



Редактирование выбросов

Описание

Компонент предназначен для автоматической корректировки выбросов и экстремальных значений в наборах данных. Для каждого поля исходного набора данных критерии определения выбросов и экстремальных значений задаются пользователем с помощью указания допустимого стандартного отклонения или интерквартильного размаха. Под выбросами при этом подразумеваются значения данных, существенно отклоняющиеся от средних, а под экстремальными — значения, которые настолько сильно отклоняются от типичных значений, что перестают соответствовать логике исследуемых процессов и явлений.

Порты

Вход

-  Входной источник данных — таблица данных.

Выход

-  Выходной набор — исходная таблица после обработки.
-  Выбросы — таблица, содержащая строки исходной таблицы, в которых были обнаружены выбросы.
-  Экстремальные значения — таблица, содержащая строки исходной таблицы, в которых были обнаружены экстремальные значения.

Мастер настройки

- **Исходные данные упорядочены** — выставить данный флаг следует в том случае, когда известно, что числовой ряд является упорядоченным, т. е. таким, значения которого упорядочены по возрастанию или убыванию (например, по дате или времени). Доступность некоторых методов обработки зависит от состояния данного флага. На обработку логических и строковых полей положение флага не влияет.
- **Область настройки методов обработки выбросов и экстремальных значений** — содержит список полей доступных для обработки. Для каждого поля можно выставить флаг, определяющий

наличие обработки, и затем, выделив поле, задать метод редактирования.

- **Определение выбросов и экстремальных значений** — доступны два метода выявления:
 - Стандартное отклонение — критерием является отклонение значения признака от среднего более, чем на заданное число стандартных отклонений. При этом данный параметр отдельно задается для выбросов и для экстремальных значений. Данный метод следует применять, если известно, что распределение данных близко к нормальному.
 - Интерквартильный размах — критерием является расстояние между 1-м и 3-м квартилями распределения значений признака. Если значение признака отклоняется от медианы более, чем на заданное число интерквартильного размаха, то оно считается аномальным. Данный параметр задается отдельно для выбросов и экстремальных значений. Этот метод можно применять и в случае, когда распределение данных отличается от нормального.

Как для выбросов, так и для экстремальных значений доступны следующие методы редактирования:

- оставить без изменения;
- удалять записи — исключить строки с аномальными значениями из набора данных;
- заменять средним — заменять аномалии средним значением столбца;
- заменять медианой — заменять аномалии медианой, вычисленной по столбцу;
- заменять наиболее вероятным — замена аномалий наиболее вероятным значением по столбцу, замена производится на среднее значение из наиболее вероятного интервала, число интервалов варьируется в зависимости от объема выборки — чем она больше, тем больше интервалов;
- заменять заданным значением — замена аномалий на значение, прописанное вручную;
- ограничивать — аномалии будут заменены значением границы, с которой начинается определение аномалии.

Для каждого поля спектр доступных методов определяется тремя характеристиками данных одновременно (см. данные):

- упорядоченностью;
- типом;
- видом.

Таблица применимости по этим характеристикам:

Метод	Неупорядоченный набор		Упорядоченный набор	
	 Дискретный	 Непрерывный	 Дискретный	 Непрерывный
Оставить без изменения	0/1 31 9.0 12 ab	31 9.0 12	0/1 31 9.0 12 ab	31 9.0 12
Удалять записи	0/1 31 9.0 12 ab	31 9.0 12		
Заменять средним	31	31 9.0 12	31	31 9.0 12
Заменять медианой	31 9.0 12	31 9.0 12	31 9.0 12	31 9.0 12

Заменять наиболее вероятным	$\begin{matrix} 0/1 & \overline{31} & 9.0 & 12 \\ & ab \end{matrix}$	$\begin{matrix} \overline{31} & 9.0 & 12 \end{matrix}$	$\begin{matrix} 0/1 & \overline{31} & 9.0 & 12 \\ & ab \end{matrix}$	$\begin{matrix} \overline{31} & 9.0 & 12 \end{matrix}$
Заменять заданным значением	$\begin{matrix} 0/1 & \overline{31} & 9.0 & 12 \\ & ab \end{matrix}$	$\begin{matrix} \overline{31} & 9.0 & 12 \end{matrix}$	$\begin{matrix} 0/1 & \overline{31} & 9.0 & 12 \\ & ab \end{matrix}$	
Ограничивать	$\begin{matrix} 0/1 & \overline{31} & 9.0 & 12 \\ & ab \end{matrix}$	$\begin{matrix} \overline{31} & 9.0 & 12 \end{matrix}$	$\begin{matrix} 0/1 & \overline{31} & 9.0 & 12 \\ & ab \end{matrix}$	$\begin{matrix} \overline{31} & 9.0 & 12 \end{matrix}$

Смотри также:

- [Заполнение пропусков](#)
- [Аномальные значения](#)