AutoVsManual

Durga Gaddam

Executive Summary:

Objective: The Main Aim of this project is (i) To determine whether automatic transmission gives more miles per gallon or manual transmission gives

(ii) Analyze different variantions of mpg and other factors

Observations: The data contains different factors that can be considered for a vehicle. In this data, few models with auto transmission and few models with manual transmission are given. It is observed that mean of manual mpg is 7.25 greater than mean of auto transmission.

Solution:

Using different Simple Linear Regression, multi linear regression and different concepts like ANOVA, Hypothesis Testing, and Cook's distance we can solve this problem. After analysis using these concepts we conclude that there is no much difference between automatic and manual transmissions, without considering other factors.

library(plotly) library(calibrate) library(datasets) library(ggplot2)

```
data=mtcars
names(mtcars)
```

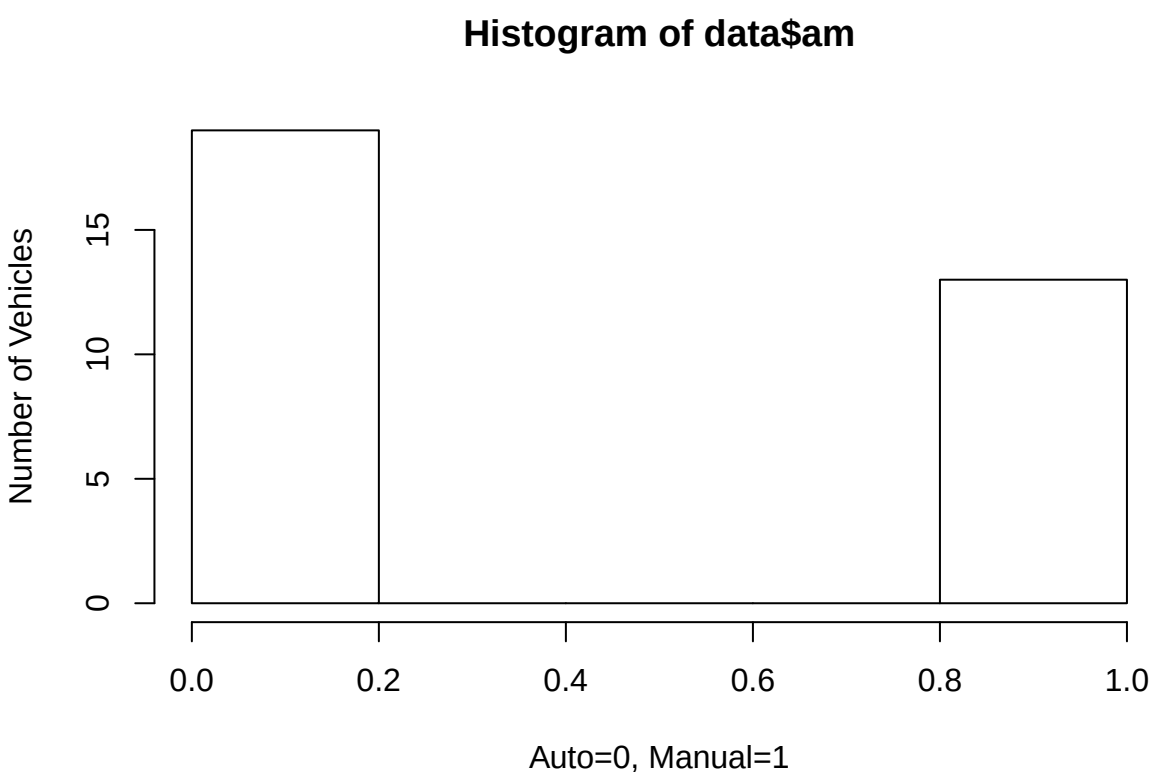
```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

The data is a collection from car magzine with different models. Each model has a set of data described using the mile per gallon(mpg), number of cylinders(cyl), displacement(disp), horse power(hp), rear axel ration (drat) , weight(wt), time taken to cover quater of mile (qsec) , velocity per substrate concentration (vs) , automatic=0 or manual=1 (am) , transmissions(gear), and Number of carberators (carb)

```
table(data$am)
```

```
##
##  0  1
## 19 13
```

```
hist(data$am, xlab= "Auto=0, Manual=1", ylab= "Number of Vehicles")
```



There are 19 automatic and 13 manual transmission vehicles.

To find the answer for the question whether automatic transmission vehicles give more mileage than manual transmission vehicles, we need to find the relation between mpg and am, and what other factors influence the mileage of a given model.

