1. **Problem definition and description**

The main purpose of the project is to evaluate the absorption isotherm of cobalt ions onto the hydroxyapatite beads with regression.

In statistical modeling, regression analysis is commonly used for estimating the relationships between a dependent variable and one or more independent variables. Therefore, to conduct regression, the correlation between the variables should be calculated to meet the condition. After building the model, various statistics were looked over to see if the model fits the data well.

1. **Core code**

# Check correlation between the variables

import pandas as pd

import numpy as np

from scipy import stats

import matplotlib.pyplot as plt

df = pd.read\_csv("AdsorptionCobaltHydroxyapatite.csv")

def calculate\_correlation(x: np.array, y: np.array):

r, p = stats.pearsonr(x, y)

print("corr:", r, "p-value:", p)

fig = plt.figure(figsize = [6, 6]) # fig

plt.scatter(df['Co (mg/L)'], df['Temperature (K)'])

# Build Linear Regression Model with statsmodels

import statsmodels.api as sm

X = sm.add\_constant(X)

# Ordinary Least Square Methods

model = sm.OLS(Y, X).fit()

predictions = model.predict(X)

print\_model = model.summary()

print(print\_model)

from mpl\_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize = [8, 8])

ax = fig.gca(projection = '3d')

ax.scatter3D(X['Co (mg/L)'], X['Temperature (K)'], Y)

# Data for a three-dimensional line

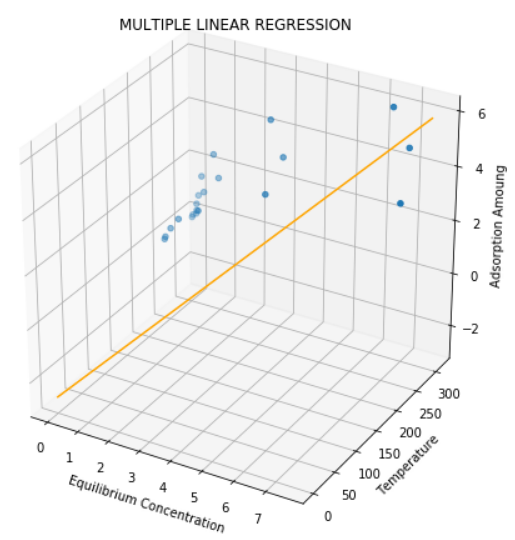
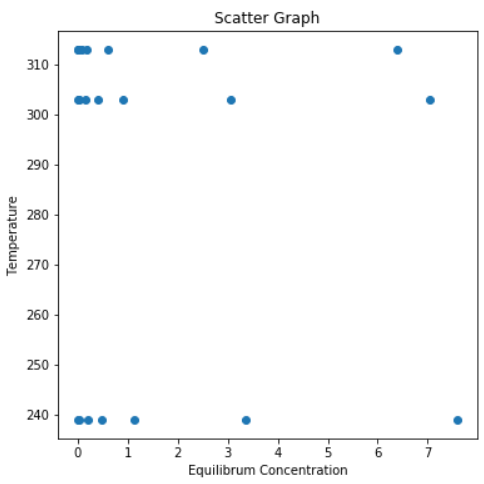
xline = np.linspace(0, X['Co (mg/L)'].max(), 1000)

yline = np.linspace(0, X['Temperature (K)'].max(), 1000)

zline = 0.639108 \* xline + 0.011671 \* yline -2.655106

ax.plot3D(xline, yline, zline, 'gray')

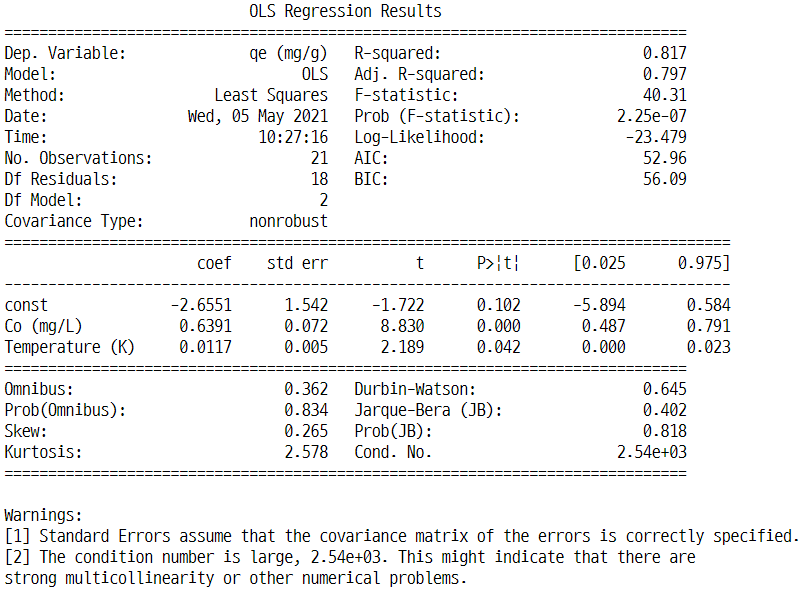
1. **Results and plots**



**Figure 1, 2 Scatter plot, Regression result**



**Figure 3. Correlation coefficient between temperature and equilibrium concentration**

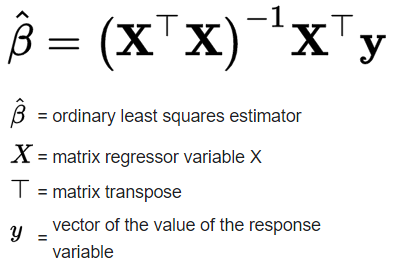


**Figure 4. Regression Results (OLS)**

1. **Discussion**

The project could be divided into two parts: i) to test if the independent variables are correlated to each other, ii) to build a model and to look over the statistics.

In part i), the pearson correlation coefficient were calculated and plotting scatterplot were conducted (Figure 1, Figure 3). It shows that the p value of the pearson correlation was 0.7828, which means the two independent variables, temperature and equilibrium concentration, are not correlated. Plus, we could see the linear relationship with the variables when we see the scatterplot. Therefore, the variables meet the regression condition.

Regression model was built with statsmodel.ols. Ordinary Least Squares Methods were used to build the model. It makes sure the square of the residuals be minimal, when figuring out the linear regression equation.

As we can see in Figure 4, the model could explain the z value (absorption isotherm of cobalt ions) for 81.7%, according to R-squared statistics. However, there were some warnings like multicollinearity or other numerical problems: this might cause the difference of the scales of each variables, and the small number of learning materials. (It would be better if there were more than 30 data)

1. **Refernces**