A Study on Fuel Efficient Transmission Type

thenuv

31 July 2018

Executive Summary

The convenience of automatic transmission in cars leads to the question of fuel inefficiency. To provide an insight on if its the case, we studied the dataset *mtcars* to explore the possible outcomes applying Linear Regression model.

If we measure the fuel efficiency with type of transmission alone (Automatic/manual), then manual seems to be more efficient. But while including other correlated variables, we could infer Automatic is similar to manual type transmission.

Data Processing

```
suppressWarnings(library(ggplot2))
suppressWarnings(library(car))
```

```
## Loading required package: carData
```

```
options(scipen = 999)
data("mtcars")
mtcars$am<-factor(mtcars$am)
levels(mtcars$am)<-c("automatic", "manual")
```

We use the mtcars dataset collected by Motor trend in 1974. There are 32 observations and 11 variables. Variable ***am*** represents the transmission type. We convert them into factor and name accordingly. Following Variables are used in the Analysis.

- **mpg** Miles/(US) gallon
- **cyl** Number of cylinders
- **wt** Weight (1000 lbs)
- **am** Transmission (0 = automatic, 1 = manual)
- **disp** Displacement (cu.in.)

Exploratory Ananylsis

```
summary(mtcars$mpg[mtcars$am=="automatic"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   14.95   17.30   17.15   19.20   24.40
```

```
summary(mtcars$mpg[mtcars$am=="manual"])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   21.00   22.80   24.39   30.40   33.90
```

```
m <- lm(mpg ~ am -1, mtcars)
summary(m)$coef
```

```
##              Estimate Std. Error  t value                 Pr(>|t|)
## amautomatic 17.14737   1.124603 15.24749 0.000000000000000113398345
## ammanual    24.39231   1.359578 17.94109 0.000000000000000001376283
```

```
# boxplot(mpg ~ am, data=mtcars, ylab = "Miles / Gallon")
# qplot(x=wt, y=mpg, color=am,  data=mtcars, ylab = "Miles / Gallon", xlab = "Weight (lbs/1000)")
t.test(mtcars$mpg[mtcars$am=="automatic"], mtcars$mpg[mtcars$am=="manual"])$p.value
```

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

The mean of miles per gallon looks higher for manual when it's looked as the only predictor (Ref: Box plot in the appendix). However we could observe from the scatter plot, as the other variables like weight and cylinders increase there is not much difference due between automatic and manual. From the t test pvalue its evident there is significance on transmission type with the mileage. We reject the null hypothesis and look for more details.

Model the Analysis

```
fit <- lm(mpg ~ ., data = mtcars)       # Fit a model with all variables with mpg as outcome
vif(fit)        # Get the variable inflation factor
```

```
##       cyl      disp        hp      drat        wt      qsec        vs
## 15.373833 21.620241  9.832037  3.374620 15.164887  7.527958  4.965873
##        am      gear      carb
##  4.648487  5.357452  7.908747
```

With the above we could see variables *cyl, wt, disp* has higher impact on mpg compared to other variables. We will model with those.

```
m1 <- lm (mpg ~ am, mtcars)             # Model m1 with mpg and Transmission type
m2 <- lm (mpg ~ am + wt , mtcars)       # Model m2 as m1 + variable weight
m3 <- lm (mpg ~ am + wt + cyl , mtcars) # Model m3 as m2 + variable cylinder
m4 <- lm (mpg ~ am + wt + cyl + disp, mtcars) # Model m4 as m3 + variable displacement
#anova(m1, m2, m3, m4) # Compare Models
```

Model m4 has a p value greater than 0.05 hence we choose the model 3 for our interpretation. *Refer to the Model summary for details in the appendix.*

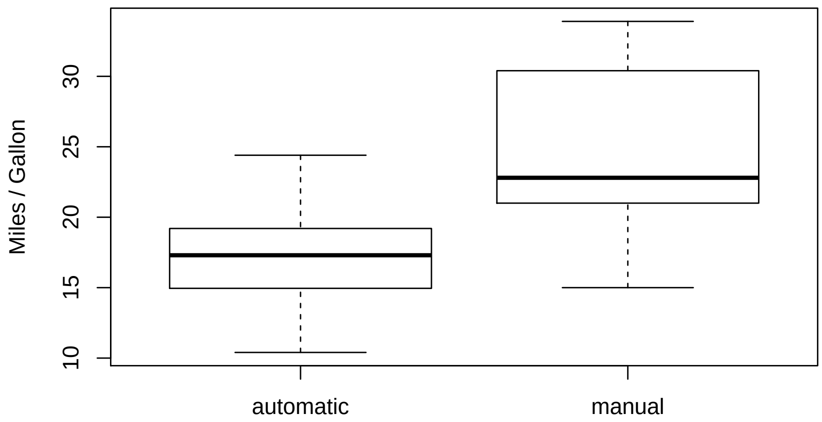
Conclusion

There was noticeable difference between manual and automatic transmission by mean (~7 mpg, Ref Appendix). Based on the linear model with impact of confounding variables *Weight & Cylinders* and a *95%* confidence interval *(-2.495555 2.848541)*, it seems Manual tranmission might yield about *0.17* miles more per gallon than a automatic. We could conclude the type of transmission yield similar mileage as the vehicle weighs more and houses a powerful engine.

