Analysis of the Impact of Transmission Type on MPG

Wiktoria Urantowka

9/10/2017

EXECUTIVE SUMMARY

This short analysis investigates the impact of the type of transition (Manual and Authomatic) on the cars meters per gallon values (mpg). It shows that the type of transition is not very relevant whenever other car's caracteristics are as well considered. What seems to have the most (statistically) significant impacts on mpg is car's weight

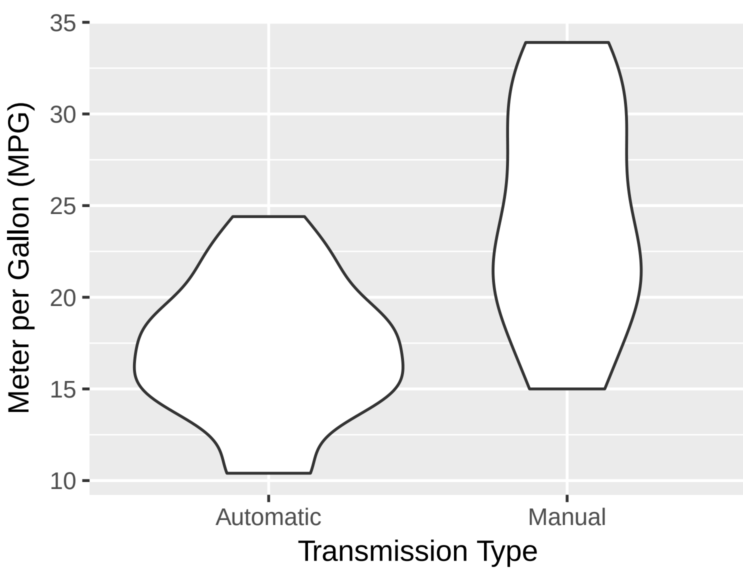
EXPLORATORY ANALYSIS

Cars with manual transition type have on avarage 7 more mpg than those with authomatic transmission as well as higher dispersion as shown below

```
data(mtcars)
aggregate(mpg ~ am, data=mtcars, mean)
```

```
##   am      mpg
## 1  0 17.14737
## 2  1 24.39231
```

```
library("ggplot2")
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
violin = ggplot(data = mtcars, aes(y = mpg, x = am))
violin = violin + geom_violin(alpha = 1)
violin = violin + xlab("Transmission Type") + ylab("Meter per Gallon (MPG)")
violin = violin + scale_fill_discrete(name = "Transmission Types", labels=c("Automatic", "Manual"))
violin
```



REGRESSION ANALYSIS

1. chosen method:

Ordinary Least Squares (Dependent variable is neither of binary nor count type (see figure 1 in appendix). It's values can only be positive, but even in this case OLS remain a resonable approach)

2. selection of controls:

Pick those with that have the highest correlation with the response variable

```
mtcars$am <- as.numeric(mtcars$am)
##mtcars$mpg <- as.numeric(mtcars$mpg)
corr <- cor(as.matrix(mtcars[,1]), as.matrix(mtcars[,-1]))
corr
```

```
##             cyl       disp         hp      drat         wt     qsec
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
##              vs        am      gear       carb
## [1,] 0.6640389 0.5998324 0.4802848 -0.5509251
```

3.Model selection:

Include covariates with the higest correlation, then add on the top the one with the second highest correlation etc. Stop at the level where adjusted R squared doesnt increase.

