MtCars Analysis

Fabiano Andrade Lima

December 22, 2015

Executive Summary

There has been a great interest on car's efficiency these days. As the prices at the pumps increases, fuel consumption becomes a matter of importance both for car manufactures and consumers. Because of that, we have been asked to make a analysis of the fuel consumption efficiency on 32 '73 and '74 car models.

One aspect of interest is whether automatic or manual transmission has significant impact on fuel consumption efficiency. So our focus is going to be figuring out if there is a relationship between this two variables, and if it's possible to state something about it.

Exploratory Analysis

First of all, As we can see on the table below, we have eleven variables for each car model.

Var	Description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (lb/1000)
qsec	1/4 mile time
vs	V/S
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

Our departure point will be a analysis of variance between all the variables in relation with mpg.

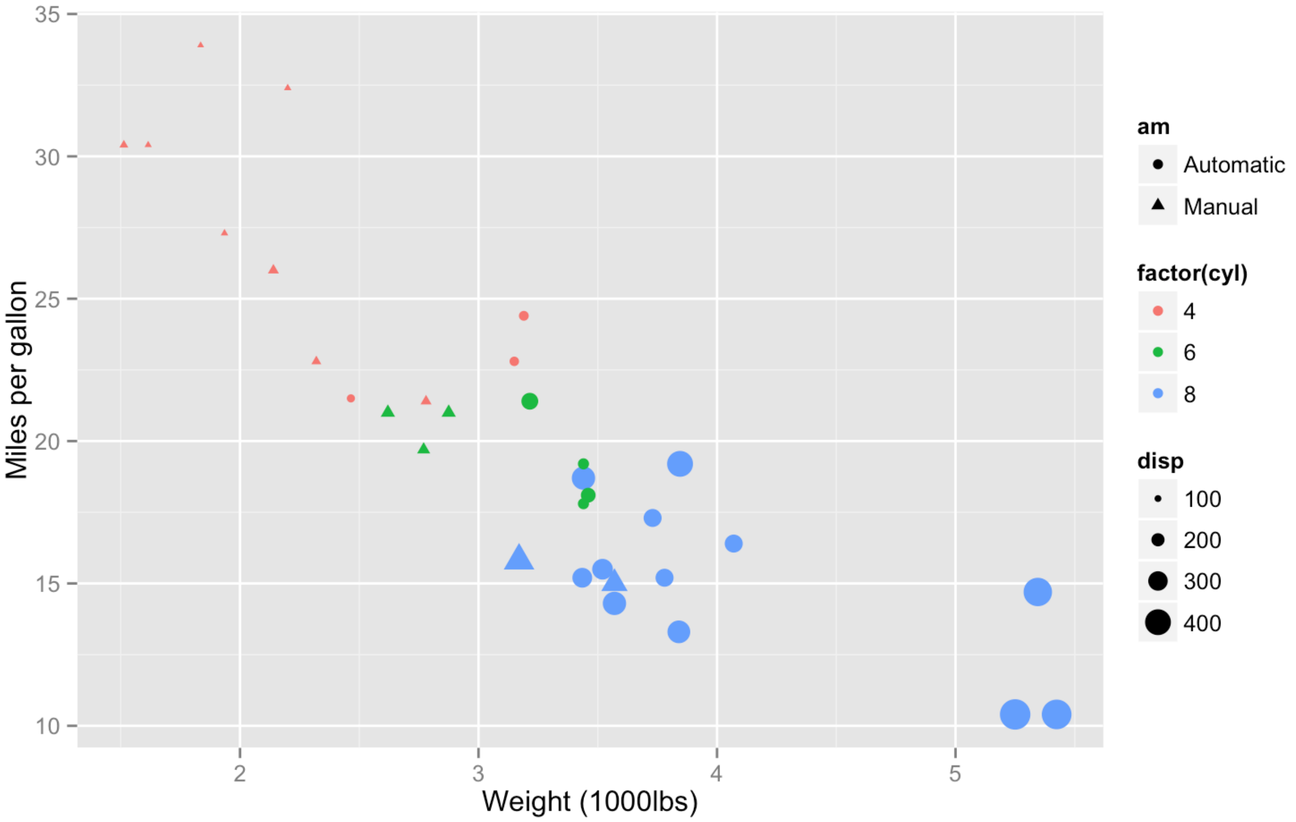
```
summary(aov(mpg ~ ., data=mtcars))
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## cyl           1  817.7   817.7 116.425 5.03e-10 ***
## disp          1   37.6    37.6   5.353  0.03091 *  
## hp            1    9.4     9.4   1.334  0.26103    
## drat          1   16.5    16.5   2.345  0.14064    
## wt            1   77.5    77.5  11.031  0.00324 ** 
## qsec          1    3.9     3.9   0.562  0.46166    
## vs            1    0.1     0.1   0.018  0.89317    
## am            1   14.5    14.5   2.061  0.16586    
## gear          1    1.0     1.0   0.138  0.71365    
## carb          1    0.4     0.4   0.058  0.81218    
## Residuals    21  147.5     7.0                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we see, tree variables seems to have a good significance level in changing the variance of mpg: Cylinders, displacement and weight.

On the plot below one can see how these variables interact with each other. The manual cars (represented by the triangles) usually have better mpg, but it seems that it's more a effect of having less weight than the automatic cars. We can see that the green/blue triangles are in the same cluster (in terms of mpg) as the green/blue points, meaning that what makes difference is not whether the transmission is automatic or manual, but if it is heavier, or has more cylinders or has a bigger displacement.

