Manual vs Auto: A Regression Model-based Analysis of Transmission Type on Fuel Efficiency

Aristide Chikando

Abstract

Automotive enthusiats and the entire automotive industry as a whole have long debated whether manual transmission provide better fuel efficiency than automatic transmission and vice versa as measured by miles per gallon (MPG). While it is well established that factors such as number of cylender (CYL), vehicle weight (WT), numbers of gears (GEAR), horse power (HP), amongst other factors, directly affect a vehicle mpg, the independent effect of transmission type (AM) is not well know. Here, *Motor Trend* has performed exploratory statistical analysis on the mtcars data set, a collection of car with the following two objectives: (1) can a statistically sound argument be made for an automatic or manaual transmission being better for MPG and (2) Is there a quantifiable MPG difference between automatic and Manual transmission? findings suggest that manual transmission holds a dicisive edge on fuel efficency over automatic transmission

Exploratory Analysis

```
library(ggplot2)
```

Here is the data set being analyzed

```
data(mtcars)
mtcars
```

```
##                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4          21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag      21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360         14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230           22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280           19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C          17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE         16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL         17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC        15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood 10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4  8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial  14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128           32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic        30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla     33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona      21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
```

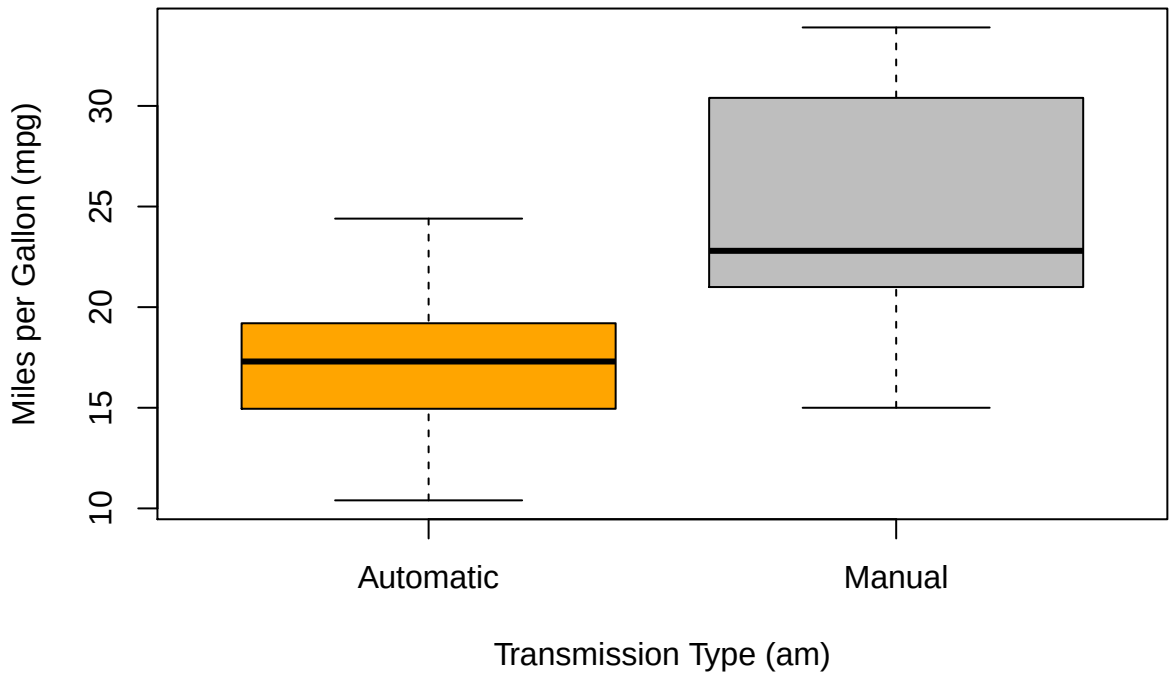
```
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

Data set consists of 32 observations of 11 variables

Does data reveal significant MPG difference between automatic and manual transmission?

Let's plot and compare the overall MPG in the two transmission type

```
mtcars$am = factor(mtcars$am, labels = c("Automatic", "Manual"))
boxplot(mtcars$mpg ~ mtcars$am, col = (c("orange", "gray")), ylab = "Miles per Gallon (mpg)", xlab = "Tr
```



Box plot reveals a noticeable MPG difference between automatic and manual transmission type vehicles in the data set.

Could we begin quantifying the difference in MPG between automatic and manual transmission type?

Lets look at the raw means

```
mpgDiff = aggregate(mpg~am, data = mtcars, mean)
mpgDiff[1,2] - mpgDiff[2,2]
```

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

Manual vehicles in the data set have 7.25 higher MPG efficiency than automatic vehicle.

Lets Assess the statistical significance of thsi MPG difference

```
AutoTrans = mtcars[mtcars$am == "Automatic",]
ManualTrans = mtcars[mtcars$am == "Manual",]
t.test(ManualTrans$mpg, AutoTrans$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  ManualTrans$mpg and AutoTrans$mpg
## 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:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

t-test p-value is less than 0.05 which implies strong statistical significance.

