DS – 510 Project 1

INVESTIGATING THE IMPACT OF A NUMBER OF AUTOMOBILE ENGINE FACTORS ON VEHICLE'S MPG

Team Members:

Aditi Agarwal

Lovepreet Kaur

Mohit Supe

Table of Contents

Our Analysis:

Code Block	Explanatory Variables	Response	Mutiple R^2 Value	Adjusted R^2 Value
Block 1	weight	mpg	0.7714	0.7706
Block 2	acceleration	mpg	0.2052	0.2025
Block 3	displacement	mpg	0.7104	0.7094
Block 4	horsepower	mpg	0.641	0.6397
Block 5	weight,acceleration	mpg	0.7748	0.7733
Block 6	weight,displacement	mpg	0.7764	0.7749
Block 7	**weight,horsepower**	**mpg**	**0.7796**	0.7781
Block 8	displacement,acceleration	mpg	0.713	0.7111
Block 9	horsepower,acceleration	mpg	0.6721	0.6699
Block 10	displacement,horsepower	mpg	0.7213	0.7194
Block 11	**displacement,horsepower,weight**	**mpg**	**0.7804**	**0.7782**
Block 12	displacement,weight,acceleration	mpg	0.7771	0.7749
Block 13	horsepower,weight,acceleration	mpg	0.7797	0.7774
Block 14	displacement,acceleration,horsepower	mpg	0.7373	0.7347
Block 15	**horsepower,weight,acceleration,displacement**	**mpg**	**0.7806**	**0.7777**

R^2 (Coefficient of Multiple Determination)

R^2 measures how close the data is to the fitted Regression Line. It represents the percentage of the response variable (MPG i.e. Miles Per Gallon) variation about its mean explained by linear model. Its value lies between 0 and 100.

R^2=Explained Variation/Total Variation

Linear Model with highest R^2 value better fits the data.

The Adjusted R^2 Value is the R-squared value that has been adjusted for the number of predictors/Explanatory variables in the model.
The adjusted R-squared value increases only if the new explanatory variable improves the model more than what would be expected by chance. It decreases when a predictor/Explanatory variable improves the model by less than what is expected by chance.

Block 1 is the best model.

- With 4 Explanatory variables, block 15 has the highest Multiple R^2 value (78.06), but its adjusted R^2 value is lesser than Block 11. It means Block 11 is a better model than Block 15. Adding one extra variable in Block 15 is not increasing R^2 value significantly so Block 11 is better model than block 15.
- If we compare Block 11 (3 Explanatory variables) with Block 7 (2 Explanatory variables), there is not much difference in their R^2 values. Adding an extra explanatory variable is not improving the model .So Block 7 is a better model than Block 11.
- Comparing Block 7 (2 Explanatory variables) with Block 1 (1 Explanatory variable), adding one extra explanatory variable (Horsepower) is just improving model by .0075 (0.7781-0.7706) R^2 value, which is not a big difference. So Block 1 is the best model.

Code Block	Explanatory Variables	Response	P values	Significant(Reject H0)	Not Significant(H0 is true)
Block 1	weight	mpg	< 2.2e-16	weight	
Block 2	acceleration	mpg	< 2.2e-16	acceleration	
Block 3	displacement	mpg	< 2.2e-16	displacement	
Block 4	horsepower	mpg	< 2.2e-16	horsepower	
Block 5	weight,acceleration	mpg	< 2.2e-16,0.0343	weight	acceleration
Block 6	weight,displacement	mpg	< 2.2e-16,0.0104	weight	displacement
Block 7	**weight,horsepower**	**mpg**	< 2.2e-16,0.00101	weight	horsepower
Block 8	displacement,acceleration	mpg	< 2.2e-16,0.104	displacement	acceleration
Block 9	horsepower,acceleration	mpg	< 2.2e-16,2.12e-07	horsepower,acceleartion	
Block 10	displacement,horsepower	mpg	< 2.2e-16,0.000758	displacement,horsepower	
Block 11	**displacement,horsepower,weight**	**mpg**	0.2945,0.0208,< 2.2e-16	weight	displacement,horsepower
Block 12	displacement,weight,acceleration	mpg	0.0811,< 2.2e-16,0.3237	weight	displacement,acceleartion
Block 13	horsepower,weight,acceleration	mpg	0.0112,< 2.2e-16,0.7376	weight	horsepower,acceleartion
Block 14	displacement,acceleration,horsepower	mpg	5.75e-16,2.86e-05,3.12e-07	displacement,horsepower& acceleration	
Block 15	**horsepower,weight,acceleration,displacement**	**mpg**	0.0304,3.22e-13,0.5655,0.2517	weight	displacement,horsepower& acceleration

P value decides the significance of variable.

H0: Explanatory Variable has no impact on MPG
Ha: Explanatory Variable is impacting MPG

=> If P value < Significance level (Alpha), then Test is significant and hence we can Reject Null Hypothesis (H0) that that variable has no impact on MPG. Variable has impact on MPG (Ha is true).
=>If P value > Significance level (Alpha), then Test is not significant, means we cannot reject H0.So, the variable has no impact on MPG.

Block 1 is the best model.
Four best models with high R^2 values are Block 15, Block 11, Block 7 and Block 1 with 4, 3, 2 and 1 explanatory variables respectively.

- In Block 15, only weight is impacting MPG majorly. Displacement, acceleration and horsepower have P values greater than Significance level .So Block 15 is not the best model.
- In block 11, only weight is impacting MPG majorly .Displacement and Horsepower have P values greater than Significance level. So Block 11 is also not the best model.
- In Block 7, only weight is impacting MPG majorly. Horsepower has P value greater than Significance level. So Block 7 is also not the best model.

So by adding Explanatory variables with Weight, R^2 is not increasing much. So Block 1 is the best model with one Explanatory variable.

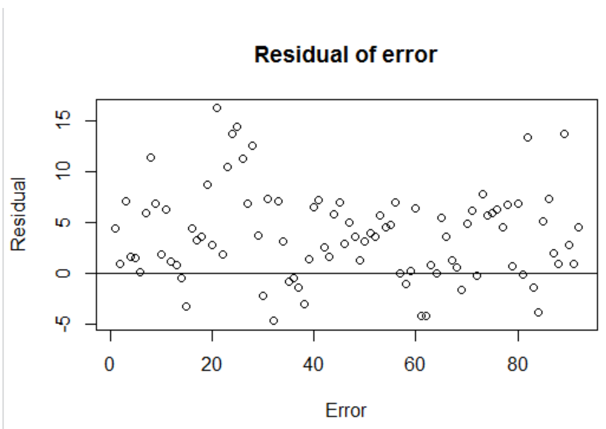
Conclusion:
Block 1 (Weight as Explanatory variable) is the First Best model with one explanatory variable.
Block 7 (Weight and horsepower as Explanatory variables) is the second Best model with two explanatory variables.

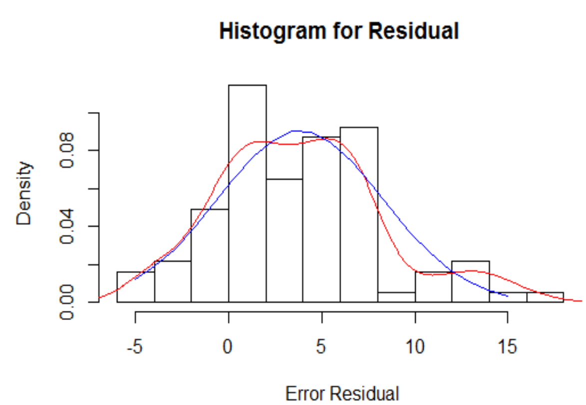
MPG decreases with increase in displacement, weight & horsepower. Horsepower is the unit to measure power of an Engine. So higher the value of horsepower, lesser is the amount of fuel engine will consume per mile (MPG).

MPG increases with acceleration. The variable acceleration is the time taken to accelerate from zero to 60 MPH (Miles Per Hour), so the higher the acceleration value, the more time it takes to accelerate and thus higher the fuel consumption (Miles Per Gallon) .

Omitting 6 rows with NULL values in Horsepower
If we include rows with NULL values, then regression against Horsepower is constrained to include only rows where all values are present. This produces incorrect results for columns with no missing data. Hence removing all the rows where the value of Horsepower is Null is the correct way of making the analysis correctly.

Predicting value of MPG for the remaining 92 rows based on selected Linear Model
Predicting the value of MPG for the remaining 92 rows based on the selected Linear Model helps us to identify the level of accuracy of the selected Linear Model. From the results obtained after calculating the predicted value of MPG using the best fit Linear Model, we can conclude that the selected Linear Model (Block 1) is fairly accurate. The below plot explains the level of accuracy of the selected Linear Model.






```
> summary(error1)
    Min.   1st Qu.   Median     Mean     3rd Qu.    Max.
-4.6710    0.8429    3.6370    3.8460    6.4650   16.2800
```

- From the plot and its corresponding summary, we can see that there exist some extreme values as the difference between the 3rd Quartile value and Maximum value is quite high. This is because there were a couple of outliers in the referenced data of 300 rows for the value of MPG.
- Since most of the residual values are positive, we can conclude that predicted values are smaller than the respective observed values.
- Observing the Histogram, we can conclude that the residual are almost normal with ~ N(3.8460, 19.44957)

Reading Data File and Arranging for Analysis

Reading the data file auto-mpg.data

```
master_file <- read.table('E:\\SPU Sem 1\\DS510\\Project 1\\auto-mpg.data')
View(master_file)
```

Renaming the columns

```
names(master_file)[1] <- paste("mpg")
names(master_file)[2] <- paste("cylinders")
names(master_file)[3] <- paste("displacement")
names(master_file)[4] <- paste("horsepower")
names(master_file)[5] <- paste("weight")
names(master_file)[6] <- paste("acceleration")
names(master_file)[7] <- paste("model year")
names(master_file)[8] <- paste("origin")
names(master_file)[9] <- paste("car name")
View(master_file)
```

Deleting empty rows

```
master_file_1 <- master_file[-c(33, 127, 331, 337, 355, 375),]
```

To Change sequence of all rows after deleting Null values

```
rownames(master_file_1) <- seq(length=nrow(master_file_1))

View(master_file_1)
```

To convert Horsepower variable into numeric data type

```
master_file_1$horsepower=as.numeric(as.character(master_file_1$horsepower))
str(master_file_1)

'data.frame':    392 obs. of  9 variables:
$ mpg         : num  18 15 18 16 17 15 14 14 14 15 ...
$ cylinders   : int  8 8 8 8 8 8 8 8 8 8 ...
$ displacement: num  307 350 318 304 302 429 454 440 455 390 ...
$ horsepower  : num  130 165 150 150 140 198 220 215 225 190 ...
$ weight      : num  3504 3693 3436 3433 3449 ...
```

$ acceleration: num 12 11.5 11 12 10.5 10 9 8.5 10 8.5 ...
$ model year : int 70 70 70 70 70 70 70 70 70 70 ...
$ origin   : int 1 1 1 1 1 1 1 1 1 1 ...
$ car name   : Factor w/ 305 levels "amc ambassador brougham",..: 50 37 232 15 162 142 55 224 242 2 …

Creating data file of 1 to 300 rows

```
data_300 <- master_file_1[1:300,]
View(data_300)
```

Creating data file of 301 to 398 rows

```
data_92 <- master_file_1[301:392,]
rownames(data_92) <- seq(length=nrow(data_92))
View(data_92)
```

Model 1: MPG vs Weight

model_1 <- lm(mpg~weight, data = data_300)
summary(model_1)

```
> summary(model_1)

Call:
lm(formula = mpg ~ weight, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.2011 -1.9157 -0.0812  1.7341 15.0246 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 40.5619792  0.6461532   62.77   <2e-16 ***
weight      -0.0062905  0.0001984  -31.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.032 on 298 degrees of freedom
Multiple R-squared:  0.7714,	Adjusted R-squared:  0.7706 
F-statistic:  1005 on 1 and 298 DF,  p-value: < 2.2e-16
```

- There is a strong negative association between MPG and Weight. MPG decreases with increase in weight and vice-versa.
- Value of R^2 is 0.7714 that means our model only explains 77.14% of variance.
- Equation of Regression Line is:

MPG = 40.5619792 – 0.0062905*weight + Ei

Where Ei ~ N(0,3.032)

B0 -> Intercept -> 40.5619792

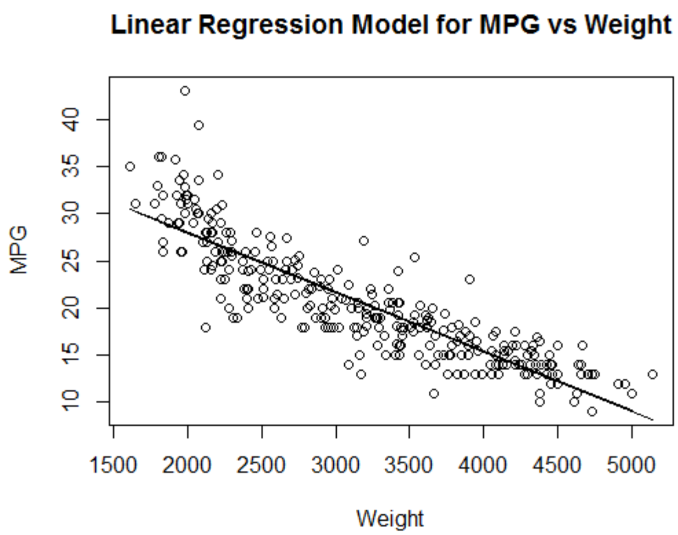
P-value: < 2.2e-16<0.001 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG.

Residual = Observed(Y) – Predicted(Y)

- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

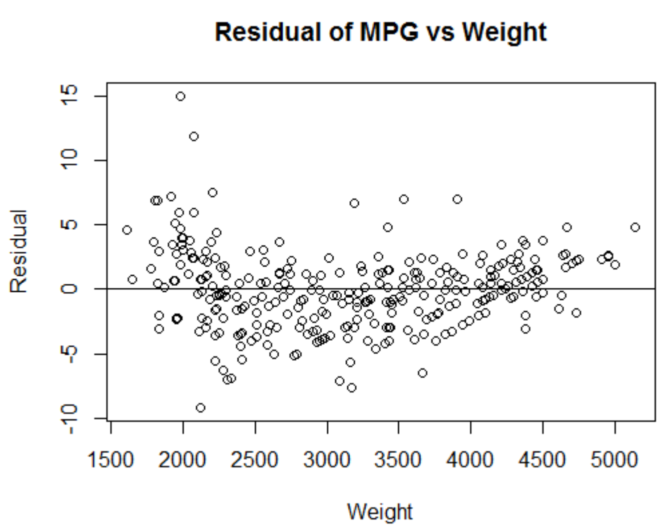
coef(model_1)[1]
(Intercept)
40.56198
coef(model_1)[2]
weight
-0.006290453

```
plot(data_300$weight, data_300$mpg,
    xlab = "Weight",
    ylab = "MPG",
    main = "Linear Regression Model for MPG vs Weight")
lines(data_300$weight, coef(model_1)[1]+coef(model_1)[2]*data_300$weight)
```



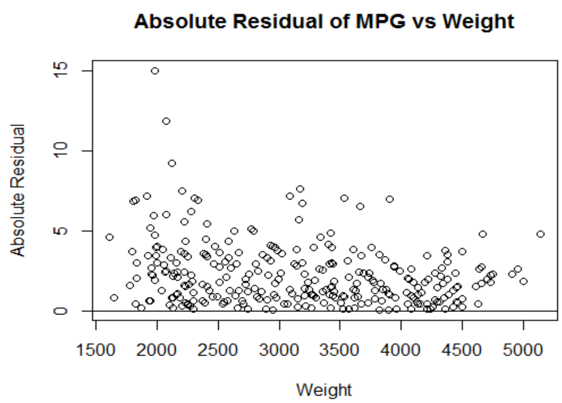
 # Residual Plot

```
res_1 <- residuals(model_1)
plot(data_300$weight, res_1,
    xlab = "Weight",
    ylab = "Residual",
    main = "Residual of MPG vs Weight")
abline(0,0)
```



Absolute Residual Plot

```
a_res_1 <- abs(res_1)
plot(data_300$weight, a_res_1,
    xlab = "Weight",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Weight")
abline(0,0)
```

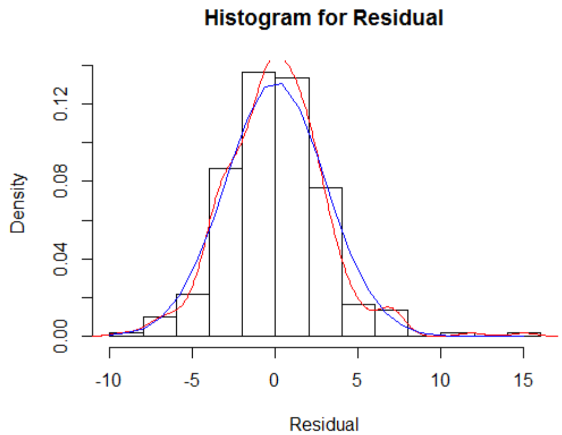


Histogram for residual

```
hist(res_1, prob=T, breaks = 10,
    main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_1), col="red")
```

Normalizing the curve

```
mu_1 <- mean(res_1)
v_1 <- var(res_1)
sd_1 <- sqrt(v_1)
x_1 <- seq(-10, 15, length=25)
y_1 <- dnorm(x_1,mu_1,sd_1)
lines(x_1,y_1,col="blue")
```



Model 2: MPG vs Acceleration

model_2 <- lm(mpg~acceleration, data = data_300)
summary(model_2)

```
> summary(model_2)

Call:
lm(formula = mpg ~ acceleration, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.202  -4.126  -1.012   3.268  16.154 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)    5.0012     1.8352   2.725  0.00681 ** 
acceleration   1.0379     0.1183   8.770  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.654 on 298 degrees of freedom
Multiple R-squared:  0.2052,	Adjusted R-squared:  0.2025 
F-statistic: 76.91 on 1 and 298 DF,  p-value: < 2.2e-16
```

- Value of R^2 is 0.2052 that means our model only explains 20.52% of variance.
- Equation of Regression Line is:

MPG = 5.0012 + 1.0379*acceleration + Ei

Where Ei ~ N(0,5.654)

B0 -> Intercept -> 5.0012

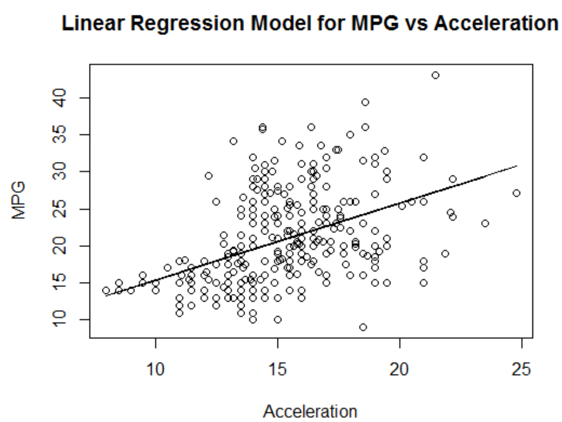
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that acceleration has no effect on MPG.

