Is Automatic or Manual transmission better for MPG?

Executive summary

This analysis is about exploring the relatioship between a set of variables and Miles Per Gallon(MPG), particluarly the effects of transmission type on MPG. Based on the regresson models built on mtcars data set, the cars with manual transmission has better MPG (~2.9 increase) than the automatic transmission.

Is an automatic or manual transmission better for MPG?

Data Cleaning

```
library(datasets)
data(mtcars)
mydata<-mtcars
#To convert the type of factor variables to factor
mydata$cyl <- as.factor(mydata$cyl); mydata$am <- as.factor(mydata$am)
mydata$vs <- as.factor(mydata$vs); mydata$gear <- as.factor(mydata$gear)
mydata$carb <- as.factor(mydata$carb)
```

A boxplot was performed of Transmission type on MPG (see Appendix for the plot). From the plot, we can see that cars with manual transmission has a better MPG than the automatic transmission.

Quantify the MPG difference between automatic and manual transmissions

Single predictor regression

Model with only transmission ("am" variable) as predictor

```
fit_s <- lm(mpg ~ am, data = mydata)
summary(fit_s)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

From the summary, we have noticed that there is a positive relationship between predictor manual transmissions (am1) and outcome MPG. The coefficient is 7.24 and significant (P<0.05). So it again confirms our previous finding that cars with manual transmission have better MPG than those with automatic transmissions (am0).

Multivariable predictor regression

```
#Model with all variable as predictors
fit_m <- lm(mpg ~., data = mydata)
```

Stepwise selection using stepAIC(), which performs stepwise model selection by exact AIC

```r
#Model with selected variables as predictors
library(MASS)
fit_AIC<- stepAIC (fit_m, direction = "both")
fit_AIC$anova
#The model we obtaind by stepwise selection is: mpg ~ wt + qsec + am
```

All-subsets regression Using leaps()

```r
#Models with selected variables as predictors
library(leaps)
fit_subsets<-regsubsets(mpg~., data=mydata, nbest=1)
summary(fit_subsets)
#The model we choice for further analysis is the 4 predictor model: mpg ~ wt + hp + am + qsec
```

Compairing Models

```r
attach(mydata)
fit1 <-lm(mpg ~ am)
fit2<-lm(mpg ~ wt + am + qsec)
fit3<-lm(mpg ~ wt + hp + am + qsec)
fit4<-lm(mpg ~ ., data = mydata)
anova(fit1, fit2, fit3, fit4)
detach(mydata)
```

Model2 (obtaind by Stepwise selection using stepAIC()) is the only one with significant p-values thus was selected as the final model. The summary is as follows:

```r
summary(fit2)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## am1          2.935837  1.4109045  2.080819 4.671551e-02
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
```

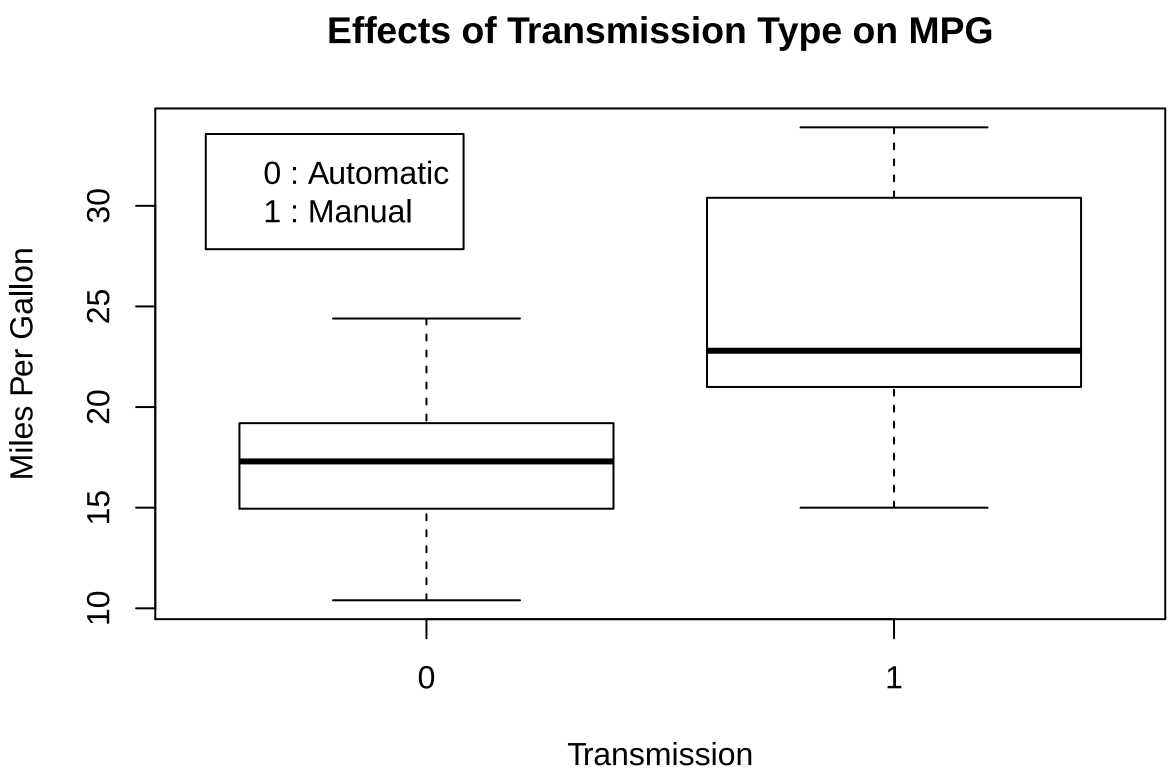
Conclution

Estimate MPG = 9.6 + 2.9*am1* - *3.9*wt + 1.2*qsec. so the cars with manual transmission have MPG 2.9 higher than cars with automatic transmission.

Appendix

Boxplot of Transmission type on MPG

