| МИНОБРНАУКИ РОССИИ |
| --- |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт кибербезопасности и цифровых технологий**

**Кафедра КБ-9 «Предметно-ориентированные информационные системы»**

**ОТЧЕТ ПО ПРАКТИЧЕСКИМ ЗАНЯТИЯМ № 15-16**

**по дисциплине «Информационные технологии анализа данных»**

| Отчет представлен к  рассмотрению:  Студент группы БСБО-16-20 | «18» декабря 2023 г. | (подпись) | Фамилия И.О. |
| --- | --- | --- | --- |
|  |  |  |  |
| Преподаватель | «18» декабря 2023 г. | (подпись) | Анфёров М.А. |

Москва, 2023 г.

**СОДЕРЖАНИЕ**

[Раздел 1. Постановка задачи и подготовка исходных данных 3](#_30j0zll)

[1.1](#_1fob9te) Экономико-математическая постановка задачи 3

[1.2](#_2et92p0) Получение исходных данных 3

[1.3](#_3dy6vkm) Подготовка исходных данных 4

[Раздел 2. Построение модели временного ряда с заданием линейного тренда 5](#_2s8eyo1)

[2.1](#_17dp8vu) Моделирование временного ряда в программе Deductor Academic 5

[2.2](#_lnxbz9) Оценка качества моделирования 13

[2.3](#_35nkun2) Выполнение прогнозных расчетов 14

[Раздел 3. Построение модели временного ряда с заданием степенного тренда 15](#_1ksv4uv)

[3.1](#_44sinio) Моделирование временного ряда в программе Deductor Academic 15

[3.2](#_2jxsxqh) Оценка качества моделирования 21

[3.3](#_z337ya) Выполнение прогнозных расчетов 22

[Раздел 4. Спектральная обработка исходных данных 24](#_3j2qqm3)

[4.1](#_1y810tw) Спектральная обработка в программе Deductor Academic 24

[4.2](#_4i7ojhp) Моделирование временного ряда в программе Deductor Academic 27

[4.3](#_2xcytpi) Оценка качества моделирования 33

[4.4](#_1ci93xb) Выполнение прогнозных расчетов 34

[Вывод 35](#_3whwml4)

# Раздел 1. Постановка задачи и подготовка исходных данных

## Экономико-математическая постановка задачи

Многофункциональный центр государственных услуг одного из районов города осуществляет прием заявлений от населения на оказание соответствующих услуг. Требуется оценить возможную зависимость объема оказываемых услуг, выраженного в количестве поданных заявлений, от фактора времени. Для этого необходимо построить модель временного ряда, с помощью которой можно выполнить прогнозирование активности населения в обращении за госуслугами.

## Получение исходных данных

Для решения данной задачи были получены исходные данные, содержащие разнообразную информацию по заявлениям. Исходные данные, полученные от преподавателя, представлены в Таблице 1.

*Таблица 1 – Исходные данные*

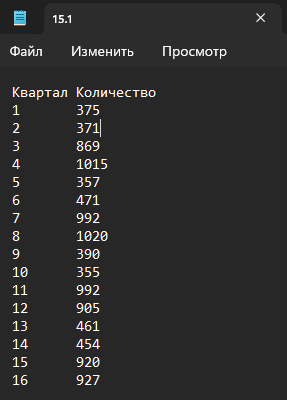
| Текущий номер квартала во временному ряду | Количество обращений за услугой |
| --- | --- |
| 1 | 375 |
| 2 | 371 |
| 3 | 869 |
| 4 | 1015 |
| 5 | 357 |
| 6 | 471 |
| 7 | 992 |
| 8 | 1020 |
| 9 | 390 |
| 10 | 355 |
| 11 | 992 |

*Таблица 1 – Исходные данные (продолжение)*

| 12 | 905 |
| --- | --- |
| 13 | 461 |
| 14 | 454 |
| 15 | 920 |
| 16 | 927 |

## Подготовка исходных данных

Для передачи исходных данных в аналитическую платформу требуется подготовить соответствующий файл в текстовом формате, скриншот представлен на Рисунке 1.

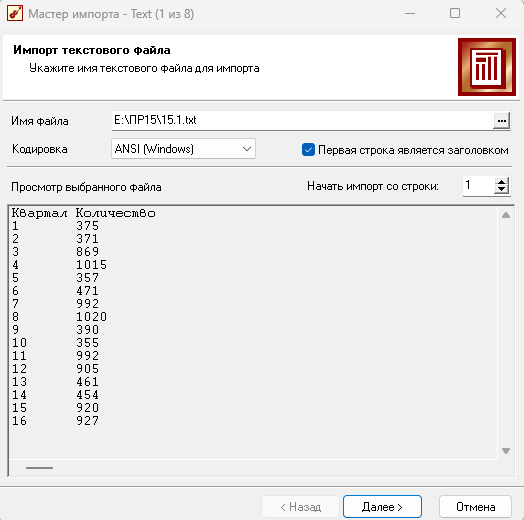


**Рисунок 1 – Исходные данные в текстовом формате**

# Раздел 2. Построение модели временного ряда с заданием линейного тренда

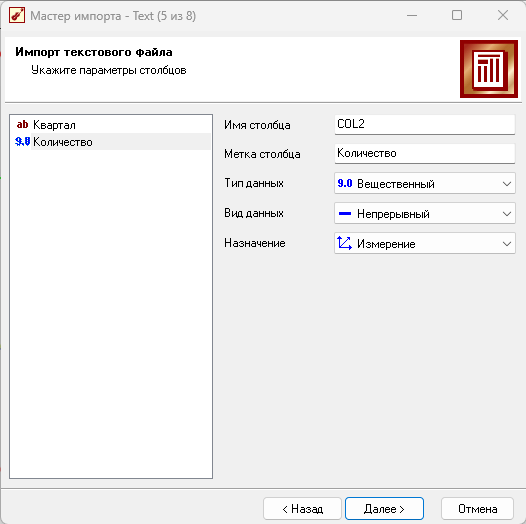
## Моделирование временного ряда в программе Deductor Academic

Необходимо запустить приложение Deductor Academic и создать новый проект. Далее необходимо кликнуть на пиктограмме «Мастер импорта». На первом шаге необходимо ввести имя файла, содержащего исходные данные, и отметить наличие заголовков в первой строке. Импортируемые данные будут отображаться в окне импорта. Первый шаг показан на Рисунке 2.

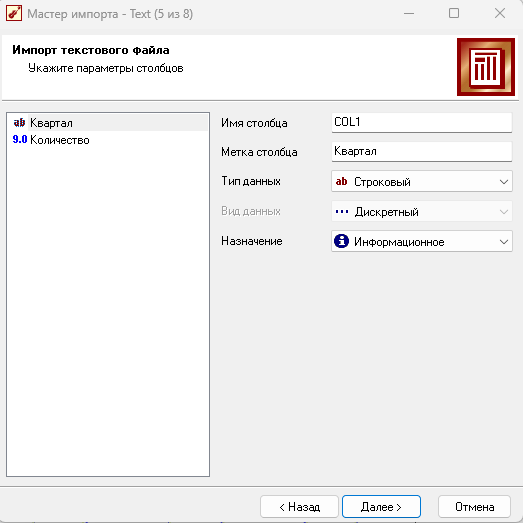


**Рисунок 2 – Импорт текстового файла**

На пятом шаге необходимо определить тип и назначение данных в каждом столбце. Данные столбца содержащего значения уровней временного ряда определяется с вещественным типом и назначением «Измерение». Данные остальных столбцов определяются ка информационные строкового типа. Пятый шаг показан на Рисунках 3-4.



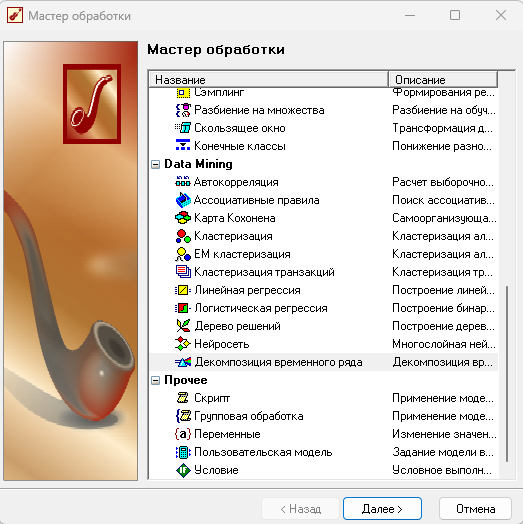
**Рисунок 3 – Идентификация данных (Количество)**



**Рисунок 4 – Идентификация данных (Квартал)**

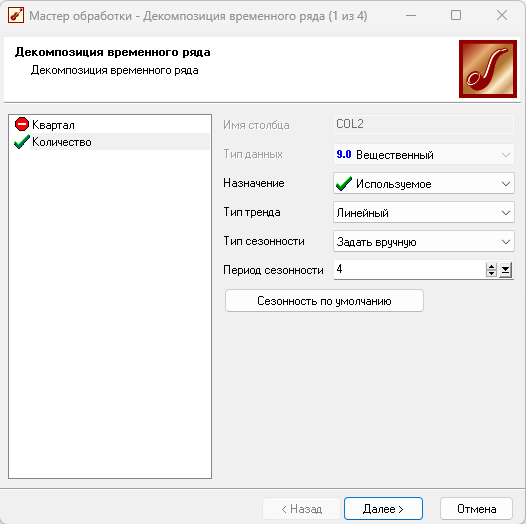
После импорта исходных данных можно приступить построению нейронной сети. Для этого необходимо активизировать «Мастер обработки».