```
cor(data$mpg,data$cyl)
```

```
## [1] -0.852162
```

```
## We can infer that there is a negative correlation between mpg and cyl. So as the number of cylinders

cor(data$mpg,data$wt)
```

```
## [1] -0.8676594
```

```
## We can infer that there is a negative correlation between mpg and wt. So as the weight of a car incr
```

```
## Mean Mileage given by Automatic and manual transmissions
aggregate(mpg~am,data,mean)
```

```
##   am      mpg
## 1  0 17.14737
## 2  1 24.39231
```

We can clearly infer that manual transmissions give 7.25 mileage more than auto transmission vehicles. However, We cannot conclude by just finding the mean of two variable without considering other factors that effect the mileage.

Hypothesis Testing

Null Hypothesis: H0 : Other factors of cyl, wt, hp, and disp does not affect the mileage, or linear models of all the parameters is same.

Simple Linear Regression

```
Fac1 <-  lm(mpg~am, data)
summary(Fac1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Understanding the coefficients

We can infer that Adj r-square= 33.8% which explains the variance(relation) between mpg and am

Multi Linear Regression

```
Fac2 <- lm(mpg~am+wt+hp,data)
Fac3 <- lm(mpg~am+wt+hp+cyl,data)
Fac4 <- lm(mpg~am+wt+hp+cyl+disp,data)

anova(Fac1,Fac2,Fac3,Fac4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + hp
## Model 3: mpg ~ am + wt + hp + cyl
## Model 4: mpg ~ am + wt + hp + cyl + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
```

```
## 2     28 180.29  2    540.61 43.0841 5.576e-09 ***
## 3     27 170.00  1     10.29  1.6407    0.2115
## 4     26 163.12  1      6.88  1.0963    0.3047
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(Fac4)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp + cyl + disp, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5952 -1.5864 -0.7157  1.2821  5.5725
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.20280    3.66910  10.412 9.08e-11 ***
## am           1.55649    1.44054   1.080  0.28984
## wt          -3.30262    1.13364  -2.913  0.00726 **
## hp          -0.02796    0.01392  -2.008  0.05510 .
## cyl         -1.10638    0.67636  -1.636  0.11393
## disp         0.01226    0.01171   1.047  0.30472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.505 on 26 degrees of freedom
## Multiple R-squared:  0.8551, Adjusted R-squared:  0.8273
## F-statistic:  30.7 on 5 and 26 DF,  p-value: 4.029e-10
```

When considering other factors, Adj R-square= 82.7%. The model Fac4 explains 82.7% of the variance.

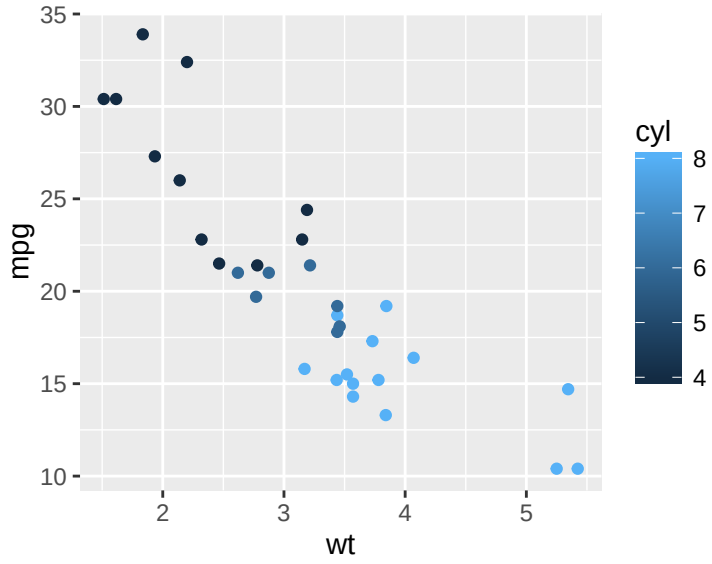
we can infer that p-value is 5.579 x 10^9, which is less than 0.05 (assumed alpha). "If p is low null will blow", we reject the null hypothesis. Hence other factors of a car does effect the mileage of a car.

