

Тимонов Артем Сергеевич
Первая повторная промежуточная
аттестация Тимона
компьютерного моделирования.

Регрессионный анализ.

Метод наименьших квадратов

Определение

Регрессионный анализ - это
статистический метод, который
позволяет моделировать зависимость
между переменными.

Метод наименьших квадратов -
- это метод оценки параметров
регрессионной модели, который
минимизирует сумму квадратов
отклонений между наблюдаемыми
значениями и предсказанными
значениями модели.

Мат. формулировка

Допустим у нас есть регрессионная модель:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

y - зависимая переменная

x - независимая переменная

β_0, β_1 - параметры модели

ε - случайная ошибка

Цель метода наименьших квадратов - найти такие значения параметров β_0 и β_1 , которые минимизируют сумму квадратов отклонений между наблюдаемыми значениями y и предсказанными значениями модели:

$$S = \sum (y_i - (\beta_0 + \beta_1 x_i))^2$$

y_i - наблюдаемое значение y

x_i - наблюдаемое значение x

i - номер наблюдения

Метод наименьших квадратов

1. Вычислить сумму квадратов отклонений S

2. Найти производные суммы квадратов отклонений по параметрам β_0 и β_1

3. Решить систему уравнений, полученную из условия минимума суммы квадратов отклонений

Пример.
Допустим у нас есть данные
зависимости между кол-вом
проданных автомобилей и их
ценой.

цена	продажи
10000	50
12000	40
15000	30
18000	20
20000	10

Задача

1. $S = \sum (y_i - (\beta_0 + \beta_1 x_i))^2$

2. Далее надо найти производные
суммы квадратов отклонений
по параметрам β_0 и β_1

$$\frac{S}{\beta_0} = -2 \sum (y_i - (\beta_0 + \beta_1 x_i)) = \frac{S}{\beta_1} =$$

$$= -2 \sum x_i (y_i - (\beta_0 + \beta_1 x_i)) = 0$$

3. Далее вычисляем сумму квадратов отклонений S и найдем производные по параметрам

β_0 и β_1

$$S = \sum (y_i - (\beta_0 + \beta_1 x_i))^2 = (50 - (\beta_0 + \beta_1 \cdot 10000))^2 + (40 - (\beta_0 + \beta_1 \cdot 12000))^2 + (30 - (\beta_0 + \beta_1 \cdot 15000))^2 + (20 - (\beta_0 + \beta_1 \cdot 18000))^2 + (10 - (\beta_0 + \beta_1 \cdot 20000))^2$$

~~Продолжение~~

Решение системы уравнений даёт нам параметры модели

$$\beta_0 = 200, \beta_1 = -0,01$$

таким образом, рекурсивная
модель имеет вид -

$$\text{Продажи} = 200 - 0,01 \cdot \text{цена}$$

Это означает что существует
обратная зависимость между
ценой и продажами - при
увеличении цены продажи
уменьшаются.