



Квантование

Описание

Компонент разбивает диапазон значений выбранных полей исходного набора на конечное число интервалов. Для разбиения возможно применение различных алгоритмов (см. далее методы квантования), а также использование внешних таблиц с заданными интервалами квантования. Квантование применяется к данным с типами: целый, вещественный и дата/время (см. типы данных).

Моделью считаются интервалы квантования, а если они не были рассчитаны, то *моделью* являются настройки метода квантования.

В процессе переобучения сначала формируются интервалы квантования, которые становятся *моделью* узла Квантование. Узел считается обученным сразу после настройки.

Вход

-  Входной источник данных (таблица данных).
-  Добавить еще один порт. Внешние диапазоны квантования (таблица данных).

Выход

-  Выходной набор данных (таблица данных). Описание структуры результирующего набора.
-  Диапазоны для квантования (таблица данных). Описание структуры таблицы диапазонов.

Мастер настройки

Мастер настройки состоит из двух основных областей: область настройки параметров квантования и область отображения результатов квантования. Обе области организованы в виде таблицы. Над ними располагается строка состояния входа.

Область настройки параметров квантования

Область представлена в виде таблицы. Над полями расположены три кнопки:

-  Редактировать — при нажатии позволяет редактировать параметры квантования для выбранного поля.

-  Уменьшить разрядность — каждое нажатие кнопки уменьшит разрядность границ интервалов на один знак после запятой.
-  Увеличить разрядность — каждое нажатие кнопки увеличит разрядность границ интервалов на один знак после запятой.

Таблица этой области состоит из нескольких столбцов:

- **Поле** — содержит поля исходного набора данных, к которым применима процедура квантования. Это поля типа: целый, вещественный, дата/время.
- **Метод** — поле представлено раскрывающимся списком, из которого необходимо выбрать метод квантования:
 - Ширина — пользователь может выбирать ширину интервала, а количество интервалов рассчитывается автоматически, как отношение разности верхней и нижней границ к заданной ширине. Выставив соответствующие флаги, можно задать:
 - Верхнюю границу — верхняя граница самого высокого интервала.
 - Нижнюю границу — нижняя граница самого низкого интервала.
 - Количество — выбирается количество интервалов, а ширина рассчитывается автоматически, как отношение разности верхней и нижней границ к заданному количеству интервалов. Для этого метода так же можно задать верхнюю и нижнюю границы.
 - Плитка — пользователь выбирает количество интервалов, а компонент задает диапазоны интервалов таким образом, чтобы в каждом интервале было примерно одинаковое количество значений.

Если включить флаг *Из сумм значений*, то размер плитки сформируется из сумм значений в интервале.

Имеется несколько способов обработки совпадающих значений:

- Добавлять в следующий — перенесет значения совпадающих наблюдений в следующий (более высокий) интервал разделения.
- Сохранять в текущем — сохраняет значения совпадающих наблюдений в текущем (более низком) интервале разделения. Этот метод может привести к тому, что всего будет создано меньше интервалов.
- Назначать случайно — типы границ интервалов будут определены случайно; возможно включение одинаковых значений в тот или иной интервал случайным образом.
- Оставить как есть — границы всех интервалов будут иметь тип $\geq$, и возможна ситуация, когда совпадающие значения окажутся в разных интервалах.
- Одинаковые плитки — достижение равного количества значений в интервалах обеспечивается не только подбором диапазонов интервалов, но и подбором типов границ для каждого интервала ($>$ или $\geq$).
- Коэффициенты СКО — разбивает значения на интервалы в зависимости от выбранного диапазона, выраженного в количестве σ (СКО).
- Внешние диапазоны.

Для всех методов квантования можно установить флаг:

- Округлять границы - округляет границы до ближайшего целого.
- Нижняя граница открыта — убирает нижнюю границу.

- Верхняя граница открыта — убирает верхнюю границу.
- **Автоматически** — установленная галочка в этом поле обеспечивает автоматическую настройку параметров квантования выбранного метода.
- **Интервалов** — количество интервалов, на которые будут разбиты значения поля.
- **Минимум** — отображается минимальное значение квантуемого поля.
- **Максимум** — отображается максимальное значение квантуемого поля.

Далее в каждой строке располагается кнопка  «рассчитать интервалы» и в шапке таблицы  «рассчитать все интервалы». При их нажатии пересчитываются параметры квантования (количество интервалов, минимум, максимум) в зависимости от изменения методов и/или настроек параметров. Этот функционал доступен только при состоянии «Вход активирован».

Область отображения результатов квантования

В этой области отображаются результаты квантования с возможностью их редактирования. Над полями таблицы расположены несколько элементов управления:

-  **Инвертировать тип** — меняет тип границ.
-  — пересчитывает гистограмму согласно новым параметрам.
- **Шаблон** — в этом поле происходит настройка шаблона для отображения метки интервала, в нем можно составить пользовательский шаблон или при нажатии на  выбрать один из готовых шаблонов. Чтобы применить шаблон необходимо нажать кнопку .
- **Образец** — при клике на эту кнопку открывается таблица обозначений, которые можно использовать при составлении шаблона.

Под элементами управления расположена таблица с результатами квантования выделенного поля, она содержит следующие поля:

- **№** — номер интервала.
- **Нижняя** — нижняя граница интервала.
- **Тип** — тип границы.
- **Верхняя** — верхняя граница интервала.
- **Метка** — метка интервала (ее можно задавать шаблоном).
- **Объём** — отображает объём значений, попавших в интервал (отображается в виде гистограммы).

Важно: При изменении границ диапазонов квантования вручную гистограмма автоматически не пересчитывается. Чтобы увидеть обновленную (пересчитанную) гистограмму, нужно нажать на кнопку  «Пересчитать гистограмму». При большом объеме данных перестройка гистограммы займет некоторое время, если нет необходимости в ее просмотре, можно сразу перейти на следующий шаг мастера настройки.

Особенности выполнения узла

1. Если интервалы квантования для поля уже рассчитаны, то при выполнении узел будет разбивать данные по имеющимся интервалам.
 2. Если для поля не были рассчитаны интервалы квантования, то при выполнении они будут рассчитываться на основе данных и настроек поля. Данные будут разбиваться по только что сформированным интервалам.
-

Статьи в разделе:

- [Внешние диапазоны](#)
- [Структура результирующего набора](#)
- [Параметры диапазонов квантования](#)