Graph between weight and MPG

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.1
```

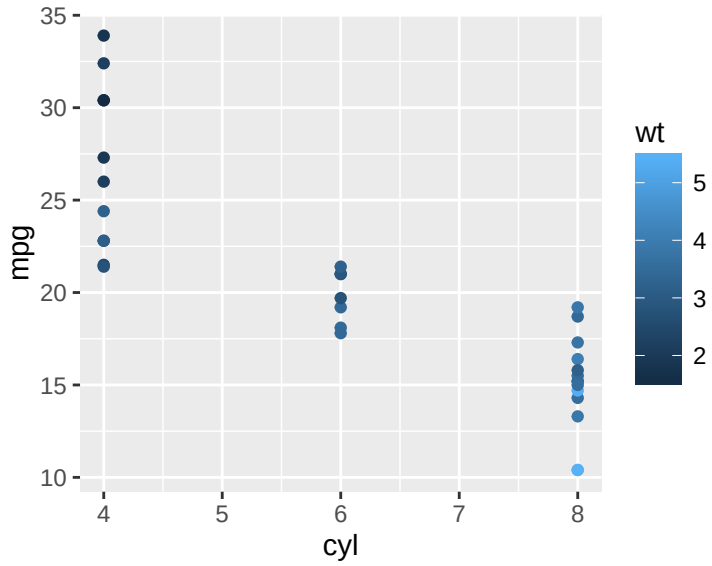
```
z <- ggplot(data=mtcars, aes(x=wt, y=mpg, color=cyl))
z <- z + geom_point()
z
```




```
### we can see that heaviest car has the least mileage
```

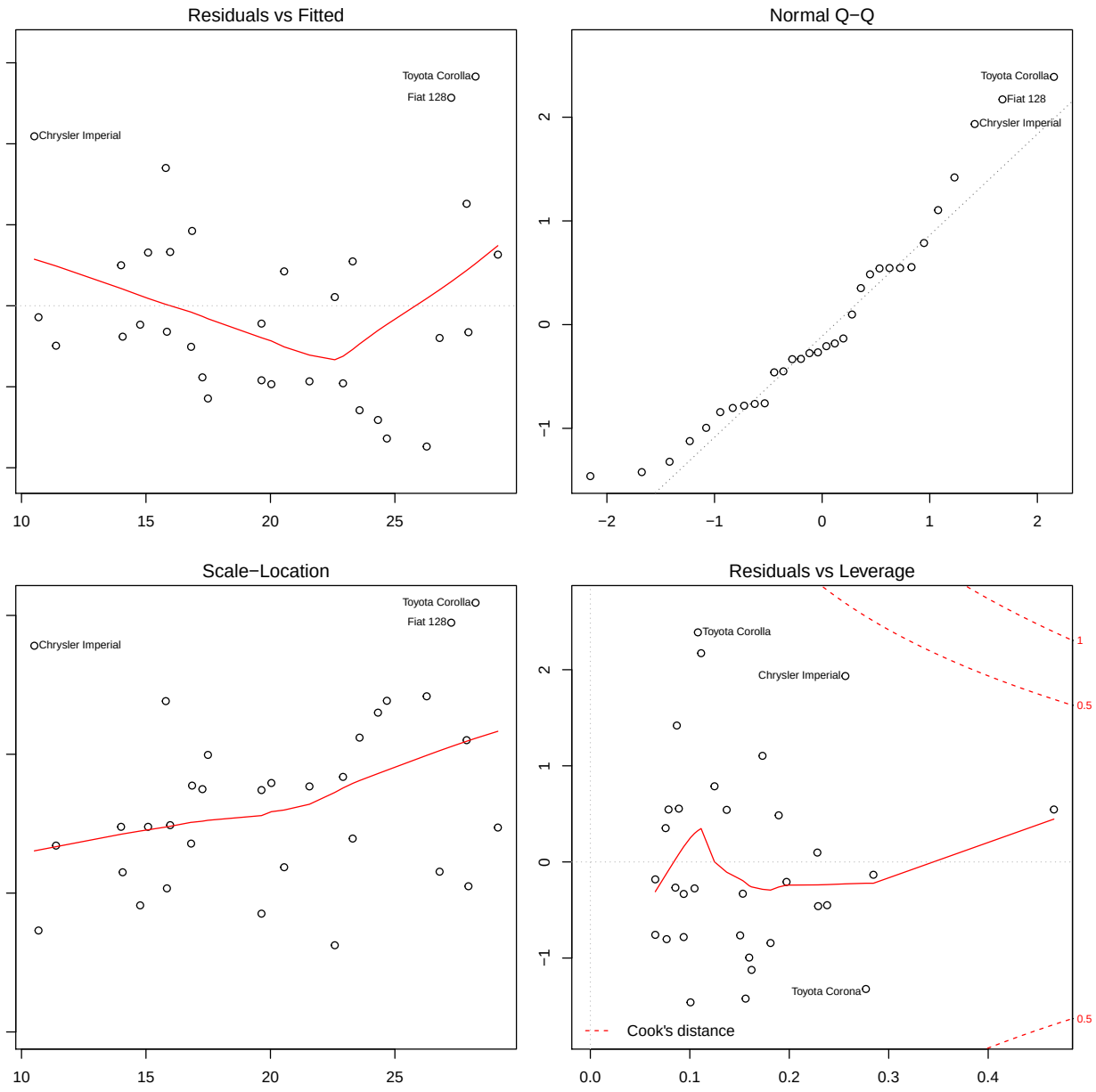
Graph between Cyl and MPG

