Analysis of Transimission Type impact on Mileage (MPG) using mtcars data set

Sreenivasulu Parimi

June 12, 2018

Coursera: Regression Models - Course Project

Executive Summary

This is a report prepared as part of the course assignment required for the Coursera Regression Models course. The instructions for this report assignment state as follows:

We 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**

In general our analysis says that Manual transmissions are better in terms of mileage (mpg) than automatic. We found that, using simple linear regression with only transmission type, Manual transmission cars increase the mileage (mpg) by 7.245 over Automatic transmission. But, the transsmission type explained only 36% of the variation in mpg.

The best model (a mutltiple linear regression model of significant variables (cyl, hp, wt, & am) determined by ANOVA) says that the manual transmission increase the mileage (mpg) by 1.80921 over Automatic transmission, however the transsmission type explained over 84% of the variation in mpg.

Data Description

The dataset **mtcars** was extracted from the 1974 Motor Trend US magazine, which comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). As per the R document https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html, the data set consists of 32 observations and 11 variables. The variables of the data set **mtcars** are:

- **mpg**: Miles/(US) gallon
- **cyl**: Number of cylinders
- **disp**: Displacement (cubic inches)

- **hp**: Gross horsepower
- **drat**: Rear axle ratio
- **wt**: Weight (1000 lbs)
- **qsec**: 1/4 mile time
- **vs**: Engine (0 = V-shaped, 1 = straight)
- **am**: Transmission (0 = automatic, 1 = manual)
- **gear**: Number of forward gears
- **carb**: Number of carburetors

Exploratory Data Analysis

Load the required packages:

```
library(ggplot2)
```

Read the data and run the basic data exploratory analysis:

```
data("mtcars")
mt_cars <- mtcars
dim(mt_cars)

## [1] 32 11

head(mt_cars)

##                    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

str(mt_cars)

## 'data.frame':    32 obs. of  11 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 ...
```

Base Statistics:

```r
summary(mt_cars)
```

```
##       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      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  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
```

```r
# Unique Values
unique(mt_cars$cyl)
```

```
## [1] 6 4 8
```

```r
unique(mt_cars$vs)
```

```
## [1] 0 1
```

```r
unique(mt_cars$am)
```

```
## [1] 1 0
```

```r
unique(mt_cars$gear)
```

```
## [1] 4 3 5
```

```r
unique(mt_cars$carb)
```

```
## [1] 4 1 2 3 6 8
```

The variables **cyl**, **vs**, **am**, **gear**, & **carb** can be converted into a factor variables as it seems that thye are rather a level than a numeric.

```r
# Convert the variables into factor from numeric
mt_cars$cyl <- factor(mt_cars$cyl)
mt_cars$vs <- factor(mt_cars$vs)
mt_cars$am <- factor(mt_cars$am,labels=c("Automatic","Manual")) # 0=automatic, 1=manual
```

```r
mt_cars$gear <- factor(mt_cars$gear)
mt_cars$carb <- factor(mt_cars$carb)
str(mt_cars)
```

```
## 'data.frame':    32 obs. of  11 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 "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

The boxplot (plot1 in the appendix) shows that Manual Transmission provides better MPG compared to Automatic Transmission.

The boxplot (plot2 in the appendix) shows that the mileage (MPG) is get decreasing drastically if the number of cylinders **cyl** increases from 4 to 6 and 8.

From all the plots (plot1, plot2, plot3 in the appendix), we can notice that variables **am**, **cyl**, **disp**, **hp**, **drat**, **wt**, and **qsec** seem to have some strong correlation with **mpg.** But we will use linear models to quantify this in the subsequent regression analysis section.

Inference Analysis

- H_0: **Mileage (MPG)** is not affected by Transmission types.
- H_a: **Mileage (MPG)** is affected by Transmission types.

```r
t.test(mpg ~ am, data = mt_cars)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  mpg by am
## 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 Automatic    mean in group Manual 
##                17.14737                24.39231
```

The above inference analysis clearly says that the p-value 0.001374 which is < 0.05 & 95 % confidence intervalthe (-11.280194 -3.209684) not contains zero and Manual & Automatic transmissions are significatively different.

Regression Analysis

We start building linear regression models based on the different variables like only with transimission type, variables selected by STEP & AOV techniques and all variables. Then find out the best fit model among them using ANOVA technique. Then finally, perform analysis of residuals.

Model with only Transmission Type

First we will run a linear regression model with **am** as independent and **mpg** as dependent variable.