```
gmtcars <- mtcars
gmtcars$am <- as.factor(ifelse(mtcars$am == 0,'Automatic','Manual'))
g = ggplot(gmtcars, aes(y = mpg, x = wt))
g <- g + geom_point(aes(shape = am, colour=factor(cyl), size = disp))
g = g + ylab("Miles per gallon")
g = g + xlab("Weight (1000lbs)")
g
```



Statistical Study

Now to statistically answer the question of whether being automatic or manual has some effect in fuel consumption, we need to find out a good model that explains well our data, and then test the hypothesis of whether transmission type has impact on mpg.

To correctly select the model, and taking in consideration the analysis of variance already applied to the data, we performed a likelihood ratio tests for analysis of variance of [mpg ~ wt] model with nested models including cylinders and displacement.

```
fit0 <- lm(mpg ~ am, data=mtcars) #for latter use

fit1 <- lm(mpg ~ wt, data=mtcars)
fit2 <- lm(mpg ~ wt + cyl, data=mtcars)
fit3 <- lm(mpg ~ wt + cyl + disp, data=mtcars)
anova(fit1, fit2, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + cyl
## Model 3: mpg ~ wt + cyl + disp
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 278.32
## 2     29 191.17  1     87.15 12.946 0.001221 **
## 3     28 188.49  1      2.68  0.398 0.533217
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we see, the best model seems to be the “fit2”, with a significance value of 0.001.

