


# Набор данных

AnyLogic позволяет сохранять данные по ходу работы модели с помощью специального элемента модели – **набора данных**. Набор данных представляет собой двумерный массив измерений (конечного размера). Каждое измерение, внесенное в набор данных, имеет два значения:  $x$  и  $y$  (типа `double`). Значение  $y$  обычно является значением наблюдаемой переменной модели.

В качестве  $X$ -значений набора данных вы можете использовать модельное время, т.е. связывать наблюдаемое значение с моментом времени, в который оно было измерено. Примерами таких величин могут быть длина очереди, координата машины, уровень жидкости и т.д. Такой набор данных называется *временным*.

Или же вы можете использовать в качестве  $X$ -значений набора данных другую наблюдаемую переменную, сохраняя таким образом данные о зависимости одного значения от другого - такие наборы данных называются *фазовыми*.

 Пожалуйста, обратите внимание, что добавление нового измерения в переполненный набор данных (который уже хранит максимально допустимое количество измерений) вызовет потерю самого старого измерения, хранимого в этом наборе данных, и если это значение было минимальным или максимальным, то вызовет заново процедуру поиска нового минимального/максимального значения, что может занять значительное время в случае наборов данных с большим количеством хранимых измерений. Поэтому в случае таких наборов данных рекомендуется задавать максимально допустимое количество измерений больше, чем фактическое количество измерений, которое будет добавлено в этот набор данных.

Набор данных не предоставляет средств для сбора статистической информации по собираемым данным (кроме как возможность узнать минимальное и максимальное из хранимых в данный момент значений (отдельно  $x$ - и  $y$ -значений)). Для этих целей используется объект [Статистика](#).

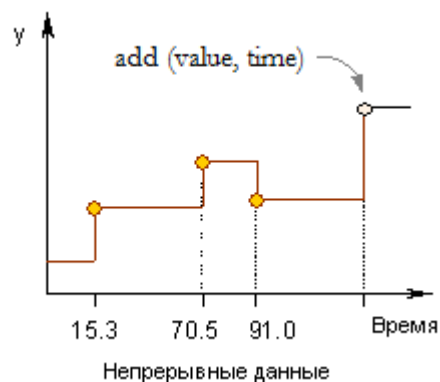
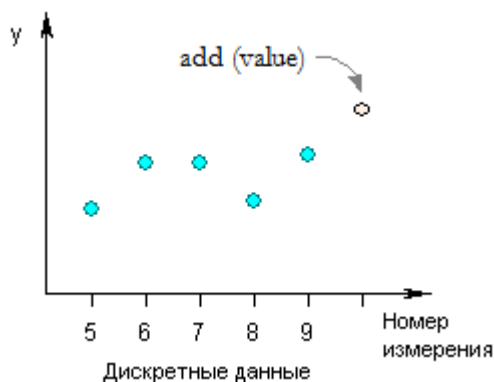
## Чтобы создать набор данных

1. Перетащите элемент **Набор данных**  из палитры **Статистика** в то место графического редактора, где вы хотите поместить набор данных.
2. Перейдите в панель **Свойства**.
3. Чтобы сделать набор данных временным, установите флажок **Использовать время в качестве значения по оси X**. Введите выражение, которое будет динамически вычисляться во время выполнения модели, чтобы получить новое значение набора данных, в поле **Значение по оси Y**.
4. Чтобы сделать набор данных фазовым, сбросьте этот флажок и задайте выражения, вычисляющие как  $X$ - , так и  $Y$ -значения, в полях **Значение по оси X** и **Значение по оси Y** соответственно.
5. Наборы данных могут иметь ограничение на максимальное количество последних измерений, которое будет храниться в наборе данных. Введите размер "хвоста" набора данных в поле **Хранить до ... последних измерений**.

# Статистика

Объект сбора статистики **Статистика** вычисляет основную статистическую информацию (среднее значение, минимум, максимум и т.д.) для последовательности измеренных значений (типа `double`).

Объект работает по-разному в зависимости от того, считаются ли собираемые данные непрерывными или дискретными.



## Непрерывные данные

Измерения имеют длительность во времени, то есть значение остается актуальным в течение какого-то времени, и изменяется только в дискретные моменты времени (например, длина очереди). В объект непрерывной статистики можно добавлять только измерения с возрастающими значениями временных меток. Методы вычисления среднего значения и т.д. подразумевают, что последнее добавленное значение было актуально до времени вызова метода.