```
base_model <- lm(mpg ~ am, data = mt_cars)
summary(base_model)

##
## Call:
## lm(formula = mpg ~ am, data = mt_cars)
##
## 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
```

It shows that the coefficient is significant, at 7.245, which we can interpret as Automatic to Manual transmission will increase the mileage (mpg) by 7.245. So, transmission type has an impact on mpg.

It also shows that the adjusted R squared value is only 0.3385 which means that only 33.8% of the regression variance can be explained by this model.

There are , however, several other predictor/independent variables that we need to look at them to see if they play any impact in the model or not.

Multivariable Regression Model using R 'step' function

Here, we perfoms stepwise model selection to select significant predictors for the model. To implement stepwise modle, we can use `step` method which runs lm multiple times to build multiple regression models and select the best variables from them using both forward selection and backward elimination methods by the AIC algorithm. The code is depicted in the section below, you can run it to see the detailed computations if required.

```
init_model <- lm(mpg ~ ., data = mt_cars)
step_model <- step(init_model, direction = "both") ## returns one by one to
final best fit model
```

```
## Start:  AIC=76.4
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
##         Df Sum of Sq    RSS    AIC
## - carb  5   13.5989 134.00 69.828
## - gear  2    3.9729 124.38 73.442
## - am    1    1.1420 121.55 74.705
## - qsec  1    1.2413 121.64 74.732
## - drat  1    1.8208 122.22 74.884
## - cyl   2   10.9314 131.33 75.184
## - vs    1    3.6299 124.03 75.354
## <none>              120.40 76.403
## - disp  1    9.9672 130.37 76.948
## - wt    1   25.5541 145.96 80.562
## - hp    1   25.6715 146.07 80.588
##
## Step:  AIC=69.83
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear
##
##         Df Sum of Sq    RSS    AIC
## - gear  2    5.0215 139.02 67.005
## - disp  1    0.9934 135.00 68.064
## - drat  1    1.1854 135.19 68.110
## - vs    1    3.6763 137.68 68.694
## - cyl   2   12.5642 146.57 68.696
## - qsec  1    5.2634 139.26 69.061
## <none>              134.00 69.828
## - am    1   11.9255 145.93 70.556
## - wt    1   19.7963 153.80 72.237
## - hp    1   22.7935 156.79 72.855
## + carb  5   13.5989 120.40 76.403
##
## Step:  AIC=67
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am
##
##         Df Sum of Sq    RSS    AIC
## - drat  1    0.9672 139.99 65.227
## - cyl   2   10.4247 149.45 65.319
## - disp  1    1.5483 140.57 65.359
## - vs    1    2.1829 141.21 65.503
## - qsec  1    3.6324 142.66 65.830
## <none>              139.02 67.005
## - am    1   16.5665 155.59 68.608
```

```
## - hp    1   18.1768 157.20 68.937
## + gear  2    5.0215 134.00 69.828
## - wt    1   31.1896 170.21 71.482
## + carb  5   14.6475 124.38 73.442
##
## Step:  AIC=65.23
## mpg ~ cyl + disp + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - disp  1    1.2474 141.24 63.511
## - vs    1    2.3403 142.33 63.757
## - cyl   2   12.3267 152.32 63.927
## - qsec  1    3.1000 143.09 63.928
## <none>              139.99 65.227
## + drat  1    0.9672 139.02 67.005
## - hp    1   17.7382 157.73 67.044
## - am    1   19.4660 159.46 67.393
## + gear  2    4.8033 135.19 68.110
## - wt    1   30.7151 170.71 69.574
## + carb  5   13.0509 126.94 72.095
##
## Step:  AIC=63.51
## mpg ~ cyl + hp + wt + qsec + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - qsec  1     2.442 143.68 62.059
## - vs    1     2.744 143.98 62.126
## - cyl   2    18.580 159.82 63.466
## <none>              141.24 63.511
## + disp  1     1.247 139.99 65.227
## + drat  1     0.666 140.57 65.359
## - hp    1    18.184 159.42 65.386
## - am    1    18.885 160.12 65.527
## + gear  2     4.684 136.55 66.431
## - wt    1    39.645 180.88 69.428
## + carb  5     2.331 138.91 72.978
##
## Step:  AIC=62.06
## mpg ~ cyl + hp + wt + vs + am
##
##        Df Sum of Sq    RSS    AIC
## - vs    1     7.346 151.03 61.655
## <none>              143.68 62.059
## - cyl   2    25.284 168.96 63.246
## + qsec  1     2.442 141.24 63.511
## - am    1    16.443 160.12 63.527
## + disp  1     0.589 143.09 63.928
## + drat  1     0.330 143.35 63.986
## + gear  2     3.437 140.24 65.284
## - hp    1    36.344 180.02 67.275
```

