Motor Trends Exploration

Jeffrey Uslan

Monday, August 17, 2015

Exectuive Summary

This report will explore data from motor trend magaizine. The data involves 10 variables describing 32 vehicles. We will discuss the effect of vehicle transmission type of fuel efficiency and to what extent does the superior mode differ. Intial results imply that manual transmission is more efficient with a mdeian mpg of 22.8 while automatic transimission vehicels have a median of 17.3. A t-test of the two classes show that the difference is statistically significant with a p-value of 0.0014. A regression analysis exploration yielded a best model of mpg on vehicle weight, quarter mile time, and transmission type (MPG = 9.6-3.92*Weight+1.2*quarter mile time+2.9*Manual Transmission).

Exploratory Data Analysis

Load in relevant libraries and view data

```
library(ggplot2)
library(dplyr)
data(mtcars)
mtcars=tbl_df(mtcars)
```

Basic summary statistics of fuel efficiency tabualted by transmission type

```
mtcars %>% group_by(Transmission) %>% summarise(min(mpg),mean(mpg),median(mpg),max(mpg))
```

```
## Source: local data frame [2 x 5]
##
##   Transmission min(mpg) mean(mpg) median(mpg) max(mpg)
## 1    Automatic     10.4  17.14737        17.3     24.4
## 2       Manual     15.0  24.39231        22.8     33.9
```

We can see that the median MPG for am automatic transmission is 17.3 mpg while the manual transmission is 22.8 mpg, initially suggesting that manual transmission has better fuel efficiency.

Perform a T-test on the two groups

```
t.test(data=mtcars,mpg~Transmission)
```

The T-test returns a p-value of 0.0014. We reject thte null hypothesis and accept that the true difference in means is not equal to zero

Regression Analysis

We begin by regressing on all numeric variables available and transimission as a categorical variable

```
mostly_numeric_model=lm(mpg~Transmission+cyl+hp+disp+drat+wt+qsec,data=mtcars)
# summary(mostly_numeric_model)
```

The above model has an adjusted R^2 value of 0.829, but has very few statistically significant coefficients.

We can use backwards stepwise selection of our previous model to omit variable with low corresponding coefficient p-values

```
stepwise_model=step(mostly_numeric_model,k=log(nrow(mtcars)))
# summary(stepwise_model)
```

The stepwise model has an adjusted R^2 value of 0.834 and has only statistically significant coefficients. It has determinied that the relevant variables are vehicle weight, quarter mile time, and transmission type.

As the previous model is a strict subset of the large model we can perform an analsis of variance to determine if the simplification was necessary.

```
anova(mostly_numeric_model,stepwise_model, test="Chi")
```

The ANOVA returns a P-value of 0.5, larger than 0.05. So, according to our criterion, we would fail to reject the simplified model.

