Regresseion Model - Assignment

Manojit Saha

28 August 2016

Executive Summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

- “Is an automatic or manual transmission better for MPG”
- “Quantify the MPG difference between automatic and manual transmissions”

Using Regeression models and Hypothesis testing we have tried to measure the impact of transmission on fuel efficiency. It is found that there are other variables too which impact the mileage.The final model results indicates that weight and quarter mile time (acceleration) have signficant impact in quantifying the difference of mpg between automatic and manual transmission cars.

The Data Set

The data set was extracted from the 1974 edition of Motor Trend US Magazine and it deals with 1973 - 1974 models. It consists of 32 observations on 11 variables:

- mpg: Miles per US gallon
- cyl: Number of cylinders
- disp: Displacement (cubic inches)
- hp: Gross horsepower
- drat: Rear axle ratio
- wt: Weight (lb / 1000)
- qsec: 1 / 4 mile time
- vs: V/S
- am: Transmission (0 = automatic, 1 = manual)
- gear: Number of forward gears
- carb: Number of carburetors

Load Dataset, Print Summary

```
library(ggplot2)
data(mtcars)
dim(mtcars)
```

```
## [1] 32 11
```

```
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

```
mtcars[1:5, ]
```

```
##                    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
```

From this dataset we need to find answers of the following questions:

- Is an automatic or manual transmission better for MPG
- Quantify the MPG difference between automatic and manual transmissions

Inference

Run a Hypothesis Test as MPG vs. transmission to check if they are co-related. Perform Two Sample T-test.

```
result <- t.test(mtcars$mpg ~ mtcars$am)
result$p.value
```

```
## [1] 0.001373638
```

```
result$estimate
```

```
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

Since p value is .00137, we reject our null hypothesis. The mileage of a manual car is 7 miles higher than that of an Automatic one.

Regression Analysis

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

```
##
## 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
## am           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
```

This model can explain about 78% of the MPG variable and none of the coeffients are 95% of the confidance level. So we need to check if inclusion of some other variable can give statistically significant result.

```
amIntWtModel<-lm(mpg ~ wt + qsec + am + wt:am, data=mtcars)
summary(amIntWtModel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am + wt:am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5076 -1.3801 -0.5588  1.0630  4.3684
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.723      5.899   1.648 0.110893
## wt            -2.937      0.666  -4.409 0.000149 ***
## qsec           1.017      0.252   4.035 0.000403 ***
## am            14.079      3.435   4.099 0.000341 ***
## wt:am         -4.141      1.197  -3.460 0.001809 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.084 on 27 degrees of freedom
## Multiple R-squared:  0.8959, Adjusted R-squared:  0.8804
## F-statistic: 58.06 on 4 and 27 DF,  p-value: 7.168e-13
```

This comes close to 88% confidence interval, but it should be closer to 95%

So let’s try another model

```
amModel<-lm(mpg ~ am, data=mtcars)
summary(amModel)
```

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

This model is not as good as the earlier one and it seems variation os mpg not only related to transmission, but there is (are) something else.

```
anova(amModel, fullModel, amIntWtModel)
confint(amIntWtModel)
```

The model with highest Adjusted R-squared value, “mpg ~ wt + qsec + am + wt:am” is the best of all options here.

```
summary(amIntWtModel)$coef
```

