Advanced Data Analysis In Python

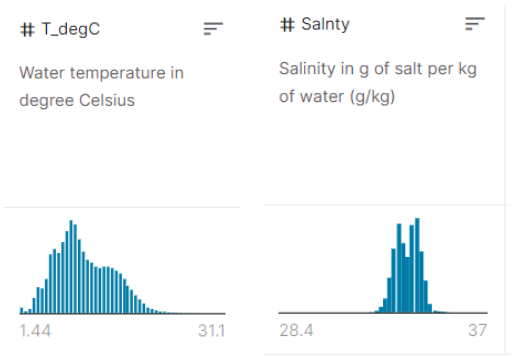
HW2 – Report **Izel Yazici - 77549**

Dataset

The CalCOFI data set represents the longest (1949-present) and most complete (more than 50,000 sampling stations) time series of oceanographic and larval fish data in the world. It includes abundance data on the larvae of over 250 species of fish; larval length frequency data and egg abundance data on key commercial species; and oceanographic and plankton data. Since it's a very large dataset, I worked with the first thousand rows.

Model

My aim in this assignment was to measure the effect of Water temperature in degree Celsius **(T_degC)** and Count - all casts consecutively numbered **(Cst_Cnt)** in estimating Salinity in g of salt per kg of water (g/kg)- **Salnty** by means of linear regression.



Linear Regression Model Formula

$$\hat{\beta} = (X'X)^{-1}X'y,$$
$$\mathrm{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1},$$
$$\sigma^2 = \frac{e'e}{n-k-1},$$
$$e = y - \hat{y},$$
$$y = X\beta + e.$$

Hypothesis

The null hypothesis states no relationship between the two variables being studied. It states that the results are due to chance and do not support the investigation's idea.

Results

Using the SciPy, we can reject the null hypothesis if the t-values we found are more significant than the t-statistics.