Собранная объектом статистика в этом случае доступна посредством программного интерфейса класса [StatisticsContinuous](#).

## Дискретные данные

Измерения не имеют длительности во времени (например, цена продуктов или длительность пребывания пациентов). Они считаются изолированными, дискретными точками во времени, так что среднее значение является просто суммой измерений, поделенной на их общее количество.

Собранная объектом статистика в этом случае доступна посредством программного интерфейса класса [StatisticsDiscrete](#).

### Чтобы добавить объект сбора статистики

1. Перетащите элемент **Статистика**  из палитры **Статистика** в то место графического редактора, где вы хотите поместить набор данных.
2. Перейдите в панель **Свойства**.
3. Введите выражение, которое будет динамически вычисляться во время выполнения модели, и результат его вычисления будет добавляться в объект сбора статистики, в поле **Значение**.
4. Укажите, хотите ли вы, чтобы собираемые данные считались непрерывными или дискретными, выбрав опцию **Непрерывная** или **Дискретная**.

# Данные гистограммы

Объект **Данные гистограммы**:

- Выполняет обычный статистический анализ добавляемых значений (вычисляет среднее значение, минимум, максимум, дисперсию, средний доверительный интервал, и т.д.).
- Производит построение функции плотности распределения вероятности и интегральной функции распределения на фиксированном или автоматически выбирающемся наборе интервалов.
- Вычисляет верхний и нижний процентиля (или рискованные значения с заданным процентом) с отклонением, равным ширине интервала.


## Процентили

В том случае, если вычисляется функция распределения, вы можете также задать процентили. Вы можете задавать любые значения нижнего и верхнего процентилей, например, 25 и 75, что будет соответствовать нижней и верхней квартилям. В этом случае будут вычисляться значения, ниже которых лежат 25 % и 75 % значений соответственно.

Собранная статистика может быть отображена с помощью [Гистограммы](#).



### Чтобы создать объект сбора данных гистограммы

1. Перетащите элемент **Данные гистограммы**  с палитры **Статистика** на диаграмму типа агентов (или эксперимента).
2. Перейдите в панель **Свойства**.
3. В поле **Значение** введите выражение, которое будет динамически вычисляться во время выполнения модели, чтобы получить новое измерение, которое будет добавлено в объект сбора данных.
4. Если вы хотите, чтобы вычислялась интегральная функция распределения, установите флажок **Считать CDF**.
5. Если вы хотите, чтобы считались процентили, установите флажок **Вычислять процентили** и задайте значения нижнего и верхнего доверительных интервалов в полях **Нижний** и **Верхний** соответственно.
6. Теперь нужно задать интервалы гистограммы. Укажите, хотите ли вы явно задать интервалы, либо использовать возможность автоматического формирования интервалов самой гистограммой, в секции свойств **Диапазон значений**.
7. Если вы знаете предполагаемые минимальное и максимальное значения измеряемых данных, то выберите в секции **Диапазон значений** опцию **Фиксированный** и задайте интервалы гистограммы, статически задав границы диапазона значений гистограммы в полях **Минимум** и **Максимум**, а количество интервалов в поле **Кол-во интервалов**.
8. Если же вы не знаете точный диапазон значений, которые будут добавляться на гистограмму, то вам нужно будет использовать не фиксированные, а адаптивные интервалы. В этом случае гистограмма не будет требовать от вас задания диапазона значений (вам нужно будет только задать количество интервалов и начальную ширину интервала). Вместо этого, она автоматически будет изменять интервалы в соответствии с добавляемыми значениями (удваивать их до тех пор, пока все значения не попадут в новый диапазон). Для этого выберите опцию **Выбирается автоматически** и задайте **Начальный размер интервала**.

## Данные двумерной гистограммы

Собирает данные (плотность вероятности, интегральную функцию распределения и т.д.) для массива гистограмм, у каждой из которых есть определенный диапазон базовых (x) значений и диапазон значений - y значений. Когда измерение (x,y) добавляется в объект **Данные двумерной гистограммы**, то вначале производится поиск той конкретной гистограммы, к которой принадлежит это измерение (это зависит от x значения), а затем y значение добавляется в эту гистограмму. Плотность вероятности и интегральная функция распределения вычисляются для каждой отдельной гистограммы в массиве. Кроме того, этот объект сбора данных поддерживает возможность вычисления *вложений* - областей, содержащих заданный процент данных в каждой простой гистограмме.