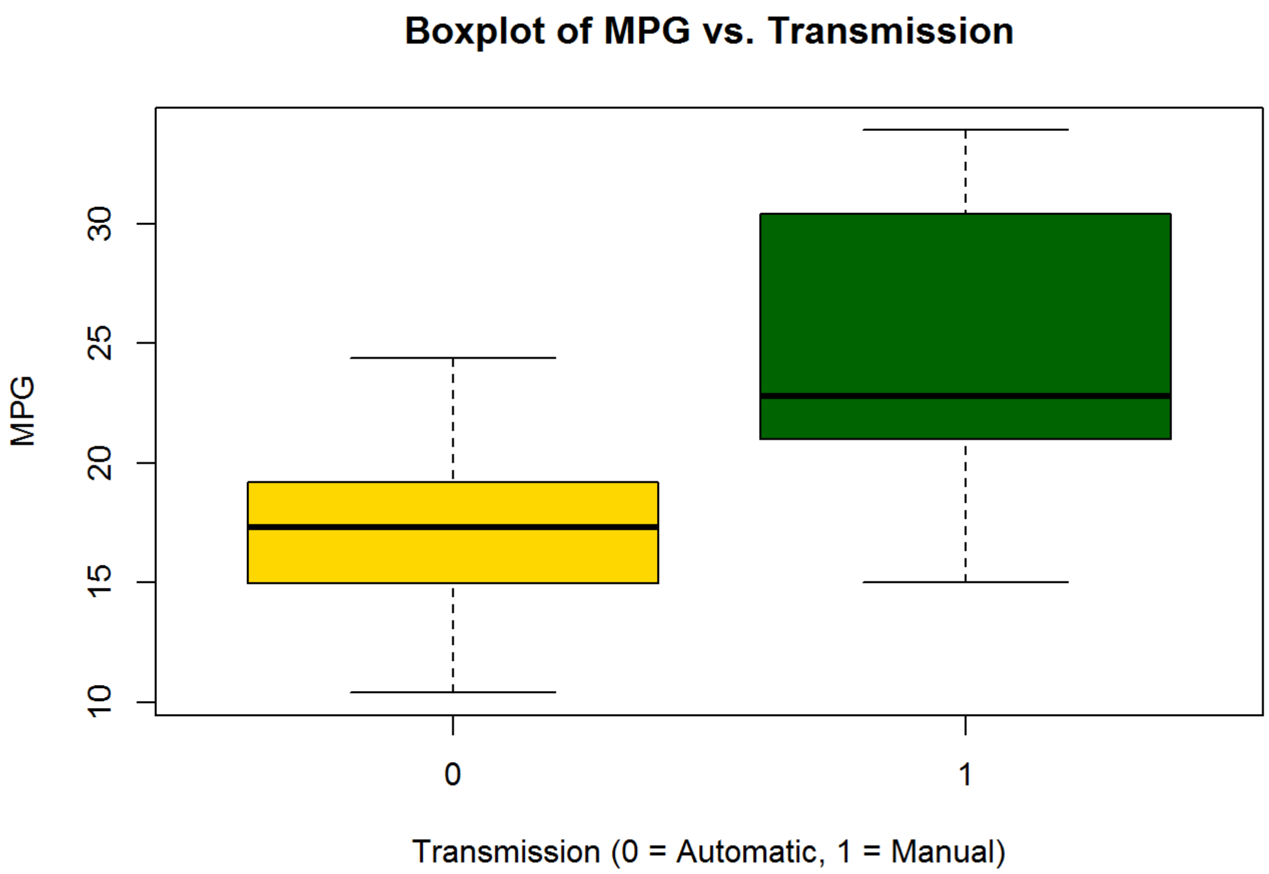
```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.723053  5.8990407  1.648243 0.1108925394
## wt          -2.936531  0.6660253 -4.409038 0.0001488947
## qsec         1.016974  0.2520152  4.035366 0.0004030165
## am          14.079428  3.4352512  4.098515 0.0003408693
## wt:am       -4.141376  1.1968119 -3.460340 0.0018085763
```

The result shows that when “wt” (weight lb/1000) and “qsec” (1/4 mile time) remain constant, cars with manual transmission add 14.079 + (-4.141)*wt more MPG (miles per gallon) on average than cars with automatic transmission.

