

# Effect of Mode of Transmissions to Fuel Efficiency (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. Variables like vs, am, and cyl are distinct values, so we can convert them to factors so ggplot can handle them better. There seems to be a distinct difference in mpg distributions between automatic and manual transmissions as shown in Figure 1 (Appendix).

## Model Selection and Interpretations

We will consider our model by starting with a variable (our variable of interest, “am”), and we will add variable 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 + hp
## Model 4: mpg ~ am + wt + hp + cyl
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      30 720.90
## 2      29 278.32  1    442.58 76.1924  3.32e-09 ***
## 3      28 180.29  1     98.03 16.8762 0.0003525 ***
## 4      26 151.03  2     29.27  2.5191 0.0999982 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. First variable to be added is weight (wt). 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. Anova gives us p-value of 3.32e-09, so we should take this variable.
2. Next is horsepower (hp). With higher potential power output, it should potentially require more fuel, which should result in worse MPG. Running anova on (wt and hp) gives a p-value of 0.0003525, so we're accepting this variable.

3. Next to be added is number of cylinders, which impacts how fast power can be generated in a car, so it should have an effect to MPG. Running anova on (wt, hp, and cyl) though gives us a p-value of 0.0999982, which is over 0.05 and that means it seems we do not need to add number of cylinders to the previous models. A possible explanation is number of cylinders is related to power, and the variation caused by this variable could have been explained by previous variable (horsepower)

As the author is not well-versed in the making of a car and the limited literature research performed has indicated that other variables are likely to also be correlated to the power of the engine, the author decides to not add any more variable. So we will use the model specified in point (2) above, which uses “am”, “wt”, and “hp” as predictors.

The coefficients of the selected model is as follows:

##	Estimate	Std. Error	t value	Pr(> t )
## (Intercept)	34.00287512	2.642659337	12.866916	2.824030e-13
## amManual	2.08371013	1.376420152	1.513862	1.412682e-01
## wt	-2.87857541	0.904970538	-3.180850	3.574031e-03
## hp	-0.03747873	0.009605422	-3.901830	5.464023e-04

The base of reference for this model is am = “Automatic”, which is a factor, so the intercept is the predicted MPG with automatic transmission, and other variables are not present (zero). The coefficients for non-factor variables (wt and hp) represents the change in MPG for each unit of change in (wt and hp) respectively, while holding other variables constants. The coefficients for amManual represents the change in MPG if the mode of transmission change to Manual from Automatic, while holding other variables constant.

Figure 2 shows the residuals of the data for the 32 observations in the data set. There does not seem to be any clear pattern.

Figure 3 shows QQ Plot of the residuals, and it does not seem to show any abnormal tendency.

## Conclusion

The coefficients for amManual, which represent the change in MPG by changing mode of transmission from automatic to manual, is positive 2.083710. That means manual transmission seems to result in more MPG (which means more efficient cars).

However while the model used looks like a good fit, the p-value of the coefficient of is 0.141268, which is outside typical 5% threshold we set for p-values. That means we are unable to reject the null hypothesis. So while the coefficient is non-zero, it is unlikely that the mode of transmission has an impact on the MPG. 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)

