

Заключение

Dynamo предоставляет гибкую и настраиваемую платформу для автоматизации различных задач в процессе проектирования, от создания геометрии до анализа и оптимизации.

Используя Dynamo, проектировщики и инженеры могут сократить время и усилия, необходимые для выполнения задач ручного проектирования, что позволяет им сосредоточиться на более важных аспектах дизайна, таких как креативность и инновации. Dynamo также позволяет интегрировать различные инструменты проектирования и анализа, упрощая изучение и оценку различных вариантов проектирования и сценариев.

Можно сказать, что использование Dynamo для автоматизации проектирования является осуществимым и практичным подходом, который может

значительно повысить эффективность и качество процесса проектирования. Благодаря своей гибкой и настраиваемой платформе Dynamo может использоваться для автоматизации широкого спектра задач проектирования.

В рамках исследования пошагово разработана технология автоматизации процесса армирования фундаментной плиты.

В целом, автоматизация процессов проектирования с использованием Dynamo Revit и кодирования на Python может повысить производительность, сократить количество ошибок и предоставить проектировщикам больше времени для творчества и инноваций.

Список литературы

1. Divin N. V. BIM by using Revit API and Dynamo. A review // AlfaBuild. – 2020. – No 2(14). – P. 1404. – DOI: 10.34910/ALF.14.4.
2. Регидо О. Автоматизированное архитектурное проектирование в интегрированной компьютерной среде Revit-Dynamo-Excel // Danish Scientific Journal. – 2020. – № 38-1. – С. 7-11.
3. Fudala T. Learn the Power of Dynamo for Automated Reinforcement Detailing – Part 1 // Autodesk Knowledge Network. Revit 2018. – URL: <https://blogs.autodesk.com/revit/2018/11/27/learn-the-power-of-dynamo-for-automated-reinforcement-detailing-part-1/> (дата обращения: 02.04.2023).
4. Sherif M., Nassar K., Hosny O., Safar O., Abotaleb I. Automated BIM-based structural design and cost optimization model for reinforced concrete buildings // Journal of Facilities Management (2Scientific Reports 12(1):21616 P. 1–18, DOI: 10.1038/s41598-022-26146-6.
5. Fudala T. Learn the Power of Dynamo for Automated Reinforcement Detailing – Part 2 // Autodesk Knowledge Network. Revit 2018. – URL: <https://blogs.autodesk.com/revit/2018/12/13/learn-the-power-of-dynamo-for-automated-reinforcement-detailing-part-2/> (дата обращения: 03.04.2023).
6. Borges M. M. BIM Modelling automation on reinforcement detailing of SLABS – Text : electronic // INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO, Porto, Portugal, 2018. – URL: <https://core.ac.uk/download/pdf/161253402.pdf> (дата обращения: 29.03.2023).
7. Цифровизация предпроектной и проектной стадий в реализации инвестиционно-строительного проекта многофункционального жилого комплекса / Т. В. Золина, Н. В. Купчикова, К. Е. Джантазаева, Е. Е. Купчиков // Инженерно-строительный вестник Прикаспия. – 2022. – № 3(41). – С. 144-148. – DOI 10.52684/2312-3702-2022-41-3-144-148. – EDN ТКАНСМ.
8. Цифровизация процессов стадии строительства в реализации инвестиционно-строительного проекта многофункционального жилого комплекса / Н. В. Купчикова, Т. В. Золина, К. Е. Джантазаева, Е. Е. Купчиков // Инженерно-строительный вестник Прикаспия. – 2022. – № 4(42). – С. 71-80. – DOI 10.52684/2312-3702-2022-42-4-71-80. – EDN ЕВХВНД.

© С. П. Заторский, К. А. Шумилов

Ссылка для цитирования:

Заторский С. П., Шумилов К. А. Автоматизация процессов проектирования в рабочей среде с помощью Dynamo Revit // Инженерно-строительный вестник Прикаспия : научно-технический журнал / Астраханский государственный архитектурно-строительный университет. Астрахань : ГАОУ АО ВО «АГАСУ», 2023. № 4 (46). С. 125–128.