Appendix: Plots and Figures

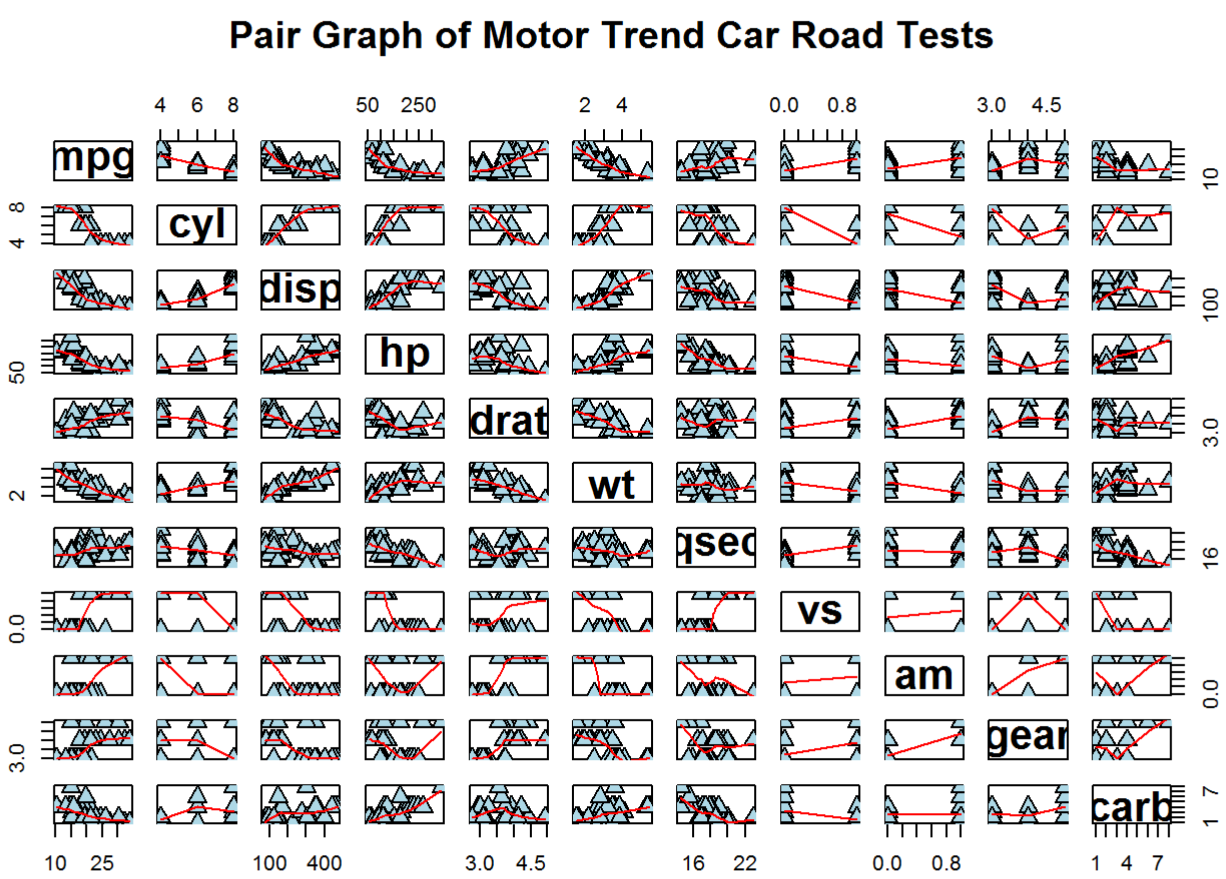
1. Boxplot - MPG vs. Transmission

```
boxplot(mtcars$mpg ~ mtcars$am, xlab="Transmission (0 = Automatic, 1 = Manual)", ylab="MPG", col=(c("gold","darkgreen")), ma
in="Boxplot of MPG vs. Transmission")
```



