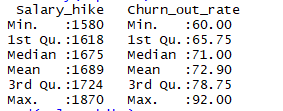
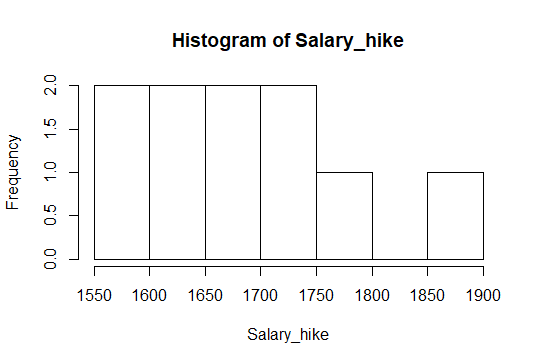
# To Build a prediction model For Churn- Out rate

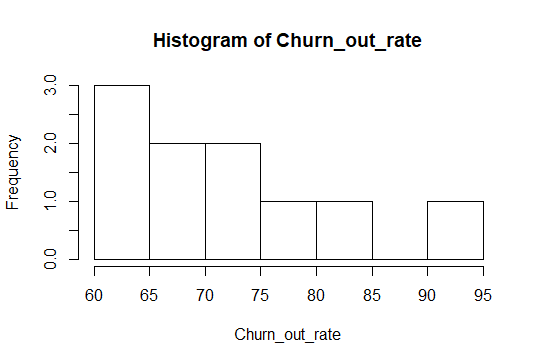
1. Data Collection: emp\_data.csv
2. EDA:

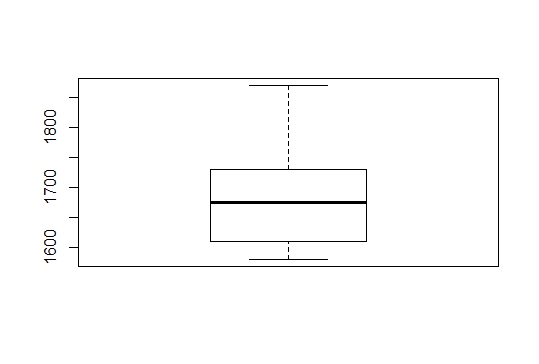


Standard deviation(Salary-hike): 92.09

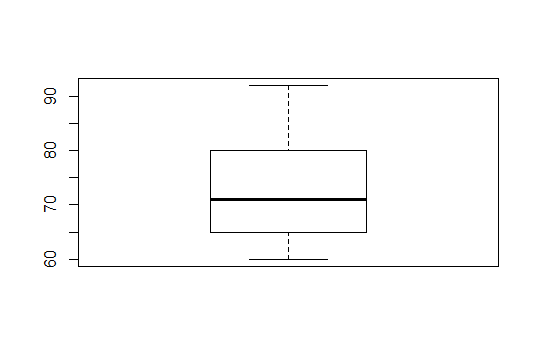
Standard deviation(Churn- out Rate): 10.26



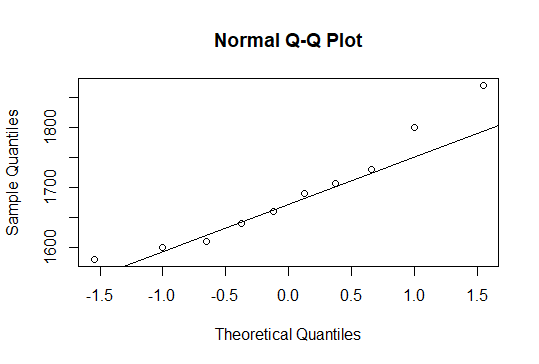




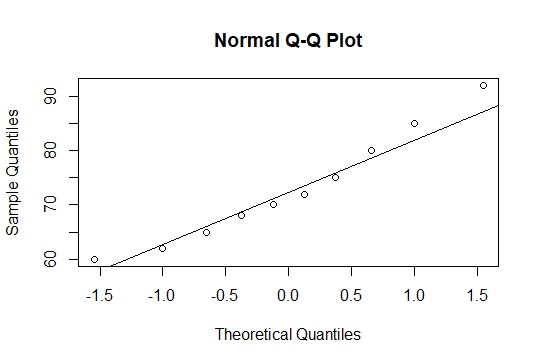
Box-Plot for Salary Hike(No Outliers)



