

MPG in Automatic vs Manual Transmissions

Daniel Cosentino

August 30, 2016

Executive Summary

This report looks to determine the difference in miles per gallon (mpg) for automatic vs manual transmission. The dataset used is the mtcars dataset in R. The data was extracted from the 1974 Motor Trend US magazine. It consists of the fuel consumption and 10 aspects of automobile design for 32 vehicles. A t-test was performed and determined that Manual cars have statistically significant higher mpg than automatic cars ($t(18)=-3.77$, $p<.05$). A multivariable regression analysis was then done to control for potential confounding variables. The result of the model found that manual transmission does have a slightly better mpg than automatic after controlling for weight, horsepower, and cylinder size, however the difference was not statistically significant ($b=1.8$, $p > .05$).

Load the Data

Mtcars has 32 cars in its dataset. Each car has its miles per gallon (mpg) and transmission (am(1= manul, 0=automatic)), among others, including number of cylinders (cyl), and V/S (vs).

```
data(mtcars)
mtcars$am <- as.factor(mtcars$am)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
```

Exploratory Data Analyses

We are interested in the difference of mpg for transmission, so I will first look at a boxplot looking at the mpg for automatic and manual transmissions.

```
library(ggplot2)
g <- ggplot(data=mtcars, aes(factor(am), mpg))
g <- g + geom_boxplot() +
  scale_x_discrete(labels=c('Automatic', 'Manual')) +
  labs(title = 'MPG for Automatic vs Manual Cars',
       x = 'Transmission',
       y = 'MPG')
```

From this boxplot (Figure 1) it appears that manual transmissions have a higher mpg than automatic. Looking at the pairs plot (Figure 2) you can see several variables that are correlated with mpg including wt, am, cyl, disp, hp, drat, and vs. We will look to fit these variables into our model that result in best model fit.

Statistical Inference

Before a regression analysis we can perform a t-test to determine if the transmission types are significantly different.

```
t.test(mpg~am, data=mtcars)
```

The t-test shows (Figure 4) that Automatic cars do have statistically significant less mpg than manual cars.

Regression Model Analysis

First, we will look at a simple linear regression model with only transmission as the coefficient.

```
simpfit <- lm(mpg~factor(am), data=mtcars)
```

Here the R-squared is 0.3385 which suggests that am accounts for roughly 34% of the variance in mpg.

Now we will look at and identify the model that results in the best fit. To determine the best model fit, I will use stepwise regression. Stepwise regression selects a model by adding or removing individual predictors based on their statistical significance.

```
library(MASS)
fit <- lm(mpg~., data=mtcars)
step_model <- stepAIC(fit, direction='both')
```

As you can see (Figure 5), this model, that includes am, wt, hp, and cyl as the coefficients, has an adjusted R^2 of .8401 meaning that this model accounts for 84% of the variance in mpg. For our interest in mpg for transmissions, manual transmissions have 1.81 mpg more than automatic transmissions, taking into account the weight, cylinders and horsepower of the cars, however this difference is not significant ($p > .05$). We can also look at the confidence intervals for our transmission coefficient.

```
confint(step_model)[6,]
```

```
##      2.5 %    97.5 %
## -1.060934  4.679356
```

You can see that the confidence intervals go through 0 which also suggests that the transmission coefficient is not significant.

Residual Plots

Now that we have our model, we can look at the residuals to determine if there is any non-normality or any high influence points.

(Figure 3) There is no pattern in the Residuals vs Fitted graph. the Q-Q plot implies that the residuals are normally distributed and the Scale-Location plot suggests constant variance. There also doesn't appear to be any outliers.

