Motor Trend: Automatic vs. Manual Transmission

Dax Gerts

July 20, 2015

Executive Summary

This report examined the data provided by R's `mtcars` dataset, originally collected by the Motor Trend US magazine in 1974. `Mtcars` details the mileage (mpg), weight (wt), and transmission type (am), as well as 8 other attributes obtained from observations of 32 automobiles. In this report, we attempted to answer the question of whether automatic or manual transmission provide more miles per gallon, and in doing so tried to quantify the predictors that led to the differences in mileage.

Preliminary analysis suggested that there was a significant difference in mileage (mpg) between automatic and manual transmission. This finding was verified through continued analysis to show that manual transmission consistently scored a higher MPG rating when compared to automatic transmission. The measured difference was found to be an approximate 1.8092 MPG increase when switching from automatic to manual transmission.

Data Preparation

The data was first extracted from `mtcars` (found in the R "datasets" package) and stored in the variable "cars." Cursory examination showed that several attributes of the data would need to be converted from "numerical" values to "factor" values before analysis could be carried out.

```
library(datasets)
library(ggplot2)
require(gridExtra)
data(mtcars)
cars <- mtcars
cars$cyl <- factor(cars$cyl)
cars$vs <- factor(cars$vs)
cars$am <- factor(cars$am,labels=c("Automatic","Manual"))
cars$gear <- factor(cars$gear)
cars$carb <- factor(cars$carb)
attach(cars)
```

Exploratory Analysis and Inference

Once the data was in suitable form, some basic exploratory analyses were conducted. The exploratory graphs can be found in the Appendix, Figures 1-3. The results suggested that manual and automatic transmission yielded distinct values of mpg, and that manual transmission tended to resut in higher mpg values. This was verified quantitatively by taking a t test of the relation between mpg and transmission, with a null hypothesis that both transmission factors were part of the same population.

```
t.test(cars$mpg ~ cars$am)$p.value
```

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

```
t.test(cars$mpg ~ cars$am)$estimate
```

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

A p-value of 0.00137 was sufficiently low enough to reject the null hypothesis, verifying the factors were unique populations. This was further verified by the approximately 7 mpg difference between automatic and manual transmissions.

Regression Model Selection

Because of the primary question of transmission vs. miles per gallon, we began with the baseline model of transmission as the sole predictor against mileage as the dependent. This model was used to test against models containing potentially confounding variables.

```
controlModel <- lm(mpg ~ am, cars)
```

With a baseline established, we built a more complex model. Starting from a linear model constructed with mpg as the dependent variable and all of `mtcars` remaining attributes as predictors, step-wise model selection by the AIC algorithm was employed to construct an optimized model. This processs systematically removed less-helpful predictors "backward" to create a smaller model, while re-checking removed predictors "forward" to ensure useful predictors were not unnecessarily removed.

```
baseModel <- lm(mpg ~.,cars)
optimizedModel <- step(baseModel,direction = "both")
summary(optimizedModel) ##results hidden to reduce length
```

With two models to test, analysis of variances (ANOVA) was used to check whether the more complex model was actually more accurate.

```
anova(controlModel,optimizedModel) ## results hidden to reduce length
```

The resulting p-value indicates that we can safely reject the null hypothesis and conclude that the addition of potentially confounding variables cyl, hp, and wt did improve the accuracy of the model.

Diagnostics

Several final check were made into the accuracy of the model by analysis of the residual, Q-Q, Scale-Location, and Residuals vs. Leverage plots of the optimized models (these plots can be found in the Appendix, Figure 4.). The following characteristics were confirmed: (1) Independence was verified by random scattering of points about 0 in Residuals vs. Fitted plot, (2) Normality was verified by the absence of significant deviations about the line of the Q-Q plot, and (3) Constant Variance was verified by constant/band scattering of points in the Scale-Location plot

Conclusion

The analysis we performed provided the following major conclusions: (1) Manual transmission yields a significant improvement in miles per gallon above automatic transmission, (2) this increase in mpg amounts to approx. 1.8092 miles per gallon when using an adjusted model including car weight, # of cylinders, and gross horsepower. Additionally, it was noted that models using car weight, # of cylinders, and horsepower proved significantly more accurate than those only using transmission type.

