Motor Trend Data Analysis: The Effect of automatic or manual transmission on miles per gallon of cars

Lau Heng Kar

January 25, 2016 (Regression Models Course Project)

Executive Summary

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). Looking at a data set of a collection of cars, this analysis is to explore the relationship between a set of variables and miles per gallon (MPG) (outcome). The analysis are particularly focus in the following two questions:

1. "Is an automatic or manual transmission better for MPG"
2. "Quantify the MPG difference between automatic and manual transmissions"

Loading and Preparing Data

```
require(datasets)
data(mtcars)
# Prepare below variable as factor
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am, labels = c('Automatic','Manual'))
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
```

Exploring and Basic summary of the data

```
summary(mtcars)
```

```
##       mpg        cyl         disp             hp             drat
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760
##  1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080
##  Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695
##  Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597
##  3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920
##  Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930
##        wt             qsec       vs             am     gear   carb
##  Min.   :1.513   Min.   :14.50   0:18   Automatic:19   3:15   1: 7
##  1st Qu.:2.581   1st Qu.:16.89   1:14   Manual   :13   4:12   2:10
##  Median :3.325   Median :17.71                         5: 5   3: 3
##  Mean   :3.217   Mean   :17.85                                4:10
##  3rd Qu.:3.610   3rd Qu.:18.90                                6: 1
##  Max.   :5.424   Max.   :22.90                                8: 1
```

```r
table(mtcars$am, mtcars$am)
```

```
##
##              Automatic Manual
##   Automatic         19      0
##   Manual             0     13
```

According to the Box plot, we see that manual transmission yields higher values of MPG in general. The mean of mpg is greater for manual (at **24.3923077**) than automatic (at **17.1473684**). **See MPG by transmission type box plot in the appendix.**

From pair plot, we can see some higher correlations between variables like "wt", "disp", "cyl" and "hp". **See Pairwise Scatterplot in the appendix.**

Inference

From the simple box plot, it seems that manual transmission is better in MPG than automatic transmission. We need to perform a t-test to confirm this hypothesis. We make the null hypothesis as the MPG of the automatic and manual transmissions are from the same population (assuming the MPG has a normal distribution)

```r
result <- t.test(mtcars$mpg~mtcars$am, conf.level = 0.95)
result # result hidden
```

with p-value = **0.0013736**, we reject the null hypothesis that there is no difference in MPG, and manual transmission looks better in MPG than automatic transmission, provided that all other conditions are the same.

Regression Analysis

First we will evalute both basic model and full model with every possible variable, and find the model with the most influential ones through the 'step' function with direction "both"

```r
# Basic Model - linear regression for automatic vs manual car
basic_fit <- lm(mpg ~ am, data = mtcars)
summary(basic_fit) # result hidden

# Full Model - linear regression for automatic vs manual car
fit <- lm(mpg ~ ., data = mtcars)
summary(fit) # result hidden

# Choose a model by AIC in a Stepwise Algorithm
full_fit <- step(fit, direction = "both")
summary(full_fit) # result hidden
```

```r
# Nested model testing for basic and full model
anova(basic_fit, full_fit)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With p-value < 0.05, reject null hypothesis. Thus Model 2: mpg ~ cyl + hp + wt + am (with 4 predictors) has statistically significant as compared to Model 1: mpg ~ am. The final model is below:

```
# Final Model - Model 2: mpg ~ cyl + hp + wt + am
summary(full_fit)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## amManual     1.80921138 1.39630450  1.295714 2.064597e-01
```

The adjusted R squared for final model is **0.8400875**. Thus, the result shows that when "cyl" (Number of cylinders), "hp" (Gross horsepower) and "wt" (weight lb/1000) remain constant, cars with manual transmission add **1.8092114** more MPG (miles per gallon) on average than cars with automatic transmission.

Residual Plot and Diagnostics

According to the residual plots, we can verify the following underlying assumptions:

1. The Residuals vs. Fitted plot shows no consistent pattern, supporting the accuracy of the independence assumption.
2. The Normal Q-Q plot indicates that the residuals are normally distributed because the points lie closely to the line.
3. The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed.
4. The Residuals vs. Leverage argues that no outliers are present, as all values fall well within the 0.5 bands. (There's just one to the right of the plot, but it isn't further than the 0.5 Cook's distance)

See Residual Plot and Diagnostics in the appendix.

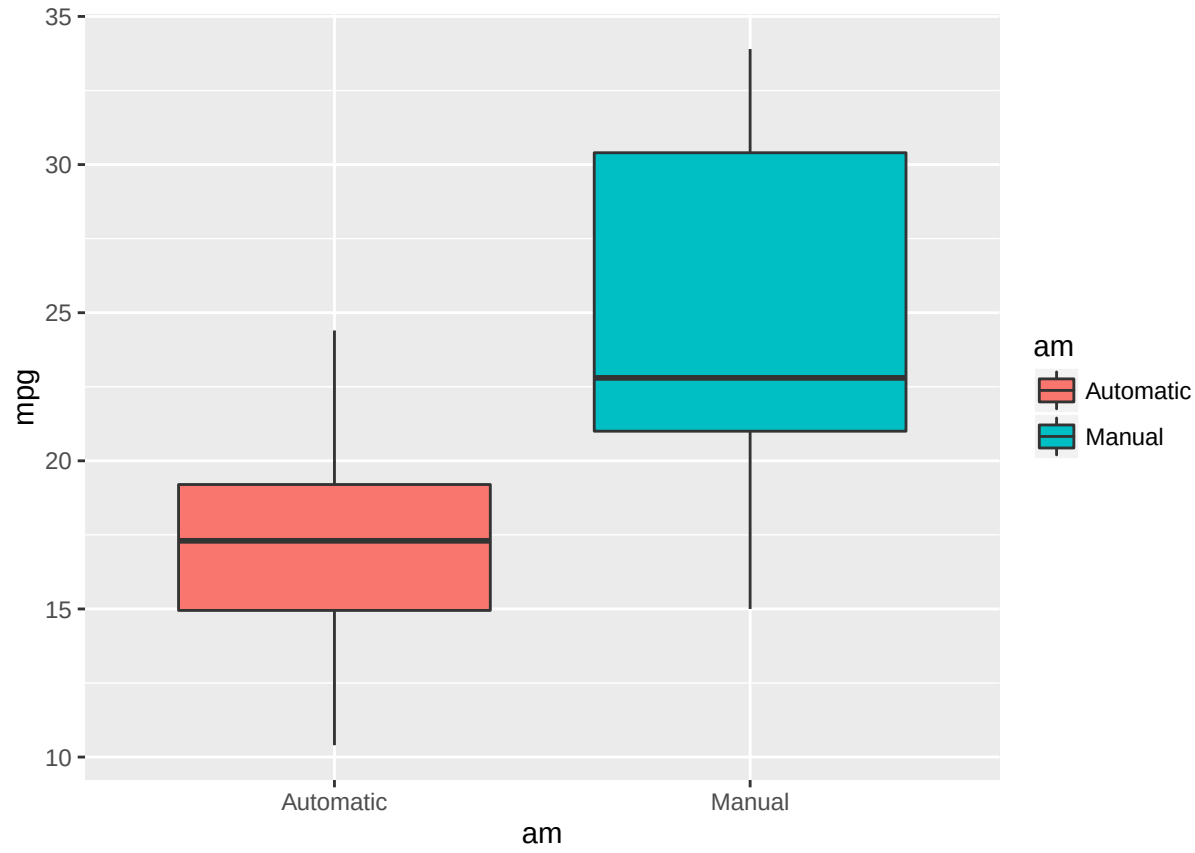
Conclusions

This model explains **84.008754**% of the variance. It may be concluded that on average, manual transmissions have **1.8092114** more mpg than automatic. The above analyses meet all basic assumptions of linear regression and well answer the questions.

Appendix : Supporting figures

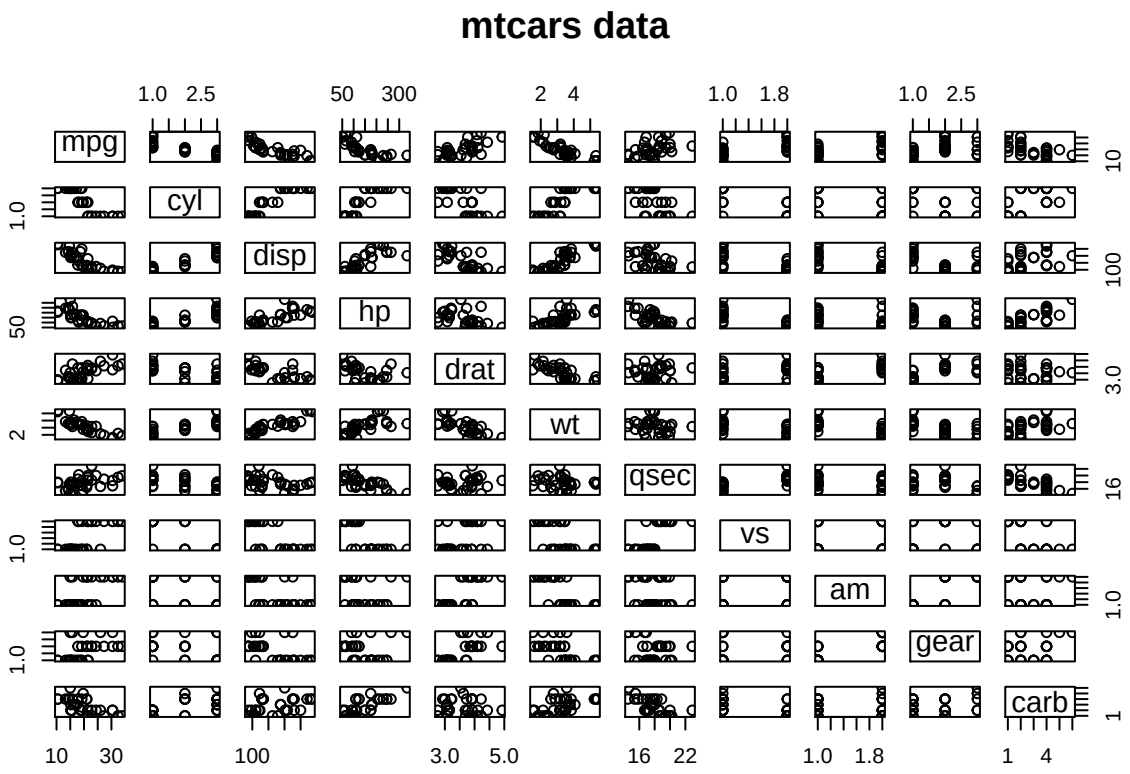
MPG by transmission type Box plot

