

Homework 2

0) Complete information below:

NAME: Valeria Aybar

PANTHER ID: 6181336

CERTIFICATION: I understand FIU's academic policies, and I certify that this
work is my own and that none of it is the work of any other person.

=====

PROGRAMMER: Valeria Aybar

PANTHER ID: 6181336

#

CLASS: CAP 4830

SECTION: RVC 1231

SEMESTER: Spring 2023

CLASSTIME: Online

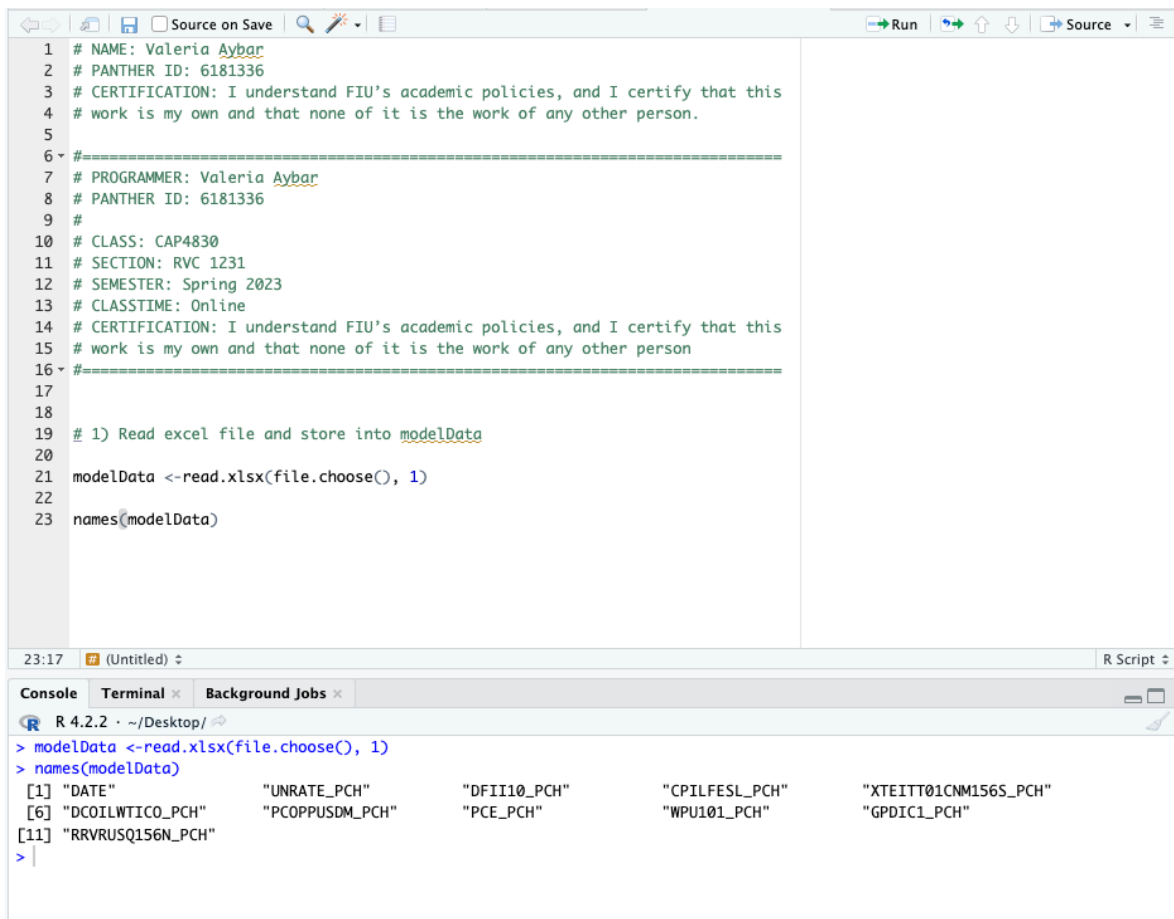
CERTIFICATION: I understand FIU's academic policies, and I certify that this

work is my own and that none of it is the work of any other person.

=====

1) Read the excel file "CAP4830_HW2_Data.xlsx" data into R and store the imported data in a variable named "modelData".

