Automobile transmission type and fuel mileage

Timo Voipio

21 Aug 2016

```
## Warning: package 'knitr' was built under R version 3.2.5
```

Executive summary

Introduction

We analyze the fuel consumption of various automobiles from early 1970s in order to determine whether the transmission type, manual or automatic, significantly affects fuel consumption. The dataset is `mtcars` provided by the R `dataset` package, and it consists of mechanical and performance data on 32 different cars from the model years 1973 and 1974.

The aim of the analysis is two answer two questions: is an automatic or manual transmission better for gas mileage (measured in mpg, miles traveled per one US gallon of fuel consumed), and how the gas mileage is quantitatively affected by the transmission type.

Data description and exploratory analysis

```r
library(datasets)
data("mtcars")

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

```r
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.2.5
```

The data consists of design and performance data of 32, as published in the *Motor Trends* magazine published in the United States in 1974. According to the dataset documentation, the variables in the following table are included. Text in *italics* gives explanatory notes not present in the original data.

Column name	Variable description and units
`mpg`	Miles/(US) gallon (*gas mileage; 235 l/100 km = 1/(1 mpg)*)
`cyl`	Number of cylinders
`disp`	Displacement (cu.in.) (*1 litre = 61.0 cu in*)
`hp`	Gross horsepower (*1 kW = 1.34 hp*)
`drat`	Rear axle ratio (*driveshaft rpm/axle rpm*[1])
`wt`	Weight (1000 lbs = *454 kg*)
`qsec`	1/4 mile time
`vs`	V/S [*V engine (0) or straight (inline) engine (1)*]
`am`	Transmission (0 = automatic, 1 = manual)
`gear`	Number of forward gears
`carb`	Number of carburetors

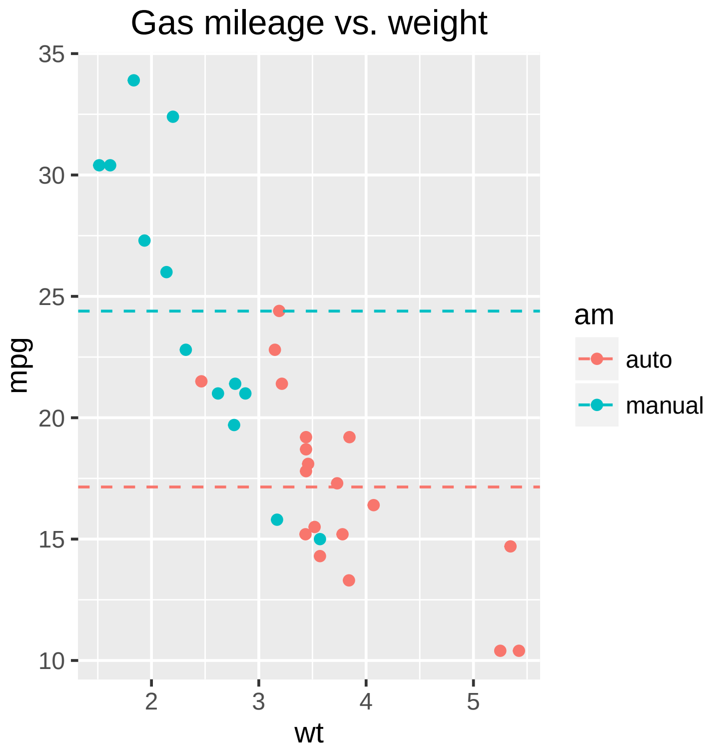
The research question stated in the project assignment exclusively asks for analysis of the gas mileage (MPG), so for consistency reasons the data is not converted into SI units, even though the conversion would make the data more accessible to most parts of the world. Additionally, analysing the consumption via the amount of fuel consumed per fixed distance (e.g., litres/100 km) would be both physically and statistically more reasonable choice[2] than gas mileage. However, as the assignment explicitly asks for effect of transmission type on gas mileage, the data will not be converted for analysis. As suggested by Henderson and Velleman, a new variable `pwr` (hp/1000 lbs) is created for the power-to-weight ratio.

```
# Convert transmission type, (cylinder count, carburetor count), and engine type
# to factor variables
mtcars$am <- factor(mtcars$am, levels = c(0, 1),
                    labels = c("auto", "manual"))
#mtcars$cyl <- factor(mtcars$cyl)
#mtcars$carb <- factor(mtcars$carb)
mtcars$vs <- factor(mtcars$vs, levels = c(0, 1),
                    labels = c("v", "straight"))

# Create a new variable for gross power/weight ratio
# (hp/1000 lb; 1 hp/1000 lb   1.64 W/kg)
mtcars$pwr <- mtcars$hp/mtcars$wt

g <- ggplot(mtcars, aes(x = wt, y = mpg))
g <- g + geom_point(aes(color = am))
g <- g + geom_hline(aes(yintercept = mpg, color = am),
                    data = aggregate(mpg ~ am, data = mtcars, mean),
```

[1] ”the drive-axle ratio is a comparison of the number of gear teeth on the ring gear of the rear axle and the pinion gear on the driveshaft. - - For example, a 4.11:1 ratio means there are 4.11 teeth on the axle’s ring gear for each tooth on the driveshaft’s pinion gear. Or, put another way, the driveshaft must turn 4.11 times to turn the rear wheels one full revolution. - - typical rule of thumb: The higher the numerical ratio, the slower the gear will be. This higher ratio gives a truck greater pulling power, but since the engine must work harder to spin the driveshaft more times for each turn of the rear wheels, top-end speed and fuel economy are sacrificed.” From [worktruckonline.com][drat]

[2] As suggested also by Henderson and Velleman

```
                    linetype = "dashed")
g <- g + ggtitle("Gas mileage vs. weight")

print(g)
```



The plot shows that the mean gas mileage of cars with manual transmission is significantly higher (i.e., better) compared to cars with automatic transmission. However, the plot similarly illustrates that the gas mileage seems to have a clear negative correlation with the weight of the vehicle.

```
fit1 <- lm(mpg ~ am, mtcars)
fitall <- lm(mpg ~ . - hp, mtcars)

summary(fit1)
```

```
##
## 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
```

```
fit2.w <- lm(mpg ~ am + wt, mtcars)
fit2.d <- lm(mpg ~ am + disp, mtcars)
fit2.h <- lm(mpg ~ am + hp, mtcars) # greatest reduction in sum(^2)

anova(fit1, fit2.w)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
##   Res.Df    RSS Df Sum of Sq      F       Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 46.115 0.0000001867 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit1, fit2.d)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + disp
##   Res.Df    RSS Df Sum of Sq      F       Pr(>F)
## 1     30 720.90
## 2     29 300.28  1    420.62 40.621 0.0000005748 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit1, fit2.h)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 56.178 2.92e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit3.w <- lm(mpg ~ am + hp + wt, mtcars) # greatest reduction in sum(^2)
fit3.d <- lm(mpg ~ am + hp + disp, mtcars)
fit3.c <- lm(mpg ~ am + hp + cyl, mtcars)

anova(fit1, fit2.h, fit3.w)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 73.841 2.445e-09 ***
## 3     28 180.29  1     65.15 10.118  0.003574 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit1, fit2.h, fit3.d)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 58.8793 2.332e-08 ***
## 3     28 226.10  1     19.34  2.3945     0.133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit1, fit2.h, fit3.c)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + cyl
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 60.3610 1.834e-08 ***
## 3     28 220.55  1     24.89  3.1594   0.08636 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit4.d <- lm(mpg ~ am + hp + wt + disp, mtcars)
fit4.c <- lm(mpg ~ am + hp + wt + cyl, mtcars) # smallest sum(^2), but fit4.c not re

anova(fit1, fit2.h, fit3.w, fit4.d)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp + wt + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 71.3552 4.646e-09 ***
## 3     28 180.29  1     65.15  9.7773    0.0042 **
## 4     27 179.91  1      0.38  0.0576    0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit1, fit2.h, fit3.w, fit4.c)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp + wt + cyl
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 75.5148 2.638e-09 ***
## 3     28 180.29  1     65.15 10.3472  0.003356 **
## 4     27 170.00  1     10.29  1.6348  0.211917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit5 <- lm(mpg ~ am + hp + wt + disp + cyl, mtcars)

anova(fit1, fit2.h, fit3.w, fit4.d, fit5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp + wt + disp
## Model 5: mpg ~ am + hp + wt + disp + cyl
```

```
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)    
## 1     30 720.90                                   
## 2     29 245.44  1    475.46 75.7841 3.499e-09 ***
## 3     28 180.29  1     65.15 10.3841  0.003408 ** 
## 4     27 179.91  1      0.38  0.0611  0.806673    
## 5     26 163.12  1     16.79  2.6758  0.113932    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit1, fit2.h, fit3.w, fit4.c, fit5)
```

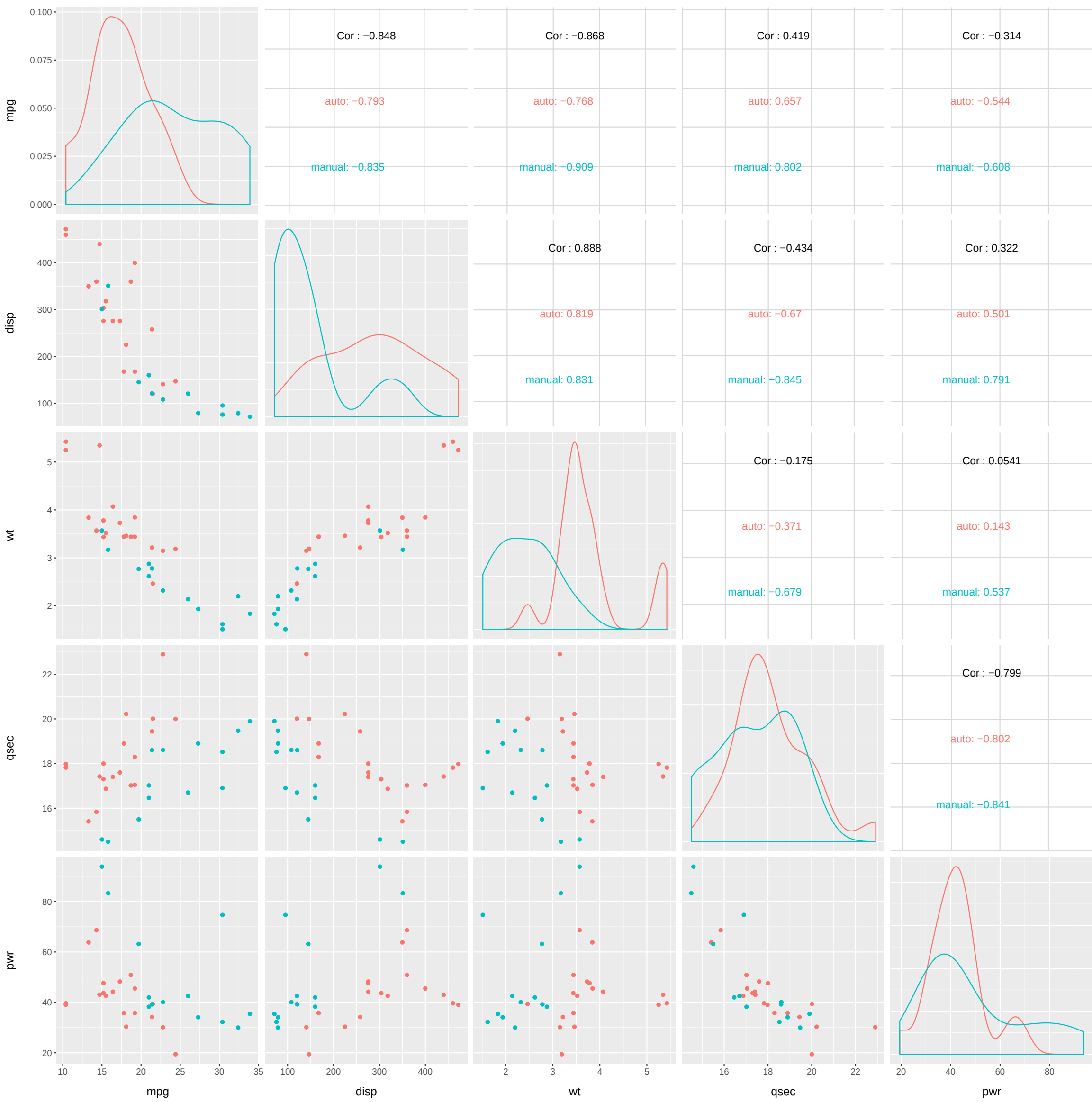
```
## Analysis of Variance Table
## 
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp + wt + cyl
## Model 5: mpg ~ am + hp + wt + disp + cyl
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)    
## 1     30 720.90                                   
## 2     29 245.44  1    475.46 75.7841 3.499e-09 ***
## 3     28 180.29  1     65.15 10.3841  0.003408 ** 
## 4     27 170.00  1     10.29  1.6407  0.211542    
## 5     26 163.12  1      6.88  1.0963  0.304719    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sources

- Hocking (1971) http://www.jstor.org/stable/2529336
- Henderson and Velleman (1981) http://www.jstor.org/stable/2530428

Appendix

```r
paircols <- c("mpg", "disp", "wt", "qsec", "pwr")
ggpairs(mtcars, aes(color = am),
        columns = match(paircols, names(mtcars)))
```


Cor : −0.848
auto: −0.793
manual: −0.835
Cor : −0.868
auto: −0.768
manual: −0.909
Cor : 0.419
auto: 0.657
manual: 0.802
Cor : −0.314
auto: −0.544
manual: −0.608
Cor : 0.888
auto: 0.819
manual: 0.831
Cor : −0.434
auto: −0.67
manual: −0.845
Cor : 0.322
auto: 0.501
manual: 0.791
Cor : −0.175
auto: −0.371
manual: −0.679
Cor : 0.0541
auto: 0.143
manual: 0.537
Cor : −0.799
auto: −0.802
manual: −0.841
mpg
disp
wt
qsec
pwr