В открывшемся окне «Мастер обработки» необходимо в разделе «Data Mining» выбрать «Декомпозиция временного ряда» и пройти по всем шагам мастера. Окно «Мастер обработки» показан на Рисунке 5.



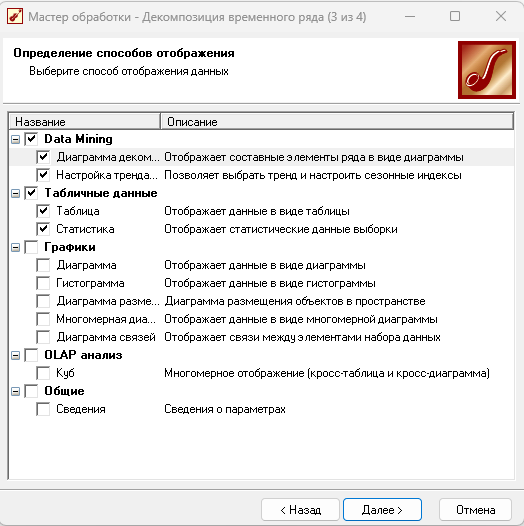
**Рисунок 5 – Мастер обработки**

На первом шаге необходимо для данных временного ряда (столбец «Количество») задать значение сезонной компоненты (значение 4), а также вид функции тренда, моделирующей тенденцию. Для рассматриваемого примера назначим линейный тренд. Первый шаг показан на Рисунке 6.



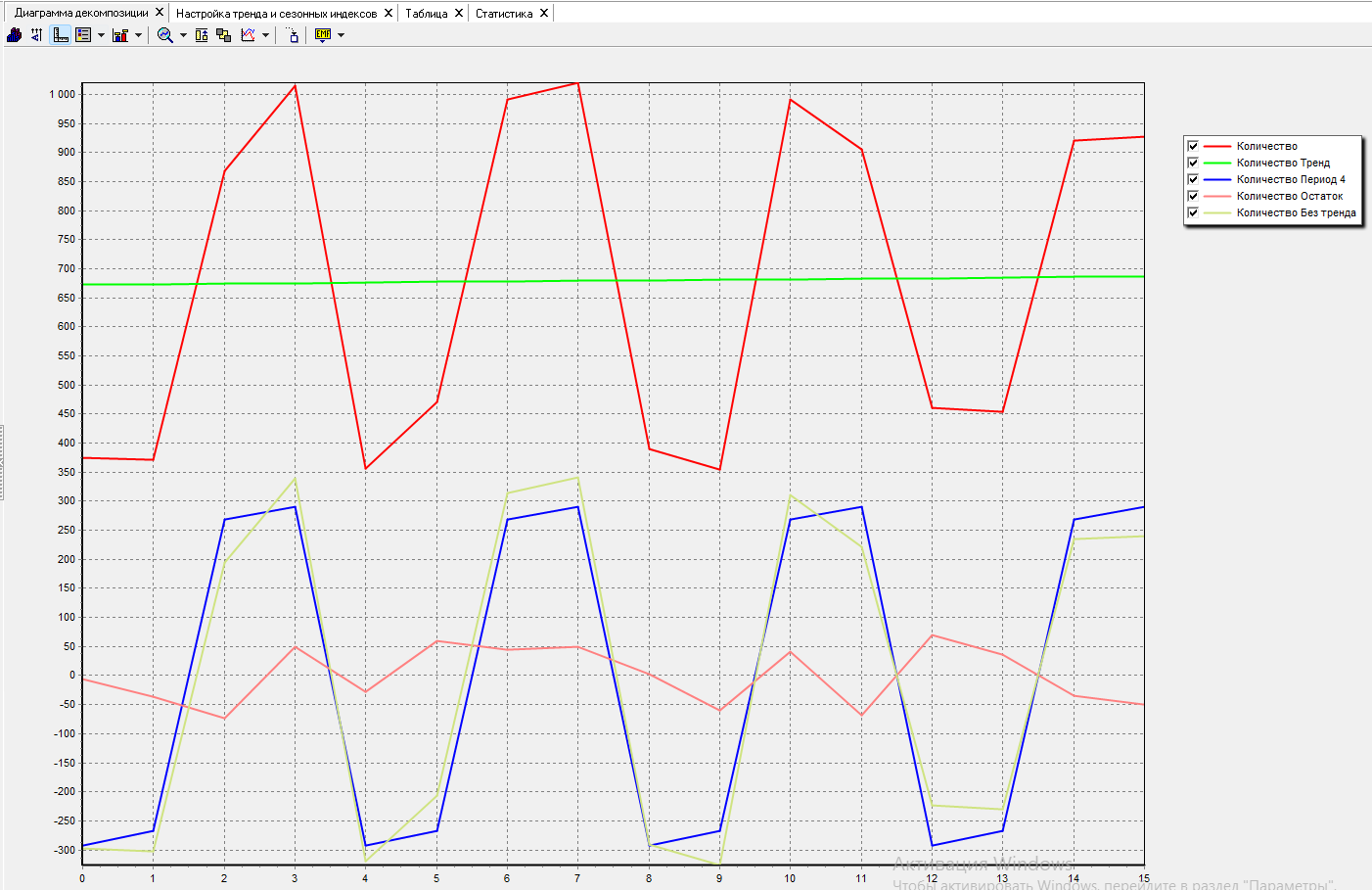
**Рисунок 6 – Настройка назначения данных**

На третьем шаге мастера обработки необходимо задать способы отображения результатов моделирования. Как минимум – это должны быть «Диаграмма декомпозиции», «Настройка тренда и сезонных индексов», «Таблица» и «Статистика». Третий шаг показан на Рисунке 7.

****

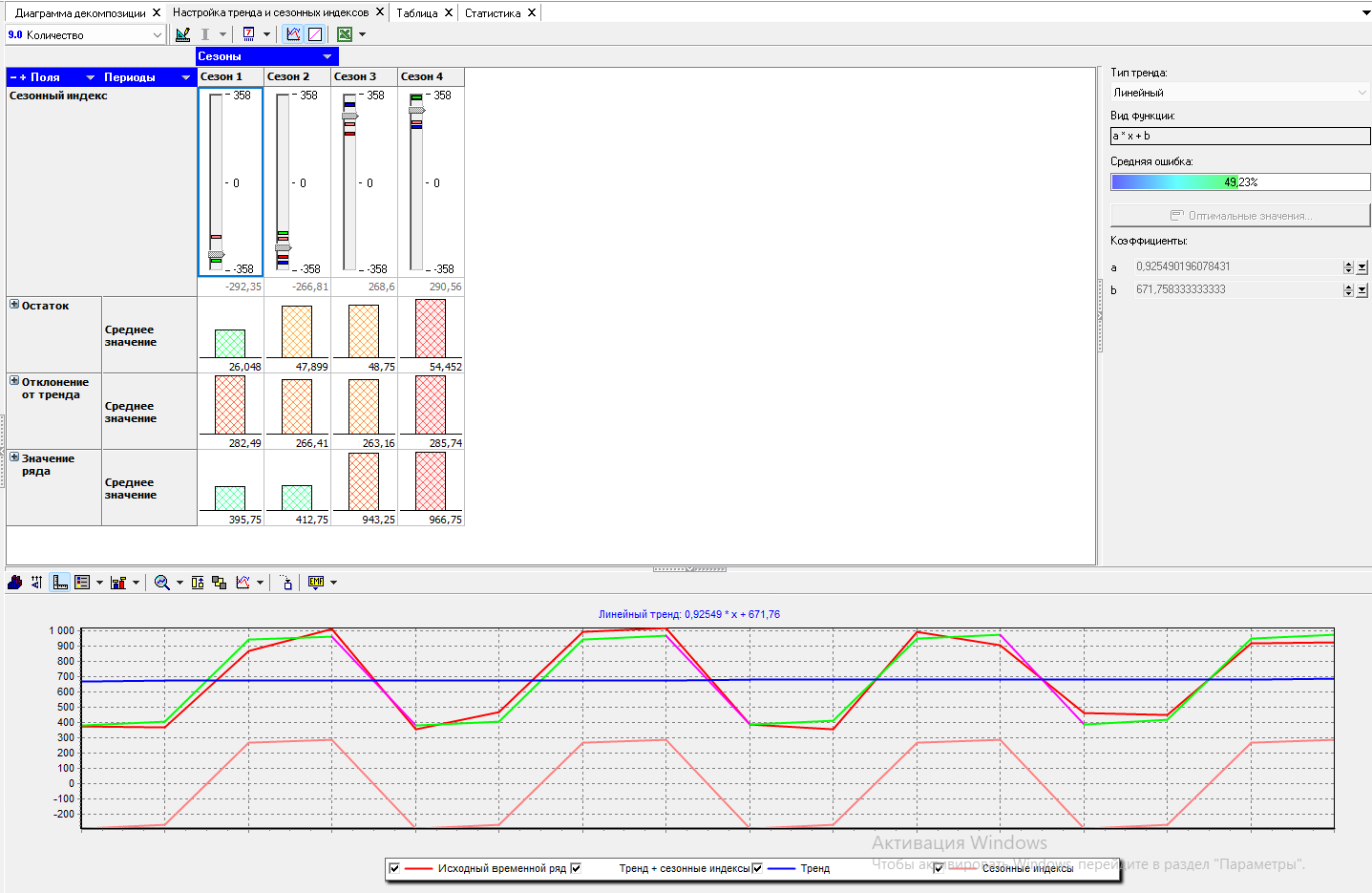
**Рисунок 7 – Определение способов отображения**

