Automatic versus Manual Transmission on Miles per Galon Performance

Alexandre Xavier Ywata Carvalho

Sunday, October 26, 2014

Executive Summary

In this project, we compared the performance, in terms of miles per galon, between automatic and manual cars, for Motor Trend magazine. Looking at a data set of a collection of cars, we are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). We are particularly interested in the following two questions: (a) is an automatic or manual transmission better for MPG? (b) Quantify the MPG difference between automatic and manual transmissions. We found that cars with manual transmission appear to have an increase of 7.24 miles per galon, when compared to cars with automatic transmission, without considering other variables. However, when we account for confounding effects, such as number of cylinders, hourse power and weight, there is no significant effect, and we can not conclude from the data that a manual transmission implies a higher or smaller average value for MPG. In the more complete regression model (with covariates), the estimated coefficient for the effect of a manual transmission was equal to 1.81, but it was not statistically significant (p-value > 20%).

Data Preparation and Initial Analysis

Initially, we uploaded the proper data, found some basic statistics, and did some exploratory data analysis, identifying some possible factor variables. For example, number of cylinders, number of carburetors and number of forward gears can be used as factor variables into the regression models. In the codes below, we added labels “automatic” and “manual” to the factor variable transmission.

```
data(mtcars);
mtcars$gear <- factor(mtcars$gear);
mtcars$cyl <- factor(mtcars$cyl);
mtcars$am <- factor(mtcars$am, labels = c("automatic", "manual"));
mtcars$carb <- factor(mtcars$carb);
summary(mtcars); str(mtcars);
```

Regression Models

The regression below has one single exploratory variable, our main variable of interest “transmission type”, into the regression for MPG. Note that the regression coefficient for transmission type is highly significant (pvalue < 0.001), and the manual transmission appears to have a positive effect on miles per galon. Cars with manual transmission appear to have an increase of 7.24 miles per galon, when compared to cars with automatic transmission (code and results in the appendix).

However, it is important to consider the impact of other variables as well, and to analyze the presence of possible confounding effects. The graphs in Figure 1 in the appendix show the effects of other variables on miles per galon performance. Note that weight, horse power, number of cylinders and number of carburetors also seem to be important in determining the values of miles per galon. The codes for these graphs are also shown in the Appendix.

We estimated a sequence of regression models starting from a more general especification, containing all variables (except MPG) in the data table, as explanatory variables. We then excluded the least significant variables (highest p-values) sequentially, until we obtained a more parsimonious especification, with only significant variables (in addition to the type of transmission variable). The codes for the sequential model estimation is provided in the appendix. We used ANOVA to test for variable exclusion.

The final model contained the variables number of cylinders, weight, hourse power and type of transmission as explanatory variables, as shown below:

```
##
## Call:
## lm(formula = mpg ~ factor(am) + factor(cyl) + wt + hp, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.939 -1.256 -0.401  1.125  5.051
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       33.7083     2.6049   12.94  7.7e-13 ***
## factor(am)manual   1.8092     1.3963    1.30   0.2065
## factor(cyl)6      -3.0313     1.4073   -2.15   0.0407 *
## factor(cyl)8      -2.1637     2.2843   -0.95   0.3523
## wt                -2.4968     0.8856   -2.82   0.0091 **
## hp                -0.0321     0.0137   -2.35   0.0269 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.866,  Adjusted R-squared:  0.84
## F-statistic: 33.6 on 5 and 26 DF,  p-value: 1.51e-10
```

As we can see from the results above, the estimated coefficient for manual transmission (against automatic transmission) resulted equal to 1.81. Therefore, all other important variables being kept equal (horsepower, weight and number of cylinders), there is an addition of 1.81 galons per miles in performance for manual cars, when compared to automatic ones. However, the estimated coefficiente is not statistically significant; the p-value is equal to 0.2065, and we do not reject the null hypothesis of null effect on MPG even with a test level of 10%.

Note that the initial regression, with only transmission type as an explanatory variable, had a highly significant and positive coefficient for the impact of manual transmision on MPG. On the other hand, the final regression, with important covariates included, did not show a positive significant effect anymore. Therefore, we have an indication of an important confounding effect of the other covariates in the model. In fact, as we can see from figure 2 in the appendix, automatic cars are in general heavier than manual ones. Besides, automatic cars also have a higher horsepower. Both weight and housepower impact negatively the MPG performance. Therefore, excluding both horsepower and weight from the regression may lead to the false conclusion that automatic transmission causes a lower value for MPG. The same argument applies to the number of cylinders as a confounding factor. Figure 3 in the appendix present some residual plots.