Residual = Observed(Y) – Predicted(Y)

- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

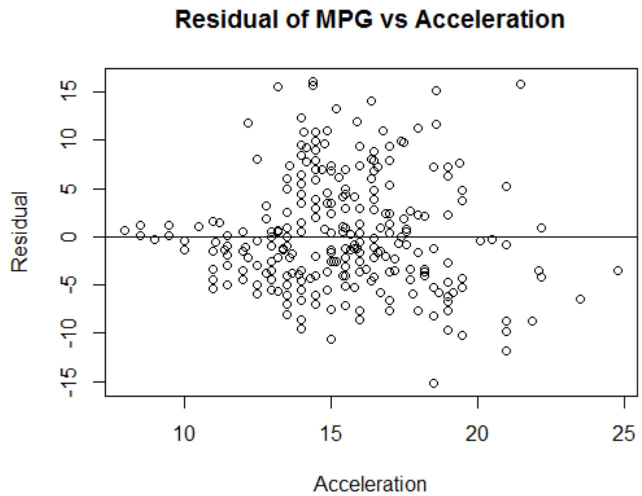
```
coef(model_2)[1]
(Intercept)
  5.001162
coef(model_2)[2]
acceleration
   1.037865
```

```
plot(data_300$acceleration, data_300$mpg,
    xlab = "Acceleration",
    ylab = "MPG",
    main = "Linear Regression Model for MPG vs Acceleration")
lines(data_300$acceleration, coef(model_2)[1]+coef(model_2)[2]*data_300$acceleration)
```



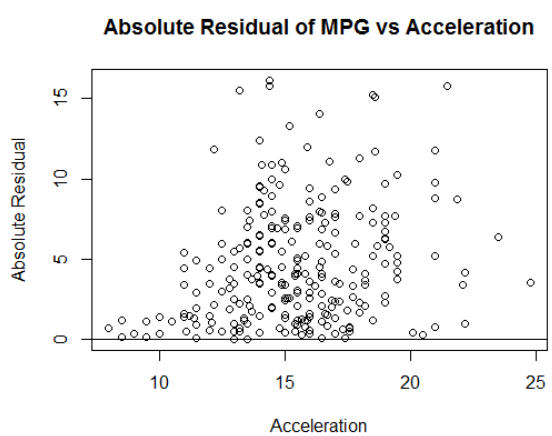
 # Residual Plot

```
res_2 <- residuals(model_2)
plot(data_300$acceleration, res_2,
    xlab = "Acceleration",
    ylab = "Residual",
    main = "Residual of MPG vs Acceleration")
abline(0,0)
```



Absolute Residual Plot

```
a_res_2 <- abs(res_2)
plot(data_300$acceleration, a_res_2,
    xlab = "Acceleration",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Acceleration")
abline(0,0)
```

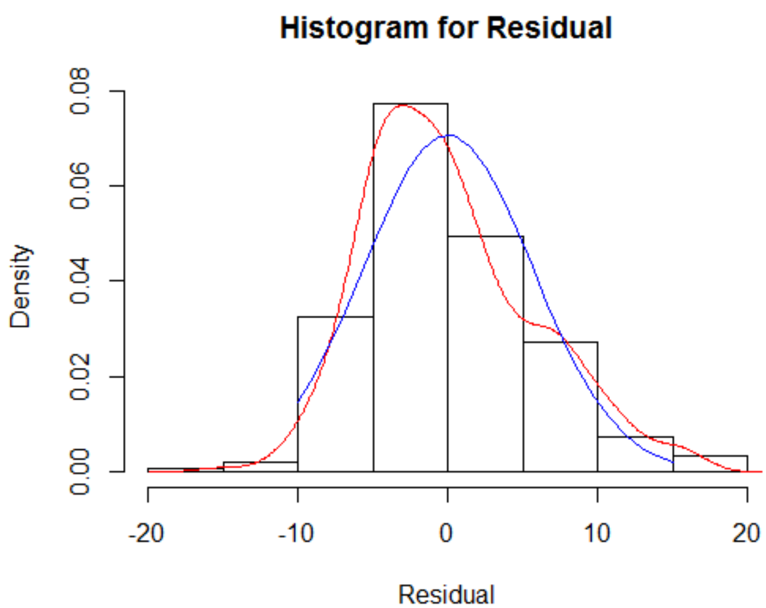


Histogram for residual

```
hist(res_2, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_2), col="red")
```

 # Normalizing the curve

```
mu_2 <- mean(res_2)
v_2 <- var(res_2)
sd_2 <- sqrt(v_2)
x_2 <- seq(-10, 15, length=25)
y_2 <- dnorm(x_2,mu_2,sd_2)
lines(x_2,y_2,col="blue")
```



Model 3: MPG vs Displacement

model_3 <- lm(mpg~displacement, data = data_300)
summary(model_3)

```
> summary(model_3)

Call:
lm(formula = mpg ~ displacement, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.9282 -2.0043 -0.5401  1.9737 16.1501 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  31.352035   0.435875   71.93   <2e-16 ***
displacement -0.048913   0.001809  -27.04   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.412 on 298 degrees of freedom
Multiple R-squared:  0.7104,	Adjusted R-squared:  0.7094 
F-statistic: 731.1 on 1 and 298 DF,  p-value: < 2.2e-16
```

- Value of R^2 is 0.7104 that means our model only explains 71.04% of variance.
- Equation of Regression Line is:

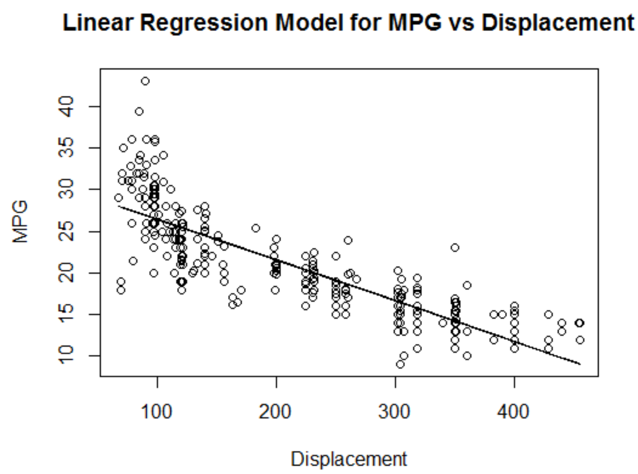
MPG = 31.352035 – 0.048913*displacement + Ei
Where Ei ~ N(0,3.412)
B0 -> Intercept -> 40.5619792
F – statistic: 731.1 on 1 and 298 DF, p-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that displacement has no effect on MPG.
Residual = Observed(Y) – Predicted(Y)

- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

```
coef(model_3)[1]
(Intercept)
  31.35204
coef(model_3)[2]
displacement
-0.04891259
```

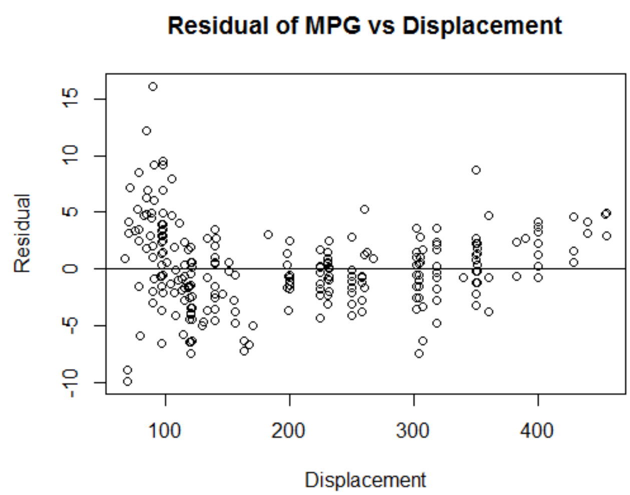
```
plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "Linear Regression Model for MPG vs Displacement")
```

```
lines(data_300$displacement, coef(model_3)[1]+coef(model_3)[2]*data_300$displacement)
```



 # Residual Plot

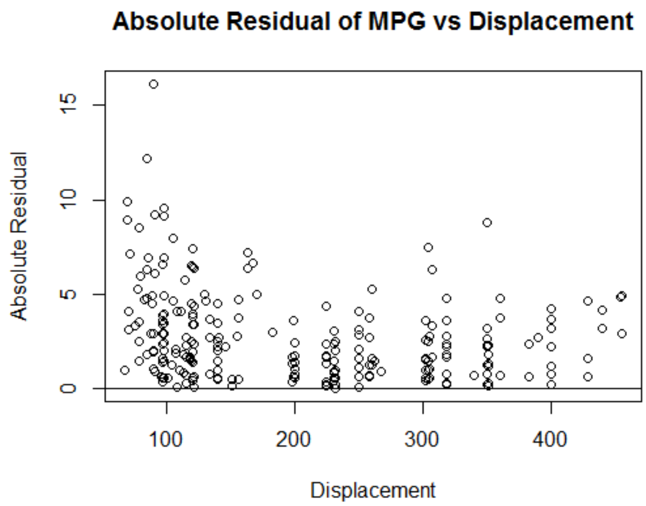
```
res_3 <- residuals(model_3)
plot(data_300$displacement, res_3,
    xlab = "Displacement",
    ylab = "Residual",
    main = "Residual of MPG vs Displacement")
abline(0,0)
```



 # Absolute Residual Plot

```
a_res_3 <- abs(res_3)
plot(data_300$displacement, a_res_3,
    xlab = "Displacement",
    ylab = "Absolute Residual",
```

```
    main = "Absolute Residual of MPG vs Displacement")
abline(0,0)
```

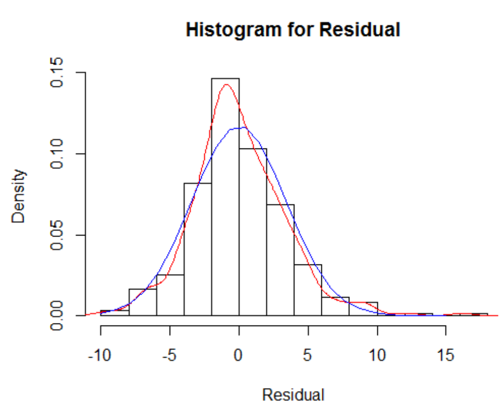


 # Histogram for residual

```
hist(res_3, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_3), col="red")
```

 # Normalizing the curve

```
mu_3 <- mean(res_3)
v_3 <- var(res_3)
sd_3 <- sqrt(v_3)
x_3 <- seq(-10, 15, length=25)
y_3 <- dnorm(x_3,mu_3,sd_3)
lines(x_3,y_3,col="blue")
```



Model 4: MPG vs Horsepower

model_4 <- lm(mpg~horsepower, data = data_300)
summary(model_4)

```
> summary(model_4)

Call:
lm(formula = mpg ~ horsepower, data = data_300)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.8442  -2.7816  -0.3376   2.4948  14.2360 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 34.903508   0.648037   53.86   <2e-16 ***
horsepower  -0.125824   0.005455  -23.07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.8 on 298 degrees of freedom
Multiple R-squared:  0.641,	Adjusted R-squared:  0.6397 
F-statistic:   532 on 1 and 298 DF,  p-value: < 2.2e-16
```

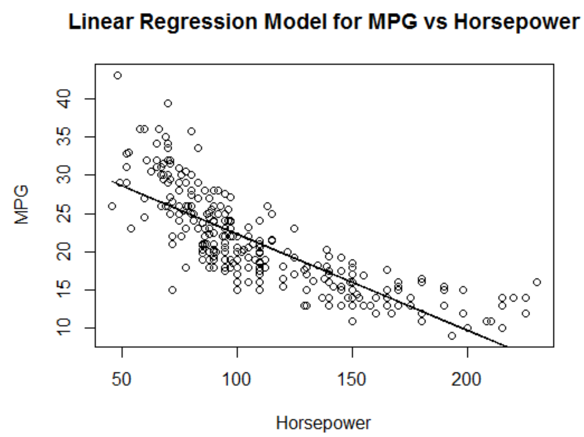
- Value of R^2 is 0.641 that means our model only explains 64.10% of variance.
- Equation of Regression Line is:

MPG = 34.903508 – 0.125824*horsepower + Ei
Where Ei ~ N(0,3.8)
B0 -> Intercept -> 34.903508
F – statistic: 532 on 1 and 298 DF, p-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that horsepower has no effect on MPG.
Residual = Observed(Y) – Predicted(Y)

- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

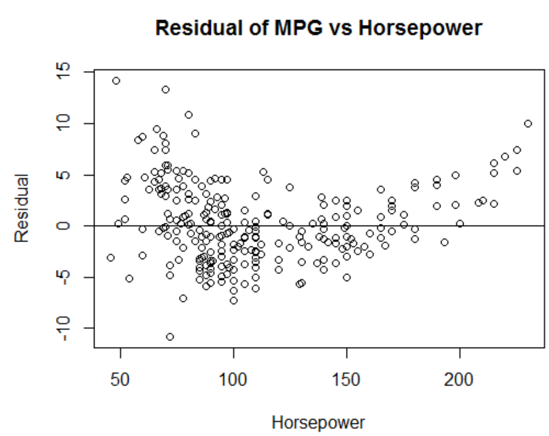
```
coef(model_4)[1]
(Intercept) 
  34.90351 
coef(model_4)[2]
horsepower 
-0.1258239 
```

```
plot(data_300$horsepower, data_300$mpg,
    xlab = "Horsepower",
    ylab = "MPG",
    main = "Linear Regression Model for MPG vs Horsepower")
lines(data_300$horsepower, coef(model_4)[1]+coef(model_4)[2]*data_300$horsepower)
```



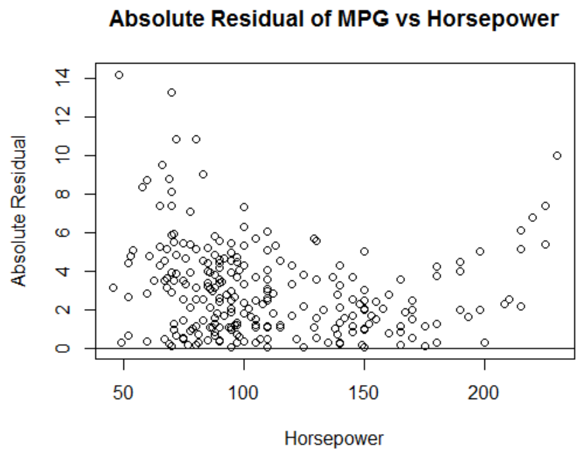
Residual Plot

```
res_4 <- residuals(model_4)
plot(data_300$horsepower, res_4,
    xlab = "Horsepower",
    ylab = "Residual",
    main = "Residual of MPG vs Horsepower")
abline(0,0)
```



Absolute Residual Plot

```
a_res_4 <- abs(res_4)
plot(data_300$horsepower, a_res_4,
    xlab = "Horsepower",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Horsepower")
abline(0,0)
```

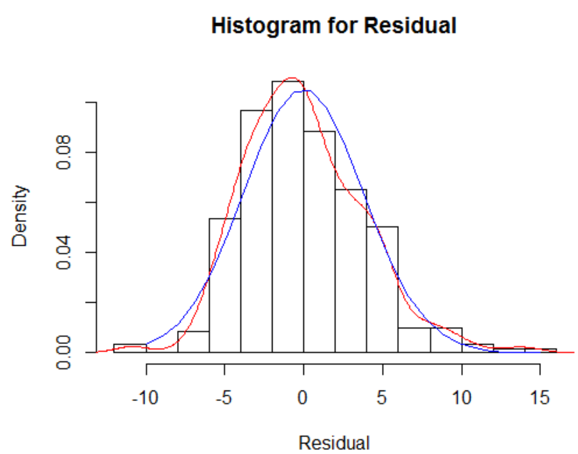


Histogram for residual

```
hist(res_4, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_4), col="red")
```

Normalizing the curve

```
mu_4 <- mean(res_4)
v_4 <- var(res_4)
sd_4 <- sqrt(v_4)
x_4 <- seq(-10, 15, length=25)
y_4 <- dnorm(x_4,mu_4,sd_4)
lines(x_4,y_4,col="blue")
```



Model 5: MPG vs Weight and Acceleration

model_5 <- lm(mpg~weight+acceleration, data = data_300)
summary(model_5)

```
> summary(model_5)

Call:
lm(formula = mpg ~ weight + acceleration, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.7181 -1.9002 -0.0528  1.7374 14.3311 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  37.5864385  1.5397024  24.411   <2e-16 ***
weight       -0.0060753  0.0002216 -27.410   <2e-16 ***
acceleration  0.1507902  0.0709118   2.126   0.0343 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.014 on 297 degrees of freedom
Multiple R-squared:  0.7748,	Adjusted R-squared:  0.7733 
F-statistic: 510.9 on 2 and 297 DF,  p-value: < 2.2e-16
```

- Value of R^2 is 0.7748 that means our model only explains 77.48% of variance.
- Equation of Regression Line is:

MPG = 37.5864385 – 0.0060753*weight + 0.1507902*acceleration + Ei
Where Ei ~ N(0,3.014)
B0 -> Intercept -> 37.5864385

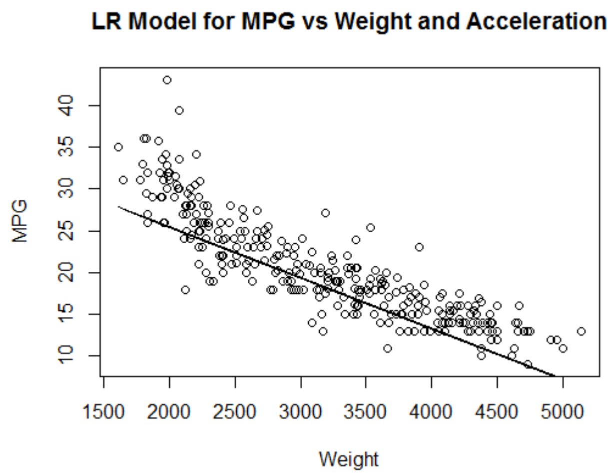
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG.
<u>But for acceleration P value is 0.0343 >0.01</u> ,so this is not significant at 1% level.Hence we can say acceleration does not impact MPG much.so we can exclude acceleration from our best model.

Residual = Observed(Y) – Predicted(Y)

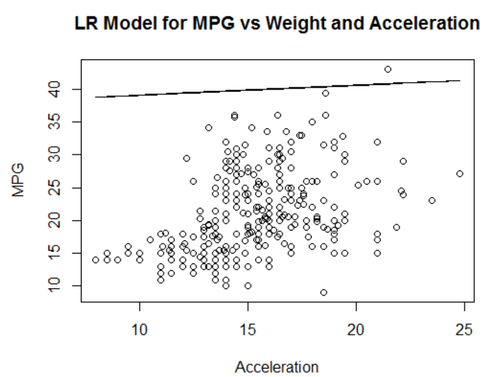
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

coef(model_5)[1]
(Intercept)
37.58644
coef(model_5)[2]
weight
-0.006075346
coef(model_5)[3]
acceleration
0.1507902

```
plot(data_300$weight, data_300$mpg,
    xlab = "Weight",
    ylab = "MPG",
    main = "LR Model for MPG vs Weight and Acceleration")
lines(data_300$weight, coef(model_5)[1]+coef(model_5)[2]*data_300$weight)
```

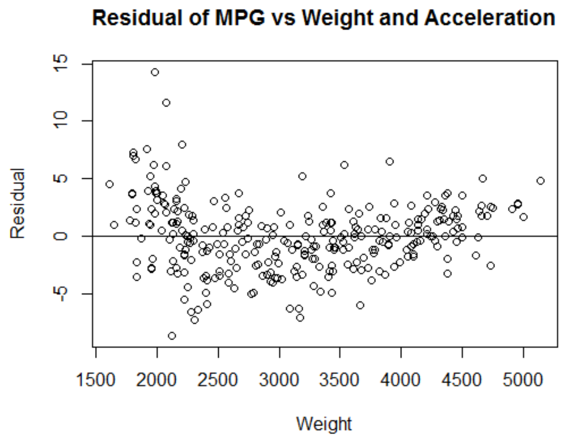



```
plot(data_300$acceleration, data_300$mpg,
    xlab = "Acceleration",
    ylab = "MPG",
    main = "LR Model for MPG vs Weight and Acceleration")
lines(data_300$acceleration, coef(model_5)[1]+coef(model_5)[3]*data_300$acceleration)
```

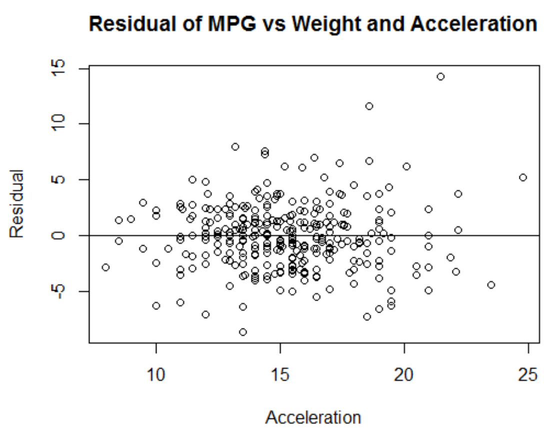


 # Residual Plot

```
res_5 <- residuals(model_5)
plot(data_300$weight, res_5,
    xlab = "Weight",
    ylab = "Residual",
    main = "Residual of MPG vs Weight and Acceleration")
abline(0,0)
```

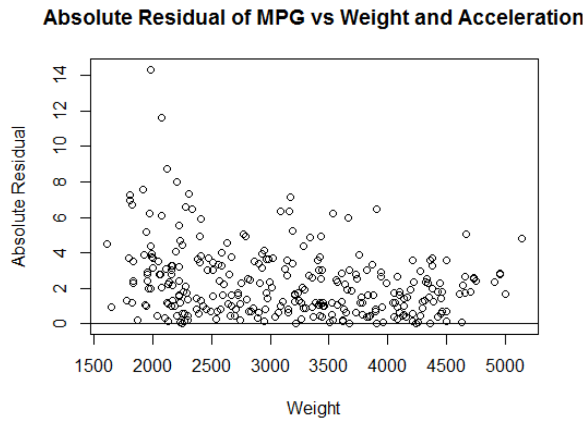



```
plot(data_300$acceleration, res_5,
    xlab = "Acceleration",
    ylab = "Residual",
    main = "Residual of MPG vs Weight and Acceleration")
abline(0,0)
```

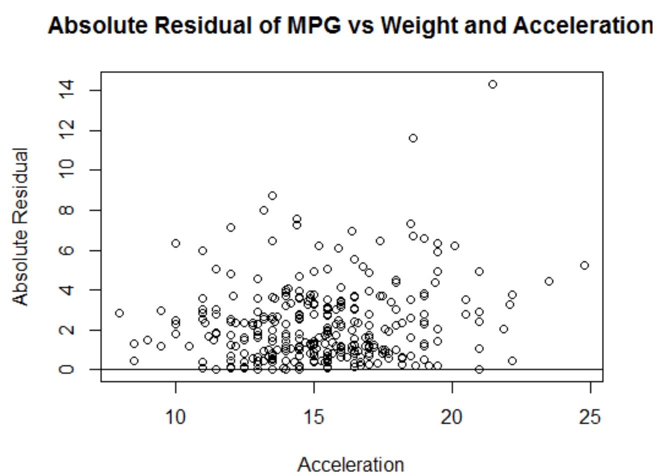


 # Absolute Residual Plot

```
a_res_5 <- abs(res_5)
plot(data_300$weight, a_res_5,
    xlab = "Weight",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Weight and Acceleration")
abline(0,0)
```




```
plot(data_300$acceleration, a_res_5,
    xlab = "Acceleration",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Weight and Acceleration")
abline(0,0)
```

