Motor Trend Analysis

Sneh Bindesh Chitalia

15/06/2020

Executive Summary

When looking at the mtcars data set, I attempted to predict Miles per Gallon using different methods. Initially, regular intference testing indicated that the null hypothesis should be rejected proving manual transmissions did indeed have a significant positive impact on MPG. Next, I wanted to predict on average how many more miles per gallon a manual transmission car gets. The linear regression analysis estimated around a 7.24 miles per gallon increase.

It is also necessary to test how other variables effect MPG along with transmission and I ran a stepwise multiple variable regression analysis, tested the difference between the two fitted models, and looked at the residuals and determined that adding the weight of the car and acceleration increase the predictability of the modeling (33% to 83%). Looking at this model it shows that Transmission has less of an impact combined with other variables compared to the linear model, increasing fuel efficiency by 2.94 MPG.

Analysis:

Variables used:

mpg = Miles/(US) gallon
wt = Weight (lb/1000)
qsec = ¼ mile time
am = Transmission (0 = Automatic, 1 = Manual)

Exploratory Analysis:

Histogram
Boxplot

Inference using t-test:

t test used to compare both the transmission types.

Regression and Residuals

Linear Regression:

Used only one variable am to predict the outcome mpg.

Stepwise Multivariate Regression:

Used all variable initially and then a final model was selected based on AIC value.

Anova Test:

This test was used to compare both the models to see their impact on mpg.

Residuals:

Residuals and diagnostics were plotted to see if there are no large outliers that govern the plot.

Load the dataset

```
library(datasets)
data(mtcars)
str(mtcars)
```

```
## '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 ...
```

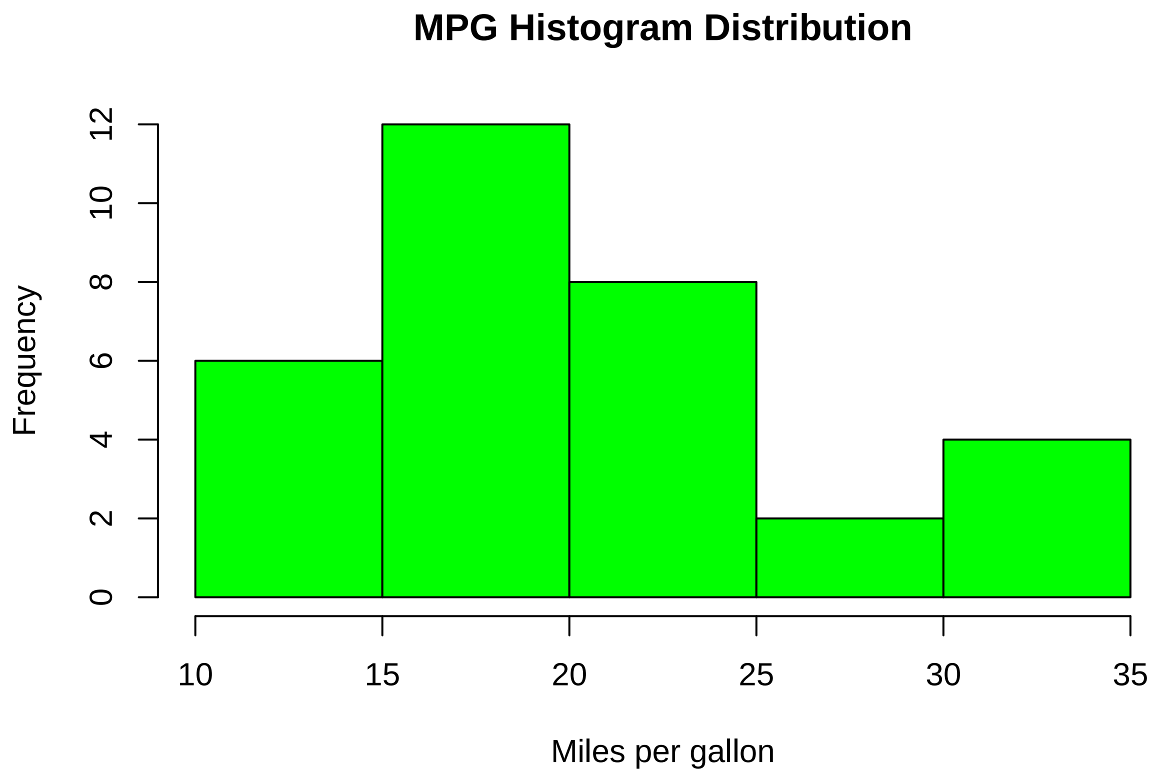
Convert am variable into factor variable

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

Appendix

Exploratory Data Analysis

```
library(ggplot2)
# histogram of MPG
hist(mtcars$mpg,col="green",xlab="Miles per gallon",main="MPG Histogram Distribution")
```




```r
# boxplot for comparing effect of transmission on mpg
ggplot(mtcars,aes(x=am,y=mpg,fill=am))+geom_boxplot()+xlab("Transmission type") +  ylab("Miles Per Gall
```

Question 1:

Is an automatic or manual transmission better for MPG

```
mpg_auto<-mtcars[mtcars$am == "Automatic",]$mpg
mpg_man<-mtcars[mtcars$am == "Manual",]$mpg
t.test(mpg_auto,mpg_man,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg_auto and mpg_man
## 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 of x mean of y
##  17.14737  24.39231
```

```
# Comparing median of both transmission types
median(mpg_auto)
```

```
## [1] 17.3
```

```r
median(mpg_man)
```

```
## [1] 22.8
```

This shows that p-value is <0.005 which indicates that the null hypothesis is rejected which the means of both transmission types are not equal. Confidence interval does not contain zero which shows that the test is significant.
Comparing the means and median of both transmission types, it is clear that MPG's for manual transmission in greater.