Appendix

Figure 1

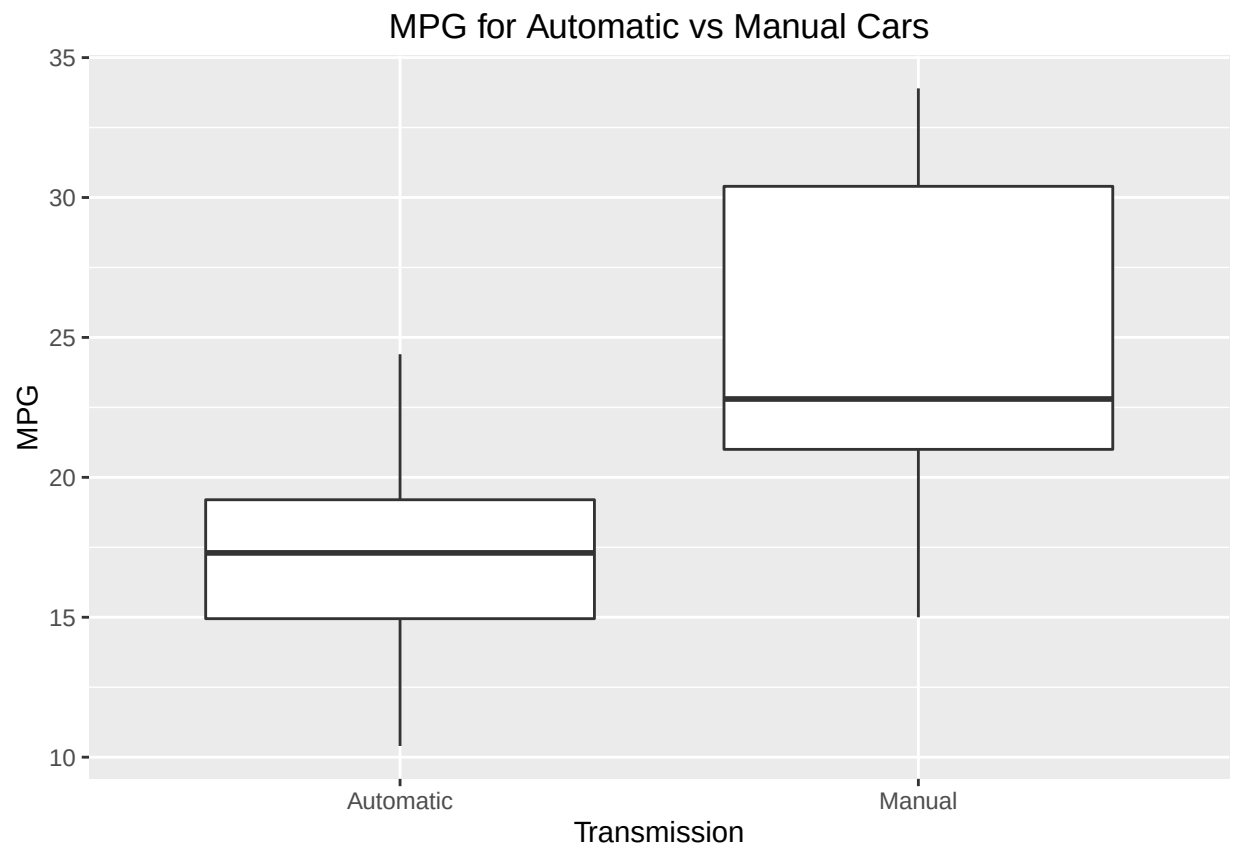


Figure 2

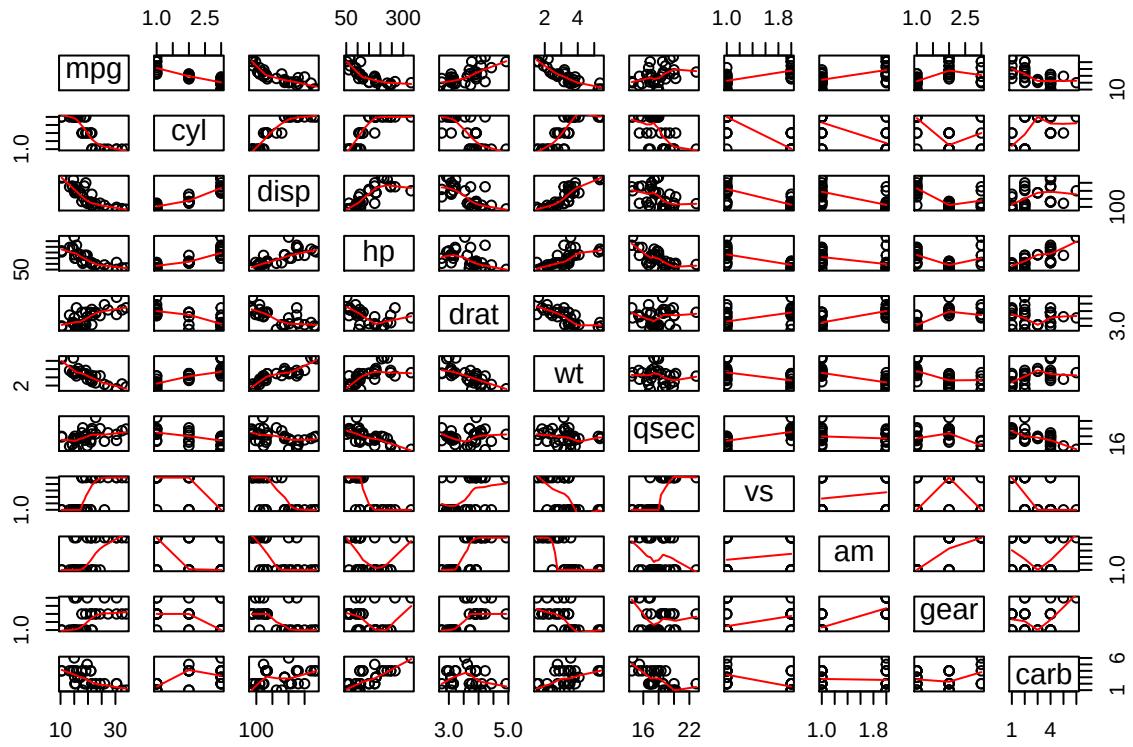


Figure 3

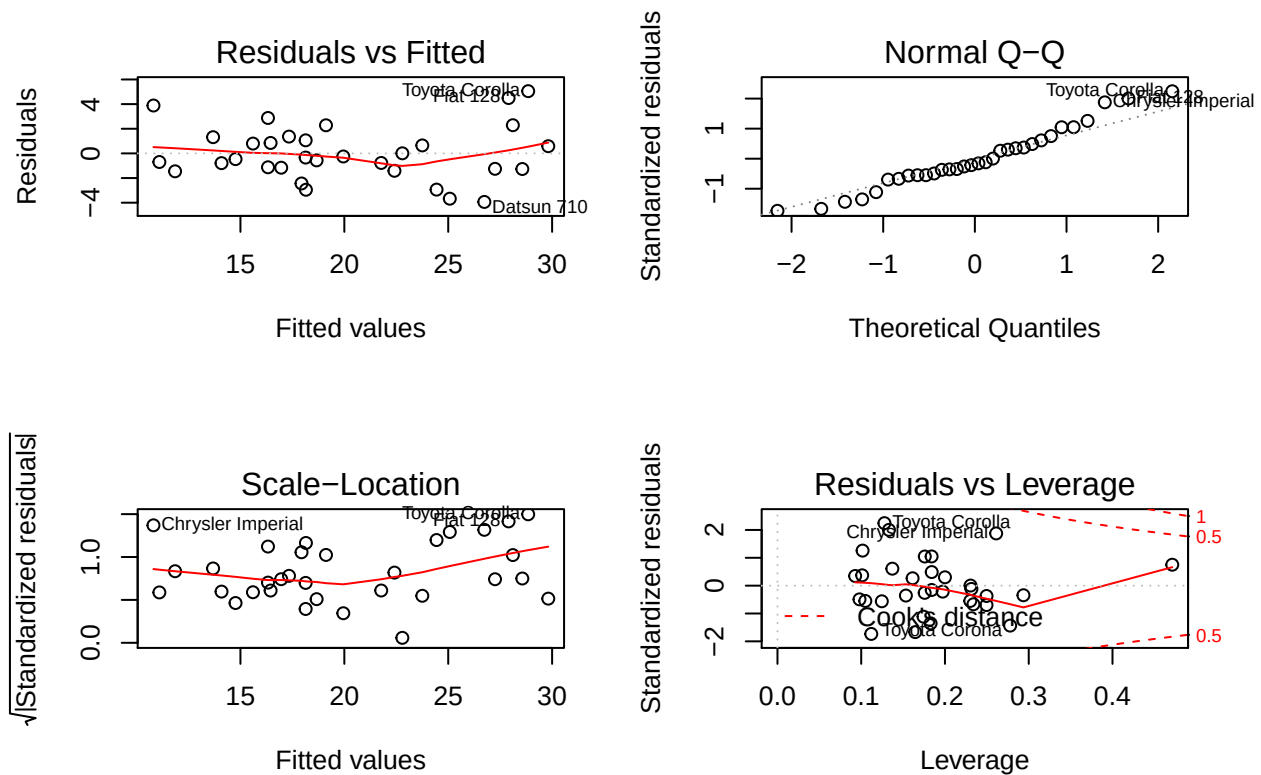


Figure 4

```
t.test(mpg~am, data=mtcars)
```

```
##
## Welch Two Sample t-test
##
## data: mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.280194 -3.209684
## sample estimates:
## mean in group 0 mean in group 1
## 17.14737 24.39231
```

Figure 5

```
summary(step_model)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.70832    2.60489   12.940 7.73e-13 ***
## cyl6         -3.03134    1.40728   -2.154  0.04068 *
## cyl8         -2.16368    2.28425   -0.947  0.35225
## hp           -0.03211    0.01369   -2.345  0.02693 *
## wt           -2.49683    0.88559   -2.819  0.00908 **
## am1          1.80921    1.39630    1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```