# Effect of Transmission Types on Fuel Efficiency

*Junyoung Kim*

*Mar. 11, 2018*


## Summary

This study examines the effect of transmission type on fuel efficiency of cars using 1974 Motor Trend US magazine data. Based on its model choice creteria, this study takes horsepower as a co-determinant of miles per gallon(mpg) while maintaining auto/manual transmission type as the main independent variable. After taking natural logs on the values of mpg and horsepower to reshape their correlation more linear, the model shows that manual transmission has some positive effect in fuel efficiency while horsepower keeps a strong negative correlation with mpg.


## 1. Exploratory Overview

It is usually believed that manual transmission of cars tend to be more fuel-efficient than automatic transmission. This article aims to examine whether this belief is true, and if so, how better it is. The data being used is 1974 *Motor Trend* US magazine data, which is composed of 32 cars with 11 variables. The variables vary in types, and highly correlated to each other.

Table 1: List of Variables

| | MPG | Cylinder | Displacement | Horsepower | Rear Axle Ratio | Weight | 1/4 Mile Time | Engine V/S | Auto/Manual | Gear | Carburetor |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb |
| Type | continuous | discrete | continuous | continuous | continuous | continuous | continuous | categorical | categorical | discrete | discrete |
| Level | - | 4, 6, 8 | - | - | - | - | - | 0, 1 | 0, 1 | 3, 4, 5 | 1 - 8 |

Table 2: Correlation Matrix

| | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb |
|---|---|---|---|---|---|---|---|---|---|---|
| cyl | 1.000 | 0.902 | 0.832 | -0.700 | 0.782 | -0.591 | -0.811 | -0.523 | -0.493 | 0.527 |
| disp | 0.902 | 1.000 | 0.791 | -0.710 | 0.888 | -0.434 | -0.710 | -0.591 | -0.556 | 0.395 |
| hp | 0.832 | 0.791 | 1.000 | -0.449 | 0.659 | -0.708 | -0.723 | -0.243 | -0.126 | 0.750 |
| drat | -0.700 | -0.710 | -0.449 | 1.000 | -0.712 | 0.091 | 0.440 | 0.713 | 0.700 | -0.091 |
| wt | 0.782 | 0.888 | 0.659 | -0.712 | 1.000 | -0.175 | -0.555 | -0.692 | -0.583 | 0.428 |
| qsec | -0.591 | -0.434 | -0.708 | 0.091 | -0.175 | 1.000 | 0.745 | -0.230 | -0.213 | -0.656 |
| vs | -0.811 | -0.710 | -0.723 | 0.440 | -0.555 | 0.745 | 1.000 | 0.168 | 0.206 | -0.570 |
| am | -0.523 | -0.591 | -0.243 | 0.713 | -0.692 | -0.230 | 0.168 | 1.000 | 0.794 | 0.058 |
| gear | -0.493 | -0.556 | -0.126 | 0.700 | -0.583 | -0.213 | 0.206 | 0.794 | 1.000 | 0.274 |
| carb | 0.527 | 0.395 | 0.750 | -0.091 | 0.428 | -0.656 | -0.570 | 0.058 | 0.274 | 1.000 |

## 2. Building Research Model

Regression of mpg on transmission type(am) shows a significant positive estimated effect(Coefficient= 7.245, $p$=0.0003, Table 3) but really does not explain much of mpg($R^2$=0.3385). It is not surprising because transmission type is a simple two-level factor, and there are several factors determine fuel efficiency in reality.

In order to capture the effect of transmission type on mpg better, we need to bring other predictors together. But, the **strong correlations among variables** matter at this point (see Table 2). Taking them all causes **overfitting**. In table 3, none of the all-variable model shows statistical significance despite high explanatory power($R^2$=0.807). A good model should, **A.** explain variance of dependent variable(mpg) well ehough, **B.** keep statiatical significance of explanatory variable(am), **C.** be theoretically resonable. Considering these creteria, this research builds a research model with two distinct factors: **1. taking transmission type and log(horsepower) as regressors**, and **2. setting log(mpg) as outcome instead of mpg**.

