Comparison of fuel efficiency between cars with automatic and manual transmission

Amade A.

December 25, 2015

Overview

Research objective

The objective of this paper is answering the question whether automatic transmission is better for MPG compared to manual transmission. In addition, the researcher also aims to quantify the difference in MPG between cars with different types of transmission.

Executive summary

This paper analyzes the influence of various car characteristics on fuel economy (measured as miles per galon of fuel). In the process of analysis 4 different linear models were fitted. They were: a simple linear model with one binary variable representing transmission type and 3 models with one additional numeric variable - *wt* (automobile's weight), *disp* (engine displacement) and *hp* (automobile's horsepower).
Based on the criterium of significance of *am* (transmission type) variable and best explanatory power, tested with likelihood ratio test, the model with 2 variables - *am* and *hp* was chosen. **The model indicates that the cars with manual transmission have better fuel efficiency (higher average MPG). The difference in fuel efficiency between the cars with manual and automatic transmission is *5.27* MPG**.

Exploratory Data Analysis

Dataset

Dataset used for the analysis is **mtcars** dataset, preloaded in R in the *datasets* package. Among the 10 variables, 5 (cyl, vs, am, gear, carb) can be characterized as discrete (factor) variables and remaining 5 (disp, hp, drat, wt, qsec), continuous (numeric) variables. Fuel consumption is represented by the continuous *mpg* variable.

MPG by transmission type

In total, there were 19 cars with automatic transmission and 13 with manual. **Graph 1** (see Appendix) shows the differences in measured fuel efficiency among cars with different transmission types.

It can be easily seen that the fuel efficiency is not the same among the cars with different transmission types, which means that *am* variable is a good candidate for independent regression variable. Median MPG for cars with automatic transmission equals **17.3** and for the ones with manual **22.8**. Values seem to be more concentrated around the median in the case of automatic transmission, compared to cars with manual transmission. Standard deviations of MPG value for two groups equaled respectively **3.83** and **6.17**.

Formal modelling

Simple model

As a first step in modeling a simple linear model with one binary variable *am* was fitted. Parameters of the regression are summarized below.

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 17.147368   1.124603 15.247492 1.133983e-15
ammanual     7.244939   1.764422  4.106127 2.850207e-04
```

Since the model has only one binary variable *am*, the intercept is interpreted as predicted value of MPG for the base level of *am* variable, which in **mtcars** dataset is "automatic transmission". Value of 17.15 informs us that cars with automatic transmission are expected to reach 17.15 miles for each galon of fuel. This value is also the average MPG value for cars with automatic transmission.

Change in MPG value for cars with manual transmission is represented by *ammanual* coefficient. Value of 7.24 tells us that the expected (average) MPG for cars with manual transmission is 7.24 higher compared to cars with automatic transmission.

P value for *ammanual* coefficient is very close to 0, informing us that the mean MPG range is significantly different for cars with different transmission types even at 99.9% confidence level. R^2 of the model equaled 0.36, meaning that 36% of the variation in MPG is explained by the regression model.

Inclusion of additional variable

Since the value of R square coefficient is relatively low (0.36) it is quite likely that there are other variables, which can help in explaining the variation in MPG. Additional variables to the model can be determined by looking at the absolute value of correlation of numeric variables with *mpg* variable.

```
     absoluteCorrelation
disp                0.85
hp                  0.78
drat                0.68
wt                  0.87
qsec                0.42
```

Two most higly correlated variables with *mpg* variable are *wt* and *disp* variables, referring respectively to automobile's weight (in 1000lbs) and engine displacement (in cu.in.).

Model including automobile's weight

```
               Estimate Std. Error     t value     Pr(>|t|)
(Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
ammanual    -0.02361522  1.5456453 -0.01527855 9.879146e-01
wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

As a result of including the weight variable to the model, *am* variable lost its explanatory power. Its value of -0.02 is very close to 0. Such low value of coefficient would indicate that the mean MPG for cars with manual transmission is lower by -0.02 compared to cars with automatic transmission. The p value of 0.99 doesn't allow us to reject the null hypothesis of no differences in mean MPG between cars with automatic and manual transmission at any conventional conficenve level.

It can be therefore concluded that the regression model with both *am* and *wt* variables is not appropriate for answering the question of impact of transmission type on automobile's fuel efficiency.

Model including engine displacement

```
              Estimate  Std. Error   t value     Pr(>|t|)
(Intercept) 27.84808111 1.834071377 15.183750 2.452658e-15
ammanual     1.83345825 1.436099585  1.276693 2.118396e-01
disp        -0.03685086 0.005781896 -6.373490 5.747528e-07
```

Also in the case of this model, the p value of 0.21 for coefficient *ammanual* indicates that it has no significant effect on exaplaining variation in MPG. Therefore, despite relatively large R^2 value of 0.73, this model is also not appropriate for explaining the impact of transmission type on automobile's MPG.

