

1. Исходные данные в файле *lin-reg.csv*, первый столбец – output, оставшиеся – переменные, количество наблюдений – 100. Требуется:

- (а) Найти оценку наименьших квадратов коэффициентов c_0, \dots, c_{10} линейной модели

$$y_i = c_0 + \sum_{j=1}^{10} c_j x_{ij} + \varepsilon_i, \quad i \in \{1, \dots, 100\},$$

$\varepsilon = (\varepsilon_1, \dots, \varepsilon_{100})^T \sim \mathcal{N}(0, \sigma^2 E_{100})$, и оценку остаточной дисперсии σ^2 .

- (b) Вычислить 95-процентные доверительные интервалы для параметров c_0, \dots, c_{10} и остаточной дисперсии σ^2 .
(c) Проверьте гипотезы вида $H_0^{(k)} : c_k = 0, k \in \{0, \dots, 10\}$.
(d) Проверьте гипотезу $H : c_i \% 5 = c_{5+(i \% 5)} = c_{10} = 0, i$ – номер в списке.

Вычисления реализовать ”руками”, разрешается пользоваться реализацией методов оптимизации функции и вычислений с матрицами, в том числе и приближенными. Гипотезы проверять через проверку попадания статистики критерия в критическую область и через вычисление p-value.

2. Исходные данные в файле *one-way-anova.txt*, в файле 10 строчек, каждая строчка из 100 наблюдений соответствует одному уровню фактора. Требуется для модели однофакторного дисперсионного анализа проверить гипотезу о равенстве среднего влияния каждого фактора (через проверку принадлежности статистики критерия критической области и вычисления p-value). Вычисления реализовать ”руками”.