

Носков С. И.
S. I. Noskov

**ПРИМЕНЕНИЕ МНОГОКРИТЕРИАЛЬНОГО МЕТОДА НАИМЕНЬШИХ МОДУЛЕЙ
ДЛЯ МОДЕЛИРОВАНИЯ КОЛИЧЕСТВА ДОРОЖНО-ТРАНСПОРТНЫХ
ПРОИСШЕСТВИЙ**

**APPLICATION OF THE MULTICRITERIA METHOD OF LEAST MODULES
FOR MODELING ROAD ACCIDENT NUMBER**

Носков Сергей Иванович – доктор технических наук, профессор кафедры «Информационные системы и защита информации» Иркутского государственного университета путей сообщения (Россия, Иркутск); Россия, 664074, ул. Чернышевского, д. 15; тел. 8(914)902-24-94. E-mail: sergey.noskov.57@mail.ru.

Sergey I. Noskov – Doctor of Technical Sciences, Professor, Information Systems and Information Protection Department, Irkutsk State University of Railway Transport (Russia, Irkutsk); Russia, 664074, st. Chernyshevsky, 15; tel. 8(914)902-24-94. E-mail: sergey.noskov.57@mail.ru.

Аннотация. Дан краткий обзор работ, посвящённых применению методов регрессионного моделирования при исследовании закономерностей, связанных с аварийностью на транспорте. Приведена вычислительная схема многокритериального метода наименьших модулей, предполагающая разделение исходной выборки данных на несколько непересекающихся подвыборок с последующей минимизацией векторной функции потерь, что приводит к необходимости решения серии задач линейного программирования. С помощью этого метода построена регрессионная модель количества дорожно-транспортных происшествий в России. В качестве независимых переменных при этом использованы: количество троллейбусов, общая численность населения, численность трудоспособного сельского населения, ежегодное потребление алкоголя, протяжённость автомобильных дорог общего назначения.

Summary. A brief review of works devoted to the application of regression modeling methods in the study of regularities related to transport accidents is given. A computational scheme of the multicriteria method of least absolute deviations is presented, which assumes the division of the initial data sample into several non-overlapping sub-samples with subsequent minimization of the vector loss function, which leads to the need to solve a series of linear programming problems. Using this method, a regression model of the number of road accidents in Russia was built. In this case, the following were used as independent variables: the number of trolleybuses, the total population, the number of able-bodied rural population, annual alcohol consumption, and the length of general roads.

Ключевые слова: регрессионная модель, многокритериальный метод наименьших модулей, вклады факторов, дорожно-транспортные происшествия, множество Парето, задача линейного программирования.

Key words: regression model, multi-criteria method of least absolute deviations, factor contributions, traffic accidents, Pareto set, linear programming problem.

УДК 330.4

Введение. Методы математического, в том числе регрессионного, моделирования активно используются в современном мире при анализе самого широкого круга различных объектов и процессов. Применяются они и при исследовании закономерностей, связанных с аварийностью на транспорте. Так, в работе [1] подтверждена эффективность моделей Пуассона, отрицательной биномиальной и отрицательной биномиальной регрессии с нулевым расширением для установления связи между дорожно-транспортными происшествиями с участием грузовиков и дорожными и геометрическими характеристиками многополосной междугородной дороги Анкара-Аксарай-Эрегли. Коэффициенты модели оценивались методом максимального правдоподобия, а отклонение и информационный критерий Акаике рассматривались как статистика согласия. Результаты

