

Effect of Transmissions Mode to Fuel MPG

Andreas Hadimulyono

Executive Summary

This analysis aims to answer two questions in particular:

1. Is an automatic or manual transmission better for MPG?
2. What is the MPG difference (if any) between manual and automatic transmission?

The analysis starts with some exploratory data analysis, followed by model selection and interpretations. Finally it attempts to try to answer the posed questions above.

Exploratory Data Analysis

The dataset contains 32 observations of 11 variables. There seems to be a distinct difference in mpg distributions between automatic and manual transmissions as shown in Figure 1 (Appendix).

The table below shows the correlations of each variable in mtcars dataset.

| ## | mpg | cyl | disp | hp | drat | wt |
|---------|-------------|------------|-------------|------------|-------------|------------|
| ## mpg | 1.0000000 | -0.8521620 | -0.8475514 | -0.7761684 | 0.68117191 | -0.8676594 |
| ## cyl | -0.8521620 | 1.0000000 | 0.9020329 | 0.8324475 | -0.69993811 | 0.7824958 |
| ## disp | -0.8475514 | 0.9020329 | 1.0000000 | 0.7909486 | -0.71021393 | 0.8879799 |
| ## hp | -0.7761684 | 0.8324475 | 0.7909486 | 1.0000000 | -0.44875912 | 0.6587479 |
| ## drat | 0.6811719 | -0.6999381 | -0.7102139 | -0.4487591 | 1.00000000 | -0.7124406 |
| ## wt | -0.8676594 | 0.7824958 | 0.8879799 | 0.6587479 | -0.71244065 | 1.0000000 |
| ## qsec | 0.4186840 | -0.5912421 | -0.4336979 | -0.7082234 | 0.09120476 | -0.1747159 |
| ## vs | 0.6640389 | -0.8108118 | -0.7104159 | -0.7230967 | 0.44027846 | -0.5549157 |
| ## am | 0.5998324 | -0.5226070 | -0.5912270 | -0.2432043 | 0.71271113 | -0.6924953 |
| ## gear | 0.4802848 | -0.4926866 | -0.5555692 | -0.1257043 | 0.69961013 | -0.5832870 |
| ## carb | -0.5509251 | 0.5269883 | 0.3949769 | 0.7498125 | -0.09078980 | 0.4276059 |
| ## | qsec | vs | am | gear | carb | |
| ## mpg | 0.41868403 | 0.6640389 | 0.59983243 | 0.4802848 | -0.55092507 | |
| ## cyl | -0.59124207 | -0.8108118 | -0.52260705 | -0.4926866 | 0.52698829 | |
| ## disp | -0.43369788 | -0.7104159 | -0.59122704 | -0.5555692 | 0.39497686 | |
| ## hp | -0.70822339 | -0.7230967 | -0.24320426 | -0.1257043 | 0.74981247 | |
| ## drat | 0.09120476 | 0.4402785 | 0.71271113 | 0.6996101 | -0.09078980 | |
| ## wt | -0.17471588 | -0.5549157 | -0.69249526 | -0.5832870 | 0.42760594 | |
| ## qsec | 1.00000000 | 0.7445354 | -0.22986086 | -0.2126822 | -0.65624923 | |
| ## vs | 0.74453544 | 1.0000000 | 0.16834512 | 0.2060233 | -0.56960714 | |
| ## am | -0.22986086 | 0.1683451 | 1.00000000 | 0.7940588 | 0.05753435 | |
| ## gear | -0.21268223 | 0.2060233 | 0.79405876 | 1.0000000 | 0.27407284 | |
| ## carb | -0.65624923 | -0.5696071 | 0.05753435 | 0.2740728 | 1.00000000 | |

Model Selection and Interpretations

We will consider our model by starting with a variable (our variable of interest, “am”) as [fit1], and we will add variables having |correlation with mpg| greater than 0.8 using nested model technique. After variables are added, we run ANOVA function to test if the variable should be added.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + cyl
## Model 4: mpg ~ am + wt + cyl + disp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1      30 720.90
## 2      29 278.32  1    442.58 63.4179 1.469e-08 ***
## 3      28 191.05  1     87.27 12.5055 0.001488 **
## 4      27 188.43  1      2.62  0.3756 0.545093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- [fit2] First variable to be added is weight (wt), with correlation value to MPG -0.8676594. Which should make sense. As a car gets heavier, it should take more fuel to travel the same distance, thus it should mean worse MPG.
- [fit3] Next to be added is number of cylinders (corr value to MPG of -0.8521620), which impacts how fast power can be generated in a car, so it should have an effect to MPG.
- [fit4] Next is displacement (corr value to MPG of -0.8475514). Displacement is also related to the car's power output.

