# Practical 8 – CP2403 – Due 11th January 2020 – 5pm

Ensure you add you name to the top of the Jupyter notebook before submission

**Part 1** – Download the Jupyter notebook for Module 8 and run the notebook

**Part 2**

Download the Jupyter Notebook Template for Prac 8 from LearnJCU. Complete the template & run the code. Refer to Module 8 Lecture Jupyter Notebook for help

Complete the questions in Part 3 as you work on the Prac 8 template

**Part 3**

**Scenario 1**

Predict co2emission(y) using relectricperperson(x1) and oilperperson(x2)

|  |
| --- |
| **1: Regression Analysis results** |
|  |
| **2: Regression equation** |
| Co2 Emissions = -1.669e-6 +3.434e +6(relectricperperson\_c) +-9.47e8(oilperperson) |

**Scenario 2**

Predict employrate (y) using linear regression. Use relectricperperson (x) as the explanatory variable

|  |
| --- |
| **1: Scatter plot** |
|  |
| **2: Regression Analysis results** |
|  |
| **3: Regression equation** |
| Employee Rate = 5.607e-15+ 0.0008(relectricperperson\_c) |

**Scenario 3**

Predict employrate(y) using linear regression. Use relectricperperson(x) – order 2 as the explanatory variable

|  |
| --- |
| **1: Scatter plot with regression line** |
|  |
| **2: Regression Analysis results** |
|  |
| **3: Regression line** |
| Employee Rate = -0.5780 + 2.037 e-7 (relectricperperson\*\*2) |

**Scenario 4**

Perform multiple and polynomial regression between oilperperson(x1), co2emissions(x2), relectricperperson(x3 - order 2) and employrate(y)

|  |
| --- |
| **1: Regression Analysis results** |
|  |
| **2: Regression line** |
| Employee Rate = -0.8489 + 0.6155(oilperperson\_c)+1.533e-11(co2emission\_c)+2.047e-7(relectricperperson\_c\*\*2) |
| **3: qqplot** |
|  |
| **4: Conclusion from qqplot** |
| The QQplot shows that the model created is not explain response variable because the quantitiles is not normally distributed and it more looks like heavy tailed. |
| **5: percentage of observations over 2 standardized deviation** |
|  |
| **6: percentage of observations over 2.5 standardized** |
|  |
| **7: Conclusion from observations over 2 std and 2.5 std** |
| 8% of observation have more than 2 standardized residual and 0% or none of the observation have 2.5 standardized residual which means and not so good model because it meets one of the conditions for up to 5%. |

**Scenario 5**

Experiment with multiple and polynomial regression between oilperperson(x1), co2emissions

(x2), relectricperperson (x3) and employrate (y)

|  |
| --- |
| **1: Regression Analysis results** |
|  |
| **2: Regression line** |
| Employment rate = -0.1582 + 8.165e-12(co2emissions\_c)+0.0014(relectricperperson\_c)+0.482(oilperperson\_cc\*\*2) |
| **3: qqplot** |
|  |
| **4: Conclusion from qqplot** |
| The QQplot shows that the model created is not explain response variable because the quantitiles is not normally distributed and it looks like heavy tailed. |
| **5: percentage of observations over 2 standardized deviation** |
|  |
| **6: percentage of observations over 2.5 standardized** |
|  |
| **7: Conclusion from observations over 2 std and 2.5 std** |
| 5% of observation have more than 2 standardized residual and 0% or none of the observation have 2.5 standardized residual means the model is example of good fit model. |