

УДК 51

**АЛГОРИТМ ВЫЯВЛЕНИЯ АНОМАЛЬНЫХ ЗНАЧЕНИЙ
КРИТЕРИЕМ ГРАББСА
В WOLFRAM MATHEMATICA**

**Гундина М. А., к.ф.-м.н, доцент,
Каменко Д. А., магистрант**

*Белорусский национальный технический университет
Минск, Республика Беларусь*

Аннотация: рассматриваются проблемы выявления аномальных значений выборки. К ним относятся такие единицы статистической совокупности, у которых значения анализируемого признака существенно отклоняются от основного массива данных. Приводится реализация одного из статистических алгоритмов – алгоритма метода Граббса. Метод реализован в компьютерной системе Wolfram Mathematica.

Ключевые слова: аномальное значение, выборка, критерий Граббса, распределение Стьюдента.

**ALGORITHM FOR DETECTING ANOMAL VALUES BY THE
GRUBBS CRITERION IN WOLFRAM MATHEMATICA**

**Hundzina M. A., Ph. d, Associate professor,
Kamenka D. A., undergraduate**

*Belarusian National Technical University
Minsk, Republic of Belarus*

Summary: the problems of detecting anomalous sample values are considered. They are such units of the statistical set, in which the values of the analyzed attribute deviate significantly from the main data array. An implementation of one of the statistical algorithms, the algorithm of the Grubbs method, is given. The method is implemented in the Wolfram Mathematica computer system.

Key words: anomalous value, sample, Grubbs' test, Student's distribution.

Единицы статистической совокупности, у которых значения анализируемого признака существенно отклоняются от основного массива, называются аномальными значениями или выбросами. Такие результаты могут появляться по причине сбоя в работе оборудования при измерениях и регистрации данных. Резкие отклонения условий наблюдений также могут привести к возникновению аномальных значений.

Игнорирование наличия аномальных значений в выборке может привести к недостоверным результатам при оценивании и контроле соответствия характеристик системы предъявляемым требованиям. Модель, построенная на основе этих данных, будет недостоверной моделью. Необходимо выявлять и устранять аномальные результаты измерений.

Одним из известных критериев, который может быть использован для определения аномальных значений, является критерий Граббса [1, 2]. Он основан на предположении о том, что исходная выборка значений нормально распределена.

Использование критерия Граббса позволяет за одну итерацию определить одно аномальное значение. Затем его необходимо удалить из списка значений. Повторяя эту процедуру, можно определить все аномальные значения за конечное число итераций.

Критерий Граббса состоит из проверки следующих гипотез:

1. H_0 : В наборе данных нет выбросов.
2. H_1 : В наборе данных присутствует как минимум один выброс.

Критерий Граббса рассчитывается как:

$$T_k = \frac{|x - \bar{x}|}{s}, \quad (1)$$

где \bar{x} – означает выборочное среднее;

s – среднеквадратичное отклонение.

Значение критерия Граббса показывает максимальное абсолютное отклонение от выборочного среднего в единицах среднеквадратичного отклонения.

Для двухстороннего теста гипотеза об отсутствии выбросов отклоняется с уровнем значимости α , если:

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t_{\alpha/(2N), N-2}^2}{N-2 + t_{\alpha/(2N), N-2}^2}}, \quad (2)$$

где $t_{\alpha/(2N), N-2}$ – максимальное критическое значение распределения Стьюдента с $N - 2$ степенями свободы и уровнем значимости $\alpha/(2N)$.

Рассмотрим исходный массив значений:

$$v = \{\{1,7.6\}, \{2,6.5\}, \{3,6\}, \{4,8\}, \{5,6\}, \{6,55\}, \{7,7\}\}.$$

Сформируем вектор значений T .

```
For [s = {}];
I = 1, i <= Length[v], I + +, s = Append [s, Abs [v [[i, 2]] - Mean [v [[All,
2]]]] / StandardDeviation [v [[All, 2]]]]].
```

Для двухстороннего критерия определим значение критического G :

```
Needs ["HypothesisTesting"]
g [n, alfa]: = (n - 1) / n ^ (1 / 2)
((StudentTPValue [alfa / (2 n), n - 2] [[2]]) ^ 2 / (n - 2 + (StudentTPValue [alfa / (2 n), n - 2] [[2]]) ^ 2)) ^ (1 / 2).
```

Выведем значение на экран для уровня значимости 5%.

```
g [Length [v], 0.05].
```

Для данной выборки эта величина равна 0.493593.

Массив T принимает значения:

```
{0.336459, 0.396849, 0.4243, 0.314499, 0.4243, 2.26581, 0.369399}.
```

Сравниваем каждое из них с критическим значением и аномальные значения вносим в массив $s1$.

```

For [s1 = {}];
I = 1, I <= Length [v], I + +, If [s [[i]] > g [Length [v], 0.05],
sI = Append [sI, v[[i]]]].

```

Удаляем аномальное значение из выборки

```
v2 = Complement [v, sI].
```

Заметим, что для одностороннего критерия критическое значение G находится следующим образом:

$$g_2 [n, \alpha] = (n - 1)/n^{1/2} ((\text{StudentTPValue}[\alpha/(n), n - 2] [[2]])^{1/2} / (n - 2 + (\text{StudentTPValue} [\alpha / (n), n - 2] [[2]])^{1/2}))^{1/2}.$$

Список использованных источников

1. Aslam M. Introducing Grubbs's test for Detecting Outliers under Neutrosophic Statistics- an Application to Medical Data / M. Aslam // Journal of King Saud University. – 2020. – №. 32. – P. 2696–2700.
2. Zeller, C. Influence diagnostics for Grubbs's model with asymmetric heavy-tailed distributions / C. Zeller, V. Lachos, F. Labra // Statistical Papers. – 2014. – № 55. – P. 671–690.