

# Gas Mileage Analysis of Motor Trend Data

*David Maust*

*06/21/2014*

## Executive Summary

When selecting a car, fuel efficiency is a common selection criteria. In this paper, we look at 1974 Motor Trend data [1] for the purpose of evaluating factors on fuel efficiency. We specifically are interested in the effects of automatic vs manual transmission on the gas mileage. By looking at several possible models, we see a relationship does exist between fuel efficiency and transmission type, but that could also be explained by other factors such as vehicle weight.

## Preprocessing the Data

We begin by loading the data, and casting the transmission type and number of cylinders to factors.

```
data(mtcars)
mtcars$am <- factor(mtcars$am, levels=c(0,1), labels=c('Automatic', 'Manual'))
mtcars$cyl <- factor(mtcars$cyl)
```

## Model Selection

First we can check that a relationship does exist with a simple linear model.

```
model1 <- lm(mpg ~ am, mtcars)
summary(model1)$coefficients
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125   15.247 1.134e-15
## amManual       7.245      1.764    4.106 2.850e-04
```

We can then see that a significant relationship does exist with a p-value  $< 0.001$  when we do not consider other factors. Vehicles in our sample with a manual transmission got on average 7.2449 more miles per gallon. We must see if other confounders could also explain the relationship.

We can then look at additional variables: weight, and number of cylinders.

```
model2 <- update(model1, mpg ~ am + wt)
model3 <- update(model1, mpg ~ am + wt + cyl)
anova(model1, model2, model3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + cyl
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
```

```
## 1      30 721
## 2      29 278 1      443 65.31 1.1e-08 ***
## 3      27 183 2      95 7.04 0.0035 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that these are also very influential, and should be included in the model.

```
summary(model2)$coefficients
```

```
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 37.32155      3.0546 12.21799 5.843e-13
## amManual    -0.02362      1.5456 -0.01528 9.879e-01
## wt          -5.35281      0.7882 -6.79081 1.867e-07
```

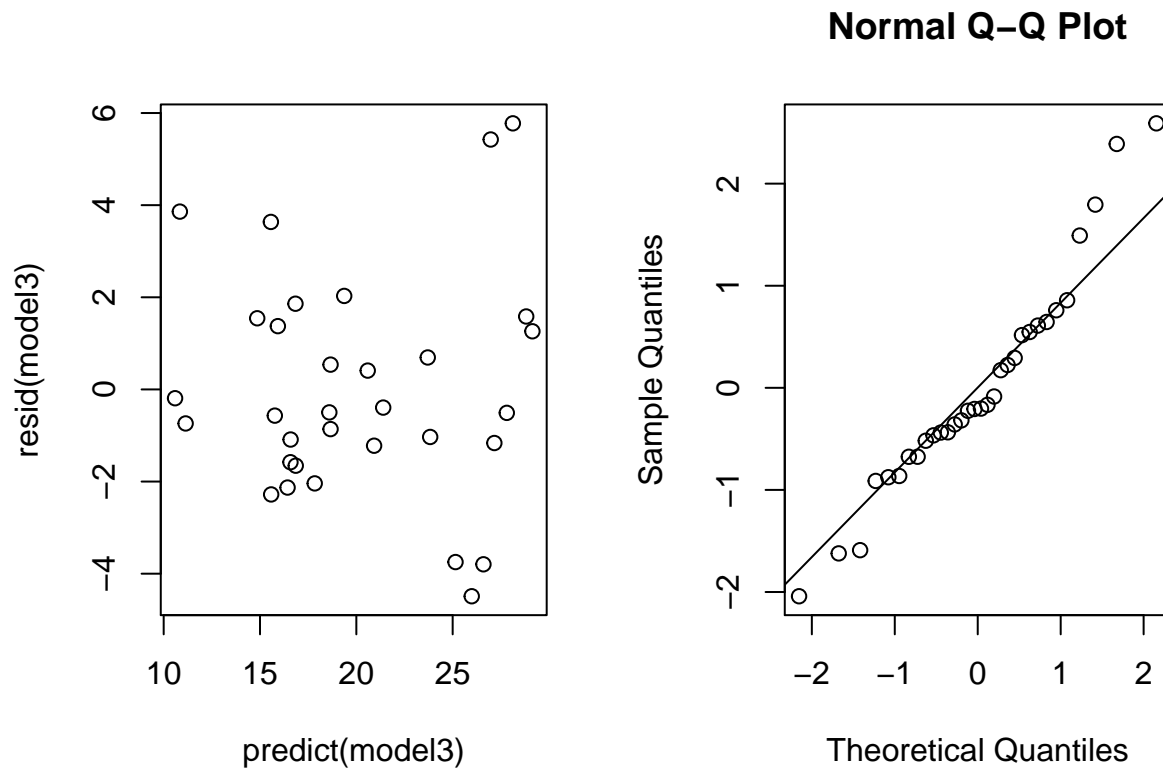
In this model with the vehicle weight added we see a high p-value for the manual transmission term. This would indicate that multicollinearity exists. Vehicle weight is correlated to both the transmission type and fuel efficiency. Also see figure 1 in the appendix for a plot of this relationship.

## Residual Analysis

Looking at a residual plot, and the Q-Q plot of the residuals vs a normal we can see the studentized residuals are approximately normal.