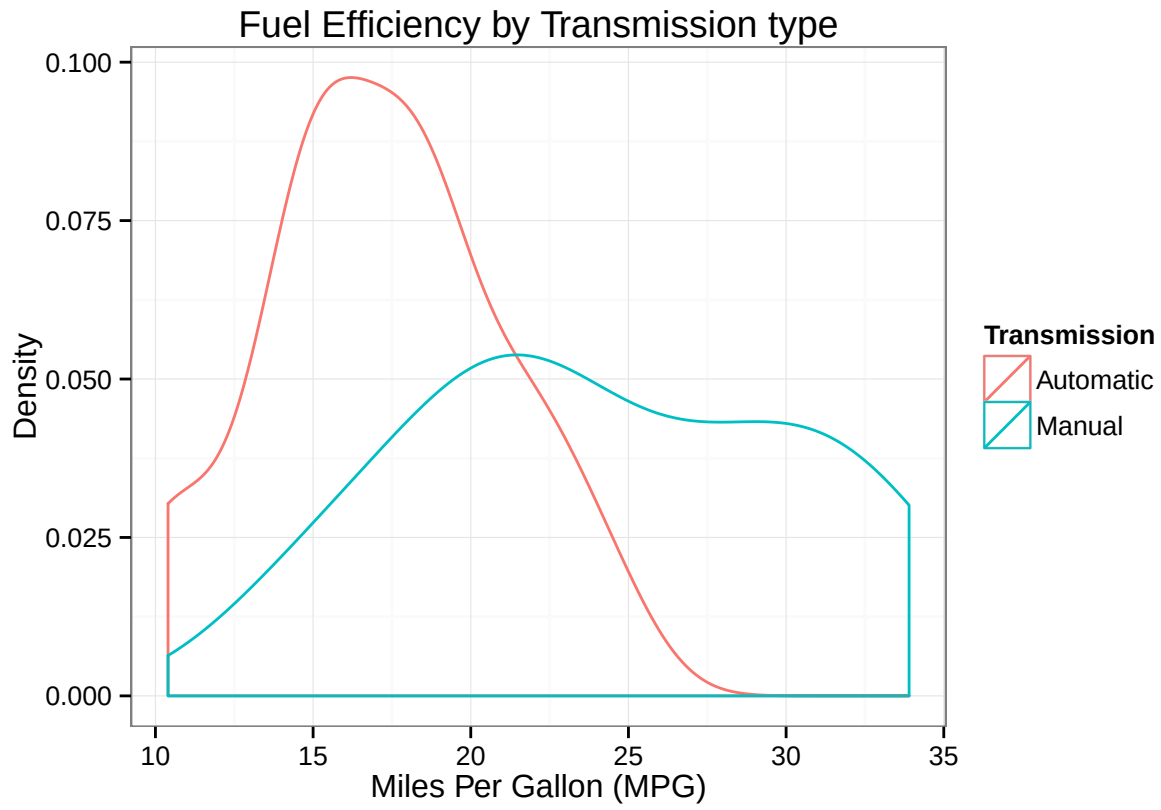
Model Residuals and Diagnostics

Please see images in the appendix for supporting material. Our regression model diagnostics yielded the following results: 1. The plot of the fitted values against residuals shows no autocorrelation nor heteroscedasticity. 2. The quantile plot of the residuls shows the residuals are approximately normally distributed.

Appendix Figures

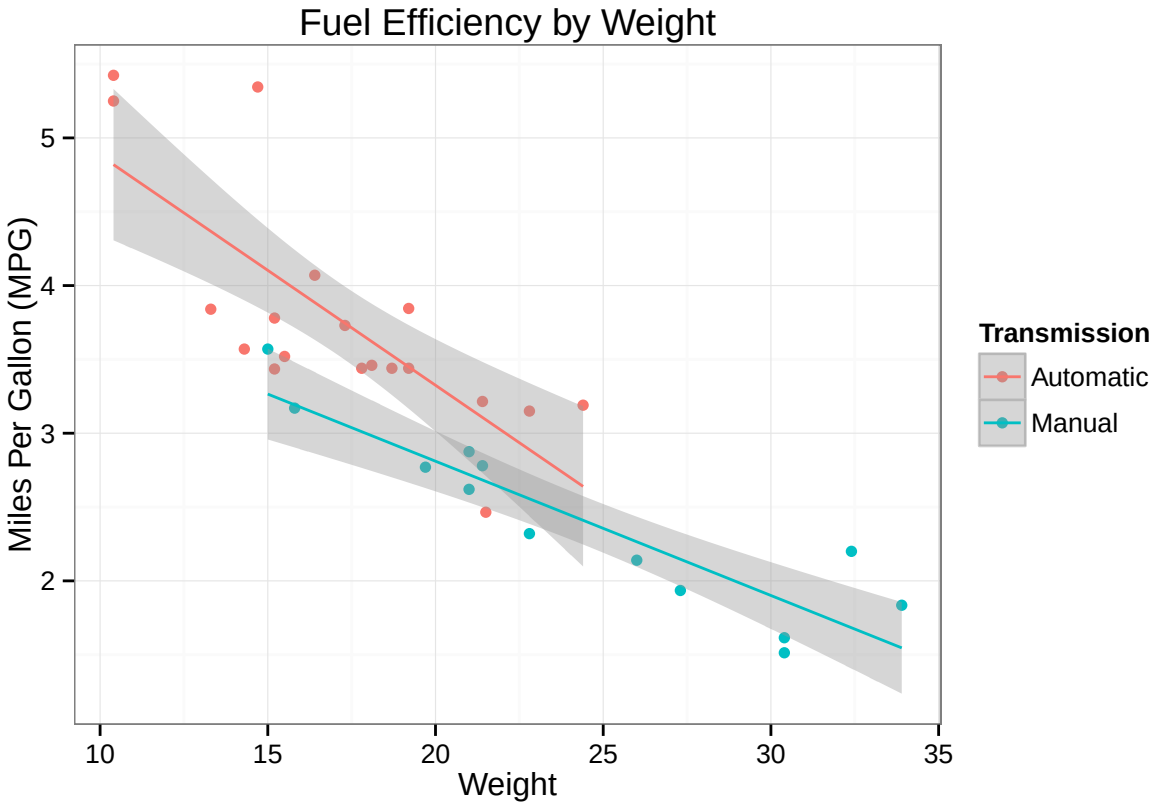
1. Density comparion of fuel efficieny against vehicle transmission type.

```
ggplot(data=mtcars)+theme_bw()+geom_density(aes(x=mpg,color=Transmission))+
  ggtitle("Fuel Efficiency by Transmission type")+xlab("Miles Per Gallon (MPG)")+ylab("Density")
```



