



Отчет по регрессии

Визуализатор *Отчет по регрессии* компонента Линейная регрессия отображает параметры и результаты статистических тестов для анализа регрессионных моделей.

Данный визуализатор состоит из трех областей:

- Информация о модели
- Коэффициенты регрессии
- Шаги построения

Интерфейс

Операции

Область Информация о модели

-  **Информация о модели** — отображение/скрытие информации о модели.

Область Коэффициенты регрессии

-  **Таблица** — переключение отображения коэффициентов регрессии в вид *Таблица*.
-  **Дерево** — переключение отображения коэффициентов регрессии в вид *Дерево*.
-  **Нулевые значения** — отображение/скрытие нулевых значений.

Область Шаги построения

-  **Шаги построения** — отображение/скрытие шагов построения модели.

Области визуализатора

Информация о модели

- Модель (выпадающий список) - позволяет выбрать одну из моделей, созданных пошаговым алгоритмом в процессе отбора факторов.

Описание показателей модели представлено в таблице:

Показатель	Описание
0/1 Константа	Информация о том содержит модель константу или нет
9.0 Логарифм функции правдоподобия	Показывает степень соответствия модели набору данных, на котором она обучается
9.0 Коэффициент детерминации	Доля дисперсии зависимой переменной, объяснённая регрессионной моделью (изменяется от 0 до 1)
9.0 Коэффициент детерминации (скорр.)	Отличается от нескорректированного коэффициента детерминации тем, что накладывает штраф за дополнительно вводимые в модель переменные (изменяется от 0 до 1)
9.0 Стандартное отклонение	Мера отклонения примеров от предсказанных моделью значений
12 Число степеней свободы ошибки	Разность между размером выборки и числом оцениваемых параметров модели
12 Число степеней свободы модели	Число независимо варьируемых <u>переменных</u> модели
9.0 F-статистика	Статистика F-теста для оценки значимости модели
9.0 Р-значение модели	Оценка <u>значимости</u> модели. Чем меньше Р-значение, тем меньше вероятность нулевой гипотезы о том, что нет статистически значимой связи между зависимой и независимыми переменными. Вычисляется по F-статистике.
9.0 Критерий Акаике	Критерий используется для сравнения моделей с разным числом параметров, когда требуется выбрать наилучший набор объясняющих переменных
9.0 Критерий Акаике (скорр.)	Модифицированный критерий Акаике, применяемый для выборок малого размера, когда отношение числа содержащихся в ней примеров к числу параметров модели меньше 40
9.0 Критерий Байеса	Критерий основан на использовании функции правдоподобия и тесно связан с информационным критерием Акаике

Показатель	Описание
9.0 Критерий Ханнана-Куинна	Критерий основан на использовании функции правдоподобия и является альтернативой критериям Акаике и Байеса

Коэффициенты регрессии

Описание коэффициентов представлено в таблице:

Коэффициент	Описание
Коэффициент	Оценка значений коэффициентов при независимых переменных
Стандартная ошибка	Стандартная ошибка оценки коэффициента
T-статистика	Статистика Т-критерия Стьюдента для оценки значимости переменной
Р-значение	Оценка значимости переменной. Чем меньше Р-значение, тем меньше вероятность нулевой гипотезы о том, что коэффициент при независимой переменной равен 0. Вычисляется по Т-статистике.
Нижняя граница ДИ	Нижняя граница доверительного интервала
Верхняя граница ДИ	Верхняя граница доверительного интервала

Коэффициенты регрессии могут быть представлены в виде  *Таблицы* (см. рисунок 1) или  *Дерева* (см. рисунок 2).

Атрибут 	
ab	Возрастная группа (до 40)
ab	Возрастная группа (старше 60)
12	Длительность регистрации (мес.) (10)
12	Длительность регистрации (мес.) (20)
9.0	Константа
ab	Пол (муж)
ab	Сегмент достатка (до 20 т.р.)
ab	Сегмент достатка (свыше 80 т.р.)
ab	Статус (Бронзовый)
ab	Статус (Золотой)
ab	Статус (Платиновый)

Рисунок 1. Режим отображения Таблица

Атрибут 	
 ab	Возрастная группа
...	ab до 40
...	ab старше 60
 12	Длительность регистрации (мес.)
...	12 10
...	12 20
...	9.0 Константа
 ab	Пол
...	ab муж
 ab	Сегмент достатка
...	ab до 20 т.р.
...	ab свыше 80 т.р.
 ab	Статус
...	ab Бронзовый
...	ab Золотой
...	ab Платиновый

Рисунок 2. Режим отображения Дерево

Шаги построения

Область  Шаги построения состоит из следующих полей:

Поле	Описание
Модель	Дерево, в котором отображается пошаговый процесс отбора факторов для построения финальной модели
Показатель	Значение выбранного показателя для текущей модели
Изменения поля	Отображает добавлен или удален фактор ("+" означает, что фактор добавлен в модель, а "-" показывает, что фактор удален из модели)
Поля	Текущие поля модели

Существует три вида моделей:

- **Нулевая модель** — начальная модель работы алгоритма.
- **Промежуточная модель** — модель, которую создал алгоритм в процессе обучения.
- **Финальная модель** — модель, которую алгоритм посчитал наилучшей.

Доступные для выбора показатели:

- F-тест (Р-значение)
- Коэф-т детерминации
- Коэф-т детерминации (скорр.)
- Критерий Акаике
- Критерий Акаике (скорр.)
- Критерий Байеса
- Критерий Ханнана-Куинна