Project - Impact of transmission type on fuel usage

Joscha Frischherz

March 12, 2016

Executive Summary

The following analysis looks at car data variables and their impact on "miles per gallon". We looked at whether an automatic or manual transmission is better and tried to quantify that difference. Our analysis shows that a manual transmission uses less fuel and a car (disregarding other factors) should be able to drive 7.25 miles further if its transmission is manual as opposed to automated. Having said that this quantification is a statistical value only, obviously many other factors would influence the exact fuel consumption of a car. But we are able to confidentially state that a manual transmission is better with regards to fuel consumption.

Analysis

Variable Selection

To start we will identify variables that are strongly correlated to "miles per gallon" (mpg). Refer to Appendix for correlation workings. As we are interested in transmission (am) impact on mpg, we use this variable's absolute correlation as lower benchmark for our initial variable selection. Therefore we are interested in analysing the impact by:

- Number of cylinders (cyl)

- Displacement (disp)
- Gross horsepower (hp)
- Rear axle ratio (drat)
- Weight (1000 lbs) (wt)
- V/S (vs)
- Transmission (0 = automatic, 1 = manual) (am)

Model Analysis

We analyse three models. The first two are multi linear regressions of correlated variables, the first one including "transmission type", while the second one exludes it. We note that some more influential variables start to cristal out, such as "weight" and "gross horse power", while "transmission type" does not seem to have an overarching influence considering the other variables. Refer to the models in the appendix.

Model of all correlated variables (Adjusted R-square)

```
## [1] 0.8180107
```

Model of all correlated variables exclusive transmission (Adjusted R-square)

```
## [1] 0.8163016
```

We note that most of the variance (>81%) is explained by the above two models. Having said that statistical significant factors are only sporadically identified.

Now performing a simple linear rgreassion only on transmission type we identify that "transmission type" does have a significant influence on "miles per gallon".

Model of transmission variable (Adjusted R-square)

```
## [1] 0.3384589
```

Comparing Models

```r
anova(fit1,fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + disp + hp + drat + wt + vs + am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + vs
## Model 3: mpg ~ am
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     24 158.65
## 2     25 166.82 -1     -8.16  1.2348    0.2775
## 3     30 720.90 -5   -554.08 16.7634 3.802e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparing the three models now, we understand that the simple linear regression provides us with the best statistical information on a variable impacting "miles per gallon". Having said that we read from the above more generic models, that other factors would likely bee impactful as well and statistical significant if analysed in isolation. This observation is substantiated by the Adjusted R-squared of 33.85%, which basically outlines that only that portion of the regression variance can be explained by the bespoke model. Hence the other factors may explain the rest.

Statistical Inference

We confirm the above via statistical inference.

```r
t.test(mpg ~ am, data = mtcars)$p.val
```

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

Based on the t-test we obtain a p-value of 0.001374 which allows us to reject the Null hypothesis that automated and manual transamission are indifferent with relation to "miles per gallon".

Conclsion

We identified that "transmission type" clearly does have a statistical influence on "miles per gallon" based on the analysed data set. Having said that we also noted that it is not the sole influencer and many other variables also have an influence on "mile per gallon". We can conclude however that an manual transmission uses less fuel than an automated one. This observation is also supported by the box plot in the Appendix. Based on the simple linear regressio, we can quantify that a manual transmission car may allow a car to drive 7.25 miles more on a gallon of petrol, disregarding all other variables. The key limitation identified is that the bespoke simple linear regression model only explains 33.85% of variances.

Appendix

R code applied for loading of data and packages

data load Procedure to load data library(datasets)

data(mtcars)

Additional tools

library(stats)

library(graphics)

Data exploration

Refer to help (?mtcars) for additional information of the data set. The help file outlines the variables as follows:

- [, 1] mpg Miles/(US) gallon
- [, 2] cyl Number of cylinders
- [, 3] disp Displacement (cu.in.)
- [, 4] hp Gross horsepower
- [, 5] drat Rear axle ratio
- [, 6] wt Weight (1000 lbs)
- [, 7] qsec 1/4 mile time
- [, 8] vs V/S
- [, 9] am Transmission (0 = automatic, 1 = manual)
- [,10] gear Number of forward gears
- [,11] carb Number of carburetors

Display top rows to understand data content and review data summary