показывают, что первая модель очень хорошо соответствует данным. Предлагаемая модель турецких разделённых многополосных междугородных дорог с высоким процентом движения грузовых автомобилей может быть полезна для выявления критических факторов и снижения количества аварий. Работа [2] посвящена применению адаптивной системы нейро-нечёткого вывода для моделирования дорожно-транспортных происшествий. Набор объясняющих переменных модели включал среднюю горизонтальную кривизну, ширину обочины, ширину дороги, землепользование, точки доступа, продольный уклон и плотность горизонтальной кривой. В [3] предложена нейронная сеть обобщённой регрессии, настроенная на самоорганизующуюся карту для оценки риска дорожно-транспортных происшествий. Эта гибридная прогностическая модель оценивает риск дорожно-транспортных происшествий, рассматривая 22 различных прогностических переменных, включая географические характеристики, временные и погодные условия, характеристики дороги, характеристики транспортного средства и водителя, рассчитанные на основе авторитетных источников данных и добровольной географической информации, сформированной путём разработки стороннего приложения, которое запускалось внутри мессенджера Telegram. В статье [4] для прогнозирования серьёзности аварии применялся метод ансамблевого ранжирования, включающий три метода ранжирования на основе использования фильтров, таких как прирост информации, коэффициент усиления и симметричная неопределённость. В результате анализа были выбраны восемь важных характеристик для прогнозирования серьёзности аварий: среднегодовой трафик, ширина дороги, процент транспортных средств, поворачивающих налево, возраст водителя, процент транспортных средств, поворачивающих направо, типы задействованных транспортных средств, точечная скорость и причины аварий. Для прогнозирования тяжести аварий на городских дорогах использовались множественная логистическая регрессия и искусственная нейронная сеть. Исследование [5] направлено на выявление временных закономерностей для прогнозирования количества аварий в Пакистане путём использования одномерного анализа временных рядов, такого как модели сезонной авторегрессии, интегрированного скользящего среднего и экспоненциального сглаживания. Результаты показывают, что последняя модель лучше соответствует данным об авариях после расчёта наименьшей средней абсолютной ошибки, среднеквадратической ошибки, средней абсолютной процентной ошибки и нормализованного байесовского информационного критерия. В работе [6] метод гедонической регрессии используется для оценки стоимости обеспечения безопасности дорожного движения с использованием информации о шведском рынке автомобилей. Результаты исследования показывают, что рыночная цена автомобиля отрицательно коррелирует с присущим ему уровнем риска, т. е. шведские потребители автомобилей платят надбавку за безопасность, приобретая более безопасные автомобили. По сравнению с предыдущими шведскими исследованиями заявленных предпочтений это исследование показывает более низкую готовность платить за дополнительную безопасность автомобиля, что может быть результатом взаимодействия между государственным вмешательством и индивидуальным самострахованием и самозащитой. Представляется, что для оценки рисков дорожно-транспортных происшествий (ДТП) может быть эффективно использован подход, описанный в [7], для управления рисками энергетических компаний.

В настоящей работе моделирование количества ДТП в Российской Федерации осуществляется с помощью многокритериального метода наименьших модулей (МНМ) [8].

Вычислительная схема многокритериального МНМ. Рассмотрим кратко алгоритм работы многокритериального МНМ [8]. Пусть задано линейное регрессионное уравнение (модель):

$$y_k = \sum_{i=1}^m \alpha_i x_{ki} + \varepsilon_k, \quad k = \overline{1, n}, \quad (1)$$

где y , x_i – соответственно зависимая и i -я независимая переменные; α_i – i -й идентифицируемый параметр; ε_k – ошибки аппроксимации, k – номер наблюдения, n – число наблюдений. Будем считать модель (1) детерминированной.

Уравнение (1) представимо в векторном виде:

$$y = X\alpha + \varepsilon,$$

где $y = (y_1, \dots, y_n)^T$; $\alpha = (\alpha_1, \dots, \alpha_m)^T$; $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$; $X - (n \times m)$ – матрица с компонентами x_{ki} .

Пусть, руководствуясь некоторыми существенными обстоятельствами, которые могут иметь самый различный характер, исследователь при построении регрессионной модели (1) для анализируемого объекта, учитывая его характеристики или суть сформулированной при моделировании задачи, может (или обязан) разделить исходную выборку данных на s непересекающихся подвыборок (групп номеров наблюдений) $P_i, i = \overline{1, s}$:

$$\bigcup_{i=1}^s P_i = \{1, 2, \dots, n\}, P_i \cap P_j = \emptyset, i \neq j.$$

