Coursera-Regression Models-Course Project

Carolina Mourao de Paula

Thursday, July 23, 2015

Executive Summary

In this report we will address the following two questions:

- Is an automatic or manual transmission better for MPG?
- Quantify the MPG difference between automatic and manual transmissions

We will use data that examines fuel efficiency and other 10 aspects of automobile design and performance for 32 automobiles. Some aspects about the data:

- Transmission types can influence the fuel efficiency.
- Manual vehicles achieved a 7.2 miles per gallon more than automatic vehicles.
- Although that can drive a conclusion, the transmission type is not a particularly good predictor of fuel efficiency.
- We were able to run analysis on the cars' variables and build models that concluded cylinders and weight are actually good predictors for fuel efficiency.

Methodology

First we looked at the data to have an idea of what we were looking at. We see that, next to mpg, we have 10 numeric columns. The other columns represented potential variables for analysis in a regression model to drive conclusions on fuel efficiency.

```
data(mtcars)
str(mtcars)  # See appendix for results
```

Exploratory data analysis

We ran ANOVA, R squared and pairs plot to identify trends in the different variables applied to a car. The patterns identified were:

- An increase in number of cylinders tends to lead to a decrease in mpg.
- An increase in hp tends to lead to a decrease in mpg.
- An increase in rear axle ratio tends (though more variation) to lead to an increase in mpg.
- An increase in weight, seems to lead to a decrease in mpg

Results

The test results listed in the appendix section presented some potential variables as regressors, however further analysis filter them down to basically four main variables that were able to relate better to the mpg index. The best model was the model 3, which combined two regressors: cylinder and weight.

Its adjusted R squared was **0.8121603**

The coefficients in model 3 were:

```
## Warning: package 'xtable' was built under R version 3.1.3
```

% latex table generated in R 3.1.2 by xtable 1.7-4 package % Fri Jul 24 14:14:56 2015

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	39.4179	2.6415	14.92	0.0000
cyl	-1.5102	0.4223	-3.58	0.0013
wt	-3.1251	0.9109	-3.43	0.0019
am	0.1765	1.3045	0.14	0.8933

This coefficient table indicated that assuming the weight stayed the same, for one additional cylinder, the mpg decreased by 1.5. Also, as the weight increased by 1000 pounds and number of cylinders stayed the same, the mpg decreased by 3.2.

Since the question raised was for the transmission type, we considered the relationship between type of transmission and fuel efficiency in Model 7:

```r
fit7 <- lm(mpg ~ am, data = mtcars)
print(xtable(fit7), type="latex")
```

% latex table generated in R 3.1.2 by xtable 1.7-4 package % Fri Jul 24 14:14:56 2015

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.1474	1.1246	15.25	0.0000
am	7.2449	1.7644	4.11	0.0003

The coefficient for transmission indicated that manual transmission achieved a fuel effiency of 7.24 miles per gallon higher than cars with an automatic transmission. However this model using transmission achieves an adjusted R squared value of **0.3384589**, which is much worse than model 3, so transmission is much poorer predictor of fuel efficiency than the number of cylinders and weight.

Conclusion

Although in this data set on average manual vehicles achieve a fuel efficency of 7.2 miles per gallon more than automatic vehicles, transmission type was not a particularly good predictor of fuel efficiency. We were able to identify that the number of cylinders and the weight of the automobile were good predictors of fuel efficiency, achieving an adjusted R squared of 0.82. If we add transmission type to this model, then the difference in fuel efficency for a manual transmission is much smaller, just 0.13 miles per gallon for a vehicle with the same weight, displacement and number of cylinders. Therefore we conclude that number of cylinders and weight are good predictors of fuel efficiency, but transmission type is not.

