**Hypothesis 1**

This is an experiment to study whether there is a relationship between the electricity cost and the sales of EV in the United States in 2017

* Null hypothesis:

There is no difference of the electricity price (group A and B) on the effect of EV sales

* Alternative hypothesis:

There is difference of the electricity price (group A and B) on the effect of EV sales

**Analysis Plan:**

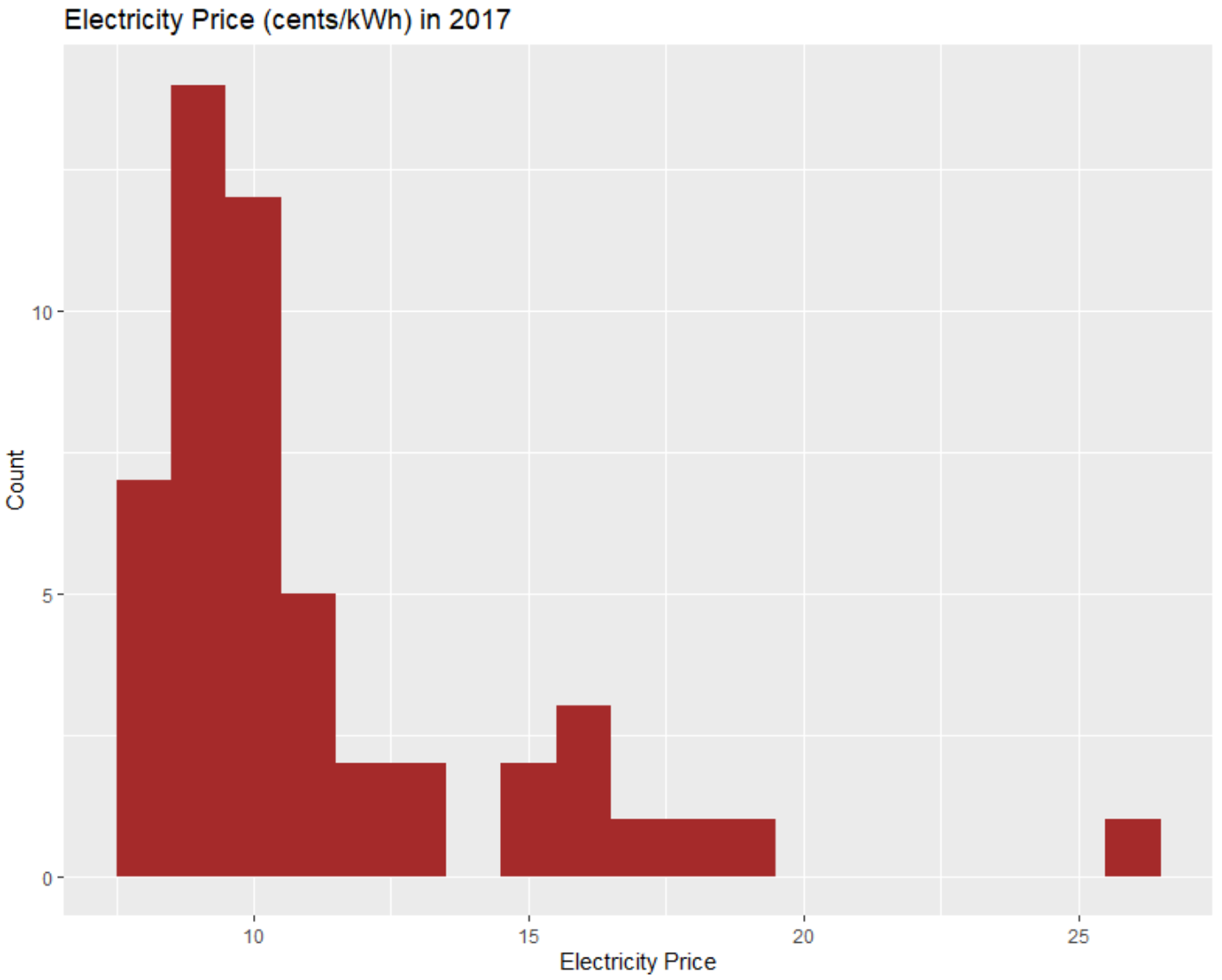
* Based on the historical data of state electricity profiles across 50 States (+1 District of Columbia) for the past 7 years, there is a clear distinction of two group electricity prices (cents/kWh) across US. Therefore, we transform the data to separate it into two groups: A & B for our experiment.
  + Group A: electricity price less than 10 cents per kWh
  + Group B: electricity price more than 10 cents per kWh
* As both sales and costs are quantitative data, we will fit a linear regression model

to predict whether there is an effect of cost on EV sales.

