What parameters give a better mileage for a car? Does type of Transmission matter?

Kuber K Gudipati

Sunday, July 26, 2015

Context

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”

Data

```
library(ggplot2)
data(mtcars)

mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)

levels(mtcars$am) <- c("Automatic", "Manual");
```

Initial Analysis

Let’s go with problem statement and draw a boxplot between mpg and am to see what type of transmission gives a better mileage. See the figures in Appendix.

It’s evdent from the boxplot that Manual Transmission gives better mileage compared to Automatic. Lets draw the conclusion from regression models. Lets go with a liner regression models ### Regression Analysis

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

R-squared: 0.3598 means, it only explaind 36% of variance in mpg. There may be other potential variables affecting mpg. Let’s plot with all the variables

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

Based on the above summary (Estimate and t-values), CylinderCount, horsepower, weight and carb2 appear to be significant. Lets do stepwise model and figure out the most significant factors

```r
stepModel <-step(fullModel,direction="backward", k=log(nrow(mtcars)))
summary(stepModel)
```

Based on the above stepwise model, wt + qsec + am seems to be significant on mpg. Let analyze this model

```r
model2<-lm(mpg ~ wt + qsec + am, data = mtcars)
summary(model2)
```

Lets compare mpgamModel and model2 and decide which one is better

```r
anova(mpgamModel,stepModel)
```

Since the p-vale $<.05$, model2 seems to give better results.

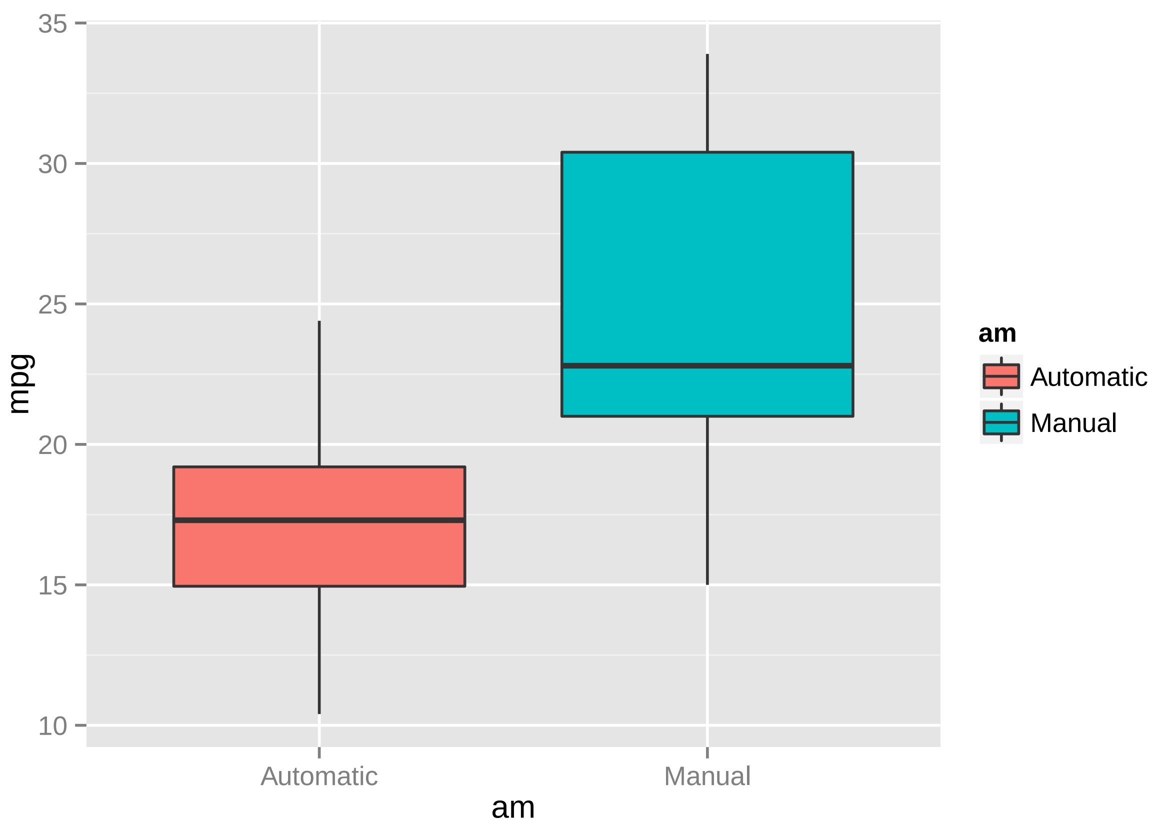
Conclusion

```r
summary(model2)
```

It appears that, keeping weight and accelertion (1/4th mile distance) constant, Manual Transmission may give 2.936 more miles per gallon

