УДК 518.852+518.853.14

9

***К.С. Перфильева*[1]**

[1]*Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация*

ПРОГРАММНЫЙ КОМПЛЕКС ПОСТРОЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ МЕТОДОМ СМЕШАННОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ

Аннотация: В статье рассматривается программный комплекс, обеспечивающий возможность оценивания параметров линейного регрессионного уравнения методами смешанного оценивания (МСО), наименьших квадратов и модулей, а также антиробастного. Рассмотрен численный пример.

Ключевые слова: Линейная регрессия, метод смешанного оценивания, метод наименьших модулей, антиробастное оценивание, программный комплекс.

***K. S. Perfilieva*[1]**

[1]*Irkutsk state University of railway engineering, Russian Federation.*

SOFTWARE PACKAGE FOR CONSTRUCTING LINEAR REGRESSION USING THE MIXED PARAMETER ESTIMATION METHOD

Annotation: The article considers a software package that provides the ability to evaluate the parameters of a linear regression equation using mixed estimation methods (MSO), least squares and modules, as well as anti-robust. A numerical example is considered.

Keyword: Linear regression, mixed estimation method, least module method, anti-robust estimation, software package.

Рассмотрим линейное регрессионное уравнение

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

где y – эндогенная, а x_i – i-ая экзогенная переменные; α_i – i-ый подлежащий оцениванию параметр; ε – ошибки аппроксимации, k – номер наблюдения, n – их число.

Представим уравнение (1) в матричной форме:

$$y = X\alpha + \varepsilon, \tag{2}$$

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

Методам оценивания параметров уравнения (2) посвящена обширная литература (см., например, [1-13]).

Широкий класс методов оценивания параметров уравнения (2) связан с поиском так называемых L_ν –оценок посредством минимизации функций потерь вида [2, 7-9]:

$$J_\nu^P(\alpha) = \sum_{k \in P} \left|\varepsilon_k\right|^\nu, P = \{1,2....n\}.$$

Каждая из этих функций характеризуется реакцией на так называемые выбросы, то есть наблюдения, не согласующиеся со всей выборкой. При этом чем больше значение ν, тем сильнее L_ν – оценка реагирует на выбросы. В регрессионном анализе методы оценивания, слабо реагирующие на выбросы, или вообще их игнорирующие, принято называть устойчивыми, или робастными.

Методом оценивания параметров уравнения (2), соответствующим $\nu = 2$, является наиболее часто используемый в регрессионном анализе метод наименьших квадратов

(МНК). При $\nu=1$ это метод наименьших модулей (МНМ), при $\nu\to\infty$ – метод антиробастного оценивания (МАО).

В [2] впервые выдвинута идея о том, что поиск вектора параметров линейной регрессии (2) может быть осуществлен с помощью минимизации суммы разных функций потерь на разных участках выборки. Этот метод назван его автором методом смешанного оценивания (МСО) параметров [14].

Общая постановка задачи определяет основные требования к построению программного комплекса метода смешанного оценивания параметров регрессионной модели, который должен:

– обеспечивать возможность полностью автоматизировать процесс построения выбранных математических моделей, начиная с анализа исходной информации и ее преобразования и заканчивая получением значений оцениваемых параметров;

– поддерживать возможность ввода исходной информации путем импорта файла;

– гарантировать высокую точность обработки исходных данных.

Для того чтобы начать работать с системой, пользователь должен выбрать файл с исходными данными. Этот файл представляет собой обычный текстовый файл с расширением *.txt, содержащий матрицу значений зависимой и независимых переменных. Значения могут быть как положительными, так и отрицательными, как целыми, так и вещественными. К этому файлу предъявляются следующие требования:

– файл не должен содержать никаких данных, кроме значений числового формата;

– при вводе численных значений столбцы матрицы отделяются друг от друга с помощью пробела;

– для вещественных чисел целая часть отделяется от дробной запятой.

Рассмотрим простой численный пример. Пусть дана выборка:

$$X=\begin{pmatrix}2,5,7,1,2,8\\9,4,9,4,9,6\\6,1,1,8,3,3\\8,3,6,5,3,2\\1,7,8,4,1,7\\5,8,5,5,0,6\\2,4,8,7,8,1\\7,2,1,2,4,0\\8,5,9,2,7,5\\9,2,8,1,6,4\end{pmatrix}.$$

Построим для нее двухфакторную линейную регрессию

$$y=\alpha_1 x_1+\alpha_2 x_2+\varepsilon$$

четырьмя упомянутыми в работе методами. При этом зависимой переменной y соответствует столбец наблюдений номер 1, а независимым – столбцы с номерами номер 2 и 5.

Исходную выборку разобьем на две подвыборки с множествами номеров наблюдений N_1={1, 2, 3, 9, 10} и N_2={4, 5, 6, 7, 8}.