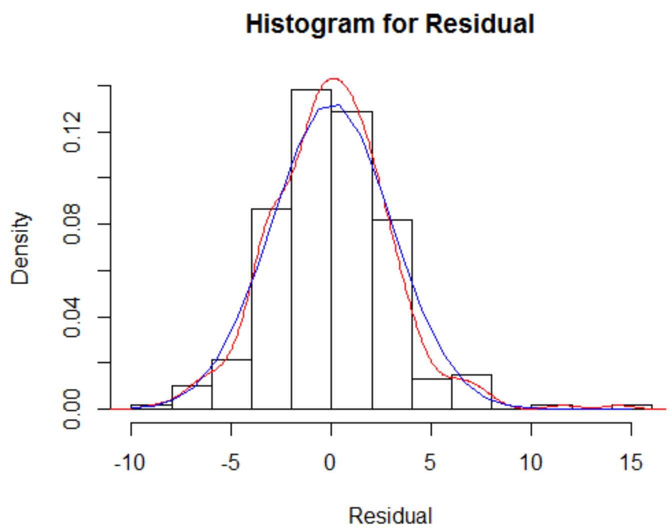


 # Histogram for residual

```
hist(res_5, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_5), col="red")
```

Normalizing the curve

```
mu_5 <- mean(res_5)
v_5 <- var(res_5)
sd_5 <- sqrt(v_5)
x_5 <- seq(-10, 15, length=25)
y_5 <- dnorm(x_5,mu_5,sd_5)
lines(x_5,y_5,col="blue")
```



Model 6: MPG vs Displacement and Weight

model_6 <- lm(mpg~weight+displacement, data = data_300)
summary(model_6)

```
> summary(model_6)

Call:
lm(formula = mpg ~ weight + displacement, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.5232 -1.8741 -0.1765  1.6766 15.1041 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 38.9422109  0.8971239  43.408   <2e-16 ***
weight      -0.0050080  0.0005351  -9.360   <2e-16 ***
displacement -0.0111720  0.0043353  -2.577   0.0104 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.004 on 297 degrees of freedom
Multiple R-squared:  0.7764,	Adjusted R-squared:  0.7749 
F-statistic: 515.6 on 2 and 297 DF,  p-value: < 2.2e-16
```

- Value of R^2 is 0.7764 that means our model only explains 77.64% of variance.
- Equation of Regression Line is:

MPG = 38.9422109 – 0.0050080*weight -0.0111720*displacement + Ei
Where Ei ~ N(0,3.004)
B0 -> Intercept -> 38.9422109

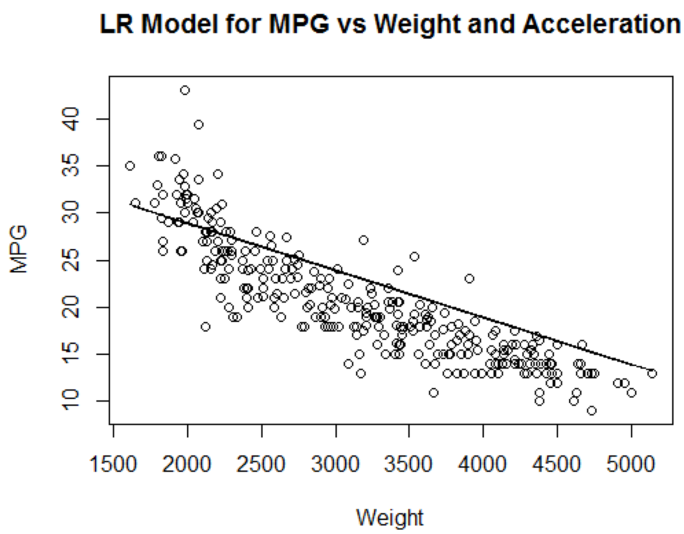
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG.
<u>But for Displacement P value is 0.0104</u> ,so we cannot reject H0.Hence we can say that displacement does not impact MPG much.

Residual = Observed(Y) – Predicted(Y)

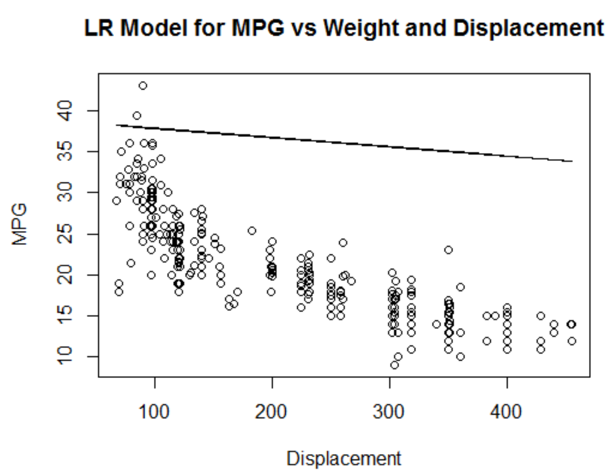
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

```
coef(model_6)[1]
(Intercept)
 38.94221
coef(model_6)[2]
    weight
-0.005007991
coef(model_6)[3]
displacement
-0.01117198
```

```
plot(data_300$weight, data_300$mpg,
    xlab = "Weight",
    ylab = "MPG",
    main = "LR Model for MPG vs Weight and Acceleration")
lines(data_300$weight, coef(model_6)[1]+coef(model_6)[2]*data_300$weight)
```



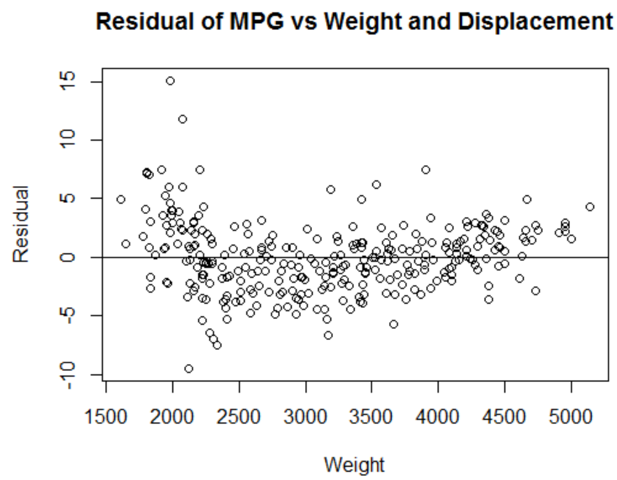

```
plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "LR Model for MPG vs Weight and Displacement")
lines(data_300$displacement, coef(model_6)[1]+coef(model_6)[3]*data_300$displacement)
```



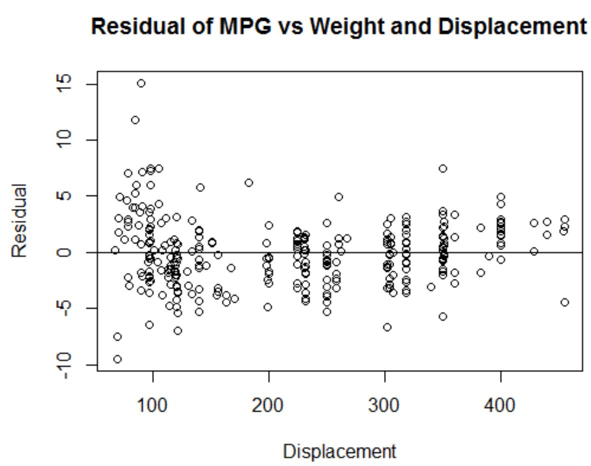
 # Residual Plot

```
res_6 <- residuals(model_6)
plot(data_300$weight, res_6,
    xlab = "Weight",
```

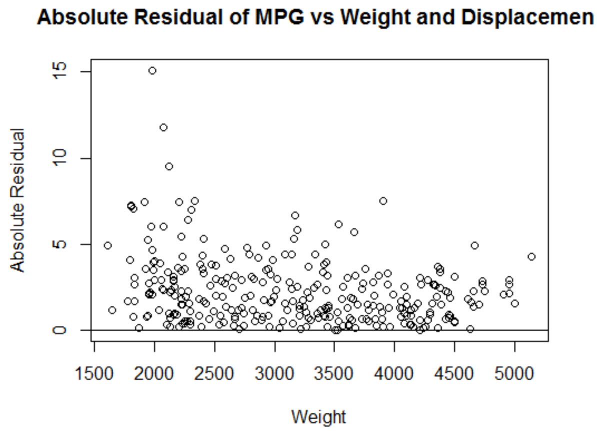
```
    ylab = "Residual",
    main = "Residual of MPG vs Weight and Displacement")
abline(0,0)
```



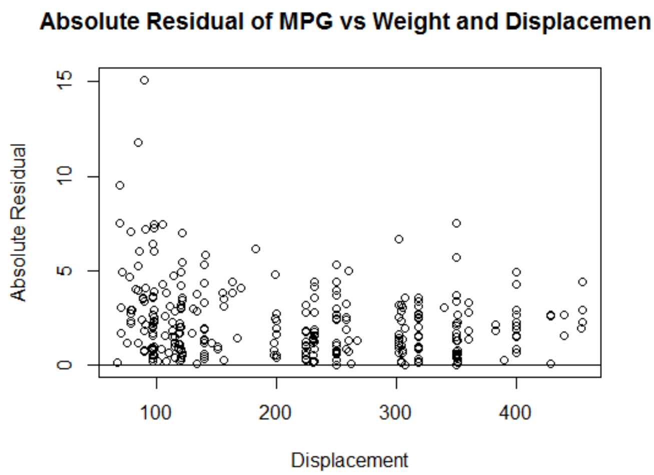

```
plot(data_300$displacement, res_6,
    xlab = "Displacement",
    ylab = "Residual",
    main = "Residual of MPG vs Weight and Displacement")
abline(0,0)
```




```
# Absolute Residual Plot

a_res_6 <- abs(res_6)
plot(data_300$weight, a_res_6,
    xlab = "Weight",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Weight and Displacement")
abline(0,0)
```




```
plot(data_300$displacement, a_res_6,
    xlab = "Displacement",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Weight and Displacement")
abline(0,0)
```

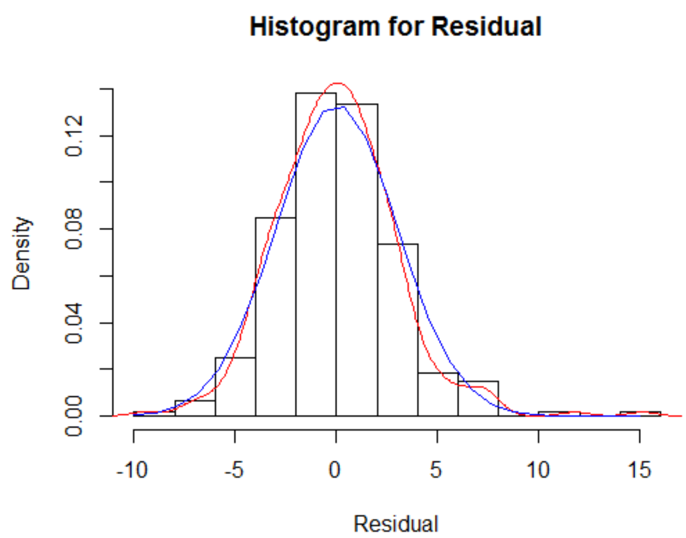


Histogram for residual

```
hist(res_6, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_6), col="red")
```

Normalizing the curve

```
mu_6 <- mean(res_6)
v_6 <- var(res_6)
sd_6 <- sqrt(v_6)
x_6 <- seq(-10, 15, length=25)
y_6 <- dnorm(x_6,mu_6,sd_6)
lines(x_6,y_6,col="blue")
```



Model 7: MPG vs Horsepower and Weight

model_7 <- lm(mpg~weight+horsepower, data = data_300)
summary(model_7)

```
> summary(model_7)

Call:
lm(formula = mpg ~ weight + horsepower, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.7069 -1.8380  0.0207  1.6877 14.5038 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 40.2587429  0.6420610  62.702  < 2e-16 ***
weight      -0.0052041  0.0003808 -13.666  < 2e-16 ***
horsepower  -0.0277594  0.0083560  -3.322  0.00101 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.982 on 297 degrees of freedom
Multiple R-squared:  0.7796,	Adjusted R-squared:  0.7781 
F-statistic: 525.2 on 2 and 297 DF,  p-value: < 2.2e-16
```

- Value of R^2 is 0.7796 that means our model only explains 77.96% of variance.
- Equation of Regression Line is:

MPG = 40.2587429 – 0.0052041*weight – 0.0277594*horsepower + Ei

Where Ei ~ N(0,2.982)

B0 -> Intercept -> 40.2587429

P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG.

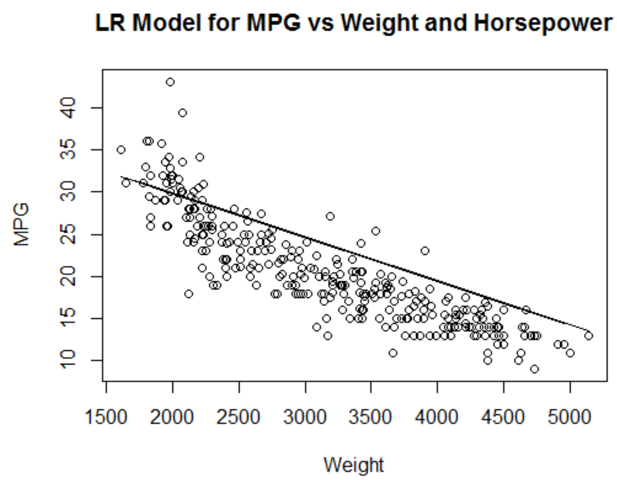
<u>But for horsepower P value is 0.00101</u>,hence we cannot Reject H0.Horsepower does not impact MPG much.

Residual = Observed(Y) – Predicted(Y)

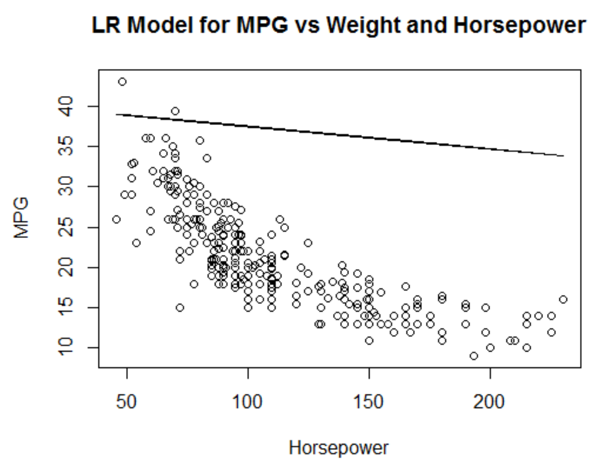
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

```
coef(model_7)[1]
(Intercept) 
  40.25874 
coef(model_7)[2]
     weight 
-0.005204094 
coef(model_7)[3]
horsepower 
-0.02775943 
```

```
plot(data_300$weight, data_300$mpg,
    xlab = "Weight",
    ylab = "MPG",
    main = "LR Model for MPG vs Weight and Horsepower")
lines(data_300$weight, coef(model_7)[1]+coef(model_7)[2]*data_300$weight)
```

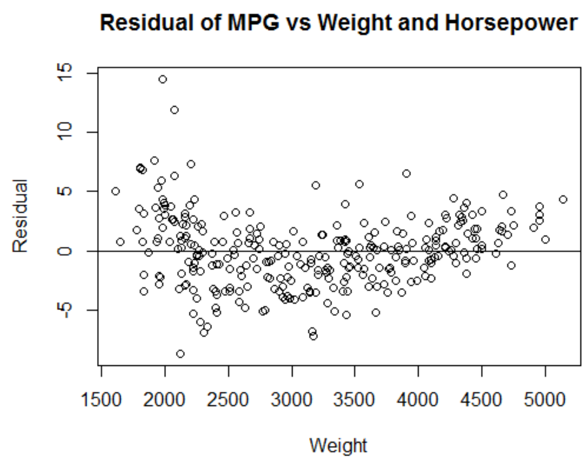



```
plot(data_300$horsepower, data_300$mpg,
    xlab = "Horsepower",
    ylab = "MPG",
    main = "LR Model for MPG vs Weight and Horsepower")
lines(data_300$horsepower, coef(model_7)[1]+coef(model_7)[3]*data_300$horsepower)
```

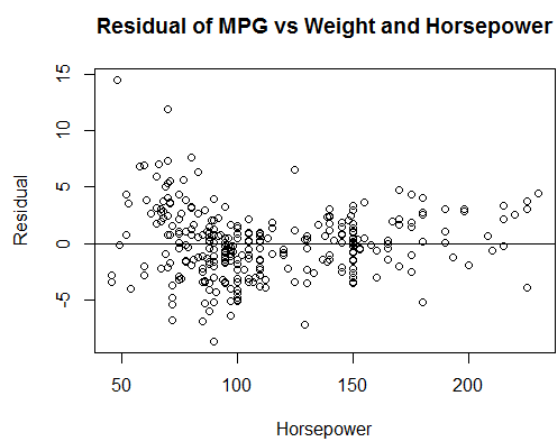


Residual Plot

```
res_7 <- residuals(model_7)
plot(data_300$weight, res_7,
    xlab = "Weight",
    ylab = "Residual",
    main = "Residual of MPG vs Weight and Horsepower")
abline(0,0)
```

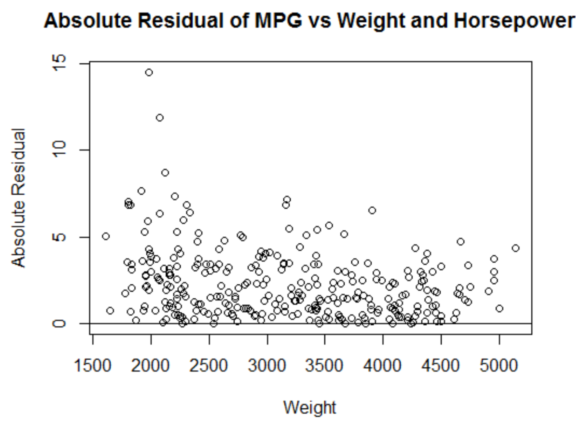



```
plot(data_300$horsepower, res_7,
    xlab = "Horsepower",
    ylab = "Residual",
    main = "Residual of MPG vs Weight and Horsepower")
abline(0,0)
```

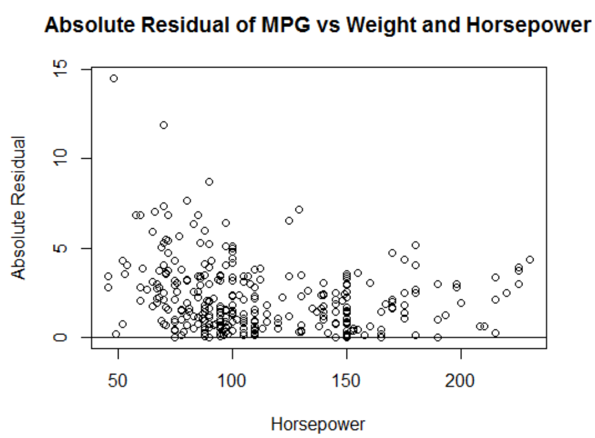


 # Absolute Residual Plot

```
a_res_7 <- abs(res_7)
plot(data_300$weight, a_res_7,
    xlab = "Weight",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Weight and Horsepower")
abline(0,0)
```




```
plot(data_300$horsepower, a_res_7,
    xlab = "Horsepower",
    ylab = "Absolute Residual",
    main = "Absolute Residual of MPG vs Weight and Horsepower")
abline(0,0)
```

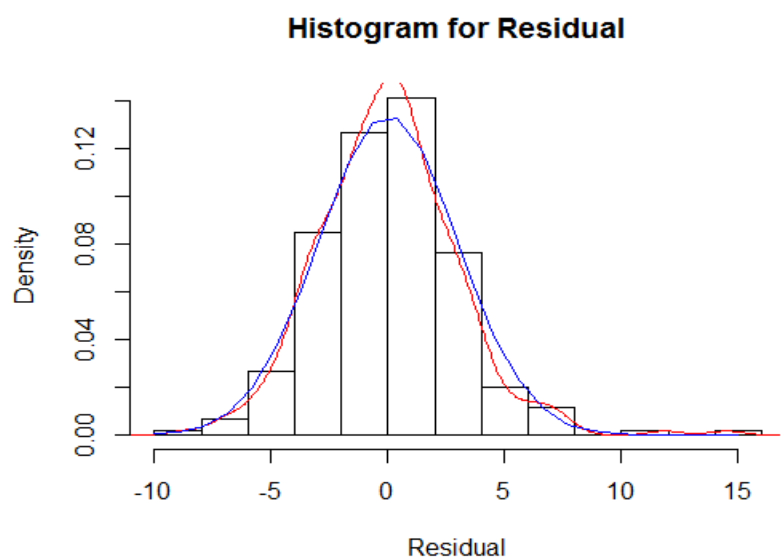


Histogram for residual

```
hist(res_7, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_7), col="red")
```

Normalizing the curve

```
mu_7 <- mean(res_7)
v_7 <- var(res_7)
sd_7 <- sqrt(v_7)
x_7 <- seq(-10, 15, length=25)
y_7 <- dnorm(x_7,mu_7,sd_7)
lines(x_7,y_7,col="blue")
```



