**Лабораторная работа №4**

**Задача 1**

По данным файла «dataset\_11508\_12.txt» постройте линейную регрессию, где - первая переменная - зависимая, вторая - независимая. В ответе выведите значения регрессионных коэффициентов  **intercept и  slope.**

**Задача 2**

Воспользуемся уже знакомыми данными diamonds. Только для бриллиантов класса Ideal (переменная cut) c числом карат равным 0.46 (переменная carat) постройте линейную регрессию, где в качестве зависимой переменной выступает price, в качестве предиктора - переменная  depth. Сохраните коэффициенты регрессии в переменную fit\_coef.

**!!!Это задание нужно решить, не используя цикл for().**

**Задача 3**

Напишите функцию fill\_na, которая принимает на вход данные с тремя переменными:

**x\_1** -  числовой вектор

**x\_2** - числовой вектор

**y** - числовой вектор с пропущенными значениями.

На первом этапе, используя только наблюдения, в которых нет пропущенных значений, мы построим регрессионную модель (без взаимодействий), где **y** — зависимая переменная, **x\_1** и**x\_2** — **независимые переменные**. Затем, используя построенную модель, мы заполним пропущенные значения предсказаниями модели.

Функция должна возвращать dataframe c новой переменной **y\_full.**Сохраните в нее переменную **y**, в которой пропущенные значения заполнены предсказанными значениями построенной модели.

  x\_1 x\_2 y y\_full

1   9 22 14 14.00000

2 14 35 9 9.00000

3 10 45 NA 13.12727

4  6 35 NA 13.30132

5 11 36 16 16.00000

6  5 27 11 11.00000

7  9 34 NA 12.83632

8  6 25 NA 12.90783

9 11 38 13 13.00000

10 12 23 NA 11.97784

**Задача 4**

В переменной df сохраните subset данных mtcars только с переменными "wt", "mpg", "disp", "drat", "hp". Воспользуйтесь множественным регрессионным анализом, чтобы предсказать вес машины (переменная "wt"). Выберите такую комбинацию независимых переменных (из "mpg", "disp", "drat", "hp"), чтобы значение R^2 adjusted было наибольшим. Взаимодействия факторов учитывать не надо.

**Задача 5**

Напишите функцию normality\_test, которая получает на вход dataframe с количественными переменными, проверяет распределения каждой переменной на нормальность с помощью теста Шапиро-Уилка. Функция должна возвращать вектор со значениями p - value, полученного в результате проверки на нормальность каждой переменной. Названия элементов вектора должны совпадать с названиями переменных.

Пример работы функции:

> normality\_test(mtcars[,1:6])

   mpg       cyl       disp         hp       drat         wt

1.228814e-01 6.058338e-06 2.080657e-02 4.880824e-02 1.100608e-01 9.265499e-02

> normality\_test(iris[,-5])

Sepal.Length Sepal.Width Petal.Length Petal.Width

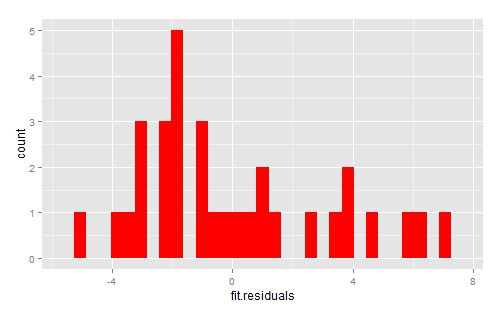
1.018116e-02 1.011543e-01 7.412263e-10 1.680465e-08

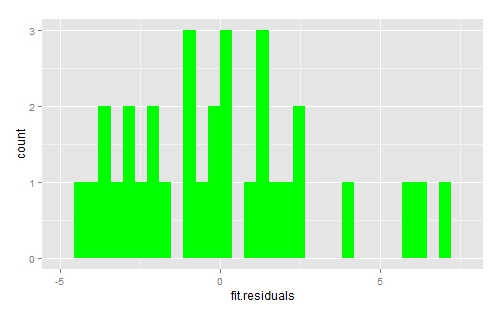
Опять же, обратите внимание функция должна работать корректно с различным количеством переменных и в независимости от их названий.

**Задача 6**

Напишите функцию resid\_norm, которая тестирует распределение остатков от модели на нормальность при помощи теста Шапиро-Уилка и создает гистограмму с красной заливкой "red", если распределение остатков значимо отличается от нормального (p < 0.05), и с зелёной заливкой "green" - если распределение остатков значимо не отличается от нормального.

На вход функция получает регрессионную модель. Функция возвращает переменную, в которой сохранен график гистограммы.





**Задача 7**

Ещё одной проблемой регрессионных моделей может стать мультиколлинеарность - ситуация, когда предикторы очень сильно коррелируют между собой. Иногда корреляция между двумя предикторами может достигать 1, например, когда два предиктора - это одна и та же переменная, измеренная в разных шкалах (x1 - рост в метрах, x2 - рост в сантиметрах)

Проверить данные на мультиколлинеарность можно по графику pairplot и посчитав корреляцию между всеми предикторами c помощью корреляции.

Напишите функцию high\_corr, которая принимает на вход датасет с произвольным числом количественных переменных и возвращает вектор с именами двух переменных с максимальным абсолютным значением коэффициента корреляции .

Примеры работы функции:

> high\_corr(swiss)

[1] "Examination" "Education"

> high\_corr(iris[,-5])

[1] "Petal.Length" "Petal.Width"

**Задача 8**

Используем данные **mtcars**. Сохраните в переменную логистическую регрессионную модель, где в качестве зависимой переменной выступает тип коробки передач (**am**), в качестве предикторов переменные**disp, vs, mpg**.

Значения **коэффициентов регрессии**сохраните в переменную log\_coef.