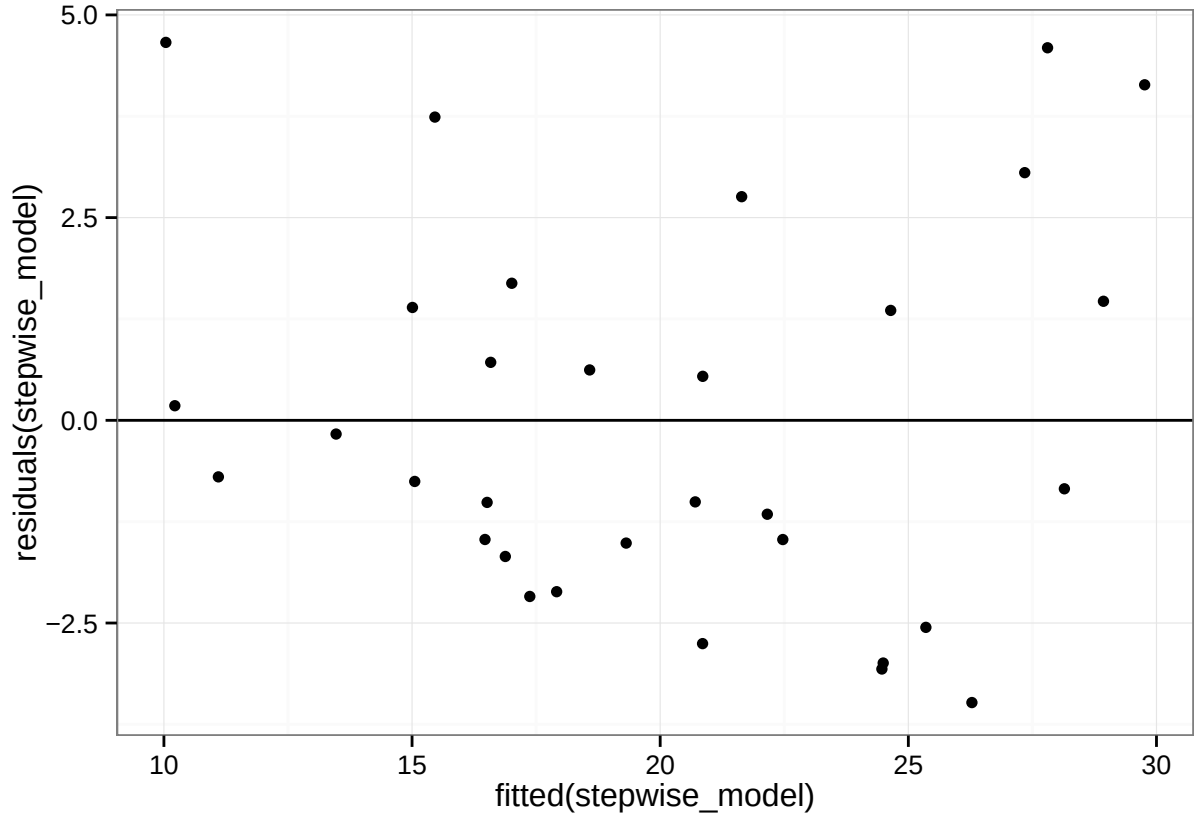
2. A comparison of fule efficiency by vehicle weight seperated by transmission type.

```
ggplot(data=mtcars)+theme_bw()+geom_point(aes(y=wt,x=mpg,color=Transmission))+
  geom_smooth(method=lm,aes(y=wt,x=mpg,color=Transmission))+
  ggtitle("Fuel Efficiency by Weight")+ylab("Miles Per Gallon (MPG)")+xlab("Weight")
```

