Министерство образования Республики Беларусь

Учреждение образования

«Белорусский Государственный Университет Информатики и Радиоэлектроники»

Инженерно-экономический факультет

Кафедра экономической информатики

Лабораторная работа №7

# «Применение инструментов Data Mining»

Проверил: Кунцевич А. А.

Выполнила: студент группы 772302

Крайняя М.Д.

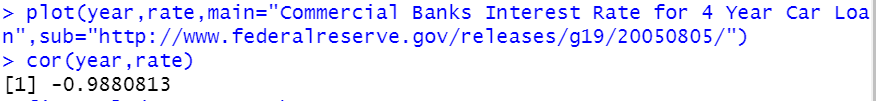
Минск 2020

Мы рассмотрели процентную ставку на четырехлетние автокредитование, и данные, которые мы использовали, происходят от [средних ставок Федеральной резервной системы США](http://translate.googleusercontent.com/translate_c?depth=1&hl=ru&prev=search&rurl=translate.google.by&sl=en&u=http://www.federalreserve.gov/releases/g19/20050805/&usg=ALkJrhgcARErF5r8objVrNleS8ocS6XPeQ).

Первое, что мы сделали, это указали данные. Здесь есть только пять пар чисел, чтобы мы могли ввести их вручную. Каждая из пяти пар состоит из года и средней процентной ставки:



Сначала построили точечный график. Найдем корреляцию между годом и средней процентной ставки:

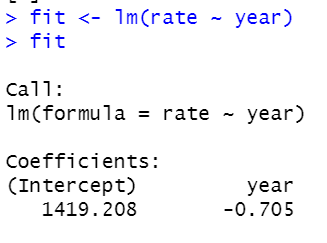


Следующий вопрос, какая прямая будет "ближе" к нашим данным? В этом случае мы будем использовать метод регрессии наименьшие квадратов как один из способов, чтобы определить искомую линию.

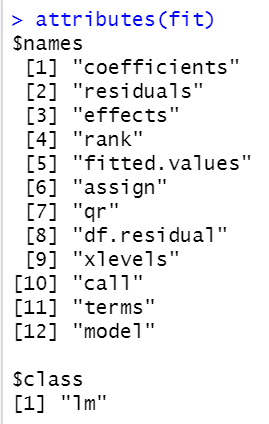
Прежде чем мы сможем найти регрессию наименьших квадратов, мы должны решить, что будет пояснительной переменной и что будет переменной отклика. В нашем случае правильнее будет выбрать пояснительной переменной год, а переменной отклика процентную ставку. Это было выбрано потому, что процентная ставка может изменяться во времени, а не время изменяется по мере изменения процентных ставок.

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

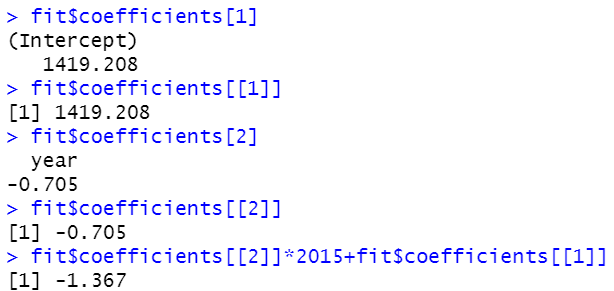
rate=(slope)year+(intercept)



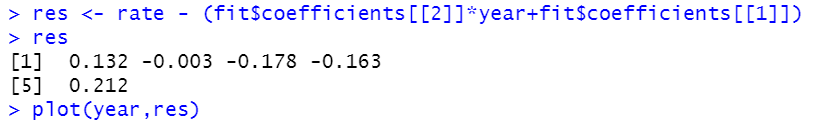
Чтобы узнать, что еще хранится в переменной, использовали команду attributes:



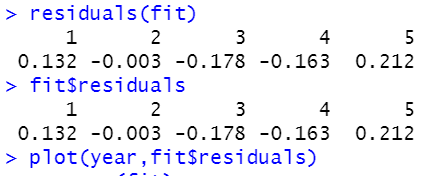
Вывели переменные коэффициенты:

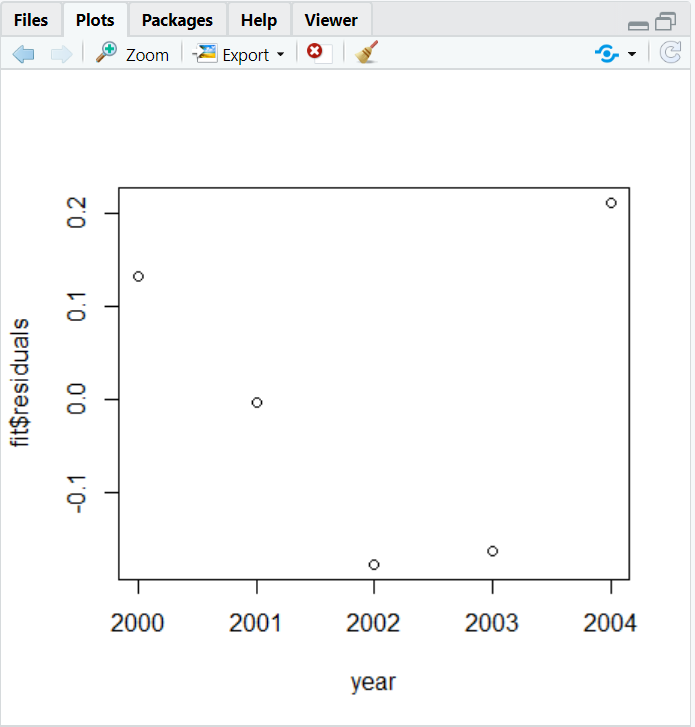


Формула для вычисления остатков:

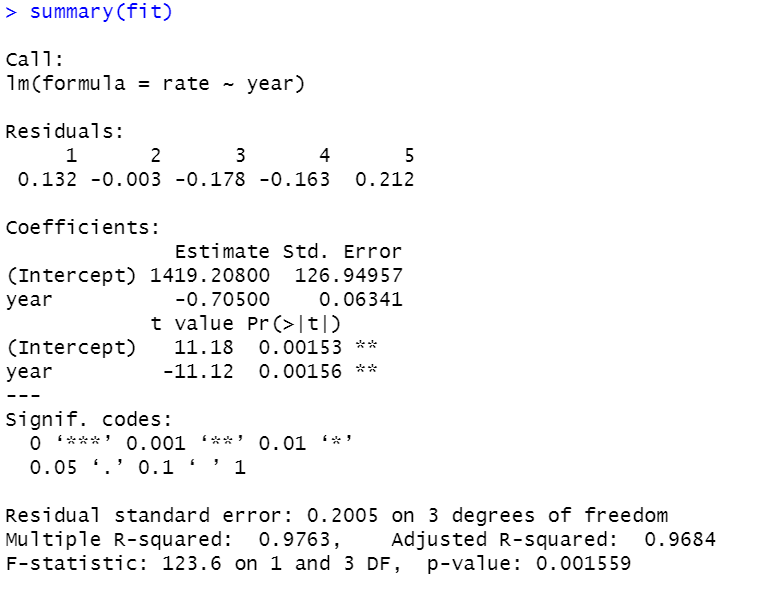


Два других способа приведены ниже:





Результаты F-теста и других тестов:



В чем разница между матрицей и фреймом данных?

*Фрейм данных может содержать неоднородные входные данные, а матрица - нет. (У нас могут быть данные с символами, целыми числами и даже с другими данными, но мы не сможем их привести к матрице - матрица должна быть одного типа.)*

Что представляет собой функция summary?

*summary - общая функция, используемая, чтобы произвести итоговые итоги результатов различных функций подбора модели. Функция вызывает конкретные методы, которые зависят от класса первого аргумента.*