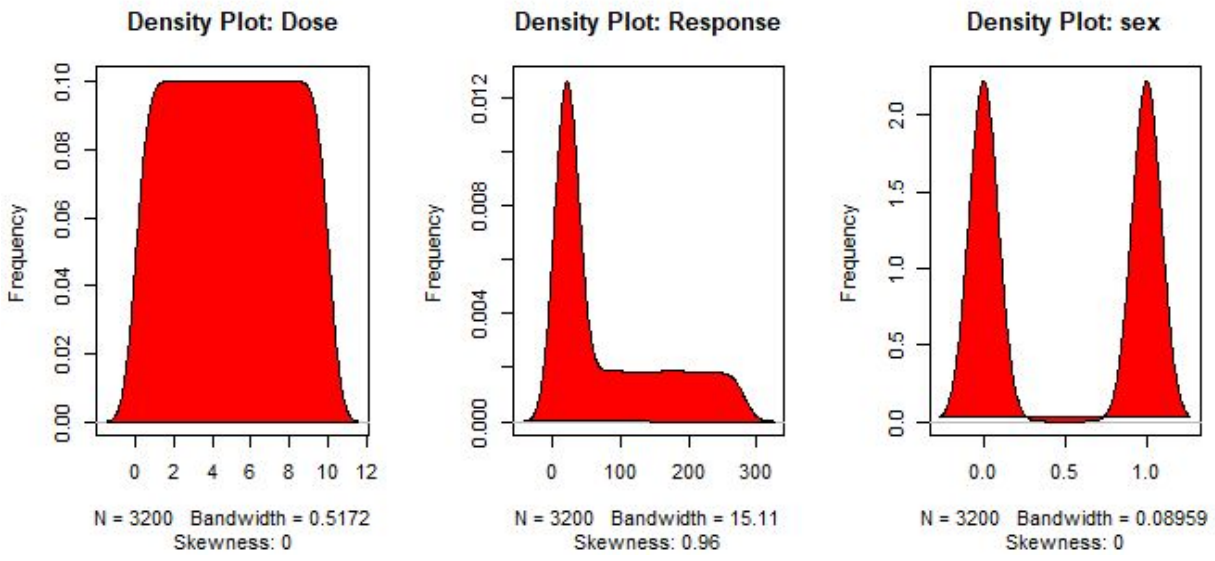
Saurabh Kumar Singh
16100051

```
>data=read.csv("C:/Users/imsau/Desktop/6th Sem/ML/ML_lab/Lab3(6-Feb)/drug2.csv")

> summary(data)
      sex          dose          response
 Min.   :0.0   Min.   : 0.100   Min.   :  1.92
 1st Qu.:0.0   1st Qu.: 2.575   1st Qu.: 19.53
 Median :0.5   Median : 5.050   Median : 32.37
 Mean   :0.5   Mean   : 5.050   Mean   : 83.01
 3rd Qu.:1.0   3rd Qu.: 7.525   3rd Qu.:147.09
 Max.   :1.0   Max.   :10.000   Max.   :280.73

> head(data)
  sex dose response
1   1  0.1    13.75
2   1  0.2    12.90
3   1  0.3    19.26
4   1  0.4    20.34
5   1  0.5    19.97
6   1  0.6    26.80

> cor(data$dose, data$response)
[1] 0.5136214
> cor(data$sex, data$response)
[1] 0.7516308
```



```
> par(mfrow=c(1, 3))  # divide graph area in 2 columns
> plot(density(data$dose), main="Density Plot: Dose", ylab="Frequency",
sub=paste("Skewness:", round(e1071::skewness(data$dose), 2)))  # density plot for 'speed'
 polygon(density(data$dose), col="red")
> plot(density(data$response), main="Density Plot: Response", ylab="Frequency",
sub=paste("Skewness:", round(e1071::skewness(data$response), 2)))  # density plot for 'dist'
 polygon(density(data$response), col="red")
> plot(density(data$sex), main="Density Plot: sex", ylab="Frequency", sub=paste("Skewness:",
round(e1071::skewness(data$sex), 2)))  # density plot for 'dist'
 polygon(density(data$sex), col="red")
```

