CP2403 - Project – Part 2 – Multiple_REGRESSION

First Name: Kantapong
Last Name: Wongsanguan

1: Scatter plots between each explanatory variable and response variable

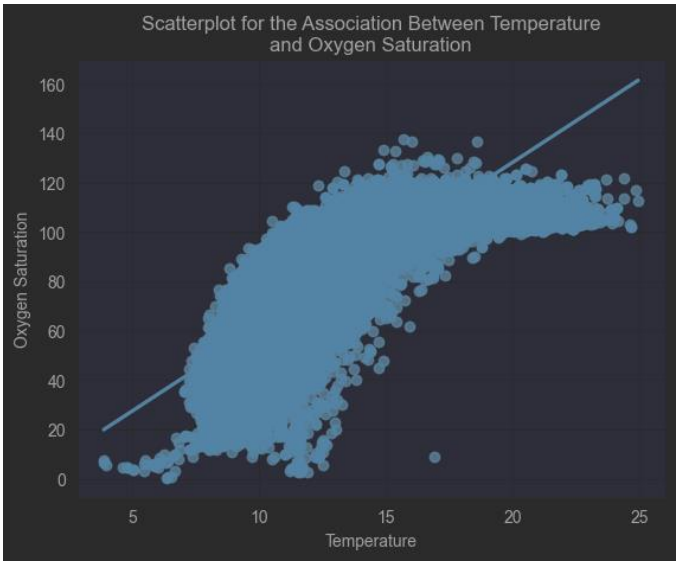


Figure 1: Scatter plot of water temperature vs oxygen saturation.

The scatter plot above represents a relationship between the water temperature and the oxygen saturation. Roughly, we can estimate that as the water temperature increase, the oxygen saturation also rises, as supported by the line of best fit through the plot.

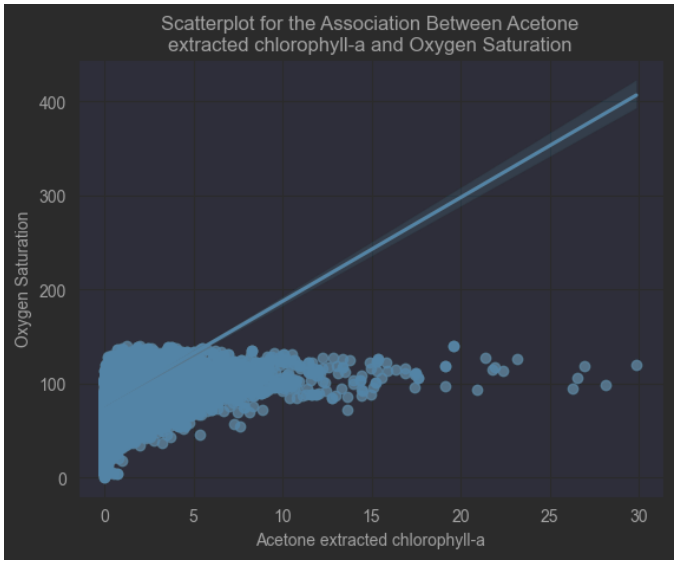


Figure 2: Scatter plot of correlation between phaeophytin concentration (x) and oxygen saturation (y).

The scatter plot above represents a relationship between the phaeophytin concentration (x) and oxygen saturation (y). This plot indicates that there is a weak positive correlation between the two variables, as one increases, the other also tend to increase.

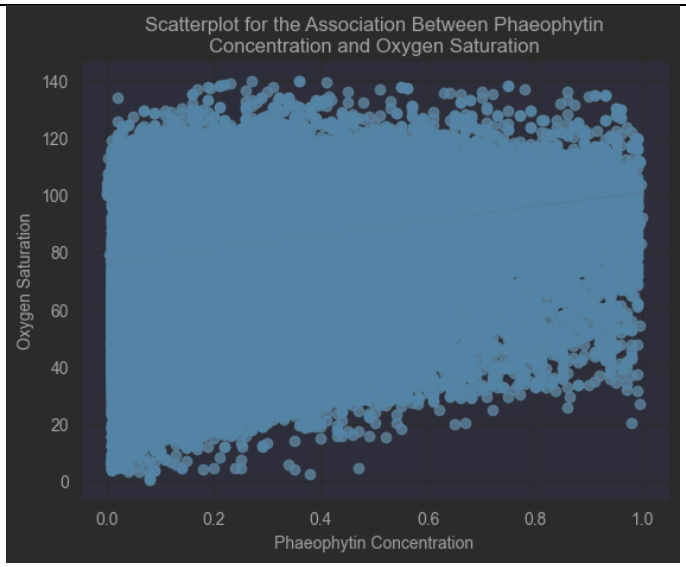


Figure 3: Scatter plot of correlation between phaeophytin concentration (x) and oxygen saturation (y).