«Диаграмма декомпозиции» позволяет наглядно графически увидеть отображение исходного временного ряда и его составляющих – тенденции (тренда) и сезонных компонентов. Также на вкладе показана диаграмма остатков, позволяющих оценить погрешность моделирования. «Диаграмма декомпозиции» показана на Рисунке 8.

****

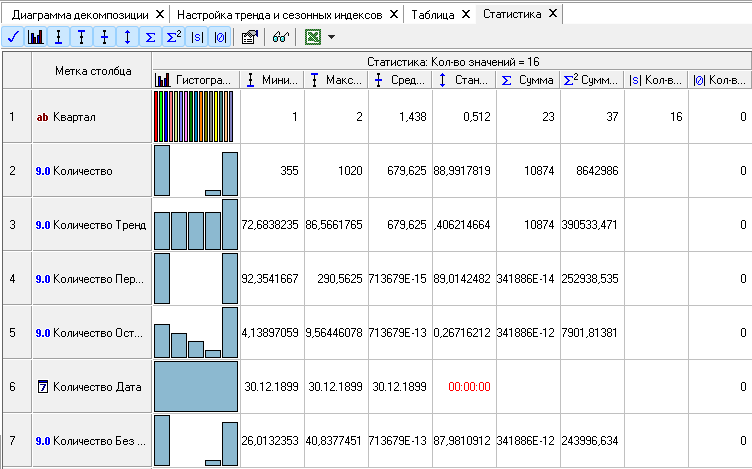
**Рисунок 8 – Инструментальное средство «Диаграмма декомпозиции»**

Вкладка «Настройка тренда и сезонных индексов» дает максимальную информацию для анализа полученной модели и выполнения прогнозных расчетов. Графически отображаются сезонные индексы по кварталам и тенденция в виде линии тренда, а также сумма этих рассчитанных составляющих временного ряда. На данный график наложена диаграмма исходного ряда для оценки расхождения найденной модели с реальными данными. «Настройка тренда и сезонных индексов» показана на Рисунке 9.

****

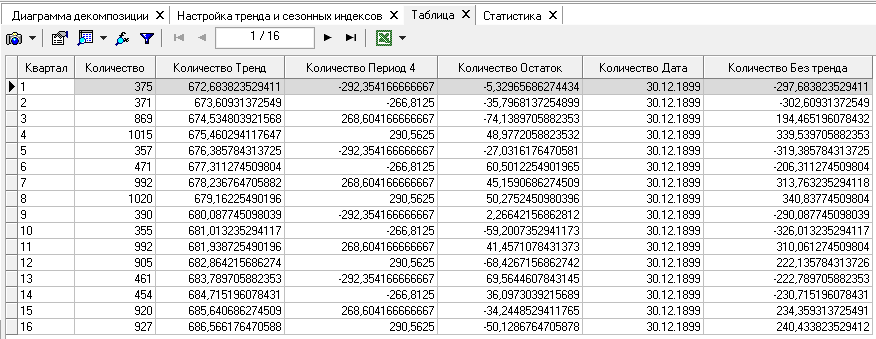
**Рисунок 9 – Инструментальное средство «Настройка тренда и сезонных индексов»**

Вкладка «Статистика» представляет информацию, позволяющую оценить качество моделирования. «Статистика» показана на Рисунке 10.

****

**Рисунок 10 – Инструментальное средство «Статистика»**

На вкладке «Таблица» полученные результаты отображаются в виде таблицы, которая будет экспортирована в MS Excel для проведения дополнительного анализа. «Таблица» показана на Рисунке 11.

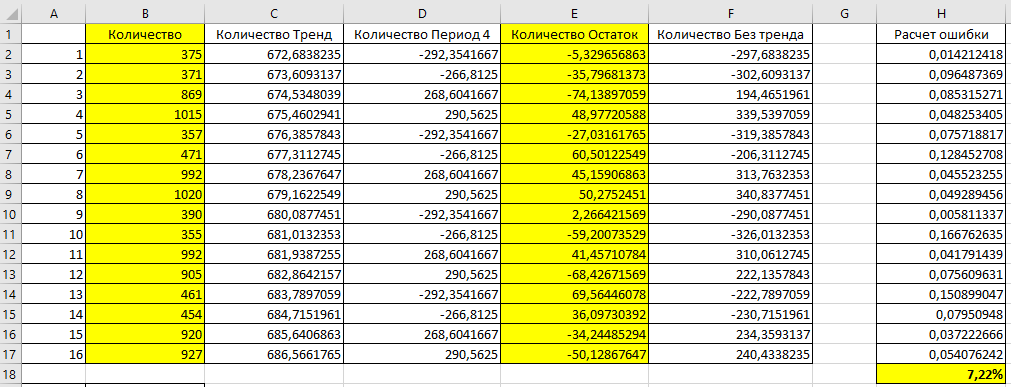
****

**Рисунок 11 – Инструментальное средство «Таблица»**

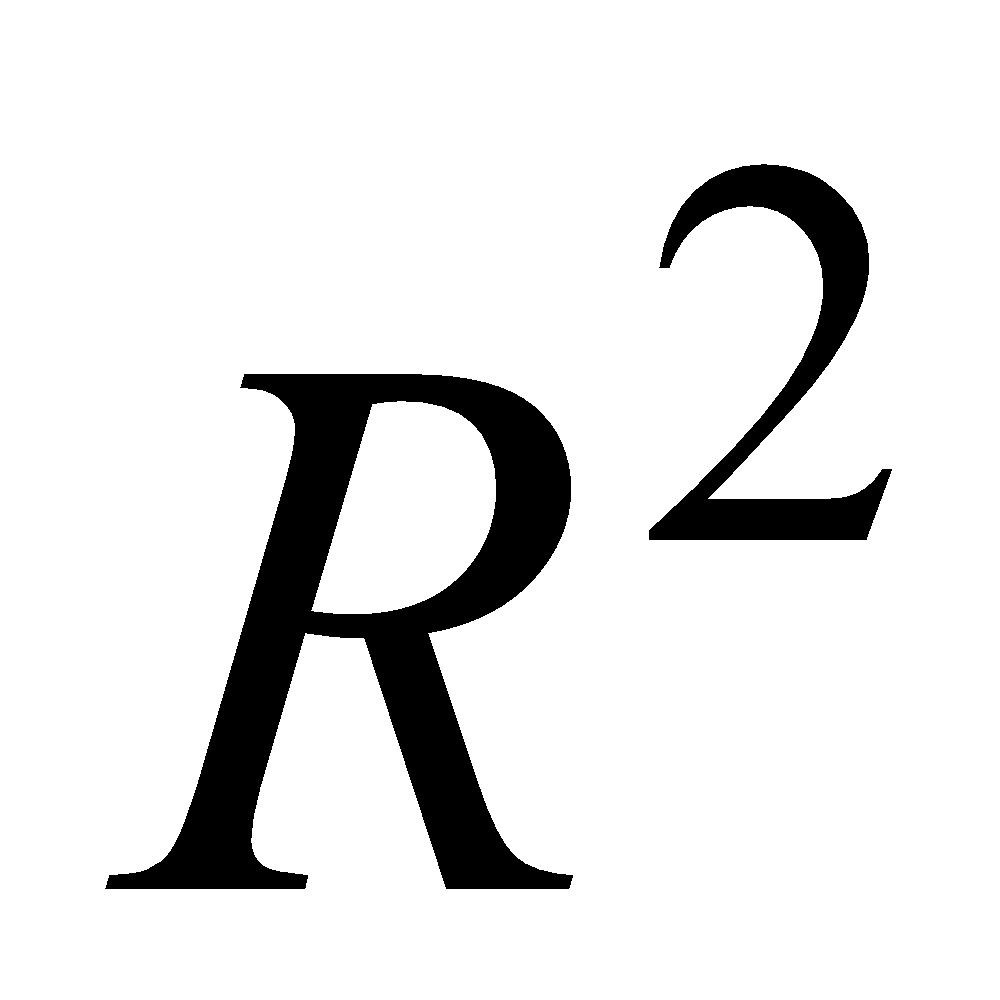
## Оценка качества моделирования

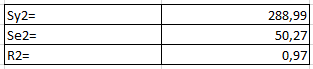
Расчет средней относительной ошибки для рассматриваемого примера показан на Рисунке 12. Для этого использован дополнительный столбец с промежуточными расчетам. Сама ошибка рассчитывается в MS Excel с использованием функции СРЗНАЧ.

