# CAP4830

# Homework 2

**Instructions**:

Download the excel file for this assignment from canvas and complete the steps below. All the needed R code for this homework can be found in Module 10.

**Important**: Do not copy someone’s code. You may be randomly selected to explain your code. If you cannot explain your code, you will get a zero for the homework grade. If you are stuck please ask me. I will gladly help you. Remember the goal is to learn. ☺

**Due Date:** If you email me this homework, I will not except it. Do not wait until 11:58pm to upload it. If you are late and you email it to me stating there was issues with canvas uploading it. I will not except it. Own your work and problems.

*The excel file has data on the following:*

| UNRATE\_PCH | Unemployment Rate, Percent Change, Quarterly, Seasonally Adjusted |
| --- | --- |
| DFII10\_PCH | 10-Year Treasury Inflation-Indexed Security, Constant Maturity, Percent Change, Quarterly, Not Seasonally Adjusted |
| CPILFESL\_PCH | Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average, Percent Change, Quarterly, Seasonally Adjusted |
| XTEITT01CNM156S\_PCH | Ratio of Exports to Imports for China, Percent Change, Quarterly, Seasonally Adjusted |
| DCOILWTICO\_PCH | Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma, Percent Change, Quarterly, Not Seasonally Adjusted |
| PCOPPUSDM\_PCH | Global price of Copper, Percent Change, Quarterly, Not Seasonally Adjusted |
| PCE\_PCH | Personal Consumption Expenditures, Percent Change, Quarterly, Seasonally Adjusted Annual Rate |
| WPU101\_PCH | Producer Price Index by Commodity: Metals and Metal Products: Iron and Steel, Percent Change, Quarterly, Not Seasonally Adjusted |
| GPDIC1\_PCH | Real Gross Private Domestic Investment, Percent Change, Quarterly, Seasonally Adjusted Annual Rate |
| RRVRUSQ156N\_PCH | Rental Vacancy Rate for the United States, Percent Change, Quarterly, Not Seasonally Adjusted |

**Please complete the following steps:**

0) Complete information below:

NAME:

PANTHER ID:

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.

Copy this header with your information filled-in on the top of your r-script.

#=============================================================================

# PROGRAMMER: Camron Cisneros

# PANTHER ID: Your panther ID

#

# CLASS: COP2210

# SECTION: Your class section: example U01

# SEMESTER: The current semester: example Spring 2021

# CLASSTIME: Your CAP4830 course meeting time :example T/TH 9:00-10:15 am

# 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.**

**Paste your R console output below:**

[1] "DATE" "UNRATE\_PCH"

[3] "DFII10\_PCH" "CPILFESL\_PCH"

[5] "XTEITT01CNM156S\_PCH" "DCOILWTICO\_PCH"

[7] "PCOPPUSDM\_PCH" "PCE\_PCH"

[9] "WPU101\_PCH" "GPDIC1\_PCH"

[11] "RRVRUSQ156N\_PCH"

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

**UNRATE\_PCH = b\_0 + b\_1\*DFII10\_PCH + b\_2 \* CPILFESL\_PCH + b\_3 \* XTEITT01CNM156S\_PCH**

**+ b\_4\* DCOILWTICO\_PCH + b\_5 \* PCOPPUSDM\_PCH + b\_6 \* PCE\_PCH**

**+ b\_7 \* WPU101\_PCH + b\_8 \* GPDIC1\_PCH + b\_9 \* RRVRUSQ156N\_PCH**

**Paste you model’s summary below:**

**Hint: use lm and summary**

Call:

lm(formula = UNRATE\_PCH ~ DFII10\_PCH + CPILFESL\_PCH + XTEITT01CNM156S\_PCH +

DCOILWTICO\_PCH + PCOPPUSDM\_PCH + PCE\_PCH + WPU101\_PCH + GPDIC1\_PCH +

RRVRUSQ156N\_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**

(Intercept) 0.006960 \*\*

DFII10\_PCH 1.83e-06 \*\*\*

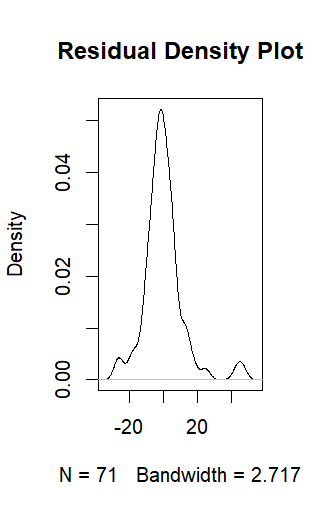
XTEITT01CNM156S\_PCH 0.000280 \*\*\*

DCOILWTICO\_PCH 0.000709 \*\*\*

PCE\_PCH 2.91e-09 \*\*\*

**5) Plot the model1’s residual Density Function**

**Paste the plot below:**



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

Shapiro-Wilk normality test

data: resid(model1)

W = 0.87998, p-value = 6.009e-06

Reject the Null hypothesis because the p-value is less than 0.05.

