Процесс выявления аномалий является очень важным вопросов в задачах предсказания поломок оборудования, выявления аномального спроса на потребляемую продукцию, выявления нестандартного поведения информационно-измерительной системы. В компьютерной системе Mathematica имеются встроенные функции анализа аномальных значений выборки, также существует возможность реализации пользовательских алгоритмов для обрабатывать массивы большого объема.

Литература

1. Шкодырев, В. П. Обзор методов обнаружения аномалий в потоках данных / В. П. Шкодырев // Second Conference on Software Engineering and Information Management. – 2017. – С. 50.

2. Большая советская энциклопедия: [В 30 т.] ; гл. ред. А. М. Прохоров. – 3-е. изд. – М.: Сов. энцикл., 1969–1978.

3. Вайнер, Э. Н. Краткий энциклопедический словарь: Адаптивная физическая культура : учебное пособие для студентов, обучающихся по специальностям "Адаптивная физическая культура" и " Физическая культура" / Э. Н. Вайнер, С. А. Кастюнин. – Москва : Флинта, 2003. – 144 с.

4. Коджаспирова, Г. М. Для студ. высш. и сред. пед. учеб. заведений // Г. М. Коджаспирова, А. Ю. Коджаспиров. – М.: Издательский центр «Академия», 2000. – 176 с.

5. Дрешин, М. Г. Норма и аномалия в социальном развитии :дис. ... канд. Филос. наук : 09.00.11 / М. Г. Дрешин. – Ростов-на-Дону, 2010. – 151 с.

УДК 004+51

ОПРЕДЕЛЕНИЕ АНОМАЛЬНЫХ ЗНАЧЕНИЙ С ПОМОЩЬЮ РАССТОЯНИЯ МАХАЛАНОБИСА

Гундина М.А.[1], Жданович М.Н.[2], Каменко Д.А.[1]

[1]Белорусский национальный технический университет
[2]«Отраслевая лаборатория новых технологий и материалов»
ОАО "ИНТЕГРАЛ" – управляющая компания холдинга "ИНТЕГРАЛ"
Минск, Республика Беларусь

Аннотация. В данной статье рассматривается реализация алгоритма выявления аномальных значений выборки с помощью расстояния Махаланобиса, реализованных в компьютерной системе Wolfram Mathematica.

Ключевые слова: аномальное значение, выборка, расстояние Махаланобиса, компьютерная система Wolfram Mathematica.

DETERMINATION OF ANOMALOUS SAMPLE VALUES USING THE MAHALANOBIS DISTANCE

Hundzina M.[1], Zhdanovich M.[2], Kamenka D.[1]

[1]Belarusian National Technical University
[2]«Branch laboratory of technologies and materials»
JSC "INTEGRAL" – the management company of the holding "INTEGRAL"
Minsk, Republic of Belarus

Abstract. This article discusses the implementation of some statistical algorithms for detecting anomalous sample values by Mahalanobis distance, implemented in a computer system Wolfram Mathematica.

Key words: anomalous value, sample, Mahaoblanis distance, computer system.

Адрес для переписки: Гундина М.А., пр. Независимости, 65, Минск 220013, Республика Беларусь
e-mail: hundzina@bntu.by

Известно, что единицы статистической совокупности, у которых значения анализируемого признака существенно отклоняются от основного набора данных, называются аномальными значениями.

Причины возникновения аномальных значений могут быть разной природы: сбои при измерениях и регистрации данных, резкие отклонения условий наблюдений и др. Наличие аномальных результатов может привести к недостоверным результатам при оценивании и контроле соответствия характеристик системы предъявляемым требованиям. Поэтому необходимо выявлять и устранять аномальные результаты измерений [1].

Степень аномальности значения может определяться по значению расстояния Махалонобиса. Эта величина в математической статистике является мерой расстояния между векторами случайных величин, обобщает понятие евклидова расстояния.

В задаче определения принадлежности точки одному из классов необходимо найти матрицы ковариации всех классов, что, обычно, делается на основе известных выборок из каждого класса.

При вычислении расстояния Махаланобиса до каждого класса выбирается тот класс, для которого это расстояние оказалось минимальным, что эквивалентно методу максимального правдоподобия.

Примем в рассмотрение предположение о нормальном законе распределения исходной выборки.

