Белорусский государственный университет

информатики и радиоэлектроники

Лабораторная работа №7 по теме:

**“**Применение инструментов Data Mining”

Вариант 2

Выполнил: Белявская А.С.

Проверил: Кунцевич А.А.

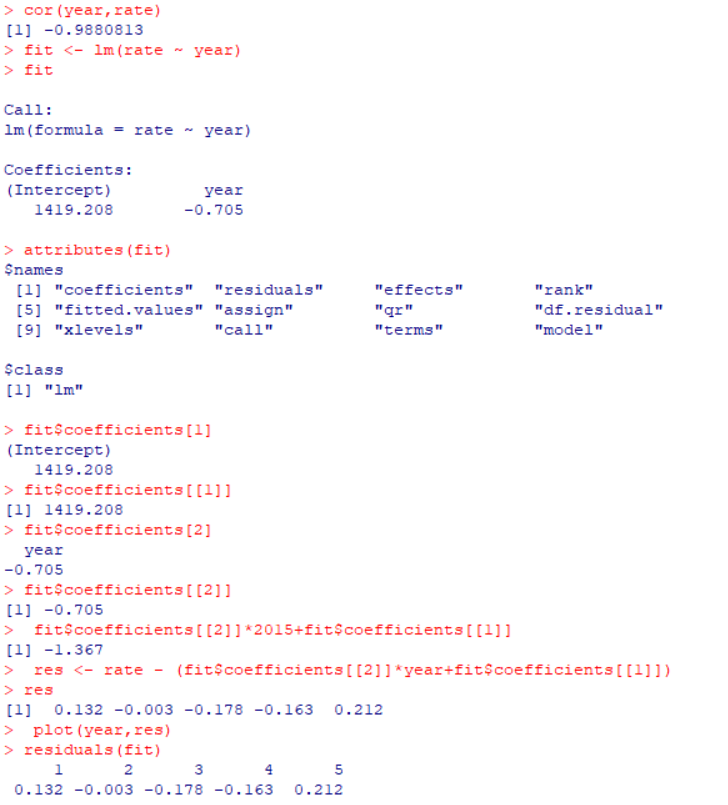
Минск, 2019

Первое, что нужно сделать, это указать данные. Каждая из пяти пар состоит из года и средней процентной ставки:



Следующее, что мы сделаем, это взглянем на данные. Сначала построим точечный график, используя наши данные и замечаем, что он выглядит линейным. Чтобы подтвердить наши подозрения, мы найдем корреляцию между годом и средней процентной ставки:





Следующим действием определяем прямую, которая будет проходить ближе к нашим данным. Для этого используем метод регрессии наименьших квадратов.

Перед нахождением регрессии определяем эндогенные и экзогенные переменные.

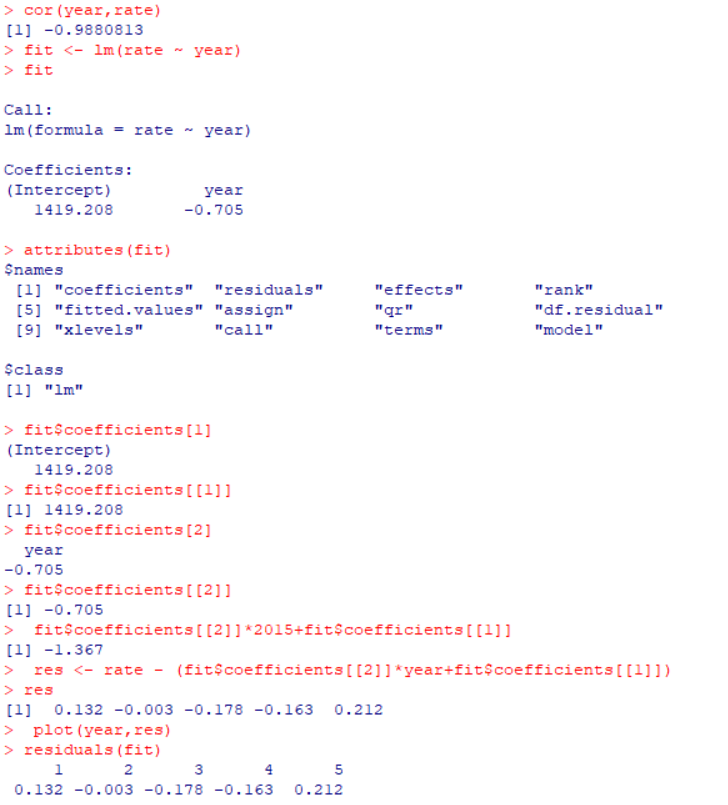
Процентная ставка может изменяться во времени, а не время изменяется по мере изменения процентных ставок, поэтому независимой переменной выбираем год, а зависимой – процентную ставку.