Model 8: MPG vs Displacement and Acceleration

model_8 <- lm(mpg~acceleration+displacement, data = data_300)
summary(model_8)

```
> summary(model_8)

Call:
lm(formula = mpg ~ acceleration + displacement, data = data_300)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.4895  -1.9515  -0.2762   1.7056  16.7794 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  33.999761   1.679392  20.245   <2e-16 ***
acceleration -0.143501   0.087918  -1.632    0.104    
displacement -0.051043   0.002227 -22.924   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.403 on 297 degrees of freedom
Multiple R-squared:  0.713,	Adjusted R-squared:  0.7111 
F-statistic: 368.9 on 2 and 297 DF,  p-value: < 2.2e-16
```

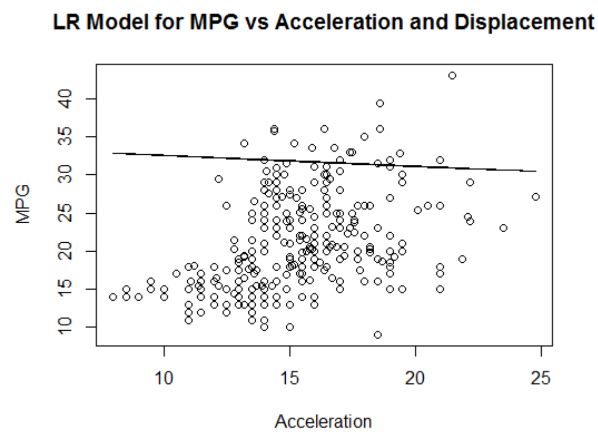
- Value of R^2 is 0.713 that means our model only explains 71.3% of variance.
- Equation of Regression Line is:

MPG = 33.999761 – 0.143501*acceleration – 0.051043*displacement + Ei
Where Ei ~ N(0,3.403)
B0 -> Intercept -> 33.999761
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis h0 that displacement has no effect on MPG.
But for acceleration P value is 0.104 ,test is not significant ,hence we can say that acceleration has not much impact on MPG
Residual = Observed(Y) – Predicted(Y)

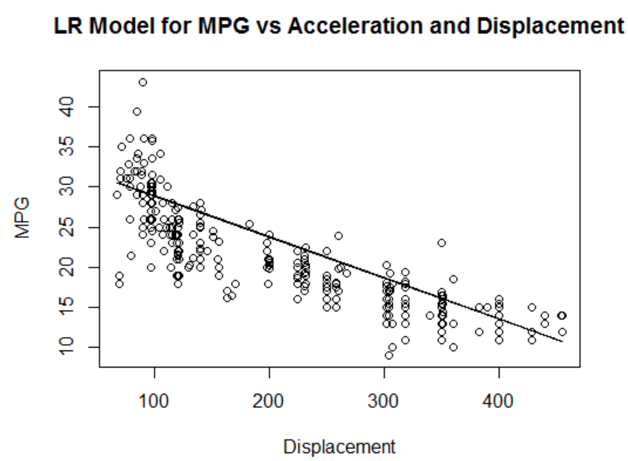
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

coef(model_8)[1]
(Intercept)
 33.99976
coef(model_8)[2]
acceleration
 -0.1435012

```
coef(model_8)[3]
displacement
-0.05104283

plot(data_300$acceleration, data_300$mpg,
    xlab = "Acceleration",
    ylab = "MPG",
    main = "LR Model for MPG vs Acceleration and Displacement")
lines(data_300$acceleration, coef(model_8)[1]+coef(model_8)[2]*data_300$acceleration)
```

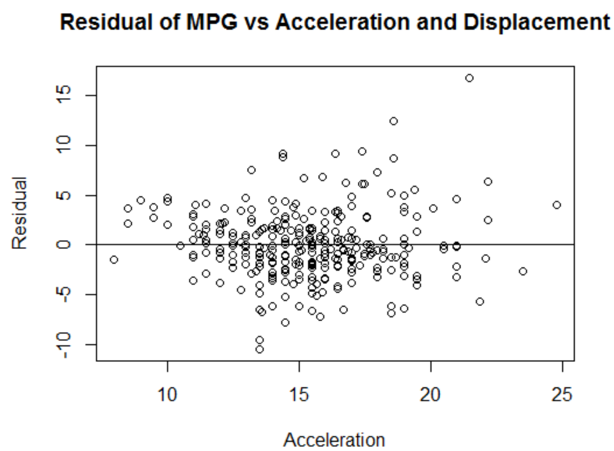



```
plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "LR Model for MPG vs Acceleration and Displacement")
lines(data_300$displacement, coef(model_8)[1]+coef(model_8)[3]*data_300$displacement)
```

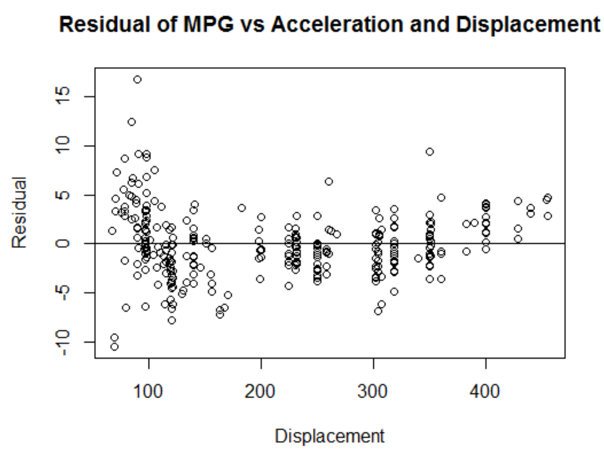


Residual Plot

```
res_8 <- residuals(model_8)
plot(data_300$acceleration, res_8,
   xlab = "Acceleration",
   ylab = "Residual",
   main = "Residual of MPG vs Acceleration and Displacement")
abline(0,0)
```

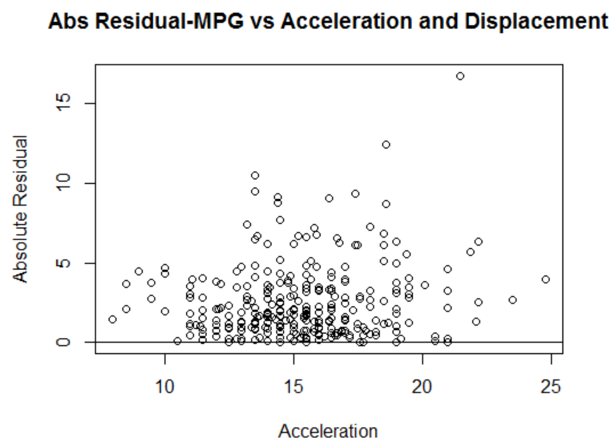



```
plot(data_300$displacement, res_8,
   xlab = "Displacement",
   ylab = "Residual",
   main = "Residual of MPG vs Acceleration and Displacement")
abline(0,0)
```

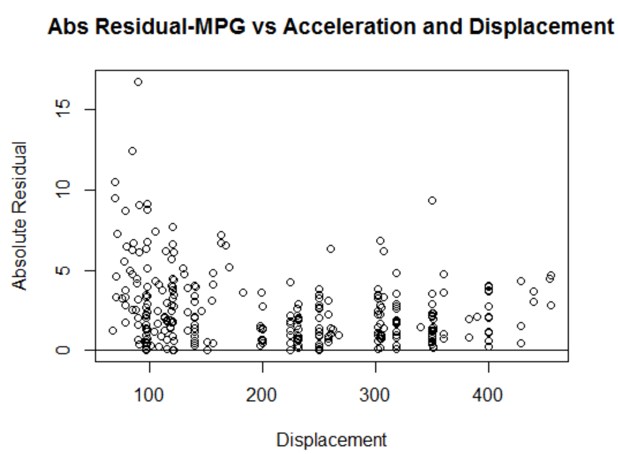


Absolute Residual Plot

```
a_res_8 <- abs(res_8)
plot(data_300$acceleration, a_res_8,
   xlab = "Acceleration",
   ylab = "Absolute Residual",
   main = "Abs Residual-MPG vs Acceleration and Displacement")
abline(0,0)
```




```
plot(data_300$displacement, a_res_8,
   xlab = "Displacement",
   ylab = "Absolute Residual",
   main = "Abs Residual-MPG vs Acceleration and Displacement")
abline(0,0)
```

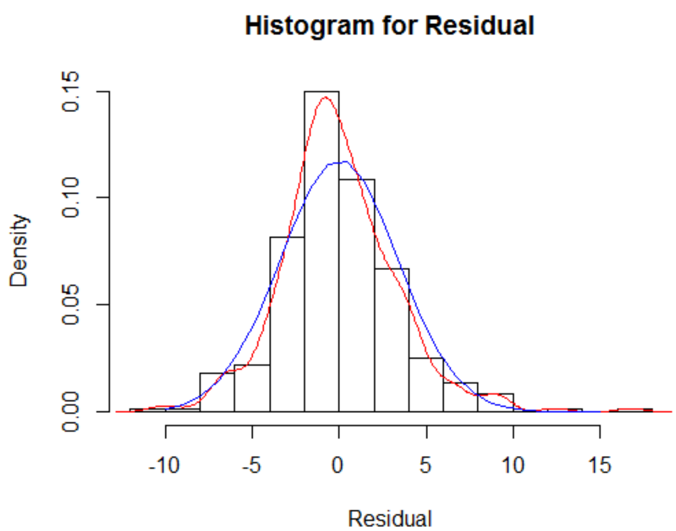


 # Histogram for residual

```
hist(res_8, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_8), col="red")
```

 # Normalizing the curve

```
mu_8 <- mean(res_8)
v_8 <- var(res_8)
sd_8 <- sqrt(v_8)
x_8 <- seq(-10, 15, length=25)
y_8 <- dnorm(x_8,mu_8,sd_8)
lines(x_8,y_8,col="blue")
```



Model 9: MPG vs Horsepower and Acceleration

```
model_9 <- lm(mpg~acceleration+horsepower, data = data_300)
summary(model_9)
```

```
> summary(model_9)

Call:
lm(formula = mpg ~ acceleration + horsepower, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.5183 -2.2855 -0.5122  2.1517 16.0371 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  46.993617   2.358577  19.925  < 2e-16 ***
acceleration -0.582022   0.109545  -5.313 2.12e-07 ***
horsepower   -0.154526   0.007513 -20.566  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.637 on 297 degrees of freedom
Multiple R-squared:  0.6721,	Adjusted R-squared:  0.6699 
F-statistic: 304.4 on 2 and 297 DF,  p-value: < 2.2e-16
```

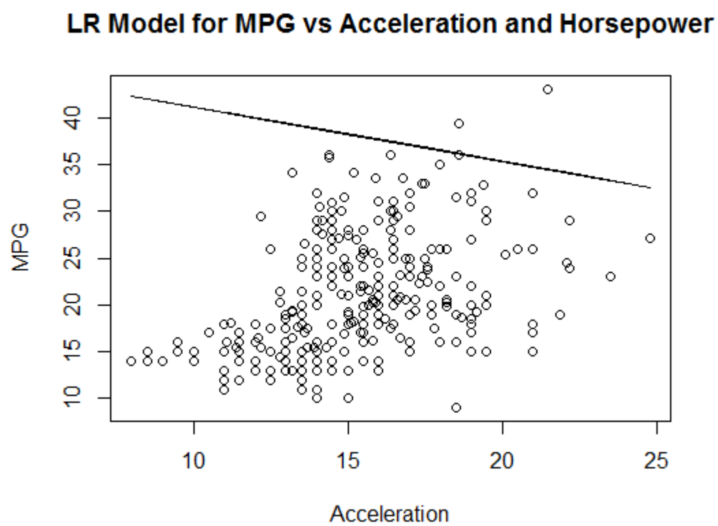
- Value of R^2 is 0.6721 that means our model only explains 67.21% of variance.
- Equation of Regression Line is:

MPG = 46.993617 – 0.582022*acceleration – 0.154526*horsepower + Ei
Where Ei ~ N(0,3.637)
B0 -> Intercept -> 46.993617
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that acceleration and horsepower has no effect on MPG.
Residual = Observed(Y) – Predicted(Y)

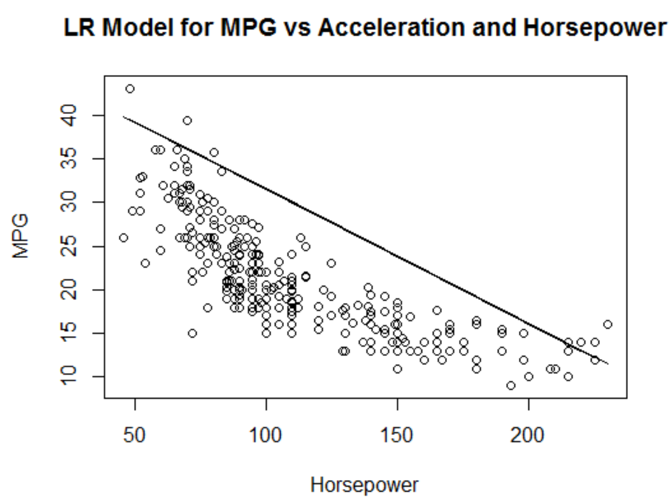
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

```
coef(model_9)[1]
(Intercept)
  46.99362
coef(model_9)[2]
acceleration
  -0.582022
coef(model_9)[3]
horsepower
-0.1545263
```

```
plot(data_300$acceleration, data_300$mpg,
   xlab = "Acceleration",
   ylab = "MPG",
   main = "LR Model for MPG vs Acceleration and Horsepower")
lines(data_300$acceleration, coef(model_9)[1]+coef(model_9)[2]*data_300$acceleration)
```

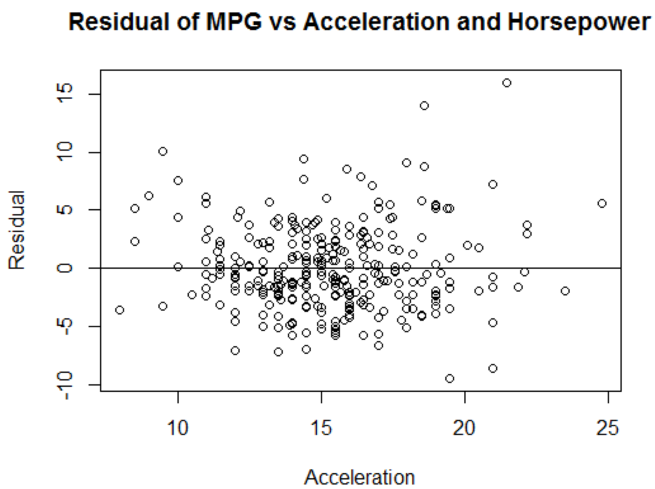



```
plot(data_300$horsepower, data_300$mpg,
   xlab = "Horsepower",
   ylab = "MPG",
   main = "LR Model for MPG vs Acceleration and Horsepower")
lines(data_300$horsepower, coef(model_9)[1]+coef(model_9)[3]*data_300$horsepower)
```

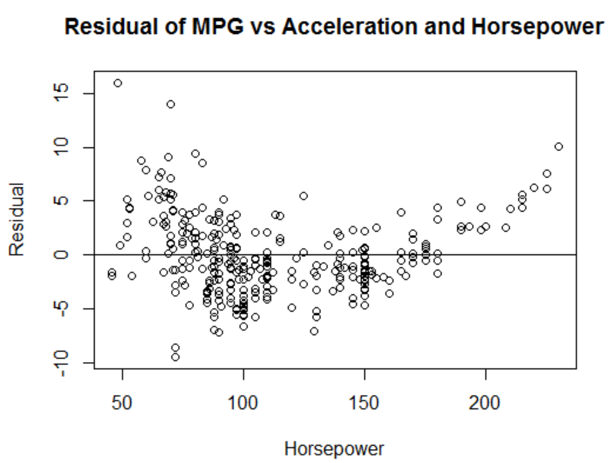


Residual Plot

```
res_9 <- residuals(model_9)
plot(data_300$acceleration, res_9,
    xlab = "Acceleration",
    ylab = "Residual",
    main = "Residual of MPG vs Acceleration and Horsepower")
abline(0,0)
```

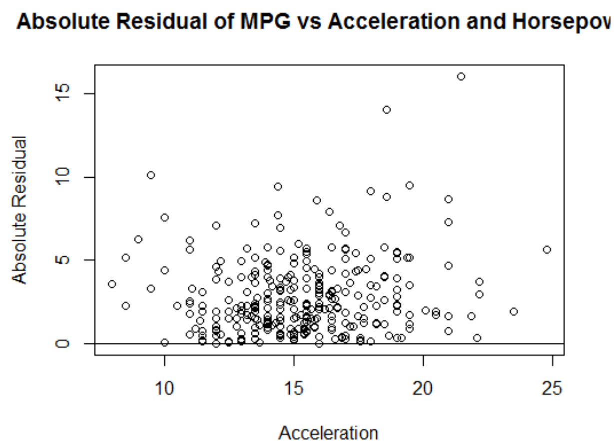



```
plot(data_300$horsepower, res_9,
    xlab = "Horsepower",
    ylab = "Residual",
    main = "Residual of MPG vs Acceleration and Horsepower")
abline(0,0)
```

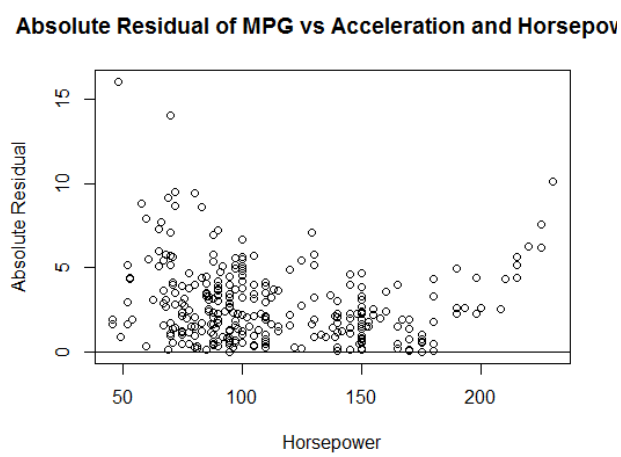


Absolute Residual Plot

```
a_res_9 <- abs(res_9)
plot(data_300$acceleration, a_res_9,
   xlab = "Acceleration",
   ylab = "Absolute Residual",
   main = "Absolute Residual of MPG vs Acceleration and Horsepower")
abline(0,0)
```




```
plot(data_300$horsepower, a_res_9,
   xlab = "Horsepower",
   ylab = "Absolute Residual",
   main = "Absolute Residual of MPG vs Acceleration and Horsepower")
abline(0,0)
```

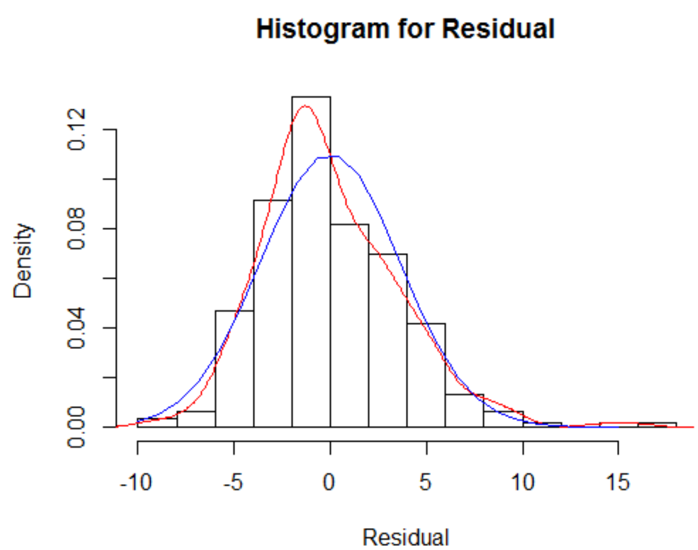


Histogram for residual

```
hist(res_9, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_9), col="red")
```

Normalizing the curve

```
mu_9 <- mean(res_9)
v_9 <- var(res_9)
sd_9 <- sqrt(v_9)
x_9 <- seq(-10, 15, length=25)
y_9 <- dnorm(x_9,mu_9,sd_9)
lines(x_9,y_9,col="blue")
```



Model 10: MPG vs Displacement and Horsepower

model_10 <- lm(mpg~displacement+horsepower, data = data_300)
summary(model_10)

```
> summary(model_10)

Call:
lm(formula = mpg ~ displacement + horsepower, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.9715 -2.1390 -0.3849  1.9555 15.3248 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  32.847122   0.613586  53.533  < 2e-16 ***
displacement -0.036803   0.003978  -9.252  < 2e-16 ***
horsepower   -0.036660   0.010773  -3.403 0.000758 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.353 on 297 degrees of freedom
Multiple R-squared:  0.7213,	Adjusted R-squared:  0.7194 
F-statistic: 384.3 on 2 and 297 DF,  p-value: < 2.2e-16
```

