Regression Models Course Project

Antonio Marquez Palacios

January 30, 2018

Executive Summary

A Regression Model Analyses were made on the Motor Trend data exposed in 1974, where it was found Manual Transmission gives more Miles Per Gallon than Automatic. Then the model was extended to include other regressors that were demonstrated are correlated to the transmission mode (number of cylinders and weight) when meassuring the miles per gallon consumption. Results show that transmission by itself are not deterministic on the evaluation for mpg, but considering together number of cylinders and weight are, having a better consumption (more miles per gallon) with less cylinders and weight.

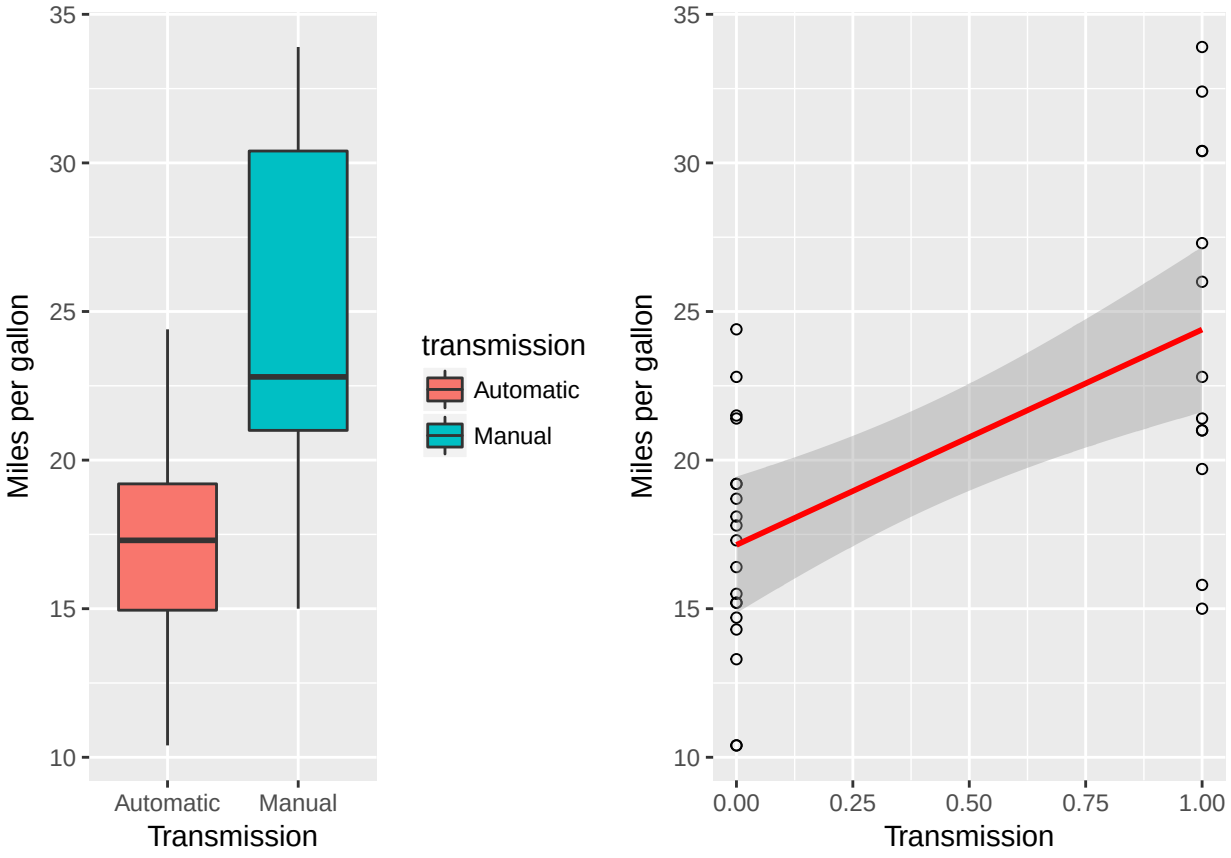
Exploratory Analyses

```
## '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         : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ 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          : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am          : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ 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 ...
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb          transmission
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000   Automatic:19  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000   Manual   :13  
##  Median :0.0000   Median :4.000   Median :2.000                 
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812                 
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000                 
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
##                   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
##                   transmission
## Mazda RX4               Manual
## Mazda RX4 Wag           Manual
## Datsun 710              Manual
## Hornet 4 Drive       Automatic
## Hornet Sportabout    Automatic
## Valiant              Automatic
```