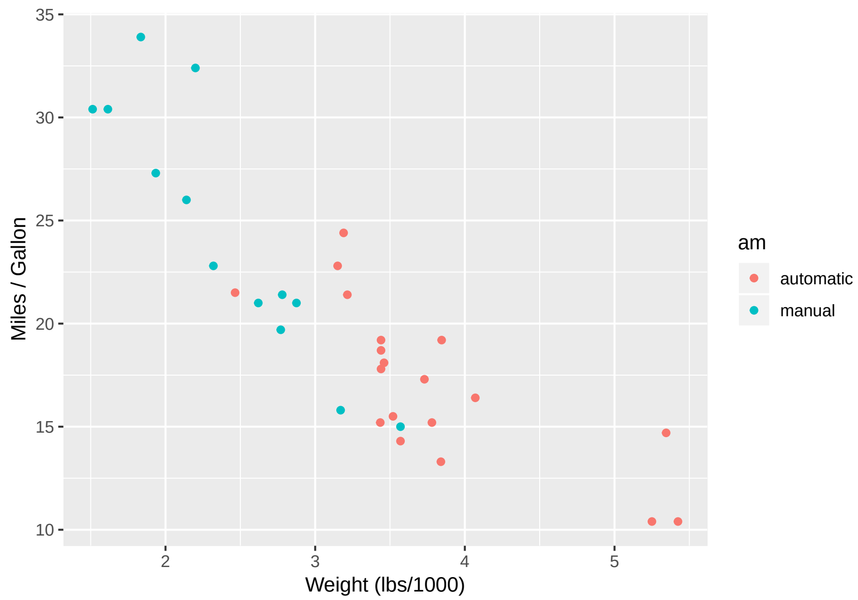
Appendix

Plot Graphs for Analysis

```r
boxplot(mpg ~ am, data=mtcars, ylab = "Miles / Gallon")
```




```r
qplot(x=wt, y=mpg, color=am,  data=mtcars, ylab = "Miles / Gallon", xlab = "Weight (lbs/1000)")
```



Compare Models

```r
anova(m1, m2, m3, m4) # Compare Models
```

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

```
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + cyl
## Model 4: mpg ~ am + wt + cyl + disp
##   Res.Df    RSS Df Sum of Sq       F        Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 63.4179 0.0000000001469 ***
## 3     28 191.05  1     87.27 12.5055      0.001488 **
## 4     27 188.43  1      2.62  0.3756      0.545093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Summary

```
summary(m1)$coef      # Model m1 with mpg and Transmission type
```

```
##              Estimate Std. Error   t value               Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 0.00000000000000113398
## ammanual     7.244939   1.764422  4.106127 0.00028502074393506777
```

```
summary(m2)$coef      # Model m2 as m1 + variable weight
```

```
##                Estimate Std. Error     t value              Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 0.0000000000005843477
## ammanual    -0.02361522  1.5456453 -0.01527855 0.9879145855484474659
## wt          -5.35281145  0.7882438 -6.79080719 0.0000001867415038084
```

```
summary(m3)      # Model m3 as m2 + variable cylinder
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + cyl, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##             Estimate Std. Error t value           Pr(>|t|)
## (Intercept)  39.4179     2.6415  14.923 0.00000000000000742 ***
## ammanual      0.1765     1.3045   0.135             0.89334
## wt           -3.1251     0.9109  -3.431             0.00189 **
## cyl          -1.5102     0.4223  -3.576             0.00129 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 0.0000000000651
```

```
summary(m4)$coef      # Model m3 as m2 + variable cylinder
```

```
##                 Estimate Std. Error    t value             Pr(>|t|)
## (Intercept) 40.898313414 3.60154037 11.3557837 0.000000000008677574
```

```
## ammanual     0.129065571 1.32151163  0.0976651 0.922919644373322634
## wt          -3.583425472 1.18650433 -3.0201537 0.005468412488207274
## cyl         -1.784173258 0.61819218 -2.8861142 0.007581533437721466
## disp         0.007403833 0.01208067  0.6128661 0.545092996564771504
```

```r
confint(m3)[2,] # Confidence Intervals for Model m3 at 95%
```

```
##      2.5 %     97.5 %
## -2.495555  2.848541
```

T Test Report

```r
t.test(mtcars$mpg[mtcars$am=="automatic"], mtcars$mpg[mtcars$am=="manual"])
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg[mtcars$am == "automatic"] and mtcars$mpg[mtcars$am == "manual"]
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

Residual Plots

```r
par(mfcol = c(2,2))
plot(m3)
```