Effect of Transmission on Miles Per Gallon

Noah DiAntonio

Executive Summary

In this analysis, I ask whether an automatic or manual transmission is better for MPG and quantify the MPG difference between automatic and manual transmissions. I find that there is a difference between automatic and manual transmission for MPG. In particular, I find that on aggregate, manual transmission is better for MPG. However, this does not tell the whole story. After fitting a model for MPG, I find that there is a downward trend for MPG as the weight of the car increases. This trend is different for automatic and manual transmission cars. Manual transmission cars have a steeper downward sloping curve, which means that around 2750 lbs, the MPG for manual transmission cars falls below that of automatic transmission cars. **The better transmission for MPG depends on the weight of the car. For lighter cars, manual transmission is better. For heavier cars, automatic transmission is better.**

Data Preparation and Exploratory Analysis

I begin by opening the data and converting the am variable into a factor variable, labeling its levels "Automatic" and "Manual" respectively. I also load the package "ggplot2," which I will use later on.

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

Next, I perform some basic exploratory analysis. You can see the analysis in **Appendix 2**.

Is an automatic or manual transmission better for MPG?

The first question I want to answer is this: "Is an automatic or manual transmission better for MPG?" I begin by determining whether there is a difference in mpg between automatic and manual transmission using a t-test.

```
t.test(mpg ~ am, data = data)$p.value
```

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

The p-value is less than 0.05 so I reject the null hypothesis that the mean mpg is the same for both automatic and manual transmission. **Appendix 1** contains a boxplot comparing mpg for both levels of am. Looking at it, it is clear that manual transmission has a higher mean for mpg than automatic transmission.

I conclude that a manual transmission is better for MPG than an automatic transmission.

Quantifying the MPG difference between automatic and manual transmissions

To quantify the MPG difference between automatic and manual transmissions, I need to select a linear model that predicts mpg using am and some number of relevant other variables. To select such a model, I create a set of models, with each one introducing one more variable. I then do an anova test on all of these models to determine whether adding each variable has a significant impact on the model. Those which do have a significant impact I will include in my initial model.

Appendix 3 contains the code used to run the anova test. The test finds that the variables cyl (number of cylinders), hp (horsepower), and wt (weight) have significant impacts on the fit. Thus, I fit my first model with am and these variables.

```
fitnew <- lm(mpg ~ am + cyl + hp + wt, data)
summary(fitnew)[4]
```

```
## $coefficients
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 36.14653575 3.10478079 11.642218 4.944804e-12
## amManual     1.47804771 1.44114927  1.025603 3.141799e-01
## cyl         -0.74515702 0.58278741 -1.278609 2.119166e-01
## hp          -0.02495106 0.01364614 -1.828433 7.855337e-02
## wt          -2.60648071 0.91983749 -2.833632 8.603218e-03
```

I notice that while some variables in this model are significant, am is not. This means that the model must be incomplete, because our earlier analysis made it evident that mpg *does* differ when am differs. That means I need to further improve the model.

To improve the model further, I go through the same anova analysis, this time starting with our existing model and then adding interaction terms of am with the other variables. The code for this analysis can be found in **Appendix 4**. This analysis finds that the interaction between am and wt is significant, while other interactions are not. I thus refit the model adding the interaction term am:wt.

```
fitint <- lm(mpg ~ am + cyl + hp + wt + am:wt, data)
summary(fitint)[4]; summary(fitint)[8]
```

```
## $coefficients
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.22944546  3.0169020 11.014426 2.732248e-11
## amManual    11.26037747  3.9198647  2.872644 7.999400e-03
## cyl         -0.83651394  0.5282624 -1.583520 1.253926e-01
## hp          -0.01245722  0.0132152 -0.942644 3.545403e-01
## wt          -2.19535776  0.8463778 -2.593827 1.538609e-02
## amManual:wt -3.72353367  1.4071368 -2.646178 1.363531e-02

## $r.squared
## [1] 0.8810632
```

This model explains 88% of the variance in mpg. This model looks better, because I see that am changing from automatic to manual has an effect. I also notice that weight and the interaction of am and weight have significant effects.

Before reporting on the meaning of these results, the residuals of our fit need to be analyzed. **Appendix 5** contains some summary plots for the residuals. Based on these plots, the residuals appear to be normally distributed and homoskedastic.

I conclude the following: 1. Cars with manual transmissions get 11.3 more miles per gallon than cars with automatic transmissions. 2. For every 1000 lbs a car with automatic transmission weighs, it gets 2.2 fewer miles per gallon. 3. For every 1000 lbs a car with a manual transmission weighs, it gets 5.9 fewer miles per gallon.

This implies that while cars with manual transmissions initially have better miles per gallon, as weight increases their miles per gallon fall faster than cars with automatic transmissions. This implies that there is a weight point at which cars with automatic transmissions actually get better miles per gallon. **Appendix 6** shows this visually. It appears that at a weight of *approximately* 2750 lbs, automatic transmission cars begin to get better gas mileage.