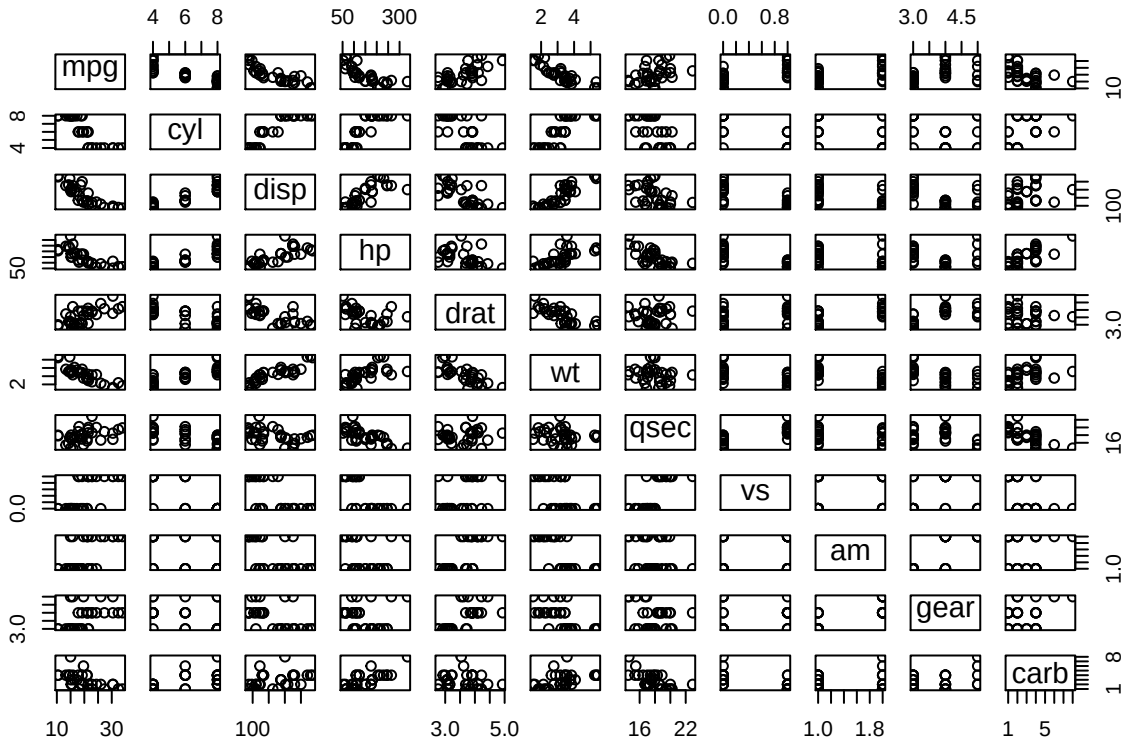
Lets use regression analysis (with MPG as predictor and transmission type as factor) to further compare how the MPG of the two transmission types stack up

```
fit1 = lm(mpg ~ am, data = mtcars)
summary(fit1)
```

```
##
## 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
```

Regression fit coefficient suggest that for every 17.15 mpg output decrease in automatic transmission, there is 7.245 mpg decrease in manual transmission. Nevertheless, $R^2 = 0.36$ suggest that the model only explains about a thrid of the variance. For a better exploration, it is thus worth performing a mutivariate analysis with other variables affecting mpg, as determined by a "pairs analysis", explicitely stated as co-factors.

```
pairs(mpg ~ ., data = mtcars)
```



The following varialbles actually have stronger correlation to MPG than transmission type and will be used in the multivariate analysis: cyl, disp, hp, drat, wt, and qsec.

```
fit2 = lm(mpg ~ am + cyl + disp + hp + drat + wt + qsec, data = mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + hp + drat + wt + qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4134 -1.6057 -0.3921  1.1730  4.6515
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.30919   15.98626   0.958  0.34779
## amManual     2.72022    1.78747   1.522  0.14112
## cyl         -0.34192    0.85076  -0.402  0.69131
## disp         0.01459    0.01177   1.240  0.22706
## hp          -0.02058    0.01521  -1.353  0.18877
## drat         0.81837    1.47926   0.553  0.58523
## wt          -3.99345    1.23289  -3.239  0.00349 **
## qsec         0.85996    0.58676   1.466  0.15573
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.492 on 24 degrees of freedom
## Multiple R-squared:  0.8676, Adjusted R-squared:  0.829
## F-statistic: 22.47 on 7 and 24 DF,  p-value: 4.407e-09
```

the mutivariate analysis reveals and even greater MPG efficiency for manual transmission over automatic

transmission when relevant co-factors are considered. P-value of nearly 0 with $R^2 = 0.87$ also implies strong statistical significance with higher scope of variance considered.

Lets verify that comparison of both model support these conclusions.

```
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl + disp + hp + drat + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     24 149.09  6    571.81 15.341 3.648e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova function p-value of less than 0.05 validates the multivariate analysis and upholds the conclusion that manual transmission vehicles in the given data set have significantly higher mpg efficiency than automatic transmission vehicles.