УДК 35.073

DOI 10.52684/2312-3702-2023-46-4-128-131

АНАЛИЗ РЕАКЦИИ ОЦЕНОК ПАРАМЕТРОВ ЛИНЕЙНОЙ РЕГРЕССИИ НА ВЫБРОСЫ В ДАННЫХ (НА ПРИМЕРЕ МОДЕЛИ ЦЕН НА ЖИЛЬЕ)

С. И. Носков

Носков Сергей Иванович, профессор кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация; e-mail: sergey.noskov.57@mail.ru

В работе дан краткий обзор публикаций по способам обработки данных с выбросами - наблюдениями, не похожими на остальные, не согласующимися с ними. К ним, в частности, относятся: генетический алгоритм, который рассматривает различные возможные группировки данных на наблюдения с выбросами и без них; метод непараметрической регрессии для настройки ковариат и преобразование проблемы обнаружения выбросов в проблему многомерной регуляризации в модели экспоненциальной регрессии с поправкой на ковариацию с цензурированными данными; метод выявления множественных выбросов в моделях линейной регрессии. Предложен способ выявления реакции на выбросы в данных оценок параметров линейной регрессионной модели, рассчитанных с помощью методов наименьших квадратов, модулей и антиробастного оценивания. Этот способ основан на использовании максимальных частных приростов соответствующих оценок. Выявлена реакция на выбросы регрессионной модели цен на жилье в Российской Федерации. При этом в качестве независимых переменных использованы средние цены железобетонных плит перекрытий, товарного бетона, песчано-гравийных смесей. Показано, что наибольшей изменчивостью обладает последний параметр.

Ключевые слова: регрессионная модель, выбросы в данных, методы наименьших квадратов, модулей и антиробастного оценивания, оценки параметров, чувствительность.

ANALYSIS OF THE RESPONSE OF LINEAR REGRESSION PARAMETER ESTIMATES TO DATA OUTPUTS
(BY THE EXAMPLE OF A HOUSING PRICE MODEL)

S. I. Noskov

Noskov Sergey Ivanovich, Professor of the Department of Information Systems and Information Security, Irkutsk State University of Transport, Irkutsk State University of Transport, Irkutsk, Russian Federation; e-mail: sergey.noskov.57@mail.ru

The paper gives a brief review of publications on data processing methods with outliers - observations that are not similar to the rest, inconsistent with them. These include, in particular: a genetic algorithm that considers various possible groupings of data for observations with and without outliers; a nonparametric regression method for adjusting covariates and converting an outlier detection problem to a multivariate regularization problem in a covariance-adjusted exponential regression model with censored data; method for detecting multiple outliers in linear regression models. A method is proposed for detecting the response to outliers in the data of estimates of the parameters of a linear regression model calculated using the least squares, moduli and anti-robust estimation methods. This method is based on the use of the maximum partial increments of the corresponding estimates. The response to outliers of the regression model of housing prices in the Russian Federation is estimated. At the same time, the average prices of reinforced concrete floor slabs, ready-mixed concrete, sand and gravel mixtures were used as independent variables. It is shown that the last parameter has the greatest variability.

Keywords: regression model, data outliers, least squares, modulus and anti-robust estimation methods, parameter estimates, sensitivity.

Введение

Весьма часто при построении математических моделей регрессионного типа для сложных объектов исследователю приходится сталкиваться с ситуациями, когда подлежащая обработке и анализу статистическая информация содержит выбросы [1-3] – наблюдения, не похожие на остальные, не согласующиеся с ними. Способам обработки данных с выбросами посвящено множество публикаций. Так, в работе [4] представлен генетический алгоритм, который рассматривает различные возможные группировки данных на наблюдения с выбросами и без них. Таким образом, все выбросы обнаруживаются одновременно. Известно, что обнаружение выбросов и выбор переменных могут влиять друг на друга и что могут быть получены разные результаты в зависимости от порядка выполнения этих двух задач. Поэтому может оказаться полезным рассматривать эти задачи одновременно, что и делается посредством указанного алгоритма. В [5] используется метод непараметрической регрессии для настройки ковариат и преобразование проблемы обнаружения выбросов в проблему многомерной регуляризации в модели экспоненциальной регрессии с поправкой на ковариацию с цензурированными данными. Предлагается метод штрафного правдоподобия с плавным усечением абсолютного отклонения для обнаружения возможных выбросов, отличающийся тем, что он может одновременно иметь дело с обнаружением выбросов и оценками коэффициентов регрессии. В статье [6] представлен новый метод обнаружения выбросов. При его использовании все возможные комбинации множественных выбросов рассматриваются как ошибки модели и принимается, что наименьшая из них дает решение для определенного количества выбросов. Затем оцениваемые ошибки модели выявляются путем сравнения с неким критическим значением. В [7] представлен обзор современных методов обнаружения выбросов. Их причинами, в частности, являются: механические неисправности, изменения в поведении системы, мошеннические действия, человеческие ошибки, ошибки приборов или просто естественные отклонения во внешней среде. В [8] разработан новый метод выявления множественных