```
## - wt    1    41.088 184.77 68.108
## + carb  5     3.480 140.20 71.275
##
## Step:  AIC=61.65
## mpg ~ cyl + hp + wt + am
##
##        Df Sum of Sq    RSS    AIC
## <none>              151.03 61.655
## - am    1     9.752 160.78 61.657
## + vs    1     7.346 143.68 62.059
## + qsec  1     7.044 143.98 62.126
## - cyl   2    29.265 180.29 63.323
## + disp  1     0.617 150.41 63.524
## + drat  1     0.220 150.81 63.608
## + gear  2     1.361 149.66 65.365
## - hp    1    31.943 182.97 65.794
## - wt    1    46.173 197.20 68.191
## + carb  5     5.633 145.39 70.438

#step_model <- step(init_model, trace=0) ## returns final best fit model
```

This analysis shows that the variables **cyl**, **hp** and **wt** as confounders and **am** as the independent variable.

```
summary(step_model)

##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mt_cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

It shows that the adjusted R squared value is 0.8401 which suggests that 84% or more of variance can be explained by this model.

P-values for **cyl**, **hp** and **wt** are below 0.05 which suggests that these are confounding variables in the relation between car Transmission Type and **mpg**.

Multivariable Regression Model using Analysis of Variance

Here, we performs an Analysis of Variance technique for the data to find best fit model.

```
T_variance <- aov(mpg ~ ., data = mt_cars)
summary(T_variance)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          2  824.8   412.4  51.377 1.94e-07 ***
## disp         1   57.6    57.6   7.181   0.0171 *
## hp           1   18.5    18.5   2.305   0.1497
## drat         1   11.9    11.9   1.484   0.2419
## wt           1   55.8    55.8   6.950   0.0187 *
## qsec         1    1.5     1.5   0.190   0.6692
## vs           1    0.3     0.3   0.038   0.8488
## am           1   16.6    16.6   2.064   0.1714
## gear         2    5.0     2.5   0.313   0.7361
## carb         5   13.6     2.7   0.339   0.8814
## Residuals   15  120.4     8.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This analysis shows that we need to consider the variables **cyl**, **disp**, and **wt** along with **am** as the p-values are less than .05 (i.e. 1.94e-07, 0.0171, and 0.0187 respectively).

```
aov_model <- lm(mpg ~ cyl + disp + wt + am, data = mt_cars)
summary(aov_model)

##
## Call:
## lm(formula = mpg ~ cyl + disp + wt + am, data = mt_cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5029 -1.2829 -0.4825  1.4954  5.7889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.816067   2.914272  11.604 8.79e-12 ***
## cyl6        -4.304782   1.492355  -2.885  0.00777 **
## cyl8        -6.318406   2.647658  -2.386  0.02458 *
## disp         0.001632   0.013757   0.119  0.90647
## wt          -3.249176   1.249098  -2.601  0.01513 *
## amManual     0.141212   1.326751   0.106  0.91605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.652 on 26 degrees of freedom
```

```
## Multiple R-squared:  0.8376, Adjusted R-squared:  0.8064
## F-statistic: 26.82 on 5 and 26 DF,  p-value: 1.73e-09
```

It shows that the adjusted R squared value is 0.8064 which suggests that 80% or more of variance can be explained by this model.

P-values for **cyl** and **wt** are below 0.05 which suggests that these are confounding variables (Confounding variables are any other variable that also has an effect on your dependent variable) in the relation between car Transmission Type and **mpg**.

Model (Multivariable Regression Model) with all Variables

Here, we performs a multivariate regression with **mpg** dependent variable and all the other variables as an independent.

