

ЛЕКЦИЯ 9. Анализ остатков и ошибки спецификации модели

Введение

К этому моменту вы уже прошли большой путь:

Вы умеете строить регрессионную модель, интерпретировать коэффициенты, оценивать значимость (через F-тест), учитывать категориальные переменные и даже анализировать R^2 .

Но настоящая аналитика начинается тогда, когда вы задаёте себе **вопросы после модели**:

- А где именно она ошибается?
- А почему она плохо предсказывает одних респондентов и хорошо других?
- А есть ли в ней что-то “подозрительное”?

Ответы на все эти вопросы лежат в **остатках модели**.

Что такое остатки?

Остаток (residual) — это **разница между фактическим и предсказанным значением y** :

$$\text{Остаток}_i = y_i - \hat{y}_i$$

Где:

- y_i — реальное значение
- \hat{y}_i — значение, предсказанное моделью

Зачем анализировать остатки?

Остатки помогают:

проверить, соответствует ли модель реальности,

найти выбросы и аномалии,
увидеть, не нарушаются ли **допущения линейной регрессии**,
понять, где модель “врёт” — и почему.

Пример

Респондент	Оценка фактическая (y)	Оценка по модели (\hat{y})	Остаток (e)
1	75	78	-3
2	82	81	+1
3	65	70	-5
4	90	89	+1
5	55	60	-5

Как интерпретировать остатки

Ситуация	Что это значит
Остатки случайные и “разбросанные”	Модель адекватна, допущения соблюдены
Остатки растут или падают с x	Нелинейность, модель плохо подогнана
Остатки сконцентрированы слева/справа	Модель систематически ошибается
Один остаток очень большой	Возможен выброс, стоит проверить

Как визуализировать остатки

Google Sheets / Excel:

1. Рассчитайте \hat{y} (по формуле модели):

$$y_{\text{pred}} = b_0 + b_1 * x_1 + b_2 * x_2 + \dots$$

2. Рассчитайте остаток = $y - y_{\text{pred}}$

3. Постройте график:

- по оси X: предсказанные значения (\hat{y})
- по оси Y: остатки (e)

Идеальный случай: **остатки идут “хаосом” вдоль нуля**

Ошибки спецификации модели

Если остатки показывают **систематическое отклонение**, возможно:

- модель **пропустила важную переменную**,
- **форма зависимости не линейная**,
- есть **нелогично закодированные переменные**,
- присутствуют **выбросы**, которые искажают результаты.

Примеры ошибок спецификации

Проблема	Как проявляется в остатках
Пропущенный фактор	Остатки "смещены" в одну сторону
Нелинейность	Остатки образуют дугу или волну
Выброс	Один остаток очень большой
Категориальная переменная не учтена	Остатки группируются по "половому" или "поведенческому" признаку

Как исправить:

Проблема	Возможное решение
Пропущенная переменная	Добавить её в модель
Нелинейность	Применить полиномиальную регрессию
Выброс	Проанализировать, удалить или обработать
Ошибка кодировки	Перепроверить шкалы и логики 0/1

Использование ИИ

Инструмент	Зачем использовать
ChatGPT	Объяснит, что значат остатки в вашем графике
Excel Copilot	Построит график остатков и выявит отклонения
Notion AI	Поможет сформулировать выводы на основе остатков

Запрещено

Не проверять остатки вообще

Говорить “модель хорошая”, глядя только на R^2

Игнорировать систематические ошибки

Оставлять выбросы “потому что они есть у всех”

Вывод

Анализ остатков — это как проверка двигателя под капотом.

Модель может выглядеть красиво, но если остатки дают сбой — она **нечестна**.

Аналитик, который читает остатки, — это аналитик, который **понимает модель, а не просто ею пользуется**.