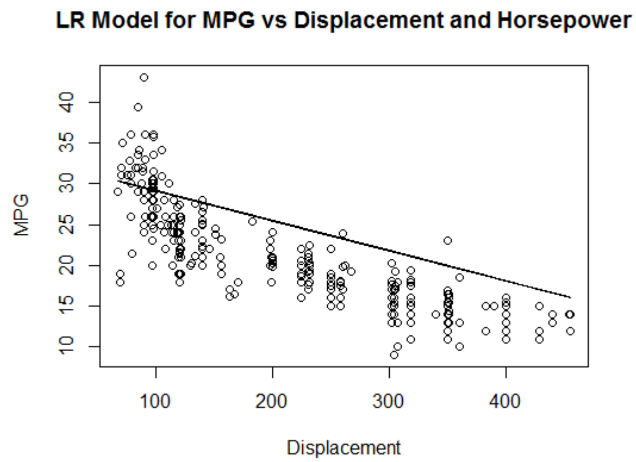
- Value of R^2 is 0.7213 that means our model only explains 72.13% of variance.
- Equation of Regression Line is:

MPG = 32.847122 – 0.036803*displacement – 0.036660*horsepower + Ei
Where Ei ~ N(0,3.353)
B0 -> Intercept -> 32.847122
P values: < 2.2e-16 and 0.000758 are lesser than significance level, clearly shows that we should Reject NULL Hypothesis that displacement and horsepower has no effect on MPG.
Residual = Observed(Y) – Predicted(Y)

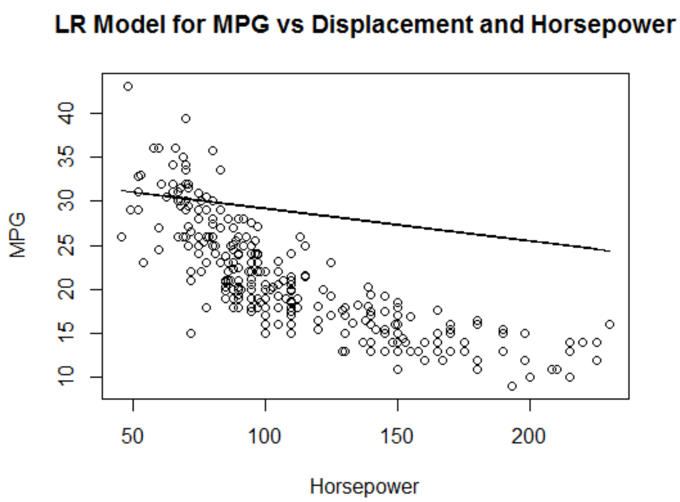
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

coef(model_10)[1]
(Intercept)
 32.84712
coef(model_10)[2]
displacement
-0.03680311
coef(model_10)[3]
horsepower
-0.03666022

```
plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "LR Model for MPG vs Displacement and Horsepower")
lines(data_300$displacement, coef(model_10)[1]+coef(model_10)[2]*data_300$displacement)
```

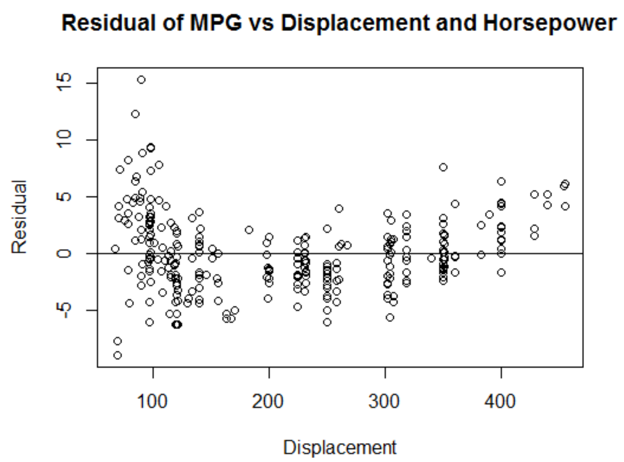



```
plot(data_300$horsepower, data_300$mpg,
    xlab = "Horsepower",
    ylab = "MPG",
    main = "LR Model for MPG vs Displacement and Horsepower")
lines(data_300$horsepower, coef(model_10)[1]+coef(model_10)[3]*data_300$horsepower)
```

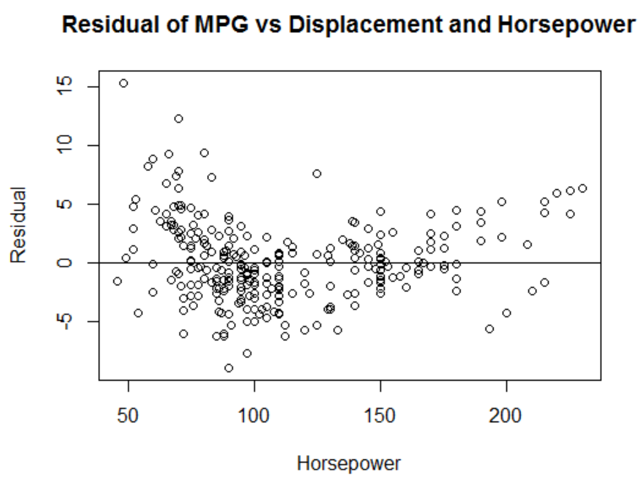


Residual Plot

```
res_10 <- residuals(model_10)
plot(data_300$displacement, res_10,
    xlab = "Displacement",
    ylab = "Residual",
    main = "Residual of MPG vs Displacement and Horsepower")
abline(0,0)
```

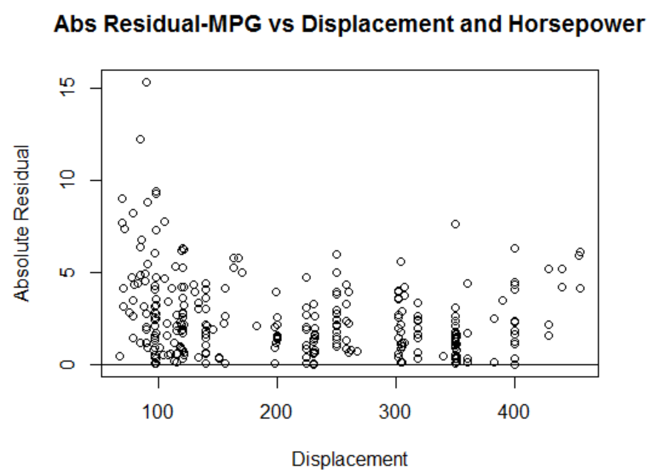



```
plot(data_300$horsepower, res_10,
    xlab = "Horsepower",
    ylab = "Residual",
    main = "Residual of MPG vs Displacement and Horsepower")
abline(0,0)
```

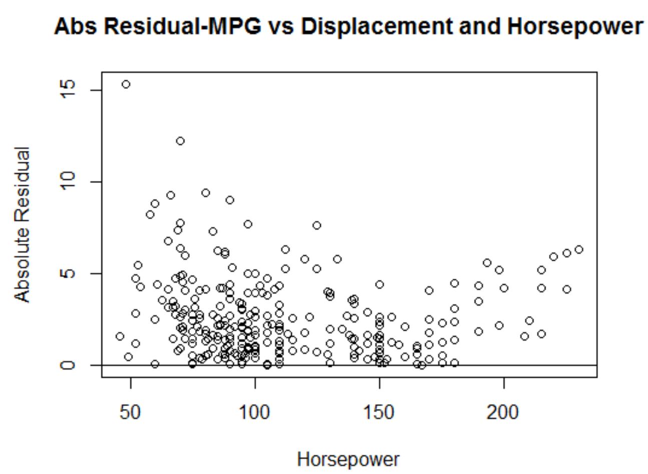


Absolute Residual Plot

```
a_res_10 <- abs(res_10)
plot(data_300$displacement, a_res_10,
    xlab = "Displacement",
    ylab = "Absolute Residual",
    main = "Abs Residual-MPG vs Displacement and Horsepower")
abline(0,0)
```




```
plot(data_300$horsepower, a_res_10,
    xlab = "Horsepower",
    ylab = "Absolute Residual",
    main = "Abs Residual-MPG vs Displacement and Horsepower")
abline(0,0)
```

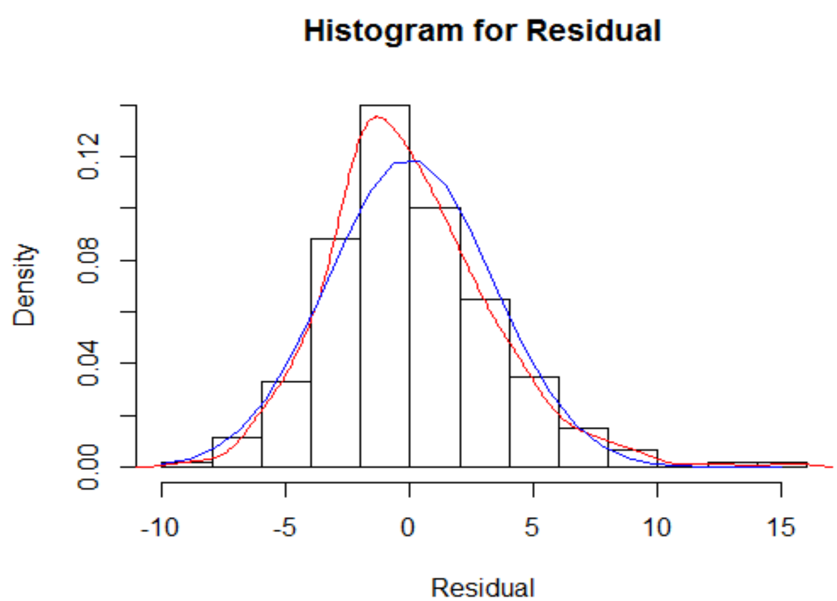


 # Histogram for residual

```
hist(res_10, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_10), col="red")
```

 # Normalizing the curve

```
mu_10 <- mean(res_10)
v_10 <- var(res_10)
sd_10 <- sqrt(v_10)
x_10 <- seq(-10, 15, length=25)
y_10 <- dnorm(x_10,mu_10,sd_10)
lines(x_10,y_10,col="blue")
```



Model 11: MPG vs Displacement, Horsepower and Weight

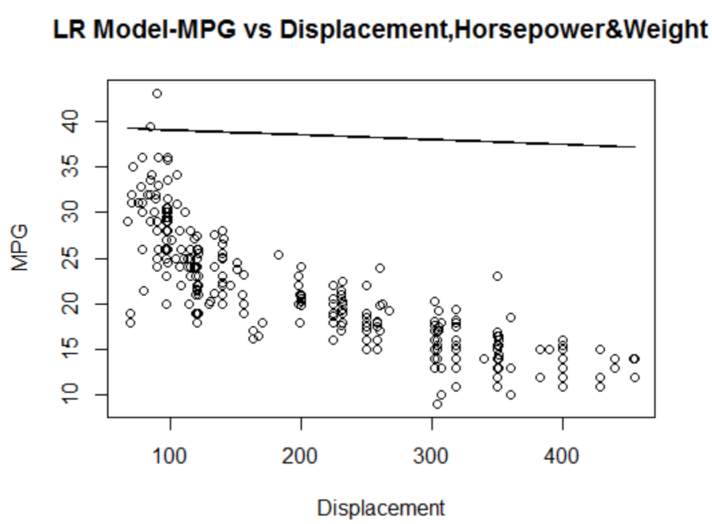
model_11 <- lm(mpg~displacement+horsepower+weight, data = data_300)
summary(model_11)

```
> summary(model_11)

Call:
lm(formula = mpg ~ displacement + horsepower + weight, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.9508 -1.8780 -0.0657  1.6311 14.6386 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  39.5540648  0.9286288  42.594   <2e-16 ***
displacement -0.0052516  0.0050006  -1.050   0.2945    
horsepower   -0.0225670  0.0097080  -2.325   0.0208 *  
weight       -0.0048045  0.0005383  -8.925   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.982 on 296 degrees of freedom
Multiple R-squared:  0.7804,	Adjusted R-squared:  0.7782 
F-statistic: 350.6 on 3 and 296 DF,  p-value: < 2.2e-16
```

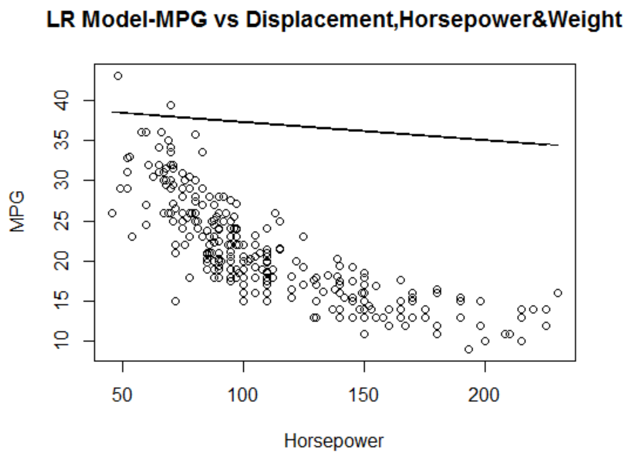
- Value of R^2 is 0.7804 that means our model only explains 78.04% of variance.
- Equation of Regression Line is:

MPG = 39.5540648 – 0.0052516*displacement – 0.0225670*horsepower – 0.0048045*weight + Ei
Where Ei ~ N(0,2.982)
B0 -> Intercept -> 39.5540648
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG.Weight has major impact on MPG(Ha).
P values : 0.2945 and 0.0208 are not significant at 1% level,hence horsepower and displacement have no impact on MPG.
Residual = Observed(Y) – Predicted(Y)
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

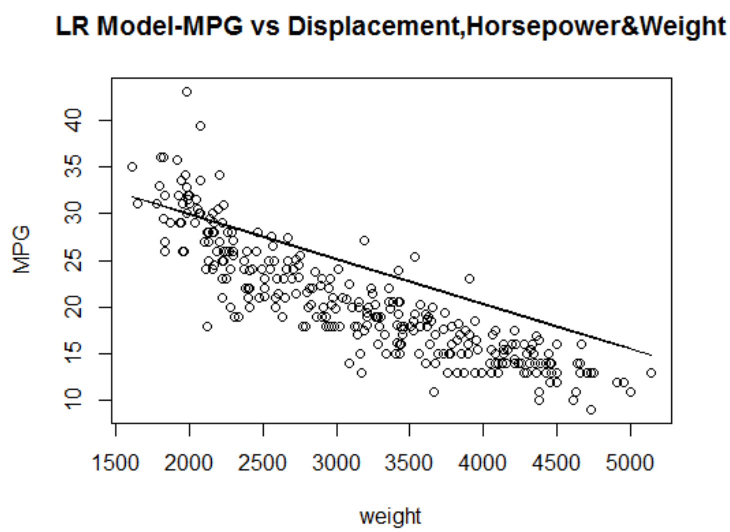
```
coef(model_11)[1]
(Intercept)
  39.55406
coef(model_11)[2]
displacement
-0.005251572
coef(model_11)[3]
horsepower
-0.022567
coef(model_11)[4]
     weight
-0.004804456

plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "LR Model-MPG vs Displacement,Horsepower&Weight")
lines(data_300$displacement, coef(model_11)[1]+coef(model_11)[2]*data_300$displacement)
```




```
plot(data_300$horsepower, data_300$mpg,
    xlab = "Horsepower",
    ylab = "MPG",
    main = "LR Model-MPG vs Displacement,Horsepower&Weight")
lines(data_300$horsepower, coef(model_11)[1]+coef(model_11)[3]*data_300$horsepower)
```

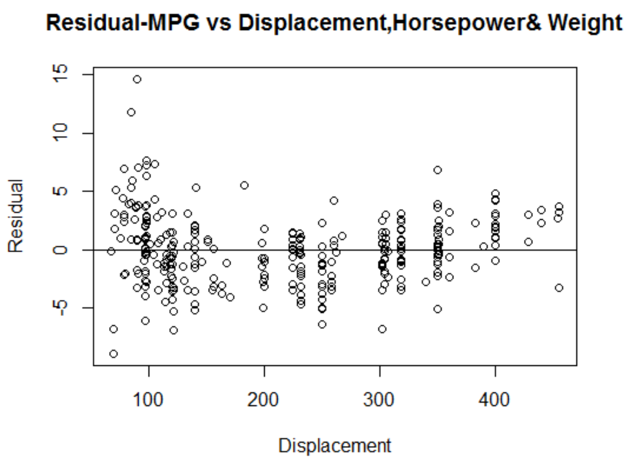



```
plot(data_300$weight, data_300$mpg,
   xlab = "weight",
   ylab = "MPG",
   main = "LR Model-MPG vs Displacement,Horsepower&Weight")
lines(data_300$weight, coef(model_11)[1]+coef(model_11)[4]*data_300$weight)
```

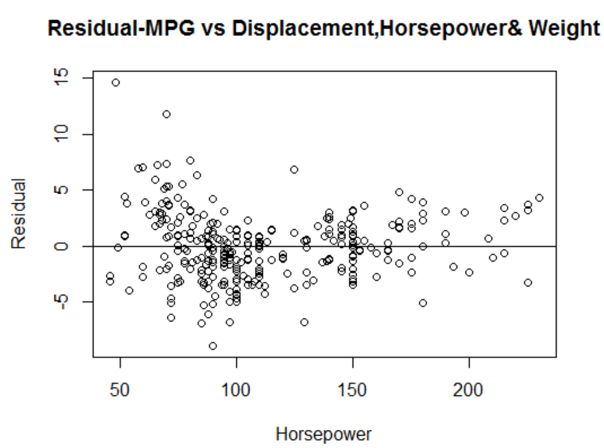


Residual Plot

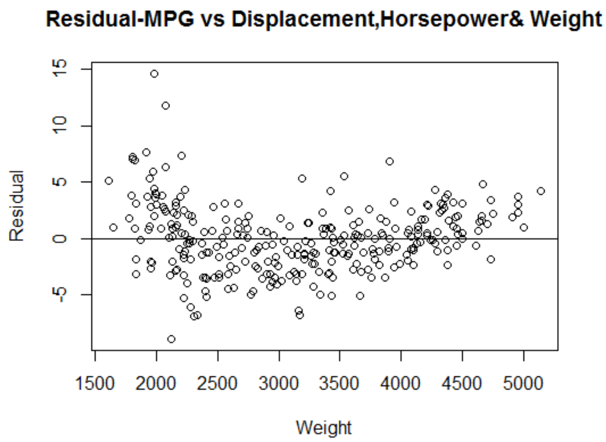
```
res_11 <- residuals(model_11)
plot(data_300$displacement, res_11,
   xlab = "Displacement",
   ylab = "Residual",
   main = "Residual-MPG vs Displacement,Horsepower& Weight")
abline(0,0)
```




```
plot(data_300$horsepower, res_11,
   xlab = "Horsepower",
   ylab = "Residual",
   main = "Residual-MPG vs Displacement,Horsepower& Weight")
abline(0,0)
```

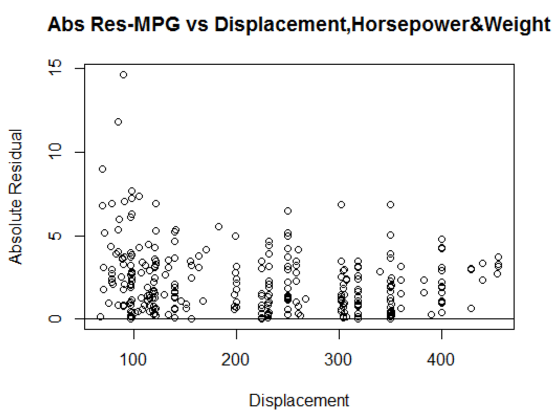


```
plot(data_300$weight, res_11,
    xlab = "Weight",
    ylab = "Residual",
    main = "Residual-MPG vs Displacement,Horsepower& Weight")
abline(0,0)
```

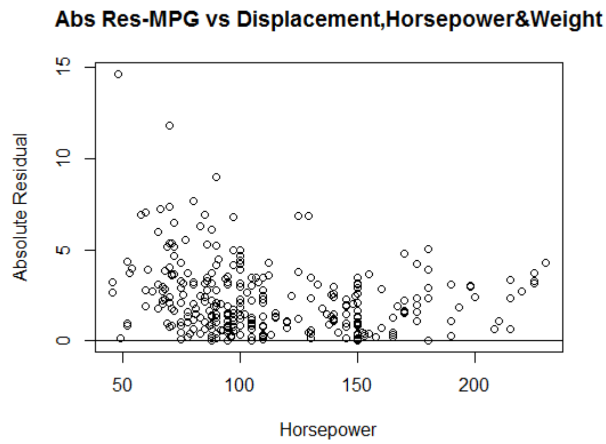


 # Absolute Residual Plot

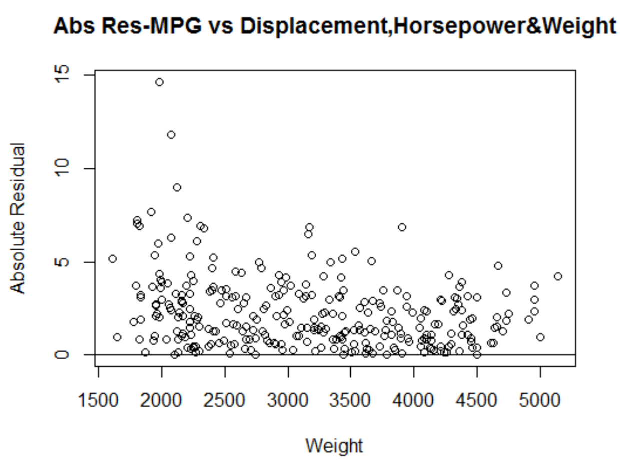
```
a_res_11 <- abs(res_11)
plot(data_300$displacement, a_res_11,
    xlab = "Displacement",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacement,Horsepower&Weight")
abline(0,0)
```



```
plot(data_300$horsepower, a_res_11,
    xlab = "Horsepower",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacement,Horsepower&Weight")
abline(0,0)
```




```
plot(data_300$weight, a_res_11,
    xlab = "Weight",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacement,Horsepower&Weight")
abline(0,0)
```