Appendix

Boxplot between Manual and Auto Transmission vs mpg



Model1: mpg~am

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

```
##
## 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 ***
## amManual       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
```

stepModel:

```r
stepModel <-step(fullModel,direction="backward", k=log(nrow(mtcars)))
summary(stepModel)
```

summary of model2:

```r
model2<-lm(mpg ~ wt + qsec + am, data = mtcars)
summary(model2)
```

```
##
## 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 ***
## amManual      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
```

Lets compare mpgamModel and model2 and decide which one is better:

```r
anova(mpgamModel,stepModel)
```

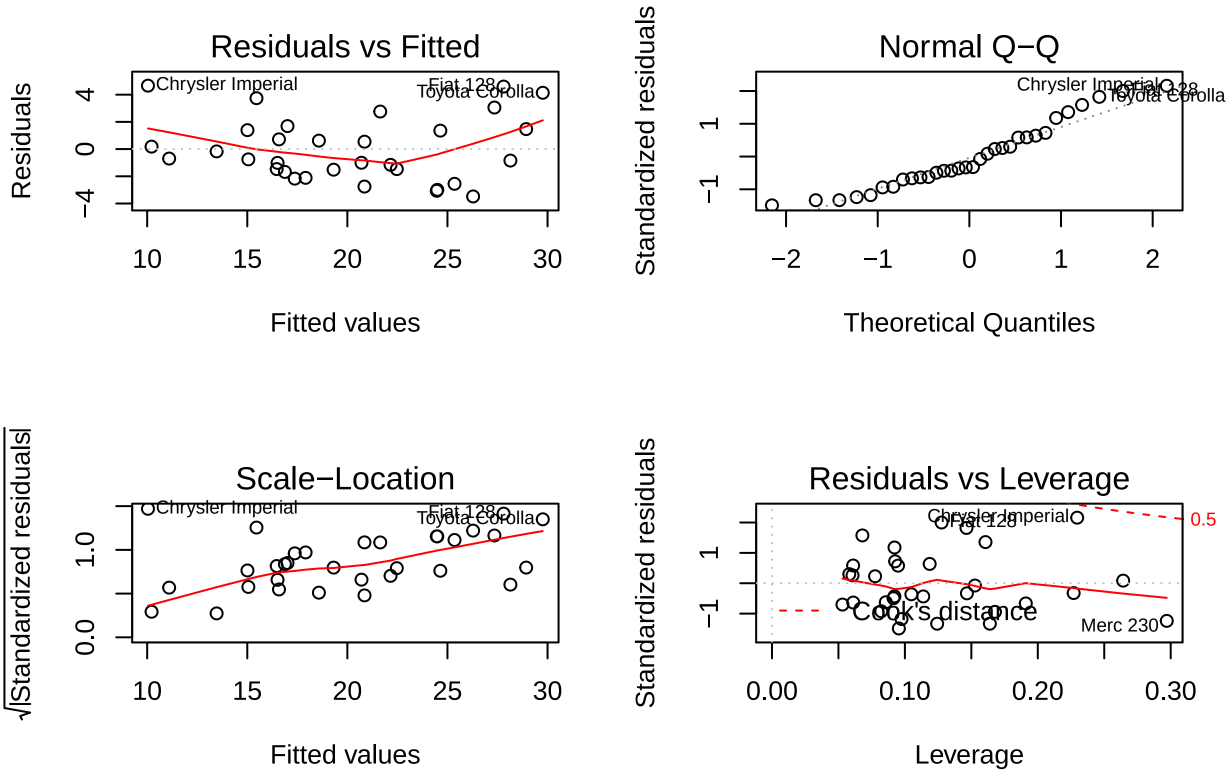
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion

```r
summary(model2)
```

```
##
## 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 ***
## amManual      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
```

```r
par(mfrow = c(2,2))
plot(model2)
```


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