

ЛЕКЦИЯ 8. Оценка регрессионной модели: что показывает R^2 и как понять, хорошая ли у вас модель

Введение

На предыдущих занятиях вы научились:

строить простую, полиномиальную и множественную регрессию,
использовать F-тест для проверки значимости,
добавлять качественные переменные,
интерпретировать коэффициенты.

Вы уже умеете **получать формулу**, но теперь пора научиться **оценивать, насколько она действительно полезна**.

Потому что регрессия — это не математика ради уравнения.

Это **инструмент предсказания и объяснения**. И нам важно знать:

Насколько хорошо эта модель объясняет данные?

Основной вопрос лекции:

“Модель построена. А она вообще работает?”

Что такое R^2 (коэффициент детерминации)?

Это **главный показатель качества линейной модели**.

Он показывает, **какую долю колебаний зависимой переменной у модель объясняет**.

 **Формула:**

$$R^2 = SSR / SST$$

Где:

SSR — сумма квадратов, объяснённых моделью (Regression)
SST — общая сумма квадратов (Total variation)

Проще:

R^2 = Насколько хорошо модель "приближается" к реальности

Интерпретация R^2

Значение R^2	Что это значит
0.9 – 1.0	Модель объясняет почти всё. Очень сильная.
0.7 – 0.89	Хорошо объясняет. Рабочая модель.
0.4 – 0.69	Средняя. Возможно, чего-то не хватает.
0.1 – 0.39	Слабая. Проверьте переменные и гипотезу.
< 0.1	Почти не объясняет. Возможно, связи нет.

Важно:

R^2 всегда от 0 до 1

- Не может быть отрицательным
- Но может быть **высоким даже у плохой модели**, если есть мультиколлинеарность
- Или **низким у модели, которая всё объясняет**, если в данных высокий шум

Пример

Гипотеза: Стресс влияет на успеваемость.

Построили модель:

Оценка = $70 - 3.5 * \text{Стресс}$

$R^2 = 0.62$

Вывод:

- 62% колебаний в оценках можно объяснить уровнем стресса
- Остальные 38% связаны с другими факторами (сон, преподаватель, мотивация и т.д.)

Как получить R^2 в Excel и Google Sheets

Google Sheets:

```
=LINEST(B2:B10, A2:A10, TRUE, TRUE)
```

R^2 будет в **четвёртом столбце второй строки** результата.

Excel (русский):

```
=ЛИНЕЙН(B2:B10; A2:A10; ИСТИНА; ИСТИНА)
```

Также вернёт массив, где R^2 отображается ниже коэффициентов.

⚠ Подводные камни:

✗ $R^2 \neq$ "насколько точна модель"

Он не показывает, **насколько хорошо она предсказывает новые данные**

✗ $R^2 \neq p\text{-value}$

Он не говорит, **значима ли модель** — за это отвечает F-тест

✗ R^2 всегда растёт при добавлении переменных

Даже если переменная не влияет — R^2 будет расти!



Поэтому используйте **корректированный R^2 (adjusted R^2)**

Что такое Adjusted R^2 ?

Это улучшенная версия R^2 , которая

- **учитывает количество переменных**
- "штрафует" модель за добавление слабых признаков

Часто используется при **множественной регрессии**.

Сравнение R^2 и Adjusted R^2

Показатель	Значение
R^2	Общая объясняющая сила модели
Adjusted R^2	Объясняющая сила с учётом числа переменных

Использование ИИ

Инструмент	Для чего использовать
Excel Copilot	Автоматически выдает R^2 , визуализации
ChatGPT	Помогает интерпретировать R^2 и сравнить модели
Google Sheets Explore	Показывает объяснённую дисперсию сразу на графике

Запрещено:

- Говорить: “модель хорошая”, основываясь только на R^2
- Не проверять F-тест и p-value Оставлять переменные ради “повышения R^2 ”
- Игнорировать Adjusted R^2 при множественной регрессии

Вывод

- R^2 — это не про красоту формулы.
- Это про то, **насколько близко ваша модель к реальности.**
- Вы научились не просто строить модель, а **оценивать её качество**, и это ключевой навык аналитика.