Motor Trend Magazine: Is an Automatic or Manual Transmission Better for MPG?

Anthony

Saturday, November 21, 2015

Executive Summary

The realization of the impact of gasoline usage on the environment as well as cost awareness of the consumer has led to fuel efficiency being a major attribute in the automobile industry. Motor Trend Magazine has performed an in-depth analysis in order to answer the question of which transmission, automatic or manual, is better for MPG? Exactly how much better or worse? Or that perhaps there is no significant difference at all?

Data Exploration

The data used in this analysis was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models).

A data frame with 32 observations on 11 variables.

Documentation on the data frame can be foundhere:

https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html)

See Appendix for pairs chart.

Fitting of Statistical Models

Based on a the pairs charts from the data exploreation, Automatic/Manual Transmission, Cylinder, Weight, Horsepower, and Displacement all appear to have some relationship with MPG. A few simple linear regression models were used to do some of the early models. These models quantify and tests the strength the relationship between MPG and the selected variables.

See Appendix for simple linear regression models.

It’s more realistic to believe that there are multiple factors, not just a single variable, that are needed to explain the dependent variable. A few different multivariate regression models were built to better explain MPG.

The models selected below had the best fit based on many different models that were tested. A multivariate linear model with Weight and HP as the explanatory variables to explain MPG was selected. A second model that added the Automatic/Manual Transmission was also created.

```
summary(lm(mpg ~ wt + hp, data=mtcars))$coeff
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.22727012 1.59878754 23.284689 2.565459e-20
## wt          -3.87783074 0.63273349 -6.128695 1.119647e-06
## hp          -0.03177295 0.00902971 -3.518712 1.451229e-03
```

See Appendix for other multivariate linear regression models and nested likelihood ratio test.

Interpreting Coefficient of Outputs

The model with Weight and HP explains the dependent variable MPG the best fit. This model estimate that MPG decreases -3.88 for every 1,000 lbs of weight. The estimate is highly significant with a p-value against the null at .000001. The model also estimates that MPG decreases -0.03 for every unit of Gross Horsepower. This estimate is also very significant with a p-value of .001.

The second model included Weight, HP, and then the dummy variable AM Transmission. This model indicates that after accounting for Weight and HP, the transimission of a car does not have a significant impact on the MPG. The p-value for the AM Transmission is 0.14 which means that the null hypothesis that AM Transmission has no effect can not be rejected.

Furthermore, a nested likelihood ratio test was performed that tested the benefit of adding additional variables to the model. MPG explained by Weight was the base model, adding in HP returns a p-value of 0.001 indicating that it is meaningful to include the variable in the model. Adding another additional variable, Automatic/Manual Transmission returns a p-value of 0.14 indicating that it is not meaningful to include the additional variable.

A summary of the model below shows that the Weight and Horsepower can explain 82.7% (R-squared value) of the dependent variable MPG.

```
summary(lm(mpg ~ wt + hp, data=mtcars))$r.squared
```

```
## [1] 0.8267855
```

Residual Plot and Diagnosis

The residuals of the fitted model using Weight and HP is plotted to diagnose if there are any discrepancies with the model. The Residuals vs Fitted and Scale-Location plot appears that there isn't any clear pattern in the residuals which is a sign of a good fit. The Normal Q-Q plot shows that generally the model is a good fit, however, there may be a couple of outliers in the data. The Residual vs Leverage shows that generally the model is a good fit but there are a couple observations with high leverage on the model.

See Appendix for Residual and Diagnosis plots.