Box- plot for Churn-out Rate(No Outliers)



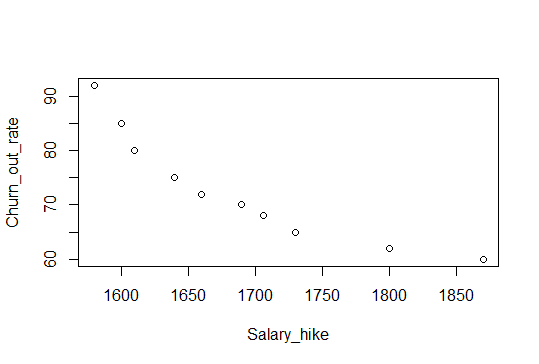
Q-Q plot for Salary-hike(Normal)



Q-Q plot for Churn-out Rate

We don’t need to transform.

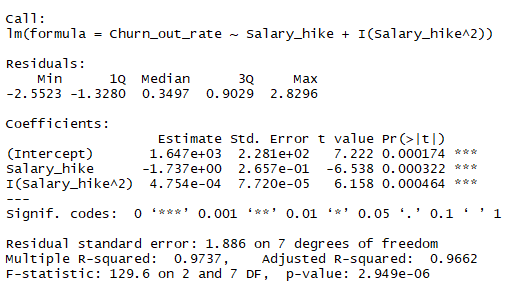
Scatter Plot:



The Scatter plot suggests :

* Negative correlation
* Non linear
* Absence of clusters

1. Data Mining:



B0 = 1.65\*exp(3) is significant

B1 = -1.737 is significant

B2 = 4.75\*exp(-4) is significant

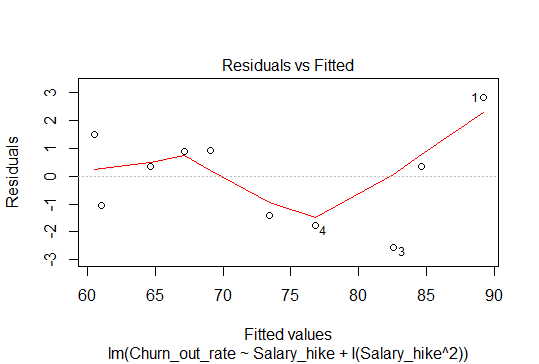
R2 value = 0.97

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

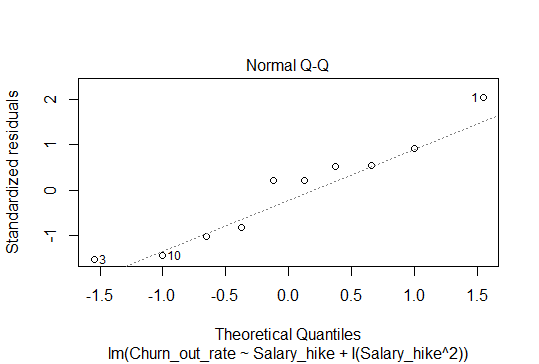
From Our R code, model is:

**Churn-out Rate = 1.64\*exp(3) – 1.74\*Salary-hike + 4.75\*exp(-4)\*Salary-hike^2**

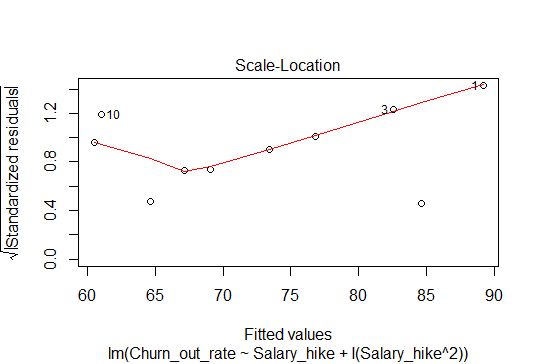
1. LINE assumptions:



The above Plot suggests that residuals are independent.



The above Plot suggests that error are normally distributed.



The above plot suggests that there is no heteroscedasticity problem.

|  |  |
| --- | --- |
| Model | R2 value |
| M1 | 0.83 |
| M2 | 0.97 |