

# ПРОВЕРКА СОГЛАСОВАННОСТИ АТРИБУТОВ НАБОРА ДАННЫХ С НОРМАЛЬНЫМ ЗАКОНОМ РАСПРЕДЕЛЕНИЯ

*А.Ю. Кайда, ассистент ОИТ,  
И.А. Рыбаченко, студент гр. 8ПМ011  
Томский политехнический университет  
E-mail: iar12@tpu.ru*

## Введение

Знание закона распределения случайной величины позволяет использовать более специализированные методы статистического анализа. Например, t-тест Стьюдента применим только при соответствии частотного распределения данных в каждой из сравниваемых групп закону нормального распределения [1].

Целью работы является проверка атрибутов набора данных на согласие с нормальным законом распределения. В данной работе из всех атрибутов набора данных рассматриваются только показания акселерометра.

## Постановка задачи

Для каждого атрибута  $i$  из набора атрибутов [аксX, аксY, аксZ] ставится отдельная и независимая задача проверки согласия величины с нормальным законом распределения.

Нулевая гипотеза – величина атрибута  $i$  имеет нормальный характер распределения.

Альтернативная гипотеза – характер распределения величины атрибута  $i$  отличен от нормального.

При проверке гипотезы уровень значимости  $\alpha$  принимается равным 0,05.

## Гистограмма плотности распределения

Базовым методом проверки величины на согласие нормальному закону распределения является визуальная оценка гистограммы плотности распределения. Построенные для показаний акселерометра гистограммы плотности распределения представлены на рисунке 1.

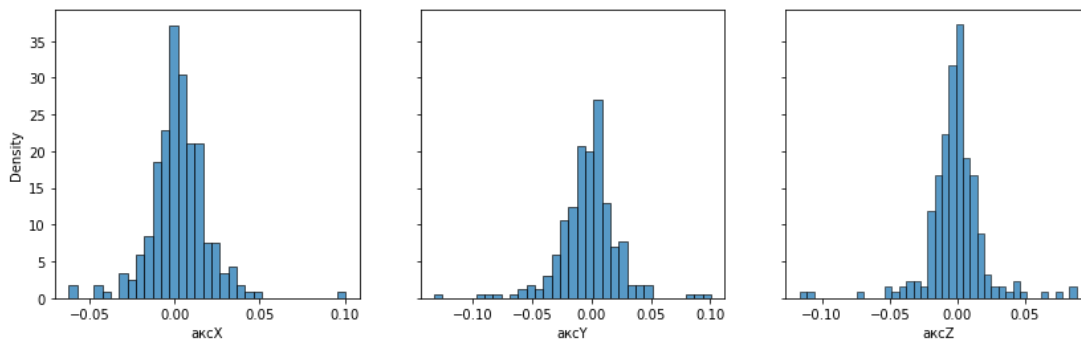


Рис. 1. Гистограммы плотности распределения показаний акселерометра.

Полученные гистограммы содержат характерный признак нормального распределения – имеют форму перевёрнутого колокола, что позволяет предположить согласие величин нормальному закону распределения [2].

Метод визуального анализа позволяет быстро оценить характер распределения, однако производимая оценка является субъективной и, следовательно, плохо воспроизводимой и не надёжной.

## Критерий Лиллиефорса

Существует большое количество формальных критериев согласия. Одним из них является критерий Лиллиефорса. Критерий Лиллиефорса также, как и критерий Колмогорова-Смирнова, вычисляется как максимальная разность функции распределения величины и функции нормального распределения. Величина считается согласной нормальному закону распределения, если значение критерия для неё меньше критического, определяемого из таблицы критических значений [3].

Результат расчёта критерия Лиллиефорса для показаний акселерометра приведён в таблице 1.

Таблица 1. Значения критерия Лиллиефорса для показаний акселерометра

Атрибут	Критерий Лиллиефорса	P-value
аксХ	0.098	0.019
аксУ	0.088	0.047
аксZ	0.121	0.002

Критическое значение для заданного уровня значимости и данного количества строк в наборе данных равно 0,057. Поскольку полученное значение критерия Лиллиефорса больше критического, нулевую гипотезу стоит отвергнуть.

Критерий Лиллиефорса объективен, но не даёт представления о причинах несогласия величины с нормальным законом распределения.

### График квантиль-квантиль

Ещё одним способом проверки характера распределения величины является построение графика квантиль-квантиль. В случае, если сравниваемые величины имеют одинаковый закон распределения, график квантиль-квантиль принимает форму прямой, идущей под углом 45 градусов к координатным осям [4].

На рисунке 2 представлены построенные графики квантиль-квантиль для показаний акселерометра.

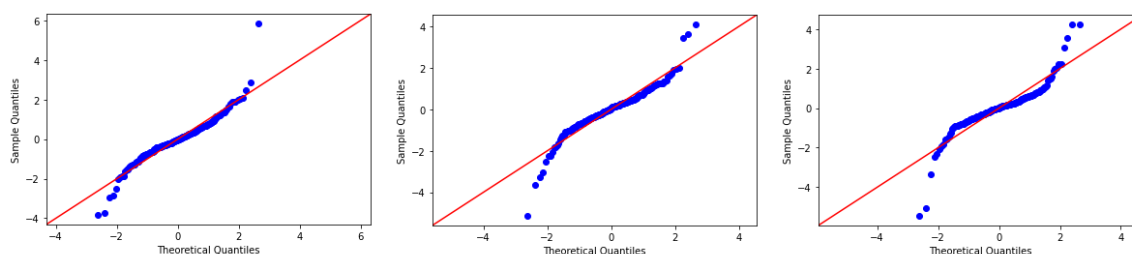


Рис. 2. Графики квантиль-квантиль для показаний акселерометра.

Полученные графики не соответствуют прямой линии, имеющей угол наклона 45 градусов.

Исходя из характера изгиба краёв прямой, образованной графиком можно сделать вывод, что исследуемая величина имеет более высокую вероятность появления больших по модулю значений, чем соответствующее нормальное распределение.

### Заключение

В результате работы была выполнена проверка показаний акселерометра на соответствие нормальному закону распределения. Два из трёх применённых методов показали, что гипотезу о нормальном распределении величины стоит отвергнуть. Таким образом, для дальнейшего статистического анализа стоит применять непараметрические методы.

### Список использованных источников

1. Проверка нормальности распределения значений признака [Электронный ресурс]. – Режим доступа: <http://vnauke.by/schkola/Osnovy-dokazatelnoj-mediciny-Biomedicinskaja-statistika/Proverka-normalnosti-raspredeleniya-znachenij-priznaka>, свободный (Дата обращения 30.01.2021)
2. Типы распределений и соответствующие им гистограммы [Электронный ресурс]. – Режим доступа: <http://sixsigmaonline.ru/baza-znaniy/22-1-0-229>, свободный (Дата обращения 30.01.2021)
3. Критерий Лиллиефорса [Электронный ресурс]. – Режим доступа: [http://wiki-org.ru/wiki/Критерий\\_Лиллиефорса](http://wiki-org.ru/wiki/Критерий_Лиллиефорса), свободный (Дата обращения 30.01.2021)
4. Нормальный график КК и общий график КК – Режим доступа: <https://desktop.arcgis.com/ru/arcmap/latest/extensions/geostatistical-analyst/normal-qq-plot-and-general-qq-plot.htm>, свободный (Дата обращения 30.01.2021)