Motor Trend Dataset

Kobra

08/09/2019

Motor Trend Data Analysis

Using the data collected from various cars, we are interested in investigating any potential relationship between a set of variables and miles per gallon (MPG) (outcome). Particularly, we will try to answer the following questions:

1) "Is an automatic or manual transmission better for MPG"
2) "Quantify the MPG difference between automatic and manual transmissions"

Based on the analysis below, it was observed that manual transmission is associated with higher MPG. However, other variables also contribute to MPG values.

Initial exploratory data analysis

Loading data itself:

```
library (dplyr)
data (mtcars)
```

Data Structure:

```
str (mtcars)
```

Variable 'am' (Transmission) needs to be converted to a factor variable and renamed properly: 0 for Atomatic and 1 for manual. Cyl, vs, gear and carb are also converted to factor variabes.

```
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
mtcars$am <- as.factor (mtcars$am)
levels (mtcars$am) <- c ("Automatic", "Manual")
```

How does average values look like for MPG between manual and automatic transmission?

he boxplot (appendix 1) shows a higher average value for manual group versus automatic group. The mean value and whether this difference is significant or not is investigated below:

```
mtcars %>% group_by (am) %>% summarise(mean_MPG = mean (mpg))
```

```
## # A tibble: 2 x 2
##   am        mean_MPG
##   <fct>        <dbl>
## 1 Automatic     17.1
## 2 Manual        24.4
```

```
t.test (mtcars$mpg ~ mtcars$am)$p.value
```

```
## [1] 0.001373638
```

The t.test above shows a significant diffence between manual and automatic cars in terms of MPG.

Regression modelling:

```
fit1 <- lm (mtcars$mpg~ mtcars$am)
summary (fit1)
```

```
##
## Call:
## lm(formula = mtcars$mpg ~ mtcars$am)
##
## 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 ***
## mtcars$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
```

P-values from above show a significant relationship between MPG and transmission type. The equation coefficient suggests MPG increases by 7.245 miles/gallon on average for manual transmission cars compared to automatic. However, R squared value shows that only %35 of the variation in the data is explained by this model. This leads us thinking that other variables might also contribute to MPG.

To find out which variables need to be included, here we use step function:

```
fit2 <- step(lm(mpg ~ ., data = mtcars), trace=0)
summary (fit2)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
```

```
##
## 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
```

Referring to p-value and R squared, respectively, this model is significant and explains %86 of the variation in the data. Residuals are also randomly scattered (appendix 2) which also confirms the suitability of our model.

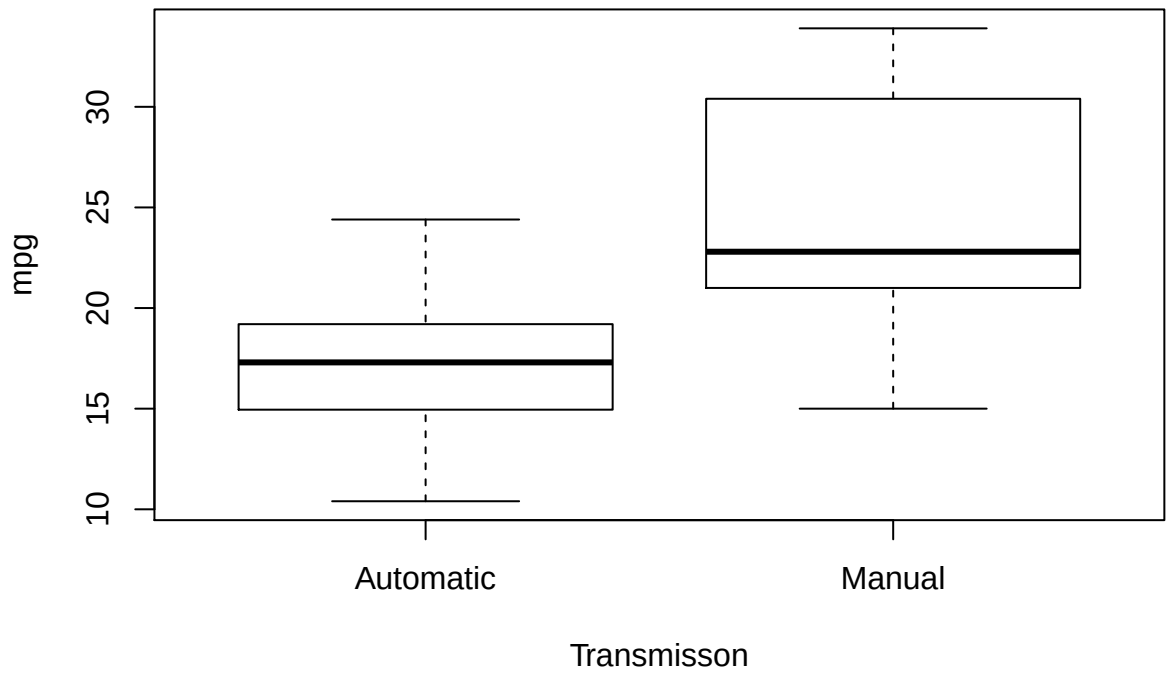
Conclusion:

Based on the results, manual cars are associated with higher MPG (Miles/(US) gallon) values compared to automatic cars. However, other variables including cyl (Number of cylinders), hp (Gross horsepower) and wt (Weight (1000 lbs)) should also be used to explain the variation in MPG.

Appendix 1:

Boxplot of MPG based on transmission type:

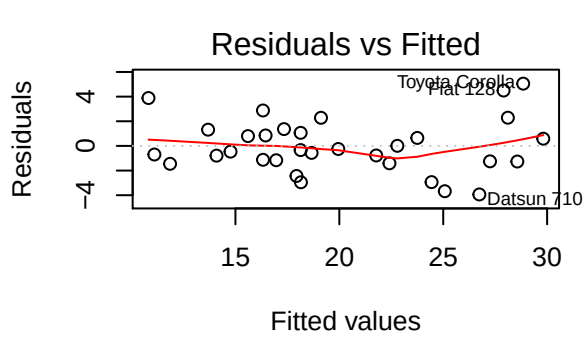
```r
boxplot(mpg~am, data =mtcars, xlab = "Transmisson")
```

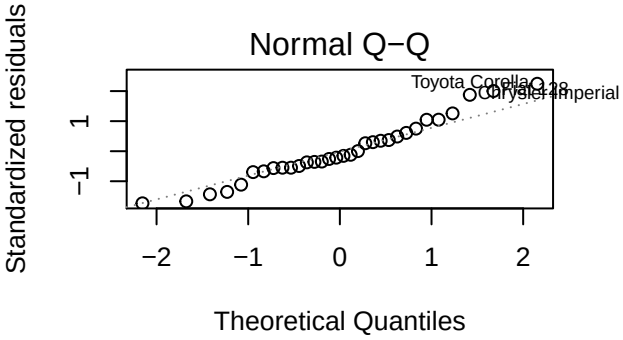


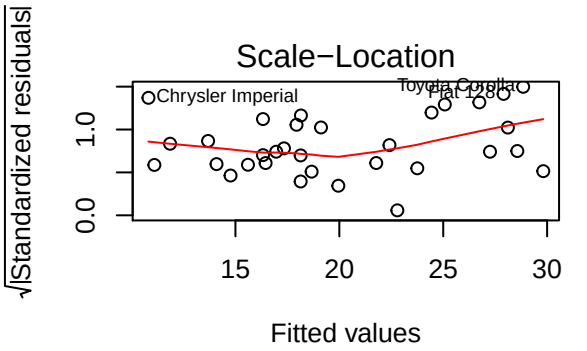
Appendix 2:

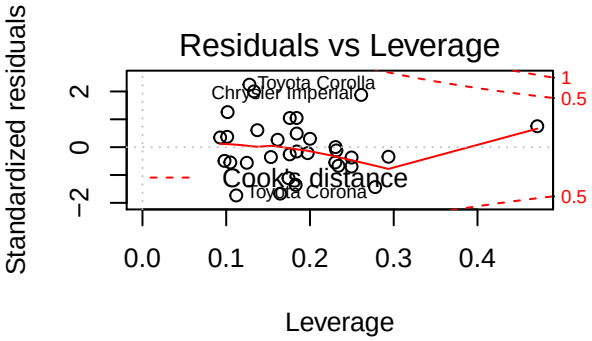
Residual plot:

```
par (mfrow= c(2,2))
plot (fit2)
```


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
15
20
25
30
−4
0
4


Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Chrysler Imperial
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
15
20
25
30
0.0
1.0


Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5
0.0
0.1
0.2
0.3
0.4
−2
0
2