```
require(ggplot2)
g <- ggplot(data = mtcars, aes(x = am, y = mpg))
g <- g + geom_boxplot(aes(fill = am))
g
```



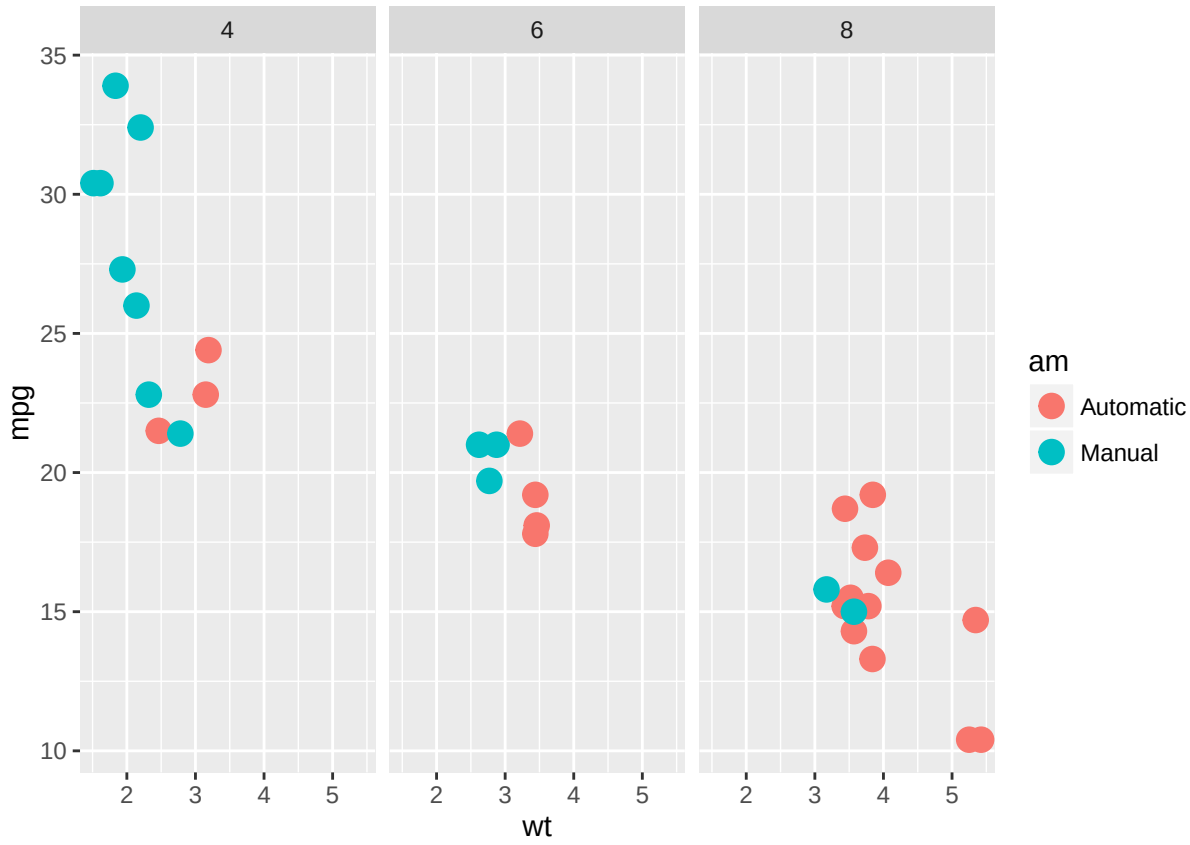
Pairwise Scatterplot for "mtcars"

