

The Future of Transmissions : Automatic or Manual?

Vedant Mane

June 25, 2020

Synopsis

The dataset is a collection of cars that the **Motor Trend** *magazine* are exploring in order to find the relationship between the set of variables and miles per gallon(MPG). The magazine is particularly interested in analysing whether automatic or manual transmission is better for MPG and wants to quantify the difference between them.

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).

Source

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

Loading Libraries

Loading the libraries we will need for further modelling and analysing the dataset.

```
library(datasets)
library(graphics)
library(MASS)
library(ggplot2)
```

Exploring the Dataset

The dataset appears to have all the numeric values for the attributes which represent the characteristics of a car and so we will need to perform some transformations in our dataset so that instead of analysing our data as numeric values we will analyse them as factor variables.

```
## 'data.frame':   32 obs. of  12 variables:
## $ mpg          : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
## $ cyl          : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
## $ disp         : num  160 160 108 258 360 ...
## $ hp           : num  110 110 93 110 175 105 245 62 95 123 ...
## $ drat         : num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt           : num  2.62 2.88 2.32 3.21 3.44 ...
```

```
## $ qsec      : num  16.5 17 18.6 19.4 17 ...
## $ vs        : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
## $ am        : Factor w/ 2 levels "0","1": 2 2 2 1 1 1 1 1 1 1 ...
## $ gear      : num   4 4 4 3 3 3 3 4 4 4 ...
## $ carb      : num   4 4 1 1 2 1 4 2 2 4 ...
## $ transmission: Factor w/ 2 levels "automatic","manual": 2 2 2 1 1 1 1 1 1 1 ...
```

Linear Model of the *MTCARS* Dataset

Let us first check which of the variables are correlated and which ones are uncorrelated with the MPG values which we are trying to predict. We will exclude the intercept from our model so that we get a clear idea about the data.

We know that including more variables potentially results in a biased prediction from our model but excluding important variables may lead to huge standard errors from our prediction model.

```
fitALL <- lm(mpg ~ . -1, data = mtcars)
summary(fitALL)$coef
```

```
##      Estimate Std. Error    t value    Pr(>|t|)
## cyl4 17.81984325 16.30602324  1.09283809 0.28745417
## cyl6 16.15953652 15.15088510  1.06657376 0.29887089
## cyl8 19.45728305 15.14640165  1.28461423 0.21360750
## disp  0.01391241  0.01740176  0.79948296 0.43340363
## hp    -0.04612835  0.02712018 -1.70088685 0.10446190
## drat  0.02635025  1.67648954  0.01571752 0.98761549
## wt    -3.80624757  1.84664309 -2.06117121 0.05252853
## qsec  0.64695710  0.72195025  0.89612421 0.38084614
## vs1   1.74738689  2.27267212  0.76886889 0.45095593
## am1   2.61726546  2.00474936  1.30553251 0.20653091
## gear  0.76402917  1.45668015  0.52450029 0.60569589
## carb  0.50935118  0.94244181  0.54045902 0.59484874
```

We see that as we have included all the variables in our prediction model, we have obtained huge standard errors which makes our model not suitable for prediction.

Linear Model MPG w.r.t Transmission Type

Let us directly address the question we are trying to answer by fitting a model that gives us the MPG value(outcome) by analysing the Type of Transmission of the car.

```
fitTransmission <- lm(mpg ~ am -1, data = mtcars)
summ <- summary(fitTransmission)$coef
summ
```

```
##      Estimate Std. Error    t value    Pr(>|t|)
## am0 17.14737    1.124603 15.24749 1.133983e-15
## am1 24.39231    1.359578 17.94109 1.376283e-17
```

```
manual <- summ[1,1] + c(-1,1) * qt(0.975, df.residual(fitTransmission)) * summ[1,2]
automatic <- summ[2,1] + c(-1,1) * qt(0.975, df.residual(fitTransmission)) * summ[2,2]
diff <- -(summ[1,1] - summ[2,1])
```

