

языке Пролог // Известия Алтайского государственного университета. – Барнаул, 2011. – №1/1 (69). – С. 117–120.

3. Фреймовая модель знаний [Электронный ресурс] // (версия от 19.05 2017). – Режим доступа: <http://helpiks.org/7-30113.html>, свободный.

**УДК 519.65**

## Исследование статистических данных на выбросы

*И.В. Пономарев*  
*АлтГУ, г. Барнаул*

При обработке статистических данных появляется необходимость оценить их на предмет выбросов. Выброс – это результат измерения, выделяющийся из общей выборки. Выбросы способны исказить статистики и результаты статистического моделирования.

В данной статье исследуются методы определения выбросов как для классических статистических регрессионных моделей, так и для моделей нечеткой регрессии.

Рассмотрим линейную регрессионную модель

$$y = a_0 + a_1x_1 + \dots + a_kx_k + \varepsilon, \quad (1)$$

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

Наиболее часто применяемым методом оценки параметров модели (1) является метод наименьших квадратов (МНК), который является очень чувствительным к выбросам, т.е. наличие в исходной выборке выбросов может привести к неверным выводам о наблюдении процесса и ставит под угрозу результаты статистического анализа модели.

В работе [1] для исследования выборки на выбросы предлагается использовать статистику Стьюдента

$$t_i = \frac{\tilde{y}_i - \tilde{y}_i^*}{s_i^2 \sqrt{1 + X_i(X_i^T X_i)^{-1} X_i^T}}, \quad (2)$$

где  $\tilde{y}_i, \tilde{y}_i^*$  – прогнозные значения зависимой переменной с учетом и без учета  $i$ -го наблюдения;  $X_i$  – матрица наблюдений без  $i$ -го наблюдения;  $s_i^2$  – оценка дисперсии ошибки регрессии без учета  $i$ -го наблюдения. Данная статистика имеет распределение Стьюдента с  $N - k - 2$  степенями свободы ( $N$  – объем выборки). Если гипотеза подтвержда-

ется, то делается вывод о том, что  $i$ -ое наблюдение является выбросом. Сложностью данного алгоритма является последовательная проверка всех наблюдений.

Аналогичный подход к определению выбросов предложен в работе [2]. Оценка критерия происходит с помощью расстояния Кука, которое имеет распределение Фишера, и работает со всеми прогнозными значениями зависимой переменной.

В работе [3] предложен подход, позволяющий проверять на выбросы сразу несколько наблюдений. Это способствует убыстрению процесса исследования выборки и уменьшению вычислительной трудности алгоритма.

Следуя логике приведенных исследований, нами предложен способ оценки выборки с использованием преобразований Лежандра. Пусть из данного множества наблюдений  $\Omega = \{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i) : i = \overline{1, N}\}$  требуется исключить фиксированный процент наблюдений так, чтобы оставшиеся  $\Omega_0$  данные имели наименьшую величину разброса  $\alpha_p(\Omega_0)$ , т.е.

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

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

Данный алгоритм может быть применен к регрессионным моделям, использующим различные нормы  $p$  при расчете функционала качества. Так, например, в работах [4,5,6] была использована нечеткая линейная регрессионная модель. Для анализа выборки в такой модели необходимо использовать чебышевскую норму  $\alpha_\infty(\Omega_0)$ .

*Работа выполнена при поддержке РФФИ (гранты: № 16–01–00336А, № 16–31–00048мол\_а).*

### 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, №1. – P. 15–18.
3. Andrews D.F., Pregibin D. Finding the outliers that matter // Journal of the Royal Statistical Society. – 1978. – Vol. 40. – P. 84–93.
4. Ponomarev I.V., Slavsky V.V. Uniformly fuzzy model of linear regression // Journal of Mathematical Sciences. – 2012. – Vol. 186. – Issue 3. – P. 478 – 494.

5. 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.

6. 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 004.4

### Разработка дополнительных модулей для информационной системы учета археологических находок

*А.С. Шевченко, Н.Н. Камышиникова, В.В. Костенко,  
К.А. Гумеров*

*РИ (филиал АлтГУ), г. Рубцовск*

Описание археологических находок является сложной технологической задачей. В течение длительного времени на практике используется технология описания вручную, которая предполагает сбор и классификацию исходной информации, составление вспомогательных таблиц, непосредственное составление описания находок, его проверку и корректировку. При такой технологии трудно учитывать огромное количество археологических находок, проверять достоверность их описания и выполнять поиск нужной находки в хранилище или музее. Поэтому проектирование и создание системы автоматизации учета археологических находок является актуальной проблемой.

На сегодняшний день большинство экспедиций использует табличный процессор MS Excel для учета археологических находок. У такого подхода есть ряд недостатков: дискомфорт в использовании, сложность визуального восприятия информации, появление ошибок при вводе информации, отсутствие возможности создания различных отчетов по определенным археологическим находкам.

Кроме Microsoft Excel создавались и специализированные прикладные программы. Примерами таких программ, являются программы «Полевая Опись» и «Autodesk Land Desktop». Но практика показала, что данные программы не полностью удовлетворяют потребностям археологов, требуют больших вычислительных мощностей от аппаратной части или же являются очень дорогостоящими решениями.

Нами в качестве среды разработки было решено использовать Visual Studio Community [1]. Данная среда была выбрана, потому что