```
all_model <- lm(mpg ~ ., data = mt_cars)
summary(all_model)

##
## Call:
## lm(formula = mpg ~ ., data = mt_cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.87913   20.06582   1.190   0.2525
## cyl6        -2.64870    3.04089  -0.871   0.3975
## cyl8        -0.33616    7.15954  -0.047   0.9632
## disp         0.03555    0.03190   1.114   0.2827
## hp          -0.07051    0.03943  -1.788   0.0939 .
## drat         1.18283    2.48348   0.476   0.6407
## wt          -4.52978    2.53875  -1.784   0.0946 .
## qsec         0.36784    0.93540   0.393   0.6997
## vs1          1.93085    2.87126   0.672   0.5115
## amManual     1.21212    3.21355   0.377   0.7113
## gear4        1.11435    3.79952   0.293   0.7733
## gear5        2.52840    3.73636   0.677   0.5089
## carb2       -0.97935    2.31797  -0.423   0.6787
## carb3        2.99964    4.29355   0.699   0.4955
## carb4        1.09142    4.44962   0.245   0.8096
## carb6        4.47757    6.38406   0.701   0.4938
## carb8        7.25041    8.36057   0.867   0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
```

```
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779 
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

It shows that the adjusted R squared value is 0.779 which suggests that 77% or more of variance can be explained by this model. But, the problem is that all the coefficients are not significative at 5% as their p-values are greather than 0.05.