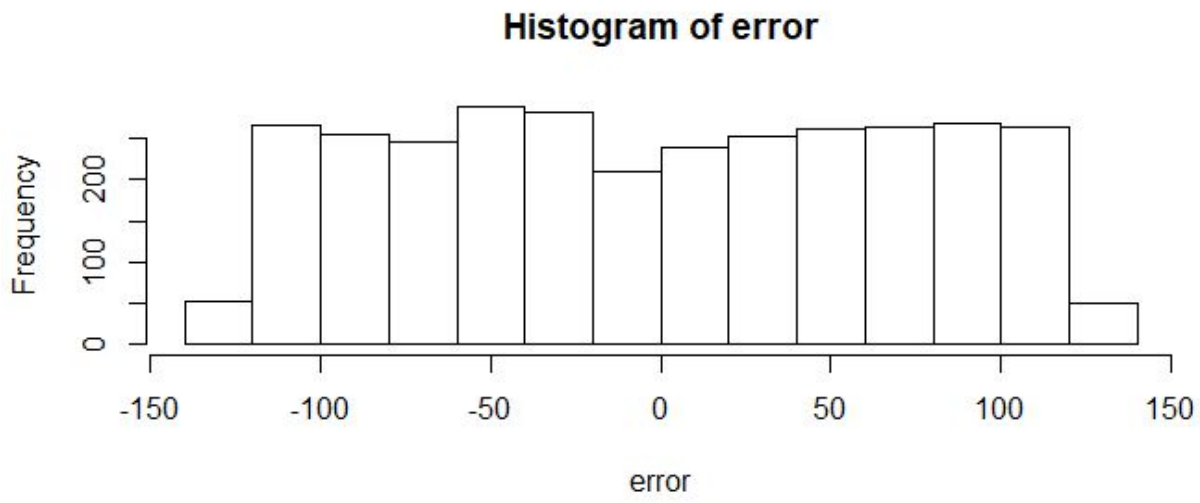



```
>scatter.smooth(x=data$dose, y=data$response, main="dose ~ response")
```

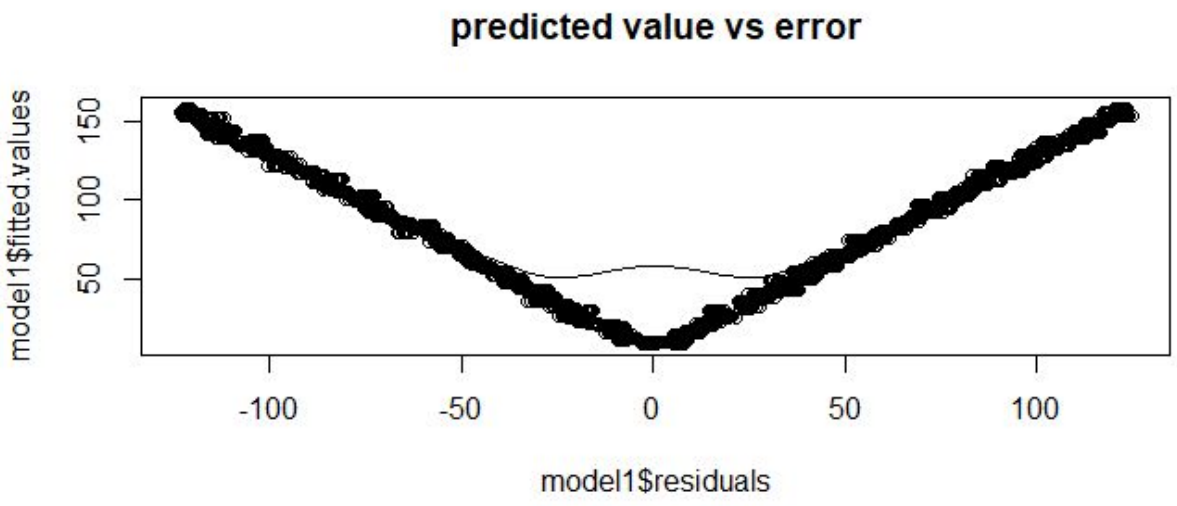

```
model1 = lm(data$response~data$dose)
> summary(model1)

Call:
lm(formula = data$response ~ data$dose)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-123.514  -62.764    0.401   63.669  124.707

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.2534     2.5778   2.814  0.00493 **
data$dose    15.0020     0.4432  33.852  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 72.36 on 3198 degrees of freedom
Multiple R-squared:  0.2638,  Adjusted R-squared:  0.2636
F-statistic:  1146 on 1 and 3198 DF,  p-value: < 2.2e-16
```