The following plots show how the miles per gallon are related to the transmission mode. It can be seen how the automatic mode gives fewer mpg than manual.



Let's see how these are related by doing the initial Linear Model as

$$Y = \beta_0 + \beta_1 X$$

where Y is the outcome mpg and X the transmission mode

From the above plot it can be seen there is a positive slope for our β_1 coefficient. Let's explore more deeply this model in the following section.

Regresion model for the Miles per Gallon consumption against the Transmission mode

```
fit <- lm(mpg ~ transmission, data=mtcars)
coefficients(fit)
```

```
##        (Intercept) transmissionManual
##          17.147368           7.244939
```

Residuals Analisys

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##          -3.3923077          -3.3923077          -1.5923077
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##           4.2526316           1.5526316           0.9526316
##          Duster 360           Merc 240D            Merc 230
##          -2.8473684           7.2526316           5.6526316
##            Merc 280           Merc 280C          Merc 450SE
##           2.0526316           0.6526316          -0.7473684
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##           0.1526316          -1.9473684          -6.7473684
## Lincoln Continental   Chrysler Imperial            Fiat 128
##          -6.7473684          -2.4473684           8.0076923
##         Honda Civic      Toyota Corolla       Toyota Corona
##           6.0076923           9.5076923           4.3526316
##    Dodge Challenger         AMC Javelin          Camaro Z28
##          -1.6473684          -1.9473684          -3.8473684
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
##           2.0526316           2.9076923           1.6076923
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##           6.0076923          -8.5923077          -4.6923077
##       Maserati Bora          Volvo 142E
##          -9.3923077          -2.9923077
```

Let's look if there is any outliers calculating the dfbetas:

$$\hat{b}_k - \hat{b}_{ki}/\sqrt{(SE_i C_{kk})}$$

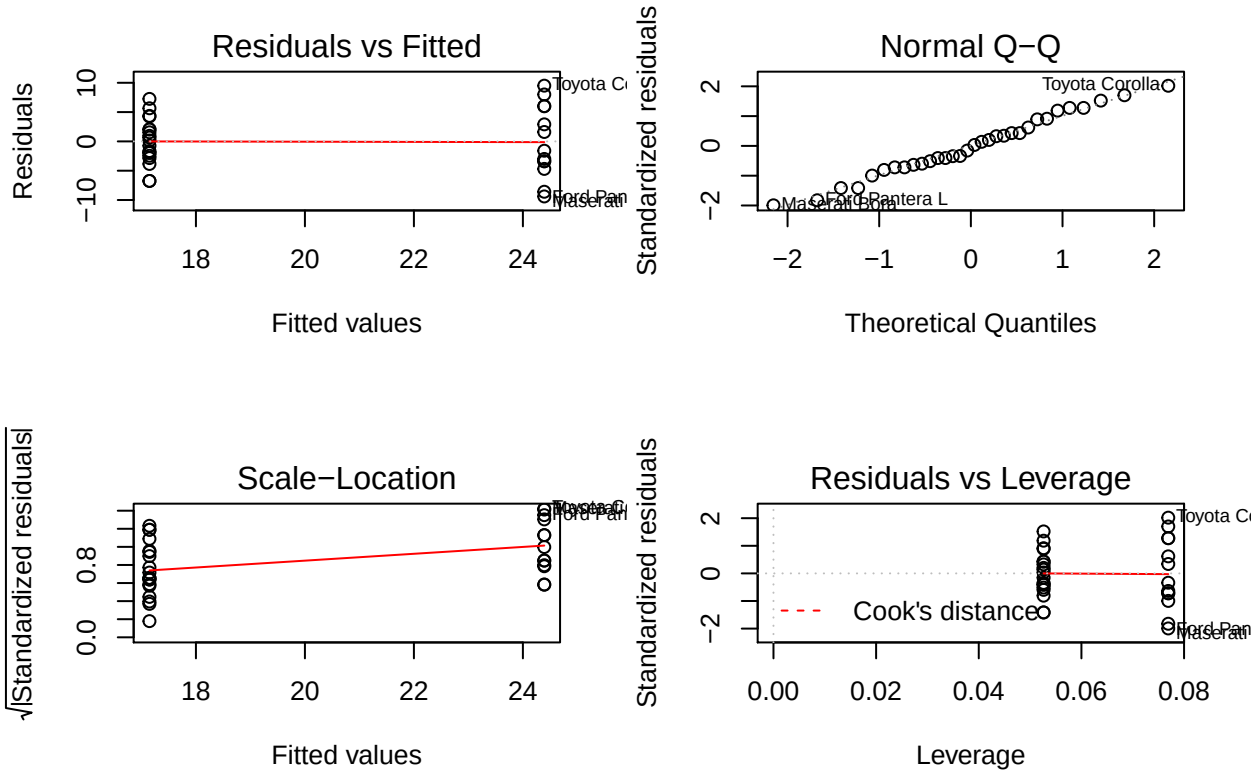
```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##              -0.159              -0.159              -0.074
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##              -0.133              -0.048              -0.030
##          Duster 360           Merc 240D            Merc 230
##               0.089              -0.234              -0.179
##            Merc 280           Merc 280C          Merc 450SE
##              -0.064              -0.020               0.023
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##              -0.005               0.060               0.216
## Lincoln Continental   Chrysler Imperial            Fiat 128
##               0.216               0.076               0.391
##         Honda Civic      Toyota Corolla       Toyota Corona
##               0.287               0.475              -0.137
##    Dodge Challenger         AMC Javelin          Camaro Z28
##               0.051               0.060               0.120
```