Question 2

Quantify the MPG difference between automatic and manual transmissions.

Linear Regression

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

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## amManual     2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

This shows that the estimate for Manual transmission is more than Automatic by 7.245. However, the R-squared value which shows the variance covered is only ~36%.

Stepwise Multivariate Regression

```
library(MASS)
fit2<-stepAIC(lm(mpg~.,data=mtcars),direction = "both",trace=FALSE)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

This model also shows that estimate for Manual transmission is more than Automatic by 2.94.
The R-squared value which shows the variance covered is also ~85%.
Which of the above two regression models should be considered to quantify the absolute difference between transmission types.
For this, we will have to compare both the models.

Null Hypothesis: weight and 1/4 mile time (acceleration) do not effect mpg

```
anova(fit,fit2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
## Model 2: mpg ~ wt + qsec + am
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     21 147.49
## 2     28 169.29 -7   -21.791 0.4432 0.8636
```

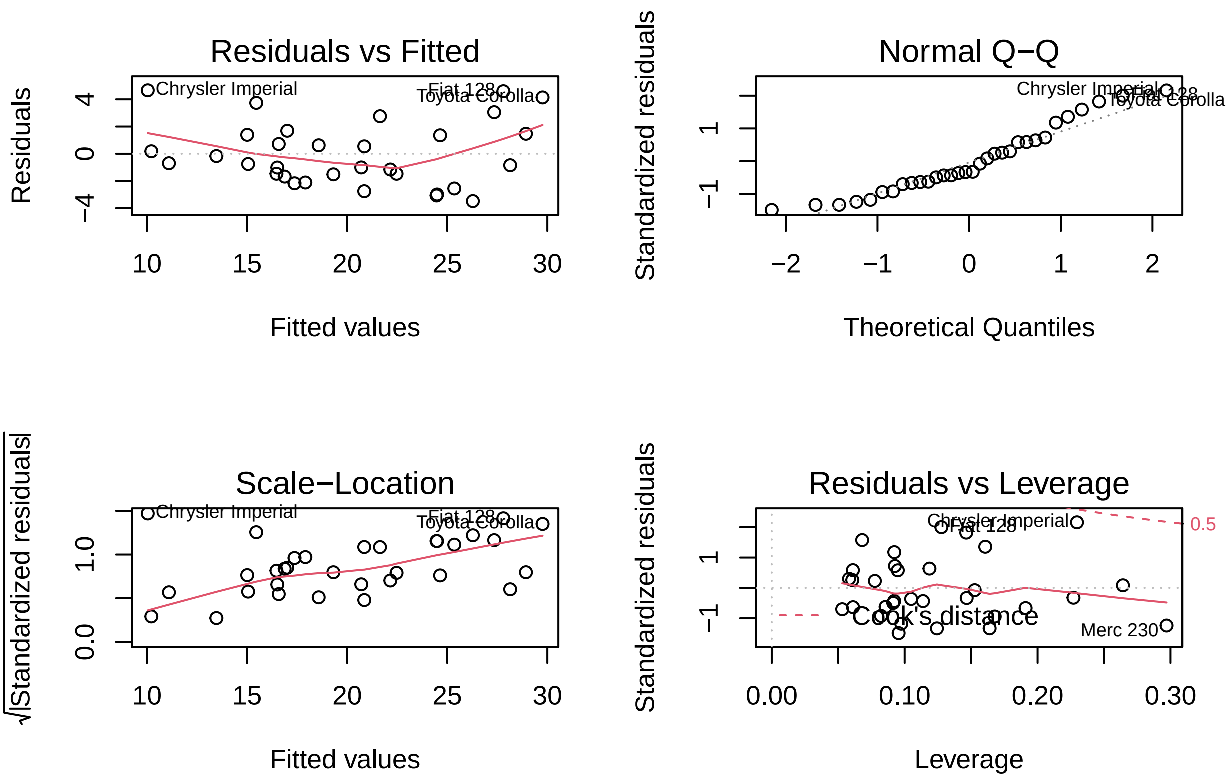
Since the p-value is <0.05, we can reject the null hypothesis and conclude that weight and acceleration do effect MPG.

Answer for question 2:

Thus, holding the weight and acceleration of the car constant, we can assume that fuel efficiency of Manual transmission cars is more than Automatic transmission cars by an estimate of ~2.94.

Comparing residuals and diagnostic

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



Thus no large outliers are being displayed.