УДК 519.688

Об одном методе проверки статистических данных на наличие выбросов

И.В. Пономарев
АлтГУ, г. Барнаул

При работе со статистическими данными возникает необходимость проверки построенной модели на предмет выбросов. Наличие выбросов может негативно сказаться на адекватности модели и ее прогностических способностях. Разработке методов поиска выбросов посвящены работы [1–4].

В данной работе рассматривается универсальный метод для исследования регрессионной модели на наличие выбросов. Универсальность этого метода заключается в том, что он может применяться к моделям, основанным на разных функционалах качества.

Пусть имеется линейная регрессионная модель

$$y = a_0 + a_1 x_1 + \ldots + a_k x_k + \varepsilon, \qquad (1)$$

где y – зависимая переменная; x_i ($i = \overline{1,k}$) – независимые переменные; ε – ошибка; a_i ($i = \overline{0,k}$) – параметры модели.

Для оценки параметров модели необходимо минимизировать функционал качества

$$F = d(Y, \hat{Y}), \qquad (2)$$

где Y – вектор наблюдаемых значений зависимой переменной; $\hat{Y}$ – вектор расчетных для модели (1) значений зависимой переменной; d – выбранная метрика.

В зависимости от выбранной метрики получаются различные типы регрессионных зависимостей [5–8]:

1) если d манхэттенская метрика, то соответствующая модель будем обозначать L_1-регрессией;

2) если d евклидова метрика, то соответствующая модель будем обозначать L_2-регрессией;

3) если d чебышевская метрика, то соответствующая модель будем обозначать L_∞-регрессией.

Будем обозначать минимальное значение функционала (2) через α_p, где p выбирается согласно используемой метрики ($1,\ 2$ или ∞).

Задачу о нахождении выбросов сформулируем следующим образом: пусть из данного множества наблюдений $\Omega=\left\{(x_{i1},x_{i2},\ldots,x_{ik},y_i):i=\overline{1,N}\right\}$ требуется исключить фиксированный процент наблюдений так, чтобы оставшиеся Ω_0 данные имели наименьшую величину разброса $\alpha_p\left(\Omega_0\right)$, т.е.

$$\alpha_p\left(\Omega_0\right)=\min\left\{\alpha_p\left(\Omega'\right):\Omega'\subset\Omega,\#\left[\Omega'\right]=N_0\right\}, \tag{3}$$

где $\#\left[\Omega'\right]$ – число элементов во множестве Ω'; $N-N_0=M_0$ – число выбросов.

Изучаемый алгоритм основан на преобразовании Лежандра.

Определение. Пусть $\Omega=\left\{A_i(x_{i1},x_{i2},\ldots,x_{ik},y_i):i=\overline{1,N}\right\}$ – конечное множество точек и дана пара натуральных чисел $1\le r,s\le N$. Обозначим через

$$MAX_r\left[\left\{c_i\right\}_{i=1}^{N}\right]=c_{i_{r+1}},\quad MIN_s\left[\left\{c_i\right\}_{i=1}^{N}\right]=c_{i_{N+s}},$$

где $\left\{c_{i_k}\right\}_{k=1}^{N}$ – перестановка последовательности $\left\{c_i\right\}_{i=1}^{N}$ в порядке убывания.

Используя введенные функции можно определить обобщенные преобразования Лежандра. Например, для L_2-регрессии преобразования Лежандра имеет вид:

$$f_r^{+}(a_0,\ldots,a_k)=MAX_r\left\{\left(a_0+a_1x_{i1}+\ldots+a_kx_{ik}-y_i\right)^2:i=1,\ldots,N\right\},$$

$$f_s^{-}(a_0,\ldots,a_k)=MIN_s\left\{\left(a_0+a_1x_{i1}+\ldots+a_kx_{ik}-y_i\right)^2:i=1,\ldots,N\right\}$$

Справедлива следующая теорема.

Теорема. Справедливо равенство

$$\min\left\{\alpha_2\left(\Omega'\right):\Omega'\subset\Omega,\#\left[\Omega'\right]=N_0\right\}=\min_{a_0,\ldots,a_k}\sum_{0\le r\le M_0-1}f_r^{-}(a_0,\ldots,a_k).$$

Библиографический список

1. Weisberg S. Applied linear regression. – 3th ed. – Jonh Wiley & Sans, Inc., 2005.
2. Cook R.D. Detection of Influential Observation in Linear Regression // Technometrics. – 1977. – Vol. 19, No. 1. – P. 15–18.
3. Andrews D.F., Pregibîn D. Finding the outliers that matter // Journal of the Royal Statistical Society. – 1978. – Vol. 40. – P. 84–93.
4. Пономарев И.В. Исследование статистических данных на выбросы // МАК: Математики – Алтайскому краю : сборник трудов всероссийской конференции по математике. – Барнаул: Изд-во Алт. ун-та, 2017. – С. 133–135.
5. Ponomarev I.V., Slavsky V.V. Uniformly fuzzy model of linear regression // Journal of Mathematical Sciences. – 2012. – Vol. 186. – Issue 3. – P. 478 – 494.
6. Пономарев И.В., Славский В.В. Нечеткая модель линейной регрессии // Доклады Академии наук. – 2009. – Т. 428, №5. – С. 598–600.
7. Пономарев И.В., Родионов Е.Д., Родионова Л.В., Славский В.В. Комплекс моделей для построения и оценки вариантов развития регионального рынка труда // Вестник Алтайской науки. – 2013. – №1. – С. 86–88.
8. Родионов Е.Д., Родионова Л.В., Славский В.В. и другие. Применение пакетов символьных вычислений к решению задач теории и практики: монография. – Концепт, Барнаул, 2014.

УДК 330.131.7

Актуализация программы капитального ремонта многоквартирных домов с использованием нейтрософских компонентов

Е.В. Токарева, С.П. Пронь
АлтГУ, г. Барнаул

Региональная программа капитального ремонта (КР) многоквартирных домов (МКД) как документ планирования, в котором указаны выборочные КР для каждого включенного в программу МКД с указанием трехлетнего планового периода (в некоторых регионах этот срок увеличен до шести лет) требует ежегодной актуализации. Формально в статье рассматривается математическая модель, позволяющая обоснованно переупорядочить массив выборочных КР МКД в текущем периоде.