Model with the engine horsepower

Third most correlated with the *mpg* variable was *hp* variable representing the horsepower of the engine.

```
              Estimate  Std. Error   t value     Pr(>|t|)
(Intercept) 26.5849137 1.425094292 18.654845 1.073954e-17
ammanual     5.2770853 1.079540576  4.888270 3.460318e-05
hp          -0.0588878 0.007856745 -7.495191 2.920375e-08
```

P value of the ammanual coefficient is very close to 0, indicating that there are statistically significant differences between mean MPG among cars with manual and automatic transmissions. Also the value of this coefficient at 5.28 is of the satisfactory order of magnitude is comparing the differences in mean MPG (MPG is higher by 5.28 for cars with manual transmission). P value for for the *hp* variable coefficient is also close to zero indicating that the coefficient is significantly different than 0. Its value of -0.06 mean that for each additional horsepower, the MPG of the car decreases by 5.28. R^2 for this model equals 0.78.

Best model selection

The most important criterion for model selection is that it must allow us to see the impact of transmission type on MPG. Therefore, such model should include *am* variable. Among the models presented above, only two of them included that variable - a simple model with one explanatory variable *am* (fit) and a model with two variables - *am* and *hp* (fit4). Those two models will be subjected to likelihood ratio test in order to determine whether the inclusion of additional variable - *hp* significantly improves the explanatory power of the model.

Test performed with the *lrtest* function from the *lmtest* package returned Chi-square statistic of 34.48. In the Chi-square distribution with 1 degree of freedom its p value is marginally close to 0 indicating that the null hypothesis that the model with 2 explanatory variables (fit4) does not offer significantly better goodness of fit compared to the model with 1 explanatory variable (fit) can be rejected at any conventional significance level (e.g. 95%). Therefore, it can be concluded that the best model for the purpose of this investigation is model *fit4* with 2 explanatory variables - *am* and *hp*.

Analysis of residuals

As can be seen in **Graph 2** (Appendix), the absolute value of residual values tends to be higher for lower values of *hp* variable, especially for the cars with manual transmission. This implicates that the assumption of homoschedasticity in residuals may not be held true.

Conclusions and interpretation

Untimately the model with 2 explanatory variables - *am* and *hp* was chosen. The intercept in this model equaled 26.58. It can be interpreted as the MPG value for cars with automatic transmission and 0 horsepower (only theoretical concept). Coefficient *ammanual* of 5.28 indicates that **cars with manual transmission have on average MPG measure higher by 26.58**. Last coefficient *hp* with the value of -0.06 represents the decrease in MPG measure for each additional horsepower.

All of the coefficients had their p values marginally close to 0. In the case of *(Intercept)* and *hp* coefficients it mean that they are significantly different than 0. I the case of *ammanual* p value close to 0 indicates that the average MPG value is significantly different between cars with manual and automatic transmission.

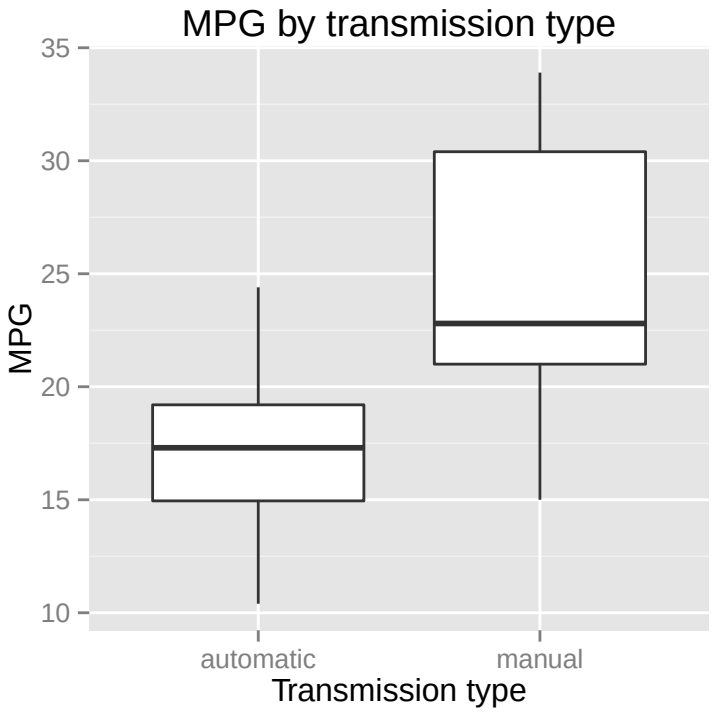
95% confidende intervals for coeffitient values are summarized below.

```
                2.5 %     97.5 %
(Intercept) 23.67026866 29.49955884
ammanual     3.06917692  7.48499369
hp          -0.07495665 -0.04281896
```

These conficende intervals can be interpreted as a range of values in which there is a 95% confidence that the true value of a coefficient lies between them. E.g. we can state that there is 95% confidence that the true value of *ammanual* coefficient lies between 3.07 and 7.48.

Appendix

Graph 1 - MPG by transmission type



Graph 2 - Horsepower and transmission type vs. residuals