Запускаем ПК и выполняем решение поставленной задачи (рис.1).

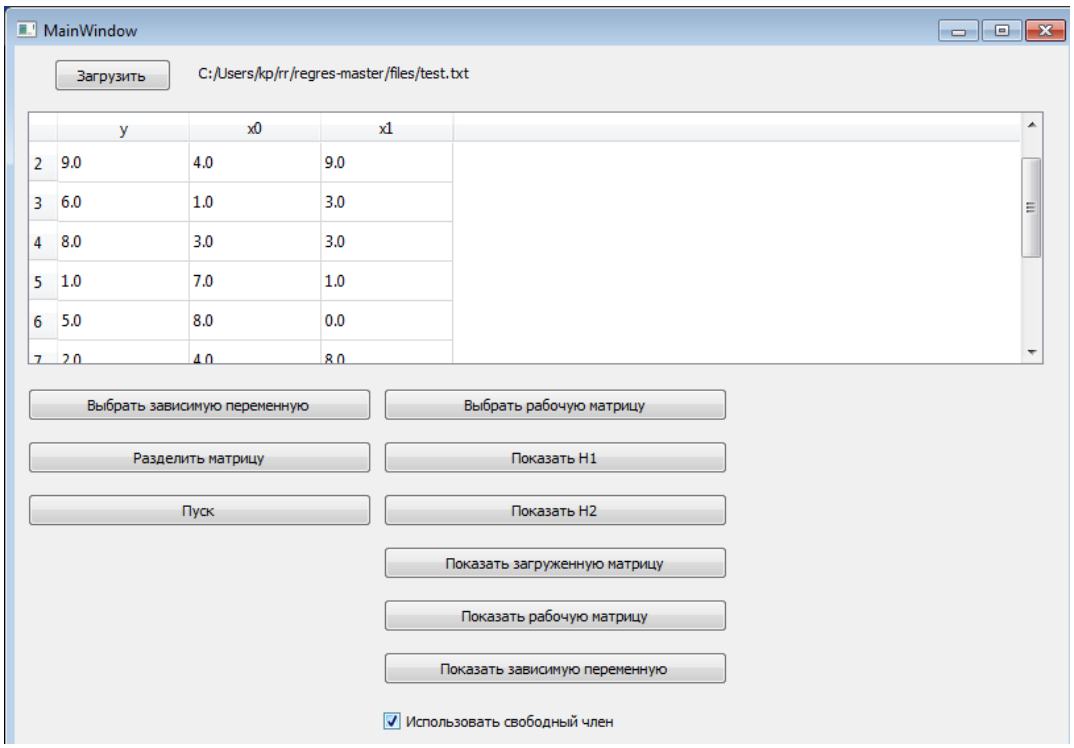

Рис.1. Начальная экранная форма программного комплекса

Чтобы выбрать файл с исходными данными, необходимо нажать кнопку «Загрузить». После этого открывается диалоговое окно, в котором используется фильтр «Тип файлов». Пользователь может выбирать только файлы *.txt.

После того, как пользователь выбрал файл с исходными данными, они будут отображаться в диалоговом окне. Далее пользователь вводит номер зависимой и независимых переменных, нажимая на кнопки «Выбрать зависимую переменную» и «Выбрать рабочую матрицу» соответственно. Далее пользователь делит рабочую матрицу на 2 подматрицы, нажимая на кнопку «Разделить матрицу». Выбор может быть осуществлен либо случайно, либо сделан пользователем (рис. 2).

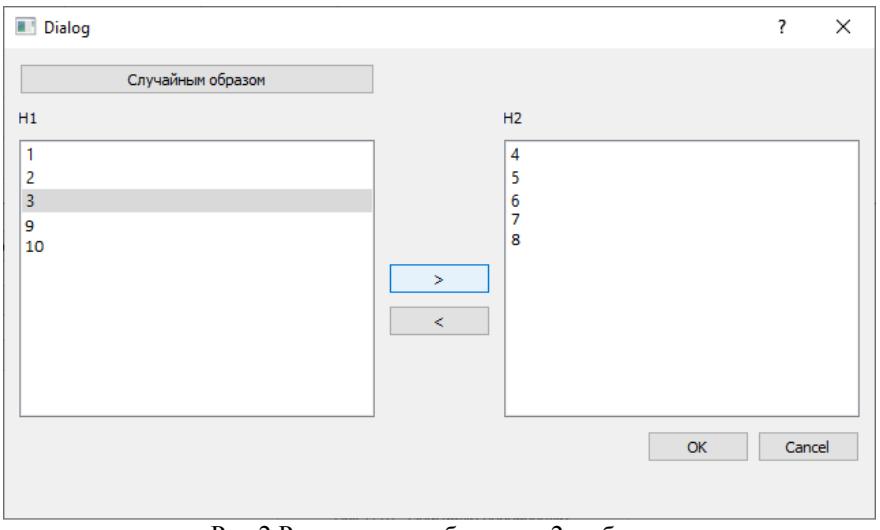

Рис.2 Разделение выборки на 2 выборки

После этого пользователь может выбрать режим построения регрессии с учетом свободного члена или без него. В нашем примере свободный член должен учитываться. Далее пользователь нажимает кнопку «Пуск» и программа открывает новое диалоговое окно с результатами расчетов. Они представлены на рис. 3.