```r
#"am" variable is the transmission type (0 = automatic, 1 = manual)
boxplot(mpg ~ am, data = mydata, main="Effects of Transmission Type on MPG", xlab="Transmission", ylab=
legend ("topleft", inset=.05, c("0 : Automatic", "1 : Manual"))
```

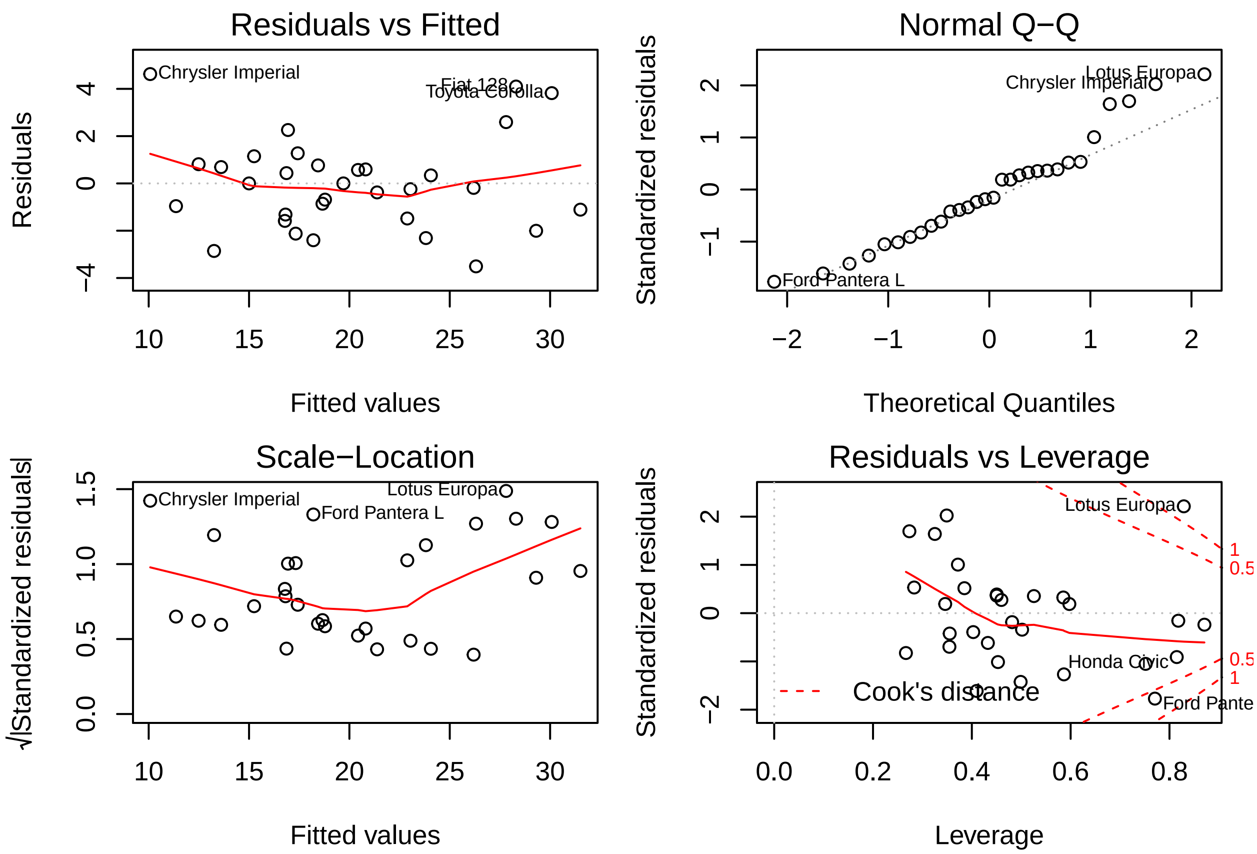


Diagnostic Plots

```
par(mfrow = c(2,2), mar = c(4,4,2,1))
plot(fit_m)
```

```
## Warning: not plotting observations with leverage one:
##   30, 31
```

```
## Warning: not plotting observations with leverage one:
##   30, 31
```



All-subsets regression Plot

```r
par(mfrow = c(1,1), mar = c(4,4,2,1))
plot(fit_subsets, scale="r2")
```