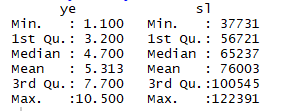
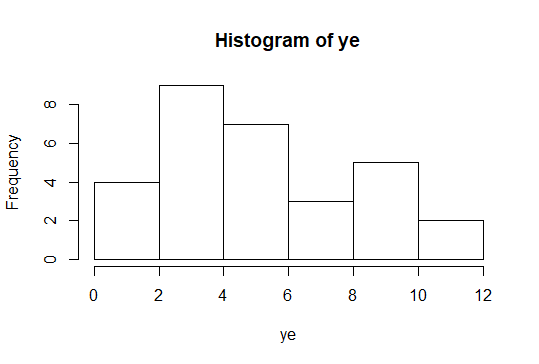
# Prediction model for Salary-hike:

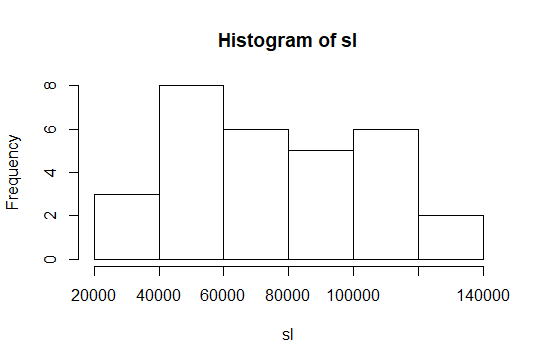
1. Data Collection: Salary\_data.csv
2. EDA:

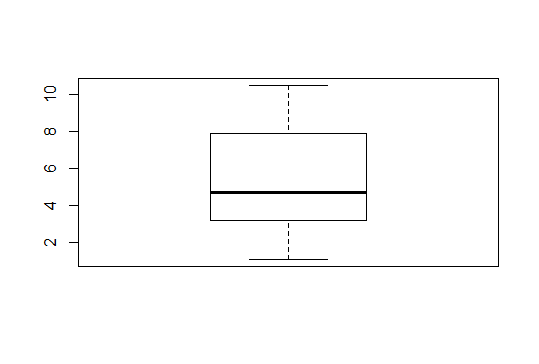


Standard deviation(Year of experience): 2.84

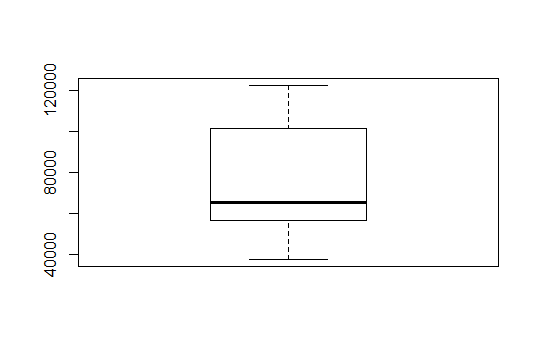
Standard deviation(Salary-hike): 27414



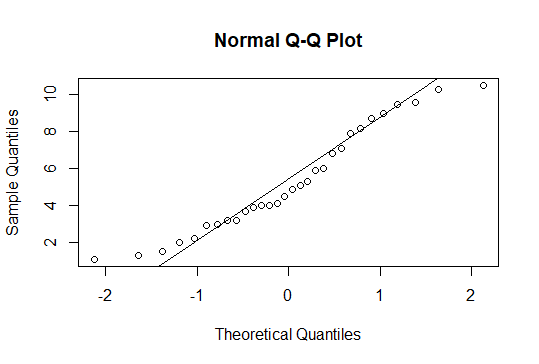




Box-plot for Year-experience(No outliers)



Box-plot for Salary-hike(No outliers)



Q-Q Plot for Year- experience(Normal)

