Analysis of mtcars data

Aijing Gao

2015.12.26

1.Basic understanding of data

```
library(ggplot2)
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 ...
```

In conclusion, there are 32 observations on 11 variables [, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (1000 lbs) [, 7] qsec 1/4 mile time [, 8] vs V/S [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears [,11] carb Number of carburetors

2. To explore which transmission (automatic or manual) is better for MPG, we perform t test under the assumption that the distribution of MPG is normally distributed. Hypothesis testing: null hypothesis: there is no difference in MPG when using different transmission methods. Alternative hypothesis: there is a difference in MPG when using different transmission methods.

```
mtcars$am<-as.factor(mtcars$am);levels(mtcars$am)<-c("at","mt")
#Compare MPG with two different transmission
mpg_at<- mtcars[mtcars$am == "at",]$mpg
mpg_mt <- mtcars[mtcars$am == "mt",]$mpg
t.test(mpg_at, mpg_mt)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg_at and mpg_mt
## 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
```

As the result indicates, p value=0.001374<0.05, thus we can reject the null hypothesis, there is significant difference in MPG when using different transmission method. Mean of MPG when using automatic transmission is 17.14737, while mean of MPG when using manual transmission is 24.39231. Thus, in conclusion, the manual transmission is better for MPG than automatic transmission (Figure 1).

3. To quantify the MPG difference between automatic and manual transmissions, we perform multivariate linear Regression.

```
fit<-lm(mpg~.,data=mtcars)
summary(fit)

##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## ammt         2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

As we can see that inclusion of all variables doesn't give us a good fitting. So here we choose a best model through step function.

```
best<-step(fit,trace=0)
summary(best)

##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ammt          2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The best fitting model includes three varibles: weight, 1/4 mile time and transmission. The coefficient for transmission is 2.9358, thus on average, manual transmission is 2.9 mpg better than automatic transmission (Residual plot is attached as Fig 2).

4. Conclusion

   the manual transmission is better for MPG than automatic transmission, and manual transmission is 2.9 mpg better than automatic transmission when including weight, 1/4 mile time and transmission to generate the best regression model for MPG.

5. Appendix Figure 1. Automatic versus Manual Transmission MPG.

```
ggplot(mtcars, aes(y=mpg, x=factor(am, labels = c("automatic", "manual
")), fill=factor(am)))+geom_boxplot()+xlab("transmission") + ylab("MPG")
```

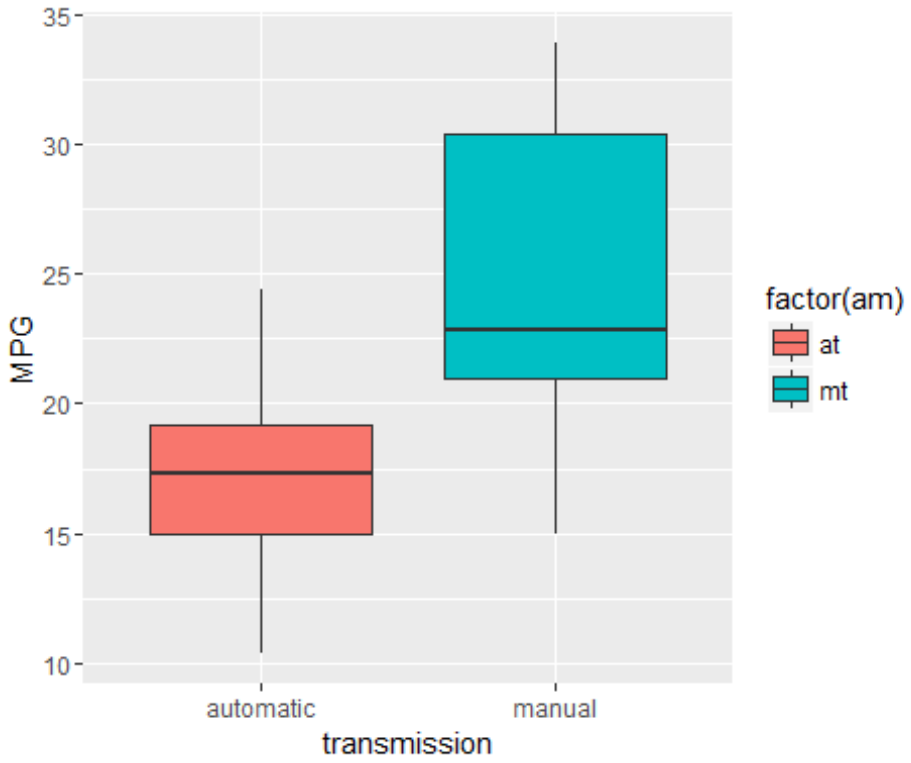
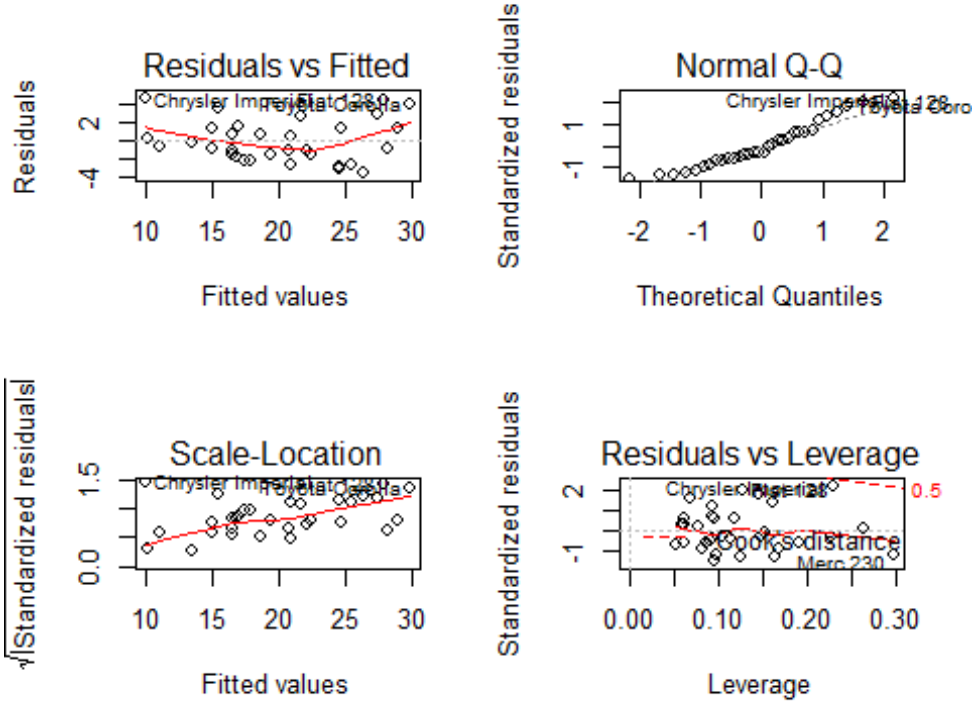


Figure 2.Residual plot for the best fit model

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


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Toyota Coro
Scale-Location
√|Standardized residuals|
Residuals vs Leverage
Cook's distance
Merc 230
0.5
Leverage