Evaluating the relationship between Miles per Gallon (MPG) and Transmission Type

Darion Lawson

Executive Summary

In the report we look at two goals:

- Determine whether an automatic or manual transmission better for MPG.
- Quantify the MPG difference between automatic and manual transmissions.

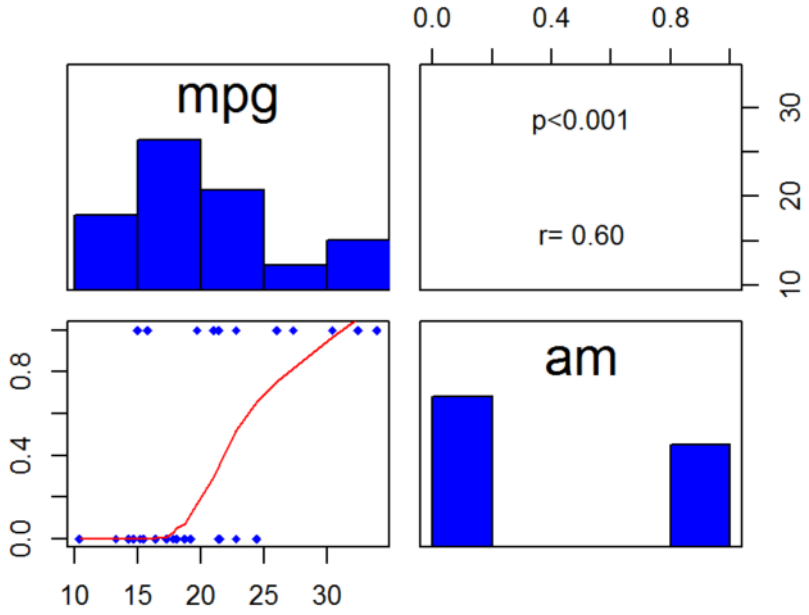
We look at correlation coefficients and the regression model to answer these questions. Upon assessment of the correlation between transmission type and gas mileage, we find that there is a correlation these two attributes. On average cars with automatic transmissions get better gas mileage.

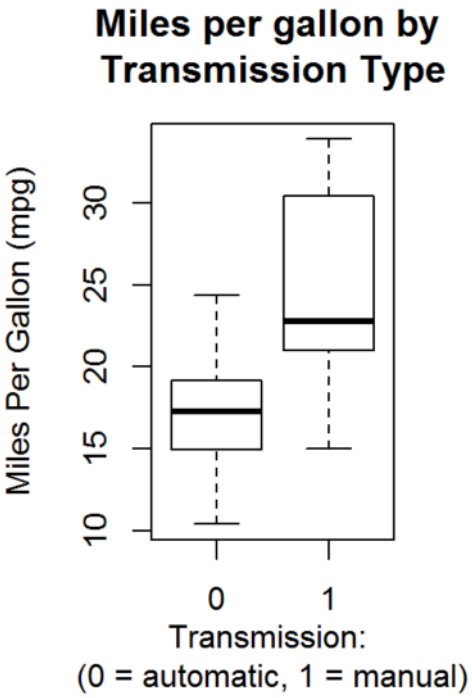
We also find that the quantified difference in gas mileage for automatic transmission (0) vs manual transmission (1) vehicles is a function of both the transmission type and the weight of the car.

Is an automatic or manual transmission better for MPG?

We're interested in evaluating the transmission type (am) against gas mileage (mpg). Transmission type appears to have significant correlation at almost 0.6, with a p-value of less then 0.001. A pairs plot of the two attributes shows the trend, correlation, and histogram.