```
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
##              -0.064               0.136               0.075
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##               0.287              -0.423              -0.222
##       Maserati Bora          Volvo 142E
##              -0.468              -0.140
```

There are multiple values that double others, so lets do a hatvalues test to verify:

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710
##               0.077               0.077               0.077
##      Hornet 4 Drive   Hornet Sportabout             Valiant
##               0.053               0.053               0.053
##          Duster 360           Merc 240D            Merc 230
##               0.053               0.053               0.053
##            Merc 280           Merc 280C           Merc 450SE
##               0.053               0.053               0.053
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood
##               0.053               0.053               0.053
## Lincoln Continental   Chrysler Imperial            Fiat 128
##               0.053               0.053               0.077
##         Honda Civic      Toyota Corolla       Toyota Corona
##               0.077               0.077               0.053
##    Dodge Challenger         AMC Javelin          Camaro Z28
##               0.053               0.053               0.053
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2
##               0.053               0.077               0.077
##        Lotus Europa      Ford Pantera L        Ferrari Dino
##               0.077               0.077               0.077
##       Maserati Bora          Volvo 142E
##               0.077               0.077
```

As noted, hatvalues is pretty normal and does not show any potential outliers in out model. Clearly it can be seen no outliers are impacting the model. Finally let's plot the residuals.



Fitting multiple models

Let's extend the linear regression model to include more regressors:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$Y = \beta_0 + \beta_1 X + \beta_2 X_2 + \beta_3 X_3$$

where

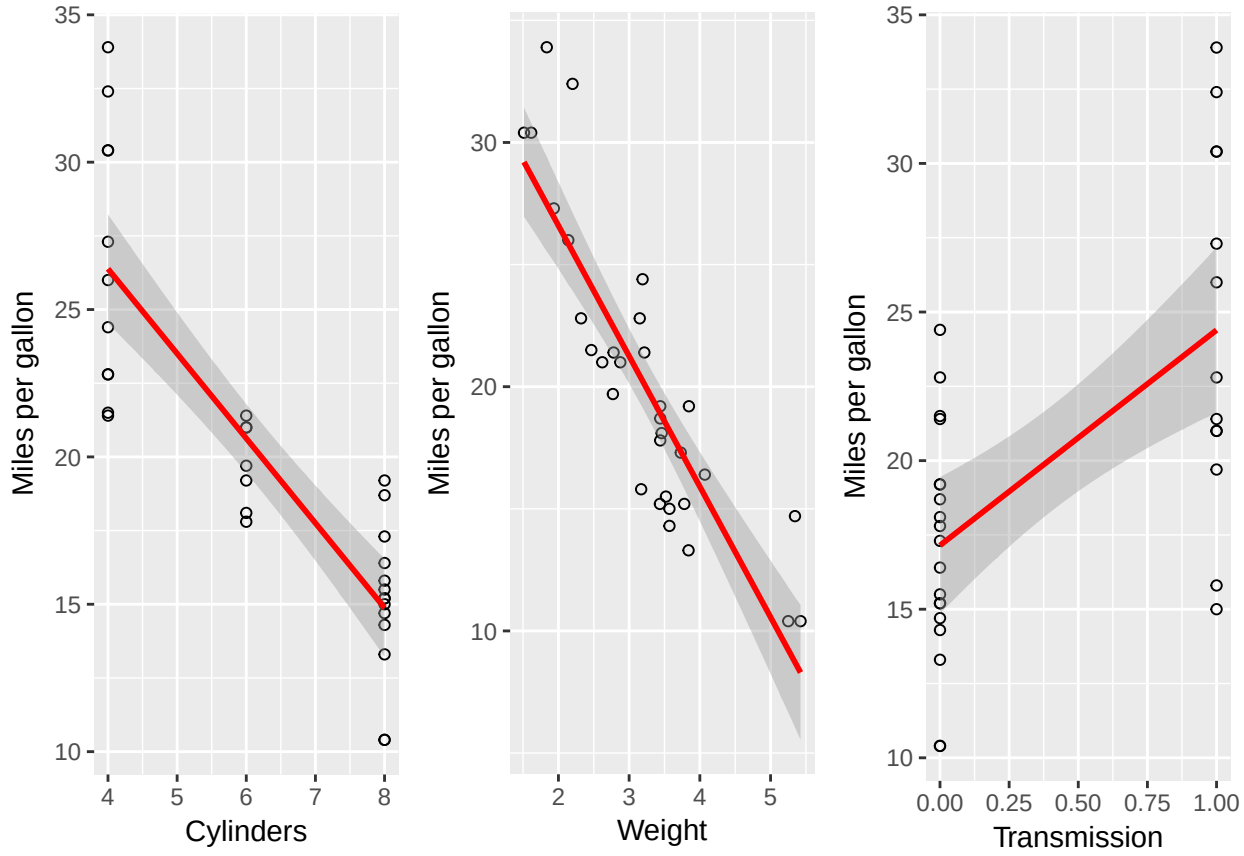
$$X_1 = am, X_2 = cyl, X_3 = wt$$

, and get the Variance Inflation Factor (VIF) of each:

```
## transmission          cyl
##     1.375739     1.375739
```

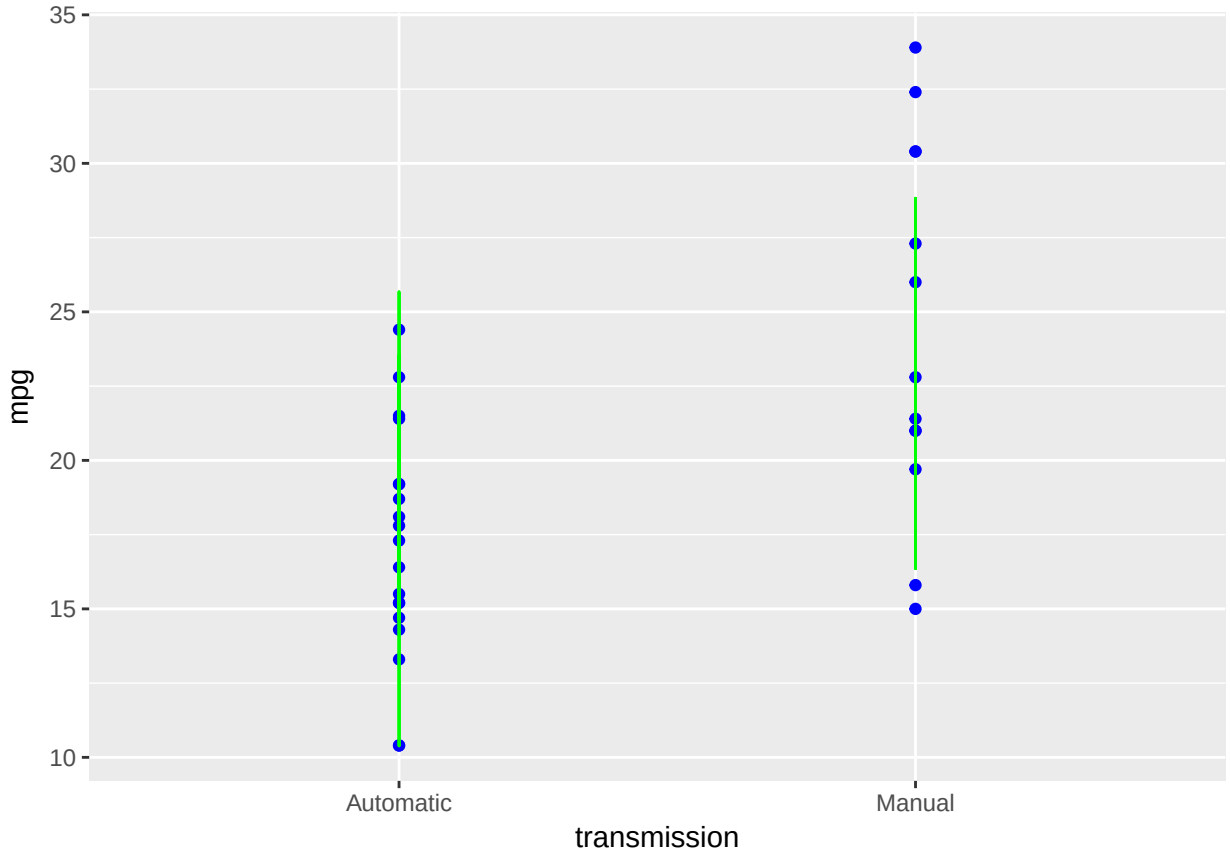
```
## transmission          cyl           wt
##     1.924955     2.584066     3.609011
```

The VIF data shows there is a relationship between cyl and weight on the miles per gallon outcome, see how these behave separately:



Now, consider out model with the three regressors

```
##
## Call:
## lm(formula = mpg ~ transmission + cyl + wt, 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 7.42e-15 ***
## transmissionManual   0.1765     1.3045   0.135  0.89334
## cyl                 -1.5102     0.4223  -3.576  0.00129 **
## wt                  -3.1251     0.9109  -3.431  0.00189 **
## ---
## 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: 6.51e-11
```



From the plot above it can be seen that Miles per Gallon is greater for Manual Transmissions than Automatic, even including other factors such as Number of Cylinders and the Weight of the car

Appendix

About data set

Motor Trend 1974 fuel comparison is available at the mtcars data frame. From the ?mtcars reference page, the following can be remarked:

mtcars is a data frame with 32 observations on 11 variables with the following structure

where the column descriptions are:

[, 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 V/S
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors
[,12] Transmission Mode. Column created based on column 9, for graphics and plotting purposes on this project