2) Output the names of the modelData dataframe.



```

1 # NAME: Valeria Aybar
2 # PANTHER ID: 6181336
3 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
4 # work is my own and that none of it is the work of any other person.
5
6 #=====
7 # PROGRAMMER: Valeria Aybar
8 # PANTHER ID: 6181336
9 #
10 # CLASS: CAP4830
11 # SECTION: RVC 1231
12 # SEMESTER: Spring 2023
13 # CLASSTIME: Online
14 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
15 # work is my own and that none of it is the work of any other person
16 #=====
17
18
19 # 1) Read excel file and store into modelData
20
21 modelData <- read.xlsx(file.choose(), 1)
22
23 names(modelData)

```

23:17 (Untitled) R Script

Console Terminal Background Jobs

```

R 4.2.2 ~ /Desktop/
> modelData <- read.xlsx(file.choose(), 1)
> names(modelData)
[1] "DATE" "UNRATE_PCH" "DFII10_PCH" "CPILFESL_PCH" "XTEITT01CNM156S_PCH"
[6] "DCOILWTICO_PCH" "PCOPUSDM_PCH" "PCE_PCH" "WPU101_PCH" "GPDIC1_PCH"
[11] "RRVRUSQ156N_PCH"
>

```

3) Create a variable with name "model1" that stores the estimate of the linear model shown below

$$\begin{aligned} \text{UNRATE_PCH} = & b_0 + b_1 \cdot \text{DFII10_PCH} + b_2 \cdot \text{CPILFESL_PCH} + b_3 \cdot \text{XTEITT01CNM156S_PCH} \\ & + b_4 \cdot \text{DCOILWTICO_PCH} + b_5 \cdot \text{PCOPPUSDM_PCH} + b_6 \cdot \text{PCE_PCH} \\ & + b_7 \cdot \text{WPU101_PCH} + b_8 \cdot \text{GPDIC1_PCH} + b_9 \cdot \text{RRVRUSQ156N_PCH} \end{aligned}$$

```
# CERTIFICATION: I understand FIU's academic policies, and I certify that this
# work is my own and that none of it is the work of any other person.
#=====
# PROGRAMMER: Valeria Aybar
# PANTHER ID: 6181336
#
# CLASS: CAP4830
# SECTION: RVC 1231
# SEMESTER: Spring 2023
# CLASSTIME: Online
# CERTIFICATION: I understand FIU's academic policies, and I certify that this
# work is my own and that none of it is the work of any other person
#=====
32:1 (Untitled) R Script
```

```
R 4.2.2 ~./Desktop/
+ + WPU101_PCH + GPDIC1_PCH + RRVUSQ156N_PCH, data = modelData)
> summary(model1)

Call:
lm(formula = UNRATE_PCH ~ DFII10_PCH + CPILFESL_PCH + XTEITT01CNM156S_PCH +
    DCOILWTICO_PCH + PCOPPUSDM_PCH + PCE_PCH + WPU101_PCH + GPDIC1_PCH +
    RRVUSQ156N_PCH, data = modelData)

Residuals:
    Min       1Q   Median       3Q      Max
-26.514  -5.337  -1.006   4.151  46.473

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.37184    5.14506   2.793 0.006960 **
DFII10_PCH    0.08225    0.01558   5.279 1.83e-06 ***
CPILFESL_PCH -5.49906   11.04301  -0.498 0.620298
XTEITT01CNM156S_PCH 1.39234    0.36109   3.856 0.000280 ***
DCOILWTICO_PCH  0.62693    0.17573   3.568 0.000709 ***
PCOPPUSDM_PCH  0.13167    0.15203   0.866 0.389834
PCE_PCH     -12.01936    1.73121  -6.943 2.91e-09 ***
WPU101_PCH   -0.19723    0.27059  -0.729 0.468844
GPDIC1_PCH   -0.52205    0.57422  -0.909 0.366849
RRVRUSQ156N_PCH -0.01746    0.41295  -0.042 0.966409
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.6 on 61 degrees of freedom
Multiple R-squared:  0.8446,    Adjusted R-squared:  0.8217
F-statistic: 36.84 on 9 and 61 DF,  p-value: < 2.2e-16

> |
```

4) List all the estimate parameters from step 3 that are statistically significant for all $\alpha \leq 0.05$

