Analysis on MPG VS automatic or manual transmission

Rdataman

Sunday, July 27, 2014

1. Executive Summary

This report explore how the MPG (miles per gallon) parameter can be influenced by the transmission type of a car.
Detailed question is :

1. Is an automatic or manual transmission better for MPG?
2. Quantifying how different is the MPG between automatic and manual transmissions?

Boxplot, t-test and multivariable regression model are performed to answer the questions. The conclusion is that a manual car tends to be 1.809 miles/gallon higher than an automatic car if other parameters of the two car are same.

2. load the Data

```
## keep a copy of the original data
data(mtcars); str(mtcars); mtcars0 <- mtcars
mtcars$am <- as.factor(mtcars$am);levels(mtcars$am) <-c("Automatic", "Manual");
```

This dataframe contains 32 obs. of 11 numeric variables.

3. Means test (ttest) to the questions

A boxplot shows that manual transmission cars do better in MPG than the automtic ones(check Appendix 1).**The answer can be validated with the t test.**

```
tt<-t.test(mtcars$mpg~mtcars$am); c(tt$p.value,tt$estimate)
```

```
##                          mean in group Automatic    mean in group Manual
##                0.001374                17.147368                24.392308
```

Clearly, the mean mpg for automatic and manual transmission cars are 17.1 and 24.4 respectively. The p-value is much smaller than 0.05. We can conclude that automatic cars has lower mpg compared with manual cars included in this dataframe.

4. Regression model for the questions.

Simple Linear Regression

```
summary(lm(mpg~am,mtcars))
```

The above code shows a simple model (mpg~am). It is poorly fitted (R-squared:0.36). The reason is that "mpg" are also highly related to other factors than am (check Appendix 2). In order to quantify the mpg difference better, a multivariable regression model should be made.

Multivariate regression model

A stepwise model selection will be helpful to determine the final multivariable.

```
## transfer some other cols into factors
mtcars$cyl <- as.factor(mtcars$cyl) ; mtcars$vs <- as.factor(mtcars$vs)
mtcars$gear <- as.factor(mtcars$gear) ; mtcars$carb <- as.factor(mtcars$carb)
modfit<-lm(mpg~., data=mtcars)
library(MASS)
step <- stepAIC(modfit, direction="both");step$anova
```

mpg ~ cyl + hp + wt + am is suggested. Make a new fitting.

```
finalfit<-lm(mpg ~ cyl + hp + wt + am, data=mtcars);summary(finalfit)$coefficients
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 33.70832    2.60489 12.9404 7.733e-13
## cyl6        -3.03134    1.40728 -2.1540 4.068e-02
## cyl8        -2.16368    2.28425 -0.9472 3.523e-01
## hp          -0.03211    0.01369 -2.3450 2.693e-02
## wt          -2.49683    0.88559 -2.8194 9.081e-03
## amManual     1.80921    1.39630  1.2957 2.065e-01
```

The final fitting seems ok judging by the value of R-squared(0.866) and the Residuals VS Fitted plots in appendix3. The Normal Q-Q plot indicates that the residuals are close to a normally distribution.
Interpretion of the oefficient

1. cly increase from 4 to 6 to 8, mpg decrease by 3 and 2.2 mile/gallon, respectively
2. mpg will decrease slighly with hp increase.
3. wt increases per 1000lb, the mpg will decrease by 2.5 mile/gallon.
4. **The coefficient of "am" in the model is 1.809. A manual car tends to be 1.809 miles/gallon higher than an automatic car if other parameters of the two car are same.**

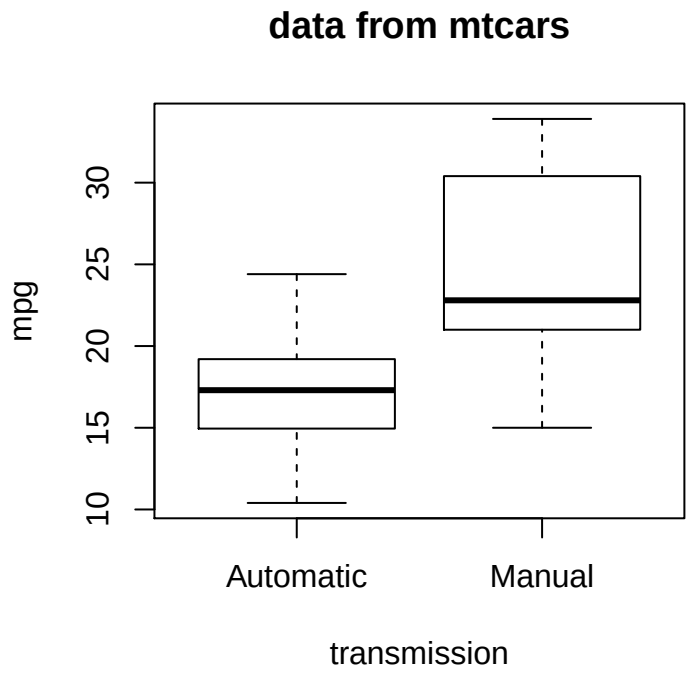
5. Conclusion

A multivariable model was built. The 4 predictors are number of cylinders, weight, gross horsepower and transmission type. This model gives a good fitting and proves that a manual car runs 1.809 miles/gallon than an automatic one.

Appendix

Appendix 1: Boxplot of mpg VS am

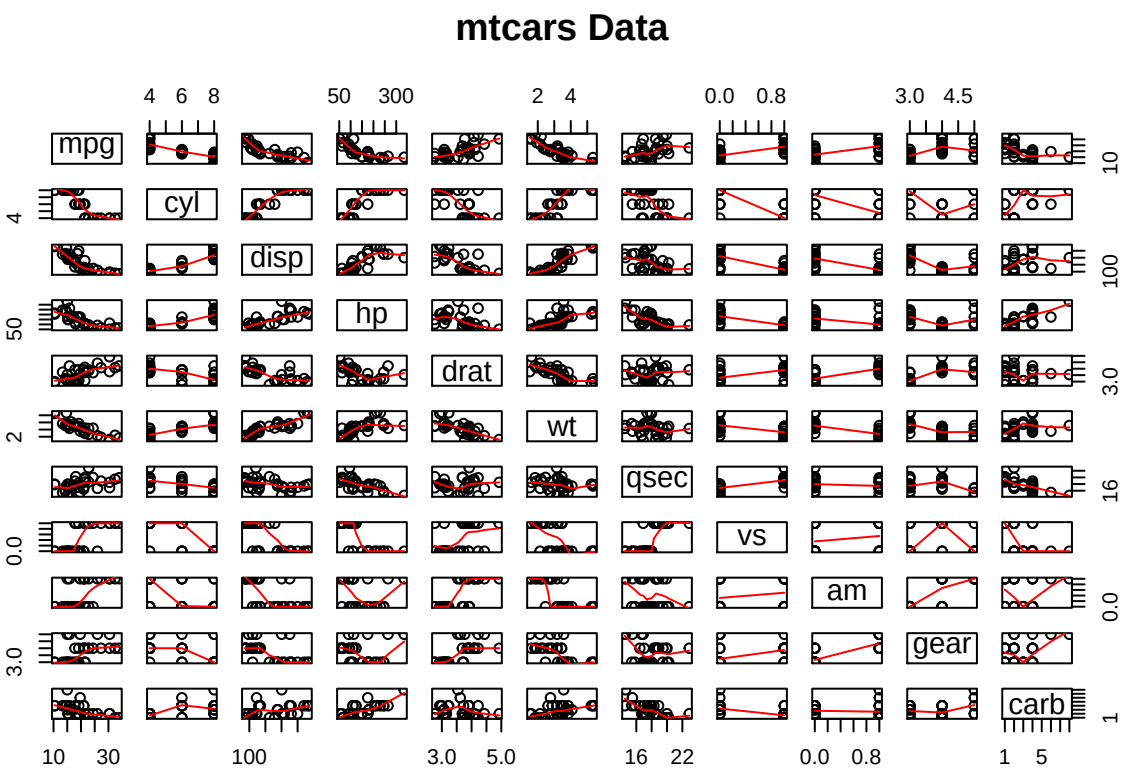
```
boxplot(mtcars$mpg ~ mtcars$am, data = mtcars, ylab="mpg",
        xlab="transmission",main="data from mtcars")
```



Appendix 2 : The dependence of mpg on other parameters.

The relationships can be shown by a scatterplot of mtcars.

```
pairs(mtcars0, panel=panel.smooth, main="mtcars Data")
```



The cor() can also show how mpg is affected by all other parameters ("am" is not included) of a car.

```
# check cor between mpg and others (no "am")
cor(mtcars$mpg,mtcars0[,-1])
```

```
##          cyl    disp      hp   drat      wt   qsec    vs     am   gear
## [1,] -0.8522 -0.8476 -0.7762 0.6812 -0.8677 0.4187 0.664 0.5998 0.4803
##         carb
## [1,] -0.5509
```

Appendix 3 : Residuals plots of our model.

```
par(mfrow=c(2,2)); plot(finalfit)
```