выбросов в моделях линейной регрессии. Он относительно прост и удобен в использовании. Метод основан на результате, дающем асимптотические свойства экстремальных невязок. Этот результат доказывается при достаточно общих условиях на процедуру оценивания и ковариантное распределение. Проведенное имитационное исследование показывает, что предлагаемый метод имеет хорошую производительность по сравнению с различными существующими методами. В работе [9] предлагается новый подход к обнаружению выбросов без учителя, который переформулирует проблему обнаружения выбросов в числовых данных как набор задач регрессионного обучения с учителем. Для каждого атрибута изучается прогностическая модель, которая предсказывает значения этого атрибута на основе значений всех других атрибутов и вычисляет отклонения между прогнозами и фактическими значениями переменных. На основе этих отклонений рассчитывается вес каждого атрибута и окончательная оценка выброса с использованием этих весов. В [10] предлагается подход, при котором вводятся штрафные издержки за отбрасывание выброса, следовательно, наилучшее соответствие для большинства данных достигается за счет отбрасывания только т.н. катастрофических наблюдений. Надежная оценка получается путем решения задачи выпуклого квадратичного смешанного целочисленного программирования, где в целевой функции минимизируется сумма квадратов невязок и штрафов за отбрасывание наблюдений. Предложенная формула математического программирования подходит для небольших выборок. Кроме того, проводится имитационное исследование, чтобы сравнить другие робастные оценки с данным подходом с точки зрения его эффективности и надежности.

Изменчивость оценок параметров линейной регрессии под влиянием выбросов в данных

Рассмотрим линейную регрессионную модель (зависимость, уравнение) вида [11, 12]:

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

где y – зависимая (выходная), x_i – i -ая независимая (входная) переменные, α_i – i -ый оцениваемый параметр; ε_k – k -ая ошибка аппроксимации, k – номер наблюдения, n – их количество.

Представим уравнение (1) в векторной форме:



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} .

Весьма широкий класс методов оценивания неизвестных параметров модели (1) связан с вычислением так называемых L_v -оценок, являющихся результатом решения задач [1]:

$$J_v(\alpha) = \sum_{k=1}^n |\varepsilon_k|^v, v \geq 1. \quad (3)$$

При этом значению $v=1$ соответствует метод наименьших модулей (МНМ), $v = 2$ - метод наименьших квадратов (МНК), а $v \rightarrow \infty$ - метод антиробастного оценивания (МАО) [13].

Эти методы обладают в числе других одним важнейшим свойством. МНМ выделяет среди всех n наблюдений ровно m , через которые и проходит гиперплоскость регрессии (т.е. для которых $\varepsilon_k = 0$). Остальные $n - m$ наблюдений попросту игнорируются, можно неявным образом полагать - считать выбросами. МАО является своего рода антиподом по отношению к МНМ, поскольку, наоборот, сильно тяготеет к выбросам, отклоняя к ним указанную гиперплоскость. МНК же занимает промежуточную позицию, являясь неким компромиссом между МНМ и МАО.

В этой связи вызывает интерес ответ на вопрос - как на выбросы реагируют оценки параметров α_i , $i = \overline{1, m}$, если для их расчета применять три упомянутых метода? Или, другими словами - как различаются оценки $\alpha_i^j, j = \text{МНМ, МНК, МАО}$?

В работе [11] предложен обобщенный ответ на этот вопрос путем введения так называемого индекса параметрической стабильности (ИПС) выборки данных (Ху):

$$\Xi = \frac{1}{ym} \sum_{i=1}^m |(\alpha_i^{\text{МНМ}} - \alpha_i^{\text{МАО}}) \bar{x}_i|,$$

где \bar{y} и \bar{x}_i - средние на ней значения соответствующих переменных:

$$\bar{y} = \frac{1}{n} \sum_{k=1}^n y_k,$$

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki}.$$

Этот обобщенный ответ может быть детализирован введением в рассмотрение максимальных частных приростов соответствующих оценок (их реакции на выбросы) ξ_i следующим образом:

$$\xi_i = \frac{\max(\alpha_i^{\text{МНМ}}, \alpha_i^{\text{МНК}}, \alpha_i^{\text{МАО}}) - \min(\alpha_i^{\text{МНМ}}, \alpha_i^{\text{МНК}}, \alpha_i^{\text{МАО}})}{\max(|\alpha_i^{\text{МНМ}}|, |\alpha_i^{\text{МНК}}|, |\alpha_i^{\text{МАО}}|)}$$

