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

Exploratory analysis

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

library(ggplot2)
```

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

```
library(GGally)
```

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

The dataset includes the following variables:

Column name	Variable description and units
`mpg`	Miles/(US) gallon (*gas mileage*)
`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])

[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." Quoted from [here][drat]

Column name	Variable description and units
`wt`	Weight (1000 lbs)
`qsec`	1/4 mile time
`vs`	V/S (*V engine or straight engine*)
`am`	Transmission (0 = automatic, 1 = manual)
`gear`	Number of forward gears
`carb`	Number of carburetors

```
# Possibly relevant covariates:
# disp, hp, cyl (likely correlated with each other)
# wt
# drat
# am (0 auto/1 manual; subject of interest)
# Outcome:
# mpg

# Convert transmission and cylinder count to factor variables
mtcars$trans <- factor(mtcars$am, levels = c(0, 1),
                       labels = c("auto", "manual"))
mtcars$cylf <- factor(mtcars$cyl)

fit1 <- lm(mpg ~ trans, mtcars)

g <- ggplot(mtcars, aes(x = trans, y = mpg))
g <- g + geom_boxplot() +
    stat_summary(fun.y = mean, geom = "point")
g <- ggtitle ("Gas mileage and transmission type")

print(g)
```

```
## $title
## [1] "Gas mileage and transmission type"
##
## attr(,"class")
## [1] "labels"
```

```
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ trans, 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 ***
## transmanual    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 ~ trans + wt, mtcars)
fit2.d <- lm(mpg ~ trans + disp, mtcars)
fit2.h <- lm(mpg ~ trans + hp, mtcars) # greatest reduction in sum(^2)

anova(fit1, fit2.w)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + 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 ~ trans
## Model 2: mpg ~ trans + 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 ~ trans
## Model 2: mpg ~ trans + 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
```

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

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

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + hp
## Model 3: mpg ~ trans + 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
```

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

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + hp
## Model 3: mpg ~ trans + 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
```

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

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + hp
## Model 3: mpg ~ trans + hp + cylf
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
```

```
## 2     29 245.44  1    475.46 65.0984 1.142e-08 ***
## 3     27 197.20  2     48.24  3.3025   0.05212 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit4.d <- lm(mpg ~ trans + hp + wt + disp, mtcars)
fit4.c <- lm(mpg ~ trans + hp + wt + cylf, mtcars) # smallest sum(^2), but fit4.c no

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

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + hp
## Model 3: mpg ~ trans + hp + wt
## Model 4: mpg ~ trans + 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
```

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

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + hp
## Model 3: mpg ~ trans + hp + wt
## Model 4: mpg ~ trans + hp + wt + cylf
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 81.8529 1.634e-09 ***
## 3     28 180.29  1     65.15 11.2157  0.002484 **
## 4     26 151.03  2     29.27  2.5191  0.099998 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit5 <- lm(mpg ~ trans + hp + wt + disp + cylf, mtcars)

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

```
## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + hp
## Model 3: mpg ~ trans + hp + wt
## Model 4: mpg ~ trans + hp + wt + disp
## Model 5: mpg ~ trans + hp + wt + disp + cylf
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 79.0275 3.26e-09 ***
## 3     28 180.29  1     65.15 10.8285 0.002973 **
## 4     27 179.91  1      0.38  0.0637 0.802747
## 5     25 150.41  2     29.50  2.4516 0.106610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans
## Model 2: mpg ~ trans + hp
## Model 3: mpg ~ trans + hp + wt
## Model 4: mpg ~ trans + hp + wt + cylf
## Model 5: mpg ~ trans + hp + wt + disp + cylf
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 79.0275 3.26e-09 ***
## 3     28 180.29  1     65.15 10.8285 0.002973 **
## 4     26 151.03  2     29.27  2.4322 0.108353
## 5     25 150.41  1      0.62  0.1025 0.751489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```