Точность моделирования временного ряда в данном примере составила 7%, что является достаточно хорошим результатом.

****

**Рисунок 12 – Средняя относительная ошибка**

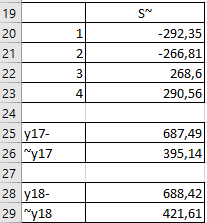
Далее рассчитаем коэффициент детерминации. В нашем случае значения представляют собой квадраты среднеквадратических отклонений уровней ряда и остатков, которые отображены на вкладке «Статистика» как стандартные отклонения: и . В результате вычислений по формуле: получаем значение коэффициента детерминации , которое показывает высокое качество моделирования. Коэффициента детерминации показан на Рисунке 13.

****

**Рисунок 13 – Коэффициента детерминации**

## Выполнение прогнозных расчетов

На вкладке «Настройка тренда и сезонных индексов» показаны значения сезонной составляющей по кварталам. Для рассматриваемого примера – это = -292,32 для I квартала, = -266,81для II квартала, = 268,6 для III квартала и = 290,56 для IV квартала. Функциональная зависимость тенденции совместно с величинами сезонных индексов позволяет выполнить интересующие расчеты , . По результатам получаем и . Прогнозный расчет показан на Рисунке 14.

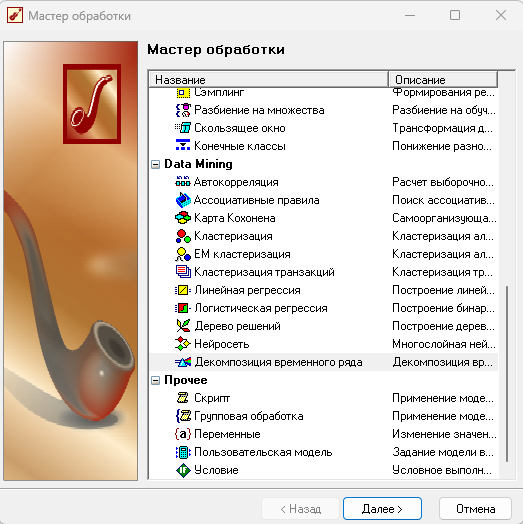


**Рисунок 14 – Прогнозный расчет**

# Раздел 3. Построение модели временного ряда с заданием степенного тренда

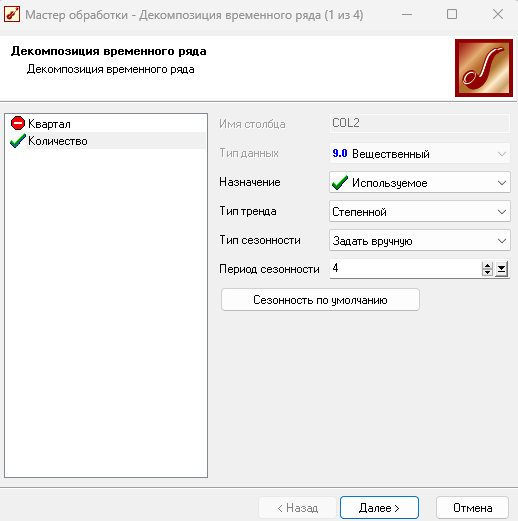
## Моделирование временного ряда в программе Deductor Academic

В открывшемся окне «Мастер обработки» необходимо в разделе «Data Mining» выбрать «Декомпозиция временного ряда» и пройти по всем шагам мастера. Окно «Мастер обработки» показан на Рисунке 15.



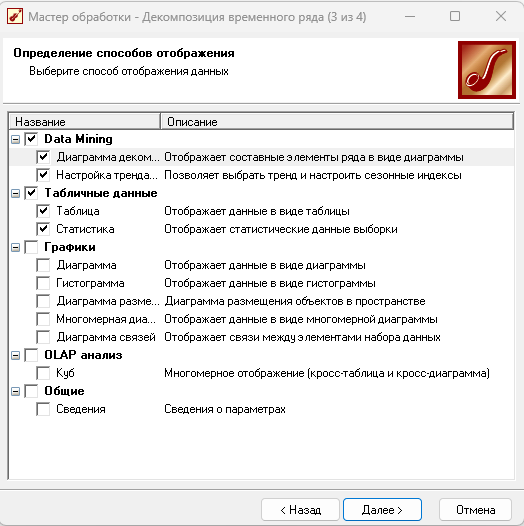
**Рисунок 15 – Мастер обработки**

На первом шаге необходимо для данных временного ряда (столбец «Количество») задать значение сезонной компоненты (значение 4), а также вид функции тренда, моделирующей тенденцию. Для рассматриваемого примера назначим степенной тренд. Первый шаг показан на Рисунке 16.



**Рисунок 16 – Настройка назначения данных**

На третьем шаге мастера обработки необходимо задать способы отображения результатов моделирования. Как минимум – это должны быть «Диаграмма декомпозиции», «Настройка тренда и сезонных индексов», «Таблица» и «Статистика». Третий шаг показан на Рисунке 17.

****

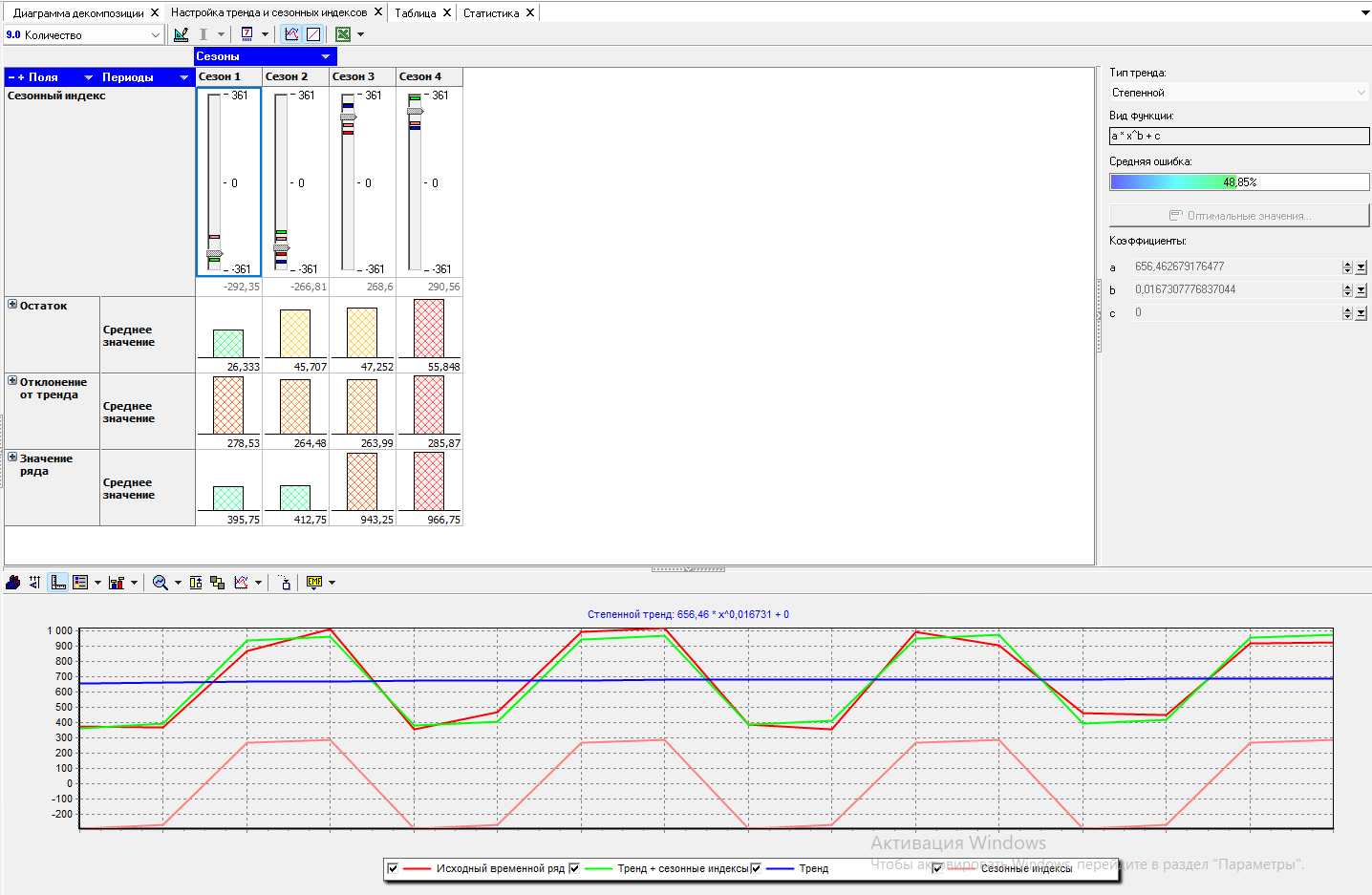
**Рисунок 17 – Определение способов отображения**

«Диаграмма декомпозиции» позволяет наглядно графически увидеть отображение исходного временного ряда и его составляющих – тенденции (тренда) и сезонных компонентов. Также на вкладе показана диаграмма остатков, позволяющих оценить погрешность моделирования. «Диаграмма декомпозиции» показана на Рисунке 18.

