

Исходные предположения регрессионного анализа

$$y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \dots + \beta_k \cdot x_{k,i} + \epsilon_i, \quad i = 1 \dots n$$

Предположения

1. (Сферический конь в вакууме)

При заданных значениях переменных x_j $j=1, \dots, k$ на отклик y не оказывают влияние никакие другие предикторы. Влияние других предикторов учитывается случайной возмущающей переменной. При этом $E\epsilon = 0$.

2. Гомоскедастичность.

Дисперсия случайной переменной ϵ должна быть для всех наблюдений одинакова и постоянна: $D\epsilon_i = \text{const} \quad i = 1 \dots n$

3. Независимость ошибок

Значения случайной переменной ϵ попарно некоррелированы или, еще более сильная предпосылка, они попарно независимы:

$$\text{corr}(\epsilon_i, \epsilon_j) = 0, \quad i, j = 1, \dots, n, i \neq j$$

4 Таблица с данными не «плоская»

Число наблюдений должно превышать число предикторов $n > k$

5

Объясняющие переменные X_j не коррелируют с ошибкой ϵ

$$\text{corr}(X_i, \epsilon) = 0$$

6 Нормальное распределение ошибок

Переменная ϵ нормально распределена.

Все обычно хорошо, если ϵ представляет собой суммарный эффект от большого числа незначительных некоррелированных влияющих на отклик переменных

Как следствие, отклик y распределен нормально.