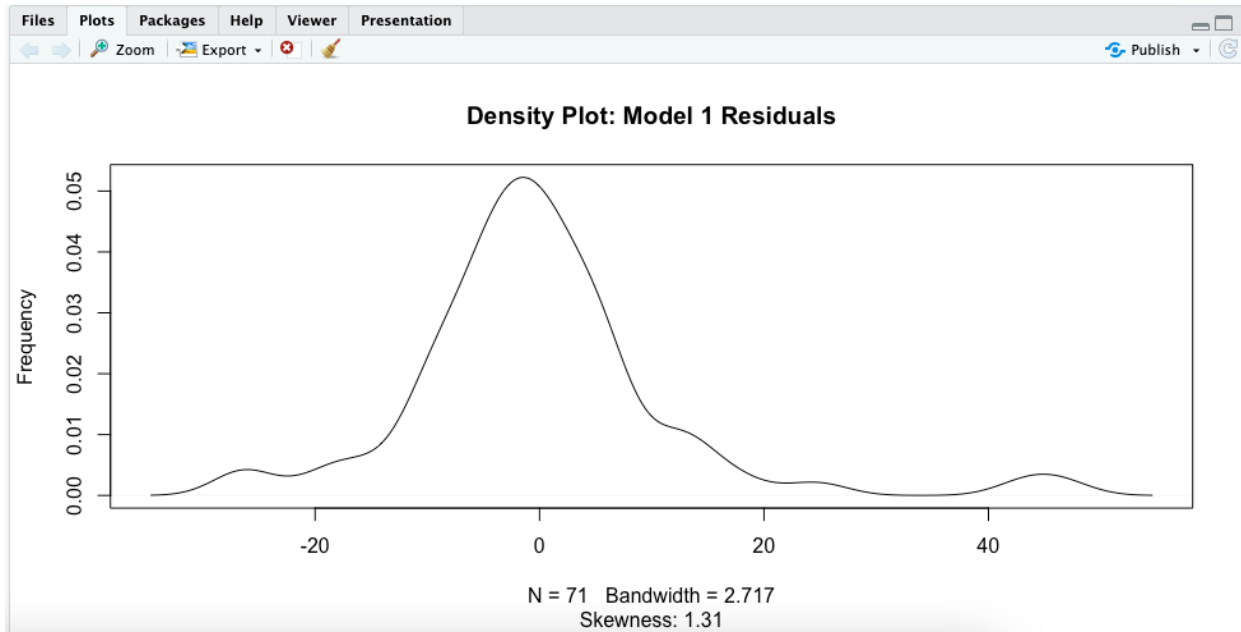
DFII10_PCH: 1.83e-06

XTEITT01CNM156S_PCH: 0.000280

DCOILWTICO_PCH: 0.000709

PCE_PCH: 2.91e-09

5) Plot the model1's residual Density Function



6) Check the model1's residual normality using the Shapiro test. Paste your results below and explain your finding in one to two sentences.

Based on the p-value, since its 6.009e-06 (0.000006009) this number is actually smaller than 0.05, so we can conclude that this is not a normal distribution.

```
# PANTHER ID: 6181336
3 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
4 # work is my own and that none of it is the work of any other person.
5
6 #=====
7 # PROGRAMMER: Valeria Aybar
8 # PANTHER ID: 6181336
9 #
10 # CLASS: CAP4830
11 # SECTION: RVC 1231
12 # SEMESTER: Spring 2023
13 # CLASSTIME: Online
14 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
15 # work is my own and that none of it is the work of any other person

32:15 (Untitled) R Script

Console Terminal Background Jobs
R 4.2.2 ~ /
> shapiro.test(model1$residuals)

Shapiro-Wilk normality test

data: model1$residuals
W = 0.87998, p-value = 6.009e-06
```

7) Create model2 which is a refinement of model1 by removing all regressors that are statistically insignificant with $\alpha = 0.55$. Hence you are removing regressors with $p\text{-value} > 0.55$. Paste your model's summary below:

```

2 # PANTHER ID: 6181336
3 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
4 # work is my own and that none of it is the work of any other person.
5
6 #=====
7 # PROGRAMMER: Valeria Aybar
8 # PANTHER ID: 6181336
9 #
10 # CLASS: CAP4830
11 # SECTION: RVC 1231
12 # SEMESTER: Spring 2023
13 # CLASSTIME: Online
14 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
15 # work is my own and that none of it is the work of any other person
16 #=====
17
18 rm(list = ls()) #clearing all objects from session
19
46:1 (Untitled) R Script

