Motor Trend Analysis - Which Transmission is Better?

Desmond Wallace

Tuesday, May 19, 2015

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? In order to answer these two questions, I analyze the data set `mtcars`, and analyze the relationship between transmission type and MPG, while controlling for other possible influences on MPG. The data used to perform this analysis is from the 1974 *Motor Trend* US magazine, including fuel consumption and 10 features of automobile design and performance for 32 automobiles. Each included automobile is either a 1973 or model. The methodology for this report consists of a combination of exploratory analysis and regression analysis. For each analysis, the null hypothesis is that there is not a significant difference in MPG based on transmission type. The alternative hypothesis is that there is in fact a statistically significant difference. Based on the analysis conducted below, it appears that cars with a manual transmission have a better MPG than cars with an automatic transmission.[1]

Exploratory Anaylsis

I begin the analysis with an exploratory look at the relationship between a car's transmission and its miles per gallon (MPG). In the original data set, the variable representing transmission type is treated only as a numeric value. Thus, I create a new variable within the data set that treats transmission type as a factor. Note that for the variable `am`, a value of 0 represents **automatic** transmissions, while a value of 1 represents **manual** transmissions.

```
data(mtcars)
library(ggplot2)
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))
```

Now I proceed to create 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 box plot in the appendix, cars with a manual transmission tend to have higher miles per gallon than cars with an automatic transmission. Thus, this initial result provide support for the hypothesis that there is in fact a significant difference in MPG, based on transmission type.

Inferential Statistics

In this section, I present an initial statistical test of whether there is a significant difference between MPG based on transmission type. For the purposes of this analysis, I conducted a two-sample t-test, where the two groups represents the two transmission types.

```
result <- t.test(mpg ~ am, data = mtcars)
result$estimate
```

[1] This report was authored in R Markdown and compiled to pdf using pdflatex (via knitr). To view the raw source, please visit the GitHub repo associated with this project

```
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

```
result$p.value
```

```
## [1] 0.001373638
```

According to the results, the average MPG for cars with an automatic transmission is 17.14737 miles per gallon. The average MPG for cars with a manual transmission is 24.39231 miles per gallon. More importantly, the p-value for this test is 0.001374. Thus, these results provide additional evidence to support the hypothesis that there is difference between transmission types with respect to MPG.

Regression Analysis

The first regression model estimated is the bivariate relationship between transmission type and miles per gallon. I choose this initial model specification because the question I am interested in answering is whether a car's transmission type has a statistically significant effect on its MPG.

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

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

The second model estimated includes all the variables from the data set. The reason for this model specification is that there are other features of cars that may influence MPG, and/or are correlated with transmission type. Failure to include these covariates would result in inefficient standard errors, and possible biased coefficient estimates. These occurrences would result in the drawing of inaccurate conclusions and inferences. Again, some of the variables are classified as numeric and require transformation to factor-type variables.

```
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)
summary(Full.Model)$coef
```

```
##               Estimate  Std. Error     t value   Pr(>|t|)
## (Intercept) 23.87913244 20.06582026  1.19004018 0.25252548
## cyl6        -2.64869528  3.04089041 -0.87102622 0.39746642
## cyl8        -0.33616298  7.15953951 -0.04695316 0.96317000
## disp         0.03554632  0.03189920  1.11433290 0.28267339
## hp          -0.07050683  0.03942556 -1.78835344 0.09393155
```

```
## drat         1.18283018  2.48348458  0.47627845 0.64073922
## wt          -4.52977584  2.53874584 -1.78425732 0.09461859
## qsec         0.36784482  0.93539569  0.39325050 0.69966720
## vs1          1.93085054  2.87125777  0.67247551 0.51150791
## amManual     1.21211570  3.21354514  0.37718957 0.71131573
## gear4        1.11435494  3.79951726  0.29328856 0.77332027
## gear5        2.52839599  3.73635801  0.67670068 0.50889747
## carb2       -0.97935432  2.31797446 -0.42250436 0.67865093
## carb3        2.99963875  4.29354611  0.69863900 0.49546781
## carb4        1.09142288  4.44961992  0.24528452 0.80956031
## carb6        4.47756921  6.38406242  0.70136677 0.49381268
## carb8        7.25041126  8.36056638  0.86721532 0.39948495
```

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.

```
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     30 720.9
## 2     15 120.4 15    600.49 4.9874 0.001759 **
## ---
## 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, it appears that there is at least one factor, besides transmission type, that significantly influences MPG. Thus, it is important that the regression model includes additional covariates besides the variable of interest.