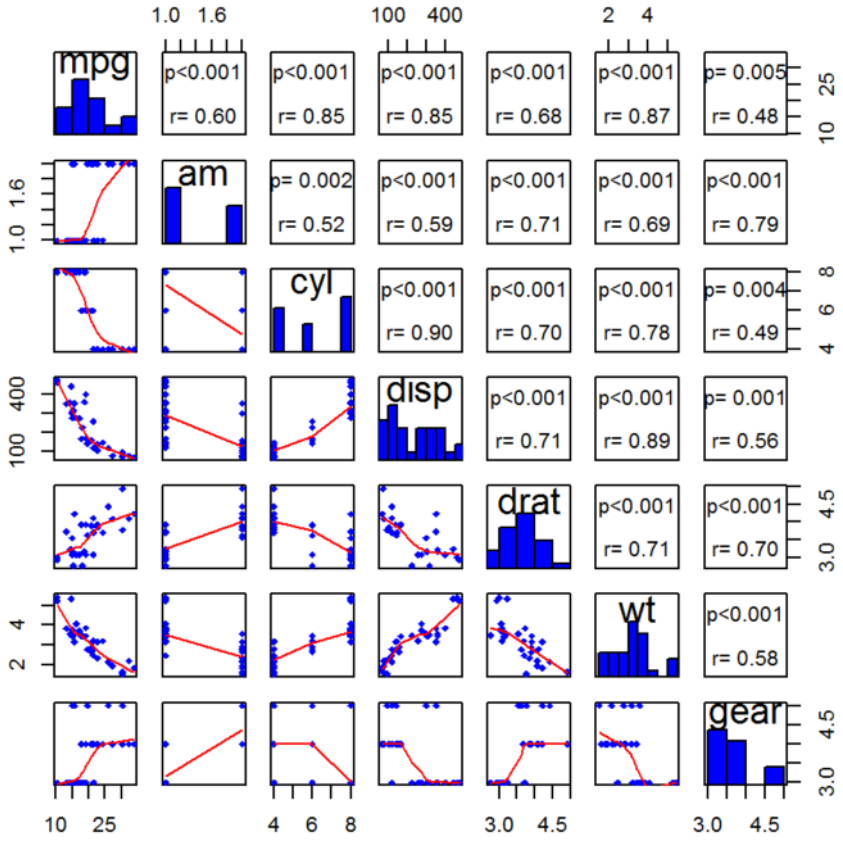
In addition, a boxplot of gas mileage against automatic transmission versus manual transmission also demonstrates the difference. The median gas mileage for manual transmission is 22.8, whereas the median gas mileage for automatic transmission is 17.3. This data suggests that a manual transmission is better for gas mileage.





Quantify the MPG difference between automatic and manual transmissions.

We've seen in the previous section that there is a correlation between transmission type (am) and gas mileage (mpg). There are also correlations between mpg when compared to cylinders (cyl), displacement (disp), rear axle ratio (drat), weight (wt), and number of forward gears (gear). We'll investigate these relationships further to identify the appropriate model to predict gas mileage from transmission type. The following pairs plot visually illustrates these relationships.



Using backward stepwise regression, we find that the parsimonious multivariate model contains the variables: weight (wt), quarter mile time (qsec), and transmission type (am). The model was adjusted to add an interaction term between am and wt because analysis indicates there was a correlation between these variables.

The difference in gas mileage (mpg) for automatic transmission (0) vs manual transmission (1) is a function of both the transmission type (am) and the weight of the car (wt). If you look only at transmission type, manual transmission vehicles get, on average, 14.079428 miles per gallon better gas mileage than automatic transmission. However, that does not tell the complete story. The prediction model includes an interaction term for transmission type (am) and weight (wt). Taking this term into account, the average difference in gas mileage is approximately 14.079428 minus 4.141376 multiplied by the weight value of the car. This means that for a 1000 lb car, holding all other variables constant, the increase in gas mileage gained by manual transmission would be on average 9.938052.

```
## lm(formula = mpg ~ wt + qsec + am + am * wt, data = mtcars)
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.723053  5.8990407  1.648243 0.1108925394
## wt          -2.936531  0.6660253 -4.409038 0.0001488947
## qsec         1.016974  0.2520152  4.035366 0.0004030165
## am1         14.079428  3.4352512  4.098515 0.0003408693
## wt:am1      -4.141376  1.1968119 -3.460340 0.0018085763
```

The adjusted r-squared is 0.8804219. The model explains approximately 88% of the variability in the response, Gas Mileage (mpg).

```
## [1] 0.8804219
```

