regression-course-project

Sheida Vanouei

August 19, 2018

Executive Summary

This report analyzed the relationship between transmission type (manual or automatic) and miles per gallon (MPG). The report set out to determine which transmission type produces a higher MPG. The mtcars dataset was used for this analysis.

Load Data

Load the dataset and convert categorical variables to factors.

```
datasets::mtcars
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
```

```
## Ferrari Dino         19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora        15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E           21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
library(ggplot2)
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
```

The dataset has 32 observations of 11 variables. We will do a quick analysis on the variables to gain some insight on the distribution of mpg and the two modes of transmission.

Exploratory Analysis

You can also embed plots, for example:

```
## [1] 0.5998324
```

```
##      mpg       cyl       disp         hp      drat         wt     qsec
## [1,]   1 -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##             vs        am      gear       carb
## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

The T-Test rejects the null hypothesis that the difference between transmission types is 0.

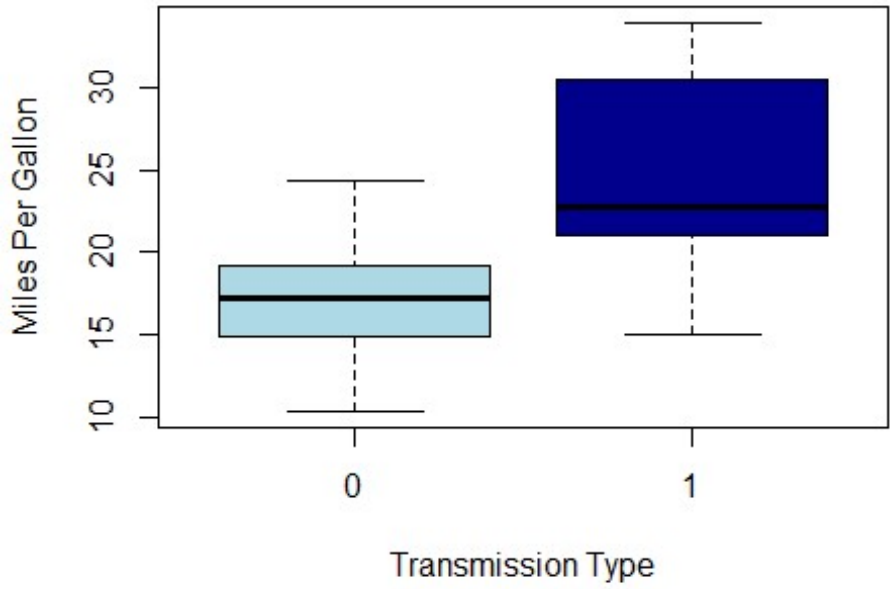
boxplot

```
Fit <- lm(mpg~am , data = mtcars)
summary(Fit)
```

```
##
## 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 ***
## am1            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
```

```
boxplot(mpg ~ am, data = mtcars, col = (c("lightblue","darkblue")), ylab =
"Miles Per Gallon", xlab = "Transmission Type")
```



The graph leads us to believe that there is a significant increase in MPG when for vehicles with a manual transmission vs automatic

##Multivariable Regression Model

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

```
##
## 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 ***
## am1          1.55649    1.44054   1.080  0.28984
## cyl         -1.10638    0.67636  -1.636  0.11393
```

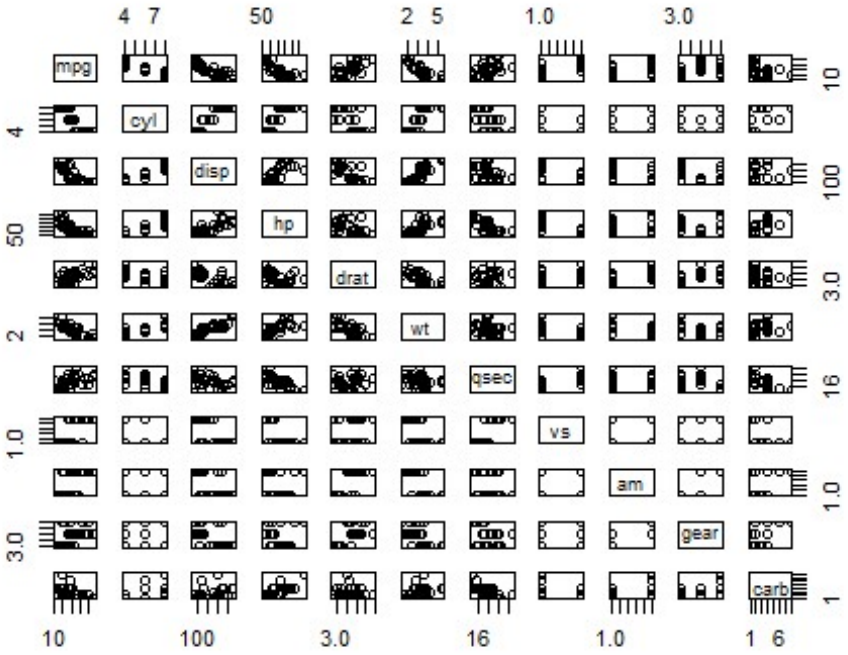
```
## disp         0.01226    0.01171   1.047  0.30472
## hp          -0.02796    0.01392  -2.008  0.05510 .
## wt          -3.30262    1.13364  -2.913  0.00726 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.505 on 26 degrees of freedom
## Multiple R-squared:  0.8551, Adjusted R-squared:  0.8273
## F-statistic:  30.7 on 5 and 26 DF,  p-value: 4.029e-10
```

The first model we will run is a linear regression model against mpg for each variable. This gives us insight into variables with coefficient significance as well as an initial attempt at explaining mpg. Additionally, we will also look at the correlation of variables with mpg to help us choose an appropriate model.