Многокритериальный МНМ состоит в решении задачи минимизации векторной функции потерь:

$$I(\alpha) = \left(\sum_{k \in P_1} |\varepsilon_k|, \sum_{k \in P_2} |\varepsilon_k|, \dots, \sum_{k \in P_s} |\varepsilon_k| \right) \rightarrow \min. \quad (2)$$

Задача (2) решается в два этапа. Вначале выделяется множество паретовских вершин N^{ex} симплекса B , задаваемого по правилу

$$B = \{(\alpha, u, v) \mid \sum_{i=1}^m \alpha_i x_{ki} + u_k - v_k = y_k, \quad u_k \geq 0, \quad v_k \geq 0, \quad k = \overline{1, n}\}.$$

Это может быть сделано посредством решения серии задач линейного программирования (ЛП):

$$\min_{(\alpha, u, v) \in B} \sum_{i=1}^s \gamma_i^l \sum_{k \in P_i} (u_k + v_k) \rightarrow \min, \quad (3)$$

где γ_i^l – узлы равномерной γ -сети s -мерного открытого единичного куба Γ :

$$\Gamma = \{\gamma = (\gamma_1, \gamma_2, \dots, \gamma_s) \mid \gamma_i > 0, i = \overline{1, s}, \sum_{i=1}^s \gamma_i = 1\}.$$

Наконец, окончательное формирование всего множества Парето N на основе построенного множества N^{ex} в [8] предлагается базировать на подходе, описанном в фундаментальной работе Yu L. и Zeleny M. [9]. Множество N представляет собой объединение паретовских выпуклых комбинаций векторов из множества N^{ex} .

Идентификация параметров модели численности ДТП с помощью многокритериального МНМ. В работе [10] на основе обработки данных официальной статистики за 2004 – 2019 гг. построена регрессионная модель динамики ДТП в России с использованием непрерывной формы метода максимальной согласованности (НММС), которая имеет вид

$$\begin{aligned} y_k &= \alpha_0 + \alpha_1 x_{k1} + \alpha_2 x_{k2} + \alpha_3 x_{k3} + \alpha_4 x_{k4} + \alpha_5 x_{k5} + \varepsilon_k = \\ &= -584\,066.39 + 13\,468.49 x_{k1} + 2810.48 x_{k2} + 6.19 x_{k3} + 6648.59 x_{k4} + 29.32 x_{k5} + \varepsilon_k = \hat{y}_k + \varepsilon_k, \quad (4) \\ k &= \overline{1, 16}, E = 2.95, M = 97\,126.21, L = 23\,131.46, \end{aligned}$$

где y – количество ДТП с пострадавшими, чел.; x_1 – количество троллейбусов, тыс. ед.; x_2 – численность населения в России, млн чел.; x_3 – численность трудоспособного сельского населения, тыс. чел.; x_4 – ежегодное потребление алкоголя, литры чистого спирта на человека в возрасте более 15 лет; x_5 – протяжённость автомобильных дорог общего назначения, тыс. км; \hat{y}_k – расчётные (вычисленные по модели) значения зависимой переменной; E – средняя относительная ошибка аппроксимации:

$$E = 100\% \sum_{k=1}^n \frac{|\varepsilon_k|}{y_k} / n;$$

M – максимальная по модулю ошибка:

$$M = \max_{k=1, \bar{n}} |\varepsilon_k|;$$

L – непрерывное представление значения критерия согласованности поведения между расчётными и фактическими значениями зависимой переменной:

$$L = \sum_{k=1}^{n-1} \sum_{s=k+1}^n l_{ks},$$

здесь

$$l_{ks} = \begin{cases} |\hat{y}_k - \hat{y}_s|, & (\hat{y}_k - \hat{y}_s)(y_k - y_s) < 0 \\ 0, & \text{в противном случае.} \end{cases}$$

Займёмся уточнением оценок параметров модели (4) с помощью многокритериального МНМ. Разобьём весь период наблюдений на две части – первую и вторую половины:

$$P_1 = \{1, 2, \dots, 8\},$$

$$P_2 = \{9, 10, \dots, 16\}.$$

Решая серию задач ЛП (3) с помощью специализированной программы [11] на основе перебора узлов равномерной γ -сети s -мерного открытого единичного куба Γ с шагом 0.01 сформируем множество N^{ex} паретовских векторов оценок параметров модели (4) (см. табл. 1).

