

Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНФОРМАТИКИ
И РАДИОЭЛЕКТРОНИКИ

Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет ИНЖЕНЕРНО – ЭКОНОМИЧЕСКИЙ

Кафедра ЭКОНОМИКИ

ОТЧЕТ

по дисциплине «СиТИОД»

Применение инструментов Data Mining

Выполнил студент

Красовский Е.П.

гр. 772303

Проверил

Кунцевич А.А.

Минск 2020

Практическая часть

Скриншоты выполнения лаб. работы:

```
> year <- c(2000, 2001, 2002, 2003, 2004)
> rate <- c(9.34, 8.50, 7.62, 6.93, 6.60)
> plot(year,rate,main="Commercial Banks Interest Rate for 4 Year Car Loan",sub="http://
> cor(year,rate)
[1] -0.9880813
> 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" "fitted.values"
[6] "assign" "qr" "df.residual" "xlevels" "call"
[11] "terms" "model"

$class
[1] "lm"

> fit$coefficients[1]
(Intercept)
    1419.208 
> fit$coefficients[[1]]
[1] 1419.208 
> fit$coefficients[2]
year
    -0.705 

> fit$coefficients[[2]]
[1] -0.705 
> fit$coefficients[[2]]*2015+fit$coefficients[[1]]
[1] -1.367 
> res <- rate - (fit$coefficients[[2]]*year+fit$coefficients[[1]])
> res
[1] 0.132 -0.003 -0.178 -0.163  0.212 
> plot(year,res)
> residuals(fit)
      1      2      3      4      5 
0.132 -0.003 -0.178 -0.163  0.212 
> fit$residuals
      1      2      3      4      5 
0.132 -0.003 -0.178 -0.163  0.212 
> plot(year,fit$residuals)
> plot(year,rate,
+       main="Commercial Banks Interest Rate for 4 Year Car Loan",
+       sub="http://www.federalreserve.gov/releases/g19/20050805/")
> abline(fit)
> 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     0.02118    -33.31  <0.0001 ***
---
Signif. codes:  0.0001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1.0 'not sig.'

>
```

Рисунок 1 – теоретическая часть консоль

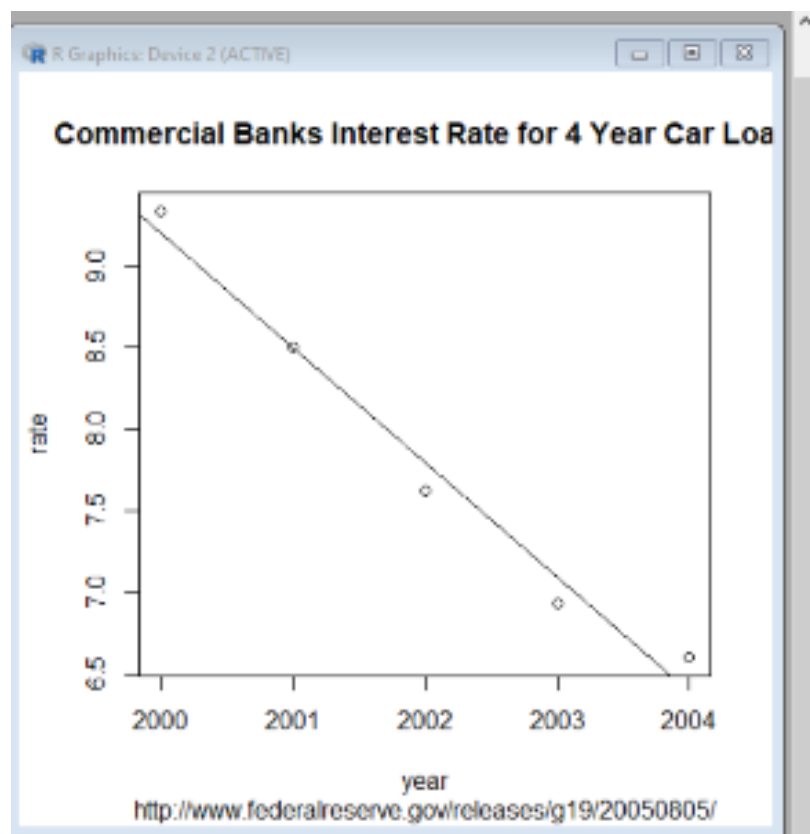


Рисунок 2 – теоретическая часть, график

```

>
> year <- c(2015, 2016, 2017, 2018, 2019)
> rate <- c(3.34, 5.50, 7.32, 6.93, 8.60)
> plot(year, rate, main="Commercial Banks Interest Rate for 4 Year Car Loan", sub="http://$
> cor(year, rate)
[1] 0.9411585
>

```

Рисунок 3 – задаем данные и проверяем корреляцию

Корреляция есть. Модель не адекватна.