****

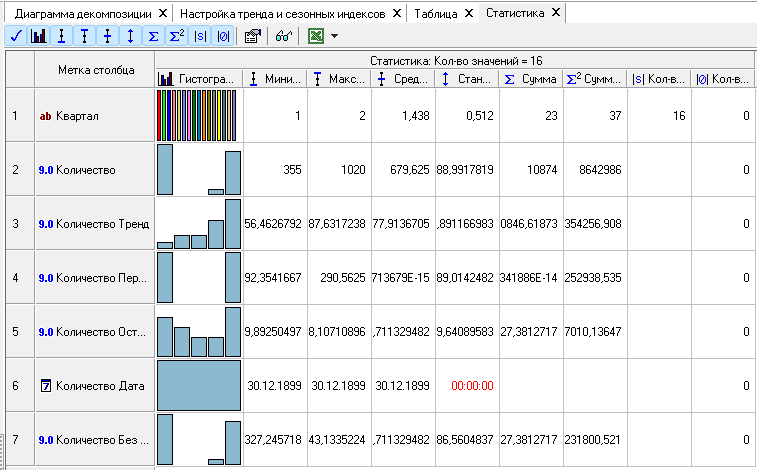
**Рисунок 18 – Инструментальное средство «Диаграмма декомпозиции»**

Вкладка «Настройка тренда и сезонных индексов» дает максимальную информацию для анализа полученной модели и выполнения прогнозных расчетов. Графически отображаются сезонные индексы по кварталам и тенденция в виде линии тренда, а также сумма этих рассчитанных составляющих временного ряда. На данный график наложена диаграмма исходного ряда для оценки расхождения найденной модели с реальными данными. «Настройка тренда и сезонных индексов» показана на Рисунке 19.

****

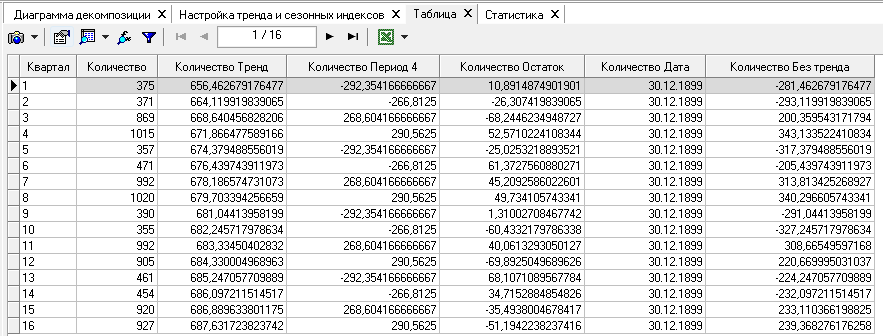
**Рисунок 19 – Инструментальное средство «Настройка тренда и сезонных индексов»**

Вкладка «Статистика» представляет информацию, позволяющую оценить качество моделирования. «Статистика» показана на Рисунке 20.

****

**Рисунок 20 – Инструментальное средство «Статистика»**

На вкладке «Таблица» полученные результаты отображаются в виде таблицы, которая будет экспортирована в MS Excel для проведения дополнительного анализа. «Таблица» показана на Рисунке 21.

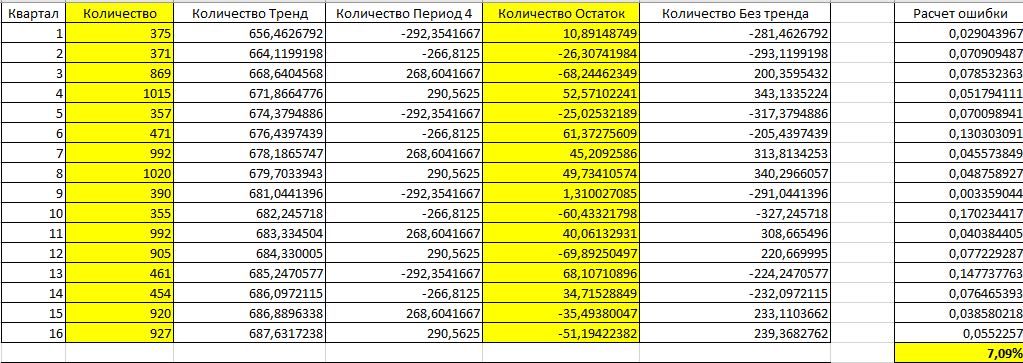
****

**Рисунок 21 – Инструментальное средство «Таблица»**

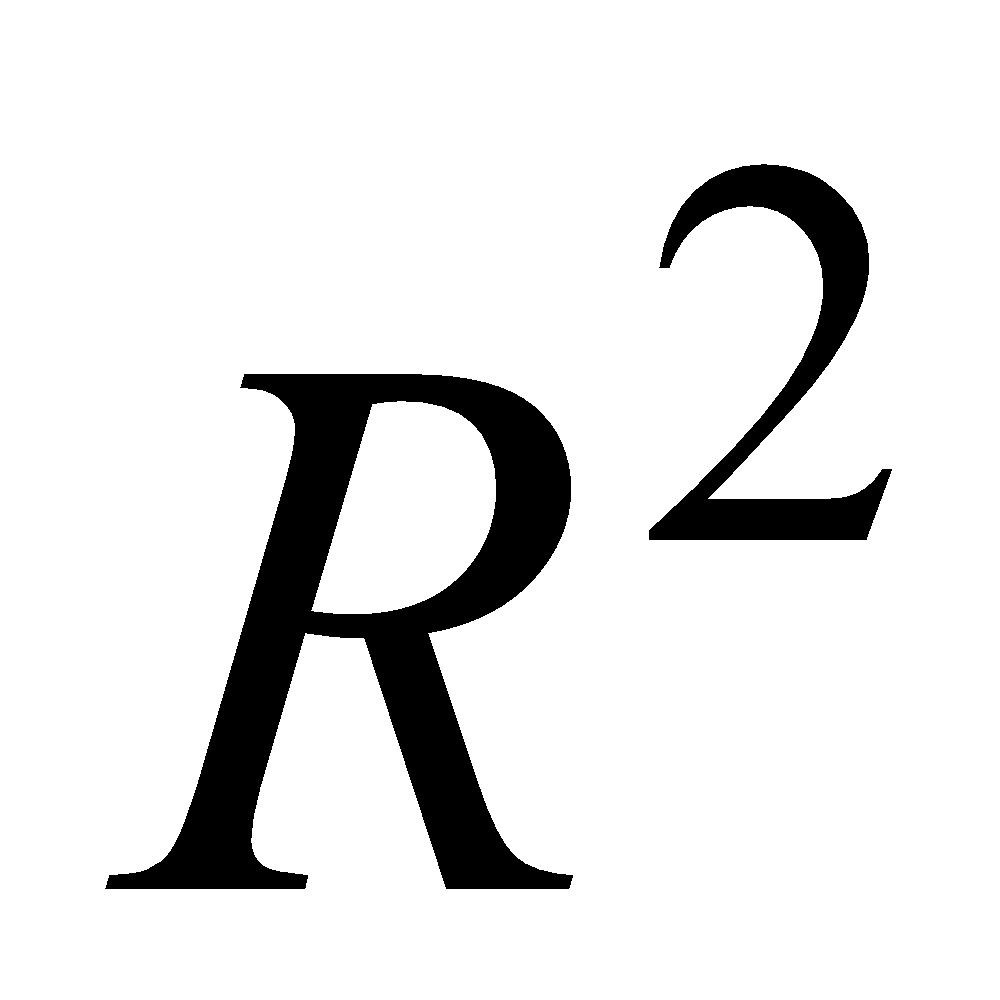
## Оценка качества моделирования

Расчет средней относительной ошибки для рассматриваемого примера показан на Рисунке 22. Для этого использован дополнительный столбец с промежуточными расчетам. Сама ошибка рассчитывается в MS Excel с использованием функции СРЗНАЧ.

