Министерство образования Республики Беларусь

Учреждение образования

Белорусский государственный университет  
информатики и радиоэлектроники

Кафедра экономической информатики

ОТЧЁТ  
по лабораторной работе №7  
на тему «Применение инструментов Data Mining»

Выполнила:

студентка группы 772303

Желнерович Т.С.

Проверил:

Кунцевич А.А.

Минск 2020

Задание

Построить модель линейной регрессии и продемонстрировать ее оценку.

Ход работы

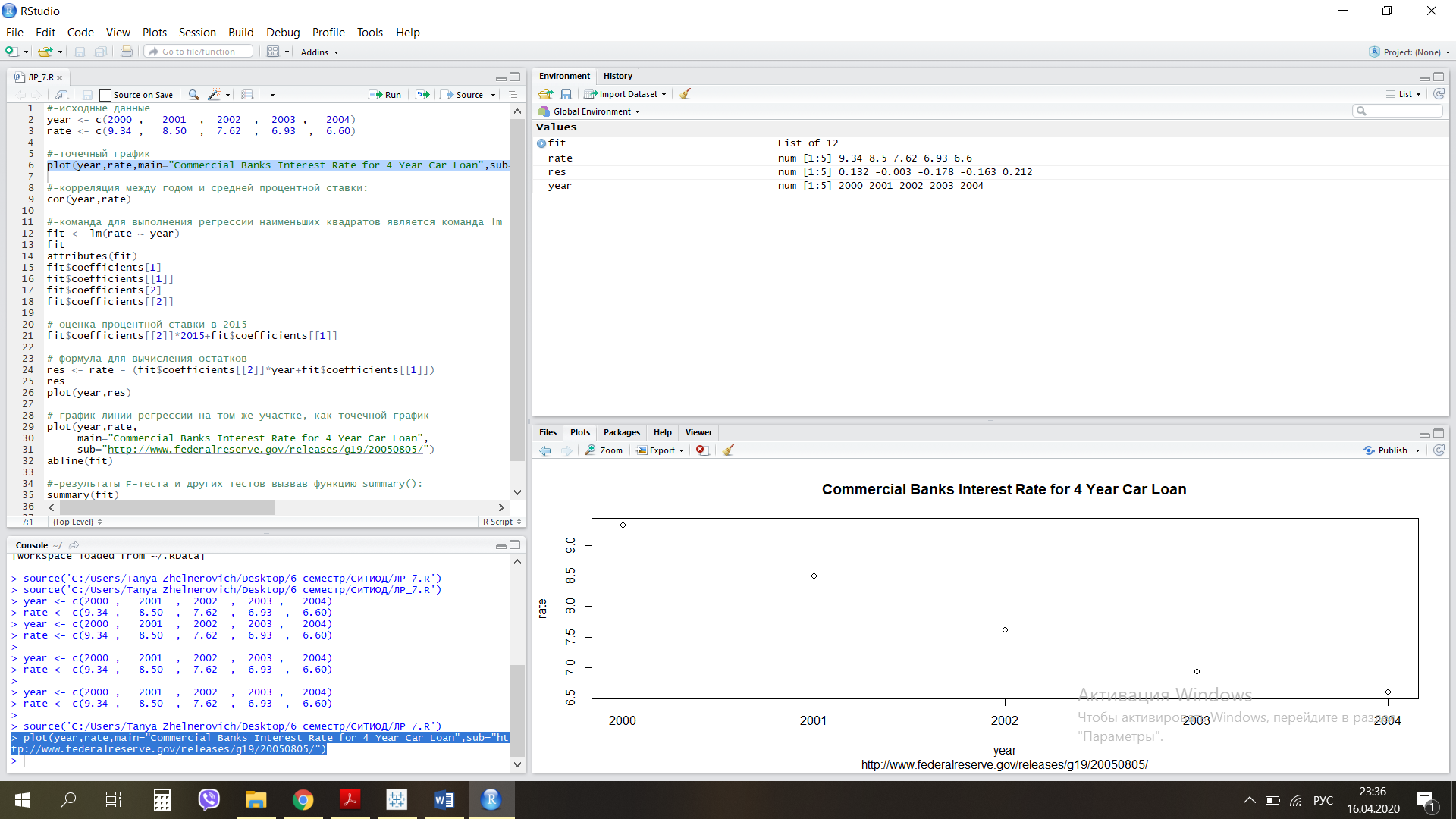
В качестве исходных данных у нас выступают год и процентная ставка:

> year <- c(2000 , 2001 , 2002 , 2003 , 2004)

> rate <- c(9.34 , 8.50 , 7.62 , 6.93 , 6.60)

С помощью функции plot строим точечный график, используя наши данные

> plot(year,rate,main="Commercial Banks Interest Rate for 4 Year Car Loan",sub="http://www.federalreserve.gov/releases/g19/20050805/")



Находим корреляцию между годом и процентной ставкой:

> cor(year,rate)

[1] -0.9880813

Выполним команду для выполнения регрессии наименьших квадратов lm:

> fit <- lm(rate ~ year)

> fit

Call:

lm(formula = rate ~ year)

Coefficients:

(Intercept) year

1419.208 -0.705

> attributes(fit)

$names

[1] "coefficients" "residuals" "effects" "rank"

[5] "fitted.values" "assign" "qr" "df.residual"

[9] "xlevels" "call" "terms" "model"

$class

[1] "lm"

> fit$coefficients[1]

(Intercept)

1419.208

> fit$coefficients[[1]]

[1] 1419.208

> fit$coefficients[2]

year

-0.705

> fit$coefficients[[2]]

[1] -0.705

Получим оценку процентной ставки в 2015 году:

fit$coefficients[[2]]\*2015+fit$coefficients[[1]]

[1] -1.367

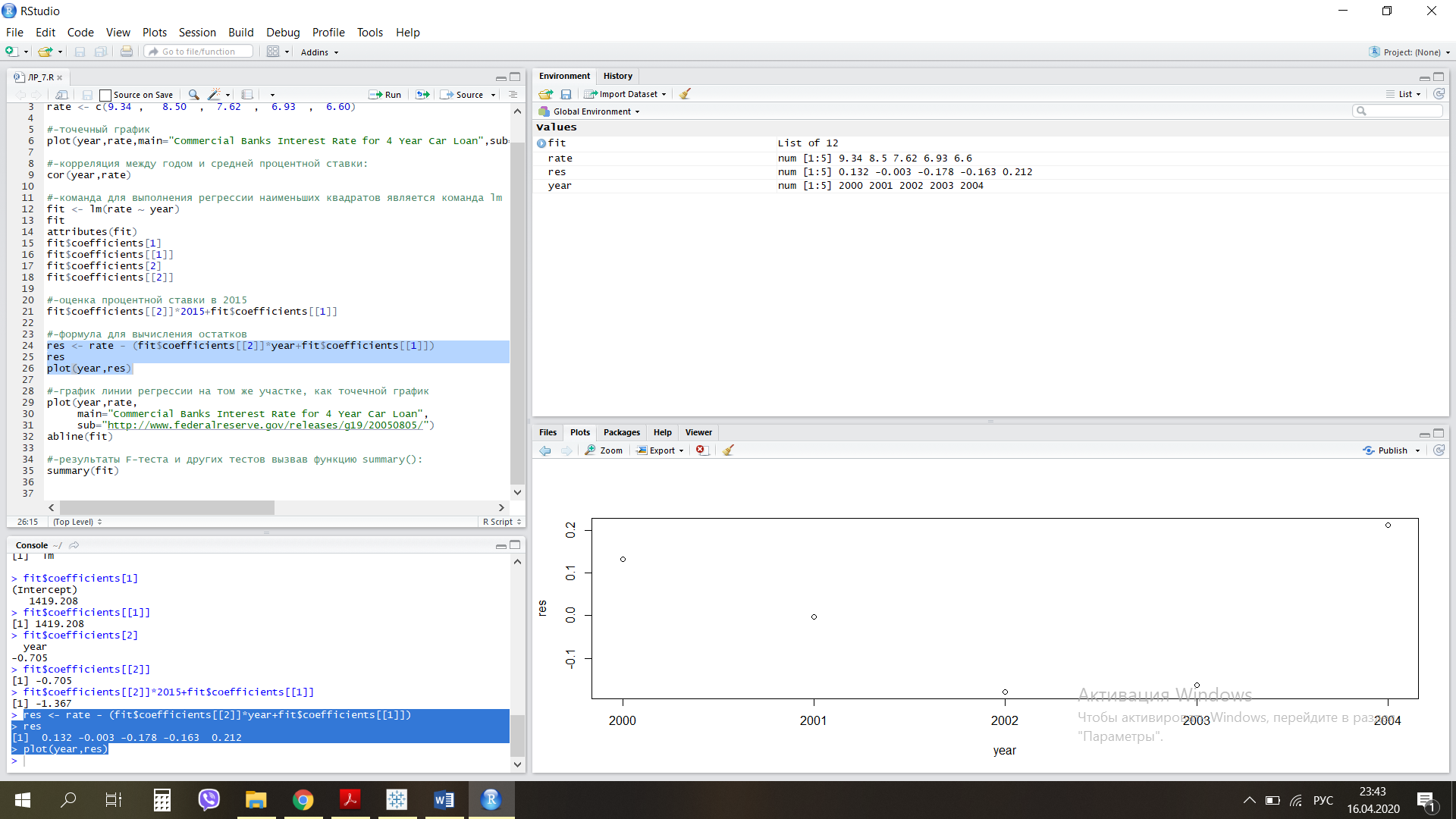
Далее вычислим остатки и построим их график:

res <- rate - (fit$coefficients[[2]]\*year+fit$coefficients[[1]])