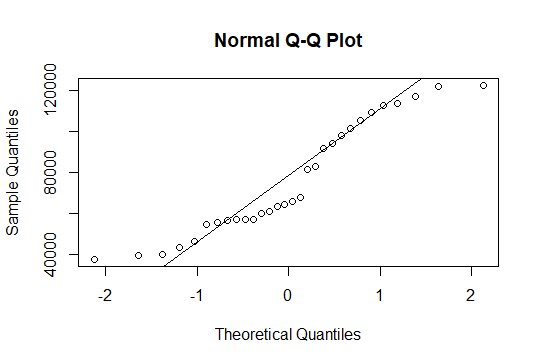
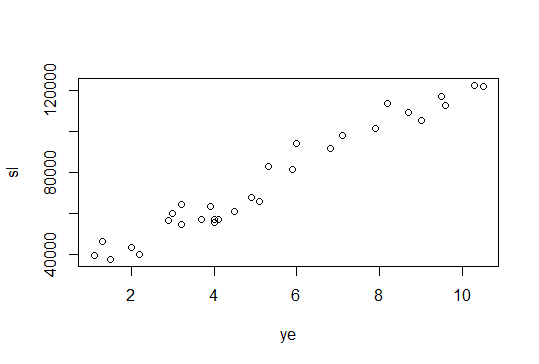


Figure 1Q-Q Plot for Salary hike(not Normal)

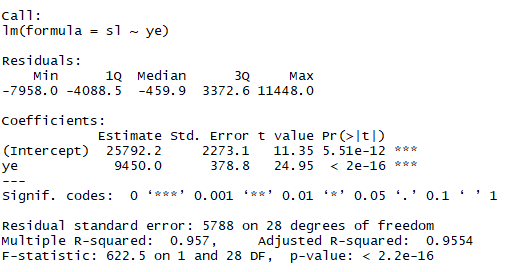
Scatter Plot:



The Scatter Plot suggests:

* Strong Positive collinearity(0.98)
* Linear
* Presence of clusters

1. Data mining:



B0 = 25792.2 is significant

B1 = 9450 is significant

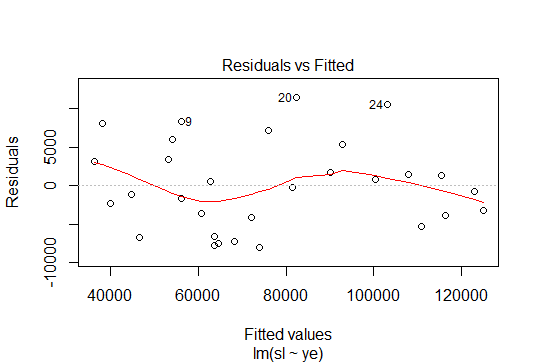
R2 value = 0.96

F-statistic and p- value suggests our model is good.

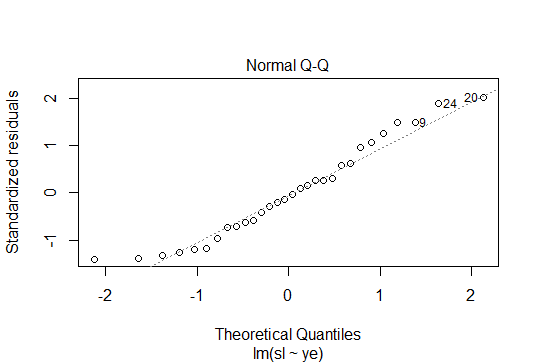
From our R code, our model can be represented as:

**Salary-hike = 25792.2 + 9450\* Year-experience**

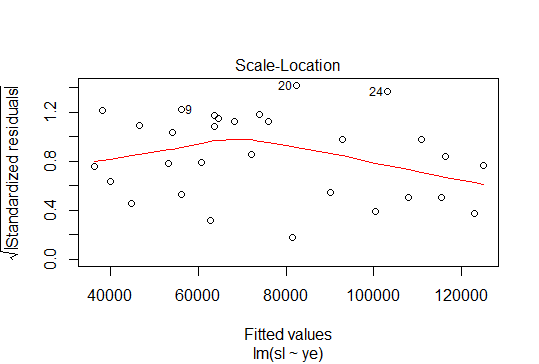
1. Line assumptions:



The above Plot suggests that residuals are independent.



The above plot suggests that errors are normally distributed



The above plot suggests that there is no heteroscedasticity problem

|  |  |
| --- | --- |
| **Model** | **R2 value** |
| **M1** | **0.96** |