Для векторизации вычисления расстояний Расстояние Махаланобиса между двумя точками – это мера расстояния между двумя случайными точками U и V, одна из которых может принадлежать некоторому классу с матрицей ковариации COV:

$$d_m(U,V,COV) = \sqrt{(U-V)COV^{-1}(U-V)^T},$$

где символ T обозначает операцию транспонирования, а под COV^{-1} подразумевается матрица, обратная ковариационной матрице.

Элементы ковариационной матрицы вычисляются следующим образом:

$$cov_{a,b} = \frac{1}{|C|-1}\sum_{x\in C}(X_1-\mu_1)(X_2-\mu_2), \quad (1)$$

где μ_1, μ_2 – математические ожидания по признакам, $|C|$ – количество точек в классе.

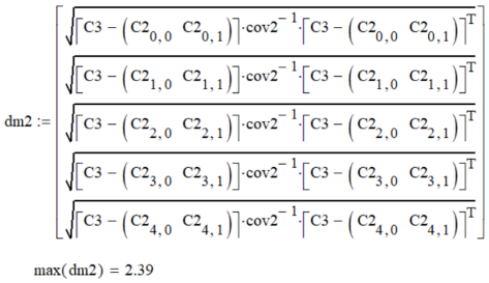

Рисунок 1 – Значение расстояния от фиксированной точки C3 до некоторого множества точек C2

Расстояние Махаланобиса широко применяется в задачах кластеризации и классификации в задачах определения соответствия точки известному классу. Оно отличается от расстояния Евклида тем, что учитывает корреляции между переменными и инвариантно масштабу [1, 2].

Точка, имеющая наибольшее расстояние Махаланобиса до остального множества точек, считается аномалией. Такая точка имеет наибольшее влияние на кривизну и на коэффициенты уравнения регрессии. Также расстояние Махаланобиса может быть использовано в задаче определения многомерных выбросов.

В системе Wolfram Mathematica определяем расстояние следующим образом [3]:

```
Dm=Compile[{{u,_Real,1},{\[Mu],_Real,1},
{s,_Real,2}},First@\
[Sqrt]((u-\[Mu]).Inverse[s].
Transpose[{u-\[Mu]}]),CompilationOptions-
>{"ExpressionOptimization"->True},
RuntimeOptions->"Quality",RuntimeAttributes-
>Listable,CompilationTarget->"C"]
```

Пусть исходная выборка имеет вид:

```
cohort={{5.04,14.22},{5.50,5.83},{5.19,4.61},
{4.78,4.12},{5.08,5.99},{4.29,4.18},{5.08,6.90}};
\[Mu]=Mean@cohort;s=Covariance@cohort;
```

На экран характеристики выборки могут быть выведены с помощью следующей команды:

```
Print["\[Mu] = ",{\[Mu]}\[Transpose]
//MatrixForm," S = ",s//MatrixForm];
```

$$\mu = \begin{pmatrix} 4.99865 \\ 6.554 \end{pmatrix} \quad S = \begin{pmatrix} 0.141418 & 0.289131 \\ 0.289131 & 12.5033 \end{pmatrix}$$

Рисунок 2 – Значения характеристик выборки

points={\[Mu]}~Join~cohort

Функция ListPlot позволяет представить точки в декартовой системе координат, дополнительно указать значения расстояния Махаланобиса для каждой точки выборки:

```
ListPlot[{cohort,Labeled[#,Round[Dm[#,
\[Mu],s],.01]]&/@points},PlotRange-
>All,AspectRatio->1,PlotStyle-
>{Darker@LightBlue,{Red,PointSize[.01]}}].
```

Результат работы алгоритма представлен на рис.3.

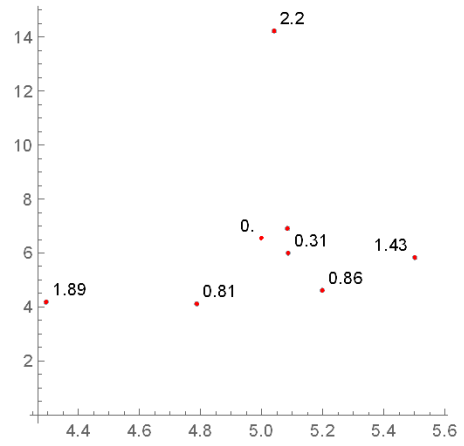

Рисунок 3 – Графическое представление исходной выборки

Анализируя полученные значения расстояний, было выявленное одно аномальное значение, величина расстояния для этого значения значительно превышает эти же расстояния для других значений выборки.

