

Гетероскедастичность

Линейная регрессия

Гетероскедастичность — нарушение одного из предположений об ошибках (об одинаковости или постоянстве их дисперсии).

Дисперсия ошибок не постоянна, а каким-то образом зависит от факторов (независимых переменных) в регрессии.

Гетероскедастичность

Почему это плохо?

Оценки параметров модели при наличии выраженной гетероскедастичности будут менее эффективны (занижает ошибки, завышает t-статистики).

Причины возникновения

- Наличие смеси в данных.
- Особенности данных (значения переменных существенно различаются для разных наблюдений).

Как обнаружить гетероскедастичность?

- Анализ совместных распределений признаков на исходных данных.
- Анализ остатков и их распределений.
- Коэффициент ранговой корреляции Спирмена (остатки-факторы):

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

в числителе разница рангов X_i и ε_i

- Другие тесты.

Как бороться с гетероскедастичностью?

- Преобразовать переменные (к примеру, логарифмировать).
- Использовать взвешенный МНК (наблюдаемые значения корректируются на соответствующие значения стандартного отклонения остатков).
- Сразу вычислять стандартные ошибки с поправкой на гетероскедастичность.