In conclusion, the better transmission for MPG depends on the weight of the car. For lighter cars, manual transmission is better. For heavier cars, automatic transmission is better

Appendices

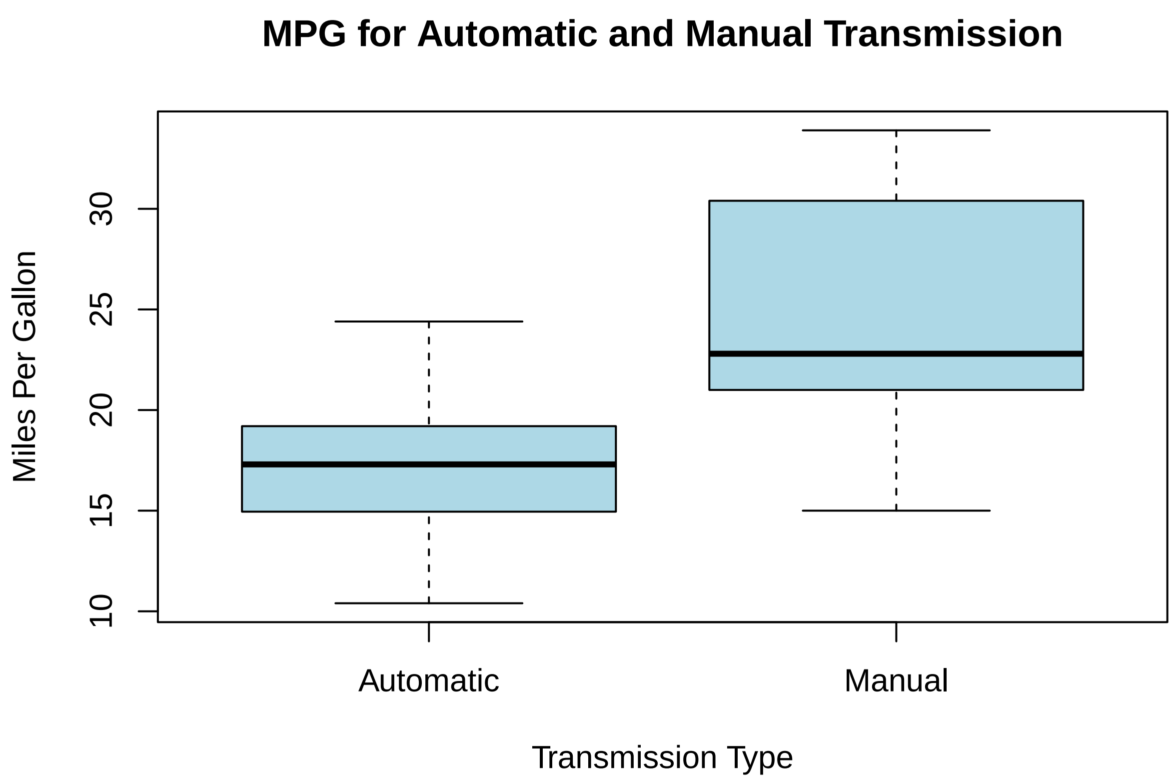
Appendix 1: Exploratory Data Analysis

```
str(data)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Appendix 2: Boxplot

```
boxplot(mpg ~ am, data = data, main = "MPG for Automatic and Manual Transmission",
        ylab = "Miles Per Gallon", xlab = "Transmission Type", col = "light blue")
```



Appendix 3: Model Selection, Part 1

```
fit1 <- lm(mpg ~ am, data)
fit2 <- lm(mpg ~ am + cyl, data)
fit3 <- lm(mpg ~ am + cyl + disp, data)
fit4 <- lm(mpg ~ am + cyl + disp + hp, data)
fit5 <- lm(mpg ~ am + cyl + disp + hp + drat, data)
fit6 <- lm(mpg ~ am + cyl + disp + hp + drat + wt, data)
fit7 <- lm(mpg ~ am + cyl + disp + hp + drat + wt + qsec, data)
fit8 <- lm(mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs, data)
fit9 <- lm(mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear, data)
fit10 <- lm(mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear + carb, data)
anova(fit1, fit2, fit3, fit4, fit5, fit6, fit7, fit8, fit9, fit10)
```

```
## Analysis of Variance Table
##
## Model  1: mpg ~ am
## Model  2: mpg ~ am + cyl
## Model  3: mpg ~ am + cyl + disp
## Model  4: mpg ~ am + cyl + disp + hp
## Model  5: mpg ~ am + cyl + disp + hp + drat
## Model  6: mpg ~ am + cyl + disp + hp + drat + wt
## Model  7: mpg ~ am + cyl + disp + hp + drat + wt + qsec
## Model  8: mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs
## Model  9: mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear
## Model 10: mpg ~ am + cyl + disp + hp + drat + wt + qsec + vs + gear + carb
##    Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1      30 720.90
## 2      29 271.36  1    449.53 64.0039 8.231e-08 ***
## 3      28 252.08  1     19.28  2.7452   0.11241
## 4      27 216.37  1     35.71  5.0849   0.03493 *
## 5      26 214.50  1      1.87  0.2663   0.61121
## 6      25 162.43  1     52.06  7.4127   0.01275 *
## 7      24 149.09  1     13.34  1.8999   0.18260
## 8      23 148.87  1      0.22  0.0309   0.86214
## 9      22 147.90  1      0.97  0.1384   0.71365
## 10     21 147.49  1      0.41  0.0579   0.81218
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 4: Model Selection, Part 2

