

Лабораторная работа №5

Задание представлено в 4 вариантах. Для каждого варианта требуется построить линейную модель, 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)

Вариант 1. В файле *cars93.csv* 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
 - Цена зависит от расхода в городе и от расхода на шоссе

Вариант 2. В файле *mobile_phones.csv* 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
 - Масса зависит и от ширины экрана, и от емкости аккумулятора

Вариант 3. В файле *MEN_SHOES_.csv* приведены данные о мужской обуви.

1. Постройте линейную модель, где в качестве независимых переменных выступают количество проданных экземпляров и цена (вместе со свободным коэффициентом), зависимой – рейтинг.
2. Проверьте следующие подозрения:
 - Чем больше продажи, тем больше рейтинг
 - Рейтинг за зависит от цены
 - Рейтинг зависит и от цены, и от количества проданных экземпляров

Вариант 4. В файле *song_data.csv* 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
 - Популярность зависит от энергичности и "танцевальности"

Ключевые понятия:

- Линейная регрессия. Основные предположения
- Метод наименьших квадратов и его свойства
- Основная теорема о линейной регрессии. Следствия из нее
- Остаточная дисперсия. Коэффициент детерминации.