This plot shows a weaker yet still increasing correlation between phaeophytin concentration and oxygen saturation, thus, a weak positive relationship between the two variables.

2: List all the explanatory variables selected for regression analysis. Justify your selection

Both temperature and chlorophyll-a concentration are logical choices as explanatory variables because they represent important physical and biological factors that can impact oxygen saturation in water.

Temperature (T_degC_c) significantly influences the oxygen solubility in water. As water temperature rises, its capacity to retain dissolved oxygen diminishes. This phenomenon is a widely recognized physical principle referred to as the temperature-solubility relationship. Consequently, incorporating temperature as an explanatory variable is imperative when investigating the factors that impact oxygen saturation in aquatic ecosystems.

Chlorophyll-a concentration (ChlorA_c) serves as a valuable indicator of phytoplankton abundance in water. Phytoplankton are primary contributors to aquatic ecosystems, and their growth and metabolic activities can exert an influence on oxygen levels. During photosynthesis, phytoplankton release oxygen into the water, bolstering oxygen saturation. Conversely, during respiration, they consume oxygen, leading to a reduction in oxygen saturation. Therefore, considering chlorophyll-a concentration is pivotal from a biological perspective when delving into the study of oxygen saturation in aquatic environments.

Phaeophytin Concentration (Phaeop): Phaeophytin is a chemical compound that plays a role in facilitating the photosynthetic process in plants and purple bacteria. Photosynthesis is the mechanism employed for generating oxygen, so alterations in phaeophytin concentration are expected to have a direct impact on the level of oxygen present in the water.

3: Regression analysis results

OLS Regression Results

Dep. Variable:	O2Sat_c	R-squared:	0.760
Model:	OLS	Adj. R-squared:	0.760
Method:	Least Squares	F-statistic:	2.269e+05
Date:	Thu, 28 Sep 2023	Prob (F-statistic):	0.000
Time:	17:11:13	Log-Likelihood:	-8.4898e+05
No. Observations:	215536	AIC:	1.698e+06
Df Residuals:	215532	BIC:	1.698e+06
Df Model:	3		
Covariance Type:	nonrobust		

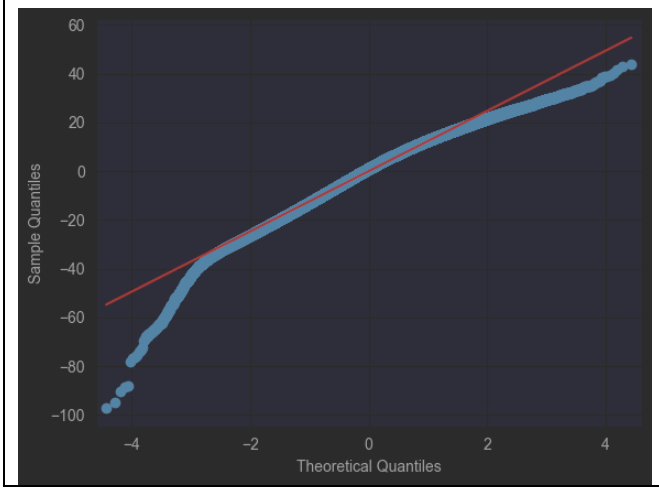
	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.817e-14	0.027	6.79e-13	1.000	-0.052	0.052
T_degC_c	6.8685	0.009	773.420	0.000	6.851	6.886
ChlorA_c	4.0366	0.049	82.567	0.000	3.941	4.132
Phaeop_c	15.3829	0.204	75.463	0.000	14.983	15.782