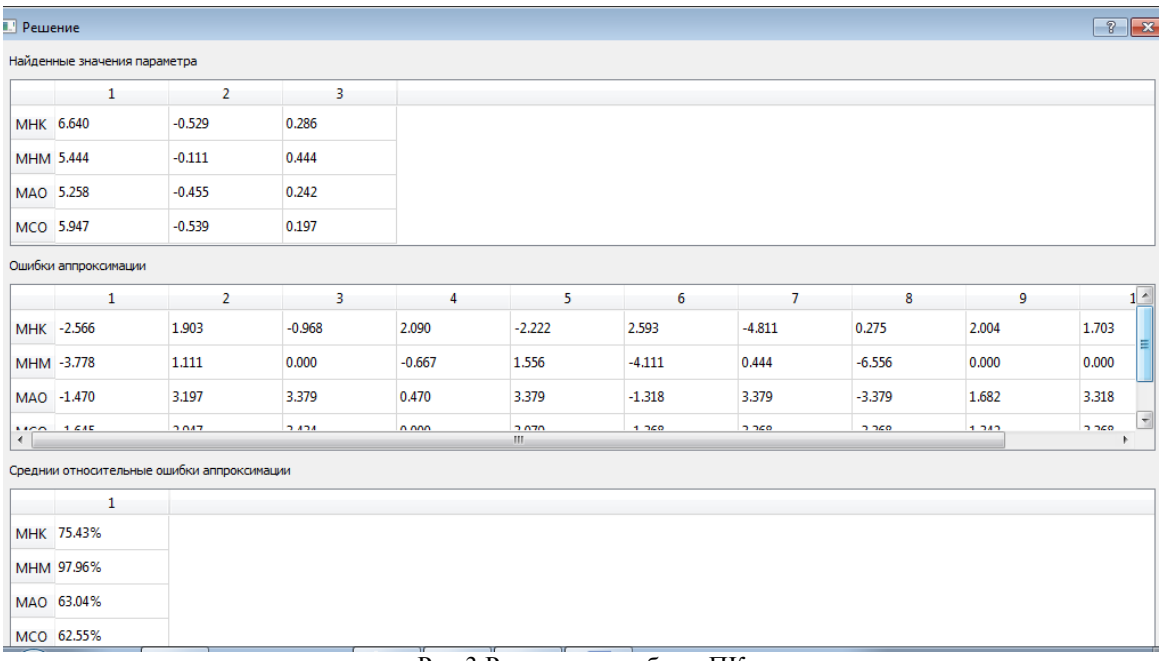

Рис.3 Результаты работы ПК

В окне «Решение» можно увидеть таблицу «Найденные значения параметров», где представлены значения a0, a1, a2 соответственно для каждого метода оценивания параметров. С помощью этих параметров можно в явном виде представить все четыре уравнения регрессии, соответствующие МНК, МНМ, МАО, МСО:

МНК: y=6,640-0,529x1+0,286x2;

МНМ: y=5,444-0,111x1+0,444x2;

МАО : y=5,258-0,455x1+0,242x2;

МСО: y=5,947-0,539x1+0,197x2.

Далее в окне «Решение» отображаются значения ошибок аппроксимации для каждого наблюдения по каждому из четырех методов, а также средние относительные ошибки аппроксимации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Финансы и статистика. 1981.Т.1. 366 с., Т. 2. 351с.

2. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. Иркутск: Облинформпечать.-1996. -320с.

3. Носков С.И., Баенхаева А.В. Множественное оценивание параметров линейного регрессионого уравнения // Современные технологии. Системный анализ. Моделирование. 2016. № 3 (51). -С. 133-138.

4. Носков С.И., Быкова О.В., Некипелова О.Е., Соколова Л.Е. Возможный способ поиска компромиссного решения в задаче линейного программирования с векторной целевой функцией // Фундаментальные исследования. 2014. № 6-3.- С. 502-505.

5. Носков С.И. Критерий «согласованность поведения» в регрессионном анализе //Современные технологии. Системный анализ. Моделирование. 2013. № 1 (37).- С. 107-110.

6. Лакеев А.В., Носков С.И. Метод наименьших модулей для линейной регрессии: число нулевых ошибок аппроксимации // Современные технологии. Системный анализ. Моделирование. 2012. № 2 (34). -С. 48-50.