```

```

Call:
lm(formula = UNRATE_PCH ~ DFII10_PCH + XTEITT01CNM156S_PCH +
    DCOILWTICO_PCH + PCOPPUSSDM_PCH + PCE_PCH + WPU101_PCH + GPDIC1_PCH,
    data = modelData)

Residuals:
    Min       1Q   Median       3Q      Max
-26.759  -5.317  -0.428   3.823  46.332

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.07281    1.85399   6.512 1.41e-08 ***
DFII10_PCH      0.08323    0.01509   5.516 6.94e-07 ***
XTEITT01CNM156S_PCH 1.39019    0.34763   3.999 0.000170 ***
DCOILWTICO_PCH  0.61623    0.16376   3.763 0.000371 ***
PCOPPUSSDM_PCH  0.13504    0.14965   0.902 0.370286
PCE_PCH        -12.39692    1.40444  -8.827 1.29e-12 ***
WPU101_PCH     -0.18866    0.26449  -0.713 0.478296
GPDIC1_PCH     -0.46499    0.55128  -0.843 0.402154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.43 on 63 degrees of freedom
Multiple R-squared:  0.844,    Adjusted R-squared:  0.8266
F-statistic: 48.68 on 7 and 63 DF,  p-value: < 2.2e-16
>

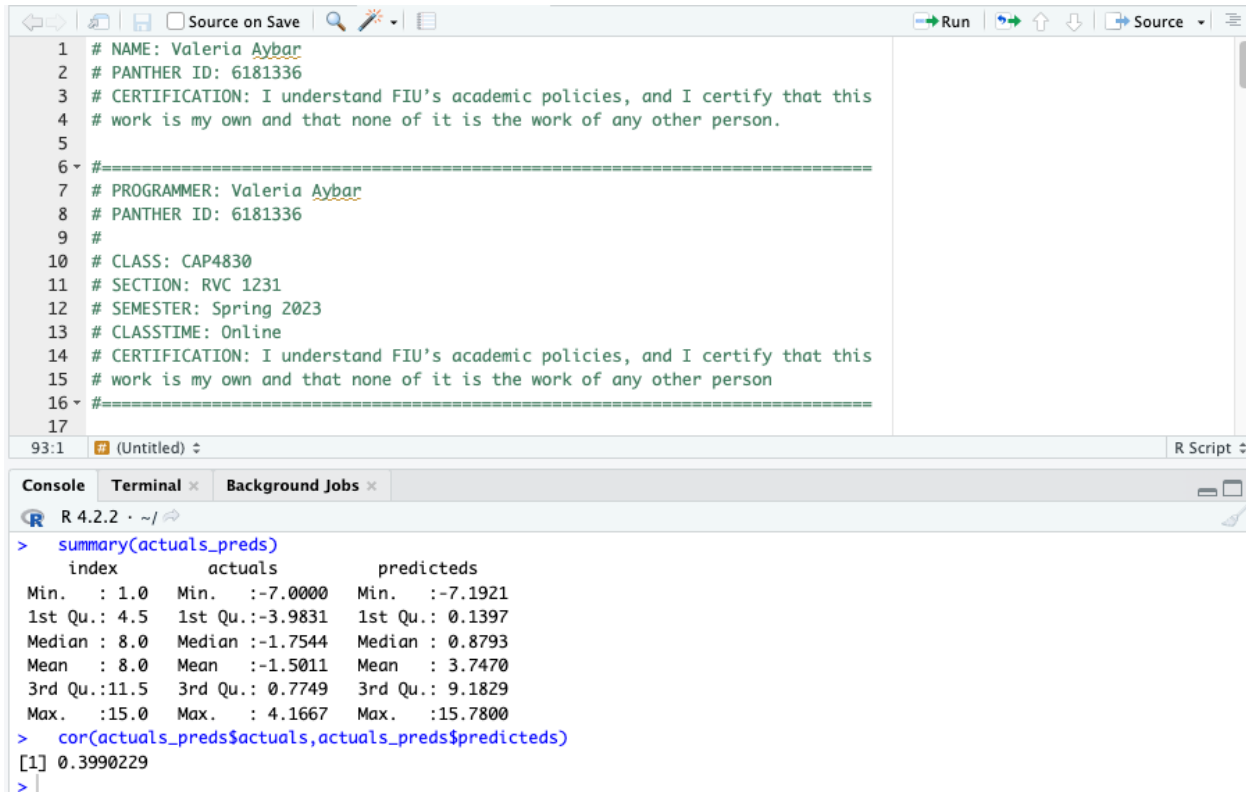
```

8) What is the difference in your Adjusted R between model1 and model2.

The difference between the adjusted R model1 0.8217 and model2 0.8266 is 0.0049

9) Calculate prediction accuracy and error rates of model2. Look at the R-script in module 10.

The errors were correlated 0.3990229.