Точность моделирования временного ряда в данном примере составила 7%, что является достаточно хорошим результатом.

****

**Рисунок 22 – Средняя относительная ошибка**

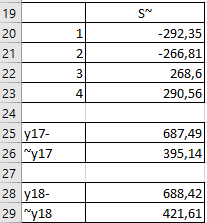
Далее рассчитаем коэффициент детерминации. В нашем случае значения представляют собой квадраты среднеквадратических отклонений уровней ряда и остатков, которые отображены на вкладке «Статистика» как стандартные отклонения: и . В результате вычислений по формуле: получаем значение коэффициента детерминации , которое показывает высокое качество моделирования. Коэффициента детерминации показан на Рисунке 23.

****

**Рисунок 23 – Коэффициента детерминации**

## Выполнение прогнозных расчетов

На вкладке «Настройка тренда и сезонных индексов» показаны значения сезонной составляющей по кварталам. Для рассматриваемого примера – это = -292,32 для I квартала, = -266,81для II квартала, = 268,6 для III квартала и = 290,56 для IV квартала. Функциональная зависимость тенденции совместно с величинами сезонных индексов позволяет выполнить интересующие расчеты , . По результатам получаем и . Прогнозный расчет показан на Рисунке 24.

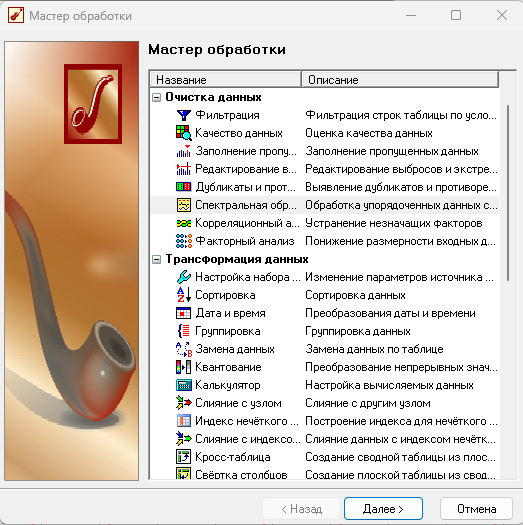


**Рисунок 24 – Прогнозный расчет**

# Раздел 4. Спектральная обработка исходных данных

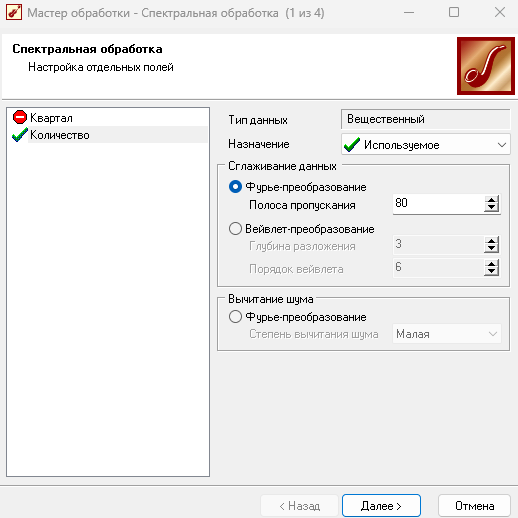
## Спектральная обработка в программе Deductor Academic

В открывшемся окне «Мастер обработки» необходимо в разделе «Data Mining» выбрать «Декомпозиция временного ряда» и пройти по всем шагам мастера. Окно «Мастер обработки» показан на Рисунке 25.



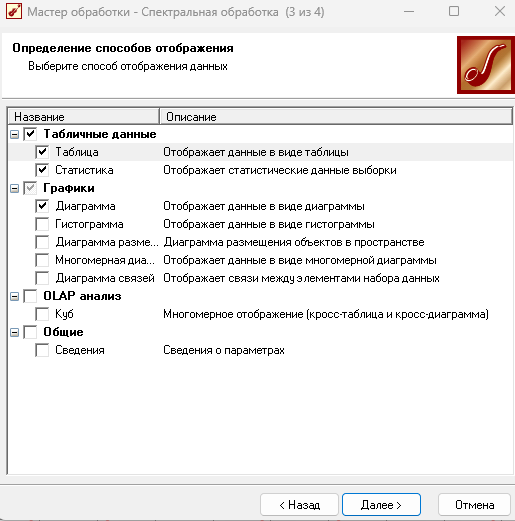
**Рисунок 25 – Мастер обработки**

На первом шаге необходимо для данных временного ряда (столбец «Количество») задать назначение «Используемое», выбрать для обработки «Фурье преобразование» и задать для него ширину частотной полосы пропускания – 80. Первый шаг показан на Рисунке 26.



**Рисунок 26 – Установление параметров преобразования**

На третьем шаге задаем способы отображения полученных результатов в виде таблицы с показом статистики и диаграмма. Третий шаг показан на Рисунке 27.



**Рисунок 27 – Способы отображения результатов спектральной обработки**

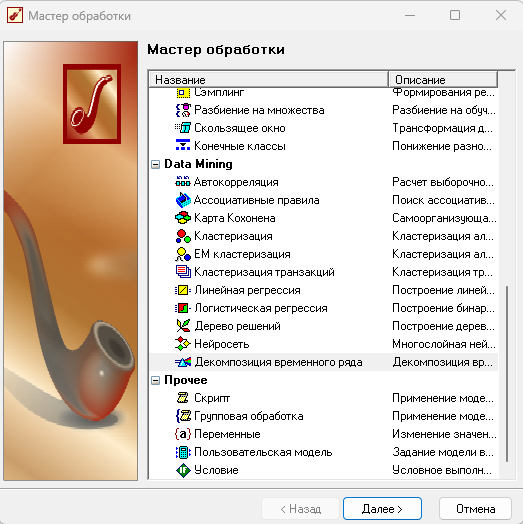
С помощью диаграммы можно качественно сравнить отфильтрованные спектральной обработкой данные с исходными. Диаграмма показана на Рисунке 28.



**Рисунок 28 – Диаграмма отфильтрованных данных**

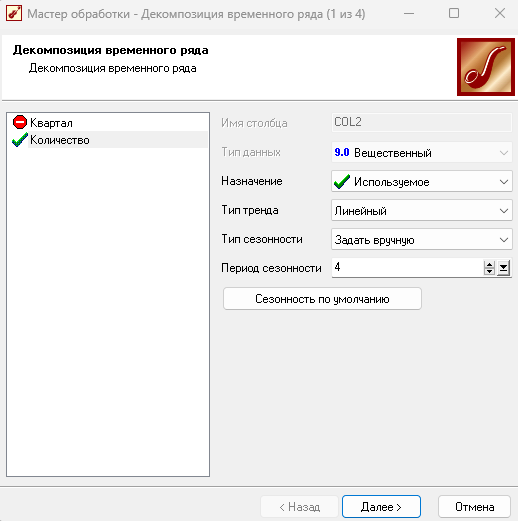
## Моделирование временного ряда в программе Deductor Academic

В открывшемся окне «Мастер обработки» необходимо в разделе «Data Mining» выбрать «Декомпозиция временного ряда» и пройти по всем шагам мастера. Окно «Мастер обработки» показан на Рисунке 29.



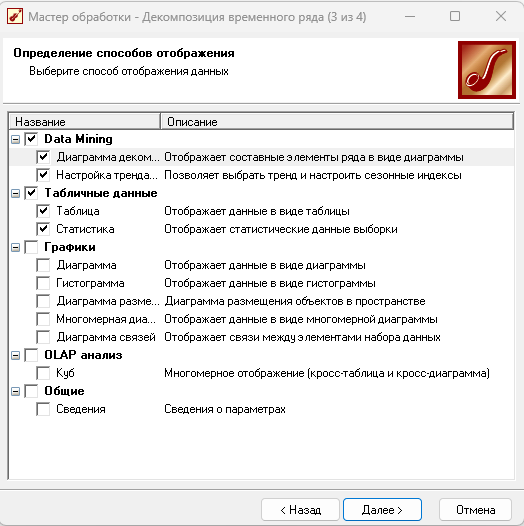
**Рисунок 29 – Мастер обработки**

На первом шаге необходимо для данных временного ряда (столбец «Количество») задать значение сезонной компоненты (значение 4), а также вид функции тренда, моделирующей тенденцию. Для рассматриваемого примера назначим линейный тренд. Первый шаг показан на Рисунке 30.



