BDSE33 數據挖掘作業

解釋下面報表

<class 'statsmodels.iolib.summary.Summary'>

"""

OLS Regression Results

==============================================================================

Dep. Variable: a1 R-squared: 0.412

Model: OLS Adj. R-squared: 0.337

Method: Least Squares F-statistic: 5.501

Date: Mon, 05 Feb 2024 Prob (F-statistic): 2.45e-08

Time: 00:15:44 Log-Likelihood: -557.87

No. Observations: 134 AIC: 1148.

Df Residuals: 118 BIC: 1194.

Df Model: 15

Covariance Type:

===================================================================================

coef std err t P>|t| [0.025 0.975]

---------------------------------------------------------------------------------------

Intercept 7.1108 34.318 0.207 0.836 -60.848 75.070

C(season)[T.spring] -1.9504 4.965 -0.393 0.695 -11.783 7.882

C(season)[T.summer] -0.3084 4.847 -0.064 0.949 -9.907 9.290

C(season)[T.winter] -1.6128 4.601 -0.351 0.727 -10.725 7.499

C(size)[T.medium] 2.7340 4.571 0.598 0.551 -6.318 11.786

C(size)[T.small] 15.5128 4.995 3.106 0.002 5.622 25.404

C(speed)[T.low] 4.0993 5.909 0.694 0.489 -7.602 15.800

C(speed)[T.medium] -6.3026 3.914 -1.610 0.110 -14.054 1.448

mxPH 1.5398 4.055 0.380 0.705 -6.491 9.570

mnO2 0.6923 0.754 0.918 0.360 -0.801 2.185

Cl -0.0214 0.035 -0.605 0.547 -0.091 0.049

NO3 -1.4059 0.591 -2.380 0.019 -2.576 -0.236

NH4 0.0018 0.001 1.716 0.089 -0.000 0.004

oPO4 -0.0620 0.048 -1.292 0.199 -0.157 0.033

PO4 -0.0056 0.037 -0.151 0.880 -0.079 0.067

Chla -0.2021 0.105 -1.926 0.057 -0.410 0.006

==============================================================================

Omnibus: 33.679 Durbin-Watson: 1.869

Prob(Omnibus): 0.000 Jarque-Bera (JB): 58.158

Skew: 1.166 Prob(JB): 2.35e-13

Kurtosis: 5.232 Cond. No. 5.85e+04

==============================================================================

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 5.85e+04. This might indicate that there are

strong multicollinearity or other numerical problems.

"""

各項解釋如下

Part1: Summary

1. Dep. Variable: a1 應變數
2. Model: OLS
3. Method: Least Squares
4. Date: 日期
5. Time: 時間
6. No. Observations: 樣本數
7. Df Residuals: 殘差樣本數 118= 134-15-1
8. Df model: 參數數量
9. Covariance Type: nonrobust
10. R-squared: 判定係數(Coefficient of determination), 表示model對應變數的解釋程度。
11. Adj. R-squared: 調整後判定係數,考慮model中的自變數數量
12. F-statistic: F統計量，檢驗model fitting 是否顯著
13. Prob (F-statistic): F統計量對應的P值。2.45-e08 非常小，表示H0=0的機率非常小，接受H1
14. Log-Likelihood: 表示model fitting 程度
15. AIC : 赤池弘次訊息準則(Akaike’s Information Criterion, AIC), 可用來衡量多個統計模型的相對品質。
16. BIC: 舒瓦茲貝氏訊息準則(Schwarz’s Baesian Information Criterion, BIC), 也是，用來衡量多個統計模型的相對品質，公式與AIC非常類似，都與殘差平方和RSS有關。

Part2: 係數報表

1. Coef: model中每個變數的係數
2. Std. err 標準誤差，表示「樣本平均數距離母體平均數多遠」，用來衡量預測值的精確度。
3. t: t統計量，用於檢驗係數是否顯著。
4. P>|t|: t統計量對應的p值，表示該係數是否顯著。P=0.002 表示H0=0的機率很小。
5. [0.025 0.975] 信賴區間 , 左右尾各2.5%, 中間是95%，為係數的估計範圍。

part3: 殘差

1. Omnibus: Omnibus檢定是現行模型針對空值（在此情況下為截取）模型的可能性比率卡方測試。 小於 0.05 的重要值指出現行模型性能優於空值模型。
2. Prob(Omnibus): Omnibus檢定對應的p值。
3. Skew: 残差的偏度。1.166 > 1, >0右偏。
4. Kurtosis: 残差的峰度。值愈高，尾愈小。
5. Durbin-Watson: 杜賓-瓦森統計量（Durbin–Watson statistic），主要可用以檢測迴歸分析中的殘差項是否存在自我相關。
6. Jarque-Bera (JB): Jarque–Bera檢定是對樣本數據是否具有符合常態分布的偏度和峰度的適合度的檢定。
7. Prob(JB): Jarque-Bera檢驗對應的p值。
8. Cond. No.: 條件數，用於檢驗變數之間是否存在多重共線性。