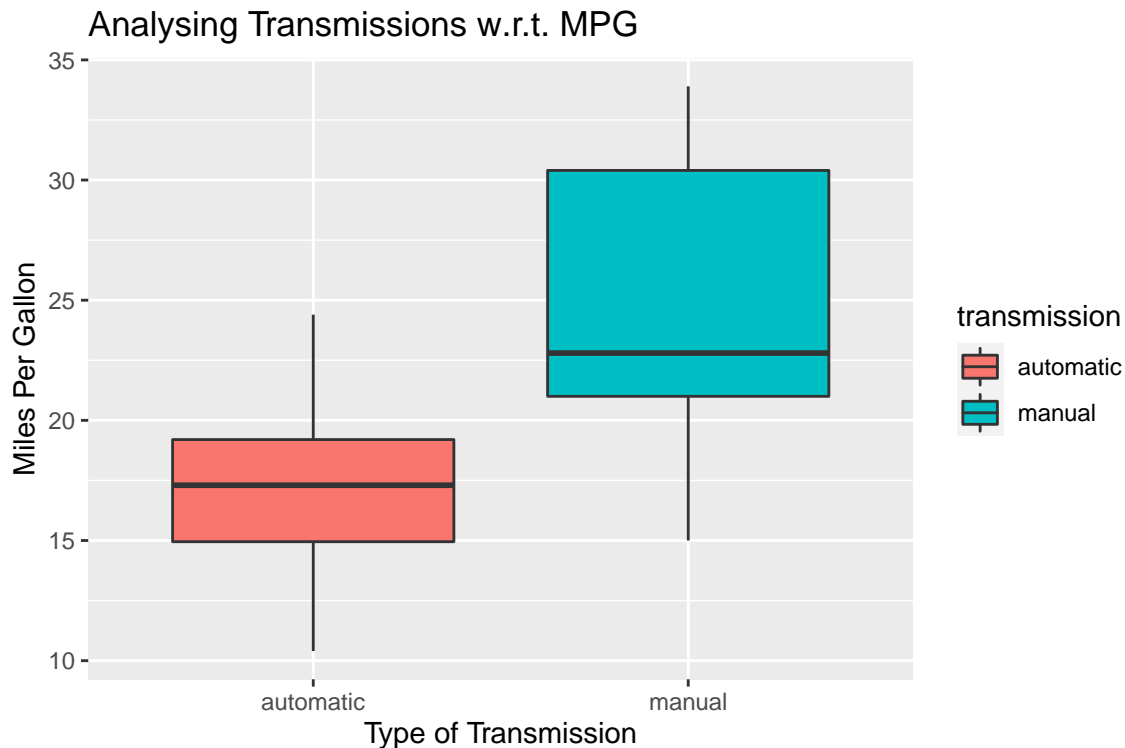
Initial assessment indicates that the **Type of transmission** has a significant impact on the **MPG** values.

Our analysis shows with 95% confidence that the cars with manual transmission have MPG values in the range **15.1145426, 19.1801942**, while cars with automatic transmission have MPG values in the range **22.3724469, 26.4121685**.

The Linear Regression Model estimates around **7.245** increase in MPG value for cars with Manual Transmission with that of Automatic Transmission.

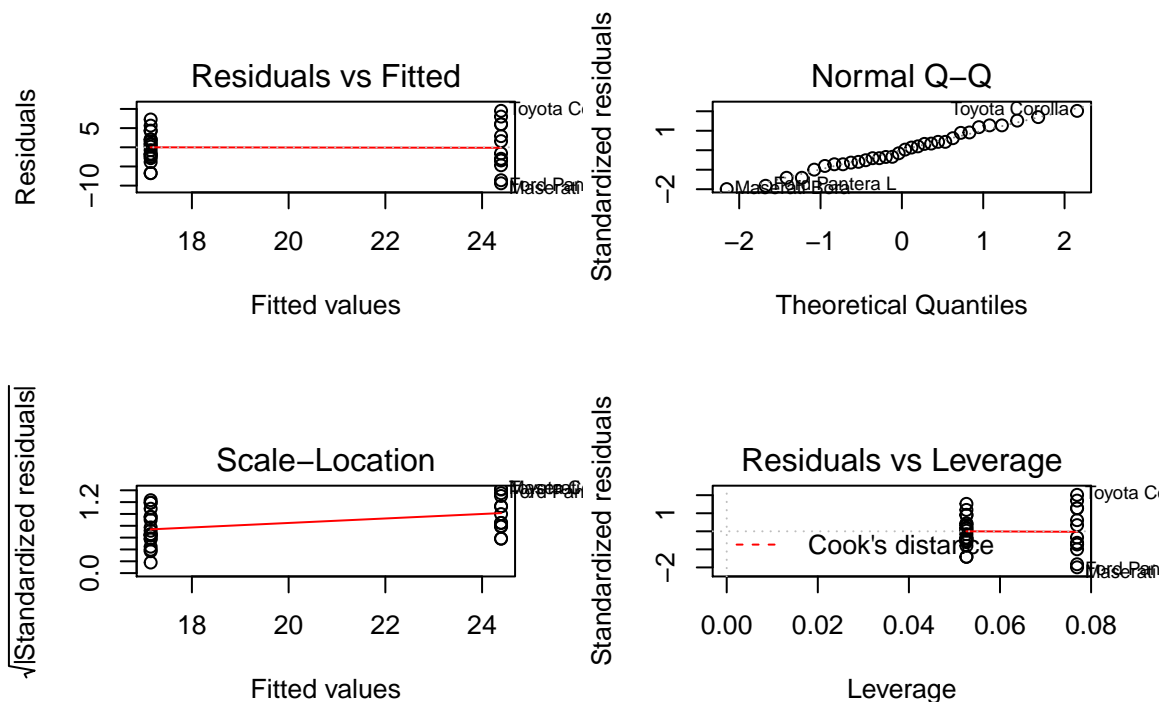
Exploring the MPG variable

```
ggplot(mtcars, aes(x = transmission, y = mpg)) + geom_boxplot(aes(fill = transmission)) +  
  xlab("Type of Transmission") + ylab("Miles Per Gallon") +  
  ggtitle("Analysing Transmissions w.r.t. MPG")
```



The plot proves that the relationship we have estimated from our model is true and that the type of transmissions causes a significant positive impact on MPG.

```
dfT <- dfbeta(fitTransmission)  
hatT <- hatvalues(fitTransmission)  
par(mfrow = c(2,2))  
plot(fitTransmission)
```



Though initial assessment of the model proves true, after diagnostics we can say that we are close to predicting the correct output. Let's see if we can do better.

