Motor Trend Analysis - Which Transmission is Better?

Desmond Wallace

Saturday, August 23, 2014

Executive Summary

For this project, I attempt to answer two important questions. First, which type of transmission is better for a car's miles per gallon (MPG), automatic or manual? Second, what is the difference in a car's MPG between the two transmission types? Based on the analysis conducted below, it appears that cars with a manual trasnmission have a better MPG than cars with an automatic transmission.

Exploratory Anaylsis

Before any statistical analysis is conducted, it is best to begin with an exploratory look at the relationship between a car's transmission and its miles per gallon (MPG). In the original dataset, the variable representing transmission type is treated only as a numeric value. Thus, I create a new variable within the dataset that treats transmission type as a factor.

```r
# Data and required packages were loaded at start of markdown document
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))
```

Now that a factor-type variable for transmission is created, I will now proceed to display a plot that displays the difference, if any, in the MPG for cars with an automatic transmission compared to cars with a manual transmission. As shown by the boxplot in the appendix, cars with a manual transmission tend to have higher miles per gallon than cars with an automatic transmission.

Regression Analysis

For the first part of the analysis, I estimate the bivariate relationship between transmission type and miles per gallon.

```r
Basic.Model <- lm(mpg ~ am, data = mtcars)
summary(Basic.Model)$coef
```

```
##             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
```

According to the results, there appears to be a significant relationship between a car's transmission type and its MPG. Cars with an automatic transmission appears to achieve 17.147 miles per gallon. However, cars with a manual transmission appear to achieve 24.392 miles per gallon. This analysis, however, spurs an additional question. Does this relationship holds when other features and aspects of cars are included in the model as controls?

For the second model, I include all the variables from the dataset into the model. Again, some of the variables are classified as numeric and require transformation to factor-type variables.

```r
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
Full.Model <- lm(mpg ~ ., data = mtcars)
Full.Model$coef
```

```
## (Intercept)        cyl6        cyl8        disp          hp        drat
##    23.87913    -2.64870    -0.33616     0.03555    -0.07051     1.18283
##          wt        qsec         vs1    amManual       gear4       gear5
##    -4.52978     0.36784     1.93085     1.21212     1.11435     2.52840
##       carb2       carb3       carb4       carb6       carb8
##    -0.97935     2.99964     1.09142     4.47757     7.25041
```

According to the full model, cars with an automatic transmission have a MPG of 23.880, holding all other variables constant. However, cars with a manual transmission have a MPG of 25.091, holding all other variables constant. However, the biggest difference between this model and the bivariate model is that the variable representing the car's transmission type has a p-Value of 0.7113, while the same variable in the bivariate model has a p-Value of 0.000285. Based on these results, I decide to conduct an F-test on the two models, to determine which model is better.

```r
anova(Basic.Model, Full.Model)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     30 721
## 2     15 120 15       600 4.99 0.0018 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the ANOVA table above, it appears that the full model has the lower p-Value of the two models. Thus, to answer the two questions posed at the beginning of this report, I rely on the results from the full model.

Conclusion

To summarize the results from above, it appears that cars with a manual transmission have a higher MPG than cars with an automatic transmission, holding all other factors constant. Numerically, cars with a manual transmission appear to have a MPG of 25.091. This MPG is approximately 1.212 MPG higher than for cars with an automatic transmission. Such cars have a MPG of 23.880. However, given that the data is from 1974, and that there are only 32 observations in the entire dataset, I do not have great confidence in this report's findings.

Appendix

Figure 1 - Boxplot of Differences in Miles per Gallon Based on Transmission Type

