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 |
| 1-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 |
| 1-XTEITT01CNM156S\_PCH | Ratio of Exports to Imports for China, Percent Change, Quarterly, Seasonally Adjusted |
| 1-DCOILWTICO\_PCH | Crude Oil Prices: West Texas Intermediate (WTI) - Cushing, Oklahoma, Percent Change, Quarterly, Not Seasonally Adjusted |
| 1-PCOPPUSDM\_PCH | Global price of Copper, Percent Change, Quarterly, Not Seasonally Adjusted |
| 1-PCE\_PCH | Personal Consumption Expenditures, Percent Change, Quarterly, Seasonally Adjusted Annual Rate |
| 1-WPU101\_PCH | Producer Price Index by Commodity: Metals and Metal Products: Iron and Steel, Percent Change, Quarterly, Not Seasonally Adjusted |
| 1-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: Garcia Milord

PANTHER ID: 6168616

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: Your name

# 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.

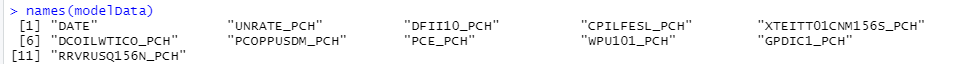
#=============================================================================

1) Read the excel file “CAP4830\_HW2\_Data.xlsx” data into R and store the imported data in a variable named “modelData”.

Sol: modelData <- read.xlsx(file.choose(), 1)

2) Output the names of the modelData dataframe.

Paste your R console output below:



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

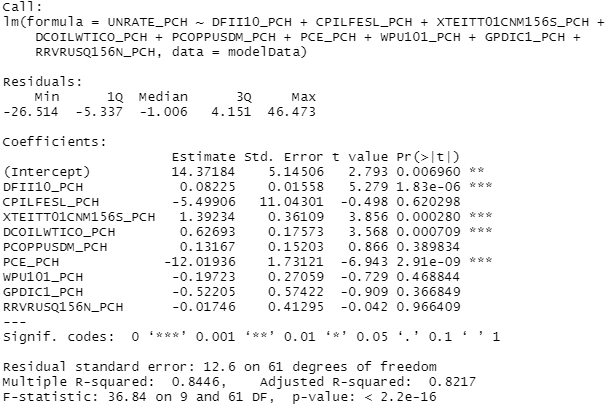
UNRATE\_PCH = b\_0 + b\_1\*DFII10\_PCH + b\_2 \* CPILFESL\_PCH + b\_3 \* XTEITT01CNM156S\_PCH

+ 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**



4) List all the estimate parameters from step 3 that are statistically significant for all

> model1\_Stats$coefficients[[1,4]]

[1] 0.006960344

> model1\_Stats$coefficients[[2,4]]

[1] 1.82546e-06

> model1\_Stats$coefficients[[4,4]]

[1] 0.0002800858

> model1\_Stats$coefficients[[5,4]]

[1] 0.0007085496

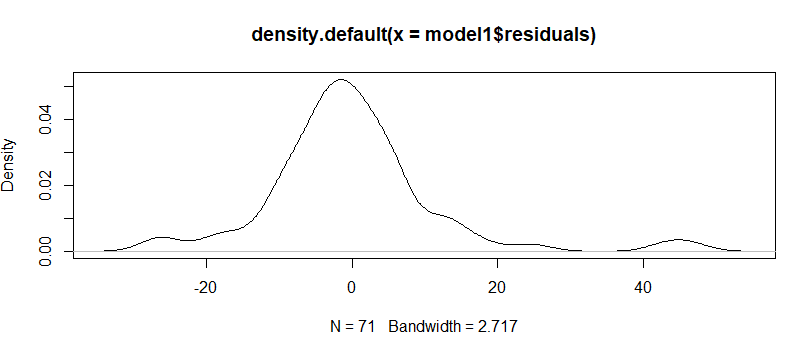
> model1\_Stats$coefficients[[7,4]]

[1] 2.906008e-09

>

5) Plot the model1’s residual Density Function

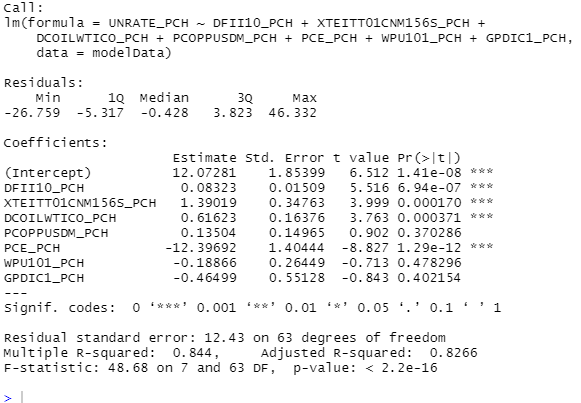
Paste the plot below:



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

From run the Shapiro test this distribution does not have normal distribution the p value has such great difference from each other.

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



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

Since removing all statistically insignificant the regressors p >0.55 from model1 was Adjusted R-squared: 0.8217 and now for model2 is Adjusted R-squared: 0.8266. difference is between the two Adjusted R-square is .0059.

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

Sol: model2\_MSE <-mean(model2$residuals^2)

model2\_MSE or this way

model2.1\_MSE <- data.frame(modelPred=predict(model2),

actual= HWData$UNRATE\_PCH)

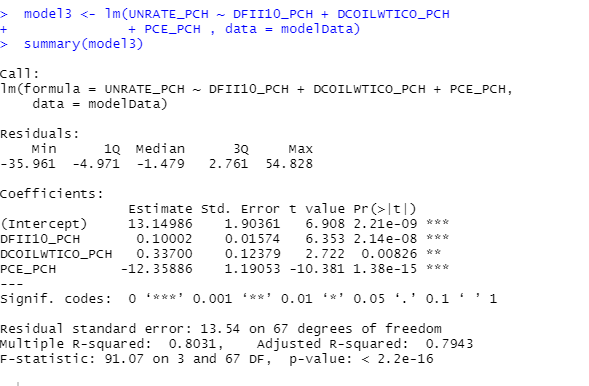
head(model2.1\_MSE)

model2.1\_MSECal <-mean ((model2.1\_MSE$actual-model2.1\_MSE$modelPred)^2)

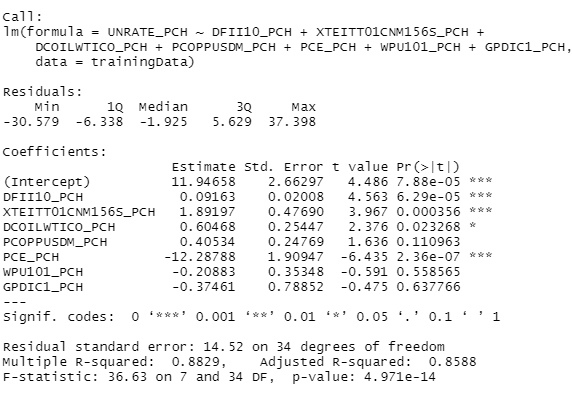
model2.1\_MSECal

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.

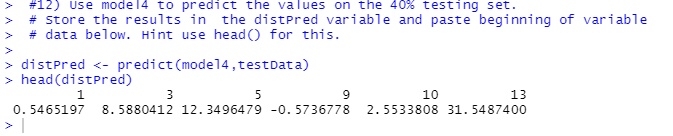
Sol: The reason I pick these three for my summary of model3 because p-value <= 0 and r squared is closest of any combination of three I had come up with least amount significant figures compare to the original dataset.



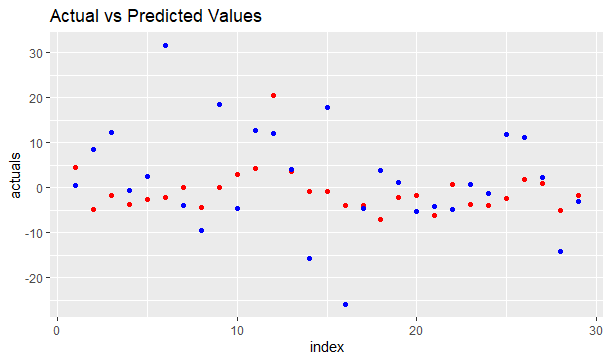
11) Create model4 that uses a manual sampling technique with a training set of 60% of the data and a testing set of 40%. Paste the summary of the model below.



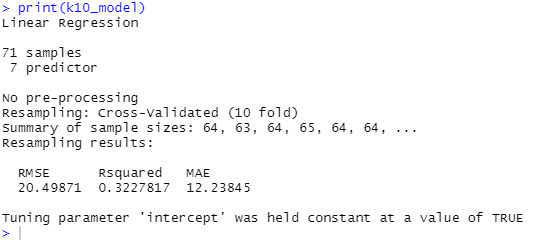
12) Use model4 to predict the values on the 40% testing set. Store the results in the distPred variable and paste beginning of variable data below. **Hint use head() for this.**



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.



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.