Impact of transmission type on Miles per Gallon

Executive Summary

Looking at a data set of a collection of cars, we are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

- Is an automatic or manual transmission better for MPG
- Quantifying how different is the MPG between automatic and manual transmissions?

Conclusions

After analysys we can conclude that the manual transmission improve eficiency for MPG; on the other hand, it can be seriously influenced by horsepower variable.

Data

We will use the `mtcars` dataset, from the 1974 Motor Trend US magazine, to compare 32 different car models.

Load Data

```
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 ...
```

We can observe that our dataset contains 32 observations and 11 variables.

Model

Initialy, we will create a ***Multivarible Regression*** with all variables. Then, we will then use the **backward elimination strategy** to eliminate the unrelated variables, throw an iterative process where we eliminate the lessest significant variables.

Multivarible Regression

Let's create a linear regression model with all variables.

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

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3.45  -1.60  -0.12   1.22   4.63
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.3034    18.7179    0.66    0.518
## cyl          -0.1114     1.0450   -0.11    0.916
## disp          0.0133     0.0179    0.75    0.463
## hp           -0.0215     0.0218   -0.99    0.335
## drat          0.7871     1.6354    0.48    0.635
## wt           -3.7153     1.8944   -1.96    0.063 .
## qsec          0.8210     0.7308    1.12    0.274
## vs            0.3178     2.1045    0.15    0.881
## am            2.5202     2.0567    1.23    0.234
## gear          0.6554     1.4933    0.44    0.665
## carb         -0.1994     0.8288   -0.24    0.812
## ---
## 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:  0.807
## F-statistic: 13.9 on 10 and 21 DF,  p-value: 3.79e-07
```

