Regression Models - Project

Saul Lugo

December 27, 2015

Executive Summary

This is my executive summary

Loading the Data

The data used for this analysis comes from the R mtcars dataset.

```
data(mtcars)
head(mtcars)
```

```
##                    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
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Data transformation

As this analysis is primary concern with the relationship between the cars transmission type (automatic or manual) and the fuel efficiency as miles per gallon (MPG), a factor variable named **transmission** was added to the dataset. Although the dataset already had the variable **am** that identifies each car's transmission type with 0 and 1, this new factor variable **transmission** is easier to visualize in a plot.

```
require(dplyr)
```

```
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
mtcars <- mutate(mtcars, transmission = ifelse(am==0,"automatic","manual"))
head(mtcars)
```

```
##    mpg cyl disp  hp drat    wt  qsec vs am gear carb transmission
## 1 21.0   6  160 110 3.90 2.620 16.46  0  1    4    4       manual
## 2 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4       manual
## 3 22.8   4  108  93 3.85 2.320 18.61  1  1    4    1       manual
## 4 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1    automatic
## 5 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2    automatic
## 6 18.1   6  225 105 2.76 3.460 20.22  1  0    3    1    automatic
```

Exploratory Data Analysis

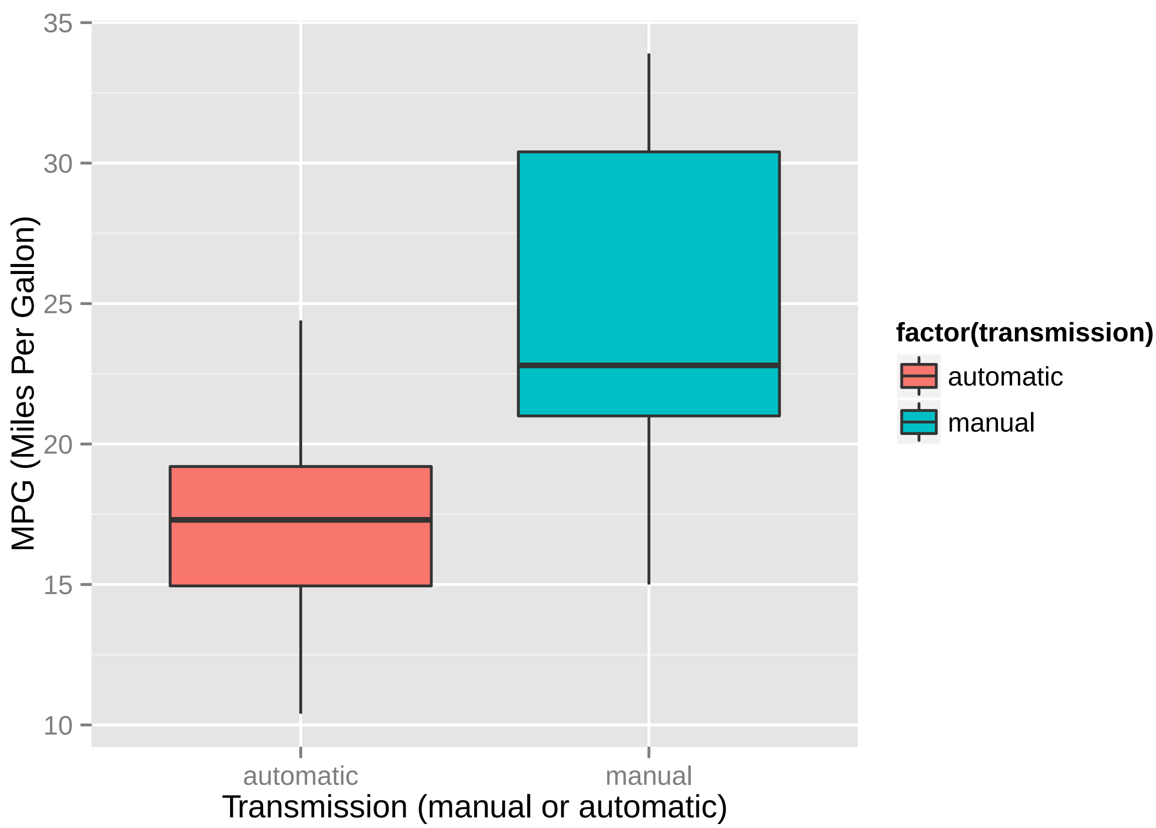
The following chart serves to visualize the posible relationship between the car transmission type and the fuel efficiency (MPG):

