Motor Trend Data: How Mileage is impacted by Transmission

Trina Dey

13/09/2020

Executive Summary

We want to analyse the trends in the motor vehicles and looking at the data set of collections of cars, we are interested in exploring the relationship between a set of variables miles per gallon (MPG) as outcome. Primarily, we want to focus on two questions -
1. Is an automatic better than the manual transmission for MPG
2. Quantify the MPG difference between automatica and manual transmission

We found out that the Manual Transmission is better than the Automatic Transmission in all cases on an average 2.084 times, the details of which are given below in the report.

Exploratory Analysis

We are looking at mtcars dataset which comprises of 32 observations on 11 numeric variables but we are mainly interested in following -
1. mpg Miles/(US) gallon
2. cyl Number of cylinders
3. disp Displacement (cu.in.)
4. hp Gross horsepower
5. wt Weight (1000 lbs)

Let's look at the summary of the data for mpg wrt to Automatic Transmission vs Manual Transmission.

Summary of Automatic Transmission

```
summary(mtcars[mtcars$am==0,]$mpg)
```

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

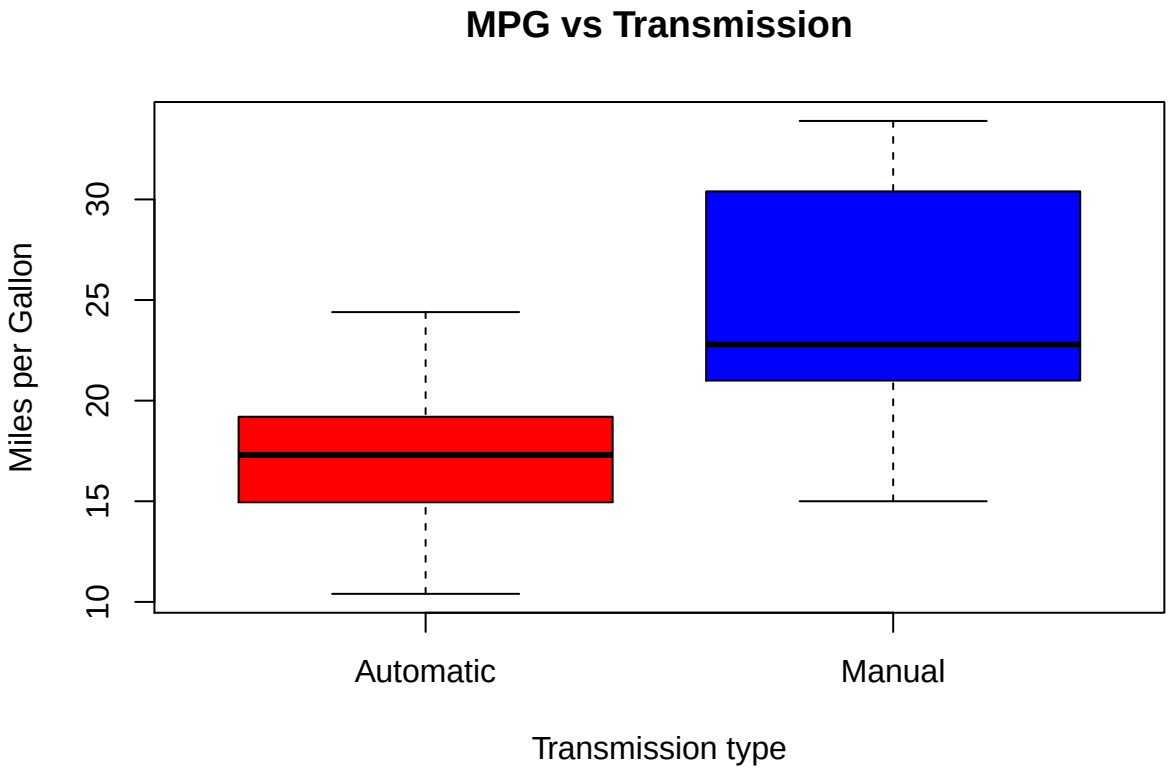
Summary of Manual Transmission

```
summary(mtcars[mtcars$am==1,]$mpg)
```