Uncertainty of Conclusions

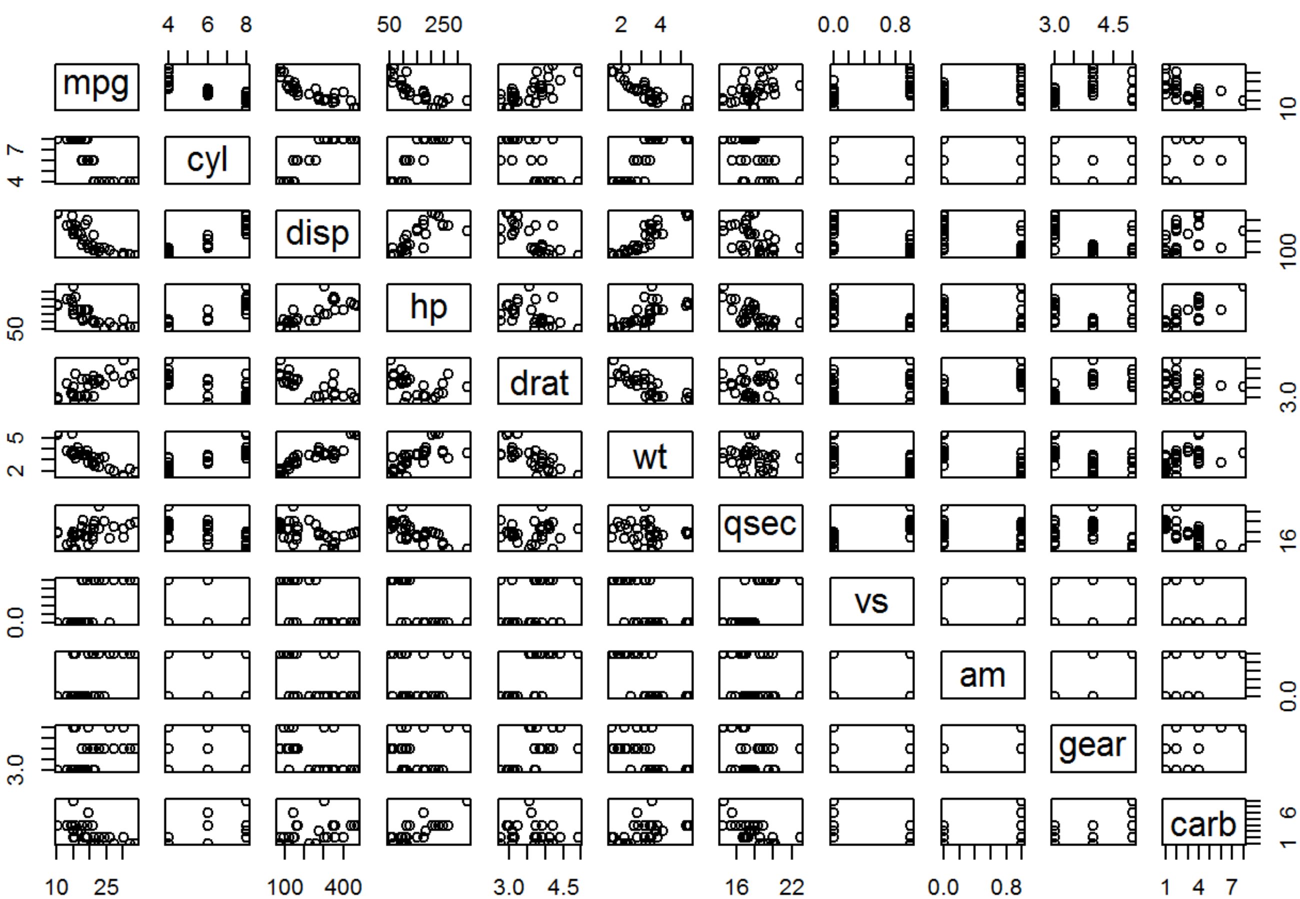
As with all cases of statistical regression modeling, there is a level of uncertainty when trying to quantify what is occuring in the real world. A model will account for things that are known and things that are known to be missing. However, that still leaves the unknown missing factors out of the model. There may be additional confounding variables or deeper levels of interactions between the variables that this model has failed to include.

The conclusion from this analysis is that Automatic and Manual Transmissions have no effect on MPG after accounting for other variables. MPG can be explained by the independent variables Weight and Horsepower. Intuitively it makes sense that MPG would be decreased as the weight and horsepower of a car increases. Advances in automatic transmission technology have increased to such a point such that it is just as efficient as a manual transmission.

Appendix

Pairs chart:

```
pairs(mtcars)
```



Simple linear regression models:

```
summary(lm(mpg ~ factor(am), data=mtcars))$coeff
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

```
summary(lm(mpg ~ wt, data=mtcars))$coeff
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.285126   1.877627 19.857575 8.241799e-19
## wt          -5.344472   0.559101 -9.559044 1.293959e-10
```

```
summary(lm(mpg ~ hp, data=mtcars))$coeff
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 30.09886054  1.6339210 18.421246 6.642736e-18
## hp          -0.06822828  0.0101193 -6.742389 1.787835e-07
```

Multivariate linear regression model including transmission:

```
summary(lm(mpg ~ wt + hp, data=mtcars))$coeff
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.22727012 1.59878754 23.284689 2.565459e-20
## wt          -3.87783074 0.63273349 -6.128695 1.119647e-06
## hp          -0.03177295 0.00902971 -3.518712 1.451229e-03
```

A Nested Likelihood Ratio Test was performed to determine if it is beneficial to sequentially add HP and then to add AM to the Weight model.

```
fit1 <- lm(mpg ~ wt, data=mtcars)
fit2 <- lm(mpg ~ wt + hp, data=mtcars)
fit3 <- lm(mpg ~ wt + hp + factor(am), data=mtcars)
anova(fit1, fit2, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + hp
## Model 3: mpg ~ wt + hp + factor(am)
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 278.32
## 2     29 195.05  1    83.274 12.9328 0.001226 **
## 3     28 180.29  1    14.757  2.2918 0.141268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual Plots:

```
fit2 <- lm(mpg ~ wt + hp, data=mtcars)
par(mfrow = c(2,2), mar = c(3,1,2,2))
plot(fit2)
```