```
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
```

Pairs plot of "miles per gallon" and "all other variables"

```
pairs(mtcars[c('mpg','cyl','disp','hp','drat','wt','vs','am')], panel=panel.smooth,
      main='Fig1: mpg vs. parameters with stronger correlation', col = 3)
```

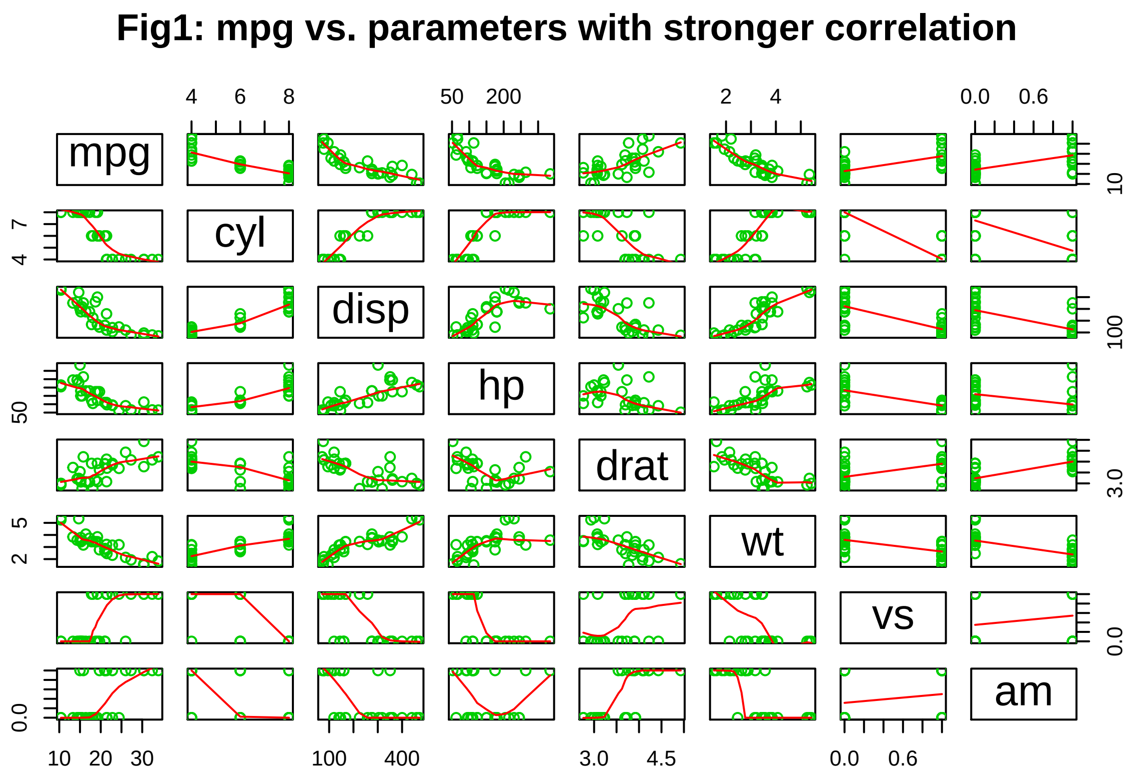
Fig1: mpg vs. parameters with stronger correlation
mpg
cyl
disp
hp
drat
wt
vs
am
4 6 8
50 200
2 4
0.0 0.6
10
100
3.0
0.0
4 7
50
2 5
0.0
10 20 30
100 400
3.0 4.5
0.0 0.6

Figure 1:

Model analysis details

```r
sapply(mtcars,function(x) cor(mtcars$mpg,x))
```

Correlated factors

```
##        mpg        cyl       disp         hp       drat         wt
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594
##       qsec         vs         am       gear       carb
##  0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
```

```r
fit1 <- lm(mpg ~ cyl + disp + hp + drat + wt + vs + am ,mtcars)
summary(fit1)$coef
```

Model of all correlated variables

