**Задания по файлу «cortest.R»**

**Задача №1**

Напишите функцию smart\_cor, которая получает на вход dataframe с двумя количественными переменными. Проверьте с помощью теста Шапиро-Уилка, что данные в обеих переменных принадлежат нормальному распределению.

Если хотя бы в одном векторе распределение переменной отличается от нормального (p - value меньше 0.05), то функция должна возвращать коэффициент корреляции Спирмена. (Числовой вектор из одного элемента).

Если в обоих векторах распределение переменных от нормального значимо не отличается, то функция должна возвращать коэффициент корреляции Пирсона.

**Задания по файлу «simple\_regr.R»**

**Задача 2**

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 3**

Воспользуемся уже знакомыми данными diamonds из библиотеки ggplot2. 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.

Памятка:

> fit <- lm(mpg ~ disp + wt, mtcars)

> fit$coefficients # коэффициенты модели

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

**Памятка**

cor.test(mtcars$mpg, mtcars$disp) # Расчет корреляции Пирсона

cor.test(~ mpg + disp, mtcars) # запись через формулу

cor.test(mtcars$mpg, mtcars$disp, method = "spearman") # Расчет корреляции Спирмена

cor.test(mtcars$mpg, mtcars$disp, method = "kendall") # Расчет корреляции Кендала

cor(iris[, -5]) # построение корреляционной матрицы

fit <- lm(mpg ~ disp, mtcars) # построение линейной регрессии

fit$coefficients # коэффициенты регрессии

fit$fitted.values # предсказанные значения зависимой переменной

При наличии одинаковых значений в переменных расчет непараметрических корреляций будет сопровождаться предупреждением о невозможности рассчитать точное значение p - value.

Если в ваших данных есть одинаковые наблюдения, но вы хотите рассчитать непараметрическую корреляцию, используйте функцию spearman\_test  из пакета coin

library(coin)

spearman\_test(~ mpg + disp, mtcars)

**Задания по файлу «multiple regression.R»**

**Задача 4**

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.

> test\_data <- read.csv("fill\_na\_test.csv")

> fill\_na(test\_data)

  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 5**

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 6**

В этом примере будем работать со встроенным датасетом *mtcars*. Переменная *am* говорит о том, какая коробка передач используется в машине: 0 - автоматическая, 1 - ручная. Сделаем эту переменную факторной.

mtcars$am <- factor(mtcars$am, labels = c('Automatic', 'Manual'))

Теперь постройте линейную модель, в которой в качестве зависимой переменной выступает расход топлива (*mpg*), а в качестве независимых - вес машины (*wt*) и коробка передач (модифицированная *am*), а также их взаимодействие. Выведите summary этой модели.  
Что отражает значение intercept в данной модели?