```
g <- ggplot(aes(am, mpg, fill = am), data = mtcars)
g + geom_boxplot() + scale_fill_manual(name = "Legend", values = c("red", "yellow"),
                                       labels = c("0" = "Automatic", "1" = "Manual")) +
  ggtitle("Miles per Gallon by Transmission") +
  xlab("Transmission Type") + ylab("Miles per Gallon")
```

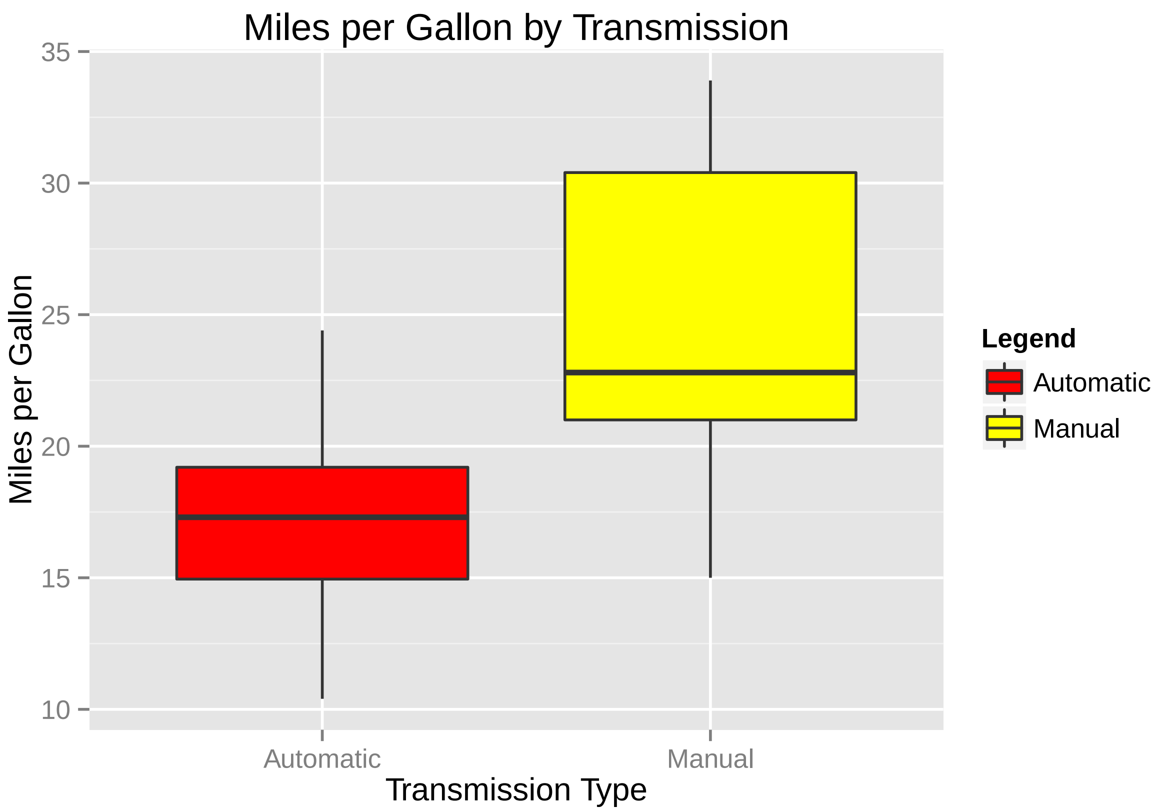


Figure 2 - Diagnostic Plots of Bivariate Regression Model

```
layout(matrix(c(1,2,3,4),2,2))
plot(Basic.Model)
```

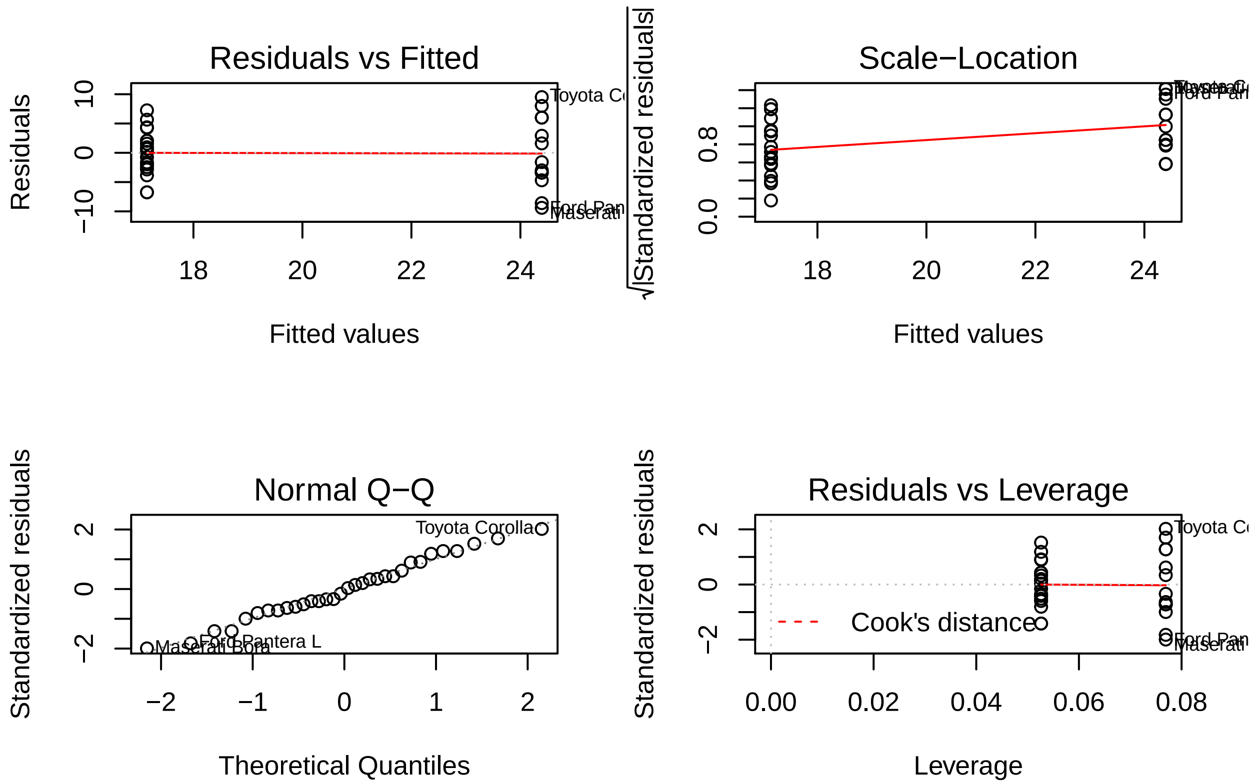
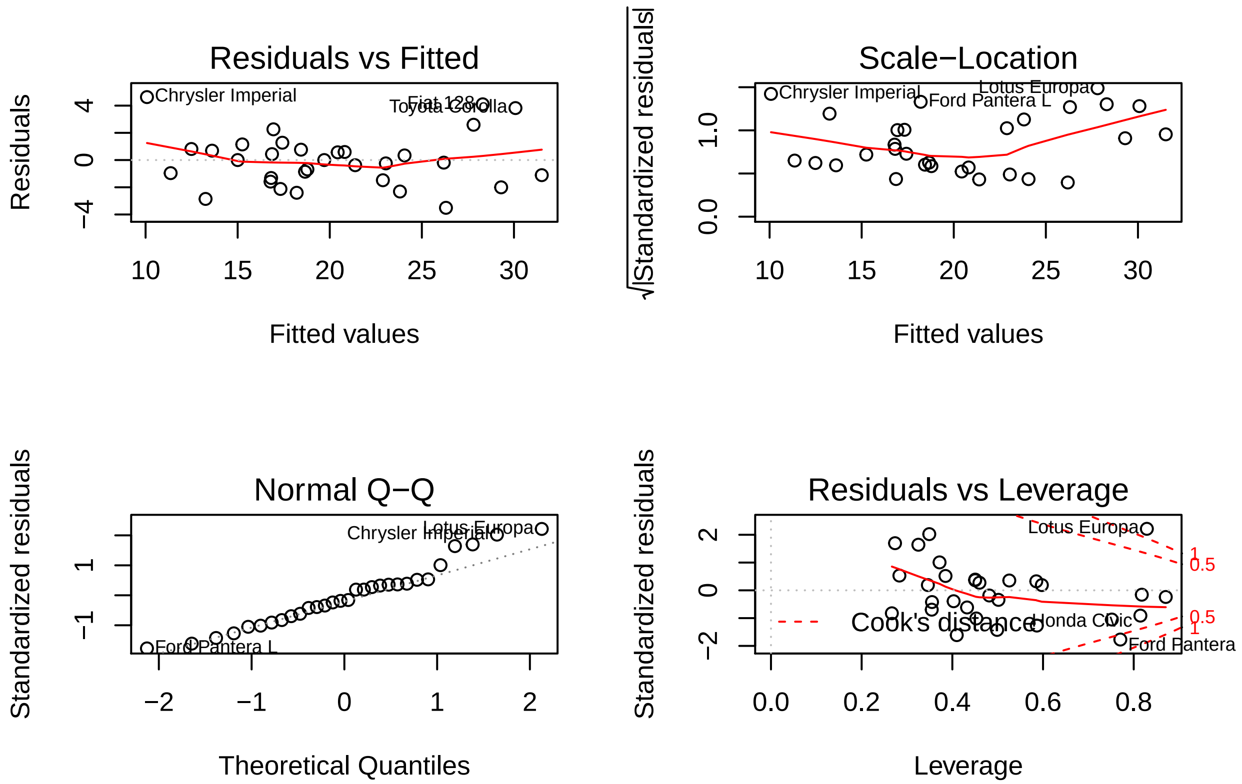


Figure 3 - Diagnostic Plots of Full Multivariate Regression Model

```
layout(matrix(c(1,2,3,4),2,2))
plot(Full.Model)
```

```
## Warning: not plotting observations with leverage one:
##   30, 31
## Warning: not plotting observations with leverage one:
##   30, 31
```



Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Lotus Europa
Ford Pantera L
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Chrysler Imperial
Lotus Europa
Ford Pantera L
Residuals vs Leverage
Standardized residuals
Leverage
Lotus Europa
Honda Civic
Ford Pantera L
Cook's distance