Table 3: Test Regression Comparison

| | All | | Transmission Only | |
|---|---|---|---|---|
| | Coeffient | p-value | Coeffient | p-value |
| (Intercept) | 12.303 | 0.518 | 17.147 | 0 |
| cyl | -0.111 | 0.916 | | |
| disp | 0.013 | 0.463 | | |
| hp | -0.021 | 0.335 | | |
| drat | 0.787 | 0.635 | | |
| wt | -3.715 | 0.063 | | |
| qsec | 0.821 | 0.274 | | |
| vs | 0.318 | 0.881 | | |
| am | 2.52 | 0.234 | 7.245 | 0.00029 |
| gear | 0.655 | 0.665 | | |
| carb | -0.199 | 0.812 | | |
| R-squared | 0.807 | | 0.338 | |

**Model:** $\ln MPG = \beta_0 + \beta_1 Transmission + \beta_2 \ln Horsepower + \varepsilon$

### Horsepower

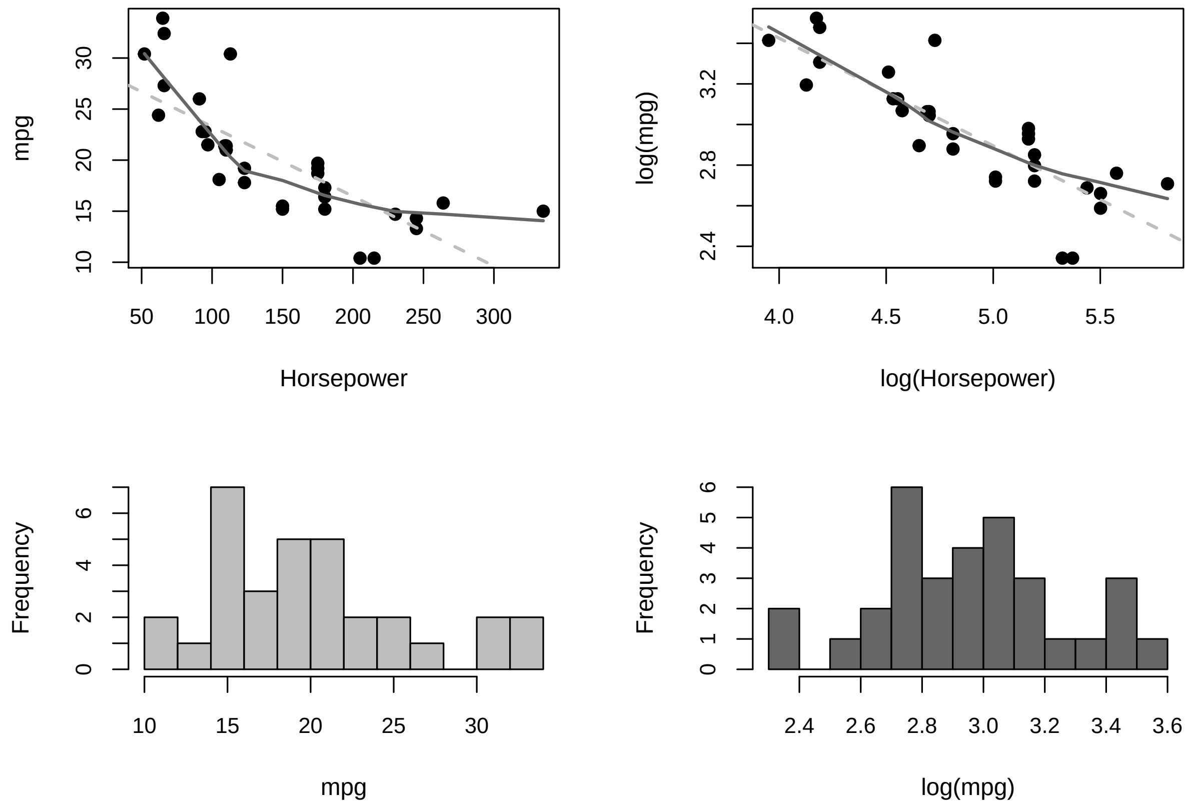
Theoretically, we can categorize determinants of fuel efficiency into engine, drivetrain, and weight. As we can see in Table 4, engine-related variables have high correlations, so, it can be a better strategy to choose one component which can represent the rest. In this model, horsepower(hp) is chosen because it has the highest correlations (total 4.096) with other engine components, and also shows relatively lower risk of multicollinearity with transmission type(VIF= 1.063, $R^2$ = 0.059).