```
par(mfcol = c(1,2))
plot(resid(model3) ~ predict(model3))

qqnorm(rstudent(model3))
qqline(rstudent(model3))
```



## References

[1] Henderson and Velleman (1981), Building multiple regression models interactively. *Biometrics*, 37, 391–411.

## Appendix

Figure 1: mpg comparision with respect to weight.

```
plot(mpg ~ wt, mtcars, col=am, xlab = 'Vehicle Weight')
abline(lm(mpg ~ wt, mtcars))
legend('topright', lty=c(1,1), col=1:2, legend = c('Automatic', 'Manual'))
```

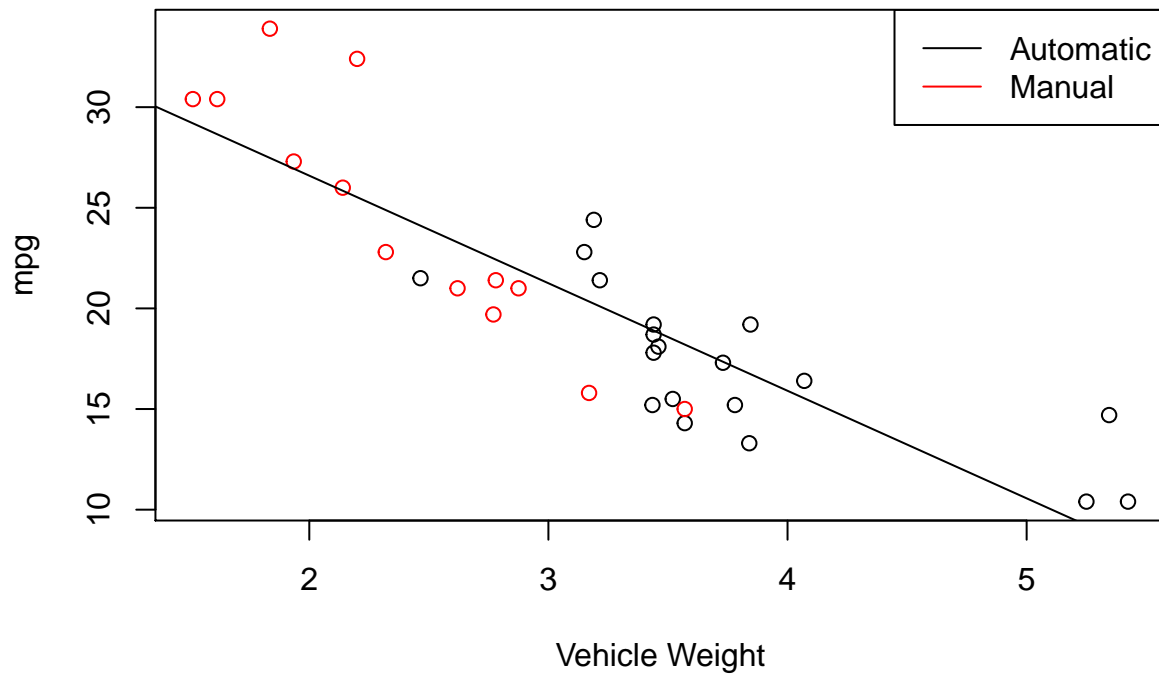


Figure 2: mpg comparison with respect to number of cylinders.