Анализ реакции на выбросы оценок параметров линейной регрессионной модели цен на жилье

В работе [14] на основе статистической информации с первого квартала 2017 г. по четвертый квартал 2020 г. была построена кусочно-линейная модель риска для цен на жилье в Российской Федерации. Построим на тех же данных (т.е. вычислим оценки параметров) с помощью МНМ, МНК и МАО линейную регрессионную модель:

$$y_k = \alpha_0 + \sum_{i=1}^3 \alpha_i x_{ki} + \varepsilon_k, \quad k = \overline{1, 16}, \quad (4)$$

где y - средняя цена одного квадратного метра общей площади квартир на рынке жилья, руб.; x_1 - средняя цена железобетонных плит перекрытий, руб./м³; x_2 - средняя цена товарного бетона, руб./м³; x_3 - средняя цена песчано-гравийных смесей, руб./м³.

В результате получим (табл. 1).

Таблица 1

Оценки параметров модели (4)

Метод оценивания	α_0	α_1	α_2	α_3
МНК	-54674.933	2.737	21.273	6.235
МНМ	-68436.420	1.307	27.862	2.752
МАО	-63795.747	3.111	23.905	-1.716

Анализ табл. 1 позволяет сделать следующие выводы. Для всех параметров, кроме α_1 , отсутствует монотонность в изменении их значений с ростом v в выражении (3). Знак параметра α_3 принимает несоответствующее смыслу переменной x_3 значение при использовании МАО.

В таблице 2 приведены фактические и расчетные значения зависимой переменной модели (4), а в таблице 3 - ошибки аппроксимации для каждого из трех используемых метода оценивания параметров.

Таблица 2

Фактические и расчетные значения зависимой переменной модели (4)

y	МНК	МНМ	МАО
56347.2	57272.7	56978.8	57979.9
56516.8	56008.9	56516.8	56772.3
56560.8	54697.7	56560.8	55161.7
56882.2	59047.8	60635.7	60036.8
58875.6	60810.8	61384	59071.6
59969.7	60568.1	60220.6	58224.5
60952.8	59999.1	59627.7	57798.3
61831.6	61693.9	61831.6	59083.9

60705.1	64465.8	66186.1	63859.7
61618.3	59030.9	60699.9	58463.7
62891.9	61672.6	62299.6	61851.6
64059.5	64396.4	64641.7	62804.8
71503.2	70126	70004.1	68772.5
73438.1	73370	73438.1	70906.8
76167.2	78006.4	80164.4	77694.3
79003	76155.6	78359.7	75848.4

Таблица 3

Ошибки аппроксимации модели (4)

МНК	МНМ	МАО
-925.494	-631.612	-1632.67
507.841	0.000	-255.566
1863.036	0.000	1399.032
-2165.65	-3753.51	-3154.56
-1935.25	-2508.44	-196.003
-598.418	-250.898	1745.178
953.698	1325.091	3154.568
137.640	0.000	2747.718
-3760.62	-5480.98	-3154.56
2587.363	918.310	3154.568
1219.381	592.363	1040.339
-336.940	-582.197	1254.677

1377.218	1499.100	2730.773
68.015	0.000	2531.244
-1839.19	-3997.18	-1527.11
2847.398	643.299	3154.568

$$\begin{aligned} \xi_0 &= 0.201, \\ \xi_1 &= 0.580, \\ \xi_2 &= 0.236, \\ \xi_3 &= 1.275. \end{aligned}$$

Их анализ указывает на соблюдение свойств МНМ и МАО [15] – для первого из них число нулевых ошибок аппроксимации равно четырем (поскольку $m=4$), а для второго число максимальных по модулю ошибок аппроксимации ($|\varepsilon_k|=3154.568$) равно пяти (поскольку $m + 1=5$). При этом, в соответствии с обоими методами, явным выбросом является наблюдение № 9.

Рассчитаем для данной выборки значение индекса параметрической стабильности \mathcal{E} :

$$\mathcal{E} = 0.183.$$

Оно указывает на относительно высокую параметрическую стабильность выборки по отношению к L_V -оценкам.