Команда для выполнения регрессии наименьших квадратов является команда *lm.*

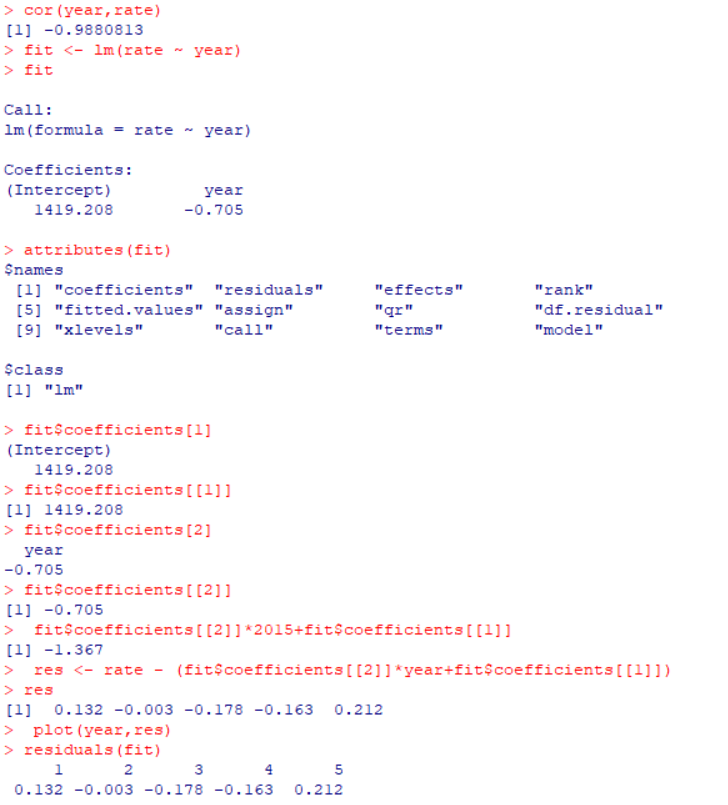
Так как мы указали, что процентная ставка будет являться зависимой, а год независимой переменной, то это означает, что линейная регрессии может быть записана в виде  slope-intercept :

rate=(slope)year+(intercept).

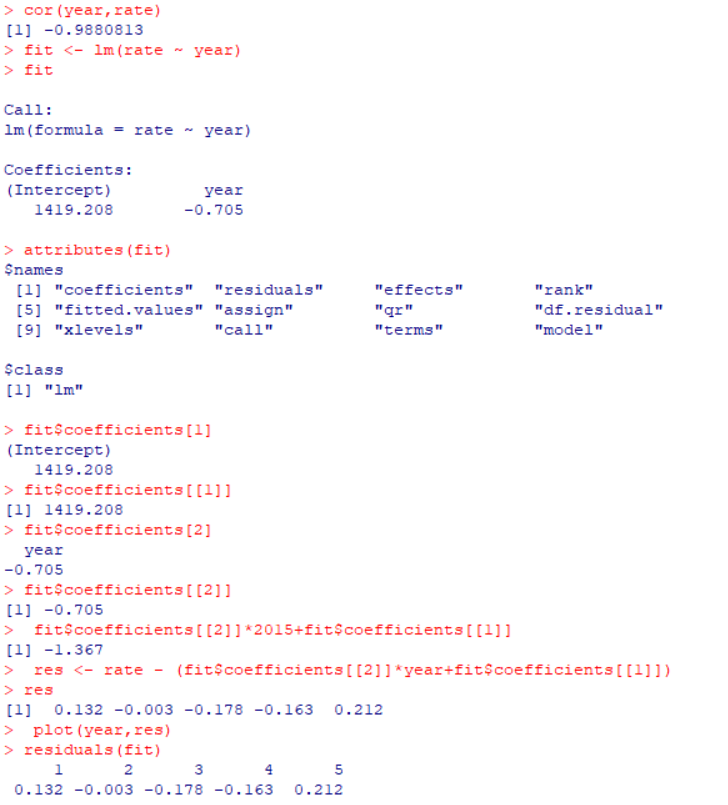
В команде *lm,* вы пишете вектор, содержащий зависимую переменную, тильда ("~"), и еще одни вектор, содержащий объясняющую переменную:

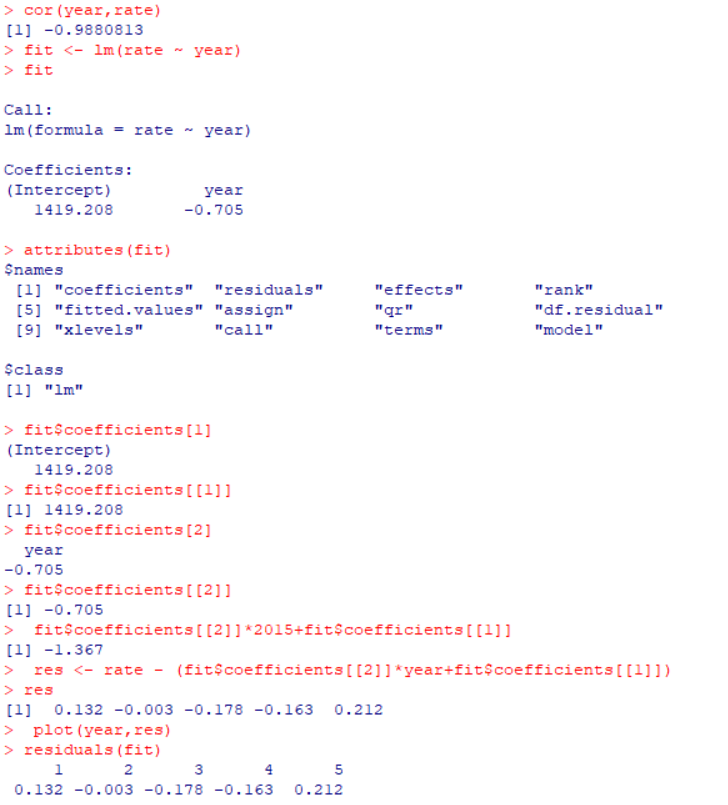


Когда вызывается *lm,* то функция возвращает переменную с большим количеством информации в нем. Если требуется узнать, что еще хранится в переменной, можно использовать команду attributes:

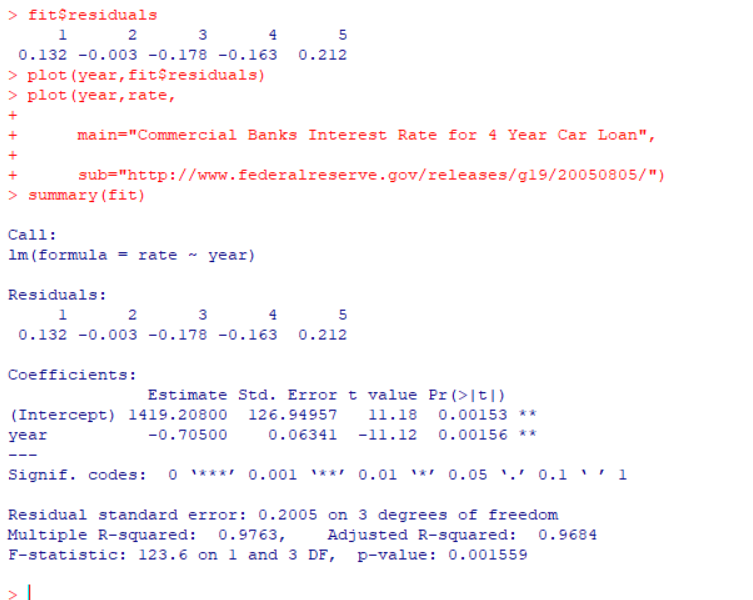


Одна из вещей, которую можно заметить, это переменные коэффициенты в fit. Можно вывести y-intercept и slope через доступ к этой части переменной:





Найдем остатки с помощью следующих операций, а чтобы построить график линии регрессии на том же участке, где находится точечной график, используем *abline()* функции наряду с переменной fit:



По итогу, чтобы получить результаты F-теста и других тестов, вызываем функцию *summary():*