Таблица 1

Множество паретовских векторов оценок параметров модели (4)

№	α_0	α_1	α_2	α_3	α_4	α_5
1	-2 950 323	27 625.71	13 251.83	41.860 03	-3411.23	68.224 32
2	-1 984 372	3262.132	9741.627	27.788 45	9701.109	40.175 78
3	-1 677 575	-4476.03	8626.744	23.319 15	13 865.74	31.267 24
4	-1 506 618	-4926.6	7860.18	19.567 51	16 107.97	34.515 69
5	-1 342 438	-6728.6	7178.958	16.8033	17 612.56	32.460 52
6	-874 808	1452.103	4782.644	9.731 316	12 727.76	25.158 79
7	-788 796	2555.959	4350.565	8.425 332	12 130.69	23.893 26
8	-721 368	2151.67	4039.332	7.384 977	12 617.34	23.146 62
9	-502 163	6773.093	2804.171	0.466 472	16 684.19	42.193 08
10	189 944.3	6376.066	-1579.86	3.363 961	7567.881	18.823 02
11	229 824.6	12 307.71	-1366.4	-2.719 42	7075.612	19.684 04
12	282 152.7	12 669.85	-1667.17	-2.911 99	6388.531	18.062 49
13	236 427.3	13 289.76	-1331.06	-3.7266	6994.111	19.826 59
14	139 516.7	16 823.13	-406.952	-7.285 54	7082.347	20.085 08
15	-66 526.4	24 335.49	1557.815	-14.8523	7269.947	20.634 66
16	435 440.6	7118.464	-4258.17	5.7257	5758.995	88.595 15

Проанализируем знаки компонент всех шестнадцати векторов параметров, составляющих множество N^{ex} . Будем иметь при этом в виду, что, в соответствии с содержательным смыслом независимых переменных, все коэффициенты модели, за исключением свободного члена, должны быть положительны, что, собственно, и проявилось в модели (4). Таким образом, допустимыми по знакам параметров являются пять входящих в состав N^{ex} векторов: № 2, 6, 7, 8, 9. Любой из них может быть выбран для замены коэффициентов модели (4) при решении исследователя о применении для идентификации её параметров многокритериального МНМ. Для однозначного выбора «лучшего» варианта можно воспользоваться теми же критериями адекватности, которые сопровождали построение этой модели в [10]. В табл. 2 приведены их значения для этих вариантов.

Значения критериев адекватности паретовских векторов параметров модели (4)

Номер варианта	E	L	M
2	3.261	117 550	99 464
6	2.167	64 195	72 507
7	2.132	58 883	72 123
8	2.139	58 881	72 129
9	2.173	56 910	73 758

Анализ табл. 2 позволяет сделать вывод, что вариант № 7 является лучшим, т. е. регрессионная модель (4) после применения многокритериального МНМ может быть представлена в виде

$$y_k = -788\,796 + 2555.959x_{k1} + 4350.565x_{k2} + 8.425x_{k3} + 12\,130.69x_{k4} + 23.893x_{k5} + \varepsilon_k, \quad k = \overline{1,16}, \quad (5)$$

$$E=2.132, M=72\,123, L=58\,883.$$

Отметим, что средняя относительная E и максимальная по модулю M ошибки аппроксимации модели (5) существенно меньше, чем модели (4). Тот факт, что лишь по критерию L модель (4), напротив, лучше модели (5), совершенно естественен, поскольку именно в соответствии с оптимизацией L и производилась идентификация параметров первой.

Проанализируем теперь, как изменилась значимость независимых переменных при переходе от (4) к (5). Отметим рост степени влияния почти в два раза факторов x_2 и x_4 , а также незначительный рост x_3 . Вместе с тем сильно уменьшилась значимость переменной x_1 (более чем в пять раз) и незначительно, на 22 %, переменной x_5 .