```
anova(Fit, bestFit)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl + disp + hp + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 163.12  4    557.78 22.226 4.507e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

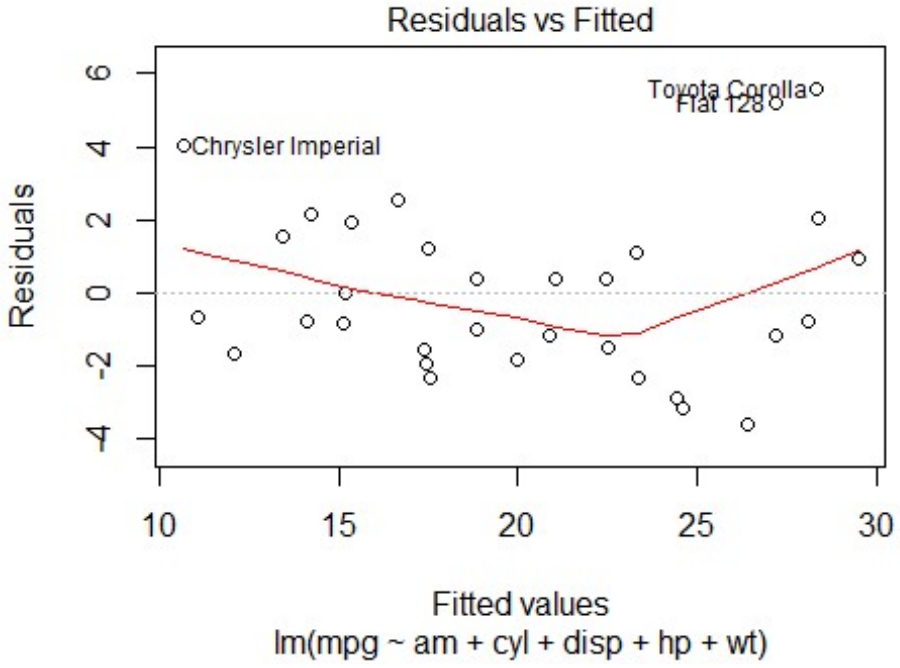
```
pairs(mpg ~ ., data = mtcars)
```

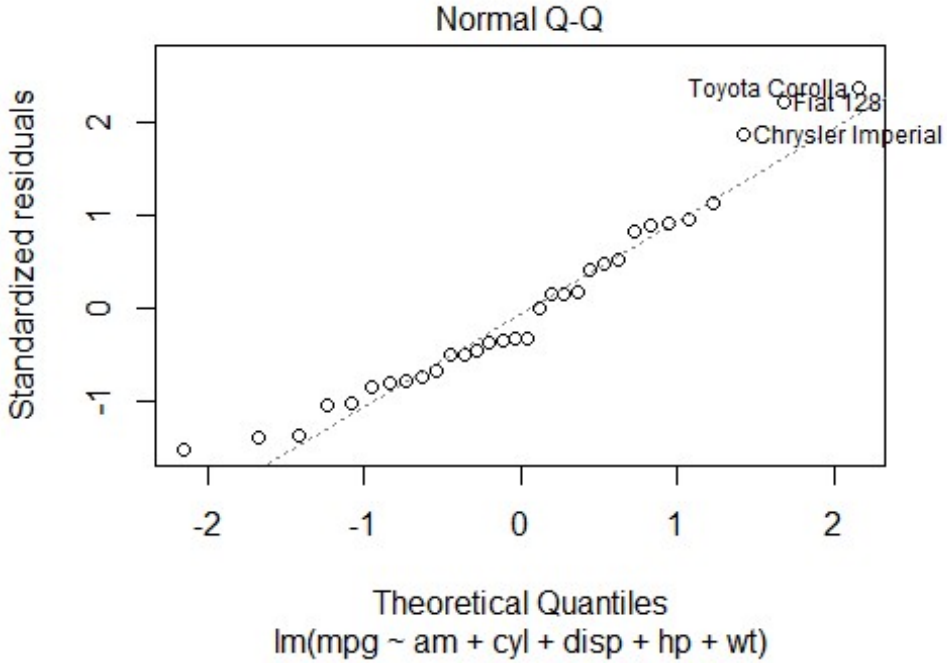


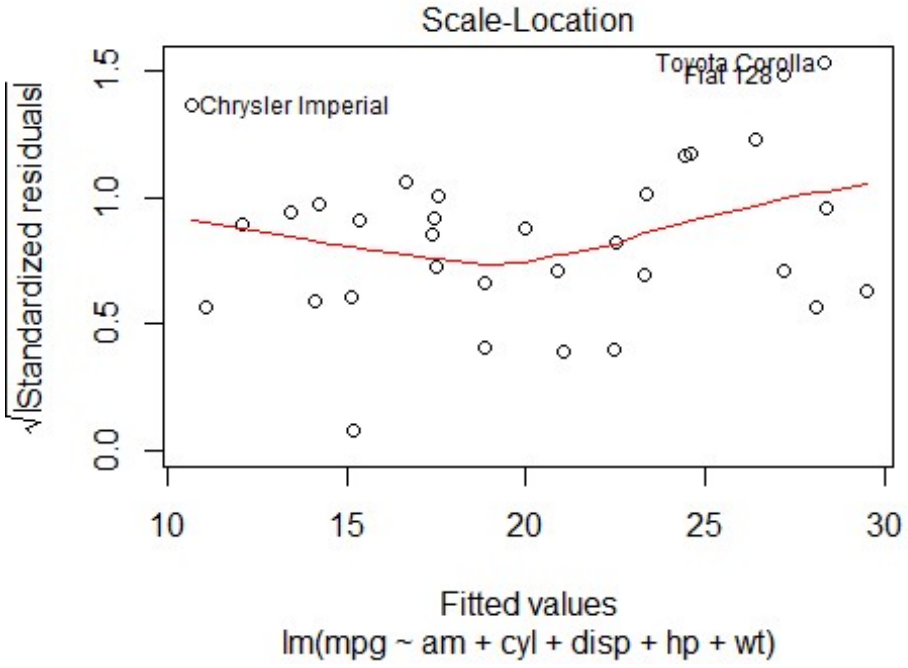
Additionally, we also plot the residuals to examine any heteroskedacity between the fitted and residual values; as well as to check for any non-normality.

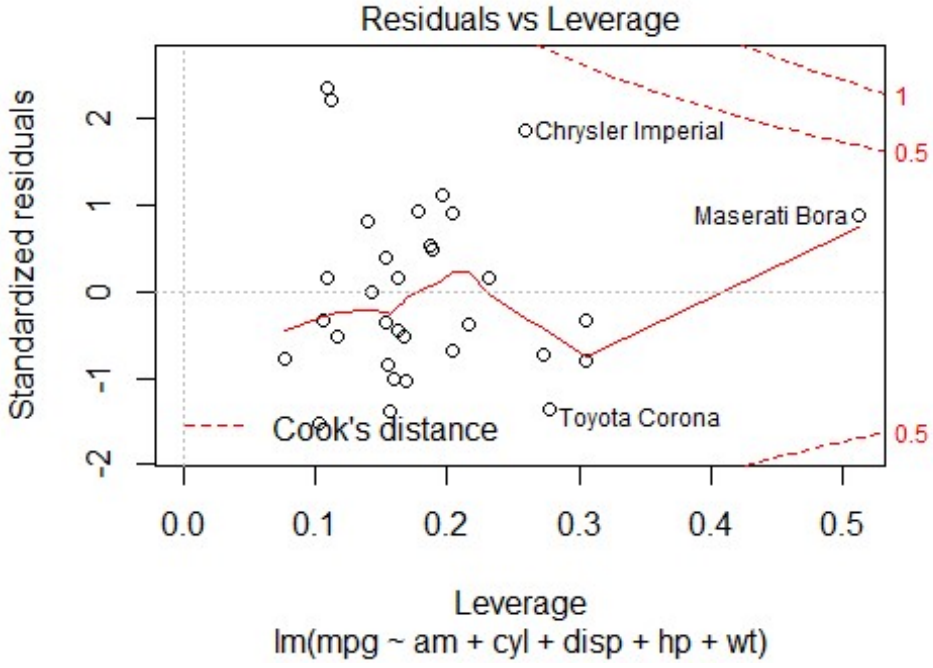
```
par(mfrow = c(2,2))

plot(bestFit)
```


Residuals vs Fitted
Toyota Corolla
Fiat 128
Chrysler Imperial
Residuals
6
4
2
0
-2
-4
10
15
20
25
30
Fitted values
lm(mpg ~ am + cyl + disp + hp + wt)


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





using the final multivariable regression model put together we can see the multiple R squared value is much higher at 0.855, where 85.5% of the regression variance can be explained by the chosen variables. We can thus conclude that 'wt' and 'cyl'are confounding variables in the

relationship between 'am and 'mpg' and that manual transmission cars on average have 1.55 miles per gallon more than automatic cars.

##Conclusion There is a difference in MPG based on transmission type. A manual transmission will have a slight MPG boost. However, it seems that weight, horsepower, & number of cylinders are more statistically significant when determining MPG.