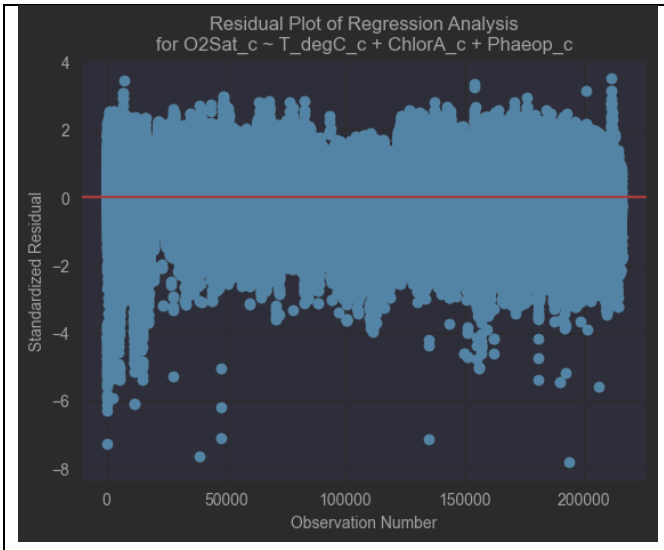
Omnibus:	5595.255	Durbin-Watson:	0.245
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6039.534
Skew:	-0.406	Prob(JB):	0.00
Kurtosis:	3.118	Cond. No.	23.7

4: Regression equation/line

O2Sat_c = 1.817e-14 + 6.8685(T_degC_c) + 4.0366(ChlorA_c) + 15.3829(Phaeop_c)

5: qqplot





6: Conclusion from qqplot

The values shown on the qqplot fit the line well for -2 < x < 2, but trail away from the line beyond these values. Although, the values do not converge on the line again. There are no major outliers, and overall, the line appears to be a good fit.

7: percentage of observations over 2 standardized deviation

3.739%

8: percentage of observations over 2.5 standardized

0.925%

9: Conclusion from observations over 2 std and 2.5 std

Standardized Residuals Over 2 Standard Deviations (2SD):
The percentage of residuals with values greater than 2 standard deviations is approximately 3.739%.
This indicates that a small proportion of the residuals deviate significantly from the expecte d values. However, this percentage is still within an acceptable range, suggesting that the m odel generally fits the data reasonably well.
Standardized Residuals Over 2.5 Standard Deviations (2.5SD):
The percentage of residuals with values greater than 2.5 standard deviations is approximate ly 0.925%
This percentage is even smaller, indicating that very few residuals exhibit extreme deviatio ns from the expected values.
Such a low percentage suggests that the model has a strong overall fit to the data.

In conclusion, the regression model demonstrates a strong fit to the data, as indicated by th e high R-squared value and the low percentage of standardized residuals exceeding 2SD an d 2.5SD. The explanatory variables, temperature (T_degC_c), Phaeophytin concentrationan d(Phaeio_c), and chlorophyll-a concentration (ChlorA_c), have a statistically significant im pact on oxygen saturation (O2Sat_c). Therefore, we can infer that temperature, Phaeophyti n concentration and chlorophyll-a concentration are important factors in explaining variatio

ns in oxygen saturation in aquatic systems.

Implications:
In terms of temperature Influence, the positive coefficient of T_degC_c indicates that as water temperature increases, oxygen saturation tends to increase as well. Ensuring proper oversight and control of water temperature is essential for preserving sufficient oxygen levels in aquatic environments. So as, recognizing the significance of Phaeop_c is essential for the efficient management and preservation of ecosystems. This recognition suggests that promoting the proliferation of plants and purple bacteria, known to produce phaeophytin, has the potential to boost oxygen levels in aquatic settings.

While the significance of Chlorophyll-a shows he positive coefficient of ChlorA_c suggesting that higher concentrations of chlorophyll-a, indicative of increased phytoplankton biomass, lead to higher oxygen saturation levels. Controlling nutrient levels to regulate chlorophyll-a concentrations becomes crucial in sustaining optimal oxygen levels within water ecosystems.

Lastly, in the aspect of the management and conservation, recognizing the impact of temperature and chlorophyll-a is fundamental for successful ecosystem preservation and management, allowing for the mitigation of detrimental factors and the promotion of aquatic habitat well-being.