Advanced Data Analysis in Python
Amir Ranjouriheravi (0071313)
Dr. Carlson

Report of Project

3.10.2019

Modelling

The estimated model is not scientifically designed one. Its subject and the variables are designed for educational purposes. Since I wanted to check the ols_reg () function for multiple regression, I looked for the data sets with more variables and also with big number of quantity (n). Finally, I find a data set related to components of wine. The dataset had analyzed different types of wines and provided the result about the ingredients of them. In the estimated model, I tried to test if there is a relation between specified ingredients with acidity level of the wine.

Variables

The control variable of the estimation is the kind of the wine. Our estimation is based on white wine analysis, so the same estimation on the red wine samples may show different outcomes. The target variable of model is “fixed acidity”, and the independent variables are quantity of 'citric acid', 'free sulfur dioxide' and 'residual sugar'. Therefore ols_reg () will calculate four beta parameters to determine if these ingredients have any relation with target value or not.

Data

The data used in this estimation has been imported from UCL machine learning repository. (https://archive.ics.uci.edu/ml/index.php)

Hypothesis

It had been assumed that the level of ‘citric acid’ and ‘sugar residuals’ have positive relation with ‘fixed acidity ’. And also, the hypothesis about ‘free sulfur dioxide’ influence is that it has not any significant effect on the level of the ‘fixed acidity’ level.

Evaluation

In the following page the result of the estimation has been shown. The result table shows that predicted amount for b_1 -the influence of citric acid- with 0.95 level estimation is in (1.83, 2.21) range. Since this range does not include 0 quantity, the hypothesis cannot be rejected, and our first assumption about its positive relation is remained.
The range of the b_2 amount-the effect of free sulfur dioxide - in the predefined condition is (-0.003, -0.006). Since the range shows weak converse relation, our first assumption in this case will be rejected. However it shows very limited relation, so if we had determined other level of testing instead of 0.95, it might not be rejected.
The range of the b_3 –the effect of sugar residuals- is (0.01, 0.02). In this case, the assumption will not been rejected, and the initial assumption will remained.

For further information, please investigate the result table.

Advanced Data Analysis in Python
Amir Ranjouriheravi (0071313)

Dr. Carlson

Report of Project

3.10.2019

You are observing the first five rows of your data frame after estimation:

	fixed acidity	volatile acidity	...	Estimated fixed acidity	Error
0	7.0	0.27	...	7.078000	-0.078000
1	6.3	0.30	...	6.902945	-0.602945
2	8.1	0.28	...	7.023367	1.076633
3	7.2	0.23	...	6.798044	0.401956
4	7.2	0.23	...	6.798044	0.401956

[5 rows x 14 columns]

Variance_Covariance matrix is equal to :

	b_0	b_1	b_2	b_3
b_0	0.001602	-0.002772	-1.298346e-05	-1.337703e-05
b_1	-0.002772	0.009092	-4.682215e-06	-1.574155e-05
b_2	-0.000013	-0.000005	5.011040e-07	-4.920360e-07
b_3	-0.000013	-0.000016	-4.920360e-07	5.634210e-06

	Estimated Parameters	Est variances	interval start	interval end
b_0	6.262653	1.601765e-03	6.184192	6.341115
b_1	2.023951	9.091771e-03	1.837021	2.210882
b_2	-0.005185	5.011040e-07	-0.006573	-0.003797
b_3	0.015461	5.634210e-06	0.010808	0.020115