Backward Strategy

Now, we eliminate the largest p-value variables in each iteration.

The final model contains `wt`, `qsec` and `am` variables.

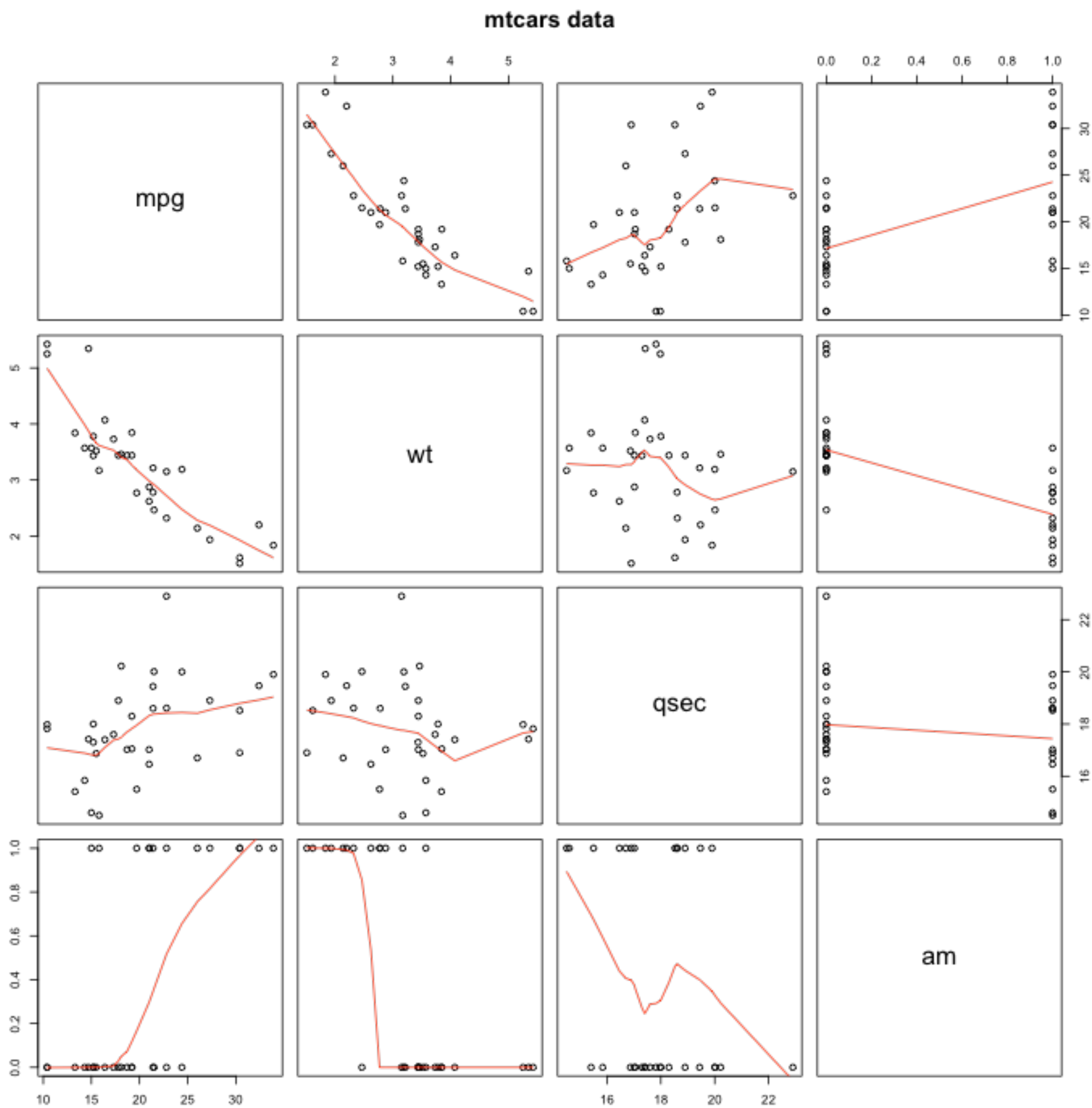
```
model <- lm(mpg ~ wt + qsec + am, data = mtcars)
summary(model)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.481 -1.556 -0.726  1.411  4.661
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.618      6.960    1.38  0.17792
## wt            -3.917      0.711   -5.51    7e-06 ***
## qsec           1.226      0.289    4.25  0.00022 ***
## am             2.936      1.411    2.08  0.04672 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.46 on 28 degrees of freedom
## Multiple R-squared:  0.85,   Adjusted R-squared:  0.834
## F-statistic: 52.7 on 3 and 28 DF,  p-value: 1.21e-11
```

Exploratory Analysis

Now we will explore the relation between `mpg` and the three variables in our final model.

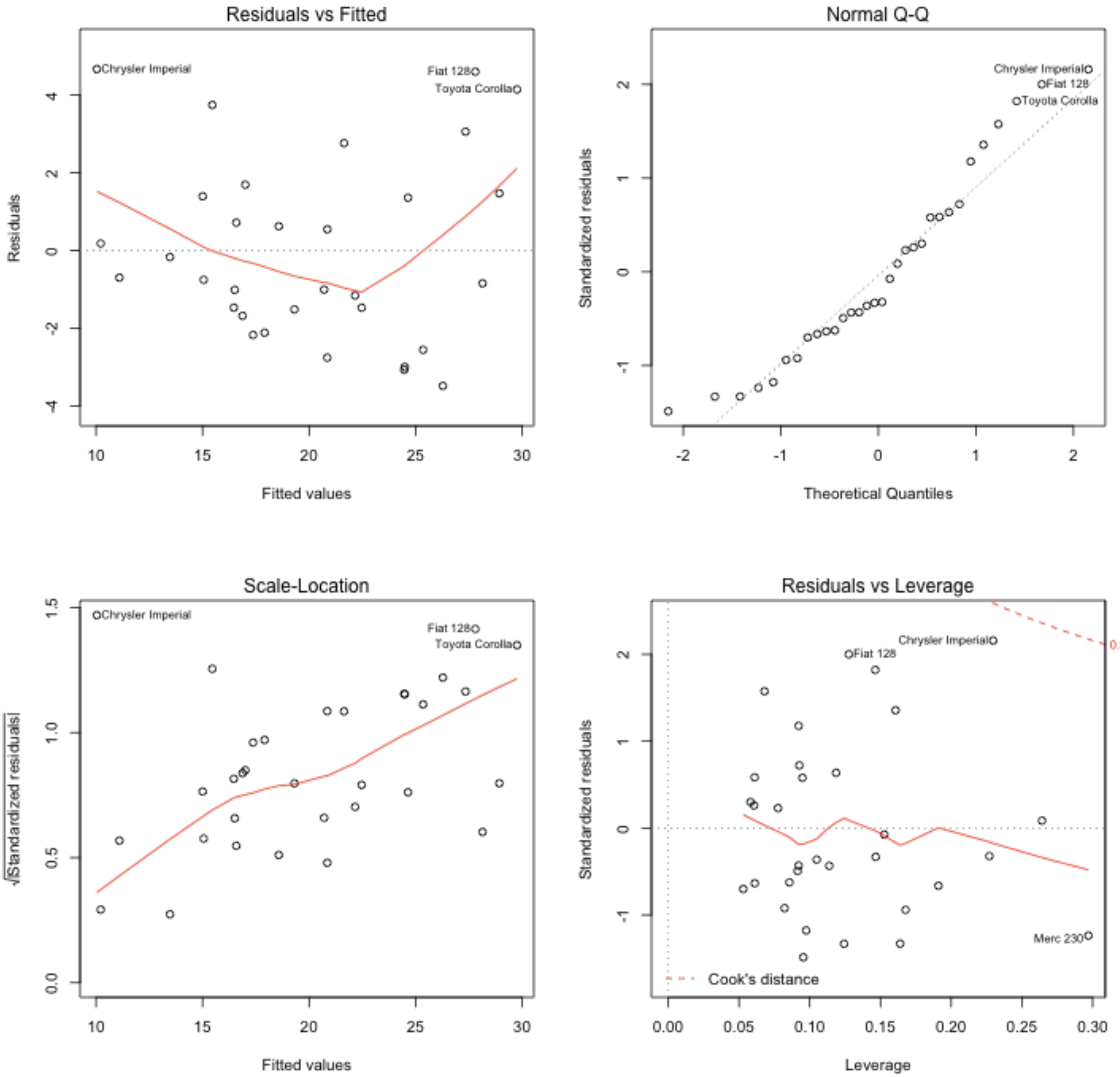
```
tidy_data <- mtcars[, c("mpg","wt", "qsec", "am")]
pairs(tidy_data, panel = panel.smooth, main = "mtcars data")
```



As we can see, there is a high relationship bewteen the three variables and the outcome.

We can also plot the residual

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



Conclusion

We can conclude that our linear model is a resonable fit.

```
sum_coefficients <- summary(model)$coefficients
result <- sum_coefficients["am", 1] + c(-1, 1) * qt(0.975, df =
model$df) * sum_coefficients["am", 2]; result
```

```
## [1] 0.04573 5.82594
```

With 95% confidence, we estimate that a the change from automatic to manual transmission results in a 0.05 to 5.83 increase in miles per gallon for the cars.

In conclusion, the manual transmission is better than automatic transmission for mpg.