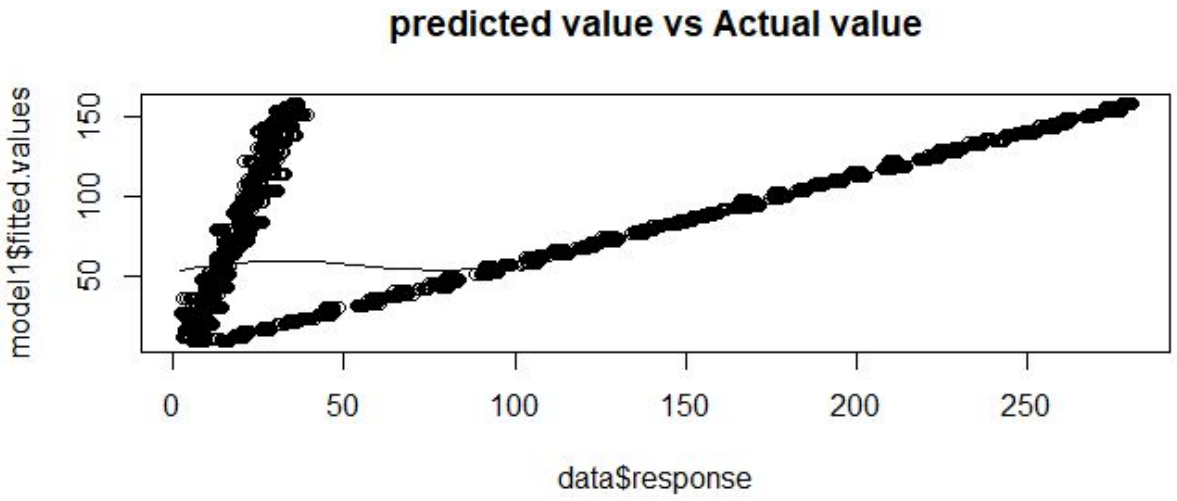
```
>error = residuals(model1)
> hist(error)
```




```
>scatter.smooth(x=model1$residuals, y=model1$fitted.values, main="predicted value vs error")
```



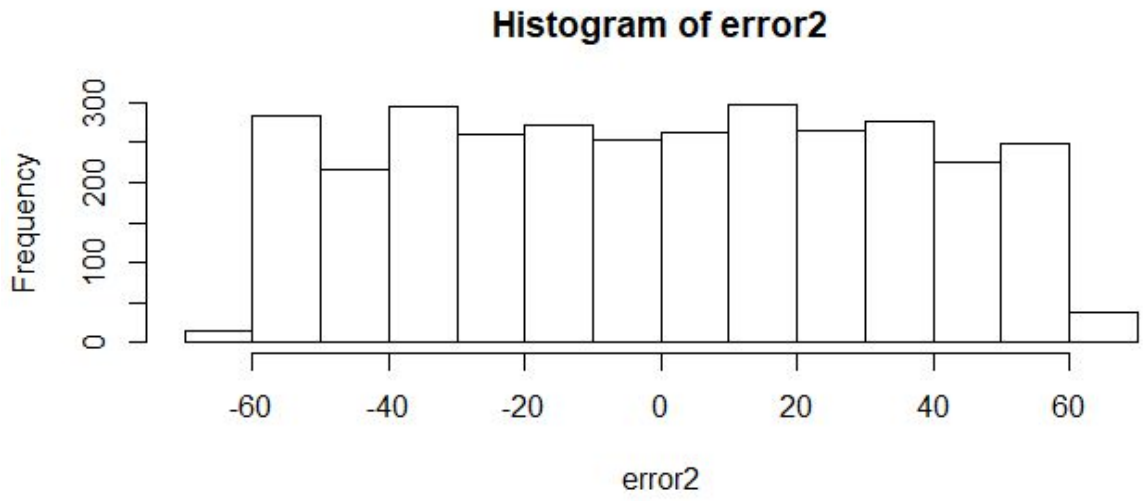

```
> scatter.smooth(x=data$response, y=model1$fitted.values, main="predicted value vs Actual value")
```



