# CP2403 - Assignment – Part 2 – Task 3: Linear Regression

First Name:

Last Name:

|  |
| --- |
| **1: Data Selection** |
| For this analysis, I selected   1. independent variable (X) 'Salnty' (Salinity) 2. dependent variable (Y) 'T\_degC' (Temperature) |
| **2: Scatter plot with regression line** |
|  |
| **3: Regression Analysis results** |
| **Regression Analysis Results:**  Slope (m): -3.5382485975085287  Intercept (b): 130.573950544648  Correlation Coefficient (r): -0.4449061526034401  P-value: 0.0  Standard Error: 0.0077492538291272055  The linear regression analysis between Salinity (`Salnty`) and Temperature (`T\_degC`) reveals a statistically significant negative correlation. The calculated correlation coefficient (r) of approximately -0.45 indicates a moderate negative linear relationship between the two variables. The negative slope (m) of -3.54 suggests that as salinity increases, the temperature tends to decrease. The p-value of 0.0 further supports the rejection of the null hypothesis, confirming the statistical significance of the observed relationship. Therefore, based on this linear regression analysis, there is evidence to suggest that changes in salinity are associated with changes in temperature, with a tendency for lower salinity values corresponding to higher temperatures in the dataset. |
| **4: Regression line – if valid** |
| Y variable = b + m(X variable)  So, the regression line for predicting temperature based on salinity is:  T\_degC = 130.57 - 3.54.Salnty |
| **4: Residual plot – if required** |
|  |
| **5: Conclusion from residual plot – if valid** |
| The residual plot shows no discernible pattern, and the residuals are randomly scattered around the mean line, indicating that the linear regression model is appropriate for predicting temperature based on salinity in the dataset. The model captures the underlying relationships between the variables without systematic errors, supporting its validity. |