7. Носков С.И. Проблема единственности Парето-оптимального решения в задаче линейного программирования с векторной целевой функцией // Современные технологии. Системный анализ. Моделирование. 2011. № S-4 (32). -С. 283-285.

8. Носков С.И. Точечная характеризация множества Парето в линейной многокритериальной задаче // Современные технологии. Системный анализ. Моделирование. 2008. № 1 (17).- С. 99-101.

9. Носков С.И. L-множество в многокритериальной задаче оценивания параметров регрессионных уравнений // Информационные технологии и проблемы математического моделирования сложных систем, 2004. № 1. -С. 64 - 69.

10. Носков С.И. Построение эконометрических зависимостей с учетом критерия «согласованность поведения» // Кибернетика и системный анализ, 1994. № 1. -С. 177 - 181.

11. Носков С.И., Удилов В.П. Управление системой обеспечения пожарной безопасности на региональном уровне. Иркутск, 2003.

12. Kreinovich V., Lakeyev A.V., Noskov S.I. Approximate linear algebra is intractable // Linear Algebra and its Applications. 1996. Т. 232. № 1-3. С. 45-54.

13. Носков С.И., Базилевский М.П. Построение регрессионных моделей с использованием аппарата линейно-булевого программирования. -Иркутск, 2018.

14. Носков С.И. О методе смешанного оценивания параметров линейной регрессии// Информационные технологии и математическое моделирование в управлении сложными системами. – 2019. – №1. – С. 14-20.

REFERENCES

1. Dreyper N., Smith G. Applied regression analysis. M.: Finance and statistics. 1981. V.1. 366 P., V. 2. 351 P.

2. Noskov S.I. Technology of modeling of objects with unstable functioning and uncertainty in data. Irkutsk: Regional Information Printing.-1996.-320 P.

3. Noskov S.I., Bayenkhayeva A.V. Multiple estimation of parameters of the linear regression equation//Modern technologies. Systems analysis. Modeling. 2016. N. 3 (51). – P.P. 133-138.

4. Noskov S.I., Bykova O.V., Nekipelova O.E., Sokolova L.E. A possible way of search of a compromise solution in a problem of linear programming with vector target function//Basic researches. 2014. N. 6-3. – P.P. 502-505.

5. Noskov S.I. Criterion "coherence of behavior" in regression analysis//Modern technologies. Systems analysis. Modeling. 2013. N. 1 (37).-P.P. 107-110.

6. Lakeev A.V., S.I. Metod Socks of the smallest modules for a linear regression: number of zero errors of approximation//Modern technologies. Systems analysis. Modeling. 2012. N. 2 (34). – P.P. 48-50.

7. Noskov S.I. A problem of uniqueness of a pareto-optimal solution in a problem of linear programming with vector target function//Modern technologies. Systems analysis. Modeling. 2011. N. 2-4 (32). – P.P. 283-285.

8. Noskov S.I. Point characterization of a Pareto set in a linear multicriteria problem//Modern technologies. Systems analysis. Modeling. 2008. N. 1 (17).-P.P. 99-101.

9. Noskov S.I. A L-set in a multicriteria problem of estimation of parameters of the regression equations//Information technologies and problems of mathematical modeling of complex systems, 2004. N. 1. – P.P. 64 - 69.

10. Noskov S.I. Creation of econometric dependences taking into account criterion "coherence of behavior"//Cybernetics and systems analysis, 1994. -N. 1. – P.P. 177 - 181.

11. Noskov S.I., Udilov V.P. Fire safety management system at the regional level. Irkutsk, 2003.

12. Kreinovich V., Lakeyev A.V., Noskov S.I. Approximate linear algebra is intractable // Linear Algebra and its Applications. 1996. Т. 232. № 1-3. С. 45-54.

13. Noskov S.I., Bazilevsky M.P. Construction of regression models using linear Boolean programming. Irkutsk, 2018.

14. Noskov S.I. About the method of mixed estimation of parameters of linear regression// Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal.-2019. N. 1. P. 14-20.

Информация об авторе

Перфильева Карина Сергеевна – аспирант кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: 552649-171233@mail.ru.

Author

Perfilieva Karina Sergeevna – Postgraduate Student, «Information systems and information security», Irkutsk State Transport University, Irkutsk, e-mail: 552649-171233@mail.ru

Для цитирования

Перфильева К.С. О методе смешанного оценивания параметров линейной регрессии// «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2020. – №1(6). – С. 9-14. DOI: 10.26731/2658-3704.2020.1(6).9-14 – Режим доступа: http://ismm-irgups.ru/toma/16-2020, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 20.01.2020)

For citation

Perfilieva K.S., About the method of mixed estimation of parameters of linear regression// Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2020. No. 1(6). P. 9-14. DOI: 10.26731/2658-3704.2020.1(6). 9-14 [Accessed 20/01/2020].