Appendix

Data set profile results

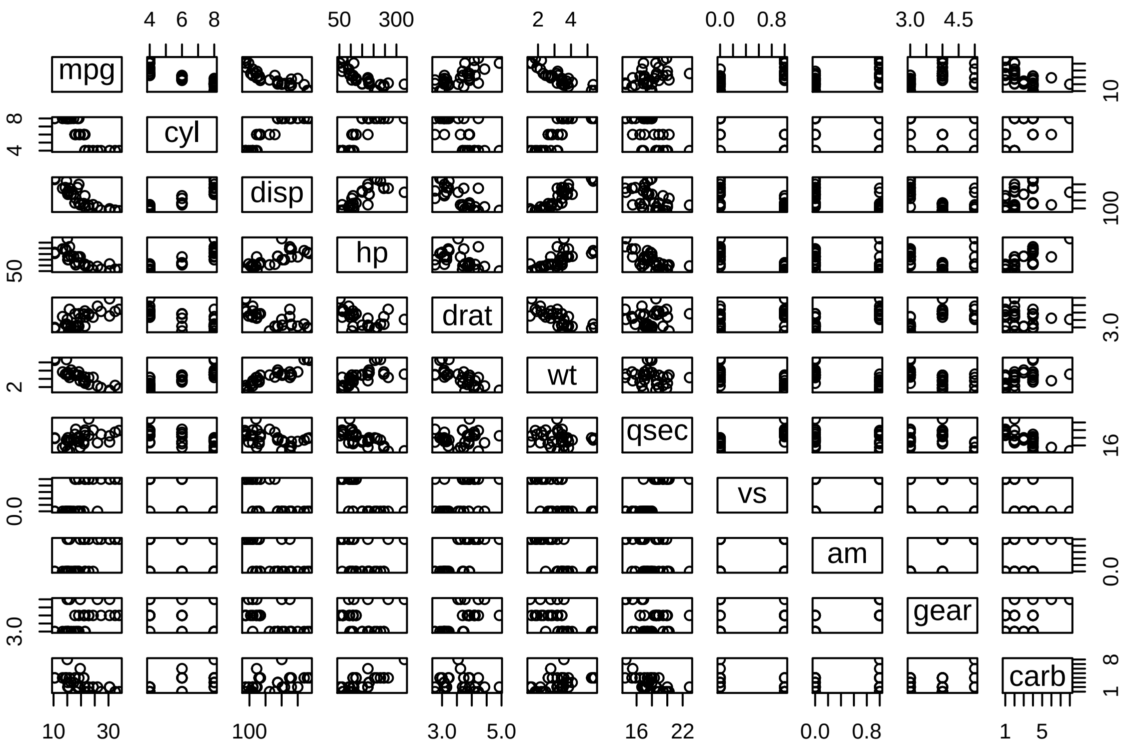
```
str(mtcars)
```

```
## '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 ...
```

Pairs plot analysis results

The plots showed a visual representation of trends in the first column, which relates to the mpg values. We noted trends of increase and decrease in mpg values in the following variables: cylinder(cyl), displacement(disp), Gross horsepower(hp), Rear axle ration(drat), weight(wt), 1/4 mile time(qsec) and Number of carburetors(carb).

```
pairs(mpg ~ ., data=mtcars)
```



Anova analysis results

Next, the analysis of variance allowed the observation of the significant values with a P-value less than 0.05. Observe cylinder(cyl), weight(wt) and displacement(disp) have significant representation and low P-value. Therefore this analysis directed us to more substancial indication of which variables to create models with.

```
data(mtcars)
options(contrasts=c("contr.sum", "contr.poly"))
aov.1 <- aov(mpg ~ ., data=mtcars)
print(xtable(aov.1), type="latex")
```

% latex table generated in R 3.1.2 by xtable 1.7-4 package % Fri Jul 24 14:14:57 2015

R squared analysis

Next we built a different number of models, based on different combinations of cylinder, weight and displacement to find the best R squared index.

Fitting different models

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
cyl	1	817.71	817.71	116.42	0.0000
disp	1	37.59	37.59	5.35	0.0309
hp	1	9.37	9.37	1.33	0.2610
drat	1	16.47	16.47	2.34	0.1406
wt	1	77.48	77.48	11.03	0.0032
qsec	1	3.95	3.95	0.56	0.4617
vs	1	0.13	0.13	0.02	0.8932
am	1	14.47	14.47	2.06	0.1659
gear	1	0.97	0.97	0.14	0.7137
carb	1	0.41	0.41	0.06	0.8122
Residuals	21	147.49	7.02		

```r
fit3 <- lm(mpg ~ cyl + wt, data = mtcars)
fit4 <- lm(mpg ~ disp, data = mtcars)
fit5 <- lm(mpg ~ disp + cyl, data = mtcars)
fit6 <- lm(mpg ~ disp + cyl + wt, data = mtcars)
```

The R squared values were:

- Model 1: Cylinders to fuel efficiency **0.7170527**.
- Model 2: Weight to fuel efficiency **0.7445939**.
- Model 3: Cylinders and weight to fuel efficiency **0.8185189**.
- Model 4: Displacement to fuel efficiency **0.7089548**.
- Model 5: Displacement and cylinders to fuel efficiency **0.7429841**.
- Model 6: Displacement, cylinders and weight to fuel efficiency **0.8146721**.

The original question though was related to the transmission variable. We created a model for the transmission to also look at its R squared value.

```r
fit7 <- lm(mpg ~ am, data = mtcars)
```

The R squared value was **0.3384589**. This low value indicated that transmission was not a good regressor.

The last model built, included all the substantial variables found and also the transmission variable.

```r
fit8 <- lm(mpg ~ cyl + wt + disp + am, data = mtcars)
print(xtable(fit8), type="latex")
```

% latex table generated in R 3.1.2 by xtable 1.7-4 package % Fri Jul 24 14:14:57 2015

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	40.8983	3.6015	11.36	0.0000
cyl	-1.7842	0.6182	-2.89	0.0076
wt	-3.5834	1.1865	-3.02	0.0055
disp	0.0074	0.0121	0.61	0.5451
am	0.1291	1.3215	0.10	0.9229

The model using cylinders, weight, displacement and transmission achieved an adjusted R squared value of **0.8078759**. This is slightly worse than model 3, that did not use transmission. If we look at the coefficients, we see that the coefficient for transmission type has decreased, indicating that for the same number of cylinders, weight and displacement, this model predicts that a car with a manual transmission would achieve an improvement in fuel efficiency of just 0.13 miles per gallon.