mtcars_Analysis

Ash Thompson

16 December 2021

Regression Models Course Project

Executive Summary

This report is an analysis of the mtcars dataset, to explore the relationship betweena set of variables and miles per gallon (MPG). Particularly interested in answering the following two questions: - Is an automatic or manual transmission better for MPG - Quantify the MPG difference between automatic and manual transmissions

Initial steps

The first steps are to load the dataset and supporting libraries, then perform some minor transformations and exploratory analysis.

```
data(mtcars)
library(ggplot2)
library(dplyr)
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
```

```
mtcars$am <- factor(mtcars$am, labels =c("Automatic", "Manual"))
```

Please see appendix 1 for a box plot of the 2 transmission types and a comparison of mpg, that shows just looking at transmission type, it appears manual transmission gives around 7 more mpg.

Using linear regression to investigate further

These steps are to investigate the hypothesis that manual transmission gives 7 more mpg. First a simple linear model.

```
model1 <- lm(mpg ~ am, data=mtcars)
summary(model1)
```

```
##
## 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 ***
## 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
```

R squared value is fairly low at 0.3385, so only around 34% of the variance can be explained by this model. the next steps are to include the other variables into the model and do an analysis of variance to determine the best variables to include in a final model.

```
model2 <- lm(mpg ~ ., data=mtcars)
summary(model2)
```

```
##
## 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
```

```
summary(aov(mpg ~ ., data=mtcars))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          1  817.7   817.7 116.425 5.03e-10 ***
## disp         1   37.6    37.6   5.353  0.03091 *
## hp           1    9.4     9.4   1.334  0.26103
## drat         1   16.5    16.5   2.345  0.14064
```

```
## wt           1   77.5    77.5  11.031  0.00324 **
## qsec         1    3.9     3.9   0.562  0.46166
## vs           1    0.1     0.1   0.018  0.89317
## am           1   14.5    14.5   2.061  0.16586
## gear         1    1.0     1.0   0.138  0.71365
## carb         1    0.4     0.4   0.058  0.81218
## Residuals   21  147.5     7.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The variables cyl, disp and wt are significant enough to include in the model, in addition to transmission type.

```
model3 <- lm(mpg ~ am + cyl + disp + wt, data=mtcars)
summary(model3)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.318 -1.362 -0.479  1.354  6.059
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.898313   3.601540  11.356 8.68e-12 ***
## amManual     0.129066   1.321512   0.098  0.92292
## cyl         -1.784173   0.618192  -2.886  0.00758 **
## disp         0.007404   0.012081   0.613  0.54509
## wt          -3.583425   1.186504  -3.020  0.00547 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.642 on 27 degrees of freedom
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8079
## F-statistic: 33.59 on 4 and 27 DF,  p-value: 4.038e-10
```

Please see appendix 2 for the plot of residuals showing they have roughly the same variance. The Normal Q-Q plot shows that the distribution of residuals is normal.

Conclusion

Manual transmission, taking other variables into account, gives a slightly higher MPG value of approximately 0.1. But number of cylinders and weight have more of an effect, cylinders lowering it by around 1.7 and wt lowering by 3.5.

Appendix 1

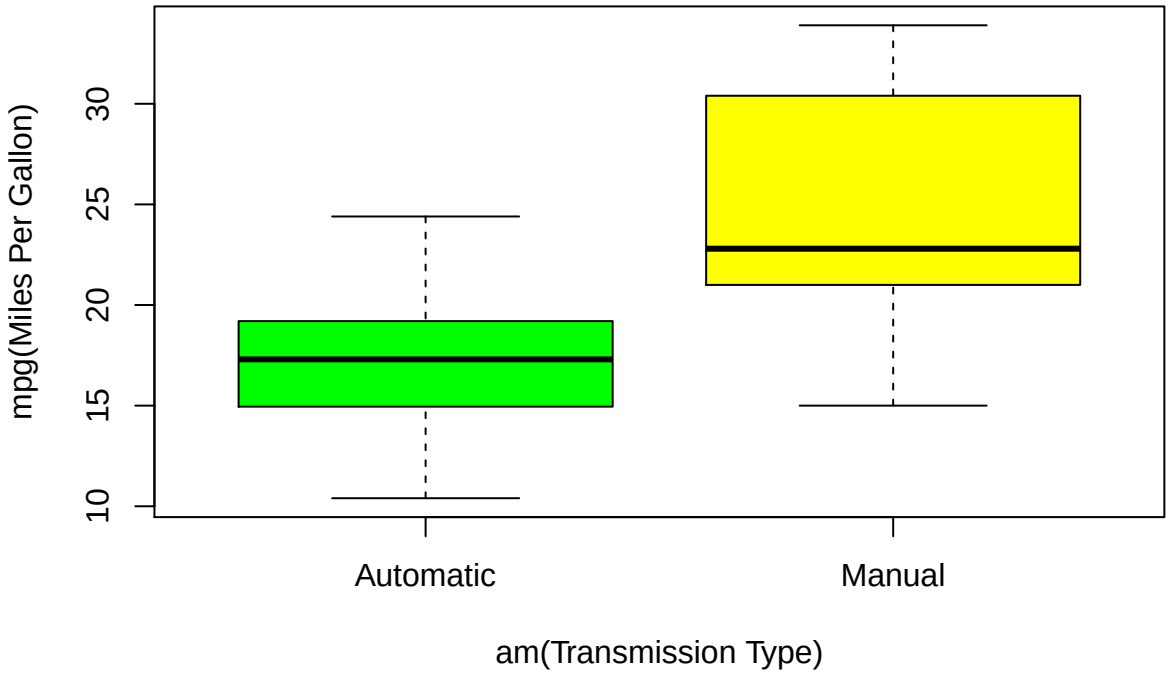
```
## [1] "Manual"
```