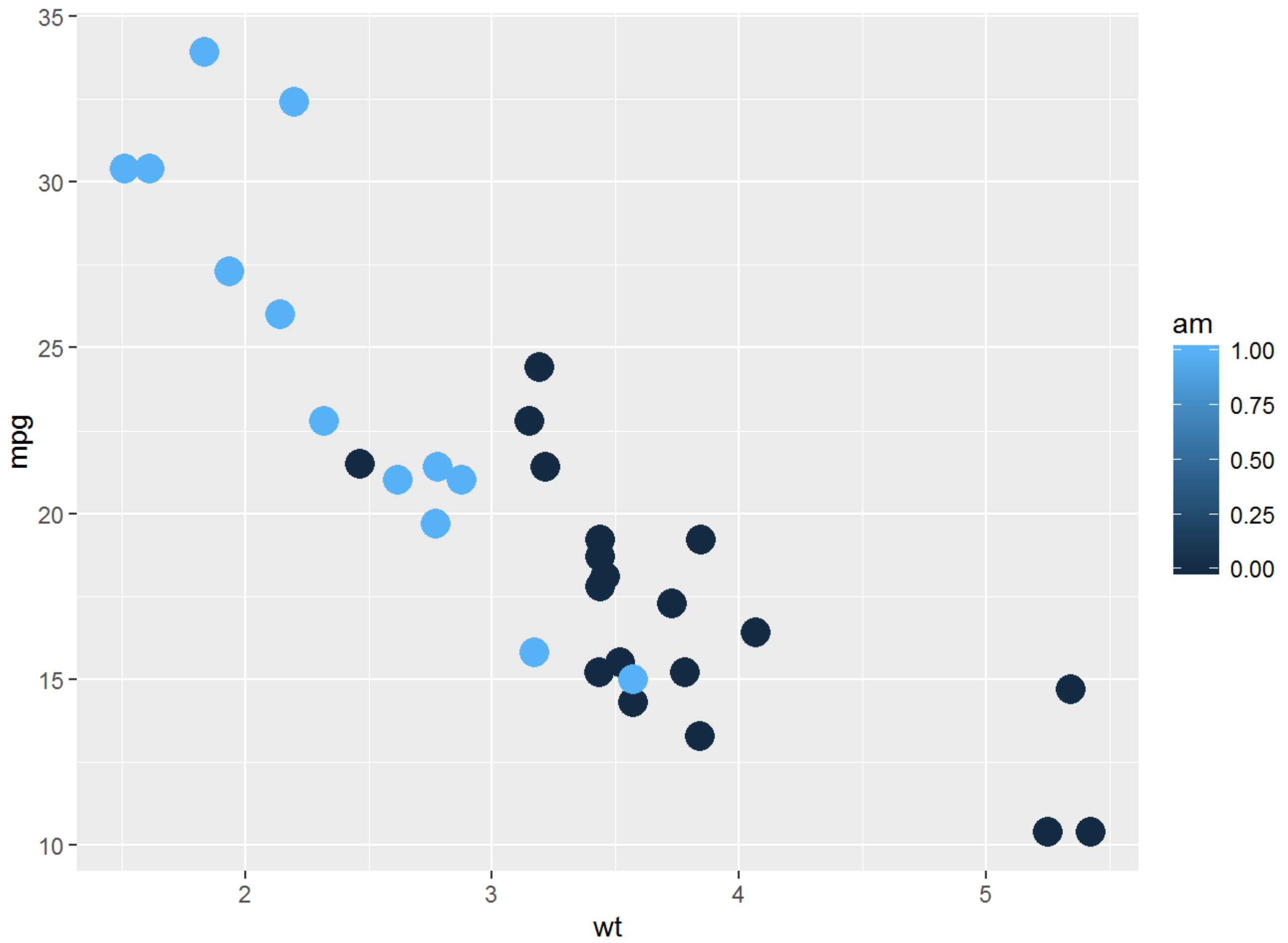
2. Pair Graph of Motor Trend Car Road Tests

```
pairs(mtcars, panel=panel.smooth, main="Pair Graph of Motor Trend Car Road Tests", cex = 1.5, pch = 24, bg = "light blue", c
ex.labels = 2, font.labels = 2)
```