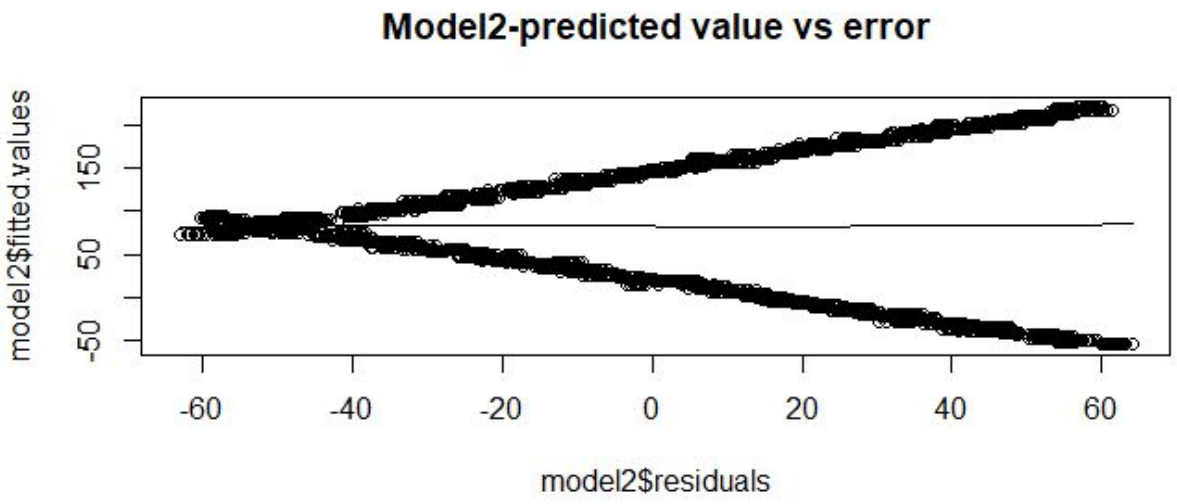

```
> model2 = lm(data$response~data$dose+data$sex)
> summary(model2)

Call:
lm(formula = data$response ~ data$dose + data$sex)

Residuals:
    Min      1Q  Median      3Q     Max
-62.986 -30.350   0.306  29.360  64.009
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -56.1189     1.3881  -40.43   <2e-16 ***
data$dose    15.0020     0.2138   70.18   <2e-16 ***
data$sex    126.7445     1.2341  102.70   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 34.91 on 3197 degrees of freedom
Multiple R-squared:  0.8288,  Adjusted R-squared:  0.8286
F-statistic:  7736 on 2 and 3197 DF,  p-value: < 2.2e-16
```

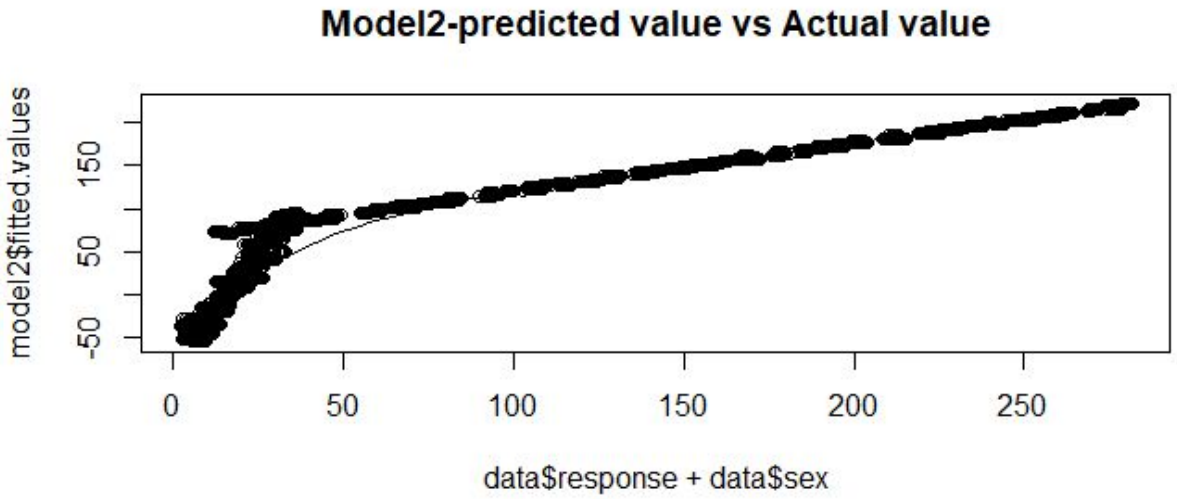
```
> error2 = residuals(model2)
> hist(error2)
```




```
> scatter.smooth(x=model2$residuals, y=model2$fitted.values, main="Model2-predicted value vs error")
```



> scatter.smooth(x=data$response+data$sex, y=model2$fitted.values, main="Model2-predicted value vs Actual value")



How to know if the model is best fit for your data?

The most common metrics to look at while selecting the model are:

STATISTIC	CRITERION
R-Squared	Higher the better *(> 0.70)*
Adj R-Squared	Higher the better
F-Statistic	Higher the better
Std. Error	Closer to zero the better
t-statistic	Should be greater 1.96 for p-value to be less than 0.05
AIC	Lower the better
BIC	Lower the better
Mallows cp	Should be close to the number of predictors in model
MAPE (Mean absolute percentage error)	Lower the better

MSE (Mean squared error)	Lower the better
Min_Max Accuracy => mean(min(actual, predicted)/max(actual, predicted))	Higher the better