В соответствии с предложенной в работе [12] методикой оценим вклады факторов δ_i , $i = \overline{1,5}$ в модели (5):

$$\delta_1 = 2.3 \%,$$

$$\delta_2 = 67 \%,$$

$$\delta_3 = 17.5 \%,$$

$$\delta_4 = 9.3 \%,$$

$$\delta_5 = 3.9 \%.$$

Таким образом, наиболее значимым фактором, определяющим динамику количества ДТП в стране, является численность её населения. Обращает на себя внимание тот факт, что, в соответствии с моделью (5), почти 10 % роста количества ДТП объясняется потреблением алкоголя.

Заключение. Регрессионная модель (5) может быть эффективно использована для решения широкого круга аналитических и прогнозных задач. При проведении многовариантных прогнозных расчётов количества ДТП предварительно необходимо разработать соответствующие прогнозные сценарии, различающиеся будущими значениями независимых переменных.

ЛИТЕРАТУРА

1. Kibar T. F., Celik T., Aytac B. P. Statistical Analysis of Truck Accidents for Divided Multilane Interurban Roads in Turkey // KSCE Journal of Civil Engineering. – 2018. – V. 22. – P. 1927-1936.
2. Hosseinpour M., Yahaya A. S., Ghadiri S. M., Prasetijo J. Application of Adaptive Neuro-fuzzy Inference System for road accident prediction // KSCE Journal of Civil Engineering. – 2013. – V. 17. – P. 1761-1772.
3. Charandabi N. K., Gholami A., Bina A. A. Road accident risk prediction using generalized regression neural network optimized with self-organizing map // Neural Computing and Applications. – 2022. – V. 34. – P. 8511-8524.
4. Sarkar A., Sarkar S. Comparative Assessment Between Statistical and Soft Computing Methods for Accident Severity Classification // Journal of The Institution of Engineers (India): Series A. – 2020. – V. 101. – P. 27-40.

5. Rabbani M. B. A., Musarat M. A., Alaloul W. S. and others. A Comparison Between Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ES) Based on Time Series Model for Forecasting Road Accidents // *Arabian Journal for Science and Engineering*. – 2021. – V. 46. – P. 11113-11138.
6. Andersson H. The Value of Safety as Revealed in the Swedish Car Market: An Application of the Hedonic Pricing Approach // *Journal of Risk and Uncertainty*. – 2005. – V. 30. – P. 211-239.
7. Ляпустина, Ю. В. Разработка алгоритма оптимизации комплекса методов управления рисками энергетических компаний / Ю. В. Ляпустина, М. А. Горькавый, В. А. Соловьев // *Учёные записки Комсомольского-на-Амуре государственного технического университета. Науки о природе и технике*. – 2017. – № II-1 (30). – С. 26-35.
8. Носков, С. И. Многокритериальный метод наименьших модулей в регрессионном анализе / С. И. Носков // *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*. – 2023. – № 1. – С. 28-36.
9. Yu L., Zeleny M. The set of all nondominated solutions in linear cases and multicriteria simplex method // *J. of Math. Anal. and Applic.* – 1975. – V. 49. – № 2. – P. 430-460.
10. Носков, С. И. Регрессионная модель динамики дорожно-транспортных происшествий в Российской Федерации / С. И. Носков, Ю. А. Бычков // *Южно-Сибирский научный вестник*. – 2023. – № 1 (47). – С. 37-42.
11. Свидетельство о государственной регистрации программ для ЭВМ № 2021613936 Программа определения паретовских оценок параметров линейной регрессии посредством применения многокритериального метода наименьших модулей / Носков С. И., Хоняков А. А. (Россия); правообладатель Федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный университет путей сообщения» (ФГБОУ ВО ИргУПС); заявка № 2022680313 28.10.2022; дата регистр. 10.11.2022.
12. Носков, С. И. Оценка динамики вкладов факторов в линейной регрессионной модели / С. И. Носков // *Вестник Воронежского государственного технического университета*. – 2021. – Т. 17. – № 5. – С. 15-19.