> res

[1] 0.132 -0.003 -0.178 -0.163 0.212

> plot(year,res)



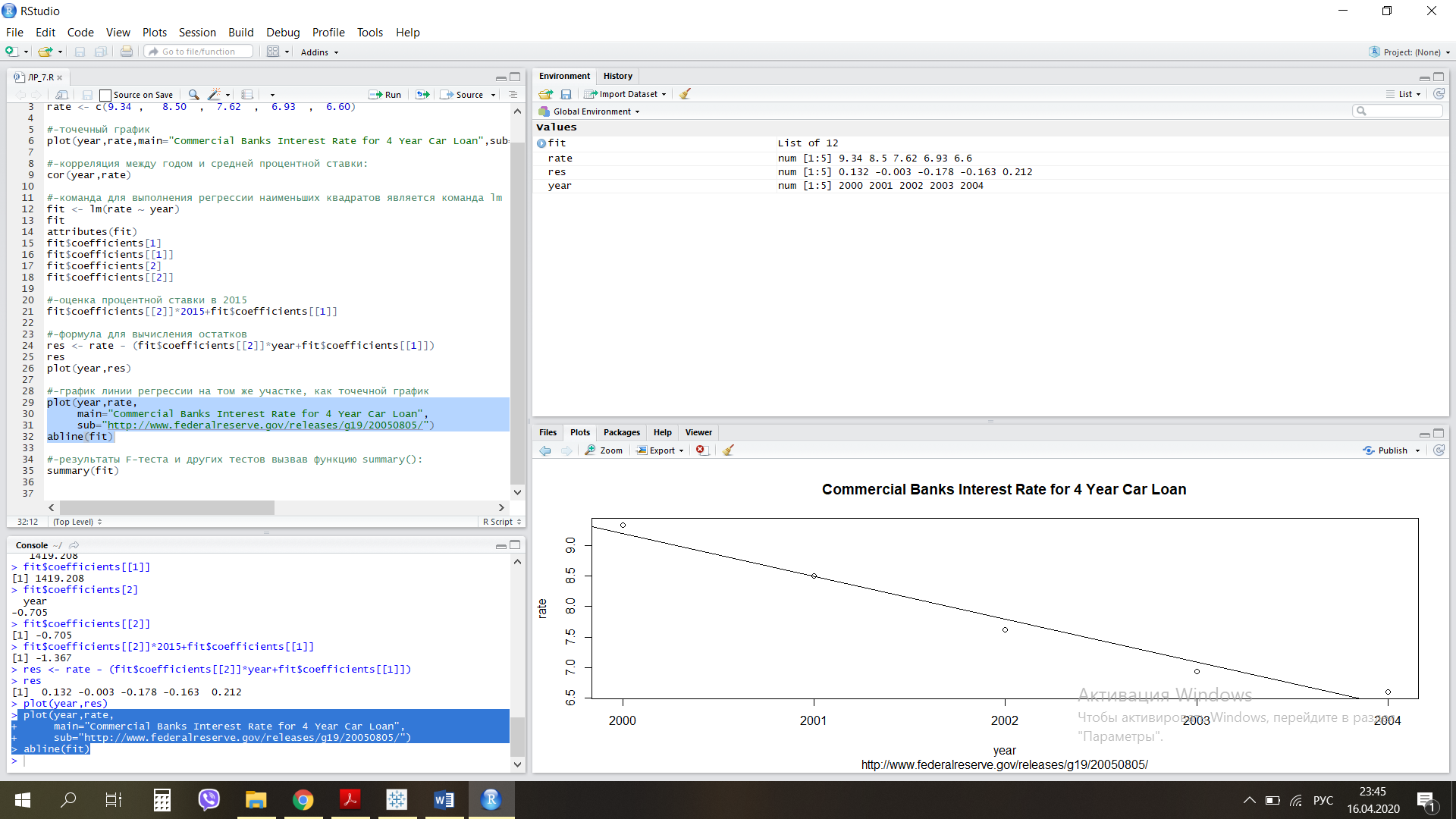
Для построения графика линии регрессии на том же участке, что и точечной график, используем abline() наряду с переменной fit:

plot(year,rate,

+ main="Commercial Banks Interest Rate for 4 Year Car Loan",

+ sub="http://www.federalreserve.gov/releases/g19/20050805/")

> abline(fit)



Вызвав функцию summary() получим результаты F- и других тестов:

> summary(fit)

Call:

lm(formula = rate ~ year)

Residuals:

1 2 3 4 5

0.132 -0.003 -0.178 -0.163 0.212

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1419.20800 126.94957 11.18 0.00153 \*\*

year -0.70500 0.06341 -11.12 0.00156 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2005 on 3 degrees of freedom

Multiple R-squared: 0.9763, Adjusted R-squared: 0.9684

F-statistic: 123.6 on 1 and 3 DF, p-value: 0.001559