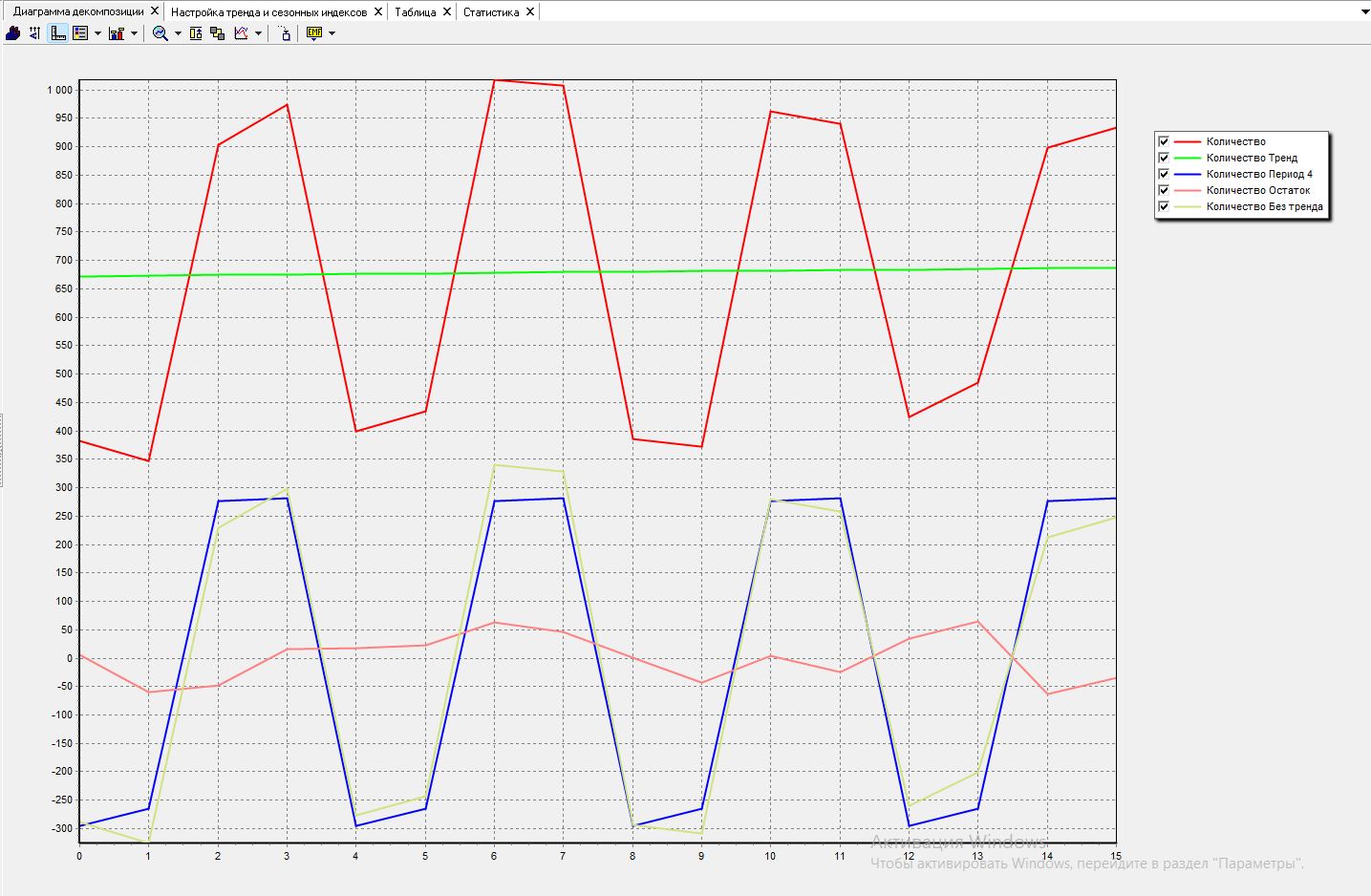
**Рисунок 30 – Настройка назначения данных**

На третьем шаге мастера обработки необходимо задать способы отображения результатов моделирования. Как минимум – это должны быть «Диаграмма декомпозиции», «Настройка тренда и сезонных индексов», «Таблица» и «Статистика». Третий шаг показан на Рисунке 31.

****

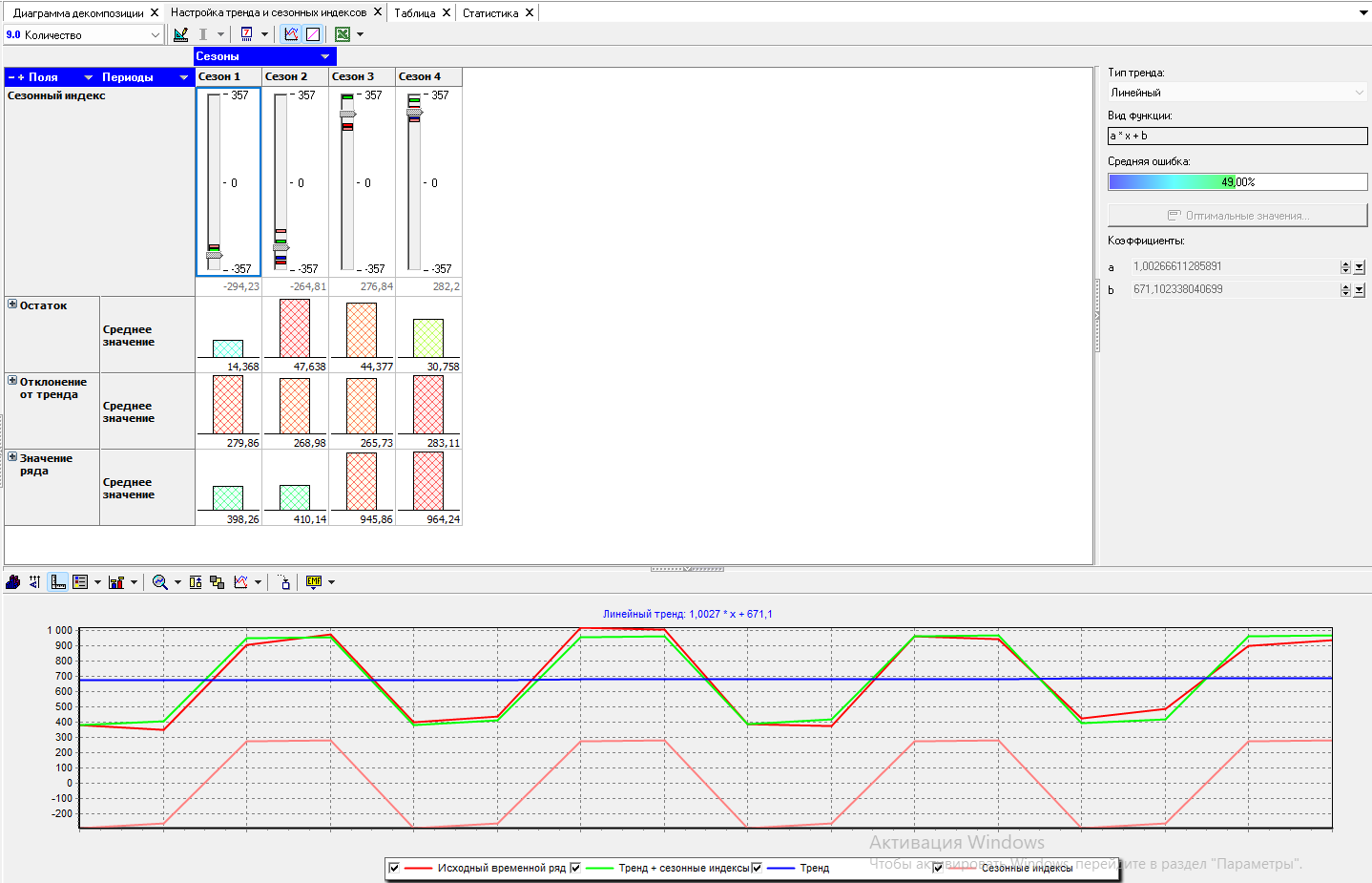
**Рисунок 31 – Определение способов отображения**

«Диаграмма декомпозиции» позволяет наглядно графически увидеть отображение исходного временного ряда и его составляющих – тенденции (тренда) и сезонных компонентов. Также на вкладе показана диаграмма остатков, позволяющих оценить погрешность моделирования. «Диаграмма декомпозиции» показана на Рисунке 32.

****

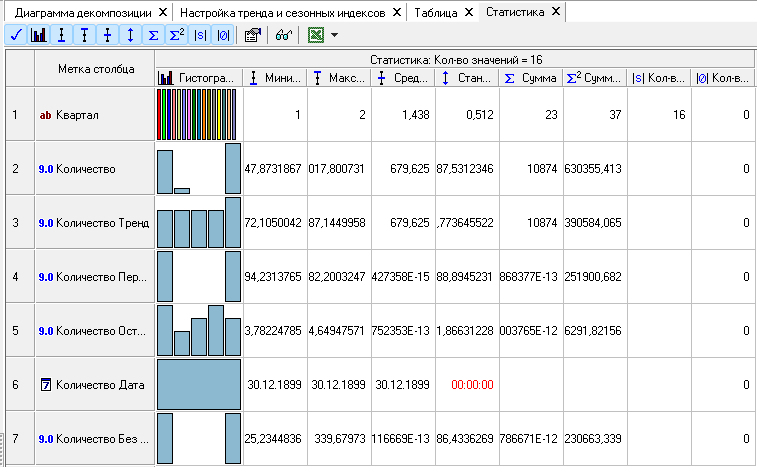
**Рисунок 32 – Инструментальное средство «Диаграмма декомпозиции»**

Вкладка «Настройка тренда и сезонных индексов» дает максимальную информацию для анализа полученной модели и выполнения прогнозных расчетов. Графически отображаются сезонные индексы по кварталам и тенденция в виде линии тренда, а также сумма этих рассчитанных составляющих временного ряда. На данный график наложена диаграмма исходного ряда для оценки расхождения найденной модели с реальными данными. «Настройка тренда и сезонных индексов» показана на Рисунке 33.

