Course_Project_7

YYC

May 28, 2017

Executive Summary

1) From my analysis, the manual transmission is better for MPG.
2) If we didn't consider other varibles and just model the am with linear regression, the cars with manual transmissions have on average 7.224 miles higher MPG than cars with automatic transmissions. If we consider the influence of major variables cyl, disp, hp and wt, the cars with manual transmissions have on average 1.556 miles higher MPG than cars with automatic transmissions

Data Exploration

```
# load the dataset and know the data structure and first 6 columns
library(datasets)
data("mtcars")
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Data Visualization

```
# plot the pairwise plot
pairs(mpg~.,data=mtcars)
```

See Figure 1

```
# plot the mpg~am
plot(as.factor(mtcars$am),mtcars$mpg,col=c("blue","red"),xlab="0-
automatic transmission, 1-manul transmission", ylab="Miles per gallon
(MPG)", main="Transmission type vs mpg")
```

See Figure 2

Statistic Inference_Hypothesis Test

Null hypothesis: The transmission type has no significant influence on the mpg.

```
 t.test(mtcars$mpg[mtcars$am==0],mtcars$mpg[mtcars$am==1])

##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg[mtcars$am == 0] and mtcars$mpg[mtcars$am == 1]
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

Conclusion: since the p-value=0.001374<0.05, we can reject the null hypothesis and conclude that there is a significance different between the two transmission types on the mpg.

Regression Analysis

```
# simple linear regression
fit1=lm(mtcars$mpg~mtcars$am)
summary(fit1)

##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$am)
##
## 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 ***
## mtcars$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
```

Conclusion: The estimate mean for automatic type is 17.1 MPG while that for manual is 24.392. But the fit1 can only explain ~36% of the variance, so other variables also should be involved and the multivariate linear regression model should be employed.

```
cor(mtcars)[1,]
```

```
##        mpg        cyl       disp         hp       drat         wt
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594
##       qsec         vs         am       gear       carb
##  0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
```

Conclusion: The variables cyl, disp, hp and wt can influence the model significantly.

```
# multivariate linear regression
fit2=lm(mpg~am+cyl+disp+hp+wt,data=mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + hp + wt, data = mtcars)
##
## 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
## cyl         -1.10638    0.67636  -1.636  0.11393
## disp         0.01226    0.01171   1.047  0.30472
## hp          -0.02796    0.01392  -2.008  0.05510 .
## wt          -3.30262    1.13364  -2.913  0.00726 **
## ---
## 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
```

Conclusion: The new model could explain ~86% of the variance. It could be concluded that the estimate mean for automatic type is 38.2 MPG while that for manual is 1.56 MPG higher than that of automatic type.

Analysis of the Residuals

```
par(mfrow=c(2,2))
plot(fit2)
```

See Figure 3 Conclusion: 1) The residuals "bounce randomly" around the 0 line. This suggests that the assumption that the relationship is linear is reasonable. 2) The residuals are normally distributed. 3) Scale-Lacation plot shows the residuals are spread equally along the ranges of predictors, indicating the assupmtion of equal variance. 4) There seems to be no influential case considering all the cases are well inside of the Cook's distance lines.