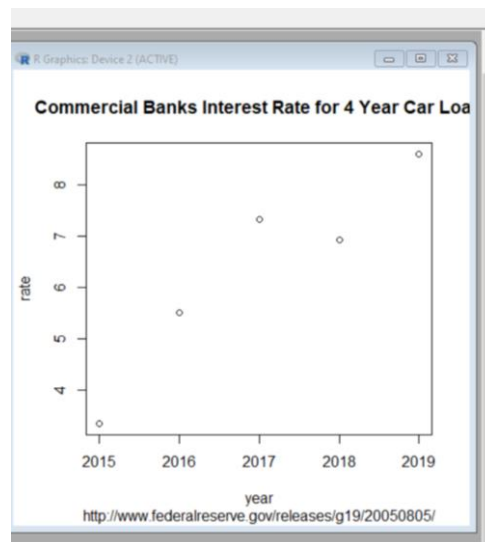


Рисунок 4 – визуализируем данные

```

> fit <- lm(rate ~ year)
> fit

Call:
lm(formula = rate ~ year)

Coefficients:
(Intercept)      year
-2403.977        1.195

```

Рисунок 5 – рассчитываем коэффициенты уравнения регрессии

Уравнение регрессии: $\text{Rate} = -2403.977 + 1.195 \text{year}$

```

res <- rate - (fit$coefficients[[2]]*year + fit$coefficients[[1]])
res
[1] -0.608  0.357  0.982 -0.603 -0.128
plot(year, rate,
      main="Commercial Banks Interest Rate for 4 Year Car Loan",
      sub="http://www.federalreserve.gov/releases/g19/20050805/")
abline(fit)

```

Рисунок 6 – получаем остатки и визуализируем их

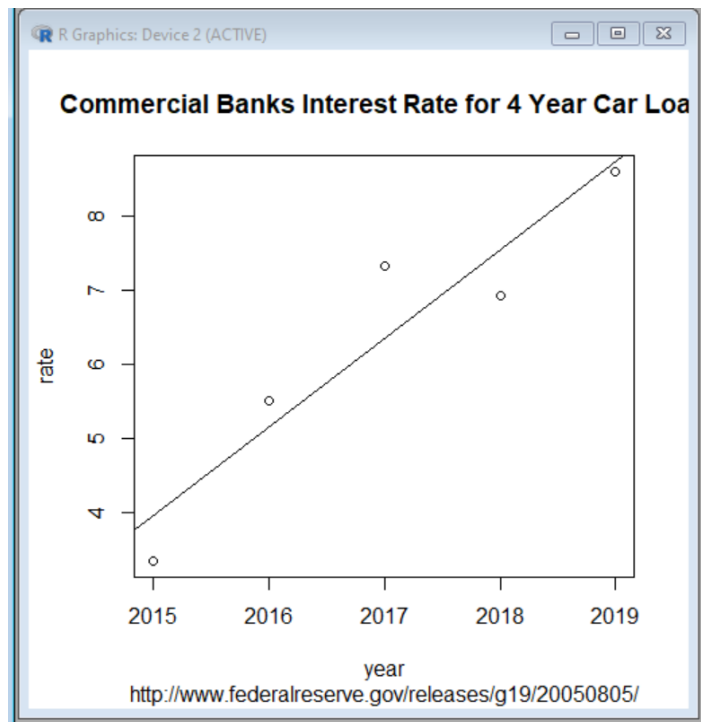


Рисунок 7 – график линий регрессии

```
>
> summary(fit)

Call:
lm(formula = rate ~ year)

Residuals:
    1     2     3     4     5 
-0.608  0.357  0.982 -0.603 -0.128 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2403.9770   499.7157  -4.811  0.0171 *
year           1.1950    0.2478   4.823  0.0170 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7835 on 3 degrees of freedom
Multiple R-squared:  0.8858,    Adjusted R-squared:  0.8477 
F-statistic: 23.26 on 1 and 3 DF,  p-value: 0.01698

> |
```

Рисунок 8 – визуализируем данные

Intercept Значение Стьюдента в норме, оценка значима. Стандартная ошибка больше нормы.

Year: Значение Стьюдента в норме, оценка значима.

3 степени свободы со значимостью 0,78. $T_{кр.}=3,82$. Статистическая значимость коэффициентов подтвердилась, т. к. $T_{кр.}< t \text{ value}$. В обоих случаях P-level меньше уровня значимости, соответственно, оба коэффициента статистически значимы с вероятностью в 99%.

совокупный коэффициент корреляции: 0,885, близка к линейной зависимости.

скорректированный (нормированный) коэффициент детерминации.: 0,84, оценки точные

F-статистика: 0,017, p-значение критерия Фишера ниже пяти сотых – модель в целом значима.