

# Расчётно-графическая работа № 4

## Задание 1

Задание представлено в 4 вариантах. Для каждого варианта требуется построить линейную модель (предполагая нормальность распределения ошибок с нулевым средним, некоррелированность компонент, гомоскедастичность), вычислить оценки коэффициентов модели и остаточной дисперсии, построить для них доверительные интервалы, вычислить коэффициент детерминации, проверить указанные в условии гипотезы с помощью построенной линейной модели.

**Указание:** из встроенных функций разрешается пользоваться квантильными функциями и средствами для квадратичной оптимизации (иными словами, готовую обертку для построения линейной модели не использовать, максимум можете сравнить вашу реализацию с готовой)

**Вариант 1.** В файле <https://drive.google.com/file/d/1vv2jGNp6EO8HHRoscDRQU90faR3j8iTN/view> представлены данные о продажах различных авто.

1. Постройте линейную модель, где в качестве независимых переменных выступают расход в городе, расход на шоссе, мощность (вместе со свободным коэффициентом), зависимой – цена.
2. Проверьте следующие подозрения:
  - Чем больше мощность, тем больше цена
  - Цена изменяется в зависимости от расхода в городе
  - Проверьте гипотезу  $H_0$  о равенстве одновременно нулю коэффициентов при расходе в городе и расходе на шоссе против альтернативы  $H_1 = \overline{H_0}$

**Вариант 2.** В файле [https://drive.google.com/file/d/1O4rFr9xg9aFmkjx4-hl\\_XOc5O9q65\\_EW/view](https://drive.google.com/file/d/1O4rFr9xg9aFmkjx4-hl_XOc5O9q65_EW/view) представлены данные о мобильных телефонах.

1. Постройте линейную модель, где в качестве независимых переменных выступают высота, ширина экрана (атрибуты `sc_h` и `sc_w`) и емкость аккумулятора (вместе со свободным коэффициентом), зависимой – масса телефона.
2. Проверьте следующие подозрения:
  - Чем больше высота экрана, тем больше масса
  - Чем больше ширина экрана, тем больше масса
  - Проверьте гипотезу  $H_0$  о равенстве одновременно нулю коэффициентов при ширине экрана и емкости аккумулятора против альтернативы  $H_1 = \overline{H_0}$

**Вариант 3.** В файле [https://drive.google.com/file/d/1KRbKtVb6Xkyc8\\_2gKT9G6N5N\\_yD7FQXC/view](https://drive.google.com/file/d/1KRbKtVb6Xkyc8_2gKT9G6N5N_yD7FQXC/view) приведены данные о музыкальных произведениях.

1. Постройте линейную модель, где в качестве независимых переменных выступают продолжительность, "танцевальность" и энергичность (вместе со свободным коэффициентом), зависимой – популярность.
2. Проверьте следующие подозрения:
  - Чем больше энергичность, тем больше популярность
  - Популярность зависит от продолжительности
  - Проверьте гипотезу  $H_0$  о равенстве одновременно нулю коэффициентов при энергичности и "танцевальности" против альтернативы  $H_1 = \overline{H_0}$

**Вариант 4** В файле <https://drive.google.com/file/d/1cx0pshptDSVmaWLJCBGS9jIJ2g-VRgT/view> приведены данные о цене на недвижимость где-то в окрестности Сиэтла

1. Постройте линейную модель, где в качестве независимых переменных выступают «жилая» площадь, «sqft\_lot», «sqft\_above» (вместе со свободным коэффициентом), зависимой – цена на недвижимость.
2. Проверьте следующие подозрения:
  - Чем больше «жилая» площадь, тем больше цена
  - Цена зависит от «sqft\_lot»
  - Проверьте гипотезу  $H_0$  о равенстве одновременно нулю коэффициентов при «жилой» площади и «sqft\_above» против альтернативы  $H_1 = \overline{H_0}$

## Задание 2

Для каждого варианта требуется проверить гипотезу о равенстве средних на каждом уровне фактора с помощью модели однофакторного дисперсионного анализа

**Указание:** реализовать самим

**Вариант 1.** В файле <https://drive.google.com/file/d/1CSChEMzjberRwgcf90BBu-J6uxMg-Qf7/view> представлены данные об ирисках. Фактор – подвид. Выходная переменная – суммарная площадь (точнее оценка площади) чашелистика и лепестка.

**Вариант 2.** В файле [https://drive.google.com/file/d/14L\\_y0LOAebuuqh8PllOw64cJQwVkmlV6/view](https://drive.google.com/file/d/14L_y0LOAebuuqh8PllOw64cJQwVkmlV6/view) представлены данные о сдаче экзаменов. Фактор – этническая/национальная группа. Выходная переменная – суммарный балл за все три экзамена.

**Вариант 3.** В файле <https://drive.google.com/file/d/1gzPRqj7gZetjsipo3xpogYGL76enZDNO/view> приведены данные о некоторых привычках и физиологических/антропологических показателях. Фактор – курит/не курит. Выходная переменная – индекс массы тела

**Вариант 4.** В файле [https://drive.google.com/file/d/1O4rFr9xg9aFmkjx4-hl\\_XOc5O9q65\\_EW/view](https://drive.google.com/file/d/1O4rFr9xg9aFmkjx4-hl_XOc5O9q65_EW/view) приведены данные о мобильных телефонах. Фактор – ценовая категория, выходная переменная – емкость аккумулятора