Effect of Transmission type in cars on fuel efficiency

Aman Bhagat

5/19/2020

Executive Summary

This report is prepared as a part of coursework required for the Coursera Regression Model course. The instruction for this report assigmnet state as follows:

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"

For this work we will use mtcars dataset to investigate the required the task.

Data Description

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

Usage mtcars Format A data frame with 32 observations on 11 (numeric) variables.

[, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (1000 lbs) [, 7] qsec 1/4 mile time [, 8] vs Engine (0 = V-shaped, 1 = straight) [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears [,11] carb Number of carburetors Note Henderson and Velleman (1981) comment in a footnote to Table 1: 'Hocking [original transcriber]'s noncrucial coding of the Mazda's rotary engine as a straight six-cylinder engine and the Porsche's flat engine as a V engine, as well as the inclusion of the diesel Mercedes 240D, have been retained to enable direct comparisons to be made with previous analyses.'

Source Henderson and Velleman (1981), Building multiple regression models interactively. Biometrics, 37, 391–411.

```
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
```

```
amf <- as.factor(mtcars$am)
cylf <- as.factor(mtcars$cyl)
```

Data analyses

We are interested in mainly the transmission gear types and their effect in the fuel efficiency.

```
amf <- as.factor(mtcars$am)
g <- ggplot(data = mtcars, aes(x = amf ,y = mpg, fill = amf))
g <- g + geom_violin()
g <- g + geom_boxplot(width=0.1)
g <- g + scale_fill_brewer(palette="Dark2")
g
```
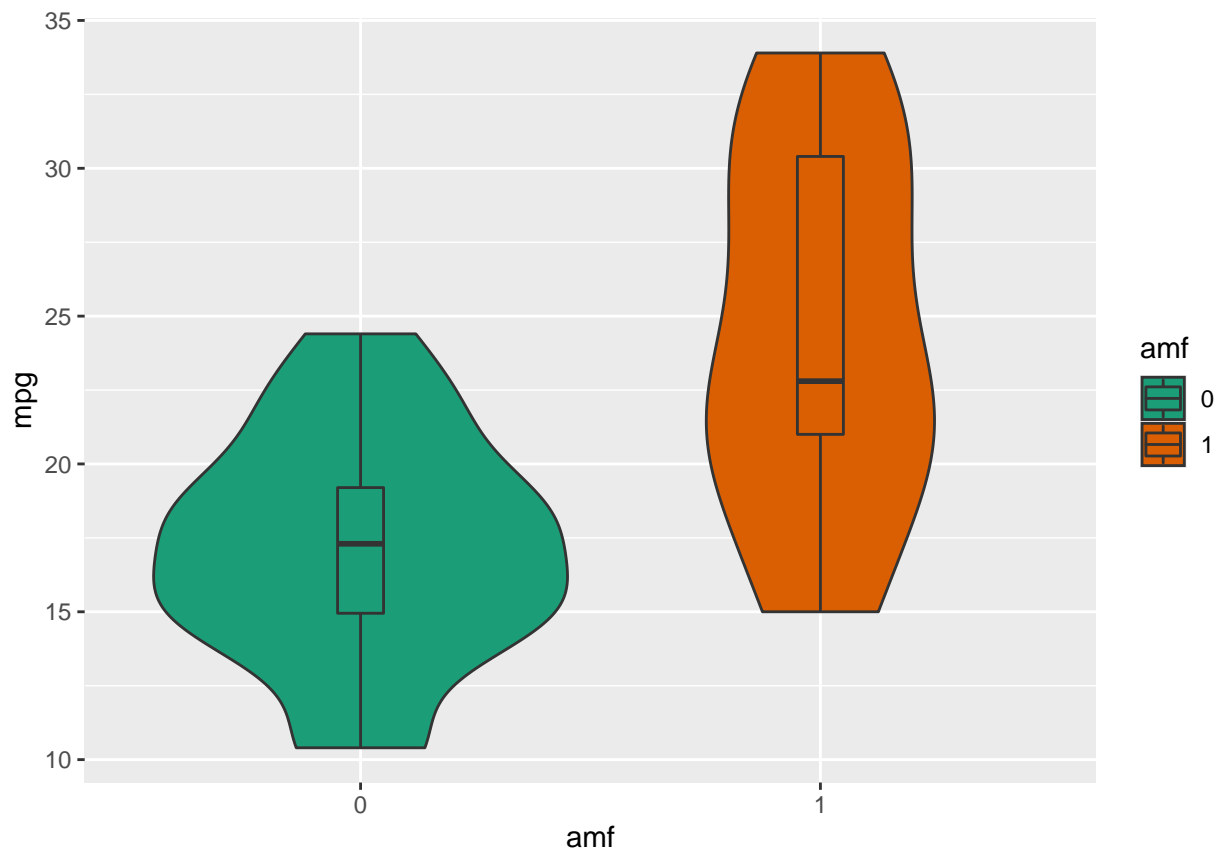