Объект **Данные двумерной гистограммы** особенно полезен для анализа набора стохастических наборов данных, например, набора реализаций стохастического процесса во времени, полученными в результате нескольких "прогонов" модели.

Собранная статистика может быть отображена с помощью [Двумерной гистограммы](#).



### Чтобы создать объект сбора данных двумерной гистограммы

1. Перетащите элемент **Данные двумерной гистограммы** из палитры **Статистика** в то место графического редактора, где вы хотите его поместить.

## Выходное значение

Элемент **Выходное значение** используется для хранения одиночного значения скалярного типа данных и отображения его во время прогона модели. Примером подобных значений могут послужить важные количественные показатели (*среднее время в системе, прибыль, расходы, качество обслуживания клиентов и т.д.*)

Элемент **Выходное значение** используется прежде всего для отображения значения в секции **Выходные данные** экрана модели в AnyLogic Cloud. Чтобы добавить элемент **Выходное значение**, а также другие элементы палитры **Статистика**, с моделью в AnyLogic Cloud, воспользуйтесь редактором [Конфигурация запуска](#).

Выходное значение задается в виде произвольного Java выражения в поле **Значение** свойств этого элемента. Это выражение вычисляется в определенный момент времени, в конце простого эксперимента Simulation, а также при вызове функции update\_outputName() в модели.

Подобно элементу **Параметр**, элемент **Выходное значение** может быть представлен любым **примитивным типом** Java (**int, double, boolean** or **String**). В отличие от элемента **Параметр**, элемент **Выходное значение** не может содержать значения класса Object. Элементу также можно задать измеряемый тип (*Время, Интенсивность, Длина, Скорость, Ускорение, Площадь, Объем* или *Скорость потока*) и указать соответствующие единицы измерения. К примеру, если в параметре **Тип** выбрана **Длина**, то будут доступны следующие единицы измерения длины: *дюйм, миля, метр*, и т.д.

**Примечание:** Если выбран измеряемый тип для выходного значения, типу данных будет неявно задан тип double.

К примеру, в вашей пешеходной модели используется элемент **Сервис с очередями** и вы хотите знать текущую длину во время выполнения модели. Это можно реализовать, введя queueName.size() (вернет длину очереди) в поле параметра **Значение** элемента **Выходное значение**. Чтобы значение обновлялось каждый раз, когда пешеход становится в очередь или покидает ее, необходимо вызвать функцию update\_outputName() в полях **При вставлении в очередь** и **При выходе из очереди** секции свойств **Действия** блока **PedService**. Если значение должно обновляться с определенной периодичностью, необходимо:

- в модель добавить [Событие, происходящее по истечении таймута](#).
- в параметре **режим** выбрать опцию *Циклический*.
- поместить вызов функции update\_outputName() в поле **Действие** события.

Элемент **Выходное значение**, как и другие элементы палитры **Статистика**, могут быть добавлены в секцию **Выходные данные** редактора [Конфигурация запуска](#).

### Чтобы создать выходное значение

1. Перетащите элемент **Выходное значение** из палитры **Статистика** в графический редактор.
2. Переместитесь в панель **Свойства**.
3. Введите имя элемента в поле **Имя**. Это имя будет использоваться для обращения к элементу и получения к нему доступа.
4. Если ваше выходное значение определяет: *время, интенсивность, длину, ускорение, скорость* или *площадь*, необходимо это явно указать, выбрав соответствующий **Тип**. Затем выберите нужную единицу измерения из выпадающего списка **Единица измерения**.
5. В противном случае, выберите необходимый тип Java (**int, double, boolean, String**) из выпадающего списка **Тип**.

6. В поле **Значение** свойств этого элемента укажите значение в виде произвольного выражения. Это значение можно позже изменить во время выполнения модели, используя функцию `update_outputName(new value)`.
7. Наконец, в параметре **Вычисляется** выберите опцию, определяющую, когда выходное значение будет пересчитываться:
  - Автоматически (в заданный момент времени или в конце простого эксперимента Simulation)
  - После вызова функции `update_outputName()`.

## Обновление объектов сбора данных

Новые данные могут добавляться в объекты сбора данных различными способами:

- [Автоматическое обновление](#). Новые измерения добавляются автоматически с заданной периодичностью.
- ["Ручное" обновление](#). Новые измерения добавляются "вручную" пользователем. Пользователь самостоятельно производит вычисление новых значений и затем добавляет их в объект сбора данных с помощью соответствующего программного интерфейса.
- [Обновление диаграмм](#). Обновление данных производится вследствие обновления диаграмм, отображающих эти объекты сбора данных.

