

ЛЕКЦИЯ 14.

Гетероскедастичность и автокорреляция: когда ошибки модели становятся предсказуемыми

Введение

Вы уже узнали, что модель — это не только формула, но и **совокупность условий**.

Если эти условия нарушаются — модель начинает “врать”.

В прошлой лекции мы говорили о **нормальности остатков** — важнейшем условии для корректности выводов.

Сегодня — продолжим проверку **качества ошибок модели**, а именно:

- **Гетероскедастичность** — когда разброс ошибок непостоянен
- **Автокорреляция** — когда ошибки “запоминают” предыдущие ошибки


1. Что такое гетероскедастичность?

 Это ситуация, когда **дисперсия остатков зависит от значений x или \hat{y}** .

Иными словами:

ошибки “растут” или “сжимаются” в зависимости от предсказаний

Как выглядит на графике остатков:

Вид графика остатков	Интерпретация
“Веер” вправо	Ошибки растут с \hat{y}
“Веер” влево	Ошибки уменьшаются
“Бабочка”	Ошибки сперва растут, потом падают
Равномерный шум вдоль нуля	 Гомоскедастичность — всё ок

Почему это важно?

Гетероскедастичность:

- делает **стандартные ошибки некорректными**,
- искажает t - и F -статистику,
- приводит к **ложным выводам**.


Как проверить гетероскедастичность

1. Постройте график остатков от предсказанных значений (\hat{y})
2. Визуально посмотрите: нет ли "веера", "трубы", "конуса"
3. (Опционально) используйте **тест Бреуша–Пагана** или **тест Уайта** (в Python/R)

Что делать, если она есть?

Подход	Что делать
Логарифмирование переменной y	$\ln(y)$ или \sqrt{y} уменьшает разброс
Модель с взвешенной регрессией	(в продвинутых инструментах)
Разделить выборку на группы	Например: "малозэкранные" и "многоэкранные"
Использовать робастные ошибки	Только в Python/R (если доступно)

2. Что такое автокорреляция?

 Это ситуация, когда **ошибки модели следуют шаблону во времени или порядке наблюдений**.

Пример:

Студенты сдавали опрос по порядку → ошибки скапливаются у первых/последних

Модель по продажам по месяцам → ошибка января похожа на ошибку февраля

Как выглядит:

Остатки "скачут" как волна

Присутствует **регулярный паттерн**

Модель **недостаточно быстро подстраивается**

Почему это важно?

Автокорреляция:

делает ошибки **предсказуемыми**

нарушает **независимость наблюдений** (одно из ключевых допущений!)

может исказить доверие к модели

Как проверить автокорреляцию?

1. Построить график остатков в порядке наблюдений

2. Использовать **тест Дарбина–Уотсона (DW)**

- $DW \approx 2 \rightarrow$ ошибок нет
- $DW < 1.5 \rightarrow$ положительная автокорреляция
- $DW > 2.5 \rightarrow$ отрицательная автокорреляция

(Доступно в Python, R, STATA, и иногда в Excel через надстройки)

Что делать?

Сценарий	Решение
Данные имеют временной характер	Используйте временные модели (ARIMA, и т.д.)
Ошибки идут "волнами"	Добавьте переменные, отражающие динамику
Есть "группы" в данных	Добавьте категориальную переменную, которая их разделит

Как объяснить студентам на пальцах:

Гетероскедастичность — когда у кого-то ошибка = ± 1 , а у кого-то = ± 10

Автокорреляция — когда ошибка сегодня зависит от вчерашней

Использование ИИ

Инструмент	Как помогает
------------	--------------

ChatGPT	Интерпретирует график остатков
Excel Copilot	Показывает "веерообразные" отклонения
Notion AI	Формулирует вывод о качестве модели

Запрещено:

- Не проверять остатки на гетероскедастичность
- Строить модель на временных данных без учёта времени
- Интерпретировать результат модели без учёта ошибок
- Считать, что R^2 и F — "всё, что нужно"

Вывод

Если вы видите гетероскедастичность или автокорреляцию — **модель теряет свою достоверность.**

Вы как исследователь обязаны не только **строить уравнение**, но и **проверять, честна ли модель по отношению к данным.**