Appendix A:

From the above plot we can infer that the manual transmission which is denoted as "1" is more efficeint than the auto transmission.

```
meanMpg <- mtcars %>% group_by(am) %>% summarize(meanMpg = mean(mpg))
meanMpg <- data.frame(meanMpg)
meanMpg
```

```
##    am   meanMpg
## 1   0  17.14737
## 2   1  24.39231
```

By calculating the average value from each transmission, it shows on an average manual gear gives more miles per gallon as compared to the auto transmission gear.

```r
sdMpg <- mtcars %>% group_by(am) %>% summarize(sdMpg = sd(mpg))
sdMpg <- data.frame(sdMpg)
sdMpg
```

```
##    am    sdMpg
## 1   0  3.833966
## 2   1  6.166504
```

We can also observe that the standard deviation from the mean of the data is quiet significant in the case of manual transmission which indicates that the collected data may not be the actual representation of the population data.

Hyposthesis Testing

In this report we are interested in finding out whether the change in the fuel efficiency is significant in both the case.

Lets assume:

H_0 : There is no significant difference in the fuel consumption in between Transmission type.
H_a : There is a significant difference in the fuel consumption in between Transmission type.

```r
t.test(mpg~amf, paired = FALSE, data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by amf
## 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 in group 0 mean in group 1
##        17.14737        24.39231
```

From the above test values we can infer that the p-value is very less than the significance level which is 0.05 therefore we reject our null hypothesis. And we can say that the difference in the fuel consumption for the two transmission type is significant. By looking at the negative sign in the confidence interval, it shows manual transmission is better than automatic transmissioon in terms of fuel efficiency.

Regression Analyses

Now, we apply various linear regression models on different variables to get the significant impact on miles per gallon by different predictor including transmission types.

```r
fit1 <- lm(mpg~ amf,data = mtcars)
fit2 <- update(fit1 , mpg~ am + cylf)
fit3 <- update(fit2 , mpg~ am + cylf + disp)
fit4 <- update(fit3 , mpg~ am + cylf + disp + hp)
fit5 <- update(fit4 , mpg~ am + cylf + disp + hp + drat)
fit6 <- update(fit5 , mpg~ am + cylf + disp + hp + drat + wt)
fit7 <- update(fit6 , mpg~ am + cylf + disp + hp + drat + wt + qsec)
fit8 <- update(fit7 , mpg~ am + cylf + disp + hp + drat + wt + qsec + vs)
fit9 <- update(fit8 , mpg~ am + cylf + disp + hp + drat + wt + qsec + vs + gear)
fit10 <- update(fit9 ,mpg~ am + cylf + disp + hp + drat + wt + qsec + vs + gear + carb)
anova(fit1, fit2, fit3,fit4,fit5,fit6,fit7,fit8,fit9,fit10)
```

```
## Analysis of Variance Table
##
## Model  1: mpg ~ amf
## Model  2: mpg ~ am + cylf
## Model  3: mpg ~ am + cylf + disp
## Model  4: mpg ~ am + cylf + disp + hp
## Model  5: mpg ~ am + cylf + disp + hp + drat
## Model  6: mpg ~ am + cylf + disp + hp + drat + wt
## Model  7: mpg ~ am + cylf + disp + hp + drat + wt + qsec
## Model  8: mpg ~ am + cylf + disp + hp + drat + wt + qsec + vs
## Model  9: mpg ~ am + cylf + disp + hp + drat + wt + qsec + vs + gear
## Model 10: mpg ~ am + cylf + disp + hp + drat + wt + qsec + vs + gear +
##     carb
##    Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1      30 720.90
## 2      28 264.50  2    456.40 34.2326 3.488e-07 ***
## 3      27 230.46  1     34.04  5.1057   0.03516 *
## 4      26 183.04  1     47.42  7.1136   0.01480 *
## 5      25 182.38  1      0.66  0.0987   0.75664
## 6      24 150.10  1     32.28  4.8425   0.03968 *
## 7      23 141.21  1      8.89  1.3343   0.26166
## 8      22 139.02  1      2.18  0.3275   0.57354
## 9      21 135.27  1      3.75  0.5629   0.46183
## 10     20 133.32  1      1.95  0.2921   0.59485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We basically fitted our model with different set of predictors and increased our degree of freedom by 1. So from the above analyses we are able to see that there are three models in particular which are significant by looking at the probability which is less than 5%. The fits are fit2, fit4 and fit6. Further if we investigate, we see that with inclusion of certain predictor making these fits more significant which are weight,horsepower and cylinder. So we futher fit new models to find out the best model for the problem.