Best Model Selection

We can use anova technique to find best model among above all the models.

```
anova(base_model, step_model, all_model, aov_model)
```

```
## Analysis of Variance Table
## 
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
## Model 3: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 4: mpg ~ cyl + disp + wt + am
##   Res.Df    RSS  Df Sum of Sq       F    Pr(>F)    
## 1     30 720.90                                    
## 2     26 151.03   4    569.87 17.7489 1.476e-05 ***
## 3     15 120.40  11     30.62  0.3468    0.9588    
## 4     26 182.87 -11    -62.47  0.7075    0.7153    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

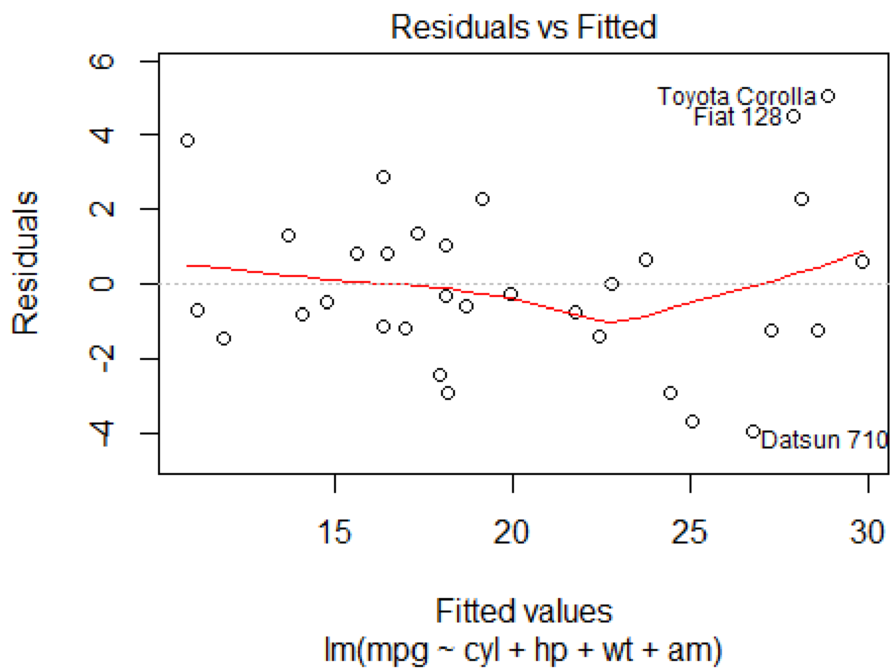
ANOVA confirms that the STEP model with 4 regressors (cyl, hp, wt, am), is the best model.

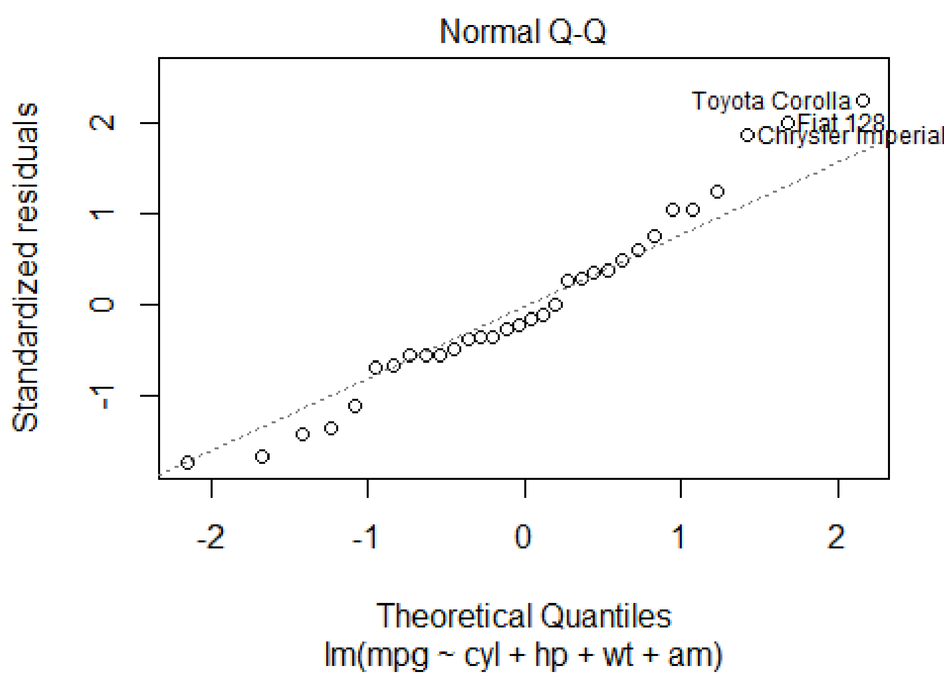
Residual and Diagnostics Analysis

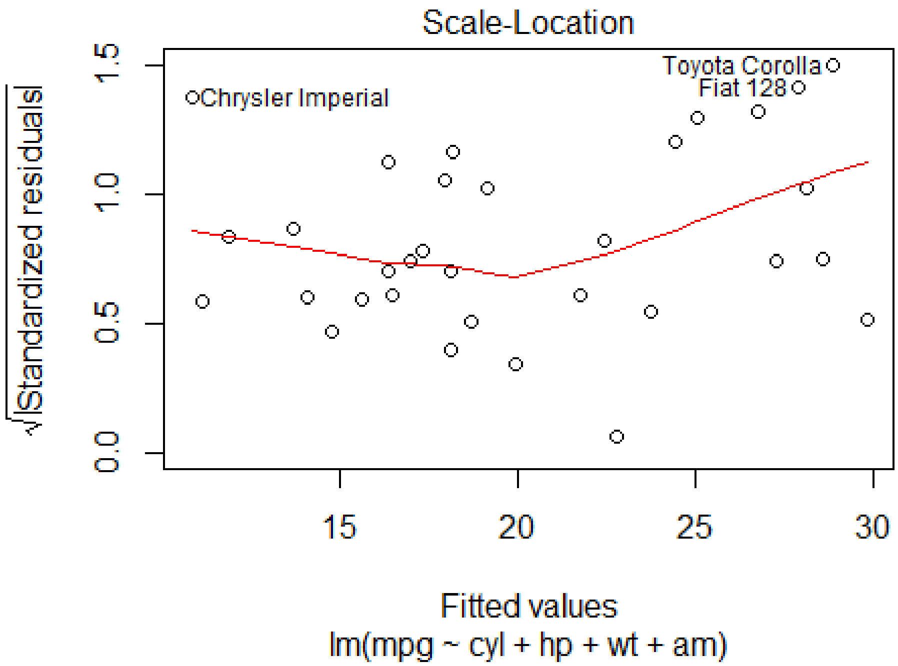
Here, we examine resdual plots of the best model (step) and compute some of its regression diagnostics to uncover outliers in the data set.

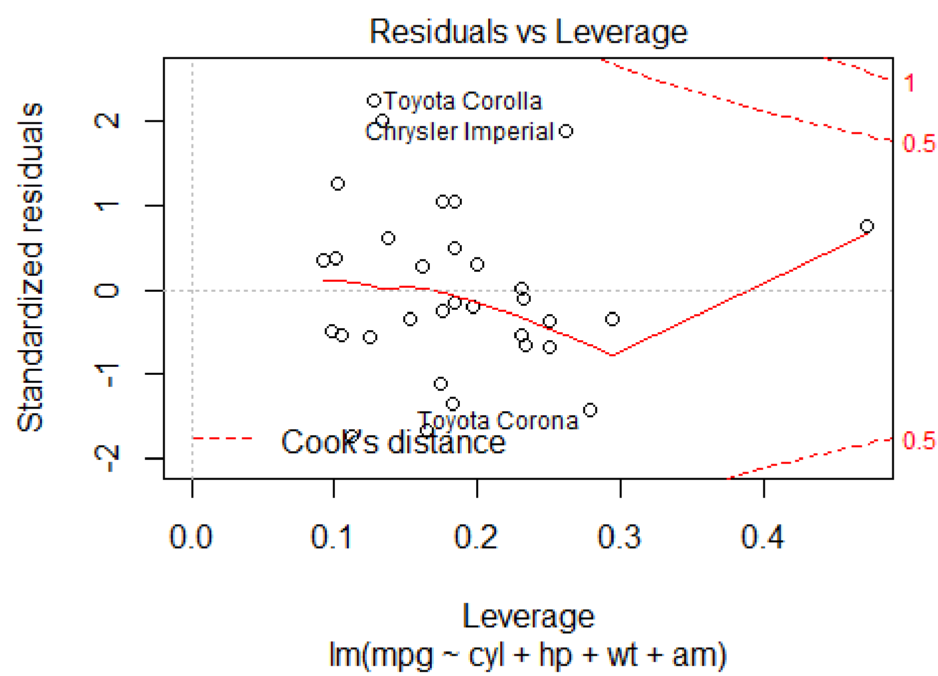
Residuals