Table 4: Correlations among Engine-related Variables

| | Cylinder | Displacement | Horsepower | V-type/Serial | Carburetor | VIF with AM | R-sqaured |
|---|---|---|---|---|---|---|---|
| Cylinder | 1.000 | 0.902 | 0.832 | -0.811 | 0.527 | 1.376 | 0.273 |
| Displacement | 0.902 | 1.000 | 0.791 | -0.710 | 0.395 | 1.537 | 0.350 |
| Horsepower | 0.832 | 0.791 | 1.000 | -0.723 | 0.750 | 1.063 | 0.059 |
| V-type/Serial | -0.811 | -0.710 | -0.723 | 1.000 | -0.570 | 1.029 | 0.028 |
| Carburetor | 0.527 | 0.395 | 0.750 | -0.570 | 1.000 | 1.003 | 0.003 |
| Total(absolute) | 4.072 | 3.798 | 4.096 | 3.814 | 3.242 | NA | NA |

### Why log-log?

The plots below compare the distribution of mpg and horsepower-mpg correlation, ones with logs and the others without. It is quite obvious that the original plots are more skewed. Taking logs on both terms linearizes the plot, normalizes the distribution, reduce effects of possible outliers, and makes interpretation of coefficients easier(percent by percent changes). Luckily, we can take logs without losing observations because all values are positive.




**Model Comparison**

In addition to engine-related variables, weight and rear axle ratio are also regarded as determinants of mpg in general. We can think of two alternative models with those two variables:

**Alternative 1:** $\ln MPG = \beta_0 + \beta_1 Transmission + \beta_2 \ln Horsepower + \beta_3 \ln Weight + \varepsilon$
**Alternative 2:** $\ln MPG = \beta_0 + \beta_1 Transmission + \beta_2 \ln Horsepower + \beta_3 \ln Weight + \ln RearAxelRatio + \varepsilon$

As mentioned above, one of the most crucial quality of a good model is statistical significance of transmission type. The alternatives show somewhat enhanced explanatory powers(adjusted $R^2$ in **Table 5**), however, they seriously weaken the statistical significance of the main independent variable($p$-value of auto/manual). Multicollinearity of log(weight) seems to be the biggest cause of this problem(**Table 6**). The VIF score of log(weight) is high($\sqrt{VIF} > 2$), and it also increases multicollinearity of other terms.

Table 5: Explanatory Power Comparison

| | adjusted R2 | Degree of Freedom | F-statistic | F-stat p-value | p-value of Auto/Manual |
|---|---|---|---|---|---|
| Model | 0.797 | 29 | 61.877 | 0 | 0.001 |
| Alternative 1 | 0.870 | 28 | 70.428 | 0 | 0.861 |
| Alternative 2 | 0.866 | 27 | 51.074 | 0 | 0.967 |

In terms of Heteroskedasticity, all three models fail to reject the null hypothesis of homogeneity of variance(Breusch-Pagan p-value in **Table 5**). But we can still find a noticeable difference that the alternatives barely lies within the assumption of homoscedasticity($p = 0.078$, $0.176$) while the model holds with much higher probability($p = 0.536$).

As we have seen so far, regressing log(mpg) on transmission type and log(horsepower) seems to be a simple but better model.

Table 6: Multicollinearity(VIF test) and Heteroskedasticity(BP test)

| | Auto/Manual | log(Horsepower) | log(Weight) | log(RearAxleRatio) | Breusch-Pagan p-value |
|---|---|---|---|---|---|
| Model | 1.136 | 1.136 | | | 0.536 |
| Alternative 1 | 2.452 | 2.527 | 4.642 | | 0.078 |
| Alternative 2 | 3.067 | 2.661 | 4.754 | 2.619 | 0.176 |

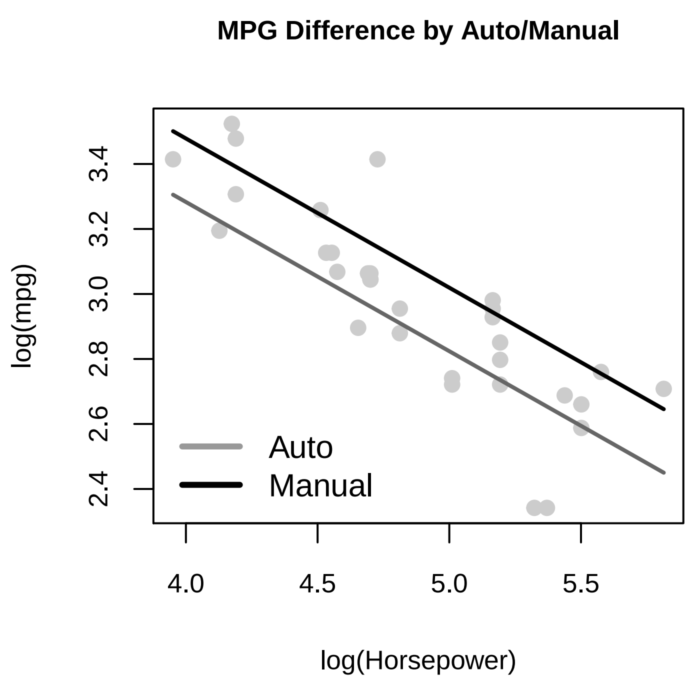
## 3. Analysis

### Regression Results and Interpretation

Table 7: Regression Results on log(MPG)

| | Coefficient | Standard Error | t-value | p-value |
|---|---|---|---|---|
| Intercept | 5.120 | 0.273 | 18.771 | 0.000 |
| Auto/Manual: 1 | 0.195 | 0.051 | 3.798 | 0.001 |
| log(Horsepower) | -0.459 | 0.054 | -8.498 | 0.000 |