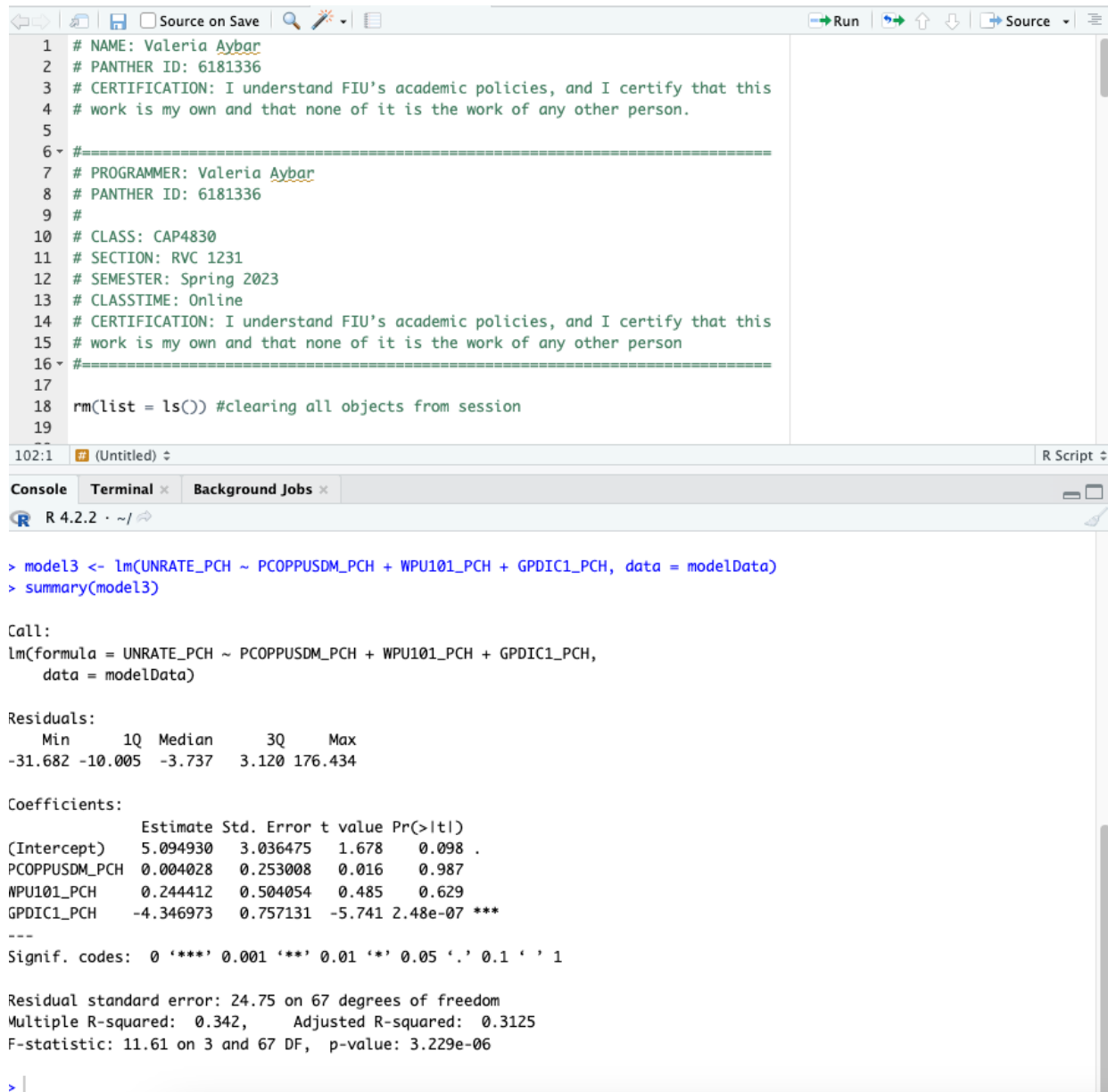


The screenshot displays the RStudio environment. The top pane shows an R script with 17 lines of code. Lines 1-4 and 7-16 contain comments identifying the user as Valeria Aybar, a student in the CAP4830 class, RVC 1231 section, Spring 2023 semester, and online class. Lines 6 and 15 are separated by green dashed lines. The bottom pane shows the R console output for the following commands:

```
> summary(actuals_preds)
  index      actuals      predicted
Min.   : 1.0   Min.   :-7.0000   Min.   :-7.1921
1st Qu.: 4.5   1st Qu.: -3.9831   1st Qu.: 0.1397
Median : 8.0   Median : -1.7544   Median : 0.8793
Mean    : 8.0   Mean    : -1.5011   Mean    : 3.7470
3rd Qu.:11.5   3rd Qu.: 0.7749   3rd Qu.: 9.1829
Max.    :15.0   Max.    : 4.1667   Max.    :15.7800
> cor(actuals_preds$actuals,actuals_preds$predicted)
[1] 0.3990229
>
```

10) Create model3 which is a refinement of model2. A requirement for model3 it must only have three regressors. How you pick the three regressor is up to you, but explain why you pick these three. Paste the summary of model3 below.

