Factors Influencing Vehicle Fuel Economy

J.E. Panzik **June 06, 2020**

Executive Summary

The analysis below shows that many of the affecting fuel economy are related to each other. Using a linear regression model on all of the variables within the mtcars dataset shows each variable does not contribute much to the overall fit ($Pr > 0.05$). **The most significant variables affecting fuel economy are vehicle weight, quarter mile time, and the transmission type of the vehicle.** Having a manual transmission instead of an automatic transmission increases fuel economy by **2.9358372 ± 1.4109045 mpg**. Fuel economy is also decreased by **-3.9165037 ± 0.7112016 mpg** for every **1000 lbs heavier** the car is, and increased by **-3.9165037 ± 0.7112016 mpg** for every 1 second faster the car finishes a quarter mile. The fit with all variables was run through An Information Criterion (AIC) step-wise regression to determine the most significant variables.

Data Exploration

```r
library(datasets); library(knitr)
data <- mtcars
table(is.na(data))
```

```
##
## FALSE
##   352
```

```r
head(data, 3)
```

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

Exploratory plots of fuel economy density distributions and influences are available in the appendix.

Model Fits

The first fit attempts to take all variables as independent and finds the best model for each variables' impact on fuel economy. The second fit takes all of the variables and only selects those that have the strongest relation to fuel economy.

```r
data$am <- as.factor(data$am)
fitAll <- lm(mpg ~., data=data)
kable(summary(fitAll)$coefficients)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.3033742	18.7178844	0.6573058	0.5181244
cyl	-0.1114405	1.0450234	-0.1066392	0.9160874
disp	0.0133352	0.0178575	0.7467585	0.4634887
hp	-0.0214821	0.0217686	-0.9868407	0.3349553
drat	0.7871110	1.6353731	0.4813036	0.6352779
wt	-3.7153039	1.8944143	-1.9611887	0.0632522
qsec	0.8210407	0.7308448	1.1234133	0.2739413
vs	0.3177628	2.1045086	0.1509915	0.8814235
am1	2.5202269	2.0566506	1.2254035	0.2339897
gear	0.6554130	1.4932600	0.4389142	0.6652064
carb	-0.1994193	0.8287525	-0.2406258	0.8121787

```
fitOpt <- step(fitAll, direction="both", trace=FALSE)
kable(summary(fitOpt)$coefficients)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.617781	6.9595930	1.381946	0.1779152
wt	-3.916504	0.7112016	-5.506882	0.0000070
qsec	1.225886	0.2886696	4.246676	0.0002162
am1	2.935837	1.4109045	2.080819	0.0467155

The summary tables above show that including all variables leads to each being fairly insignificant in the fit ($Pr > 0.05$). By doing a step-wise selection, the most significant variables are selected.

The table below shows that the optimal fit is better that the fit to all variables based on the differences of statistical parameters.

