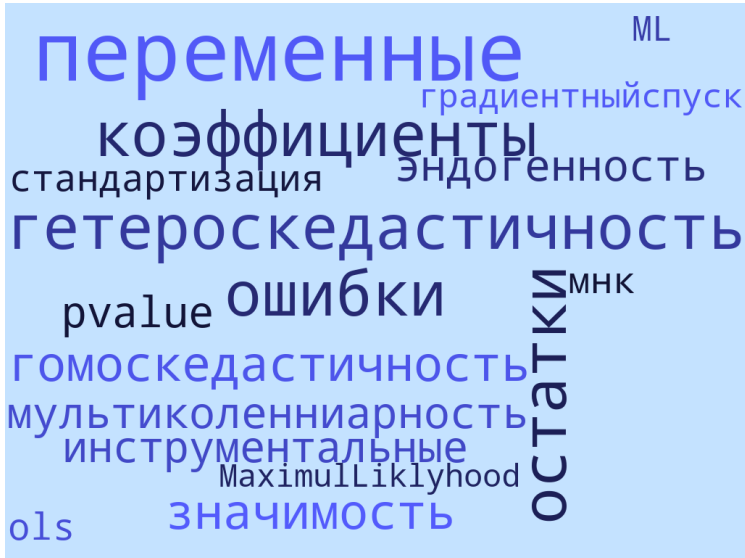


Regression. Beginning.

simple examples

Chuvakin Sergey

May 3, 2019



Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлинейность

Гетероскедастичность

Эндогенность

Значимость

Какой план

Нужна ли нам регрессия?

Что такое регрессия, в принципе..

Методы нахождения коэффициентов (Estimators)

Методы оптимизации

Проблемы и диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость

Нужна ли нам регрессия?

- ▶ Как устроен мир? (Inferential statistics)
- ▶ Что будет завтра (Machine learning)

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость

Что такое регрессия, в принципе..

$$y = \beta X$$

- ▶ y - зависимая, то что нам интересно
- ▶ β - наши параметры, коэффициенты
- ▶ X - набор независимых переменных, предикторов, регрессоры

$$\hat{y} = \hat{\beta}X + \hat{\epsilon}$$

- ▶ $\hat{\epsilon}$ - ошибка, остаток, «человеческий фактор», «случайность»

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость

Методы нахождения коэффициентов (Estimators)

- ▶ МНК (OLS): $\beta = \Sigma(\hat{y} - y)^2 \rightarrow \min$
- ▶ ММП (ML): $\theta = P(X|\theta) \rightarrow \max$ (сильно огрубленно)
- ▶ Moments method (MM)
- ▶ Баясовские решения

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость

Методы оптимизации

- ▶ Аналитическое решение: $\beta = (X^T X)^{-1} X^T y$
- ▶ Градиентный спуск (GD)
- ▶ Стохастический градиентный спуск (SGD)
- ▶ MCMC

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

**Методы
оптимизации**

Проблемы и
диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость

Проблемы и диагностика

- ▶ Мультиколлиниарность
- ▶ Гетероскедастичность
- ▶ Эндогенность

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

**Проблемы и
диагностика**

Мультиколлиниарность

Гетероскедастичность

Эндогенность

Значимость

Мультиколлиниарность

- ▶ Зависимость векторов
- ▶ Мешает оптимизации, строгая коллиниарность «вырождает» матрицу X
- ▶ Не дает сделать валидные выводы
- ▶ $vif < 5$

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам регрессия?

Что такое регрессия, в принципе..

Методы нахождения коэффициентов (Estimators)

Методы оптимизации

Проблемы и диагностика

Мультиколлиниарность

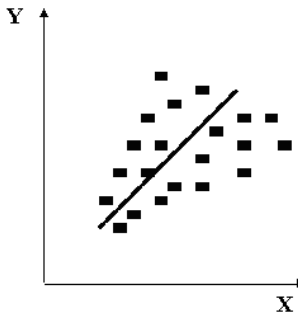
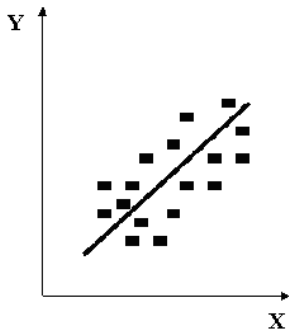
Гетероскедастичность

Эндогенность

Значимость

Гетероскедастичность

► Рассеянность остатков



Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость

Эндогенность

- ▶ Строго говоря: $\text{Cov}(\epsilon_i, x_i) \neq 0$
- ▶ Для прогнозирования абсолютно не важно
- ▶ Причины
 - ▶ ошибка измерения регрессора ($x^* = x_i + u_i$)
 - ▶ пропущенный регрессор
 - ▶ взаимное влияние регрессоров
- ▶ Сложность при интерпретации
- ▶ Двухшаговый МНК (IV)

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлиниарность

Гетероскедастичность

Эндогенность

Значимость

Значимость

1. Для прогнозирования не важно
2. Рассчитывается на основе стандартной ошибки для β
→ t-value
3. Для удобства были придуманы p-value

Regression.
Beginning.

Chuvakin Sergey

Нужна ли нам
регрессия?

Что такое
регрессия, в
принципе..

Методы
нахождения
коэффициентов
(Estimators)

Методы
оптимизации

Проблемы и
диагностика

Мультиколлинеарность

Гетероскедастичность

Эндогенность

Значимость