Appendix of Figures

Figure 1

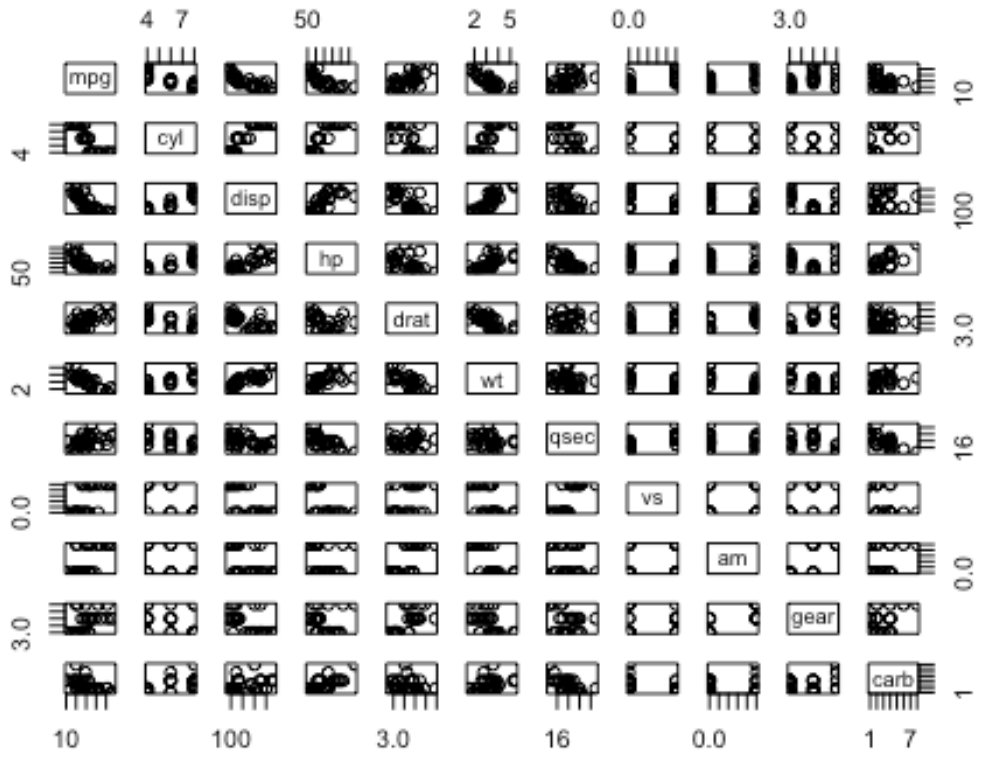


Figure 2

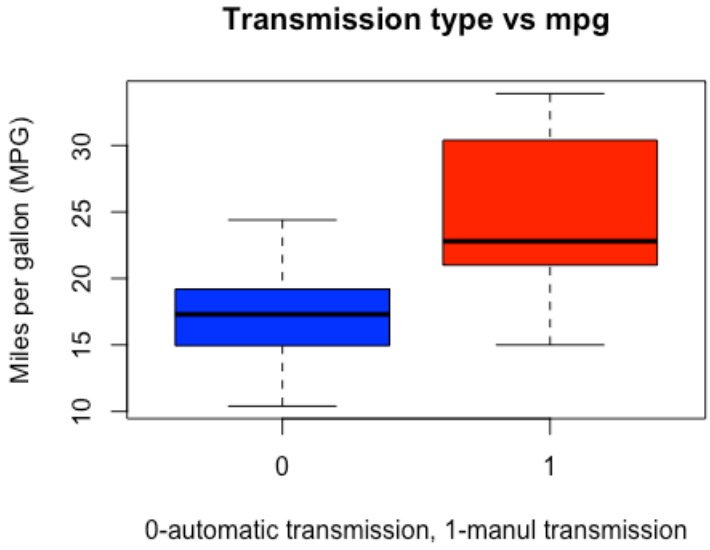
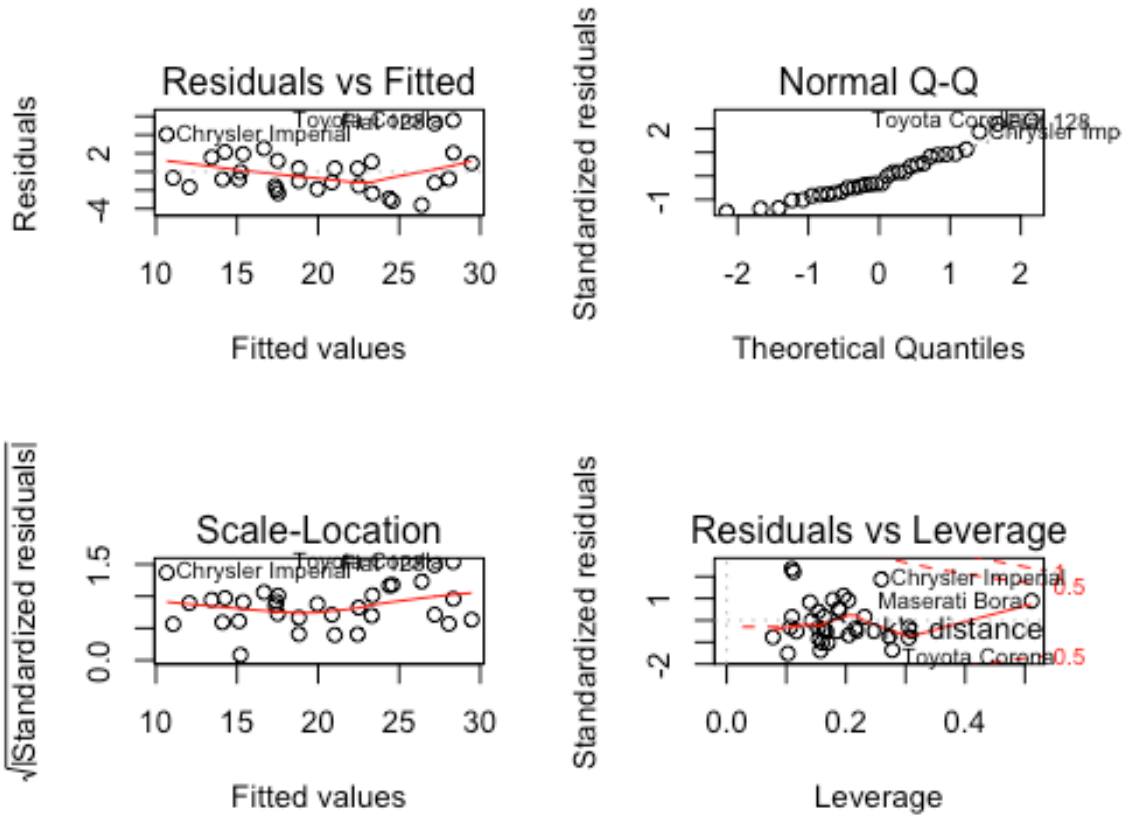

Transmission type vs mpg
Miles per gallon (MPG)
10
15
20
25
30
0
1
0-automatic transmission, 1-manul transmission

Figure 3


Residuals vs Fitted
Residuals
Chrysler Imperial
Fitted values
Normal Q-Q
Standardized residuals
Toyota Corolla
Chrysler Imp
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Chrysler Imperial
Fitted values
Residuals vs Leverage
Standardized residuals
Chrysler Imperial
Maserati Bora
Cook's distance
Toyota Corona
Leverage