```r
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```r
p <- ggplot(mtcars, aes(x = factor(transmission), y = mpg))
p <- p + geom_boxplot(aes(fill = factor(transmission)))
p <- p + xlab("Transmission (manual or automatic)") + ylab("MPG (Miles Per Gallon)")
p
```



```
autoMean <- mean(mtcars[mtcars$transmission=="automatic",]$mpg)
manualMean <- mean(mtcars[mtcars$transmission=="manual",]$mpg)
autoSD <- sd(mtcars[mtcars$transmission=="automatic",]$mpg)
manualSD <- sd(mtcars[mtcars$transmission=="manual",]$mpg)
autoMean
```

```
## [1] 17.14737
```

```
autoSD
```

```
## [1] 3.833966
```

```
manualMean
```

```
## [1] 24.39231
```

```
manualSD
```

```
## [1] 6.166504
```

```
manualMean - autoMean
```

```
## [1] 7.244939
```

As the previous chart shows, there is an aparent relationship between the car transmission type and the fuel efficiency. The evidence analysed so far indicates that the manual transmission cars have in average a higher fuel efficiency as that of the automatic transmission cars.

In average, the manual transmission cars are 7.25 MGP more efficient than the automatic transmission cars.

Linear Models

The first model fitted in this analysis has MPG as the outcome and the factor of transmission type as the regressor:

```
fit1 <- lm(mpg ~ transmission, data = mtcars)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ transmission, 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 ***
## transmissionmanual    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
```

```r
co1 <- summary(fit1)$coefficients
co1
```

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

```r
b0_1 <- co1[1,1]
b1_1 <- co1[2,1]
autoCI <- b0_1 + c(-1,1) * pt(.975, df=fit1$df)*co1[1,2]
autoCI
```

```
## [1] 16.21246 18.08227
```

```r
manualCI <- b1_1 + c(-1,1) * pt(.975, df=fit1$df)*co1[2,2]
manualCI
```

```
## [1] 5.778138 8.711741
```

For this model, the hypothesis of whether the model factors are 0 vs the model factors are different from 0 can be tested calculating the 95% confidence interval.

As it can be observed from the confidence interval calculation of the last R code chunck, neither of the intervals include zero, so the model factors must be different from zero and the estimates of 17.15 for $\beta 0$ and 7.24 for $\beta 1$ must be correct.

Comparison with other Linear Models

One could think that the transmission type is not the only variable that has an influence in the fuel efficiency of a car. For instance, other variables included in the mtcars dataset might have a relevant influence on it. The following model takes in consideration also the cylinders, weight, gears and carburators of the car:

```r
summary(lm(mpg ~ transmission + factor(cyl) + wt + gear + carb, data = mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ transmission + factor(cyl) + wt + gear + carb,
##     data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5571 -1.3941 -0.2116  1.3974  5.3330
```

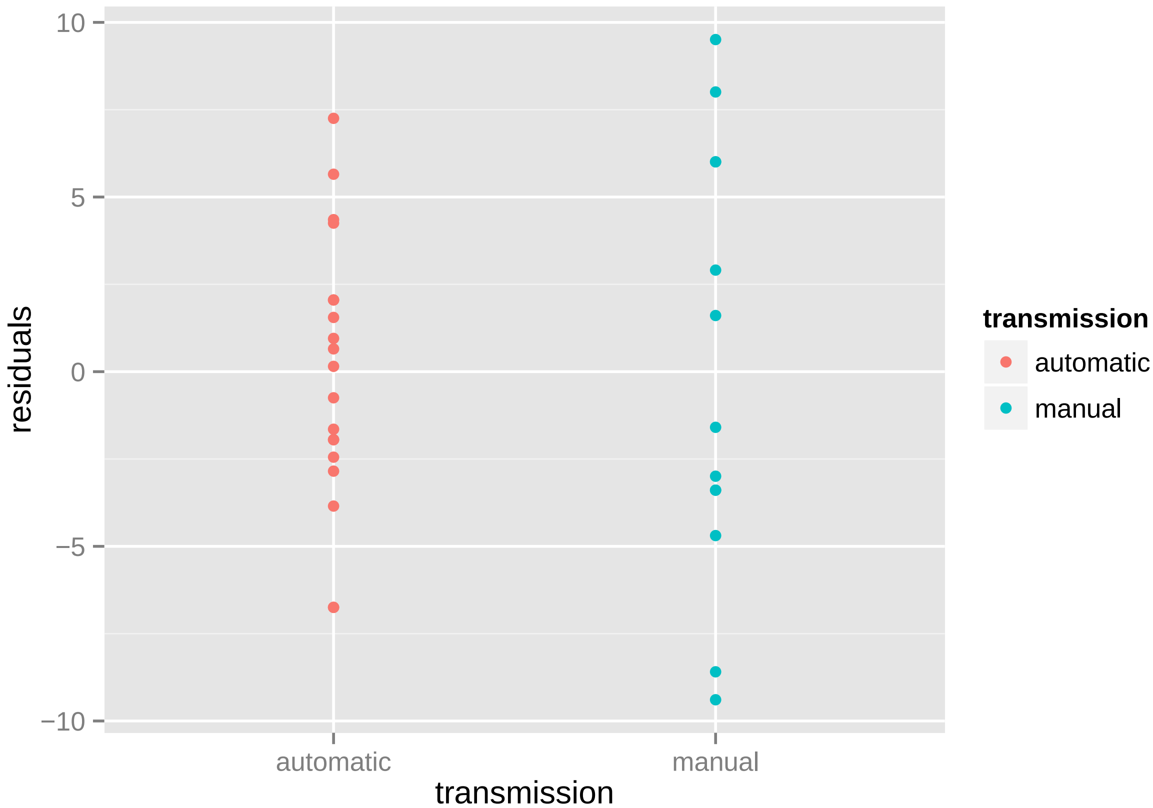
```
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)    
## (Intercept)         31.2644     6.0569   5.162 2.45e-05 ***
## transmissionmanual   1.4487     1.7670   0.820   0.4200    
## factor(cyl)6        -2.9265     1.7077  -1.714   0.0990 .  
## factor(cyl)8        -4.7321     2.0610  -2.296   0.0303 *  
## wt                  -2.4946     1.0215  -2.442   0.0220 *  
## gear                 0.2961     1.4129   0.210   0.8357    
## carb                -0.7531     0.5801  -1.298   0.2060    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.583 on 25 degrees of freedom
## Multiple R-squared:  0.8519, Adjusted R-squared:  0.8163 
## F-statistic: 23.97 on 6 and 25 DF,  p-value: 3.126e-09
```

Eventhough the other factors besides the transmission type has an influence in the outcome (the MPG), still the manual transmission has a better fuel efficiency in this model than that of the automatic transmission. So, it can be concluded that the transmission type is a strong predictor of the MPG.

Residuals Calculation

From the first model calculated with MPG as an outcome and transmission type as the regressor, the residual calculation and plotting is as follow:

```
e = resid(fit1)
mypoints <- data.frame(transmission = mtcars$transmission, residuals = e)
ep <- ggplot(mypoints, aes(x = transmission, y = residuals))
ep <- ep + geom_point(aes(colour = transmission))
ep
```


10
5
0
−5
−10
residuals
automatic
manual
transmission
transmission
automatic
manual