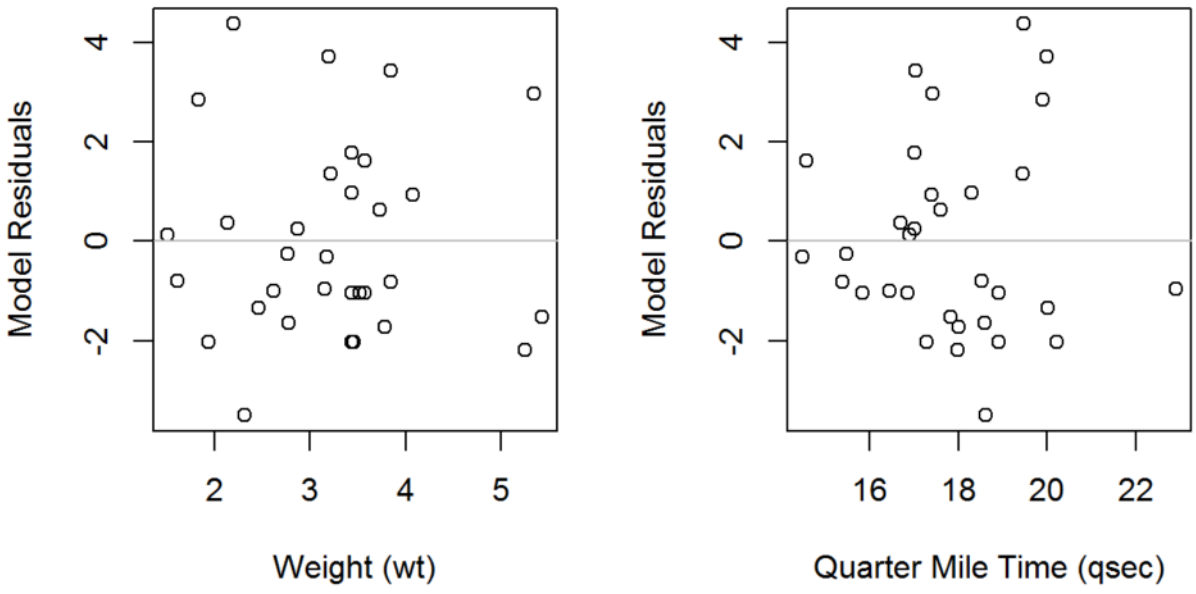
Appendix

Residuals Analysis

This section evaluates the model fit by looking at model residuals plots.

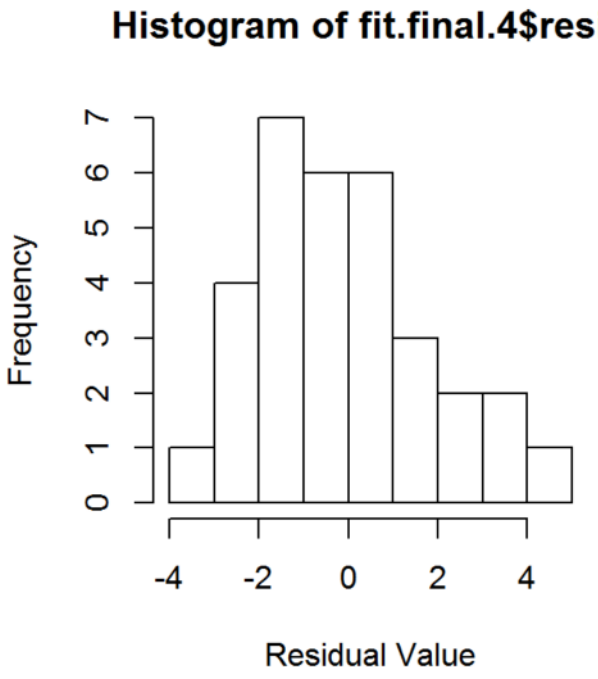
Does the Model Capture the Relationship?

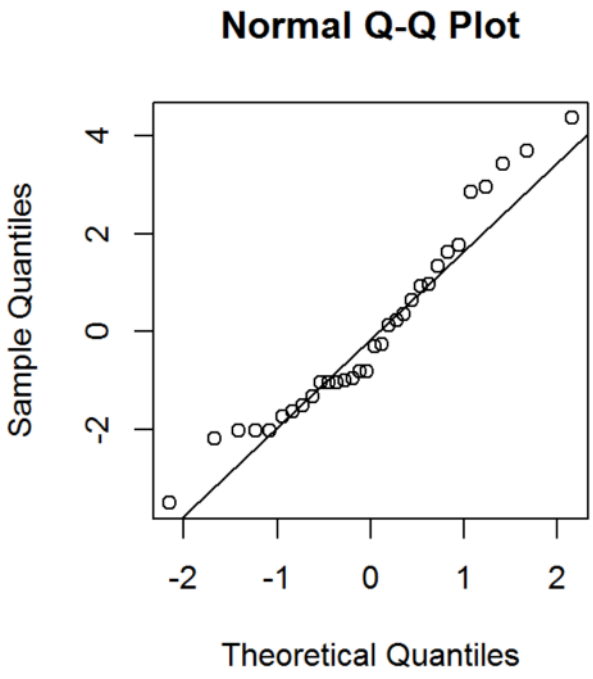
The residuals plotted against the numeric variables appear randomly scattered, so the model seems to capture the relationship.



Are the Residuals Normally Distributed?