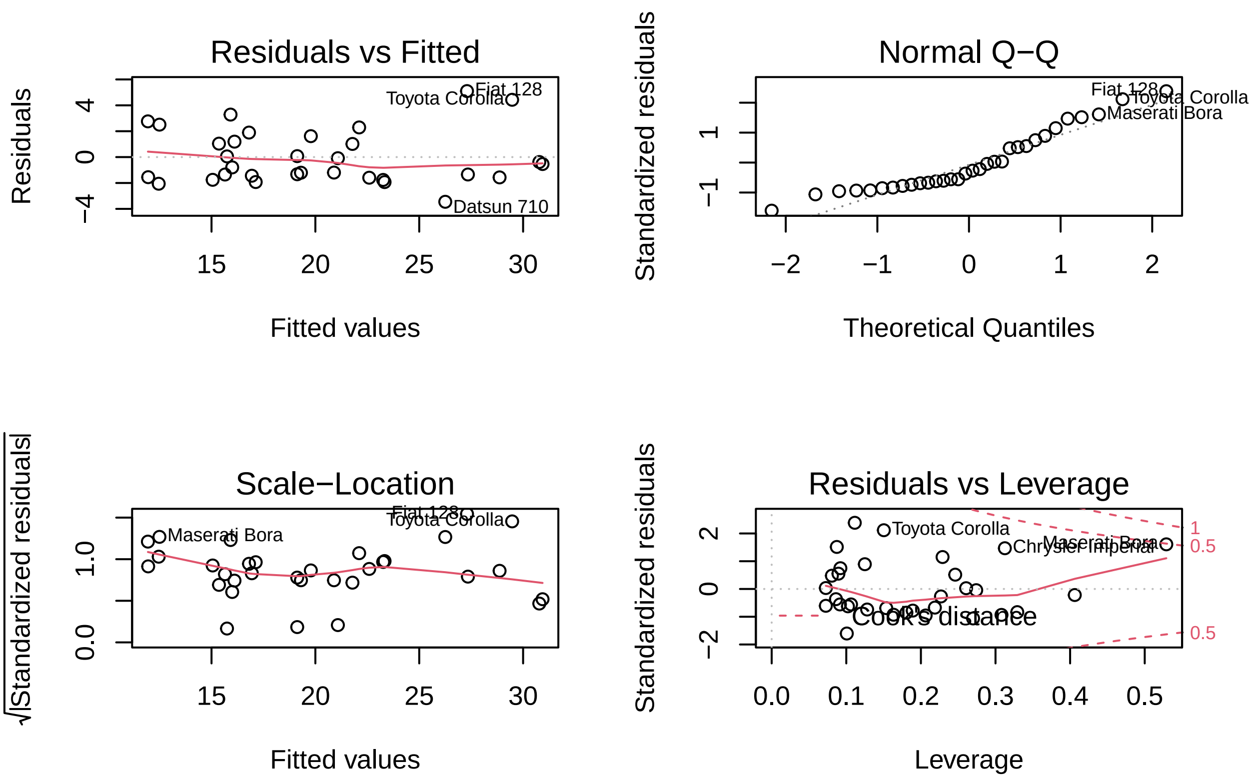
```
fit11 <- lm(mpg ~ am + cyl + hp + wt, data)
fit12 <- lm(mpg ~ am + cyl + hp + wt + am:cyl, data)
fit13 <- lm(mpg ~ am + cyl + hp + wt + am:cyl + am:hp, data)
fit14 <- lm(mpg ~ am + cyl + hp + wt + am:cyl + am:hp + am:wt, data)
anova(fit11, fit12, fit13, fit14)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ am + cyl + hp + wt
## Model 2: mpg ~ am + cyl + hp + wt + am:cyl
## Model 3: mpg ~ am + cyl + hp + wt + am:cyl + am:hp
## Model 4: mpg ~ am + cyl + hp + wt + am:cyl + am:hp + am:wt
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 170.00
## 2     26 158.50  1   11.5002 2.1464 0.1559
## 3     25 158.50  1    0.0011 0.0002 0.9885
## 4     24 128.59  1   29.9051 5.5814 0.0266 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 5: Residual Plots

```r
par(mfrow = c(2,2))
plot(fitint)
```



Appendix 6: Different MPG ~ Weight Slopes For Different Transmissions

```r
qplot(x = data$wt, y = data$mpg, col = data$am, geom = c("point", "smooth"), method = "lm",
      main = "Effect of Weight on MPG for Automatic and Manual Transmission Cars",
      xlab = "Weight (1000 lbs)", ylab = "Miles Per Gallon")
```

```
## Warning: Ignoring unknown parameters: method
```

```
## `geom_smooth()` using formula 'y ~ x'
```