Appendix

Sequence of estimated regression models

```
m0 <- lm(mpg ~ am, data = mtcars);
summary(m0);
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## ammanual        7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

```
m0 <- lm(mpg ~ am, data = mtcars);
summary(m0);
m1 <- lm(mpg ~ am + factor(cyl) + wt + disp +
               hp + drat + factor(gear) + qsec +
               vs + factor(carb), data = mtcars)
m2 <- lm(mpg ~ factor(am) + factor(cyl) + wt + disp +
             hp + drat + factor(gear) + qsec +
             vs, data = mtcars)
m3 <- lm(mpg ~ factor(am) + factor(cyl) + wt + disp +
             hp + drat + qsec +
             vs, data = mtcars);
m4 <- lm(mpg ~ factor(am) + factor(cyl) + wt +
             hp, data = mtcars);
m5 <- lm(mpg ~ factor(am) + wt +
             hp, data = mtcars);
anova(m1, m2, m3, m4, m5)
summary(m5)
anova(m4, m5)
summary(m4)
```

Figure 1 - Initial plots

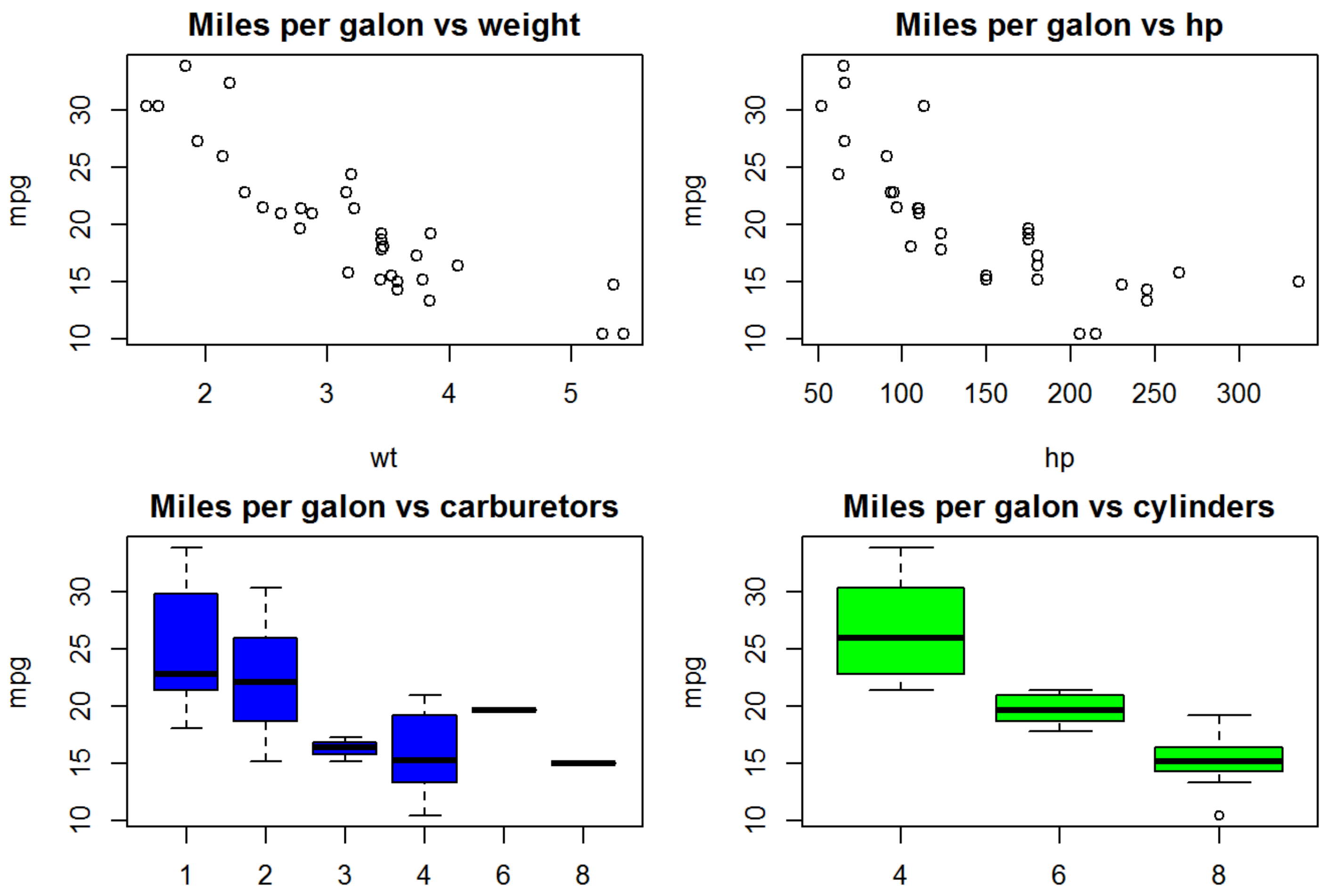

Figure 2 - Interpreting regression results

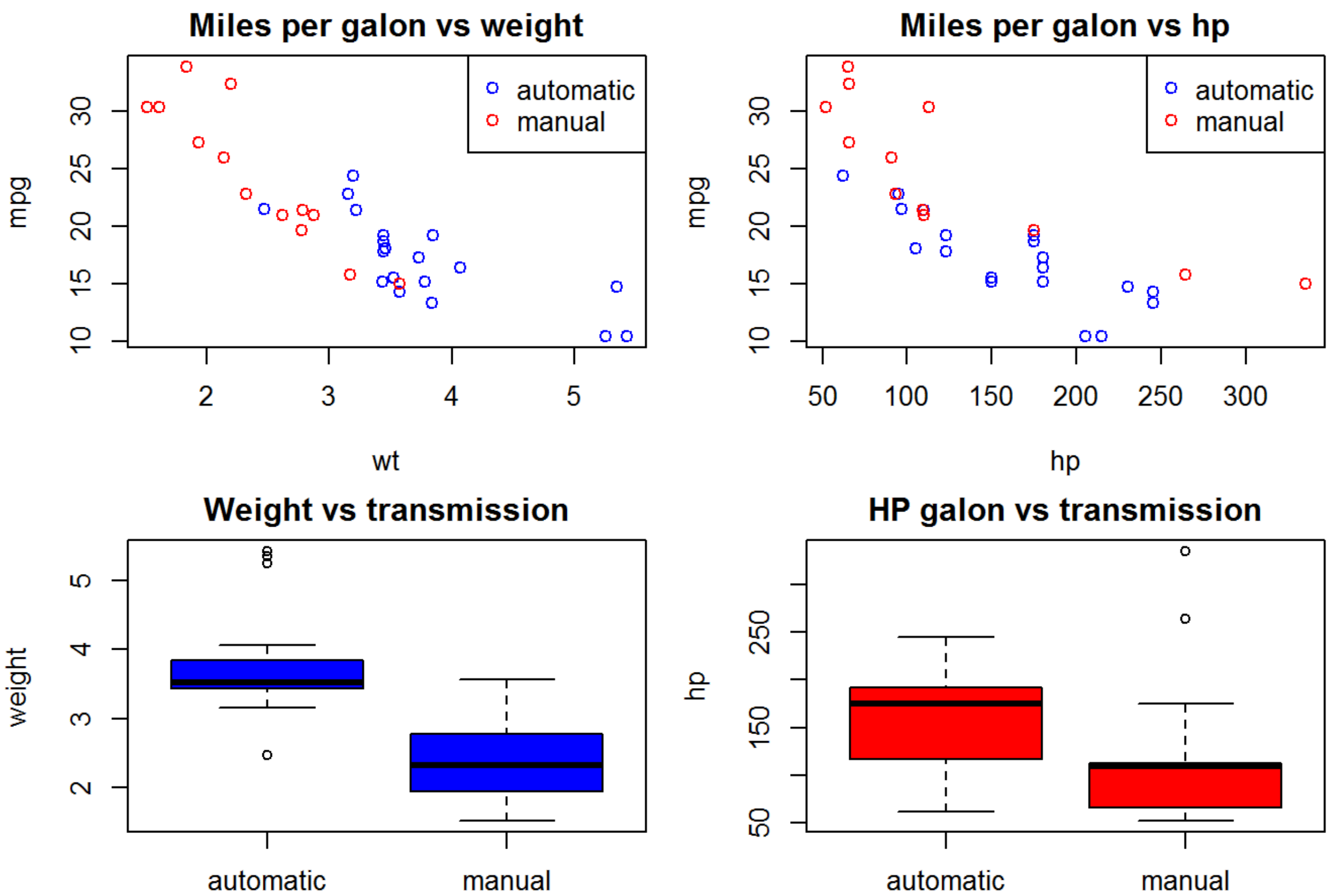

Figure 3 - Residual plots