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 data set, 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")
```

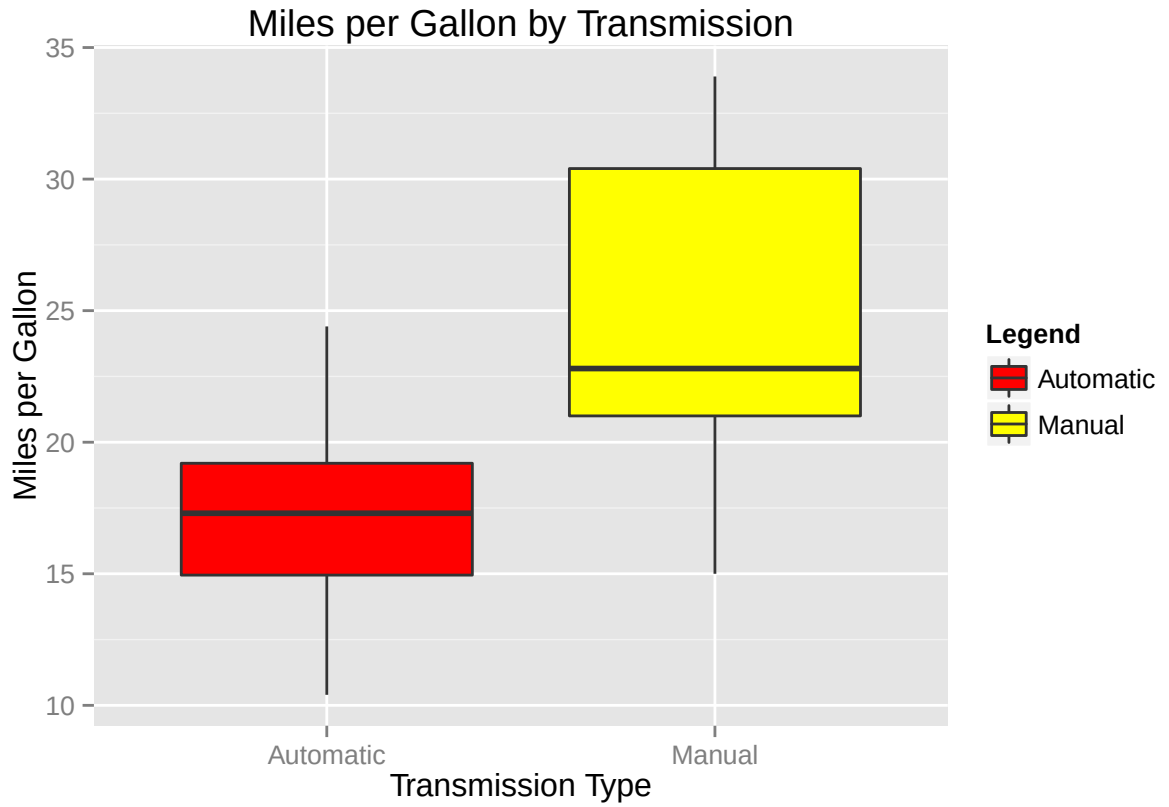
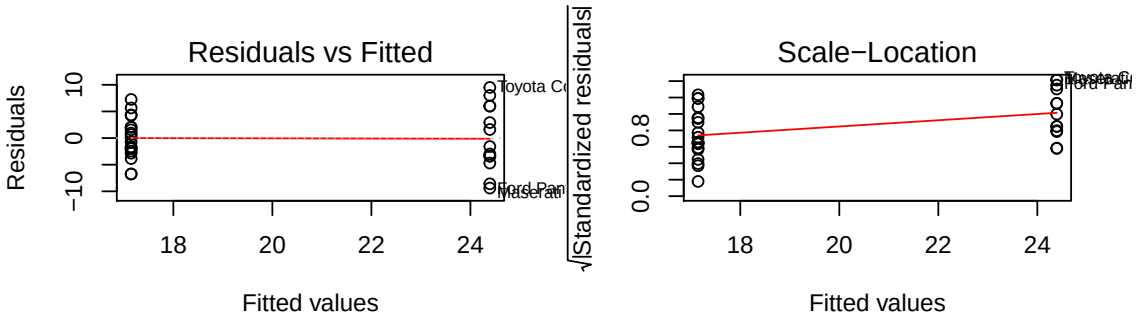
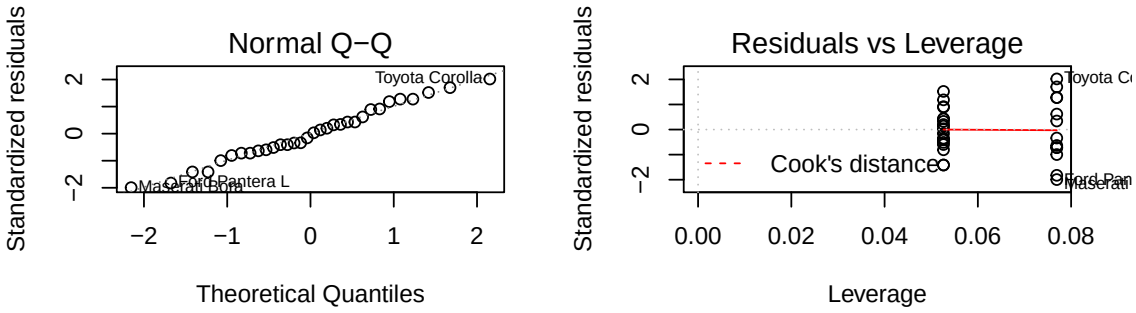


Figure 2 - Diagnostic Plots of Bivariate and Multivariate Regression Models

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






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