```
##       mpg
##  Min.   :15.00
##  1st Qu.:21.00
##  Median :22.80
##  Mean   :24.39
```

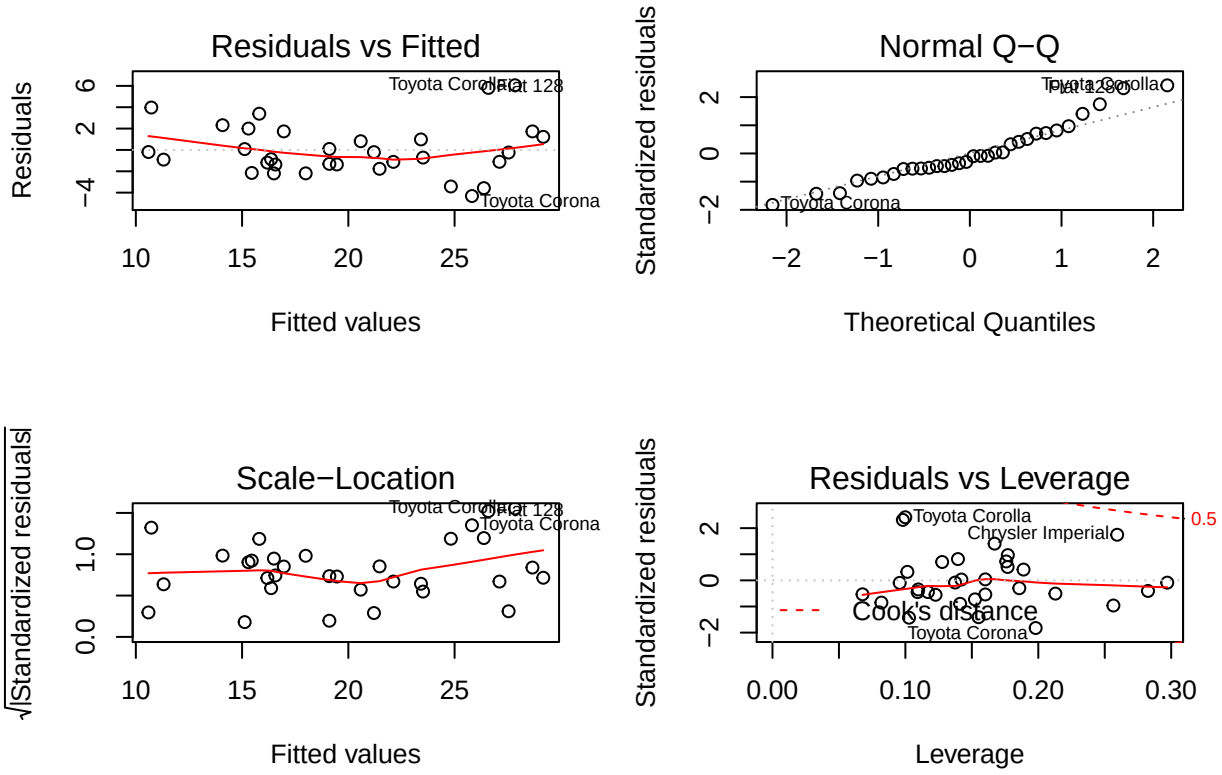
```
##  3rd Qu.:30.40
##  Max.   :33.90
```

```
## [1] "Automatic"
```

```
##       mpg
##  Min.   :10.40
##  1st Qu.:14.95
##  Median :17.30
##  Mean   :17.15
##  3rd Qu.:19.20
##  Max.   :24.40
```



Appendix 2


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Chrysler Imperial
Cook's distance
0.5
Leverage