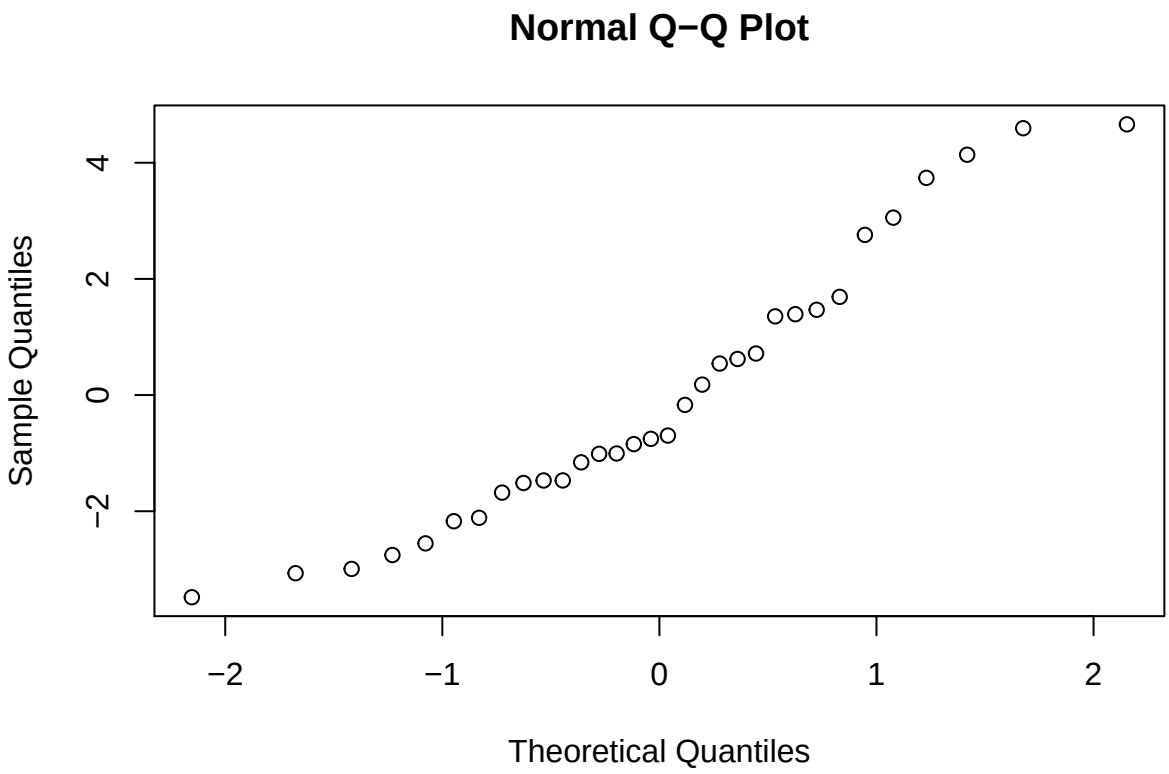


3. Residual diagnostics

```
qplot(fitted(stepwise_model),residuals(stepwise_model))+geom_hline(yintercept=0)+theme_bw()
```




```
qqnorm(residuals(stepwise_model))
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−2
0
2
4
−2
−1
0
1
2