From the Anova table, we should not add displacement as the p-value is above 5%. So we will use Model 3 in the anova table, which is based on am, wt, and cyl. The coefficients of the selected model is as follows:

```
##           Estimate Std. Error   t value    Pr(>|t|)
## (Intercept) 39.4179334  2.6414573 14.9227979 7.424998e-15
## am          0.1764932  1.3044515  0.1353007 8.933421e-01
## wt         -3.1251422  0.9108827 -3.4308942 1.885894e-03
## cyl        -1.5102457  0.4222792 -3.5764148 1.291605e-03
```

For am, 0 represents automatic and 1 represents manual. The intercept is the predicted MPG value when all predictors are not present. The am coefficient means the expected change in MPG for each change in am, while holding other variables constants. The coefficients for wt represents the change in MPG for each unit of change in wt, while holding other variables constants. The same for cyl, the coefficient means the expected change to MPG for each unit of change in cylinder, holding other variables constant.

Figure 2 (Appendix) shows the residual plot of the data for the 32 observations in the data set. There does not seem to be any clear pattern to suggest abnormality (e.g. heteroskedasticity).

Figure 3 (Appendix) shows QQ Plot of the residuals. It does not seem to have obvious abnormality.

Conclusion

The coefficients for am is positive, which represent positive change in MPG by changing mode of transmission from automatic to manual by 0.1764932. That means manual transmission seems to result in more MPG (which means more efficient cars).

However while the coefficient is non-zero, the p-value of the coefficient is 8.933421e-01, which is outside typical 5% threshold we set for p-values. As we are unable to reject the null hypothesis, that means this observation is not statistically significant. Thus this analysis is unable to answer the questions posed in the executive summary based on the data provided and model selected.

Appendix

Figure 1. Comparison of Mode of Transmission Effect on MPG
(Red Point Represents the Mean)



Figure 2. Residual Plot

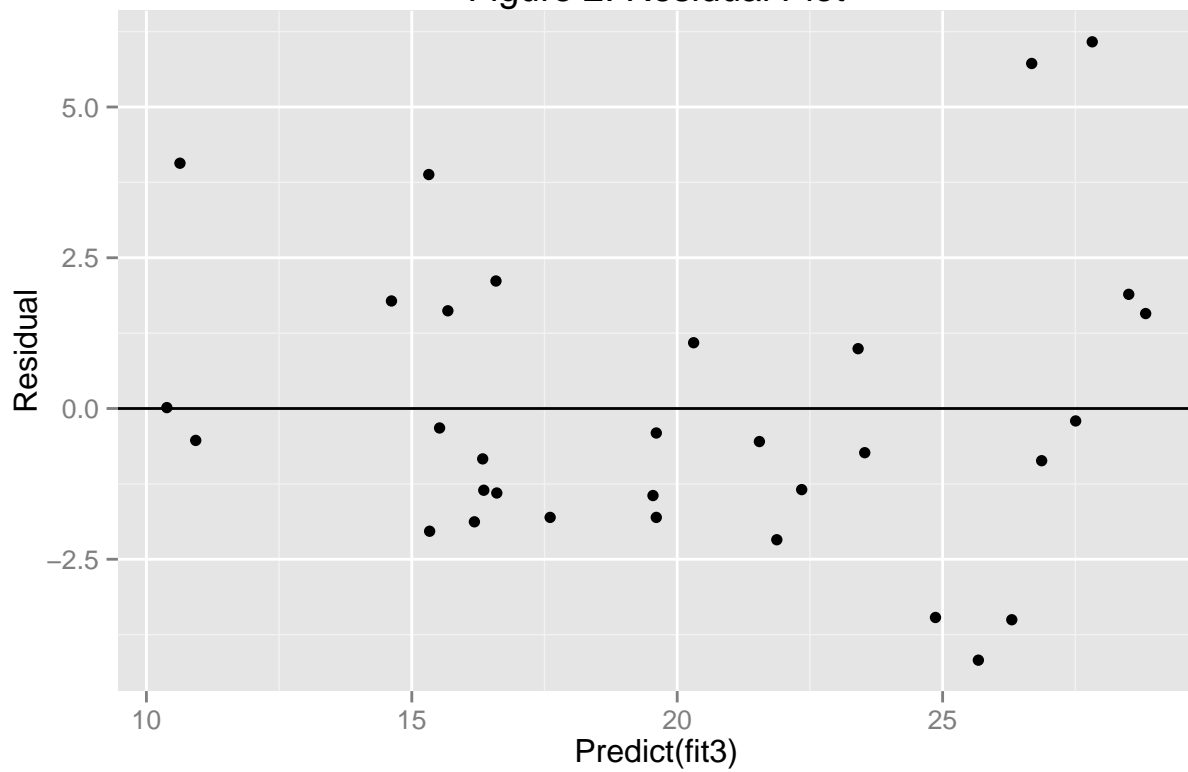


Figure 3. QQ Plot