Литература

1. Лукин, В. Л. Статистический метод выявления аномальных результатов измерений характеристик технических систем / В. Л. Лукин, Б. И. Сухорученков, В. И. Кузнецов // Двойные технологии. – 2010. – № 2 (51). – С. 32–40.

2. Демяненко, Я. М. Компьютерное зрение и обработка изображений. Лекция 12. Детекторы и дескрипторы / Южнофедеральный университет, 2019 [Электронный ресурс] – Режим доступа: edu.mmcs.sfedu.ru. – Дата доступа. – 1.02.2022.

3. Гороховатский, В. А. Структурный анализ и интеллектуальная обработка данных в компьютерном зрении / В. А. Гороховатский. – Харьков: Компания СМИТ, 2014. – 316 с.

4. Жданович, М. Н. Обработка изображений. Расстояние Махаланобиса / М. Н. Жданович, М. А. Гундина // Новые направления развития приборостроения : материалы 15-й Междунар. научн.-технич. конфер. мол. учен. и студ., Минск, 20−22 апреля 2022 г. / Белорусский национальный технический университет ; редкол.: О. К. Гусев (пред. редкол.) [и др.]. – Минск : БНТУ, 2022. – С. 205–206.

УДК 51

РЕАЛИЗАЦИЯ СТАТИСТИЧЕСКИХ АЛГОРИТМОВ ОПРЕДЕЛЕНИЯ АНОМАЛЬНЫХ ЗНАЧЕНИЙ ВЫБОРКИ В WOLFRAM MATHEMATICA

Гундина М.А., Кондратьева Н.А., Каменко Д.А.

Белорусский национальный технический университет
Минск, Республика Беларусь

Аннотация. В данной статье рассматривается реализация некоторых статистических алгоритмов выявления аномальных значений выборки, реализованных в компьютерной системе WolframMathematica.
Ключевые слова: аномальное значение, выборка, компьютерная система Wolfram Mathematica.

IMPLEMENTATION OF STATISTICAL ALGORITHMS FOR DETERMINATION ANOMALOUS SAMPLE VALUES IN WOLFRAM MATHEMATICA

Hundzina M., Kondratyeva N., Kamenko D.

Belarusian National Technical University
Minsk, Republic of Belarus

Abstract. This article discusses the implementation of some statistical algorithms for detecting anomalous sample values, implemented in a computer system Wolfram Mathematica.
Key words: anomalous value, sample, computer system.

Адрес для переписки: Гундина М.А., пр. Независимости, 65, Минск 220013, Республика Беларусь
e-mail: hundzina@bntu.by

Задача автоматизации процесса выявления аномальных значений выборки решалась в инженерной практике и математической статистике и не теряет свою актуальность несколько десятилетий [1–4]. Известно, что аномальные значения способны существенно исказить функционирование математических моделей анализа данных, что может привести к снижению надежности и некорректной работе всей системы [5, 6].

Единицы статистической совокупности, у которых значения анализируемого признака существенно отклоняются от основного массива, называются аномальными явлениями, «грубыми ошибками» или выбросами. Возможны и аномальные результаты, обусловленные сбоями при измерениях и регистрации данных, резкими отклонениями условий наблюдений, нештатной работой оборудования, ошибками операторов и др. Наличие аномальных результатов может привести к недостоверным результатам при оценивании и контроле соответствия характеристик системы предъявляемым требованиям. Поэтому необходимо выявлять и устранять аномальные результаты измерений [7].

Аномальные значения определяются исходя из их характеристики по отношению ко всей совокупности. Эти значения можно разделить на три группы:

1) значения, отражающие объективное развитие процесса, но сильно отклоняющиеся от общей тенденции;

2) значения, возникающие вследствие изменения методики расчетов;

3) значения, возникающие из-за ошибок при измерении показателя, при записи, передаче информации и т.д.

Поэтому процесс выявления и затем удаления этих значений состоит из нескольких этапов [8]. Вначале выявляются значения, которые выходят за границы интервала возможного варьирования характеристики признака.

Исходя из физического смысла исследуемой величины, могут рассматриваться как аномальные значения те, которые не соответствуют монотонному характеру изменения величины при последовательных наблюдениях, а также значения, приращения которых превышают предельно возможную скорость изменения величины.

Одной из групп методов выявления аномальных значений является группа статистических методов [2, 4].

Реализация метода поиска аномалий с помощью правила сигм. С помощью этого метода можно осуществлять контроль за нахождением параметра в допустимых границах, что удобно в производственных процессах. Анализ выбросов в