Model 1

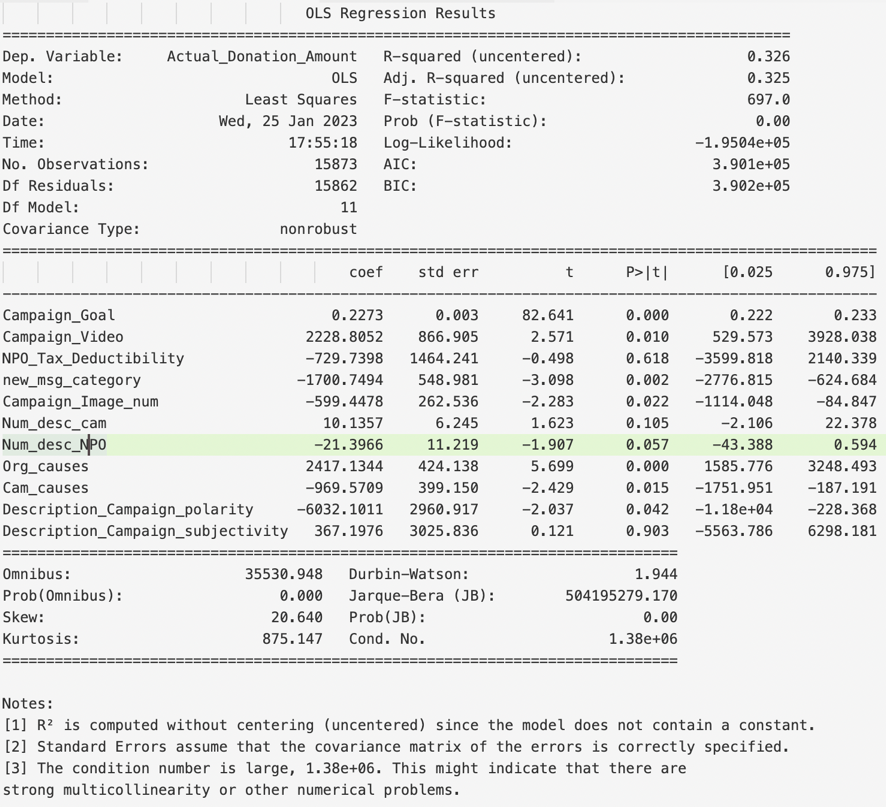
Actual\_Donation\_Amount = Campaign\_Goal + NPO\_Tax\_Deductibility + Campaign\_Image\_num + Campaign\_Video + new\_msg\_category + Num\_desc\_cam + Num\_desc\_NPO + Cam\_causes + Org\_causes + Description\_Campaign\_polarity + Description\_Campaign\_subjectivity

With a const

A picture containing text, receipt

Description automatically generated

Without a const



Conclusion: Constant is workless

Model2

Verify the reason why some coefficients are so large.

Normalize the data: Put all the data range from 0-1 by using the function

Assuming x(1), x(2), x(3)… is a list of data x

Normalized data

After normalization:

Table

Description automatically generated

The model result is:

Table

Description automatically generated

**VIF and**

'Actual\_Donation\_Amount' [(1, 1.4853126972296733),

'Campaign\_Goal' (2, 1.6428530355802324),

'NPO\_Tax\_Deductibility' (3, 11.59915109937583),

'duration\_day' (4, 2.122255706166321),

'Campaign\_Video' (5, 1.752525902819659),

'new\_msg\_category' (6, 1.717867597145572),

'Campaign\_Image\_num' (7, 4.311848192157986),

'Num\_desc\_cam' (8, 5.278887399454132),

'Num\_desc\_NPO' (9, 9.725330433578161),

'Org\_causes' (10, 11.995919799132848),

'Cam\_causes' (11, 11.837710985945904),

'Description\_Campaign\_polarity' (12, 2.589528325419086),

'Description\_Campaign\_subjectivity' (13, 10.518294251768054)]

Remove all the variables which have high VIF and then the model is:

Actual\_Donation\_Amount = ['Campaign\_Goal','Campaign\_Video','new\_msg\_category',

'Campaign\_Image\_num', 'Num\_desc\_cam', 'Description\_Campaign\_polarity']

Table

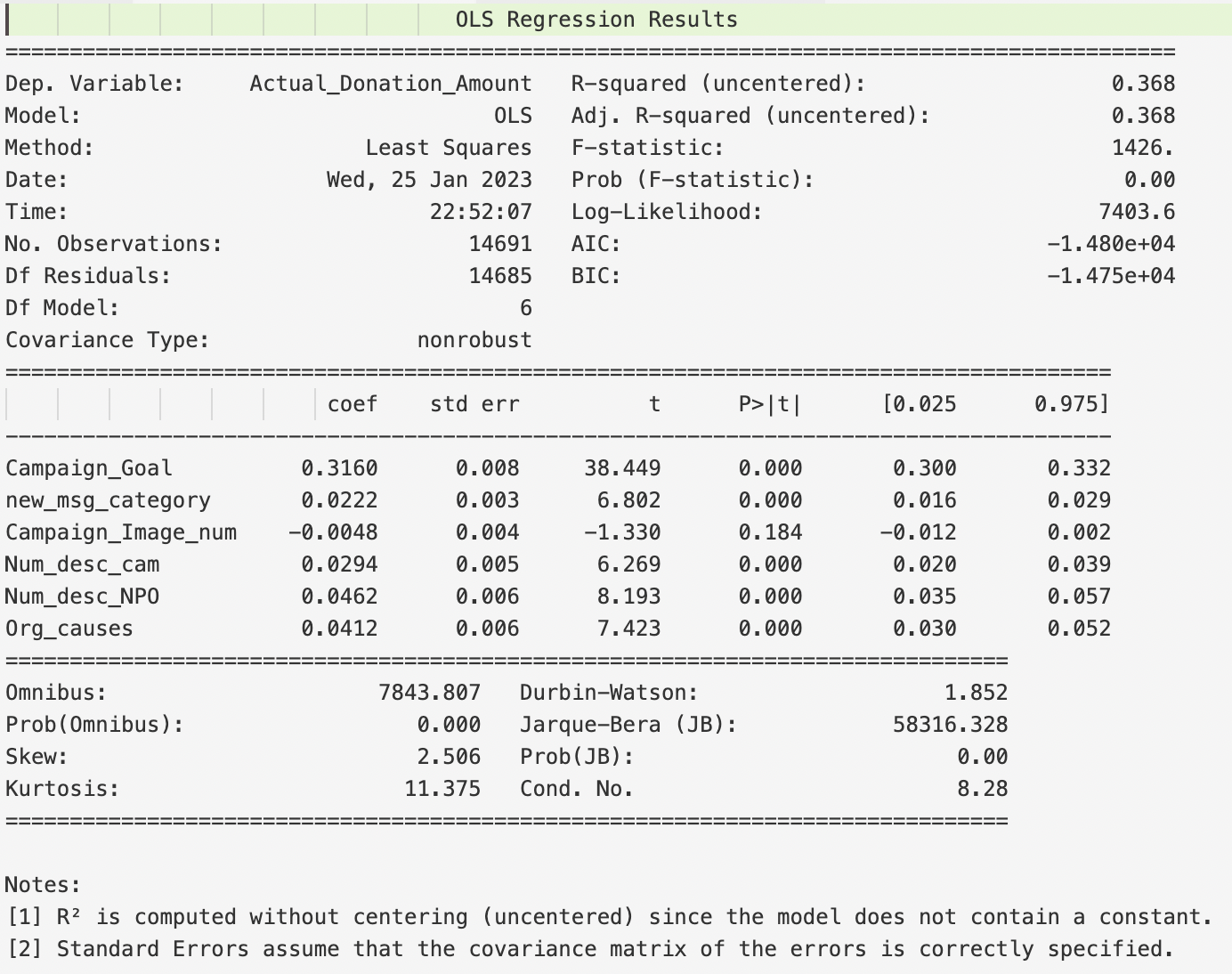
Description automatically generated

The result doesn’t change much.

Model 3

In order to improve the performance, I tried remove some abnormal data by using bucket plot. And the result is much better. I am not sure whether it useful or not. You can just take it as reference. If this way works we can try other ways to remove or clean some abnormal data.

The variables are the same with model 1:



Model 4

Besides, I tried to remove some variables which have high VIF, I found the R-square will reduce.