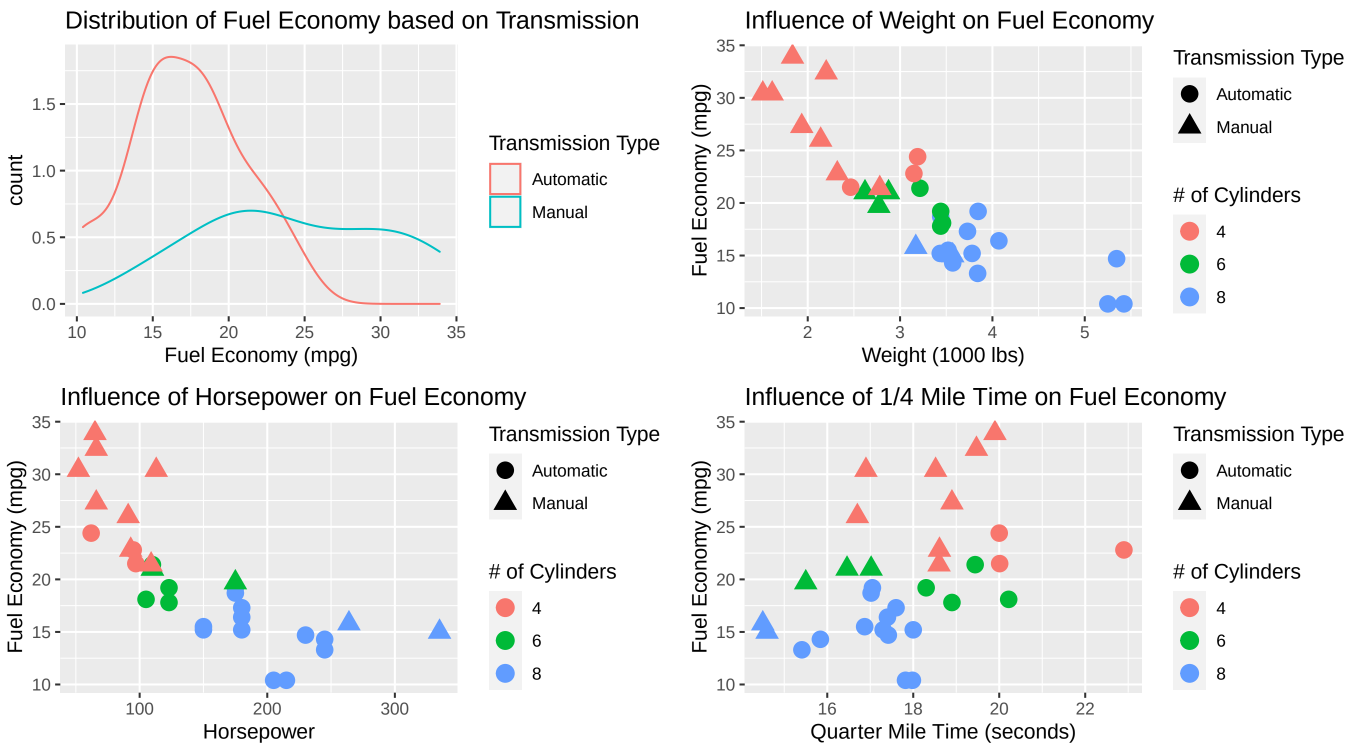
	All Fit	Optimal Fit
Residual Standard Error	2.65	2.459
R Squared	0.869	0.8497
Adjusted R Squared	0.8066	0.8336
F Statistic	13.93	52.75
p Value	3.793e-07	1.21e-11

Conclusions

- Transmission type **does** influence fuel economy with manual transmissions providing and increase of ~3.
- Many of the variables are correlated to each other and their addition adds no significance to interpreting the influences on fuel economy.
- The most significant variable to estimating fuel economy is the overall vehicle weight.
  - Heavier vehicles decrease fuel economy
- The diagnostic plots for the optimal fit shown in the Appendix do not displayw any obvious patterns of concern.

Appendix Figures

```r
library(ggplot2); library(gridExtra)

p1 <- ggplot(data, aes(x=mpg, color=am)) +
      geom_density(aes(y = ..count..)) +
      xlab("Fuel Economy (mpg)") +
      ggtitle("Distribution of Fuel Economy based on Transmission") +
      scale_color_discrete(name = "Transmission Type", labels=c("Automatic", "Manual") )

p2 <- ggplot(data, aes(x=wt, y=mpg, color=as.factor(cyl), shape=am)) +
      geom_point(size=4) +
      xlab("Weight (1000 lbs)") +
      ylab("Fuel Economy (mpg)") +
      ggtitle("Influence of Weight on Fuel Economy") +
      scale_color_discrete(name = "# of Cylinders" ) +
      scale_shape_discrete("Transmission Type", labels=c("Automatic", "Manual"))

p3 <- ggplot(data, aes(x=hp, y=mpg, color=as.factor(cyl), shape=am)) +
      geom_point(size=4)+
      xlab("Horsepower") +
      ylab("Fuel Economy (mpg)") +
      ggtitle("Influence of Horsepower on Fuel Economy") +
      scale_color_discrete(name = "# of Cylinders" ) +
      scale_shape_discrete("Transmission Type", labels=c("Automatic", "Manual"))

p4 <- ggplot(data, aes(x=qsec, y=mpg, color=as.factor(cyl), shape=am)) +
      geom_point(size=4) +
      xlab("Quarter Mile Time (seconds)") +
      ylab("Fuel Economy (mpg)") +
      ggtitle("Influence of 1/4 Mile Time on Fuel Economy") +
      scale_color_discrete(name = "# of Cylinders" ) +
      scale_shape_discrete("Transmission Type", labels=c("Automatic", "Manual"))

grid.arrange(p1, p2, p3, p4, nrow=2)
```




```r
par(mfrow = c(2, 2))
plot(fitOpt)
```