This tells us that residuals are not normally distributed.

**7) Create model2 which is a refinement of model1 by removing all regressors that are statistically insignificant with . Hence you are removing regressors with p-value > 0.55. Paste you model’s summary below:**

Call:

lm(formula = UNRATE\_PCH ~ DFII10\_PCH + XTEITT01CNM156S\_PCH +

DCOILWTICO\_PCH + PCOPPUSDM\_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 \*\*\*

PCOPPUSDM\_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.**

Adjusted R-Squared(model1) = .8217 sqrt(.8217) = .90647

Adjusted R-Squared(model2) = .8266 sqrt(.8293) = .90918

0. 8266 - 0.8217 = 0.0049 . 90918- .90647 = .00271

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

Accuracy: 27.67 %

Error rate: 72.33 %

**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.**

Call:

lm(formula = UNRATE\_PCH ~ DFII10\_PCH + XTEITT01CNM156S\_PCH +

PCE\_PCH, data = modelData)

Residuals:

Min 1Q Median 3Q Max

-31.819 -4.999 -1.995 3.302 57.723

Coefficients:

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

(Intercept) 11.46894 1.96518 5.836 1.71e-07 \*\*\*

DFII10\_PCH 0.09794 0.01661 5.897 1.34e-07 \*\*\*

XTEITT01CNM156S\_PCH 0.43884 0.31563 1.390 0.169

PCE\_PCH -10.21733 1.04090 -9.816 1.35e-14 \*\*\*

---

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

Residual standard error: 14.07 on 67 degrees of freedom

Multiple R-squared: 0.7874, Adjusted R-squared: 0.7779

F-statistic: 82.73 on 3 and 67 DF, p-value: < 2.2e-16

I chose to use these 3 because they have the smallest P-value

**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.**

Call:

lm(formula = UNRATE\_PCH ~ DFII10\_PCH + CPILFESL\_PCH + XTEITT01CNM156S\_PCH +

DCOILWTICO\_PCH + PCOPPUSDM\_PCH + PCE\_PCH + WPU101\_PCH + GPDIC1\_PCH +

RRVRUSQ156N\_PCH, data = trainingData)

Residuals:

Min 1Q Median 3Q Max

-4.5596 -1.2634 -0.2741 1.2463 8.7434

Coefficients:

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

(Intercept) 3.523021 1.843832 1.911 0.0650 .

DFII10\_PCH 0.003768 0.005565 0.677 0.5032

CPILFESL\_PCH -0.659798 3.906936 -0.169 0.8670

XTEITT01CNM156S\_PCH 0.347497 0.139799 2.486 0.0183 \*

DCOILWTICO\_PCH 0.146528 0.075681 1.936 0.0617 .

PCOPPUSDM\_PCH 0.060353 0.065212 0.925 0.3616

PCE\_PCH -3.373137 0.701397 -4.809 3.46e-05 \*\*\*

WPU101\_PCH 0.008084 0.099872 0.081 0.9360

GPDIC1\_PCH -0.946340 0.190738 -4.961 2.22e-05 \*\*\*

RRVRUSQ156N\_PCH 0.251574 0.143991 1.747 0.0902 .

---

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

Residual standard error: 3.037 on 32 degrees of freedom

Multiple R-squared: 0.8582, Adjusted R-squared: 0.8184

F-statistic: 21.53 on 9 and 32 DF, p-value: 3.58e-11

**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.**

$coefficients

(Intercept) DFII10\_PCH CPILFESL\_PCH

5.255946820 0.002786556 -3.373838299

XTEITT01CNM156S\_PCH DCOILWTICO\_PCH PCOPPUSDM\_PCH

0.161751939 0.154821389 0.007511849

PCE\_PCH WPU101\_PCH GPDIC1\_PCH

-4.481817939 0.060859700 -0.648719831

RRVRUSQ156N\_PCH

-0.008273218

$residuals

30 49 2 11 47

1.125868996 -0.334693971 7.277133719 4.949106315 -1.342865799

1 32 9 61 57

6.688954734 -5.978452090 -1.414918812 -0.506563308 -3.531140771

13 68 36 58 14

-2.338249069 -1.122746680 -0.924824947 -0.395433031 2.509897822

42 20 15 54 55

-1.893170211 0.006327087 -3.052852065 0.846831638 -1.436600456

50 45 67 18 38

-2.306173755 -2.543898666 -0.451132423 3.036503058 -2.718517009

25 52 60 63 22

2.825067767 -0.555113081 -1.401505306 2.886475994 9.202914809

23 39 31 62 37

0.708351504 -1.865245828 0.159609420 -1.835829295 -0.568445109

12 46 19 44 59

-1.649774417 1.305294327 2.688465311 -4.464218437 -0.708651206

51 10

-1.430185010 0.554398252

$effects

(Intercept) DFII10\_PCH CPILFESL\_PCH

4.475668963 -1.227387581 -1.271261677

XTEITT01CNM156S\_PCH DCOILWTICO\_PCH PCOPPUSDM\_PCH

-4.443220118 -7.422132674 3.391420826

PCE\_PCH WPU101\_PCH GPDIC1\_PCH

20.810520199 -0.201792859 -7.701725121

RRVRUSQ156N\_PCH

0.191563620 -0.737836842 -4.299828286

0.155225022 -1.332076083 6.471648280

0.001120583 1.241096895 -1.304604751

2.435737565 -2.581006959 2.542958794

-1.014267016 -1.823104197 3.961635339

-3.364641740 -3.558756434 2.183452937

1.581435393 2.482295089 10.680690874

-1.474842885 -1.288394381 -6.442856413

-0.367001480 -0.167113246 0.915450671

1.970401650 6.084693088 -5.875928706

-0.985496908 -0.489603255 -0.206463916

$rank

[1] 10

$fitted.values

30 49 2 11 47 1

-2.8559690 -1.4725360 -7.2771337 -4.9491063 -4.7010942 -2.1435047

32 9 61 57 13 68

1.0661721 -2.3586612 -1.9727767 -0.1184992 0.2255691 6.6783067

36 58 14 42 20 15

-3.3222751 -1.8772970 -2.5098978 -2.0891298 4.1603429 -1.2636979

54 55 50 45 67 18

-1.5225116 -1.2844895 -3.8287962 -4.4561013 0.4511324 0.6671969

38 25 52 60 63 22

0.6859970 9.6749322 -2.0938969 -1.0178447 -1.1165660 3.2970852

23 39 31 62 37 12

13.7360885 -1.0393142 0.1925006 -2.4014607 -0.2380049 -3.0482156

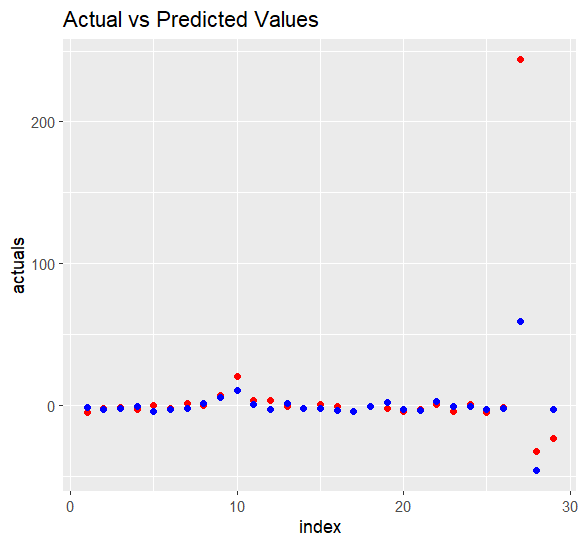
46 19 44 59 51 10

-3.4558343 0.1686747 0.6180684 -3.1673188 0.1229950 -3.1687783

$assign

[1] 0 1 2 3 4 5 6 7 8 9

**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.**

Linear Regression

71 samples

7 predictor

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 64, 63, 65, 64, 63, 66, ...

Resampling results:

RMSE Rsquared MAE

21.21515 0.4221049 12.28693

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.**