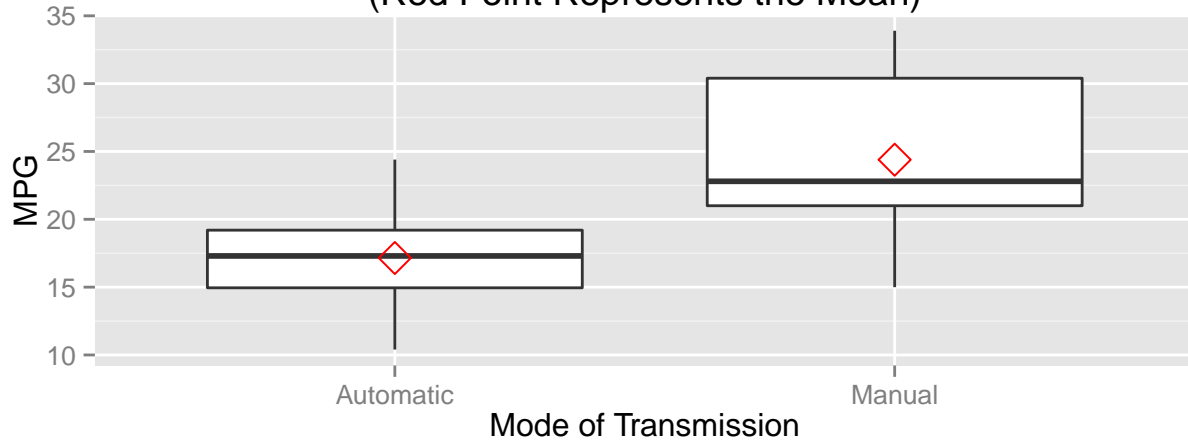
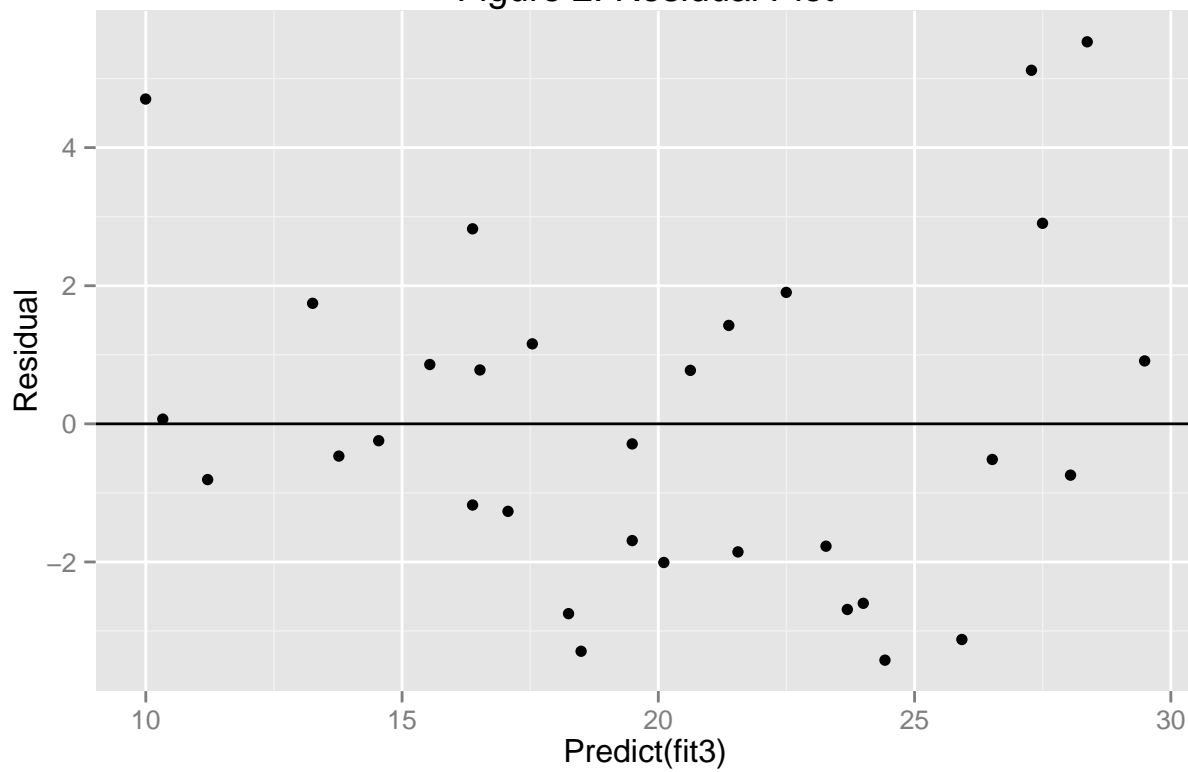


Figure 2. Residual Plot



**Figure 3. QQ Plot**