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

Clearly we can see that the values of MPG for Manual Transmission is higher than the Automatic Transmission. This difference can be seen via a box plot as well.

```r
boxplot(mpg ~am, data = mtcars, col = c("red","blue"), names=c("Automatic","Manual"),
        xlab="Transmission type", ylab = "Miles per Gallon", main = "MPG vs Transmission")
```



Regression Models

We are going to consider two regression models based on different predictors.
Model 1 - This predicts mpg based on variable am only.
Model 2 - This predicts mpg based on variables am, cyl, disp, hp, wt.

We are only going to look at the coefficients - Intercept and Slope. The detailed summary of the models can be found in the appendix section

Model 1

```r
mdl1 <- lm(mpg~am, data = mtcars)
coef(mdl1)
```

```
## (Intercept)          am
##   17.147368    7.244939
```

Clearly we see the slope to be `7.2449393` that predicts the Manual is better than the Automatic Transmission. Let's look at other models.

Model 2

```
mdl2 <- lm(mpg~am + cyl + disp + hp + wt, data = mtcars)
coef(mdl2)
```

```
## (Intercept)          am         cyl        disp          hp          wt
## 38.20279869  1.55649163 -1.10637984  0.01225708 -0.02796002 -3.30262301
```

We see that adding more predictors has clearly changed the intercept but the slope isnt varying as much as with predictor am only. The addition of other predictors underfits the model. Also, we can see from the summary of model 2 in appendix that the predictors cyl and disp are less significant. So we are going to remove these predictors from our model and use a model with predictors am, hp and wt only.

Model 3

```
mdl3 <- lm(mpg~am + wt + hp, data = mtcars)
coef(mdl3)
```

```
## (Intercept)          am          wt          hp
## 34.00287512  2.08371013 -2.87857541 -0.03747873
```

The above model has more significant codes which makes this model better than the model 1 or 2. This can be seen in the summary of model 3 in the appendix section.

Conclusion

To conclude, let's revisit the original two questions:

1. Is an automatic or manual transmission better for MPG? As we can see above from the MPG mean and all three regression models that manual transmission is clearly better for MPG than automatic transmission, by varying degrees.
2. Quantify the difference between automatic and manual transmissions. Our regression models have quantified how much better manual transmission is than automatic transmission by three different amounts:

Slope From Model 1 is `7.2449393`
Slope From Model 2 is `2.0837101`
Slope From Model 3 is `2.0837101`
Since mdl 3 is the most reliable from the three models, we can say with confidence that manual transmission is better than automatic transmission by approximately 2.084 on average. All values in this model are statistically significant and account for ~98% of the variance. Looking at the residual plots (which can be found in the appendix), there does not appear to be any apparent patterns or outliers that create an undue influence on the prediction.

Appendix

Below are the summary of the three models and some residual plots for the model 3

Model 1

```
##
## 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
```

Model 2

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + hp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5952 -1.5864 -0.7157  1.2821  5.5725
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.20280    3.66910  10.412 9.08e-11 ***
## am           1.55649    1.44054   1.080  0.28984
## cyl         -1.10638    0.67636  -1.636  0.11393
## disp         0.01226    0.01171   1.047  0.30472
## hp          -0.02796    0.01392  -2.008  0.05510 .
## wt          -3.30262    1.13364  -2.913  0.00726 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.505 on 26 degrees of freedom
## Multiple R-squared:  0.8551, Adjusted R-squared:  0.8273
## F-statistic:  30.7 on 5 and 26 DF,  p-value: 4.029e-10
```

Model 3

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

Plot for residuals for Model 3

```
par(mfrow = c(2, 2))
plot(mdl3)
```