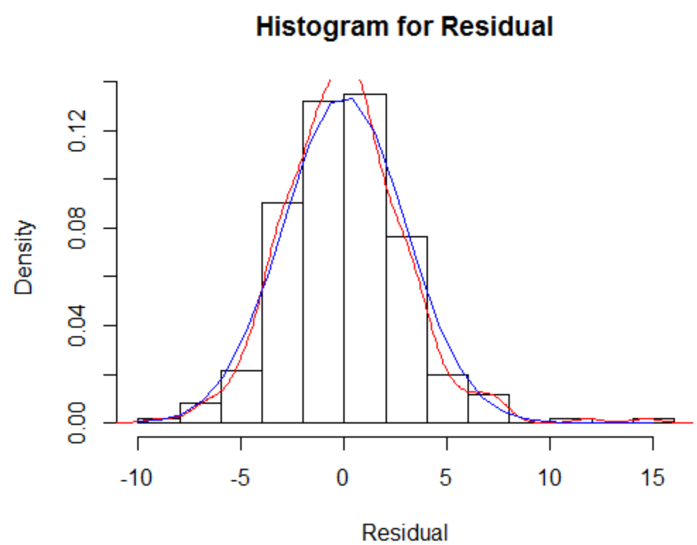


 # Histogram for residual

```
hist(res_11, prob=T, breaks = 10, main = "Histogram for Residual",
   xlab = "Residual",
   ylab = "Density")
lines(density(res_11), col="red")
```

 # Normalizing the curve

```
mu_11 <- mean(res_11)
v_11 <- var(res_11)
sd_11 <- sqrt(v_11)
x_11 <- seq(-10, 15, length=25)
y_11 <- dnorm(x_11,mu_11,sd_11)
lines(x_11,y_11,col="blue")
```



Model 12: MPG vs Displacement, Weight and Acceleration

model_12 <- lm(mpg~displacement+weight+acceleration, data = data_300)
summary(model_12)

```
> summary(model_12)

Call:
lm(formula = mpg ~ displacement + weight + acceleration, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.1954 -1.8965 -0.0852  1.7040 14.7171 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  37.7097417  1.5359949  24.551   <2e-16 ***
displacement -0.0087326  0.0049885  -1.751   0.0811 .  
weight       -0.0051733  0.0005606  -9.228   <2e-16 ***
acceleration  0.0803801  0.0813115   0.989   0.3237    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.004 on 296 degrees of freedom
Multiple R-squared:  0.7771,	Adjusted R-squared:  0.7749 
F-statistic:   344 on 3 and 296 DF,  p-value: < 2.2e-16
```

- Value of R^2 is 0.7771 that means our model only explains 77.71% of variance.
- Equation of Regression Line is:

MPG = 37.7097417 – 0.0087326*displacement – 0.0051733*weight + 0.0803801*acceleration + Ei
Where Ei ~ N(0,3.004)
B0 -> Intercept -> 37.7097417

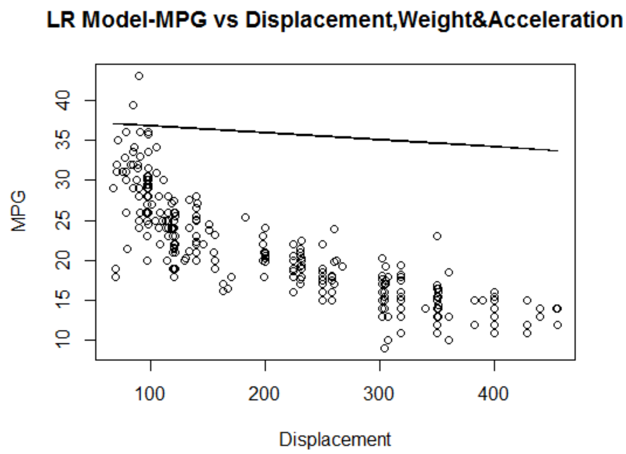
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG.Weight is impacting MPG majorly.
P values: 0.0811 and 0.3237 are greater than significance level,hence we cannot reject Null Hypothesis H0. Displacement and acceleration has no impact on MPG.

Residual = Observed(Y) – Predicted(Y)

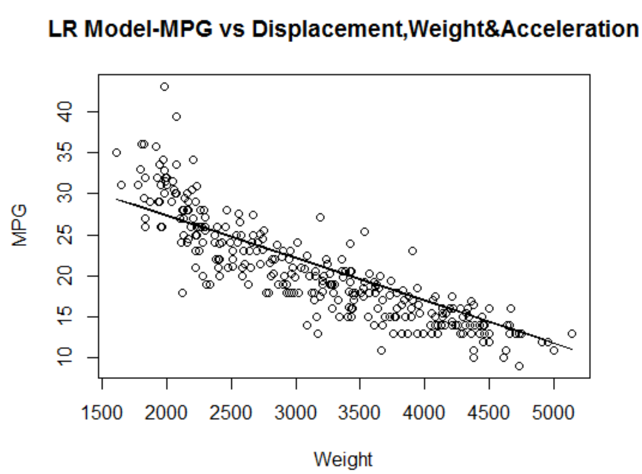
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

```
coef(model_12)[1]
(Intercept) 
  37.70974 
coef(model_12)[2]
displacement 
-0.008732619
```

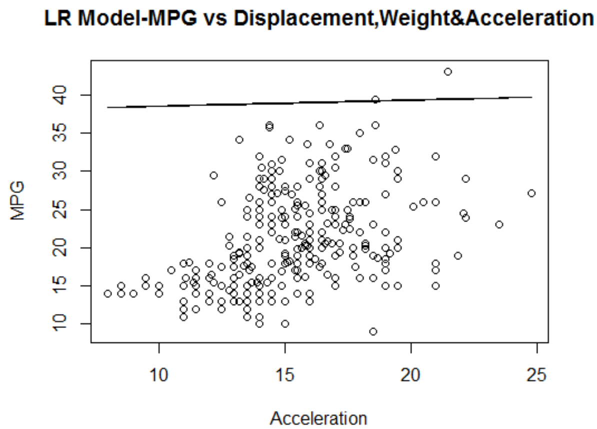
```
coef(model_12)[3]
     weight
-0.005173346
coef(model_12)[4]
acceleration
  0.0803801

plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "LR Model-MPG vs Displacement,Weight&Acceleration")
lines(data_300$displacement, coef(model_12)[1]+coef(model_12)[2]*data_300$displacement)
```




```
plot(data_300$weight, data_300$mpg,
    xlab = "Weight",
    ylab = "MPG",
    main = "LR Model-MPG vs Displacement,Weight&Acceleration")
lines(data_300$weight, coef(model_12)[1]+coef(model_12)[3]*data_300$weight)
```

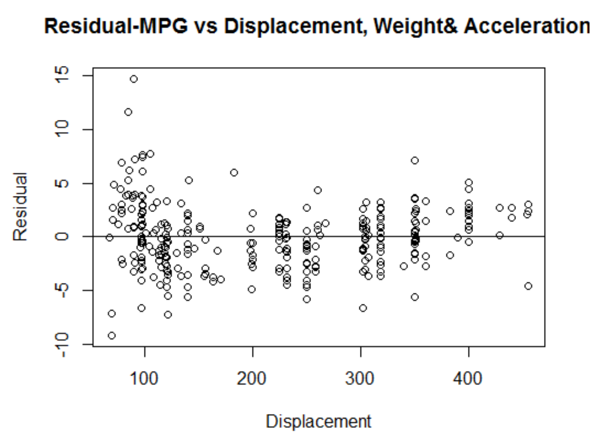


```
plot(data_300$acceleration, data_300$mpg,
    xlab = "Acceleration",
    ylab = "MPG",
    main = "LR Model-MPG vs Displacement,Weight&Acceleration")
lines(data_300$acceleration, coef(model_12)[1]+coef(model_12)[4]*data_300$acceleration)
```

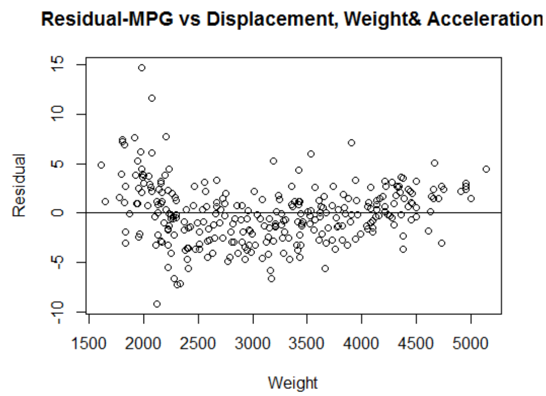


 # Residual Plot

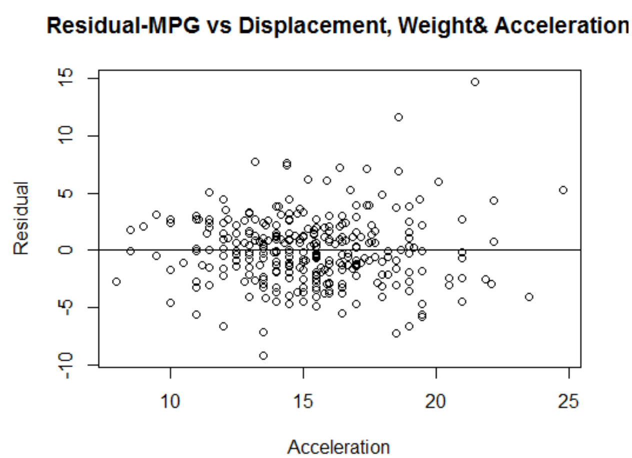
```
res_12 <- residuals(model_12)
plot(data_300$displacement, res_12,
    xlab = "Displacement",
    ylab = "Residual",
    main = "Residual-MPG vs Displacement, Weight& Acceleration")
abline(0,0)
```



```
plot(data_300$weight, res_12,
    xlab = "Weight",
    ylab = "Residual",
    main = "Residual-MPG vs Displacement, Weight& Acceleration")
abline(0,0)
```

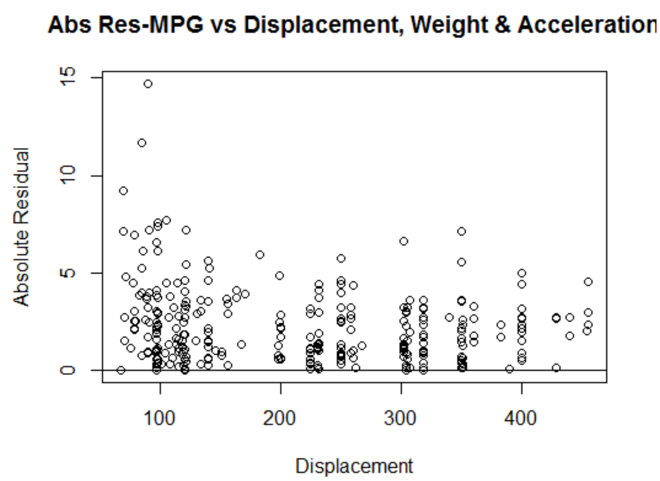



```
plot(data_300$acceleration, res_12,
    xlab = "Acceleration",
    ylab = "Residual",
    main = "Residual-MPG vs Displacement, Weight& Acceleration")
abline(0,0)
```

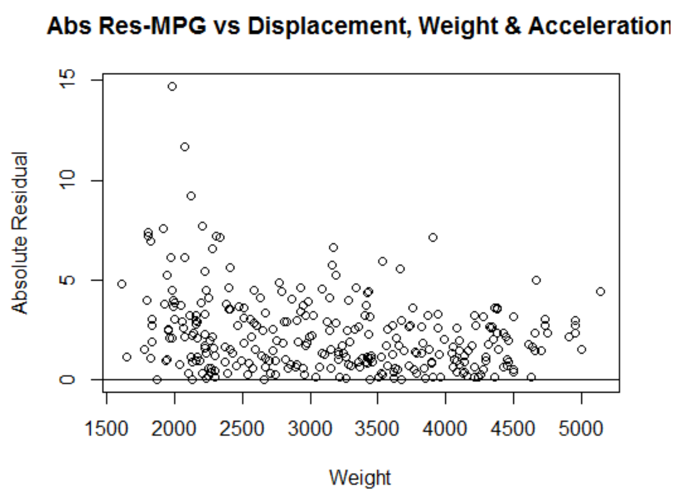


Absolute Residual Plot

```
a_res_12 <- abs(res_12)
plot(data_300$displacement, a_res_12,
    xlab = "Displacement",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacement, Weight & Acceleration")
abline(0,0)
```

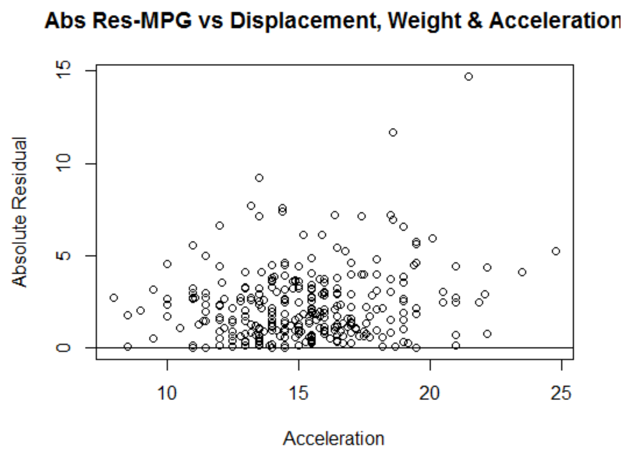



```
plot(data_300$weight, a_res_12,
    xlab = "Weight",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacement, Weight & Acceleration")
abline(0,0)
```




```
plot(data_300$acceleration, a_res_12,
    xlab = "Acceleration",
    ylab = "Absolute Residual",
```

```
    main = "Abs Res-MPG vs Displacement, Weight & Acceleration")
abline(0,0)
```

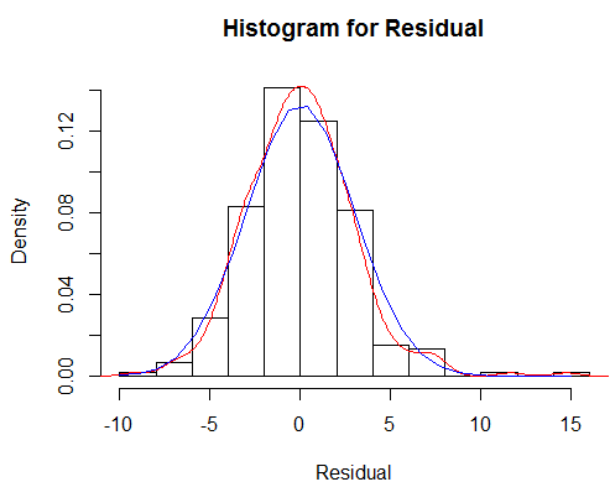


 # Histogram for residual

```
hist(res_12, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_12), col="red")
```

 # Normalizing the curve

```
mu_12 <- mean(res_12)
v_12 <- var(res_12)
sd_12 <- sqrt(v_12)
x_12 <- seq(-10, 15, length=25)
y_12 <- dnorm(x_12,mu_12,sd_12)
lines(x_12,y_12,col="blue")
```



Model 13: MPG vs Horsepower, Weight and Acceleration

model_13 <- lm(mpg~horsepower+weight+acceleration, data = data_300)
summary(model_13)

```
> summary(model_13)

Call:
lm(formula = mpg ~ horsepower + weight + acceleration, data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.7638 -1.7818  0.0114  1.6741 14.6052 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  40.8946033  2.0021356  20.425   <2e-16 ***
horsepower   -0.0306494  0.0120123  -2.552   0.0112 *  
weight       -0.0051392  0.0004276 -12.019   <2e-16 ***
acceleration -0.0338230  0.1008566  -0.335   0.7376    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.987 on 296 degrees of freedom
Multiple R-squared:  0.7797,	Adjusted R-squared:  0.7774 
F-statistic: 349.1 on 3 and 296 DF,  p-value: < 2.2e-16
```

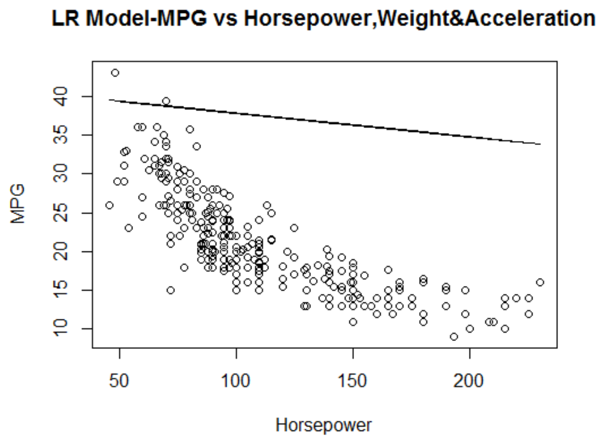
- Value of R^2 is 0.7797 that means our model only explains 77.97% of variance.
- Equation of Regression Line is:

MPG = 40.8946033 – 0.0306494*horsepower - 0.0051392*weight – 0.0338230*acceleration + Ei
Where Ei ~ N(0,2.987)
B0 -> Intercept -> 40.8946033
P value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG. Weight is impacting MPG(Ha) .
<u>P values: 0.0112 and 0.7376</u> are greater than significance level,hence we cannot reject H0.Horsepower and acceleration have no impact on MPG
Residual = Observed(Y) – Predicted(Y)

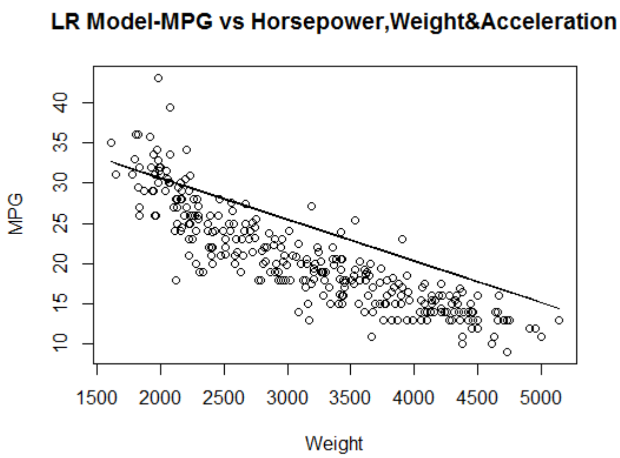
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

coef(model_13)[1]
(Intercept)
 40.8946
coef(model_13)[2]
horsepower
-0.0306494

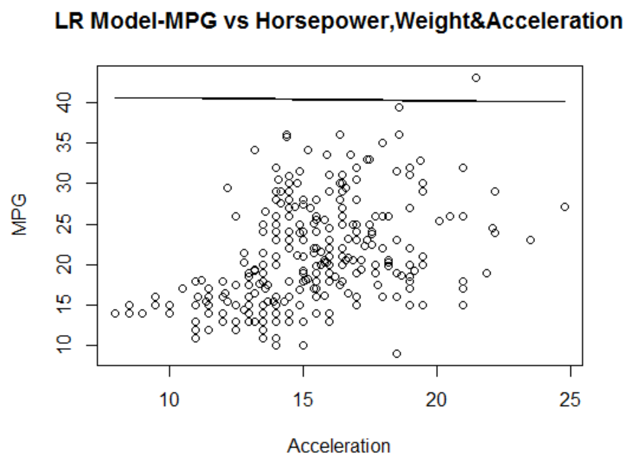
```
coef(model_13)[3]
     weight
-0.005139246
coef(model_13)[4]
acceleration
-0.03382304

plot(data_300$horsepower, data_300$mpg,
    xlab = "Horsepower",
    ylab = "MPG",
    main = "LR Model-MPG vs Horsepower,Weight&Acceleration")
lines(data_300$horsepower, coef(model_13)[1]+coef(model_13)[2]*data_300$horsepower)
```




```
plot(data_300$weight, data_300$mpg,
    xlab = "Weight",
    ylab = "MPG",
    main = "LR Model-MPG vs Horsepower,Weight&Acceleration")
lines(data_300$weight, coef(model_13)[1]+coef(model_13)[3]*data_300$weight)
```

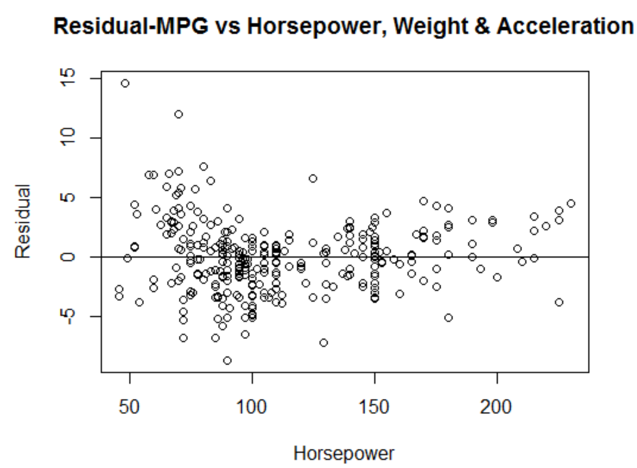


```
plot(data_300$acceleration, data_300$mpg,
    xlab = "Acceleration",
    ylab = "MPG",
    main = "LR Model-MPG vs Horsepower,Weight&Acceleration")
lines(data_300$acceleration, coef(model_13)[1]+coef(model_13)[4]*data_300$acceleration)
```