```
plot(mpg ~ cyl, mtcars, col=am)
```

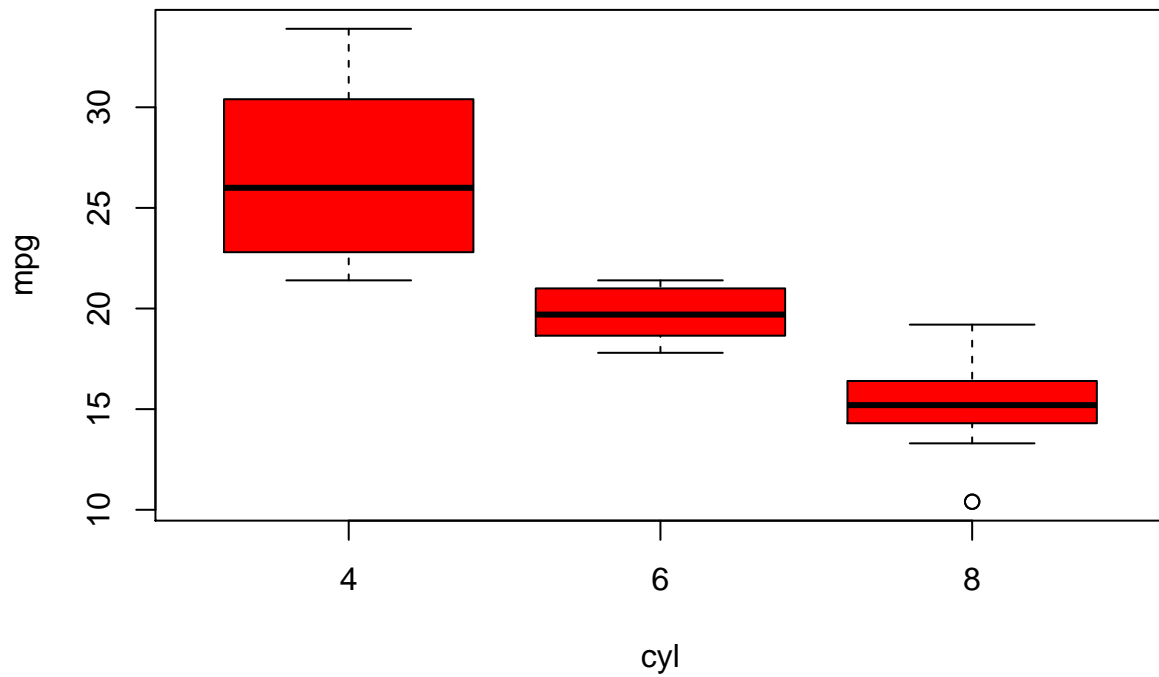


Figure 3: mpg comparison with respect to horsepower.

```
plot(mpg ~ hp, mtcars, col=am)
abline(lm(mpg ~ hp, mtcars))
legend('topright', lty=c(1,1), col=1:2, legend = c('Automatic', 'Manual'))
```