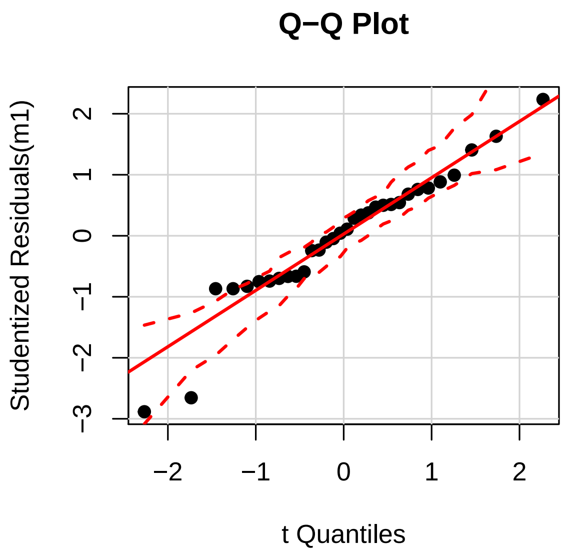
Table 6 shows the results. While log(horsepower) shows a strong negative relation with log(mpg)(-0.459), auto/manual holds a modreate positive coefficient(0.195). Low p-values of both regressors(0.001, 0.000) imply their solid statistical significance. Since both horsepower and mpg are log-transformed, the meaning of coefficient should be interpreted as changes in percent values: **for 1 percent increase of horsepower, mpg decreases approximately by 0.459 percent on average**. The transmission type(0 or 1) is not log-transformed, therefore, the effect should be interpreted as **manual transmissions(value = 1) have approximately 19.5%**($100 * \beta_1$%) **better fuel efficiency on average than automatic transmissions.**

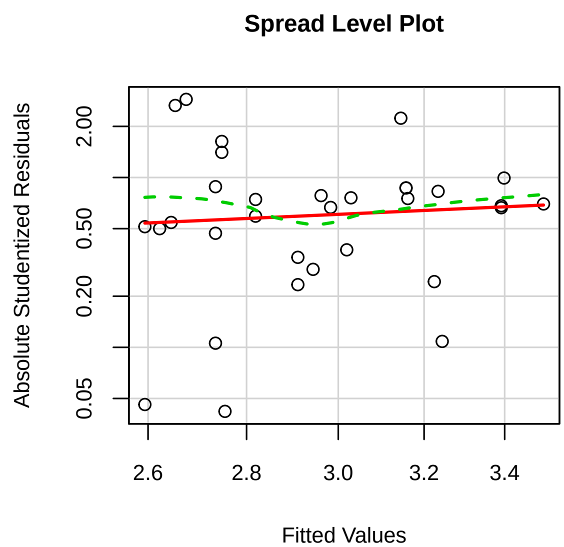


**Residual Diagnostics**

The plots below examine the residuals of the research model in order to detect violation of Gauss-Markov Theorem. As we can see in the Q-Q Plot, the residuals are distributed almost normally. The residual fit(green dot line) of in the Spread Level Plot is flat-like, and it implies the homogeniety of the residual variance. We can also look back **Table 6** to see the result of heteroskedasticity test($p = 0.536$), which is high enough to hold the $H_0$: Homoscedasticity.

**Model:** *Transmission + ln(Horsepower) + ln(Weight)*




## 4. Conclusion

**To sum up, this research shows that manual transmission is more fuel efficient than automatic transmission, by 19.5% on average.** The uncertainty of the model is very small as the p-values of estimates assure(0.000, 0.001. see **Table 7**), and the confidence interval of each estimate does not include zero, which means that the probability of each regressor having no effect on mpg is very small. (**Table 8**) However, we should avoid generalizing the outcomes over the car models today because the data is very small and limited within some models of specific years(1973-74). From selection bias to epistemic uncertainty, there are undeniable possiblity of drawing biased results, and it can be cleared only with using more elaborate models based on extensive data.

Table 8: Confidence Interval of Estimates

| | 2.5% | 97.5% |
|---|---|---|
| Auto/Manual | 0.09 | 0.301 |
| log(Horsepower) | -0.57 | -0.349 |