Обратите внимание, что вы можете комбинировать эти подходы, например, использовать режим автоматического обновления и к тому же добавлять "вручную" дополнительные данные в ключевые моменты времени.

### Автоматическое обновление

Используйте этот режим, если вам нужно добавлять новые данные с заданной периодичностью. В этом случае выражение, заданное в поле **Значение** объекта сбора данных, будет вычисляться через заданный период времени, и полученное в результате вычисления значение будет добавляться в объект сбора данных.

#### Чтобы включить автоматическое обновление

1. Выберите объект сбора данных в графическом редакторе или в панели **Проекты**.
2. Перейдите в секцию **Обновление данных** панели **Свойства**.
3. Установите флажок **Обновлять данные автоматически**, а затем задайте период обновления в поле **Период**.

Если вы не хотите обновлять данные автоматически, вы можете отключить автоматическое обновление данных.

#### Чтобы отключить автоматическое обновление

1. Выберите объект сбора данных в графическом редакторе или в панели **Проекты**.
2. Перейдите в секцию **Обновление данных** панели **Свойства**.
3. Установите флажок **Не обновлять данные автоматически**.

### "Ручное" обновление

Используйте этот режим, если вам нужно выполнять обновление данных в какие-то определенные моменты времени, например, в те моменты когда нужные значения формируются в результате каких-то событий в модели.

Чтобы добавить новое значение в объект сбора данных, пользователь должен самостоятельно произвести вычисление нового значения и затем добавить его с помощью соответствующего программного интерфейса.

- В случае объектов с одной размерностью ([Статистика](#), [Данные гистограммы](#)), новые значения добавляются с помощью метода `add(double value)`.
- В случае объектов с двумя размерностями ([Набор данных](#), [Данные двумерной гистограммы](#)) вы должны будете передать объекту пару значений с помощью метода `add(double xValue, double yValue)`.

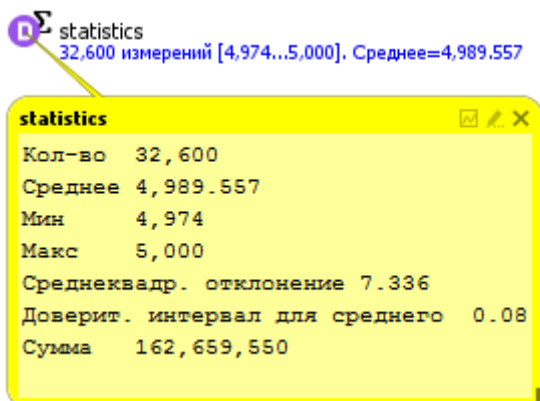
## Обновление диаграмм

Если объект сбора данных отображается на диаграмме, то при обновлении такой диаграммы будет происходить и обновление объекта сбора данных, а именно - в объект сбора данных будут добавляться новые значения, измеренные в момент обновления соответствующей диаграммы. Информацию об обновлении диаграмм см. здесь: [Обновление диаграмм](#).

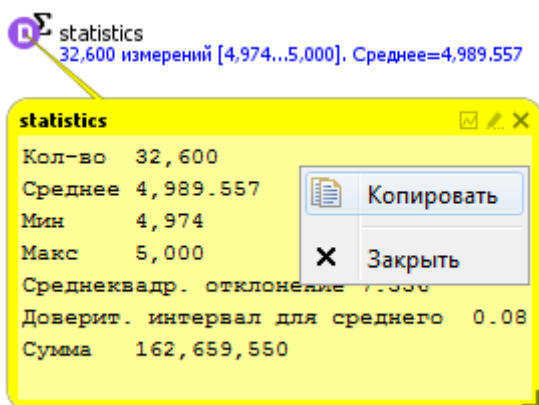
# Просмотр и копирование собранных данных

## Чтобы просмотреть собранные данные

1. Во время выполнения модели, щелкните значок элемента **Статистика** в окне презентации. Откроется окно инспекта, в котором будут отображены основные статистические данные, собранные элементом к этому моменту выполнения модели:



2. Если вы хотите скопировать эти данные статистики в Буфер обмена, щелкните окно инспекта правой кнопкой мыши (Mac OS: Ctrl+щелчок) и выберите опцию **Копировать** из контекстного меню.



3. Данные будут скопированы в Буфер обмена и вы сможете вставить их в стороннее приложение, например, Excel, чтобы их анализировать.