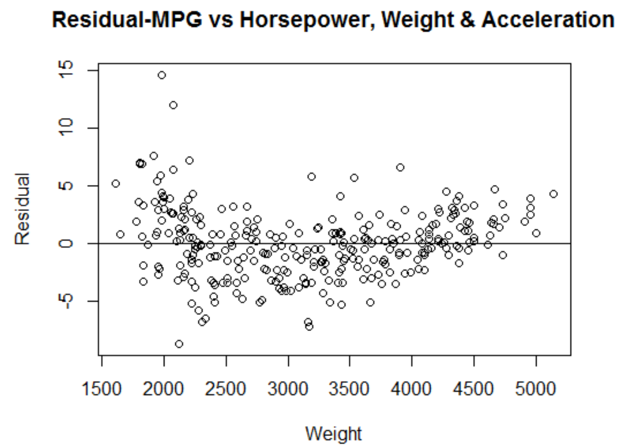


 # Residual Plot

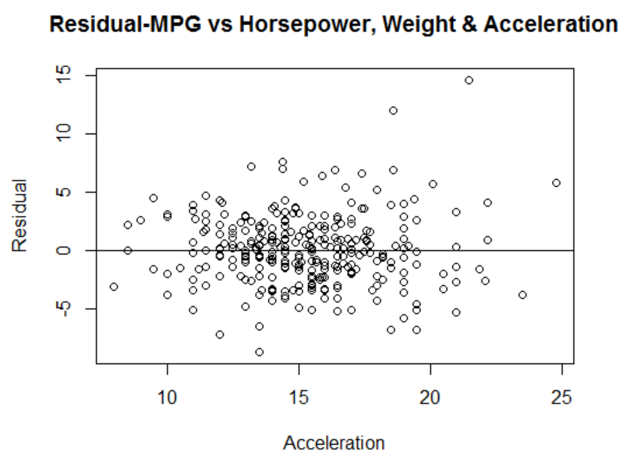
```
res_13 <- residuals(model_13)
plot(data_300$horsepower, res_13,
    xlab = "Horsepower",
    ylab = "Residual",
    main = "Residual-MPG vs Horsepower, Weight & Acceleration")
abline(0,0)
```



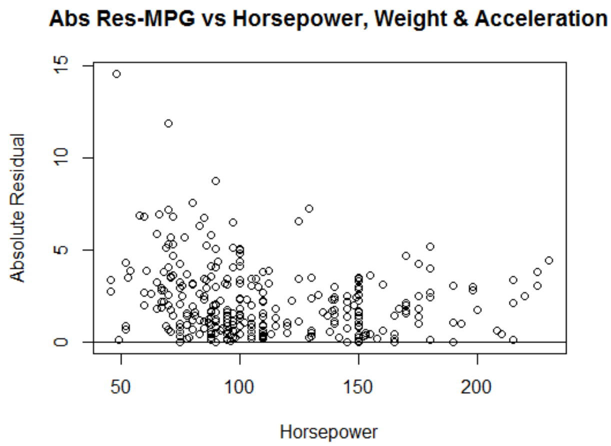
```
plot(data_300$weight, res_13,
    xlab = "Weight",
    ylab = "Residual",
    main = "Residual-MPG vs Horsepower, Weight & Acceleration")
abline(0,0)
```



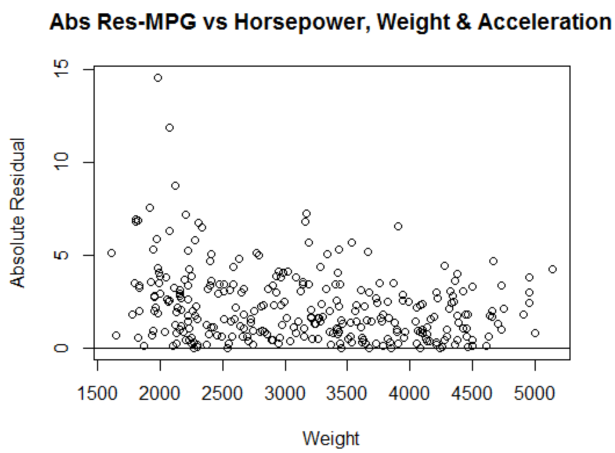

```
plot(data_300$acceleration, res_13,
    xlab = "Acceleration",
    ylab = "Residual",
    main = "Residual-MPG vs Horsepower, Weight & Acceleration")
abline(0,0)
```



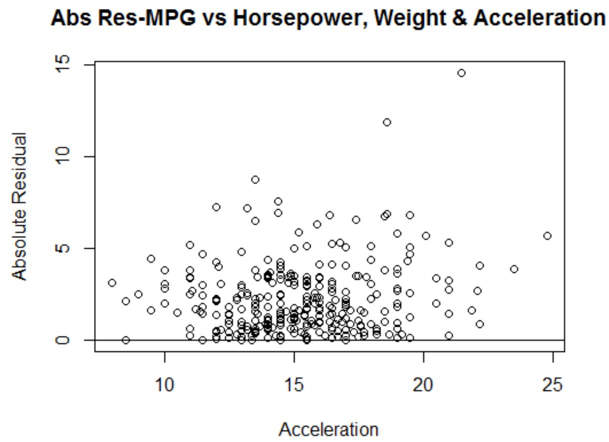

```
# Absolute Residual Plot

a_res_13 <- abs(res_13)
plot(data_300$horsepower, a_res_13,
    xlab = "Horsepower",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Horsepower, Weight & Acceleration")
abline(0,0)
```




```
plot(data_300$weight, a_res_13,
   xlab = "Weight",
   ylab = "Absolute Residual",
   main = "Abs Res-MPG vs Horsepower, Weight & Acceleration")
abline(0,0)
```




```
plot(data_300$acceleration, a_res_13,
   xlab = "Acceleration",
   ylab = "Absolute Residual",
   main = "Abs Res-MPG vs Horsepower, Weight & Acceleration")
abline(0,0)
```

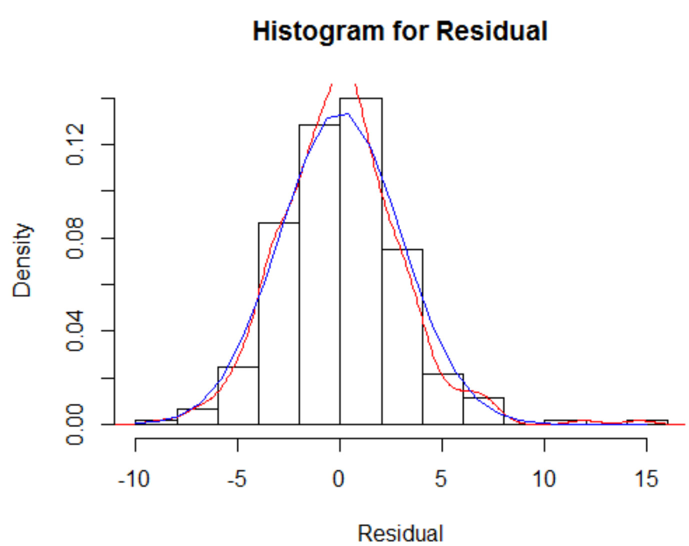


 # Histogram for residual

```
hist(res_13, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_13), col="red")
```

 # Normalizing the curve

```
mu_13 <- mean(res_13)
v_13 <- var(res_13)
sd_13 <- sqrt(v_13)
x_13 <- seq(-10, 15, length=25)
y_13 <- dnorm(x_13,mu_13,sd_13)
lines(x_13,y_13,col="blue")
```



Model 14: MPG vs Displacement, Acceleration and Horsepower

model_14 <- lm(mpg~displacement+horsepower+acceleration, data = data_300)
summary(model_14)

```
> summary(model_14)

Call:
lm(formula = mpg ~ displacement + horsepower + acceleration,
    data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max
-9.8921 -2.0869 -0.3065  1.4988 16.5483

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  41.841346   2.198400  19.033  < 2e-16 ***
displacement -0.033731   0.003935  -8.572 5.75e-16 ***
horsepower   -0.065049   0.012423  -5.236 3.12e-07 ***
acceleration -0.424721   0.099915  -4.251 2.86e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.261 on 296 degrees of freedom
Multiple R-squared:  0.7373,    Adjusted R-squared:  0.7347
F-statistic:   277 on 3 and 296 DF,  p-value: < 2.2e-16
```

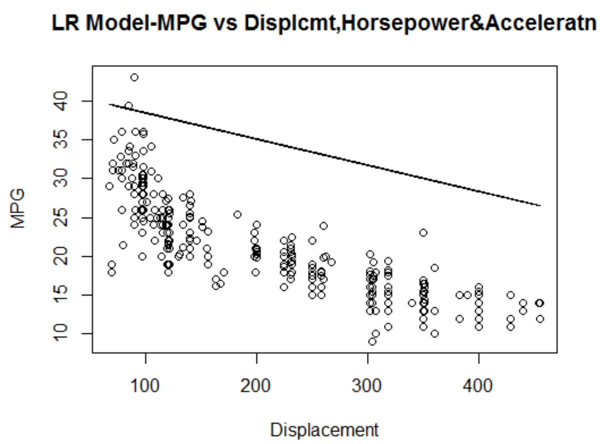
- Value of R^2 is 0.7373 that means our model only explains 73.73% of variance.
- Equation of Regression Line is:

MPG = 41.841346 – 0.033731*displacement – 0.065049*horsepower – 0.424721*acceleration + Ei
Where Ei ~ N(0,3.261)
B0 -> Intercept -> 41.841346
P-value: < 2.2e-16 clearly shows that we should Reject NULL Hypothesis that displacement, horsepower and acceleration has no effect on MPG(Ha is true).
Residual = Observed(Y) – Predicted(Y)

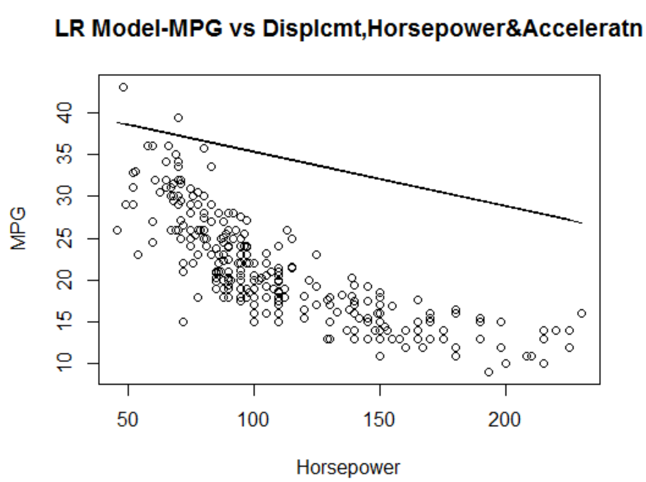
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

coef(model_14)[1]
(Intercept)
  41.84135
coef(model_14)[2]
displacement
-0.03373087

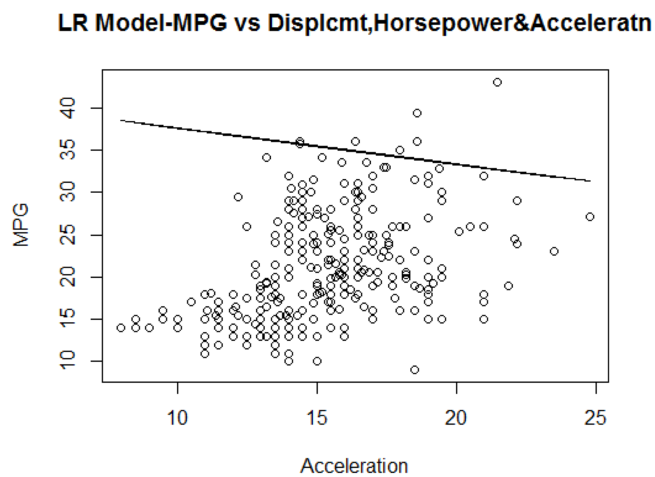
```
coef(model_14)[3]
horsepower
-0.06504852
coef(model_14)[4]
acceleration
 -0.4247212

plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "LR Model-MPG vs Displcmt,Horsepower&Acceleratn")
lines(data_300$displacement, coef(model_14)[1]+coef(model_14)[2]*data_300$displacement)
```




```
plot(data_300$horsepower, data_300$mpg,
    xlab = "Horsepower",
    ylab = "MPG",
    main = "LR Model-MPG vs Displcmt,Horsepower&Acceleratn")
lines(data_300$horsepower, coef(model_14)[1]+coef(model_14)[3]*data_300$horsepower)
```

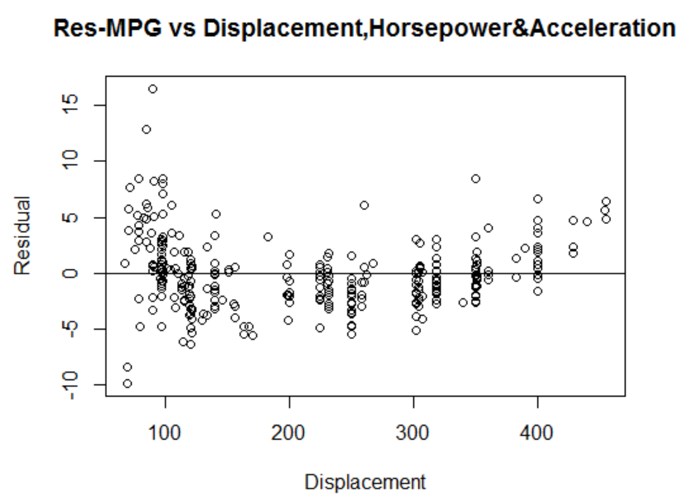


```
plot(data_300$acceleration, data_300$mpg,
    xlab = "Acceleration",
    ylab = "MPG",
    main = "LR Model-MPG vs Displcmt,Horsepower&Acceleratn")
lines(data_300$acceleration, coef(model_14)[1]+coef(model_14)[4]*data_300$acceleration)
```

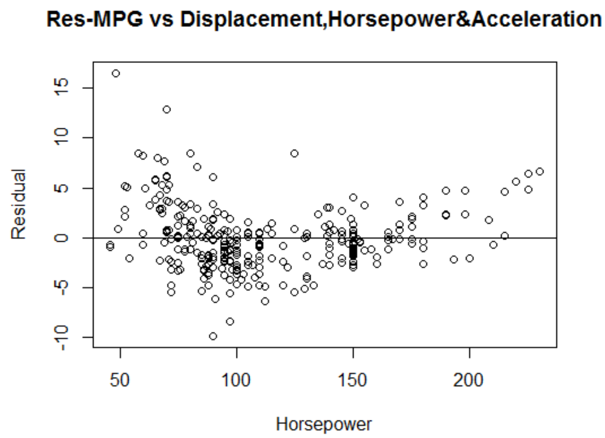


 # Residual Plot

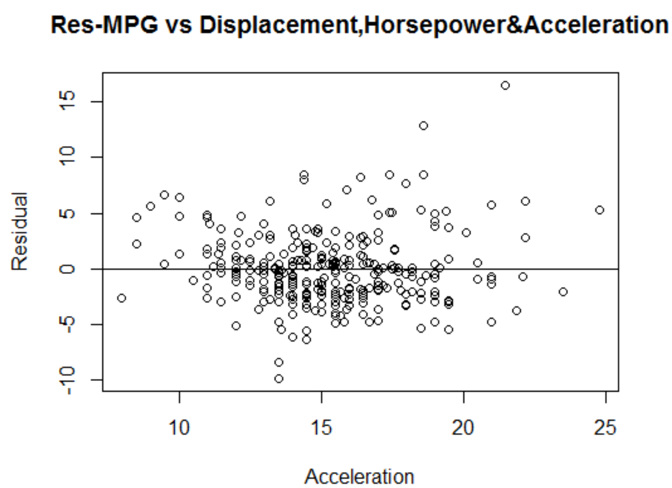
```
res_14 <- residuals(model_14)
plot(data_300$displacement, res_14,
    xlab = "Displacement",
    ylab = "Residual",
    main = "Res-MPG vs Displacement,Horsepower&Acceleration")
abline(0,0)
```



```
plot(data_300$horsepower, res_14,
    xlab = "Horsepower",
    ylab = "Residual",
    main = "Res-MPG vs Displacement,Horsepower&Acceleration")
abline(0,0)
```

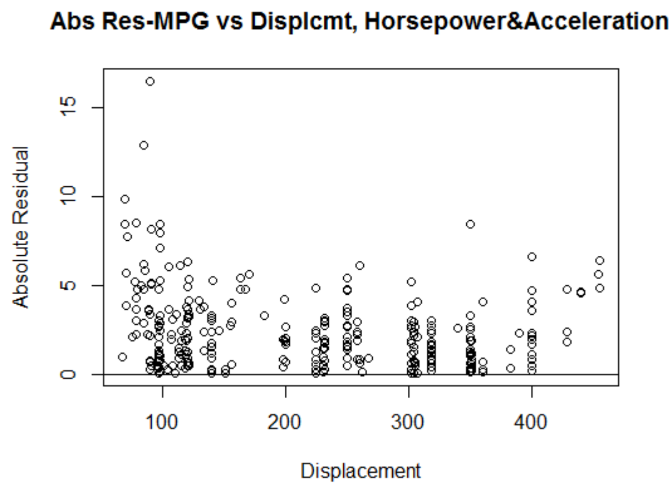



```
plot(data_300$acceleration, res_14,
    xlab = "Acceleration",
    ylab = "Residual",
    main = "Res-MPG vs Displacement,Horsepower&Acceleration")
abline(0,0)
```

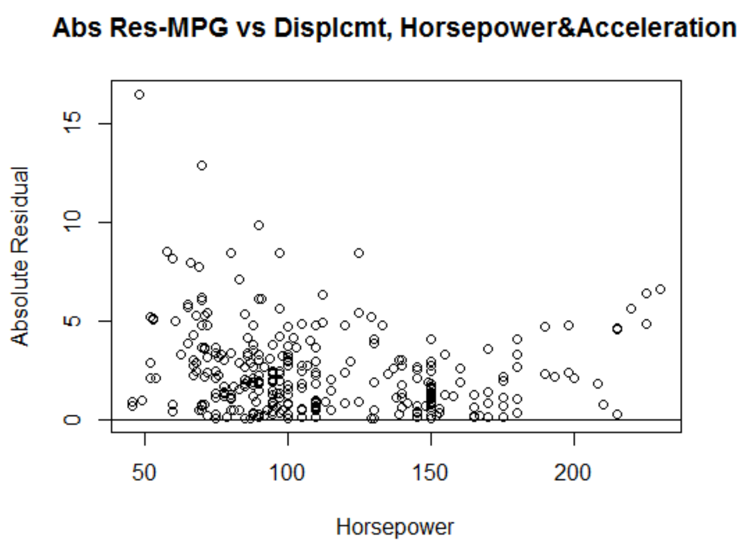


Absolute Residual Plot

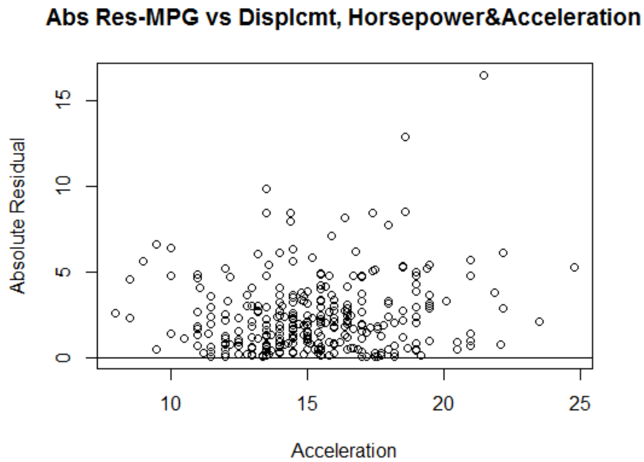
```
a_res_14 <- abs(res_14)
plot(data_300$displacement, a_res_14,
    xlab = "Displacement",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displcmt, Horsepower&Acceleration")
abline(0,0)
```




```
plot(data_300$horsepower, a_res_14,
    xlab = "Horsepower",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displcmt, Horsepower&Acceleration")
abline(0,0)
```



```
plot(data_300$acceleration, a_res_14,
    xlab = "Acceleration",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displcmt, Horsepower&Acceleration")
abline(0,0)
```

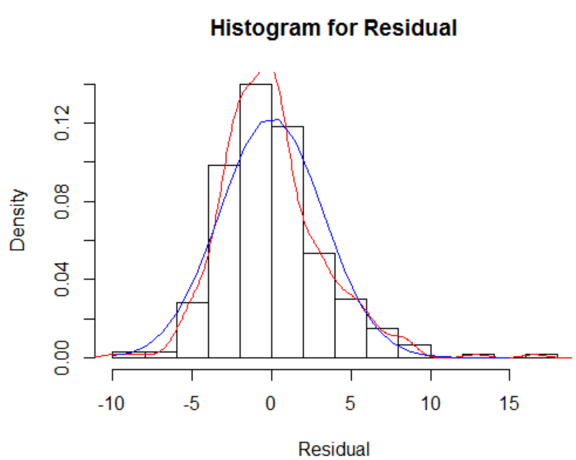


 # Histogram for residual

```
hist(res_14, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_14), col="red")
```

 # Normalizing the curve

```
mu_14 <- mean(res_14)
v_14 <- var(res_14)
sd_14 <- sqrt(v_14)
x_14 <- seq(-10, 15, length=25)
y_14 <- dnorm(x_14,mu_14,sd_14)
lines(x_14,y_14,col="blue")
```