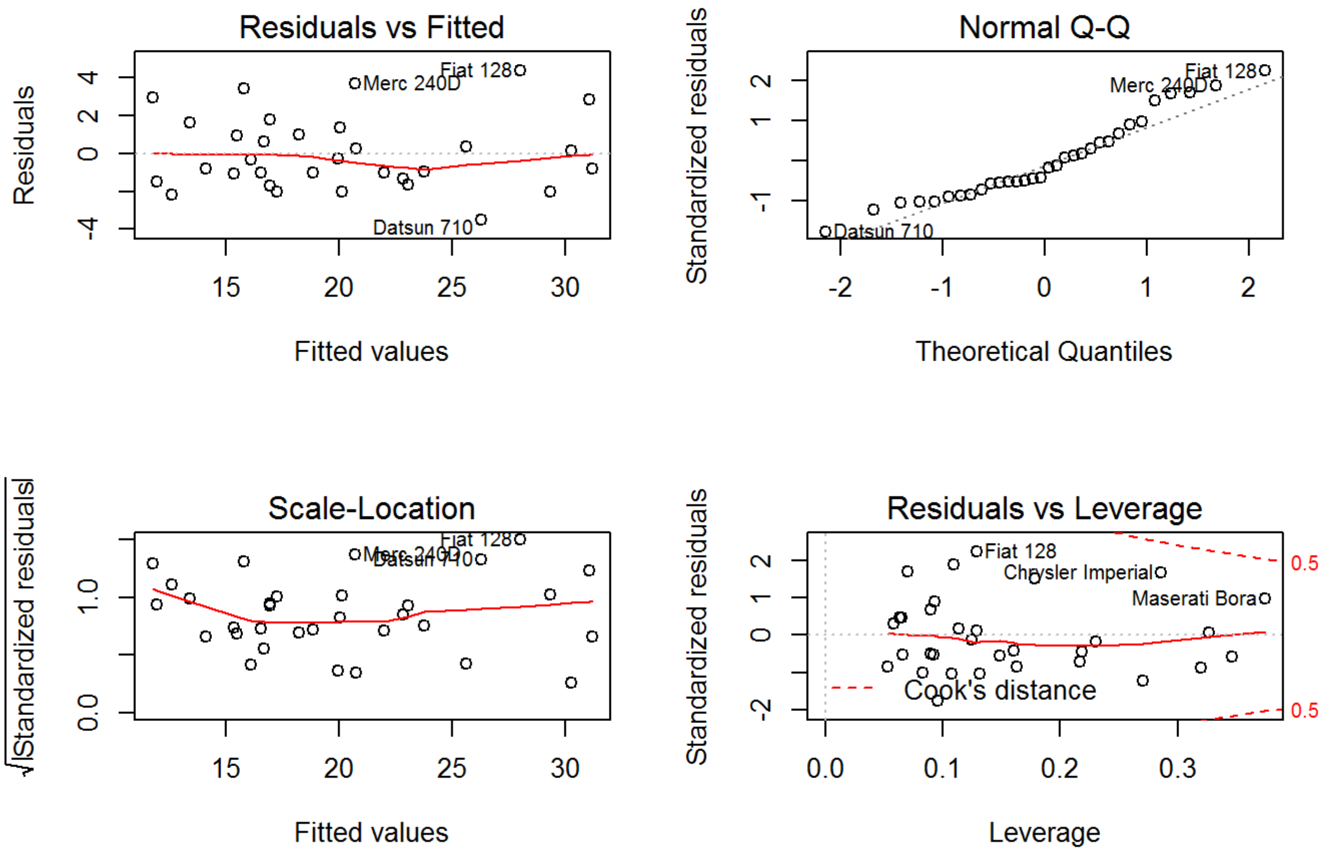
3. Scatter Plot of MPG vs. Weight by Transmission

```
g2 <- ggplot(mtcars,aes(x=wt,y=mpg,colour=am))+geom_point(size=5)
g2
```



4. Residual Plots

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



Residuals vs Fitted
Fitted values
Residuals
Fiat 128
Merc 240D
Datsun 710
Normal Q-Q
Theoretical Quantiles
Standardized residuals
Scale-Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Chrysler Imperial
Maserati Bora
Cook's distance
0.5