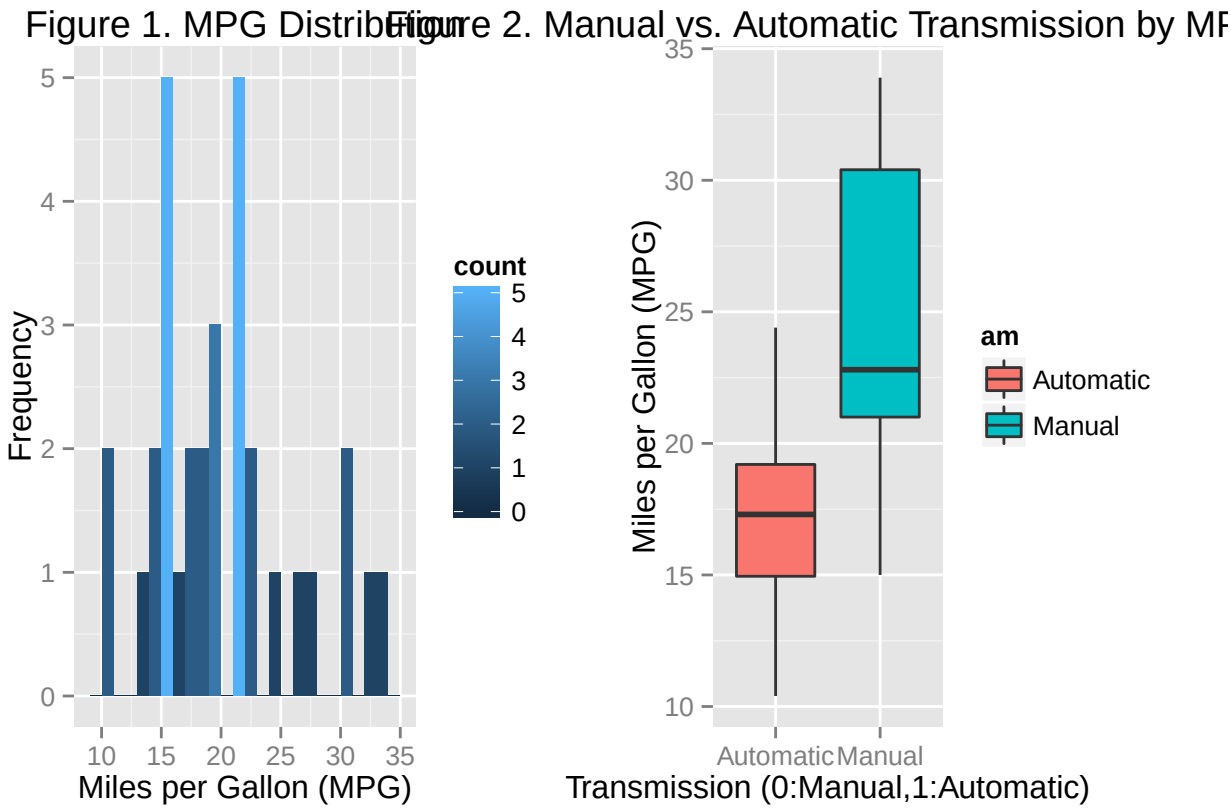
Appendix

```
par(mfrow = c(2,2))
p1 <- ggplot(cars,aes(x=mpg)) +
  geom_histogram(binwidth=1,aes(fill = ..count..)) +
  labs(title="Figure 1. MPG Distribution",x="Miles per Gallon (MPG)",y="Frequency")
```

Figure 1. Histogram (MPG Distribution)

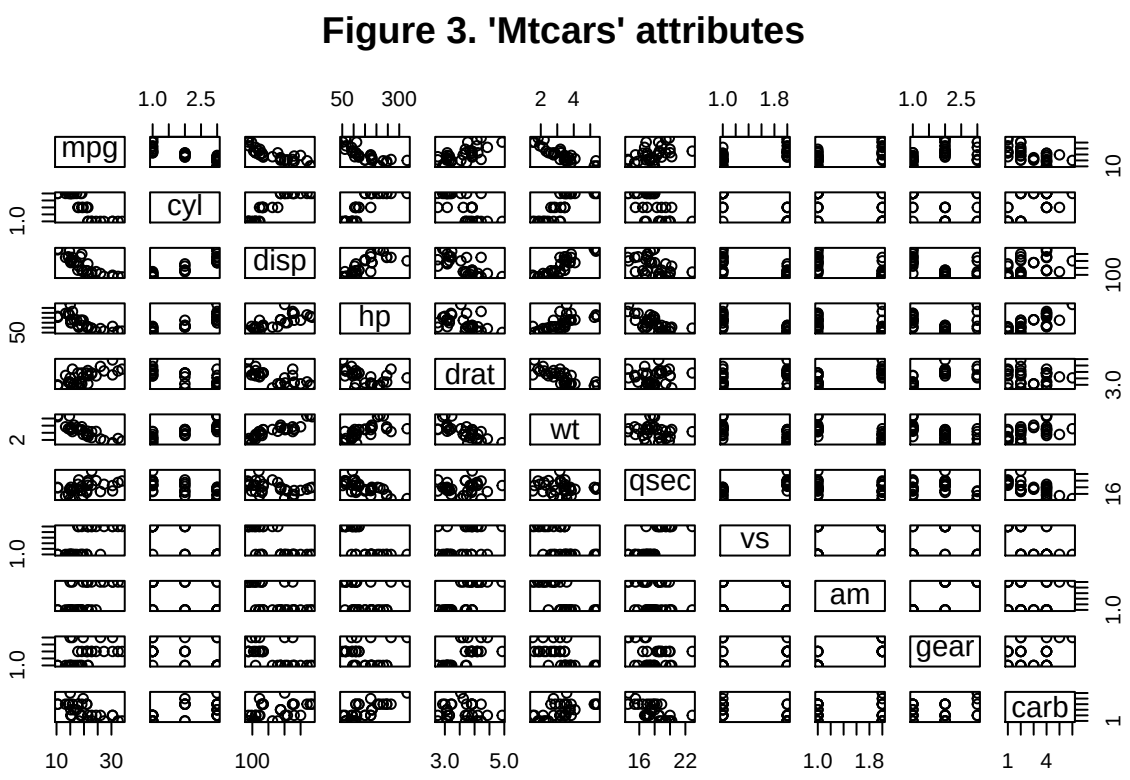
```
p2 <- ggplot(cars,aes(am,mpg)) +
  geom_boxplot(aes(fill=am)) +
  labs(title="Figure 2. Manual vs. Automatic Transmission by MPG") +
  labs(x="Transmission (0:Manual,1:Automatic)",y="Miles per Gallon (MPG)")

grid.arrange(p1,p2,ncol=2)
```

Figure 2. Boxplot: Manual vs. Automatic Transmission by MPG



```
pairs(mpg ~.,cars,main="Figure 3. 'Mtcars' attributes")
```

Figure 3. Pairs plot (All 'mtcars' attributes)



```r
par(mfrow=c(2,2))
plot(optimizedModel,main="Figure 4. Optimized Model Diagnostics")
```

Figure 4. Combined Residual, Q-Q, Scale-Location, Residuals vs. Leverage Plots of Optimized Model (mpg ~ am + cyl + hp + wt)