```
##                Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) 32.45671453 9.02383422  3.5967765 0.001449177
## cyl         -0.63991888 0.89673705 -0.7136082 0.482351934
## disp         0.01347790 0.01211503  1.1124935 0.276945640
## hp          -0.03031547 0.01469183 -2.0634240 0.050048639
## drat         0.54696337 1.51008586  0.3622068 0.720367421
## wt          -3.24531236 1.16753530 -2.7796268 0.010409694
## vs           1.39760892 1.84843389  0.7561044 0.456944578
## am           1.95200821 1.75665421  1.1112080 0.277487428
```

```r
summary(fit1)$adj.r.squared
```

```
## [1] 0.8180107
```

```r
fit2 <- lm(mpg ~ cyl + disp + hp + drat + wt + vs ,mtcars)
summary(fit2)$coef
```

Model of all correlated variables exclusive transmission

```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 34.70295206 8.83570776  3.9275803 0.0005965849
## cyl         -0.97646291 0.84799864 -1.1514911 0.2604241161
## disp         0.01265481 0.01214902  1.0416323 0.3075512215
## hp          -0.02380382 0.01353570 -1.7585955 0.0908834346
## drat         1.04506123 1.44877907  0.7213393 0.4773932259
## wt          -3.72312480 1.09055461 -3.4139737 0.0021892403
## vs           0.50540298 1.67274263  0.3021403 0.7650444642
```

```r
summary(fit2)$adj.r.squared
```

```
## [1] 0.8163016
```

Model of transmission variable

```r
fit3 <- lm(mpg ~ am ,mtcars)
summary(fit3)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

```r
summary(fit3)$adj.r.squared
```

```
## [1] 0.3384589
```

Boxplot showing impact of “transmission type” on “miles per gallon”

The below graph helps us to understand the impact of the “transmission type” on “miles per gallon.”

```r
boxplot(mpg ~ am, data = mtcars, col = (c("green","blue")), ylab = "Miles Per Gallon", xlab = "Transmis
```

Residual plots

The below residual plots outline a bit of a curve in residuals vs. fitted but for the purpose of the analysis we accept the data as normally distributed.

```r
par(mfrow=c(2,2))
plot(fit1)
```