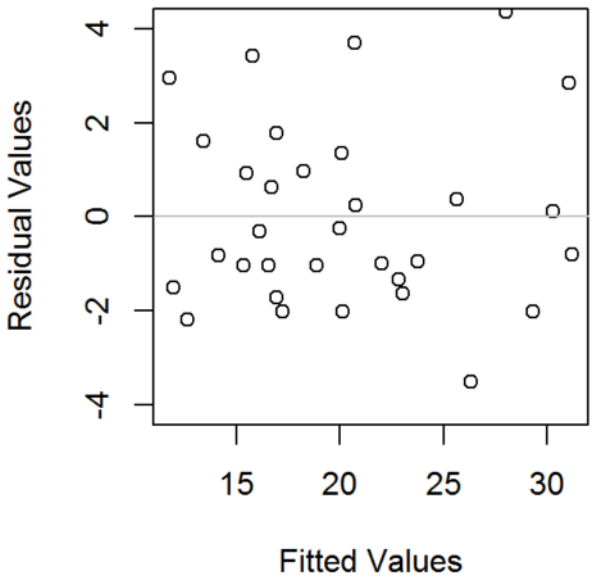
The residuals histogram appears to be near normally distributed. There is slight skew in the residuals distribution.

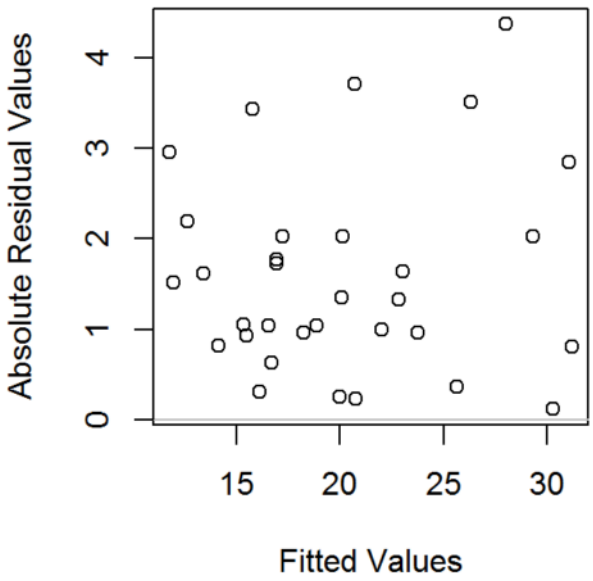




Is the Variability Constant?

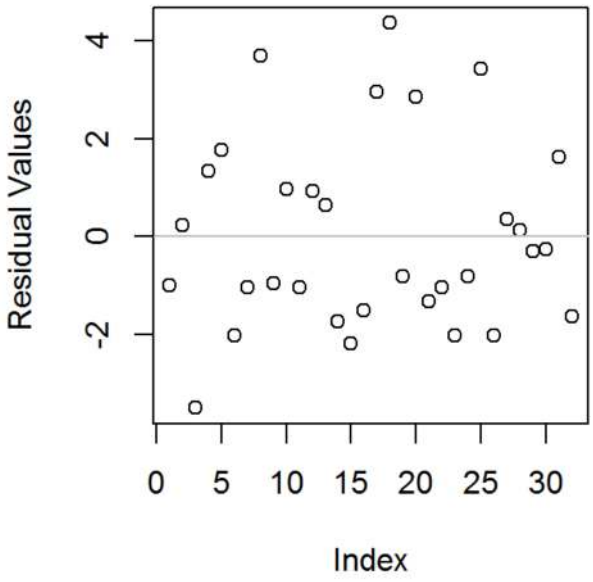
The variability of the residuals appears to be constant.





Are the Residuals Independent?

The residuals appear to be Independent, and there does not appear to be a time-based trend.



Description of Variables

The data is from the 1974 Motor Trend US magazine and contains data about fuel consumption with 10 additional attributes for 32 automobiles (all are 1973-74 models). The data frame has 32 observations with 11 variables.

Column	Name	Description
[, 1]	mpg	Miles/(US) gallon
[, 2]	cyl	Number of cylinders
[, 3]	disp	Displacement (cu.in.)
[, 4]	hp	Gross horsepower
[, 5]	drat	Rear axle ratio
[, 6]	wt	Weight (lb/1000)
[, 7]	qsec	1/4 mile time in seconds. The time it takes for the car to go from a standstill to a 1/4 of a mile.
[, 8]	vs	V/S whether the car has a V engine (http://en.wikipedia.org/wiki/V_engine) or a straight engine (http://en.wikipedia.org/wiki/Straight_engine).
[, 9]	am	Transmission (0 = automatic, 1 = manual)
[,10]	gear	Number of forward gears
[,11]	carb	Number of carburetors

Exploratory Data Analysis

Here is a pairs plot with histogram and correlation coefficients showing variable relationships.

Evaluating the relationship between Miles per Gallon (MPG) and Transmission Type