```r
fit21 <- update(fit1 , mpg~ amf + cylf)
fit22 <- update(fit21, mpg ~ amf + cylf + wt)
fit23 <- update(fit21, mpg ~ amf + cylf + wt + hp)
anova(fit1, fit21,fit22,fit23)
```

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

```
## Model 1: mpg ~ amf
## Model 2: mpg ~ amf + cylf
## Model 3: mpg ~ amf + cylf + wt
## Model 4: mpg ~ amf + cylf + wt + hp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 264.50  2    456.40 39.2861 1.388e-08 ***
## 3     27 182.97  1     81.53 14.0354 0.0009026 ***
## 4     26 151.03  1     31.94  5.4991 0.0269346 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above ananlyses we can see that fit21 is more significant and it has probability very less that our significance probability which means by keeping the cylinder variable constant we can interpret our coefficient more effectively.Now lets see at the coefficients of that model:

```r
summary(fit21)$coef
```

```
##                 Estimate Std. Error    t value       Pr(>|t|)
## (Intercept)    24.801852   1.322615  18.752135   2.182425e-17
## amf1            2.559954   1.297579   1.972869   5.845717e-02
## cylf6          -6.156118   1.535723  -4.008612   4.106131e-04
## cylf8         -10.067560   1.452082  -6.933187   1.546574e-07
```

Therefore we can interpret our coefficient as follows:

Intercept: The automatic transmission gear gives approx 24.8019 mpg in the case where the car has 4 cylinders while keeping other variables constant.
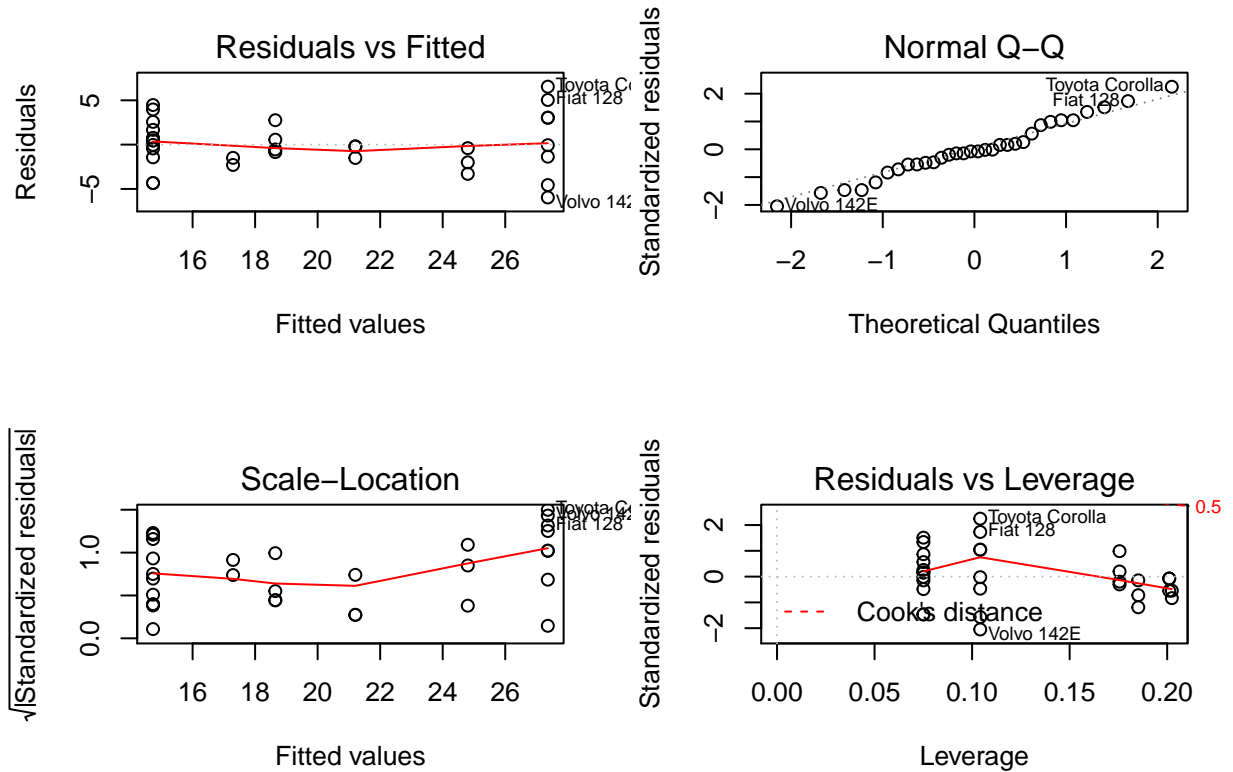
amf1: It states that there is an increase of 2.56 in the mpg as compared to the automatic transmission type if the car has manual transmission gear and while keeping the other variables constant.

cylf6 The coefficient of this variable says that there is a decrease in mpg by 6.1561 as compared to the cars which have 4 cylinders.

cylf8 The coefficient of this variable says that there is a decrease in mpg by 10.0675 as compared to the cars which have 4 cylinders.

Diagnostic Plots

```r
par(mfrow = c(2,2))
plot(fit21)
```

Residuals vs Fitted

Residuals

Toyota Co
Fiat 128
Volvo 142

16 18 20 22 24 26

Fitted values

Normal Q–Q

Standardized residuals

Toyota Corolla
Fiat 128
Volvo 142E

−2 −1 0 1 2

Theoretical Quantiles

Scale–Location

√|Standardized residuals|

Toyota C
Fiat 128

16 18 20 22 24 26

Fitted values

Residuals vs Leverage

Standardized residuals

0.5

Toyota Corolla
Fiat 128

- - - Cook's distance

Volvo 142E

0.00 0.05 0.10 0.15 0.20

Leverage

Appendix B

Above residual plots indicates that our model fits well as there is no significant patterns in the residual plot which infer something unusual. Therefore residual variance doesn't follow any significant variance which interferes with our model fit.

Summary

In this particular project we are interested in finding out which transmission is better for MPG(Miles per gallon). From our hypothesis test; we found out that there is a significant difference between the gear types to mpg which also indicated that the manual transmission produces more miles per gallon.

Further we investigated to fit a model in the data and found out that inclusion of the predictor cylinder gives us more insight to the data and more accurate interpretation of the data. By which we found out that with the increase in the number of cylinders in a car reduces miles per gallon. Therefore we found out that there is a 2.56 rise in mpg of manual transmission as compared to the automatic transmission for the cars which have 4 cylinders.

Appendix

Appendix	Defination
Appendix A:	Violon Plot between transmission type and mpg
Appendix B:	Various Residual plots for the diagnosis of the model