I decided to go with the coefficients: PCOPPUUSD_M_PCH, WPU101_PCH and GPDIC1_PCH because these did not have the signif.code for 0 ***



```
1 # NAME: Valeria Aybar
2 # PANTHER ID: 6181336
3 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
4 # work is my own and that none of it is the work of any other person.
5
6 #=====
7 # PROGRAMMER: Valeria Aybar
8 # PANTHER ID: 6181336
9 #
10 # CLASS: CAP4830
11 # SECTION: RVC 1231
12 # SEMESTER: Spring 2023
13 # CLASSTIME: Online
14 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
15 # work is my own and that none of it is the work of any other person
16 #=====
17
18 rm(list = ls()) #clearing all objects from session
19
102:1 (Untitled) R Script
```

```
> model3 <- lm(UNRATE_PCH ~ PCOPPUUSD_M_PCH + WPU101_PCH + GPDIC1_PCH, data = modelData)
> summary(model3)

Call:
lm(formula = UNRATE_PCH ~ PCOPPUUSD_M_PCH + WPU101_PCH + GPDIC1_PCH,
    data = modelData)

Residuals:
    Min       1Q   Median       3Q      Max
-31.682 -10.005  -3.737   3.120  176.434

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.094930   3.036475   1.678   0.098 .
PCOPPUUSD_M_PCH  0.004028   0.253008   0.016   0.987
WPU101_PCH    0.244412   0.504054   0.485   0.629
GPDIC1_PCH   -4.346973   0.757131  -5.741 2.48e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.75 on 67 degrees of freedom
Multiple R-squared:  0.342,    Adjusted R-squared:  0.3125
F-statistic: 11.61 on 3 and 67 DF,  p-value: 3.229e-06

>
```

11) Create model4 that uses a manual sampling technique with a training set of 60% of the data and a testing set of 40%. Paste the summary of the model below.

```

1 # NAME: Valeria Aybar
2 # PANTHER ID: 6181336