Thus, as we see below, holding weight and cylinders constant, the significance value for changing from manual to automatic is 0.89. If we consider a maximum value of 0.05% as the threshold to state that transmission type has a effect on mpg, we can discard the alternative hypothesis in favor of the null hypothesis: transmission type has no effect on mpg.

```
fit5 <- lm(mpg ~ wt + cyl + factor(am), data = mtcars)
summary(fit5)
```

```
##
## Call:
## lm(formula = mpg ~ wt + cyl + factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.4179     2.6415  14.923 7.42e-15 ***
## wt           -3.1251     0.9109  -3.431  0.00189 **
## cyl          -1.5102     0.4223  -3.576  0.00129 **
## factor(am)1   0.1765     1.3045   0.135  0.89334
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```

We also can see in the plot 1 at the appendix a comparison of residuals with a linear model with mpg and transmission, mpg, weight and cylinders and mpg, weight, cylinders and transmission. The last two remains practically the same, showing that indeed transmission type has no effect on mpg.

Conclusion

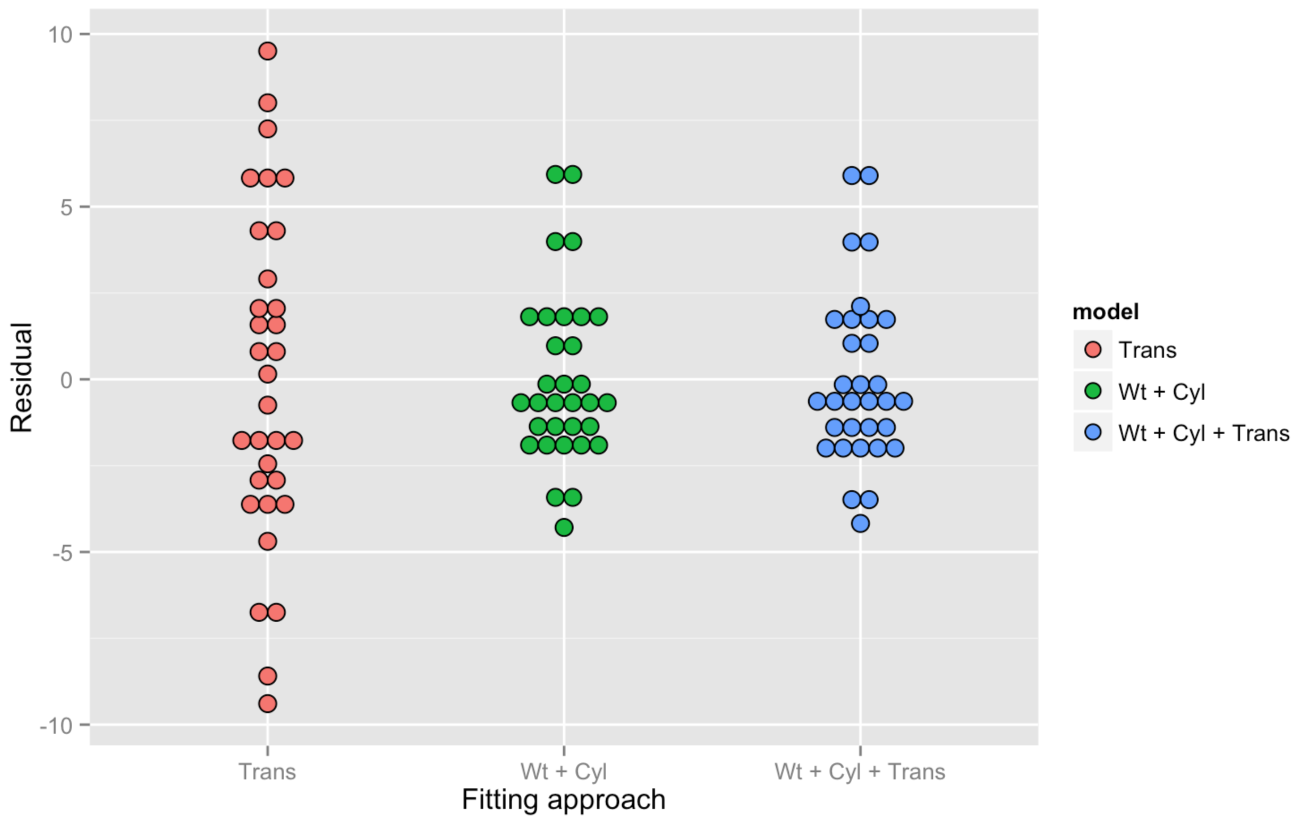
Transmission type has no significance effect on fuel consumption efficiency. If we fit a linear regression to explain the variation in mpg only by the transmission type we get 33.85% of explained variance, but that doesn’t hold wen taking in account other variables.

When we took in consideration the weight and the cylinders, we were able to explain 81.85% of the variation in mpg, and both variables with high level of significance.

We also saw in plot 1 (at the appendix) that the residuals of a model with mpg, weight, cylinders and transmission almost didn’t change from that with just mpg, weight and cylinders. This confirms that when taking in consideration the weight and cylinders, transmission type has no effect on for mpg.

Appendix

```
residuals <- c(resid(fit0), resid(fit2), resid(fit5))
model <- factor(c(rep('Trans',nrow(mtcars)), rep('Wt + Cyl', nrow(mtcars)), re
p('Wt + Cyl + Trans', nrow(mtcars))))
g = ggplot(data.frame(residuals = residuals, model = model), aes(y = residuals,
x = model, fill = model))
g = g + geom_dotplot(binaxis = "y", size = 2, stackdir = "center", binwidth =
0.5)
g = g + xlab("Fitting approach")
g = g + ylab("Residual")
g
```



Plot 1: Showing difference in residuals between a model with only transmission type as the predictor and a model with weight and cylinders as predictors for mpg.