Automobile Industry. Miles/gallon vs Transmission

Coursera, Regression Models, Project

Malgorzata MS

January 31, 2017

Executive Summary

The analyzed data `mtcars` are data available in R datasets package. 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).

This document explores the relationship between a set of variables and miles per gallon (MPG) (outcome). In particular it quantifies the MPG difference between automatic and manual transmission and answers a question if an automatic or manual transmission is better for MPG?

The MPG difference between automatic and manual transmission was quantified with `t.test`. The final model, for the MPG prediction includes weight and horsepower, and is adjusted according to the transmission. Transmission variable (`am`) is included as a factor (0=automatic, 1=manual) and is treated as an interaction term. Hence, the model coefficients - intercept and slope - depend on the transmission. In other words, two sets of coefficients are used to fit the `mtcars` dataset - one set of coefficients for automatic transmission and other set for manual transmission. The transmission effect depends on a value of weight (`wt`) and horsepower (`hp`). The impact of transmission reverses itself depending on this two variables. Therefore, it is not possible to make a general answer if automatic or manual transmission is better for MPG. The transmission effect can be provided only for fixed `wt` and `hp`.

Exploratory Analysis

```r
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 ...
```

The MPG difference between automatic and manual transmission

The difference was quantified with t.test.

```r
with(mtcars, t.test(mpg~am))
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## 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 in group 0 mean in group 1
##        17.14737        24.39231
```

The t.test indicates that the difference in MPG for the two transmissions is statistically siginificant. The p-value for the t.test is equal 0.001374, therefore with significance level alfa equal 0.05, the H0 (no difference) is rejected.

The MPG mean of cars with automatic transmission is: 17.14737 The MPG mean of cars with manual transmission is: 24.39231. Consequently the difference in mpg between automatic and manual transmission is: -7.244939, indicating automiatic transmission being better for MPG.

The confidence interval for the MPG difference is -11.280194 -3.209684.
The boxplot of the mpg dependence on transmission is presented in Figure 1.

Is an automatic or manual transmission better for MPG?

When transmission is considered as the only factor influencing MPG the difference between automatic and manual transmission is significant and indicates lower MPG for cars with automatic transmission. In this part a prediction model for MPG, that consideres other predictors is constructed.

`mtcars` data set has two kinds of variables: continuous (`mpg`, `disp`, `hp`, `drat`, `wt`, and `qsec`), and factors (`cyl`, `vs`, `am`, `gear`, `carb`). First only the continuous variables are considered as predictors for MPG.

Graphical analysis of continuous variables interactions + correlations

First the data are graphically investigated (Appendix, Figure 2). Next the correlations between the `mpg` and other variables is analyzed.

The correlation between `mpg` and other variables decreases in the following order: `wt` (-0.868) > `disp` (-0.848) > `hp` (-0.776) > `drat` (0.681) > `qsec` (0.419)

Testing nested models with anova

Below a function`anova()` is used to test nested models for`mpg`prediction. The nested models are constructed in the order of correlation decrease.

```r
fit1 <- lm(mpg~wt, data=mtcars)
fit2 <- lm(mpg~wt + disp, data=mtcars)
fit3 <- lm(mpg~wt + disp + hp, data=mtcars)
fit4 <- lm(mpg~wt + disp + hp + drat, data=mtcars)
fit5 <- lm(mpg~wt + disp + hp + drat+ qsec, data=mtcars)

anova(fit1, fit2, fit3, fit4, fit5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + disp
## Model 3: mpg ~ wt + disp + hp
## Model 4: mpg ~ wt + disp + hp + drat
## Model 5: mpg ~ wt + disp + hp + drat + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 278.32
## 2     29 246.68  1    31.639 4.8353 0.036978 *
## 3     28 194.99  1    51.692 7.8998 0.009272 **
## 4     27 182.84  1    12.153 1.8573 0.184625
## 5     26 170.13  1    12.708 1.9422 0.175233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova analysis indicates that adding additional variable after Model 3 is not statistically significant (for significance level alfa=0.05). Therefore Model 3 (`fit3`) is considered in the further analysis.

```
summary(fit3)$coeff
```

```
##                  Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.1055052690 2.11081525 17.57875558 1.161936e-16
## wt          -3.8008905826 1.06619064 -3.56492586 1.330991e-03
## disp        -0.0009370091 0.01034974 -0.09053451 9.285070e-01
## hp          -0.0311565508 0.01143579 -2.72447633 1.097103e-02
```

The coefficient of `disp` is very small (-0.000937), almost 0. Therefore have negligible influence. Hence, model: `mpg ~ wt + hp`, is selected as the best model.

```
fit6 <- lm(mpg~wt+hp, data=mtcars); summary(fit6)
```

```
##
## Call:
## lm(formula = mpg ~ wt + hp, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.941 -1.600 -0.182  1.050  5.854
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.22727    1.59879  23.285  < 2e-16 ***
## wt          -3.87783    0.63273  -6.129 1.12e-06 ***
## hp          -0.03177    0.00903  -3.519  0.00145 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.593 on 29 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8148
## F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

Model adjustment for transmission - including `am` as factor variable (interaction term)

```
fit_final <- lm(mpg~(wt + hp) * factor(am), data=mtcars); summary(fit_final)$coef
```

```
##                    Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     30.70392721 2.67515435 11.477441 1.117089e-11
## wt              -1.85591121 0.94510642 -1.963706 6.034159e-02
## hp              -0.04094406 0.01362921 -3.004142 5.826559e-03
## factor(am)1     13.74000384 4.22337051  3.253327 3.155621e-03
## wt:factor(am)1 -5.76894729 2.07200930 -2.784228 9.870579e-03
## hp:factor(am)1  0.02779357 0.01920705  1.447050 1.598330e-01
```

The diagnosic plots for the final fit are presented in Figure 3.

Constructing coefficients for automatic (`am` = 0) and manual (`am` = 1) transmission

```
coefTable<-fit_final$coef
b0 <- coefTable[1]
b1 <- coefTable[2]
b2 <- coefTable[3]
b3 <- coefTable[4]
b4 <- coefTable[5]
b5 <- coefTable[6]

# coefficients for automatic, am = 0
intercept_0 <- b0; slope_0_wt <- b1; slope_0_hp <- b2
# coefficients for manual, am = 1
intercept_1 <- b0 + b3; slope_1_wt <- b1 + b4; slope_1_hp <- b2+ b5
```

Coefficients for the final model:

Coefficient for automatic transmission:

```
##   intercept          wt          hp
## 30.70392721 -1.85591121 -0.04094406
```

Coefficient for manual transmission:

```
##   intercept          wt          hp
## 44.44393104 -7.62485850 -0.01315049
```

For automatic transmission MPG decreases by 1.86 with 1000 lb weight increase and by -0.04 when horsepower increases by 1. For manual transmission the MPG decreases by 7.62 with 1000 lb weight increase and by -0.01 when horsepower increases by 1.

The fitted model (planes) can be seen in Appendix Figure 4.

Appendix

Figure 1

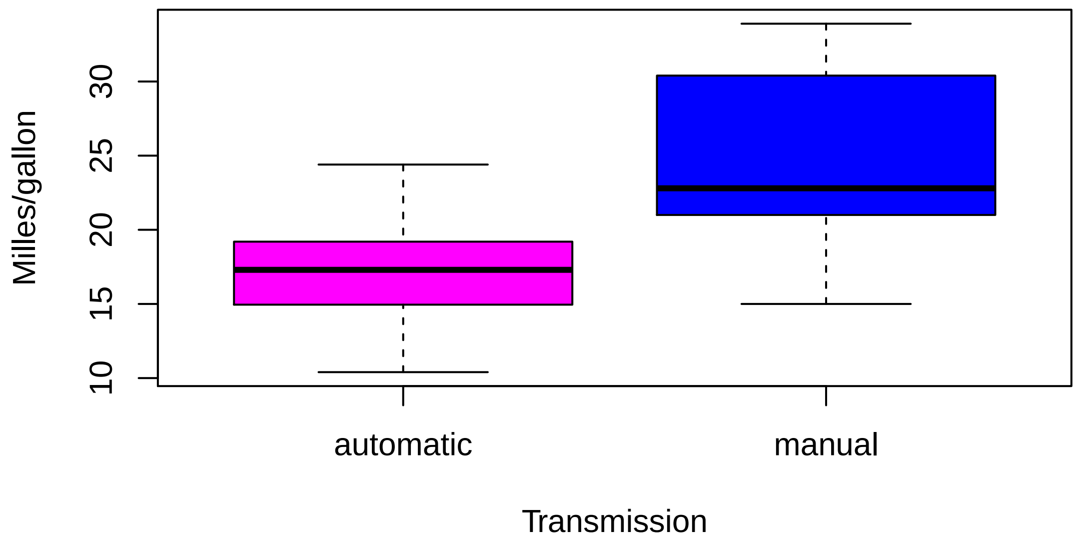


Figure 2

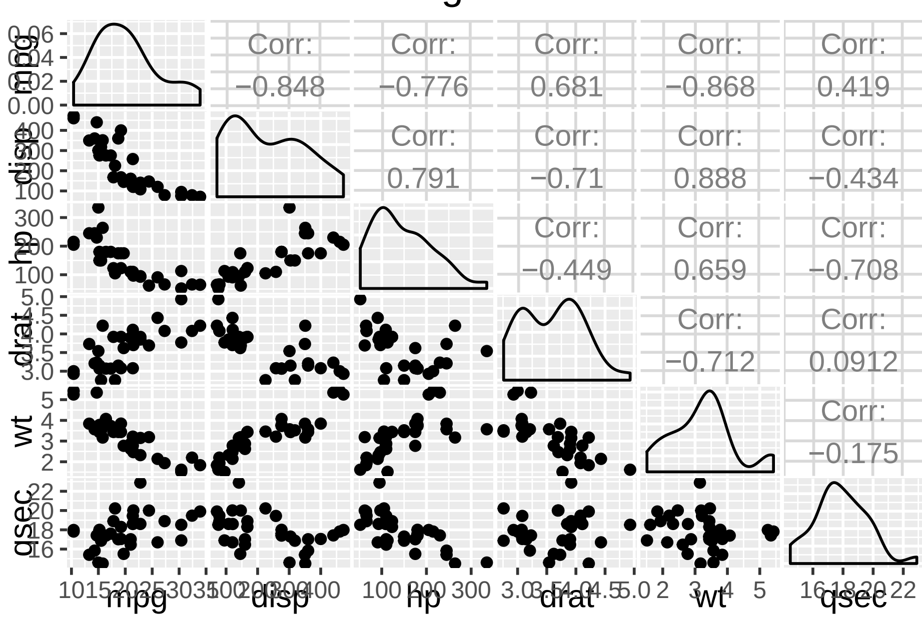
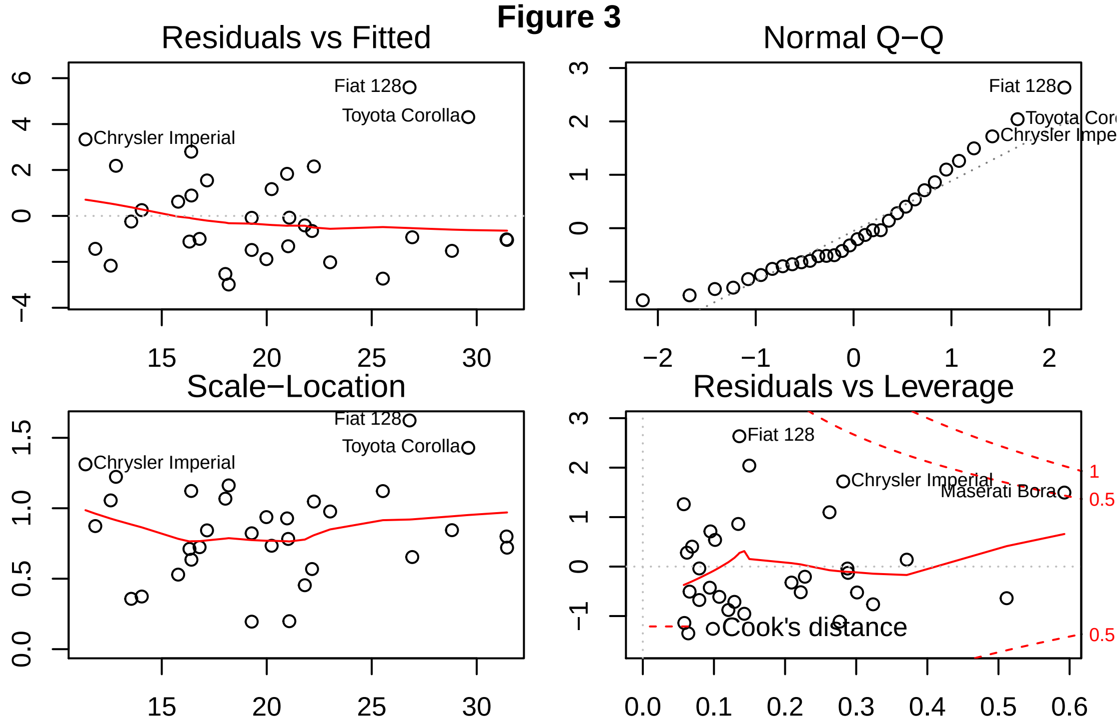


Figure 3



Residuals vs Fitted
Fiat 128
Toyota Corolla
Chrysler Imperial
Normal Q−Q
Scale−Location
Residuals vs Leverage
Maserati Bora
Cook's distance

Figure 4
Visualisation of fit: mpg~(wt + hp) * factor(am)