Nested models:
0. mpg_i = alpha + betha * am_i + error_i (benchmark regression)
1. mpg_i = alpha + betha * am_i + wt_i + error_i
2. mpg_i = alpha + betha * am_i + wt_i + cyl_i + error_i
3. mpg_i = alpha + betha * am_i + wt_i + cyl_i + disp_i + error_i
4. mpg_i = alpha + betha * am_i + wt_i + cyl_i + disp_i + hp_i + error_i
5. mpg_i = alpha + betha * am_i + wt_i + cyl_i + disp_i + hp_i + drat_i + error_i

where:

mpg = gallons per meter
am = Transition type, binary factor variable: authomatic or manual
wt = car's weight, numerical
cyl = number of cylinders, factor variable
disp = displacement, numerical

RESULTS AND DISCUSSION

Benchmark regression has statistically significant both constant term (Authomatic transmission type) and the regressor (Manual transmission type) with the coefficient of 7, meaning the switch from Authomatic to Manual will be associated with the increase of mpg of 7 units.

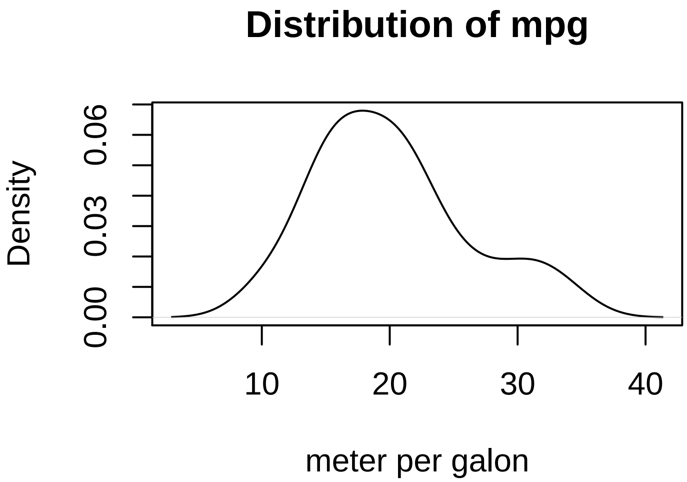
When we add more controls to the regression however, the first coefficient (Manual transition type) looses its statistical significance and passes it forward onto new variables.

Model 4. mpg_i = betha * am_i + wt_i + cyl_i + disp_i + hp_i + error_i chosen according to discussed above criteria (results of selection not shown, residual plot in appendix not displaying any apparent patterns) illustrates (results in appendix) that what seems the most relevant in explaining mps is the car's weigth and not the type of trasmission. While isolating from the effects of am, cyl, disp, hp, the increase of 1000 lbs in weight is associated with decrease of 3 mpg. The switch from authomatic to manual transition (as long as a car has a transition mode) doesn't not seem to have impact on mpg.

APPENDIX

FIGURE 1

```r
plot(density(mtcars$mpg), main = "Distribution of mpg", xlab = "meter per galon")
```



RESULTS OF BENCHMARK REGRESSION

```r
fit0<-lm(formula = mpg ~ factor(am), data = mtcars)
summary(fit0)
```

```
##
## Call:
## lm(formula = mpg ~ factor(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 ***
## factor(am)2    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
```

RESULTS OF REGRESSION 4

```r
fit4<-lm(mpg ~ factor(am) + wt + cyl + disp + hp, data = mtcars)
summary(fit4)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt + cyl + disp + hp, data = mtcars)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5952 -1.5864 -0.7157  1.2821  5.5725
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.20280    3.66910  10.412 9.08e-11 ***
## factor(am)2  1.55649    1.44054   1.080  0.28984
## wt          -3.30262    1.13364  -2.913  0.00726 **
## cyl         -1.10638    0.67636  -1.636  0.11393
## disp         0.01226    0.01171   1.047  0.30472
## hp          -0.02796    0.01392  -2.008  0.05510 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.505 on 26 degrees of freedom
## Multiple R-squared:  0.8551, Adjusted R-squared:  0.8273
## F-statistic:  30.7 on 5 and 26 DF,  p-value: 4.029e-10
```

RESIDUALS VS FITTED VALUES

```
fit4<-lm(mpg ~ factor(am) + wt + cyl + disp + hp, data = mtcars)
resid<-resid(fit4)
predict<-predict(fit4, data=mtcars)
plot(predict, resid)
```