```
#par(mfrow = c(2, 2))
plot(step_model)
```


Residuals vs Fitted
Residuals
6
4
2
0
-2
-4
Toyota Corolla
Fiat 128
Datsun 710
15
20
25
30
Fitted values
lm(mpg ~ cyl + hp + wt + am)


Normal Q-Q
Standardized residuals
2
1
0
-1
Toyota Corolla
Fiat 128
Chrysler Imperial
-2
-1
0
1
2
Theoretical Quantiles
lm(mpg ~ cyl + hp + wt + am)





Diagnostics

```
leverage <- hatvalues(step_model)
tail(sort(leverage),3)
```

```
##       Toyota Corona Lincoln Continental       Maserati Bora
##           0.2777872           0.2936819           0.4713671

influential <- dfbetas(step_model)
tail(sort(influential[,6]),3)

## Chrysler Imperial          Fiat 128     Toyota Corona
##         0.3507458         0.4292043         0.7305402
```

By looking at the above cars, we can see that our analysis was correct since the same cars are mentioned in the residual plots.

Conclusion

Is an automatic or manual transmission better for MPG?

When we consider only Transimission type as predictor, it shows that Manual transimission cars are better mileages compared to automatic cars. But when we modeled by considering confounding variables, the difference is not as significant as it seems with only transimission type since a major part of the difference is explained by other variables.

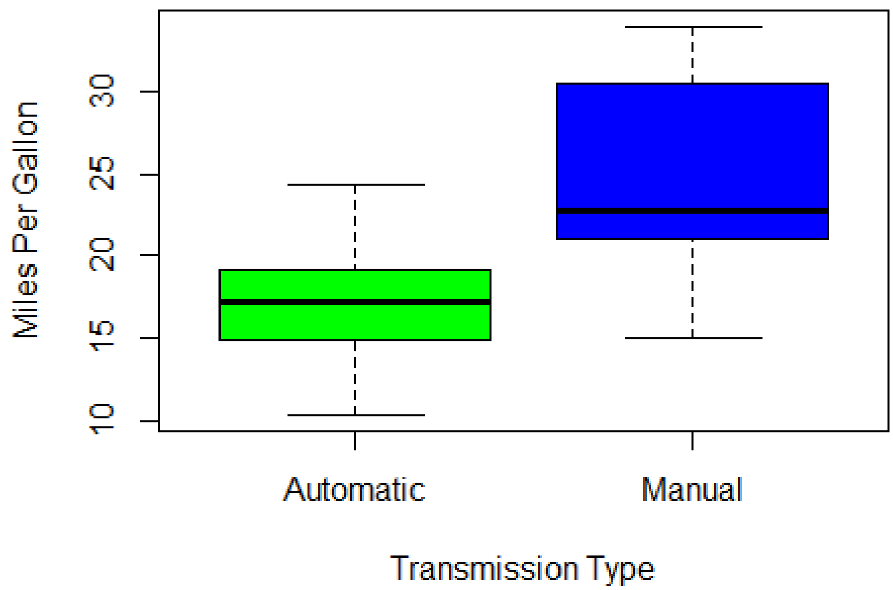
Quantify the MPG difference between automatic and manual transmissions

Our analysis confirms that when we considered only transimission type in the model, manual cars increase the mileage (mpg) by 7.245. But when we modeled by considering confounding variables (cyl + hp + wt) or (cyl + disp + wt) with transimission type, the Manual car's mileage advantage drops to 1.80921 or 0.141212 respectively.

Appendix

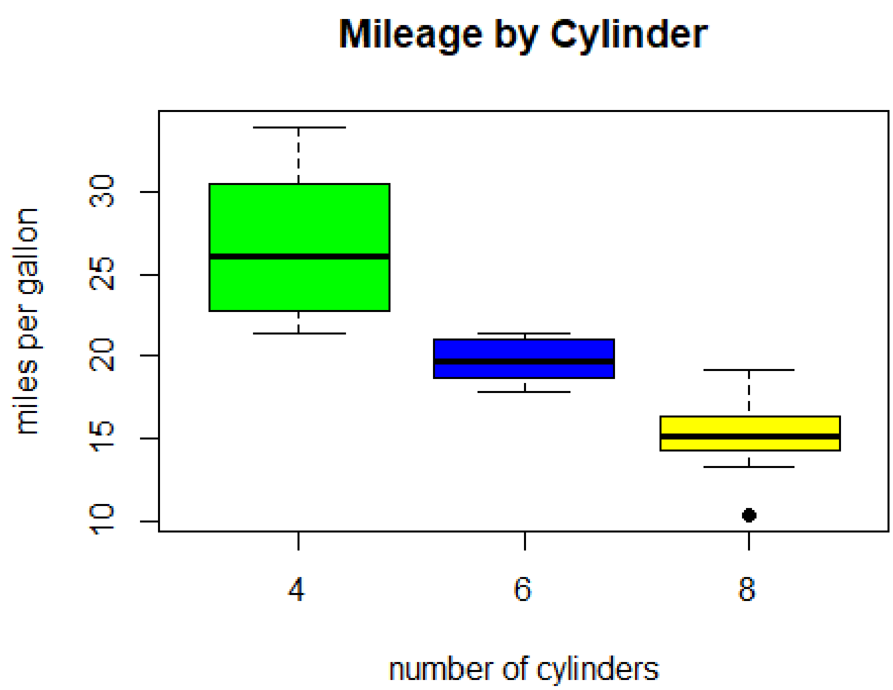
plot1: Boxplot of MPG by transmission type

```
boxplot(mpg ~ am, data = mt_cars, col = (c("green","blue")), ylab = "Miles
Per Gallon", xlab = "Transmission Type")
```



plot2: Boxplot of Mileage by Cylinder

```
boxplot(mt_cars$mpg ~ mt_cars$cyl, data=mt_cars, outpch = 19, col=(c("green",
"blue", "yellow")), ylab="miles per gallon", xlab="number of cylinders",
main="Mileage by Cylinder")
```



plot3: Scatter plot matrix

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