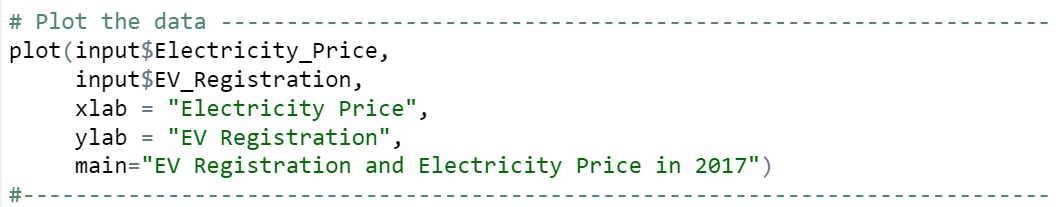
* We will also perform a hypothesis test for the regression coefficient of ‘cost’ to test whether it is zero, i.e. no significant difference in the sales due to electricity costs

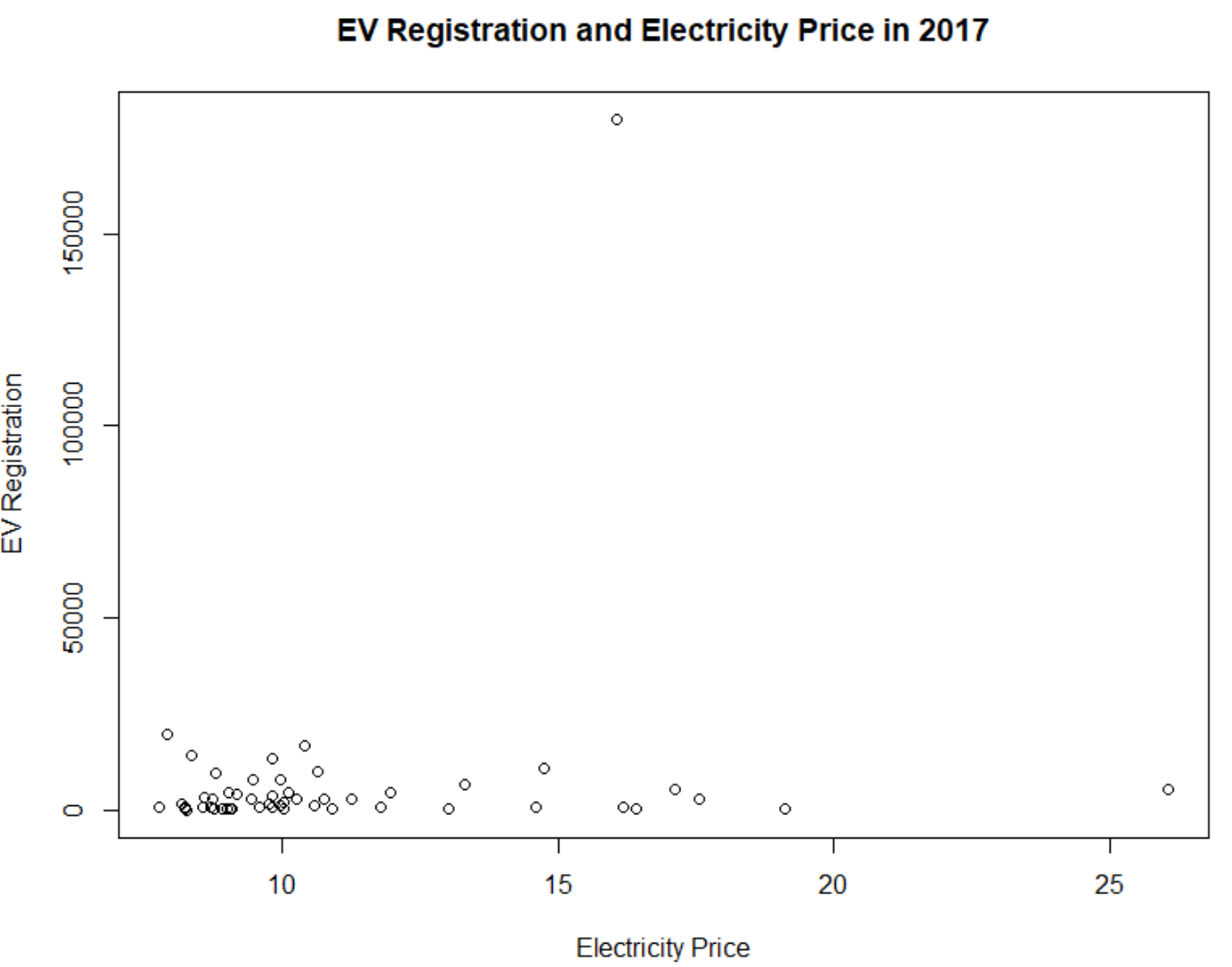
--------------------------------------------------------------------------------------------------------------------------------------



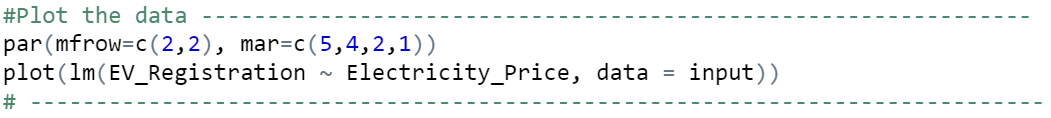


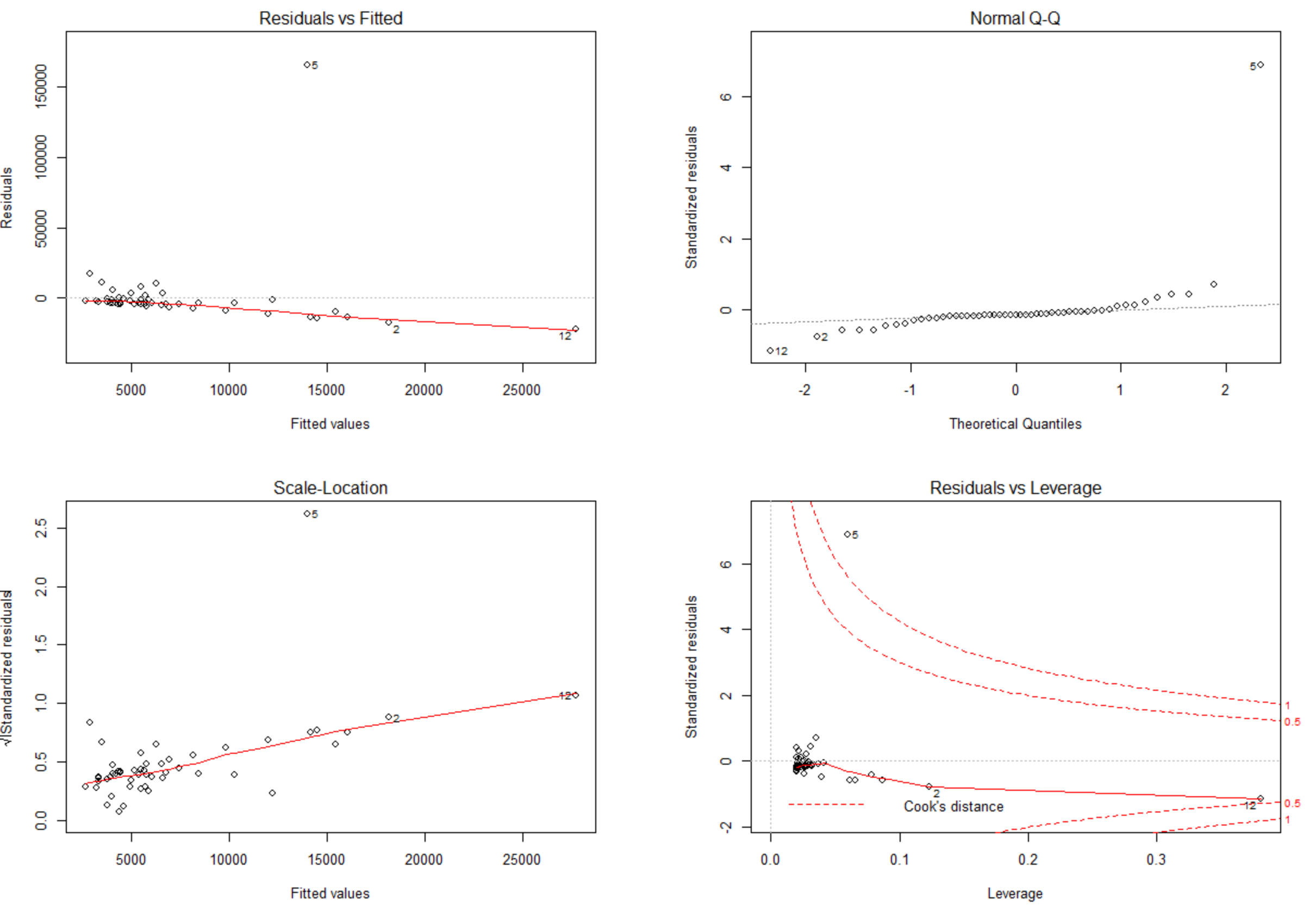
The histogram of electricity price is right skew, this demonstrates the majority of electricity price is less than 10 cents / kWh in the US in 2017





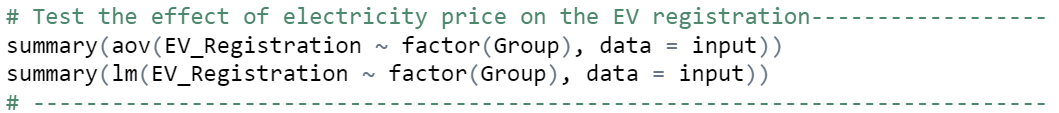
There are two outliers from the graph. One is California and one is Hawaii. The majority of the EV registrations lies on the states which have low cost of electricity price. The graph also shows that people tends to buy EV car when the electricity price is low in comparison to gasoline.

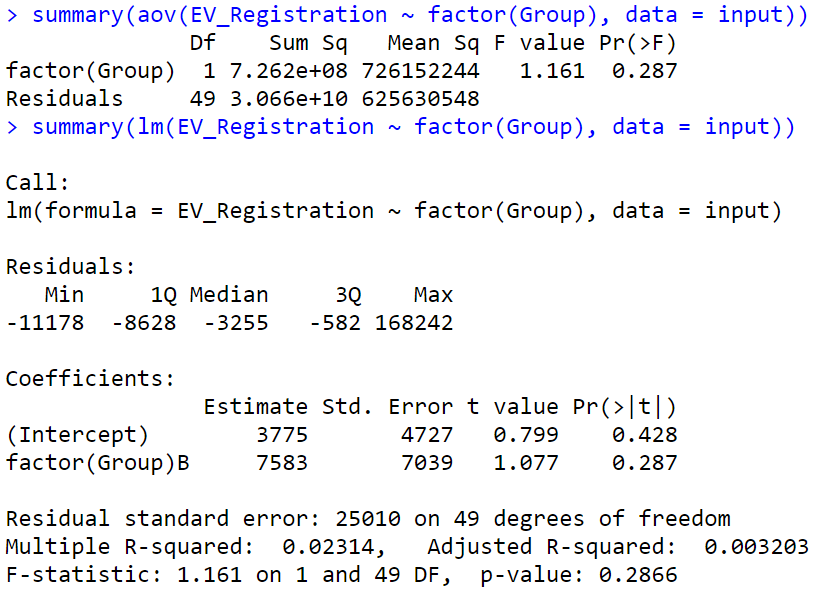




Perform the fit linear regression, we can see the majority of the points lie on the straight line of the Normal Q-Q plot. There are small deviations from the straight line but are minimal. Therefore, the degree of non-normality of the error is not severe.

On the “Residuals vs Fitted” and “Scale-Location” plots, there are not much evidence of non-constant variance. Majority of the data points closes to the red line and not far away from the red line (except for 1 data point).





Both the ANOVA and linear regression give the same value for P-value of 0.287. Because P-value > 0.05, therefore, we fail to reject the null hypothesis.