```
pairs(mtcars, main = "mtcars data")
```



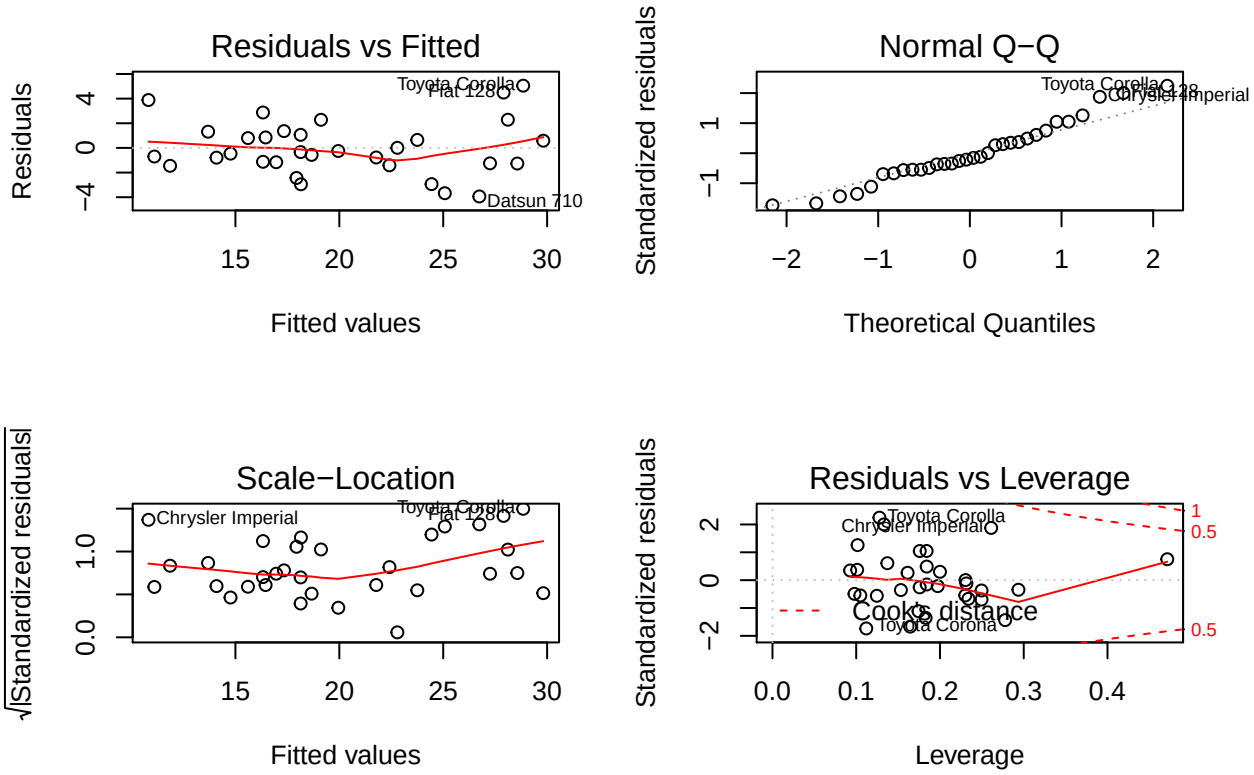
MPG vs. Weight by Transmission

```
require(ggplot2)
g <- ggplot(data = mtcars, aes(x = wt, y = mpg, colour = am))
g <- g + facet_grid(.~cyl) + geom_point(size = 4)
g
```



Residual Plot and Diagnostics

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


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Chrysler Imperial
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Toyota Corolla
Chrysler Imperial
Toyota Corona