Floorplan Stage of IC Physical Design. *Izvestiya vysshikh uchebnykh zavedenii. Elektromekhanika = Russian electromechanics*, 2022, vol. 27, no. 6, pp. 763–773. (In Russian). EDN: YQEXUG. DOI: 10.24151/1561-5405-2022-27-6-763-773.

19. Bazilevskii M.P., Karaulova A.V. A Method for Measuring the Nonlinearity Degree of Multivariate Polynomial and Posinomial Regression Models. *Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami = Information Technology and Mathematical Modeling in the Management of Complex Systems*, 2022, no. 4, pp. 1–9. (In Russian). EDN: WFVGGR. DOI: 10.26731/2658-3704.2022.4(16).1-9.

20. Bazilevskiy M.P. Nonlinear Criteria of Quasilinear Regression Models. *Modelirovanie, optimizacija i informacionnye tehnologii = Modeling, Optimization and Information Technology*, 2018, vol. 6, no. 4, pp. 185–195. (In Russian). EDN: YZSOFF. DOI: 10.26102/2310-6018/2018.23.4.015.

21. Kolyshev A.S. Train Weight as a Factor of Increasing the Efficiency of Freight Transport. *Nauka i biznes: puti razvitiya= Science and Business: Ways of Development*, 2016, no. 10, pp. 68–71. (In Russian). EDN: XAAOCF.

22. Krakovskii Yu.M., Popova N.N. Multi-factor forecasting of cargo turnover according to the data of the Far Eastern Railway. *Transportnaya infrastruktura Sibirskogo regiona = Transport infrastructure of the Siberian region*, 2019, vol. 1, pp. 337–341. (In Russian). EDN: ANMFCD.

Информация об авторах

Базилевский Михаил Павлович — кандидат технических наук, доцент кафедры математики, Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация, e-mail: mik2178@yandex.ru.

Караулова Анна Витальевна — аспирант, кафедра информационных систем и защиты информации, Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация, e-mail: anuta_kav@mail.ru.

Information about the Authors

Mikhail P. Bazilevskiy — PhD in Technical Sciences, Associate Professor of the Department of Mathematics, Irkutsk State Transport University, Irkutsk, Russian Federation, e-mail: mik2178@yandex.ru.

Anna V. Karaulova — PhD Student, Department of Information Systems and Information Protection, Irkutsk State Transport University, Irkutsk, Russian Federation, e-mail: anuta_kav@mail.ru.

Вклад авторов

Все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the Authors

The authors contributed equally to this article. The authors declare no conflicts of interests.

Для цитирования

Базилевский М.П. Алгоритмы и программное обеспечение численного оценивания позиномиальных регрессионных моделей с помощью метода наименьших квадратов / М.П. Базилевский, А.В. Караулова. — DOI 10.17150/2713-1734.2023.5(2).97-114 — EDN XFYMSJ // System Analysis & Mathematical Modeling. — 2023. — Т. 5, № 2. — С. 97–114.

For Citation

Bazilevskiy M.P., Karaulova A.V. Algorithms and Software for Numerical Estimation of Posinomial Regression Models Using the Least Squares Method. *System Analysis & Mathematical Modeling*, 2023, vol. 5, no. 2, pp. 97–114. (In Russian). EDN: XFYMSJ. DOI: 10.17150/2713-1734.2023.5(2).97-114.