****

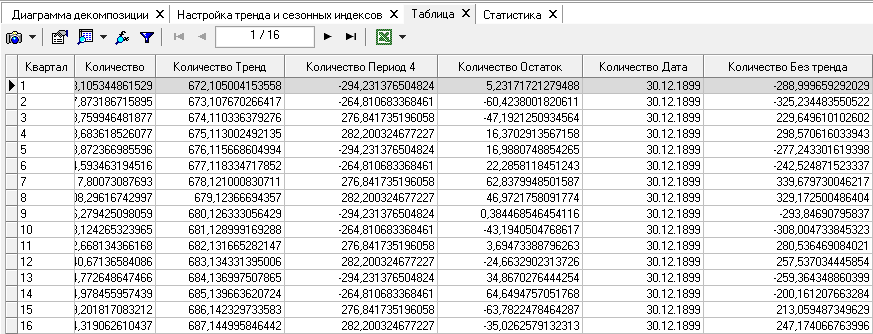
**Рисунок 33 – Инструментальное средство «Настройка тренда и сезонных индексов»**

Вкладка «Статистика» представляет информацию, позволяющую оценить качество моделирования. «Статистика» показана на Рисунке 34.

****

**Рисунок 34 – Инструментальное средство «Статистика»**

На вкладке «Таблица» полученные результаты отображаются в виде таблицы, которая будет экспортирована в MS Excel для проведения дополнительного анализа. «Таблица» показана на Рисунке 35.

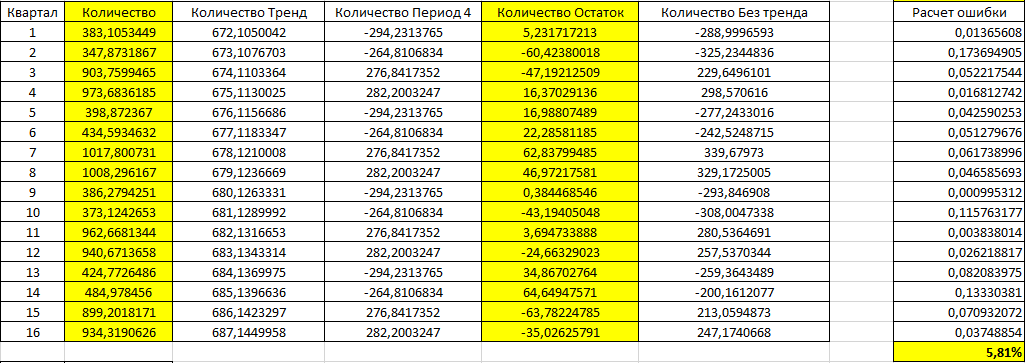
****

**Рисунок 35 – Инструментальное средство «Таблица»**

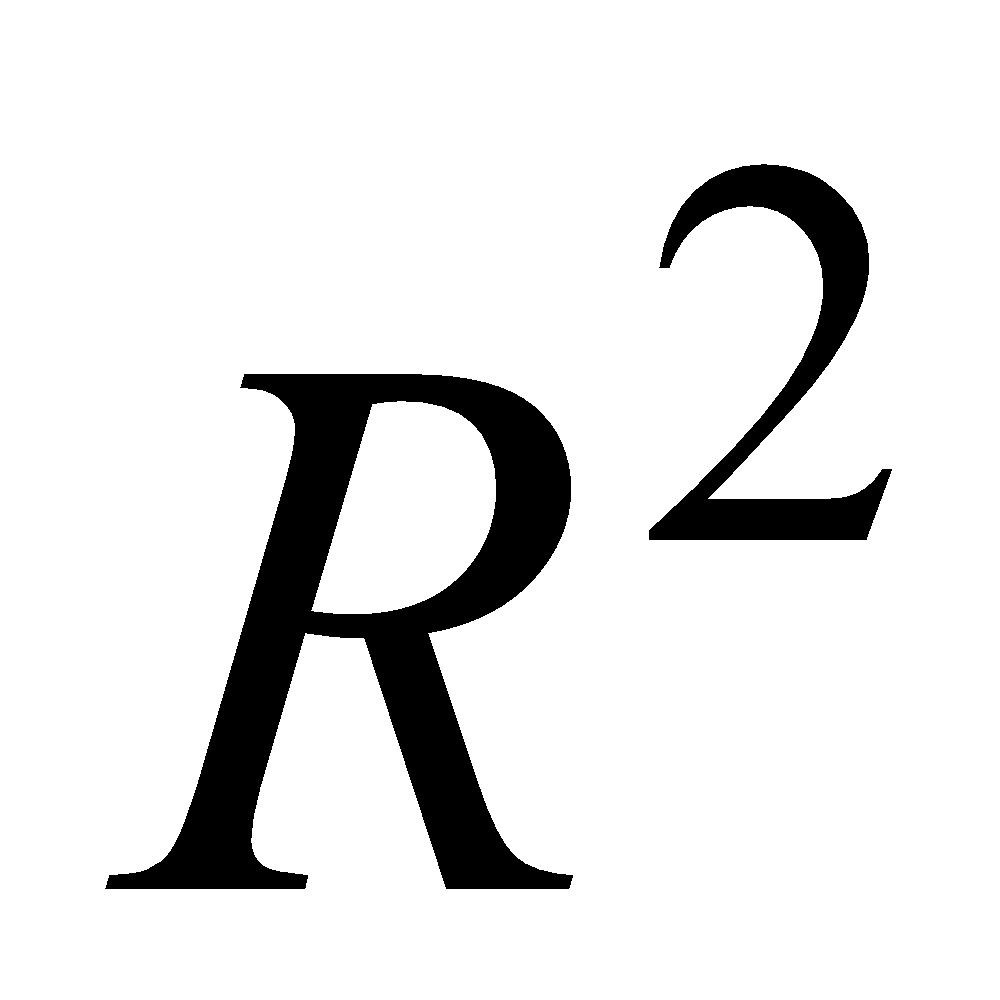
## Оценка качества моделирования

Расчет средней относительной ошибки для рассматриваемого примера показан на Рисунке 36. Для этого использован дополнительный столбец с промежуточными расчетам. Сама ошибка рассчитывается в MS Excel с использованием функции СРЗНАЧ.

Точность моделирования временного ряда в данном примере составила 5,8%, что является достаточно хорошим результатом.

****

**Рисунок 36 – Средняя относительная ошибка**

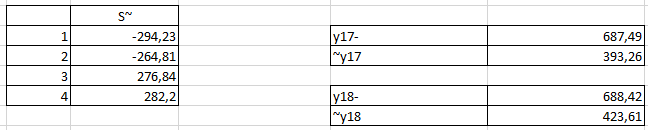
Далее рассчитаем коэффициент детерминации. В нашем случае значения представляют собой квадраты среднеквадратических отклонений уровней ряда и остатков, которые отображены на вкладке «Статистика» как стандартные отклонения: и . В результате вычислений по формуле: получаем значение коэффициента детерминации , которое показывает высокое качество моделирования. Коэффициента детерминации показан на Рисунке 37.

****

**Рисунок 37 – Коэффициента детерминации**

## Выполнение прогнозных расчетов

На вкладке «Настройка тренда и сезонных индексов» показаны значения сезонной составляющей по кварталам. Для рассматриваемого примера – это = -294,23 для I квартала, = -264,81для II квартала, = 276,84 для III квартала и = 282,2 для IV квартала. Функциональная зависимость тенденции совместно с величинами сезонных индексов позволяет выполнить интересующие расчеты , . По результатам получаем и . Прогнозный расчет показан на Рисунке 38.



**Рисунок 38 – Прогнозный расчет**

# Вывод

Исследуя возможности «Декомпозиция временного ряда» в функции «Data Mining» были построены модели рядов линейных и степенных. При сравнении моделей, было выявлено, что самым наилучшим оказался степенной тренд временного ряда. Также была проведена спектральная обработка данных, после которой результат моделирования временного ряда улучшился.

Deductor Studio Academic является мощным инструментом анализа данных, обладает способностью обрабатывать большие объемы информации, выявлять закономерности и делать выводы.