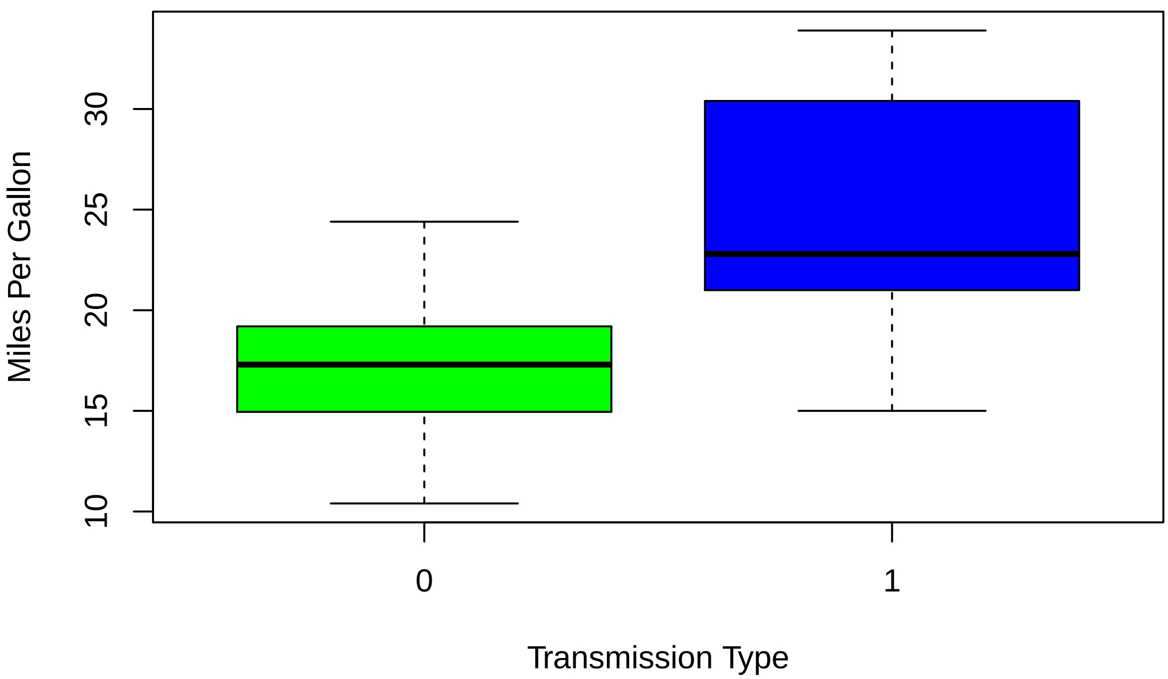

Figure 2:

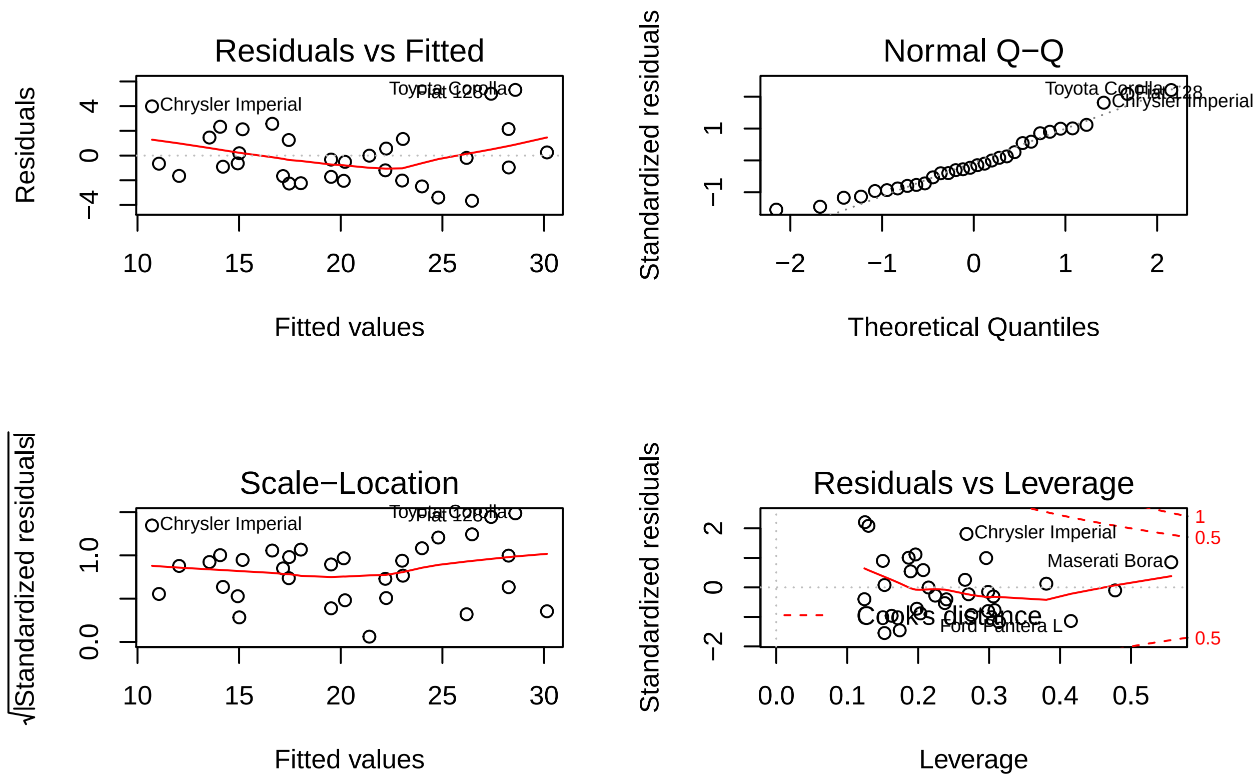
Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Chrysler Imperial
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Leverage
Chrysler Imperial
Maserati Bora
Ford Pantera L
Cook's distance
1
0.5
0.5

Figure 3: