

Линейная регрессия

1. Вывести непосредственно формулу для МНК-оценки параметра θ в регрессионной модели вида $Y_i = \theta X_i + \varepsilon_i, i = 1, \dots, n$.
2. Имеются данные о себестоимости Y (в у.е.) экземпляра книги в зависимости от тиража X (в тыс. экземпляров). Данные представлены в таблице:

тираж	1	2	3	4	5
себестоимость	6	5	4	4	3

Предполагается, что справедлива модель вида $Y_i = a + bX_i + \varepsilon_i, i = 1, \dots, n$, где вектор $\varepsilon \sim N(0, \sigma^2 I)$. Постройте:

- 1) МНК-оценки параметров a и b (с использованием и без использования матричной техники вычислений);
- 2) оценку дисперсии σ_ε^2 ;
- 3) проверьте гипотезу $H_0: b = 0$;

Постройте точечную и интервальную (уровня надёжности 0.95) оценку для себестоимости, если тираж $X=6$.

3. Имеются данные (из «Основы химии» Д.И. Менделеева) о количестве (Y) азотнатриевой соли NaNO_3 , которое можно растворить в 100 граммах воды в зависимости от температуры t

t	0	4	10	15	21	29	36	51	68
Y	66,7	71,0	76,3	80,6	85,7	92,9	99,4	113,6	125,1

Оцените по МНК параметры регрессионной модели $Y_i = \theta_0 + \theta_1 t_i + \theta_2 t_i^2 + \varepsilon_i$, где вектор $\varepsilon \sim N(0, \sigma^2 I)$. Проверьте гипотезу о том, что параметр $\theta_2 = 0$. Следует ли для описания данного явления перейти к более простой регрессионной модели? Оцените коэффициент детерминации для «короткой» и «длинной» модели.

Домашнее задание

1. В таблице указана динамика веса поросят

Возраст X (недели)	0	1	2	3	4	5	6
Вес Y (кг)	1,2	2,5	3,9	5,2	6,4	7,7	9,2

Предполагается, что справедлива модель вида $Y_i = a + bX_i + \varepsilon_i$, где вектор $\varepsilon \sim N(0, \sigma^2 I)$. Постройте:

- 1) МНК-оценки параметров a и b ;
- 2) оценку дисперсии σ_ε^2 ;
- 3) проверьте гипотезу $H_0: b = 0$;
- 4) постройте точечную и интервальную (уровня надёжности 0.95) веса в точках $X_0=3$ и $X_0=6$;
- 5) сделайте прогноз для значения Y в точке $X=8$.

2. Бюджетное обследование пяти случайным образом выбранных семей дало следующие результаты (в тыс. у.е.):

Семья	Накопления (Y)	Доход (X)	Имущество (Z)
1	3,0	40	60
2	6,0	55	36
3	5,0	45	36
4	3,5	30	15
5	1,5	30	90

Предполагается, что справедлива модель вида $Y_i = a + b_1X_i + b_2Z + \varepsilon_i$, где вектор $\varepsilon \sim N(0, \sigma^2 I)$.

- 1) Оцените параметры модели;
- 2) Спрогнозируйте накопления семьи, имеющей доход 40 тыс у.е. и имущество 25 тыс у.е.;
- 3) Предположим, что доход семьи вырос на 10 тыс у.е. в то время, как стоимость имущества не изменилась. Оцените, как вырастут накопления семьи;
- 4) Оцените, как вырастут накопления семьи, если её доход увеличился на 5 тыс, а имущество на 15 тысяч;
- 5) Проверьте гипотезы (на уровне значимости 0.05) о том, что а) b_1 и b_2 равны нулю (т.е. модель является тривиальной), б) $b_1 = 0$ (т.е. величина дохода не существенна), в) $b_2 = 0$ (т.е. стоимость имущества не существенна), г) $b_1 = 0,8$ (такое значение было вычислено согласно данным по другой стране)