3 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
4 # work is my own and that none of it is the work of any other person.
5
6 #=====
7 # PROGRAMMER: Valeria Aybar
122:1 (Untitled)
R Script

Console Terminal Background Jobs
R 4.2.2 ~ /
> # model - training data
> trainingData <- modelData[model4RowIndex , ]
> # test data
> testData <- modelData[-model4RowIndex , ]
> # building model4
> model4 <- lm(UNRATE_PCH ~ DFII10_PCH + CPILFESL_PCH + XTEITT01CNM156S_PCH
+             + DCOILWTICO_PCH + PCOPPU5DM_PCH + PCE_PCH
+             + WPU101_PCH + GPDIC1_PCH + RRVRSQ156N_PCH, data = trainingData)
> summary(model4)

Call:
lm(formula = UNRATE_PCH ~ DFII10_PCH + CPILFESL_PCH + XTEITT01CNM156S_PCH +
    DCOILWTICO_PCH + PCOPPU5DM_PCH + PCE_PCH + WPU101_PCH + GPDIC1_PCH +
    RRVRSQ156N_PCH, data = trainingData)

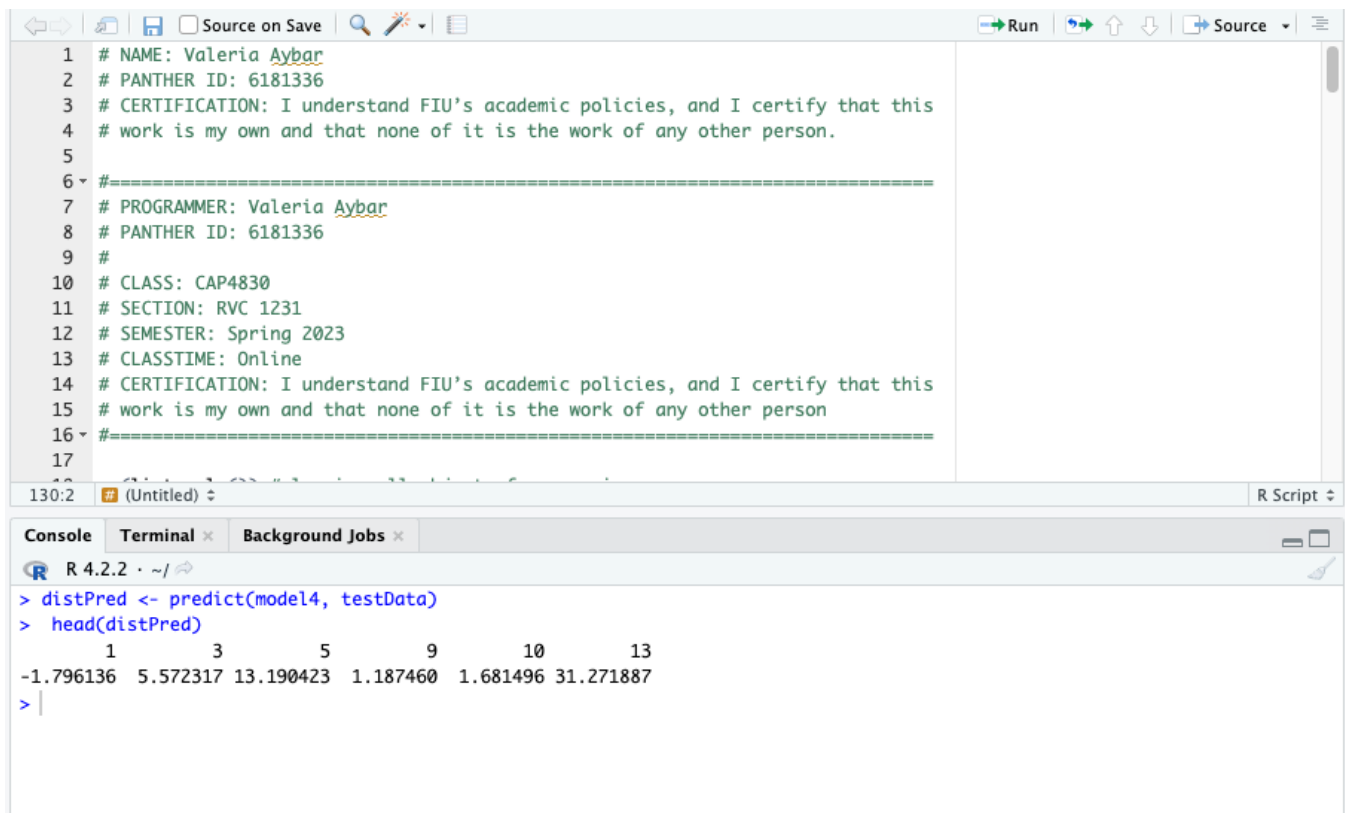
Residuals:
    Min       1Q   Median       3Q      Max
-30.385  -5.978  -2.204   5.342  38.232

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.62129    8.08342   1.190 0.242704
DFII10_PCH     0.09563    0.02145   4.458 9.52e-05 ***
CPILFESL_PCH   3.74501   18.20765   0.206 0.838341
XTEITT01CNM156S_PCH 1.81930    0.50156   3.627 0.000985 ***
DCOILWTICO_PCH  0.58847    0.26204   2.246 0.031755 *
PCOPPU5DM_PCH  0.38680    0.25550   1.514 0.139867
PCE_PCH       -11.69329    2.56478  -4.559 7.12e-05 ***
WPU101_PCH    -0.21318    0.36451  -0.585 0.562749
GPDIC1_PCH    -0.45500    0.83117  -0.547 0.587888
RRVRSQ156N_PCH -0.48251    0.73991  -0.652 0.518983
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.86 on 32 degrees of freedom
Multiple R-squared:  0.8845,    Adjusted R-squared:  0.8521
F-statistic: 27.24 on 9 and 32 DF,  p-value: 1.483e-12
>

