Motor_Trend Data Analysis

Bert

5 juni 2019

Executive Summary

This report examines the relationship between a set of variables and fuel consumption measured in miles per gallon (MPG). With the focus on the question if an automatic or manual transmission is having an impact on the fuel consumption expressed in Miles Per Gallon (MPG). The mtcars dataset was used for this analysis.

A t-test between automatic and manual transmission vehicles shows that manual transmission vehicles have a 7.245 greater MPG than automatic transmission vehicles.

After fitting multiple linear regressions, analysis showed that the impact of weight, horsepower and the number of cylinders have a higher impact on the fuel consumption. Further analysis shows that a non linear approach furter improves the prediction model.

Load Data mtcars

Load the dataset and convert categorical variables to factors. Refer to appendix for code and description dataset

Exploratory Data Analysis

Fit linear model with transmission as only variable.

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

Statistical Inference test on relation mpg and transmission type

With a p.value of 0.0013736 which is below the 5% threshold the ttest confirms that a manual transmission has a signfcant higher range in MPG than the automatic transmission.

The ttest indicates a 7.24 MPG higher mean for the Manual transmission versus the Automatic transmission. In the Appendix a boxplot of the results is given, confirming the above coclusion.

Examining the impact of the other variables on the fuel consumption

Fit the model using all available variables

```
fullmodel <- lm(mpg ~ ., mtcars)
summary(fullmodel) # result is hidden
```

The summary results indicate that in a full model none of the variables have a p-value < 5%. Therefor no conclusions can be drawn and further investigation is needed.

The further investigation is done by using the step() function which is a step by step backward selection function optimizing Akaike's 'An Information Criterion' (AIC).

```
stepfitmpg <- step(lm(mpg ~ ., mtcars))
```

```
summary(stepfitmpg)$call
```

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

```r
AIC(fit, stepfitmpg)
```

```
##            df      AIC
## fit         3 196.4844
## stepfitmpg  7 154.4669
```

The model with the lowest AIC score has 4 variables (cylinders, horsepower, weight and transmission). The model is supporting the the conclusion that the manual transmission has a lower fuel consumption compared to the automatic transmission. However the impact has been reduced to an expected increase of 1.81 Miles per Gallon, down from 7.24 mpg in the simplest model with transmission as only variable.

Residual plot of model with cylinders, horsepower, weight and transmission as variables

The plot suggests that a non linear model might be a better fit. If weight square is added in the model we get a slight improvement.

```r
stepFitmpgsquare <- step(lm(mpg ~ .+I(wt^2), mtcars))
```

```r
summary(stepFitmpgsquare)$call
```

```
## lm(formula = mpg ~ wt + qsec + I(wt^2), data = mtcars)
```

In this model the transmission type is not anymore selected as one of the significant variables. qsec (the 1/4 mile time) is in this model more significant than the transmission type.

```r
AIC(fit, stepfitmpg, stepFitmpgsquare)
```

```
##                  df      AIC
## fit               3 196.4844
## stepfitmpg        7 154.4669
## stepFitmpgsquare  5 146.4754
```

When comparing the AIC score we see a slight improvement.

Also the residual plot of the quadratic model shows a better fit. See appendix for the comparison of the plots.

Conclusion

A t-test between automatic and manual transmission vehicles shows that manual transmission vehicles have a 7.245 greater MPG than automatic transmission vehicles.

After fitting multiple linear regressions, analysis showed that the impact of weight, horsepower and the number of cylinders have a higher impact on the fuel consumption.

Further analysis shows that a non linear approach furter improves the prediction model in which the transmission type only has a limited impact.

Appendix

Load Data mtcars

Load the dataset and convert categorical variables to factors.

```
data(mtcars)
#head(mtcars)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Description mtcars

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).

Format A data frame with 32 observations on 11 (numeric) variables

Variables	description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
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

Statistical Inference test on relation mpg and transmission type

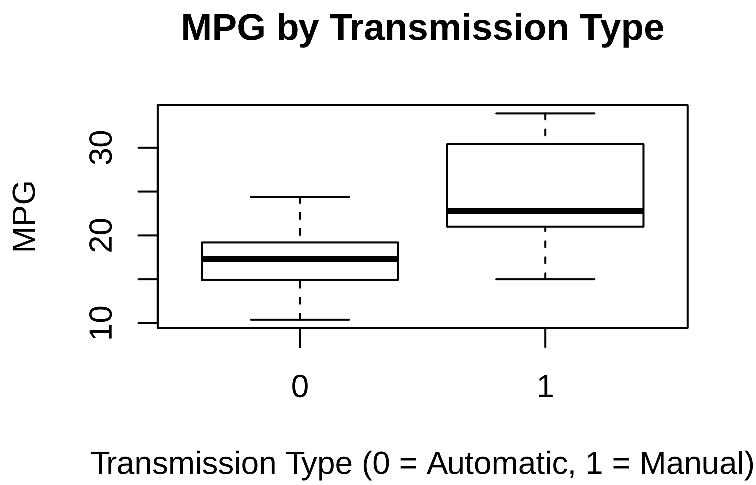
```
ttest <-t.test(mpg~am, mtcars)
ttest$p.value
```

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