Linear Model Selecting Best Variable for Predicting MPG

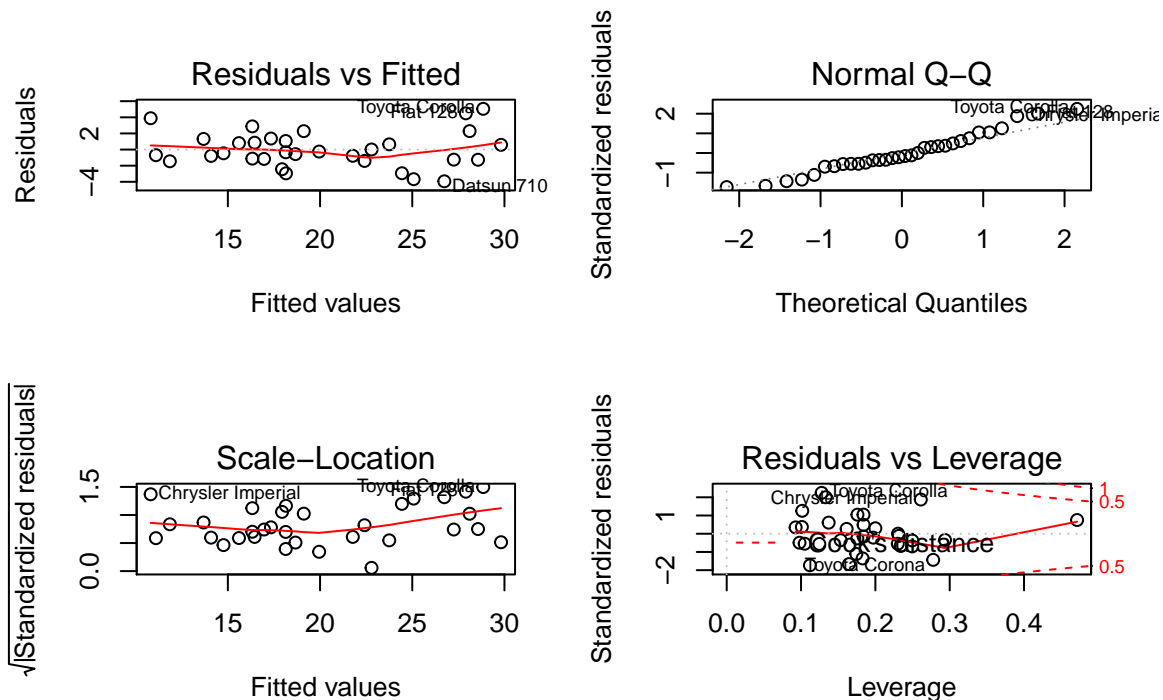
Now, let us see whether any other variables help us improve our model and reducing the standard error rate. We can do this using the STEPWISE function which automatically fits our model with the relevant attributes to accurately estimate the outcome and reduce the standard error in prediction.

```
fitSelective <- stepAIC(fitALL, direction = "both", trace = FALSE)
sumn <- summary(fitSelective)$coef
sumn
```

```
##      Estimate Std. Error  t value    Pr(>|t|)
## cyl4 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6 30.67697941 3.10834881  9.869220 2.786700e-10
## cyl8 31.54464858 3.88460652  8.120423 1.335191e-08
## hp   -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt   -2.49682942 0.88558779 -2.819404 9.081408e-03
## am1   1.80921138 1.39630450  1.295714 2.064597e-01
```

We see that the function chooses the following variables while fitting our model: **am**, **cyl**, **hp**, **wt**

```
dfS <- dfbeta(fitSelective)
hatS <- hatvalues(fitSelective)
par(mfrow = c(2,2))
plot(fitSelective)
```



Our diagnostics show that when we add a few variables to our model, the impact of the type of transmission lowers down to **1.809**.

Conclusion

Let us compare our models:

```
anova(fitTransmission, fitSelective, fitALL)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am - 1
## Model 2: mpg ~ cyl + hp + wt + am - 1
## Model 3: mpg ~ (cyl + disp + hp + drat + wt + qsec + vs + am + gear +
##          carb + transmission) - 1
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      30 720.90
## 2      26 151.03  4    569.87 21.3717 5.464e-07 ***
## 3      20 133.32  6     17.70  0.4426  0.8416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that when comparing our models, the p-value of the Selective Models drops with respect to the Transmission Model which states that while Manual Transmissions are significant when determining MPG, it has lower value significance when adding Number of Cylinders, Horse Power and Weight.