```

12) Use model4 to predict the values on the 40% testing set. Store the results in the distPred variable and paste beginning of variable data below. Hint use head() for this.



The screenshot shows the RStudio environment. The top pane contains an R script with the following content:

```
1 # NAME: Valeria Aybar
2 # PANTHER ID: 6181336
3 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
4 # work is my own and that none of it is the work of any other person.
5
6 #=====
7 # PROGRAMMER: Valeria Aybar
8 # PANTHER ID: 6181336
9 #
10 # CLASS: CAP4830
11 # SECTION: RVC 1231
12 # SEMESTER: Spring 2023
13 # CLASSTIME: Online
14 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
15 # work is my own and that none of it is the work of any other person
16 #=====
17
```

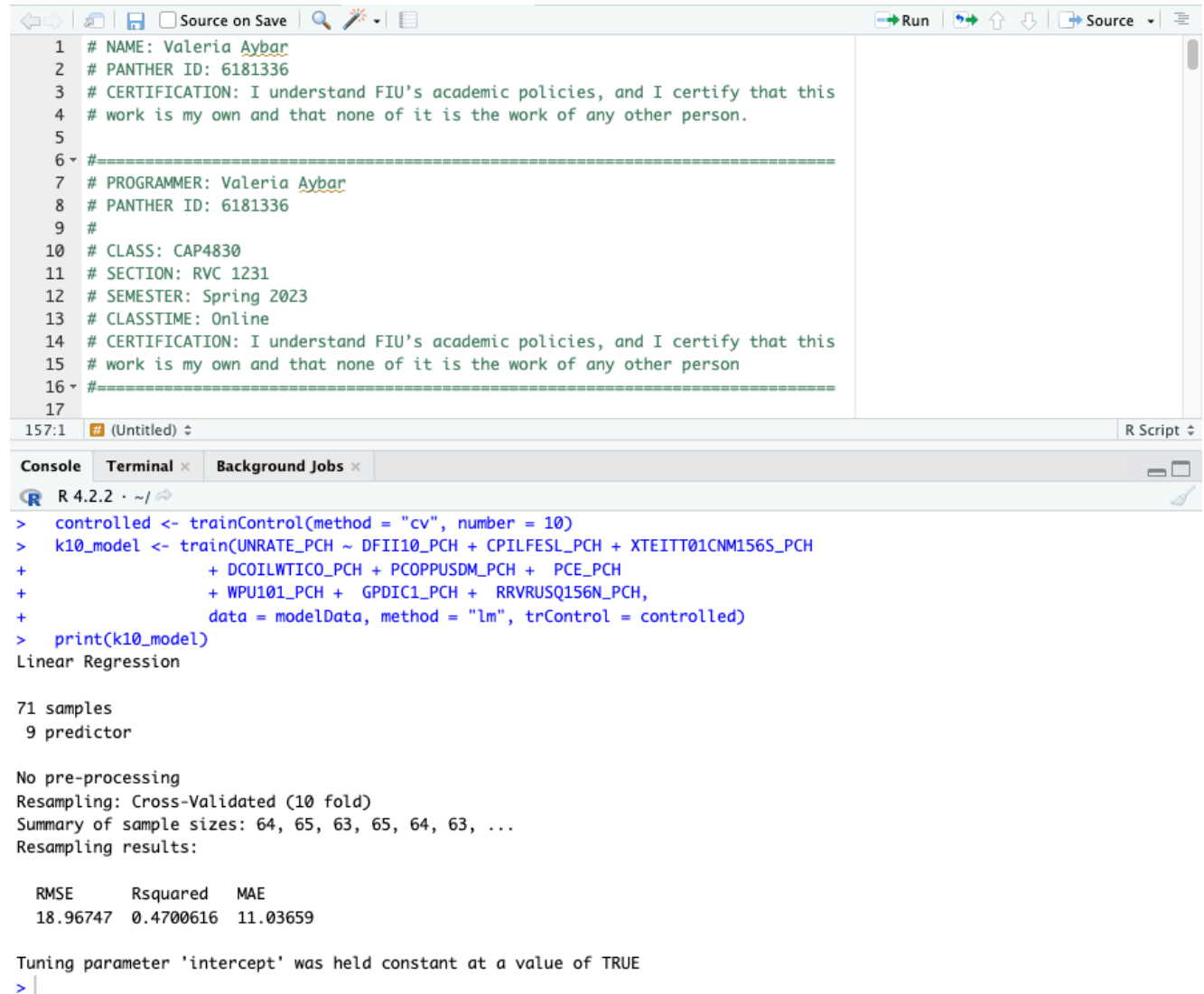
The bottom pane shows the R console with the following output:

```
> distPred <- predict(model4, testData)
> head(distPred)
      1      3      5      9     10     13
-1.796136  5.572317 13.190423  1.187460  1.681496 31.271887
> |
```


13) Using model4 calculate prediction accuracy and error rates then use ggplot that shows actual vs Predicted values. Paste your plot below.



14) Run a k-fold cross validation with k=10. Paste the print of the model below.



The screenshot shows the R Studio interface. The script editor at the top contains a file named 'Untitled' with the following content:

```
1 # NAME: Valeria Aybar
2 # PANTHER ID: 6181336
3 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
4 # work is my own and that none of it is the work of any other person.
5
6 #=====
7 # PROGRAMMER: Valeria Aybar
8 # PANTHER ID: 6181336
9 #
10 # CLASS: CAP4830
11 # SECTION: RVC 1231
12 # SEMESTER: Spring 2023
13 # CLASSTIME: Online
14 # CERTIFICATION: I understand FIU's academic policies, and I certify that this
15 # work is my own and that none of it is the work of any other person
16 #=====
17
```

The console window at the bottom shows the output of the R script:

```
R 4.2.2 ~ /
> controlled <- trainControl(method = "cv", number = 10)
> k10_model <- train(UNRATE_PCH ~ DFII10_PCH + CPILFESL_PCH + XTEITT01CNM156S_PCH
+                   + DCOILWTICO_PCH + PCOPUSDM_PCH + PCE_PCH
+                   + WPU101_PCH + GPDIC1_PCH + RRVRSQ156N_PCH,
+                   data = modelData, method = "lm", trControl = controlled)
> print(k10_model)
Linear Regression

71 samples
9 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 64, 65, 63, 65, 64, 63, ...
Resampling results:

      RMSE      Rsquared    MAE
18.96747  0.4700616  11.03659

Tuning parameter 'intercept' was held constant at a value of TRUE
>
```

15) Put this file and your r-script in a folder name CAP4830HW2 and zip the folder then upload the zipped folder to canvas. Make sure you completed step 0. If you do not have this done you will have 40% deducted from your homework grade.