Model 15: MPG vs Horsepower, Weight, Acceleration and Displacement

model_15 <- lm(mpg~displacement+horsepower+weight+acceleration, data = data_300)
summary(model_15)

```
> summary(model_15)

Call:
lm(formula = mpg ~ displacement + horsepower + weight + acceleration,
    data = data_300)

Residuals:
    Min      1Q  Median      3Q     Max 
-9.0802 -1.8601 -0.0355  1.5691 14.8329 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  40.5851720  2.0191187  20.100  < 2e-16 ***
displacement -0.0058876  0.0051269  -1.148   0.2517    
horsepower   -0.0270124  0.0124165  -2.176   0.0304 *  
weight       -0.0046422  0.0006083  -7.632 3.22e-13 ***
acceleration -0.0593869  0.1032312  -0.575   0.5655    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.985 on 295 degrees of freedom
Multiple R-squared:  0.7806,	Adjusted R-squared:  0.7777 
F-statistic: 262.4 on 4 and 295 DF,  p-value: < 2.2e-16
```

- Value of R^2 is 0.7806 that means our model only explains 78.06% of variance.
- Equation of Regression Line is:

MPG = 40.5851720 – 0.0058876*displacement – 0.0270124*horsepower – 0.0046422*weight – 0.0593869*acceleration + Ei
Where Ei ~ N(0,2.985)
B0 -> Intercept -> 40.5851720

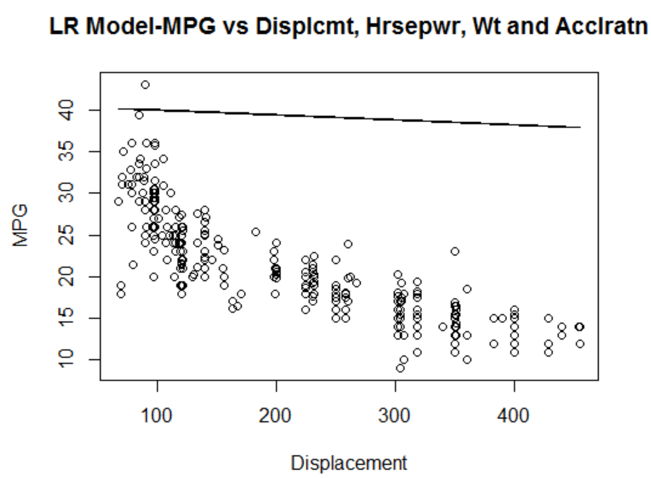
P-value: 3.22e-13 clearly shows that we should Reject NULL Hypothesis that weight has no effect on MPG.Weight has major impact on MPG(Ha is true).
<u>P values: 0.2517,0.0304,0.5655</u> are greater than significance level,hence test is not significant for displacement,horsepower & acceleration.We can say that displacement,horsepower & acceleration has no impact on MPG.

Residual = Observed(Y) – Predicted(Y)

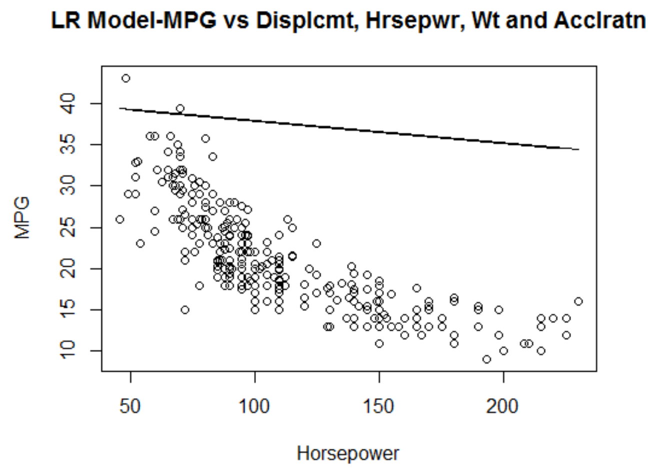
- Positive values for the residual (on the y axis) mean the prediction was too low.
- Negative values means the prediction was too high.
- Zero means the guess was exactly correct.
- Histogram of MPG Residual seems to be normal.

coef(model_15)[1]
(Intercept)
  40.58517

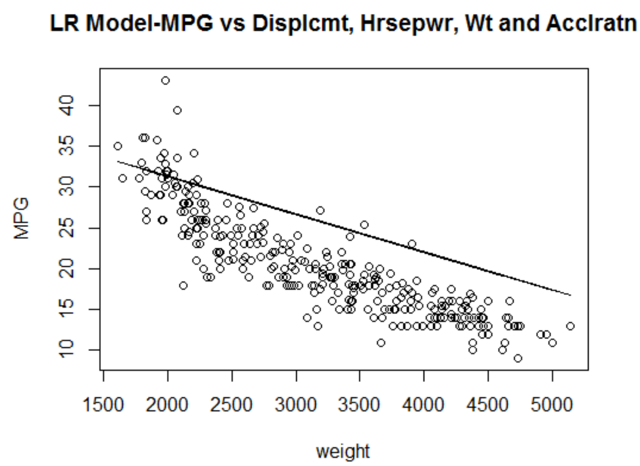
```
coef(model_15)[2]
displacement
-0.005887596
coef(model_15)[3]
horsepower
-0.02701239
coef(model_15)[4]
     weight
-0.004642193
coef(model_15)[5]
acceleration
-0.05938692

plot(data_300$displacement, data_300$mpg,
    xlab = "Displacement",
    ylab = "MPG",
    main = "LR Model-MPG vs Displcmt, Hrsepwr, Wt and Acclratn")
lines(data_300$displacement, coef(model_15)[1]+coef(model_15)[2]*data_300$displacement)
```



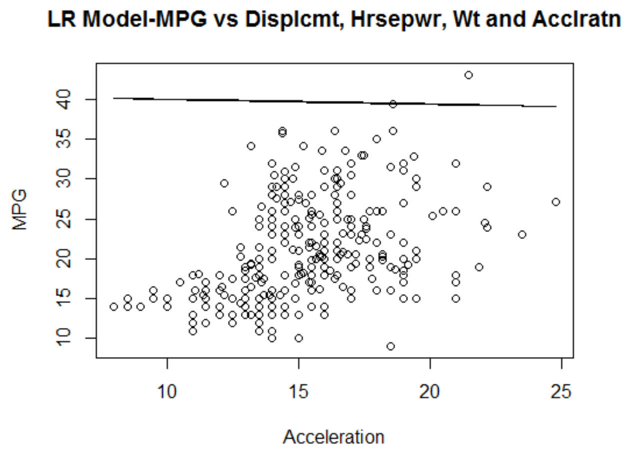

```
plot(data_300$horsepower, data_300$mpg,
    xlab = "Horsepower",
    ylab = "MPG",
    main = "LR Model-MPG vs Displcmt, Hrsepwr, Wt and Acclratn")
lines(data_300$horsepower, coef(model_15)[1]+coef(model_15)[3]*data_300$horsepower)
```




```
plot(data_300$weight, data_300$mpg,
   xlab = "weight",
   ylab = "MPG",
   main = "LR Model-MPG vs Displcmt, Hrsepwr, Wt and Acclratn")
lines(data_300$weight, coef(model_15)[1]+coef(model_15)[4]*data_300$weight)
```

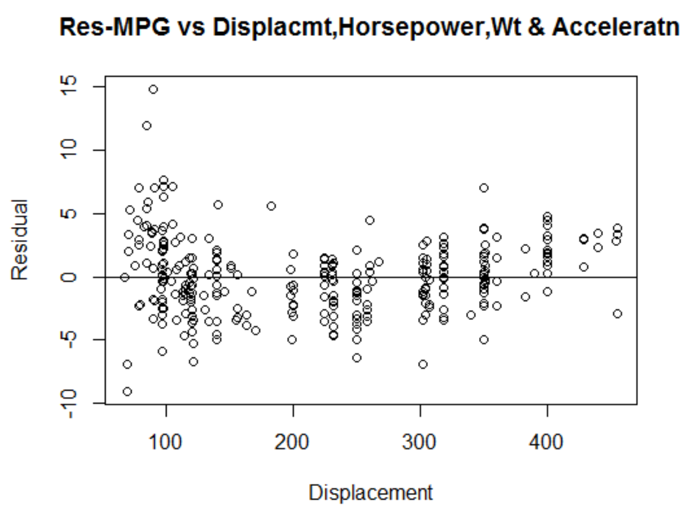



```
plot(data_300$acceleration, data_300$mpg,
   xlab = "Acceleration",
   ylab = "MPG",
   main = "LR Model-MPG vs Displcmt, Hrsepwr, Wt and Acclratn")
lines(data_300$acceleration, coef(model_15)[1]+coef(model_15)[5]*data_300$acceleration)
```

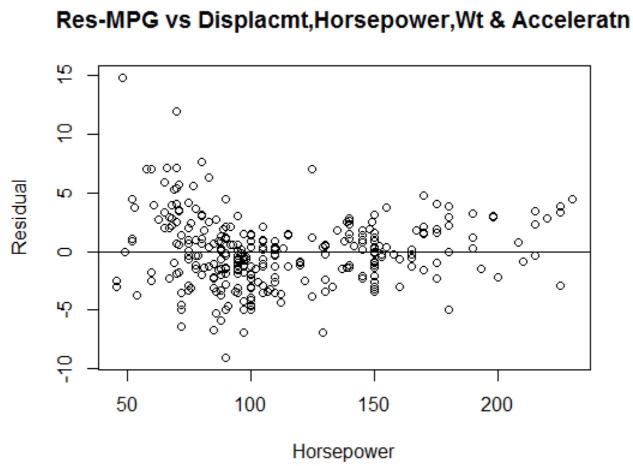


Residual Plot

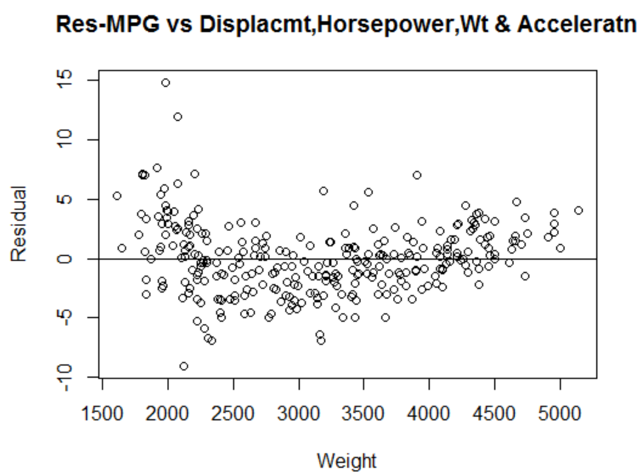
```
res_15 <- residuals(model_15)
plot(data_300$displacement, res_15,
    xlab = "Displacement",
    ylab = "Residual",
    main = "Res-MPG vs Displacmt,Horsepower,Wt & Acceleratn")
abline(0,0)
```



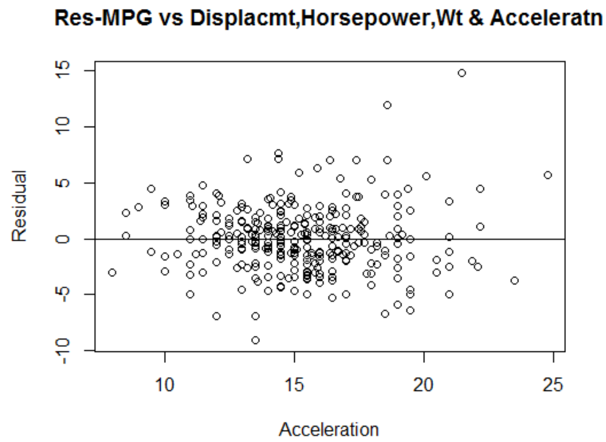

```
plot(data_300$horsepower, res_15,
    xlab = "Horsepower",
    ylab = "Residual",
    main = "Res-MPG vs Displacmt,Horsepower,Wt & Acceleratn")
abline(0,0)
```




```
plot(data_300$weight, res_15,
    xlab = "Weight",
    ylab = "Residual",
    main = "Res-MPG vs Displacmt,Horsepower,Wt & Acceleratn")
abline(0,0)
```

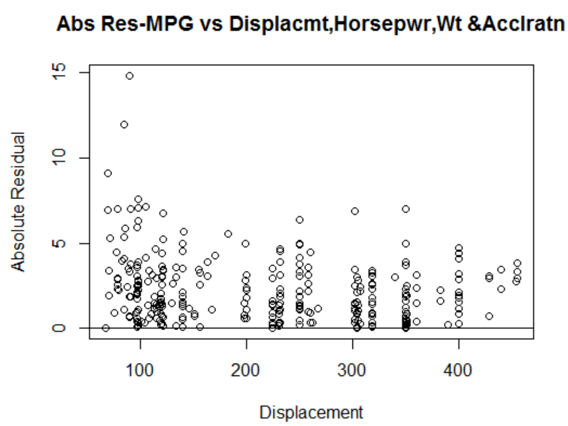



```
plot(data_300$acceleration, res_15,
    xlab = "Acceleration",
    ylab = "Residual",
    main = "Res-MPG vs Displacmt,Horsepower,Wt & Acceleratn")
abline(0,0)
```

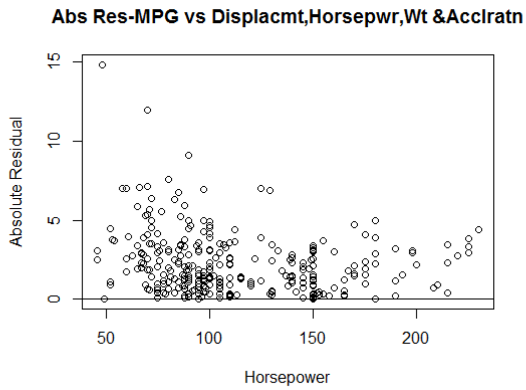


Absolute Residual Plot

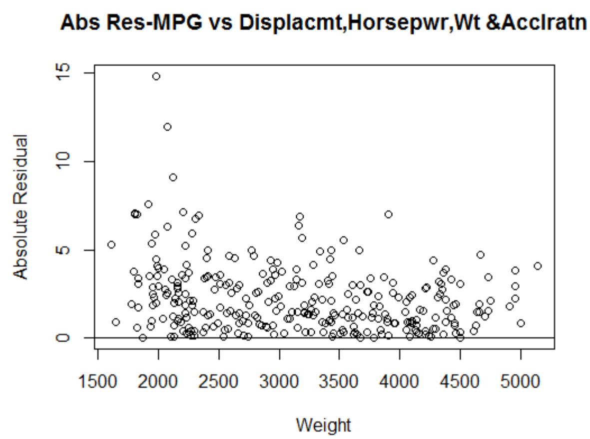
```
a_res_15 <- abs(res_15)
plot(data_300$displacement, a_res_15,
    xlab = "Displacement",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacmt,Horsepwr,Wt &Acclratn")
abline(0,0)
```



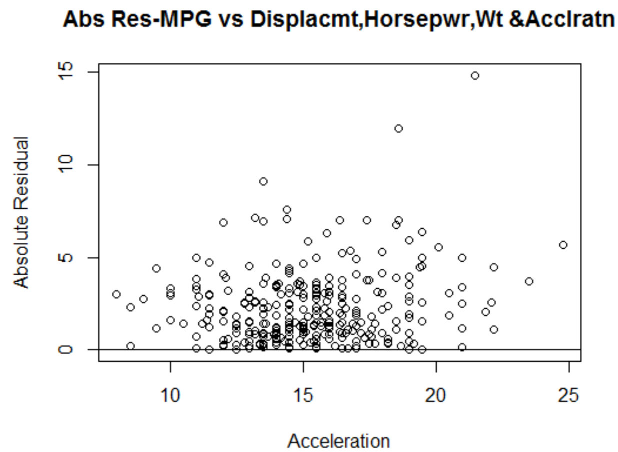

```
plot(data_300$horsepower, a_res_15,
    xlab = "Horsepower",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacmt,Horsepwr,Wt &Acclratn")
abline(0,0)
```




```
plot(data_300$weight, a_res_15,
    xlab = "Weight",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacmt,Horsepwr,Wt &Acclratn")
abline(0,0)
```




```
plot(data_300$acceleration, a_res_15,
    xlab = "Acceleration",
    ylab = "Absolute Residual",
    main = "Abs Res-MPG vs Displacmt,Horsepwr,Wt &Acclratn")
abline(0,0)
```

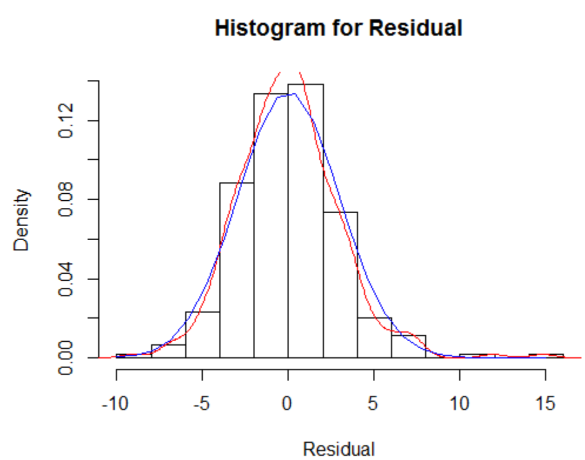


 # Histogram for residual

```
hist(res_15, prob=T, breaks = 10, main = "Histogram for Residual",
    xlab = "Residual",
    ylab = "Density")
lines(density(res_15), col="red")
```

 # Normalizing the curve

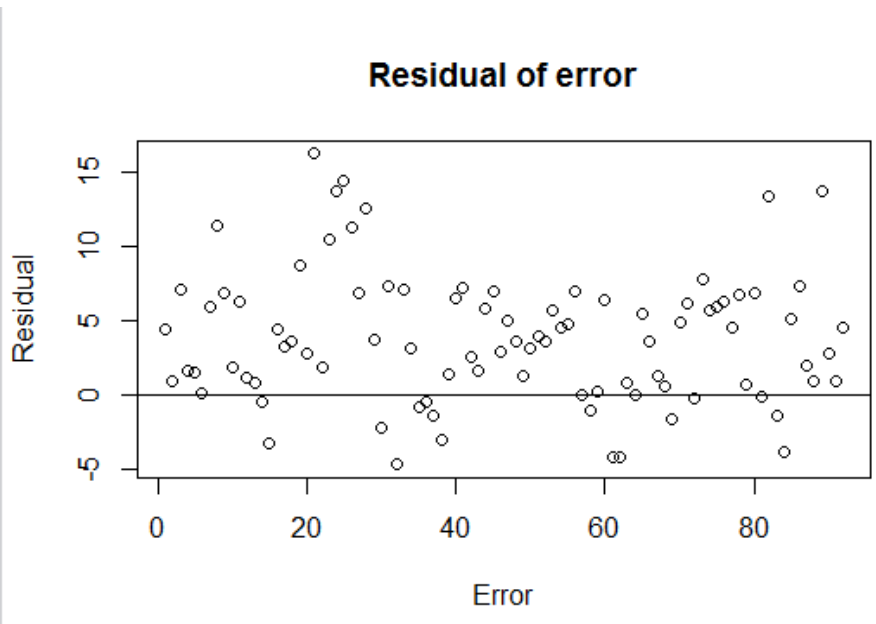
```
mu_15 <- mean(res_15)
v_15 <- var(res_15)
sd_15 <- sqrt(v_15)
x_15 <- seq(-10, 15, length=25)
y_15 <- dnorm(x_15,mu_15,sd_15)
lines(x_15,y_15,col="blue")
```



Predicting MPG for 92 data

```
# Predicting Values of MPG vs Weight.
predicted1 <- coef(model_1)[1]+coef(model_1)[2]*data_92$weight
View(predicted1)

# Extracting MPG data of last 92 rows
data_92_mpg <- data_92[,1]
View(data_92_mpg)

# Calculating difference between oberved and predicted values
error1 <- data_92_mpg[] - predicted1[]
View(error1)

# Plotting all errors
plot(error1,
    xlab = "Error",
    ylab = "Residual",
    main = "Residual of error"
    )
abline(0,0)

summary(error1)
```

```
> summary(error1)
    Min.   1st Qu.   Median   Mean     3rd Qu.   Max.
-4.6710   0.8429   3.6370   3.8460   6.4650   16.2800
```



 # Histogram for Error

```
hist(error1, prob=T, breaks = 10, main = "Histogram for Residual",
     xlab = " Error Residual",
     ylab = "Density")
lines(density(error1), col="red")
```

 # Normalizing the curve

```
mu_e <- mean(error1)
v_e <- var(error1)
sd_e <- sqrt(v_e)
x_e <- seq(-5, 15, length=20)
y_e <- dnorm(x_e,mu_e,sd_e)
lines(x_e, y_e, col="blue")
```