```
#(0 =Automatic, 1=Manual)
x <- round(ttest$estimate[2]-ttest$estimate[1], 2)
x
```

```
## mean in group 1
##            7.24
```

```
  boxplot(mpg ~ am, data=mtcars,
          xlab="Transmission Type (0 = Automatic, 1 = Manual)",
          ylab="MPG",
          main="MPG by Transmission Type")
```



Code and result of the step backward selection function optimizing Akaike's 'An Information Criterion' (AIC).

```
stepfitmpg <- step(lm(mpg ~ ., mtcars))
stepFitmpgsquare <- step(lm(mpg ~ . + I(wt^2), mtcars))
```

```
summary(stepfitmpg)$call
```

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

```
coeftransmission <- round(summary(stepfitmpg)$coef[6,1],2)
coeftransmission
```

```
## [1] 1.81
```

```
summary(stepFitmpgsquare)$call
```

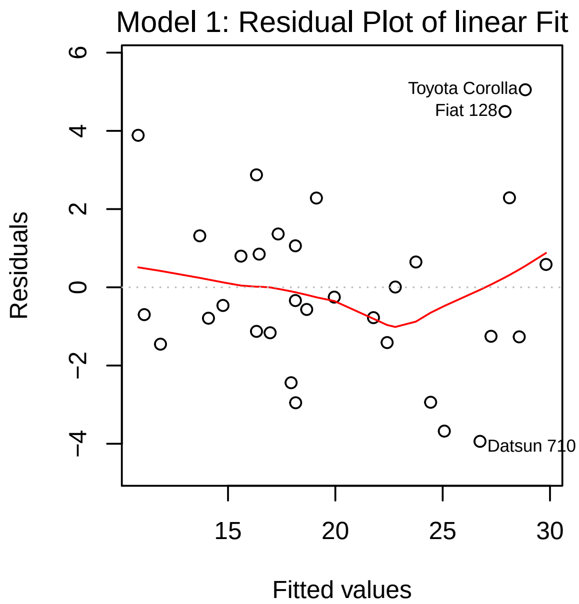
```
## lm(formula = mpg ~ wt + qsec + I(wt^2), data = mtcars)
```

Residual plots

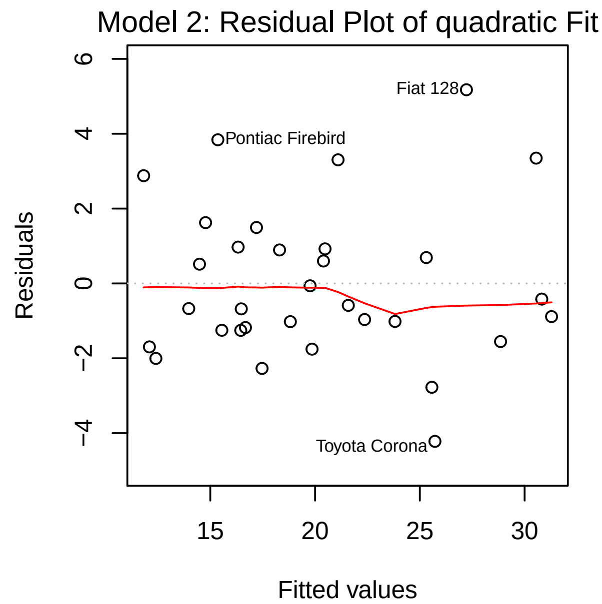
Model 1 : lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)

Model 2 : lm(formula = mpg ~ wt + qsec + I(wt^2), data = mtcars)

```
par(mfrow = c(2, 2))
plot(stepfitmpg, which=1,  caption = "Model 1: Residual Plot of linear Fit")
plot(stepFitmpgsquare, which=1, caption = "Model 2: Residual Plot of quadratic Fit")
```



Model 1: Residual Plot of linear Fit
Toyota Corolla
Fiat 128
Datsun 710
Residuals
Fitted values



Model 2: Residual Plot of quadratic Fit
Fiat 128
Pontiac Firebird
Toyota Corona
Residuals
Fitted values