```
g<-ggplot(mtcars, aes(x=cyl, y=mpg, color= wt)) + geom_point()
g
```

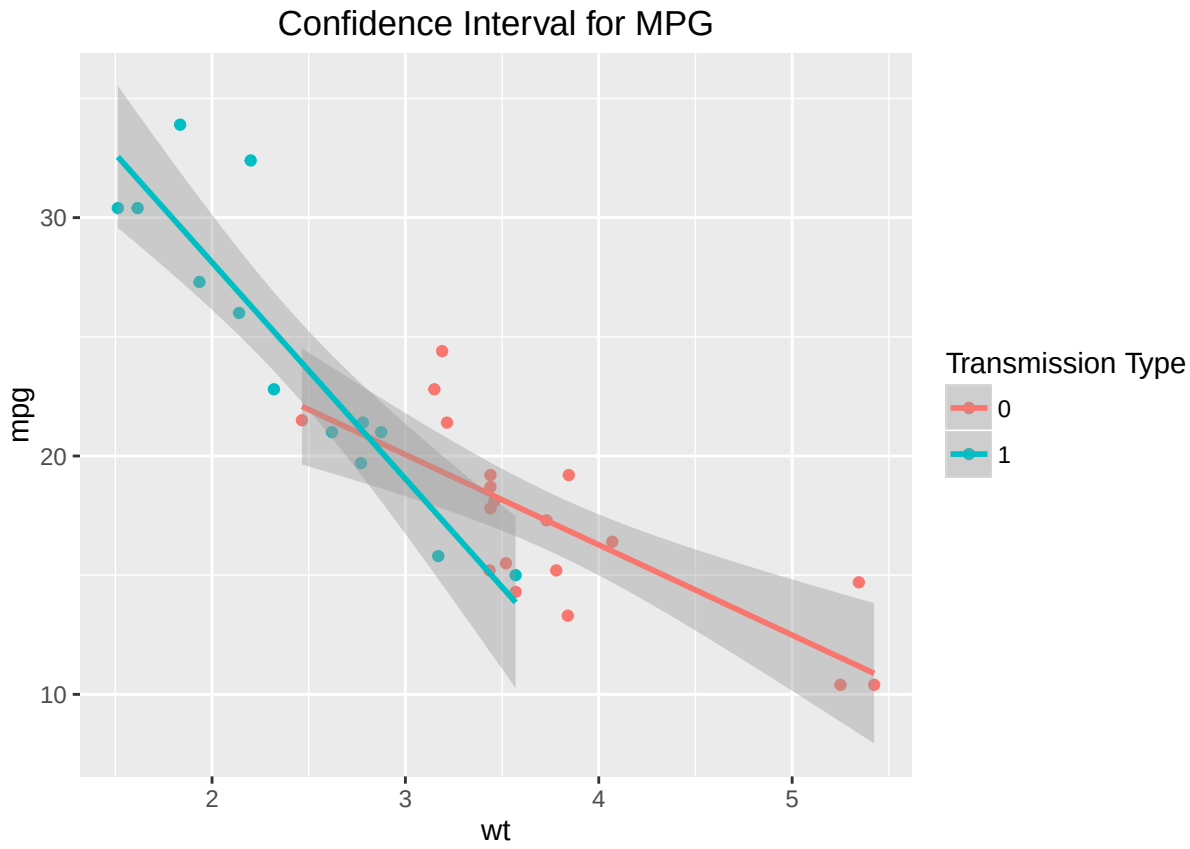


Understanding the Residuals

```
ResidualGraphs <- lm(mpg ~ am + wt + hp + cyl  , data)
par(mfrow = c(2,2), mar = c(3,1,2,2))
plot(ResidualGraphs)
```




```
h <- ggplot( data, aes(wt, mpg, color= factor(am)))
h <- h + geom_point()
h <- h+stat_smooth(method = lm)  + labs(x = "wt", y = "mpg", title = "Confidence Interval for MPG") + s
h
```



Conclusion:

Due to the overlapping of Confidence intervals of automatic transmissions and manual transmissions we can conclude that there is no much difference between mpg of Automatic and manual transmission, without considering other factors.