Теперь проанализируем, как на выбросы реагируют оценки параметров $\alpha_i^j, i = \overline{0, m}, j = \text{МНМ, МНК, МАО}$, считав их максимальные частные приросты ξ_i :

Наиболее чувствителен к выбросам параметр α_3 , наименее – параметры α_0 и α_2 . При этом тот факт, что $\xi_3 > 1$, указывает на несовпадение знаков оценок $\alpha_3^j, j = \text{МНМ, МНК, МАО}$.

Заключение

В работе предложен способ выявления реакции на выбросы в данных оценок параметров линейной регрессионной модели, рассчитанных с помощью методов наименьших квадратов, модулей и антиробастного оценивания. Этот способ основан на использовании максимальных частных приростов соответствующих оценок. Выявлена реакция на выбросы регрессионной модели цен на жилье в Российской Федерации.

Список литературы

1. Демиденко Е.З. Линейная и нелинейная регрессии. - М.: -Финансы и статистика. -1981. - 302 с.
2. Баенхаева А.В., Базилевский М.П., Носков С.И. Моделирование валового регионального продукта Иркутской области на основе применения методики множественного оценивания регрессионных параметров // Фундаментальные исследования. - 2016.- № 10-1. - С. 9- 14.
3. Носков С.И., Удилов В.П. Управление системой обеспечения пожарной безопасности на региональном уровне. -Иркутск: ВСИ МВД России. - 2003. -151 с.
4. Tolvi J. Genetic algorithms for outlier detection and variable selection in linear regression models // Soft Computing. - 2004. - V. 8. - P.527-533.
5. Yingli Pan, Zhan Liu, Guangyu Song. Outlier detection under a covariate-adjusted exponential regression model with censored data // Computational Statistics. – 2021. – V. 36. – P. 961-976.
6. Hekimoglu S., Erdogan B., Erenoglu R. C. A new outlier detection method considering outliers as model errors // Experimental Techniques. – 2015. – V. 39. - P. 57-68.
7. Hodge V., Austin J. A Survey of Outlier Detection Methodologies // Artificial Intelligence Review, - 2004. – V. 22. – P. 85-126.
8. Bagdonavičius V., Petkevičius L. A new multiple outliers identification method in linear regression // Metrika. – 2020. – V. 83. – P. 275-296.
9. Paulheim H., Meusel R. A decomposition of the outlier detection problem into a set of supervised learning problems // Machine Learning. - 2015. – V. 100. – P. 509-531.
10. Zioutas G., Avramidis A. Deleting Outliers in Robust Regression with Mixed Integer Programming // Acta Mathematicae Applicatae Sinica. – 2005. – V. 2. – P. 323-334.
11. Носков С.И. Индекс параметрической стабильности регрессионной модели // Вестник транспорта Поволжья. - 2021. - № 2 (86). - С. 74-77.
12. Носков С.И., Глухов Н.И., Кириллова Т.К., Попов Е.С. Математическая модель динамики дорожного строительства в Российской Федерации // Инженерно-строительный вестник Прикаспия. - 2022. - № 4 (42). - С. 95-99.
13. Носков С. И. Компромиссные паретовские оценки параметров линейной регрессии // Математическое моделирование. - 2020. - Т. 32. - № 11. - С. 70-78.
14. Носков С. И., Хоняков А. А. Применение функции риска для модельного описания ожидаемых цен на рынке недвижимости // Инженерно-строительный вестник Прикаспия. - 2021. - № 3 (37). - С. 77-82.
15. Носков С.И. Метод антиробастного оценивания параметров линейной регрессии: число максимальных по модулю ошибок аппроксимации // Южно-Сибирский научный вестник. - 2020. - № 1. - С. 51-54.

© С. И. Носков

Ссылка для цитирования:

Носков С. И. Анализ реакции оценок параметров линейной регрессии на выбросы в данных (на примере модели цен на жилье) // Инженерно-строительный вестник Прикаспия : научно-технический журнал / Астраханский государственный архитектурно-строительный университет. Астрахань : ГАОУ АО ВО «АГАСУ», 2023. № 4 (46). С. 128–131.

УДК 004.94

DOI 10.52684/2312-3702-2023-46-4-131-136

РАСЧЕТНО- ЭКСПЕРИМЕНТАЛЬНЫЙ МЕТОД СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МАЛЫХ ВЫБОРОК

Р. З. Хайруллин

Хайруллин Рустам Зиннатуллович, доктор физико-математических наук, ведущий научный сотрудник Главного метрологического центра Минобороны России; профессор кафедры фундаментального образования, Московский государственный строительный университет; тел. +7 (926) 405-22-17; e-mail: zrkzrk@list.ru