BT2101 Problem-set 2: OLS

# Problem 1: Use Python to solve the problem [Number of Items and Total Spending]

# We are interested in the association between the number of items and total spending. Through the short survey to NUS students, we obtain the data below.

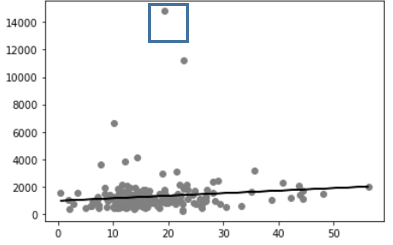
# Number of items = [4,6,6,8,8,8,9,9,10,12],

# Total spending = [9,10,18,20,15,17,20,22,25,30].

# Structure the simple regression model to measure the association between the number of items and total spending.

# Problem 2: Use Python to solve the problem [CEO Salary and Return on Equity]

We are interested in the impact of ROE (return on investment in percent) on CEO salary. We are using the data set CEOSAL1. Structure the simple regression model to measure the impact of ROE on CEO salary. Can we say that ROE is associated with an increase in CEO salary? To visualize this association, draw an OLS regression line by fitting the values. After drawing the line, we found a very clear outlier (see below, x= roe, y=ceo salary). After removing this outlier, does ROE still have a positive association with CEO salary? Do you think that an outlier has an important meaning in this regression?



ROE is associated as 18.501186 roe

After removing outlier, ROE will still have positive association

Outlier important? Not really, removing the positive association will still lead to a positive correlation so theres not much difference.