MPG Impact of Automatic Transmission

Ben Daniel

September 24, 2015

Summary

Often car afficienados have debated on whether automatic or manual transmissions get better gas mileage. My first car had a manual transmission, and I believed I was getting better gas mileage because I had complete control over the car. But nostalgia for my first ride aside, this study will look at the data in the mtcars dataset and plot it on a linear model to examine the projected effect of an automatic transmission on fuel efficiency measured in miles per gallon (MPG).

Data Analysis

First, let's fit a linear model associating MPG and whether or not the transmission is automatic or manual.

```
library(knitr)
data(mtcars)
fit<-lm(mpg ~ am, data=mtcars)

summary(fit)
```

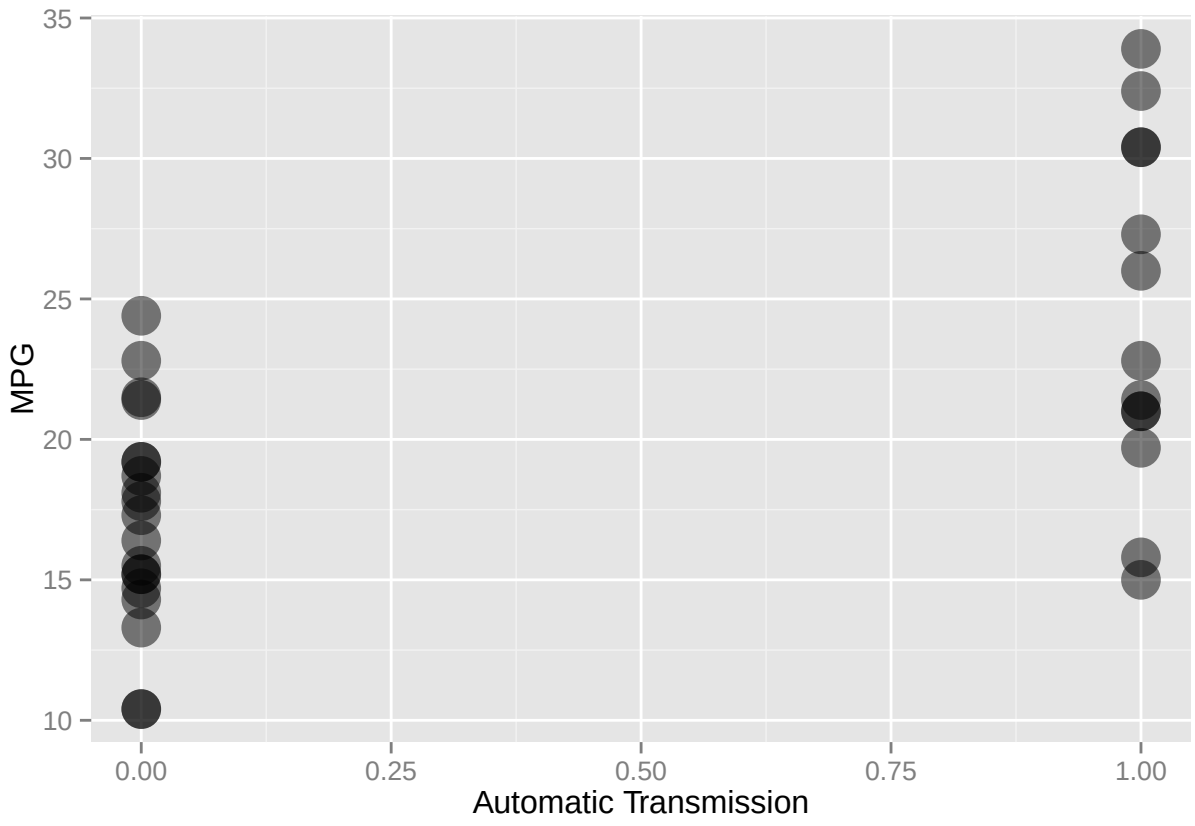
```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

From the coeffients stated above, the intercept states when the car has a manual transmission. In our model: Model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$.

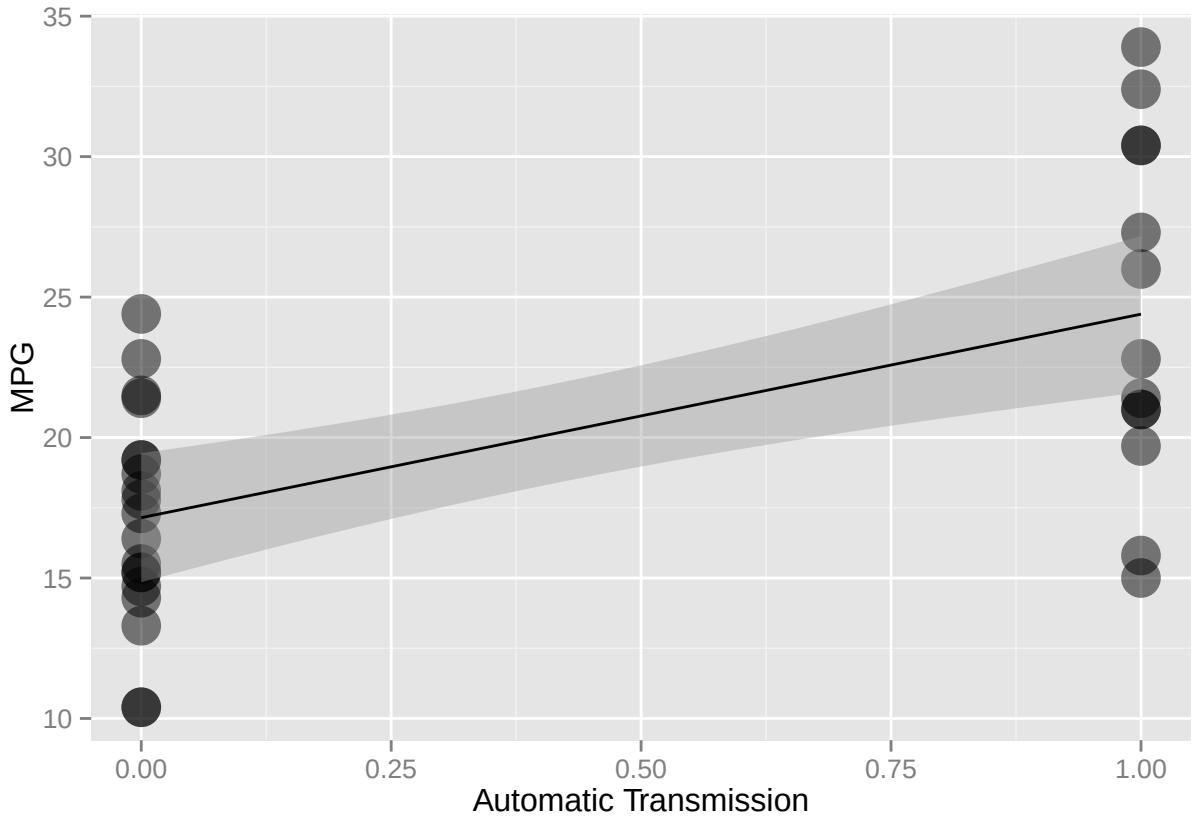
The β_0 because the variable in this equation is 0, indicating that the car is NOT equipped with an automatic transmission. Hence, the mean MPG for a car with a manual transmission is 17.147. The am variable, with an estimated coefficient of 7.245 says that a car equipped with an automatic transmission has a mean that is 17.147+7.245 or roughly 24.4 MPG.

Take a look at the plot below. The principles of the intercept being hte mean of the mpg's of cars without automatic transmissions and the mean of the mpg's of cars with automatic transmisions can be shown in the graph easily because the end points of the line are inthe middle of the two data sets.

```
library(ggplot2)
data(mtcars)
#R Code is provided here to show work
g<-ggplot(mtcars, aes(x=am, y=mpg))
 g<-g + xlab("Automatic Transmission")
 g<-g + ylab("MPG")
 g<-g + geom_point(size=7,color="black",alpha=0.5)
 g
```




```
g <- g + geom_smooth(method="lm", color="black")
g
```



We have confidence in the predictiveness of these variables because their p-values are smaller than 5%, and hence we have 95% confidence that we are not making a Type II error. The model also allows us to estimate a confidence interval of the true mean of the automatic transmission being between the following two numbers.

```
## [1] 20.864 27.920
```

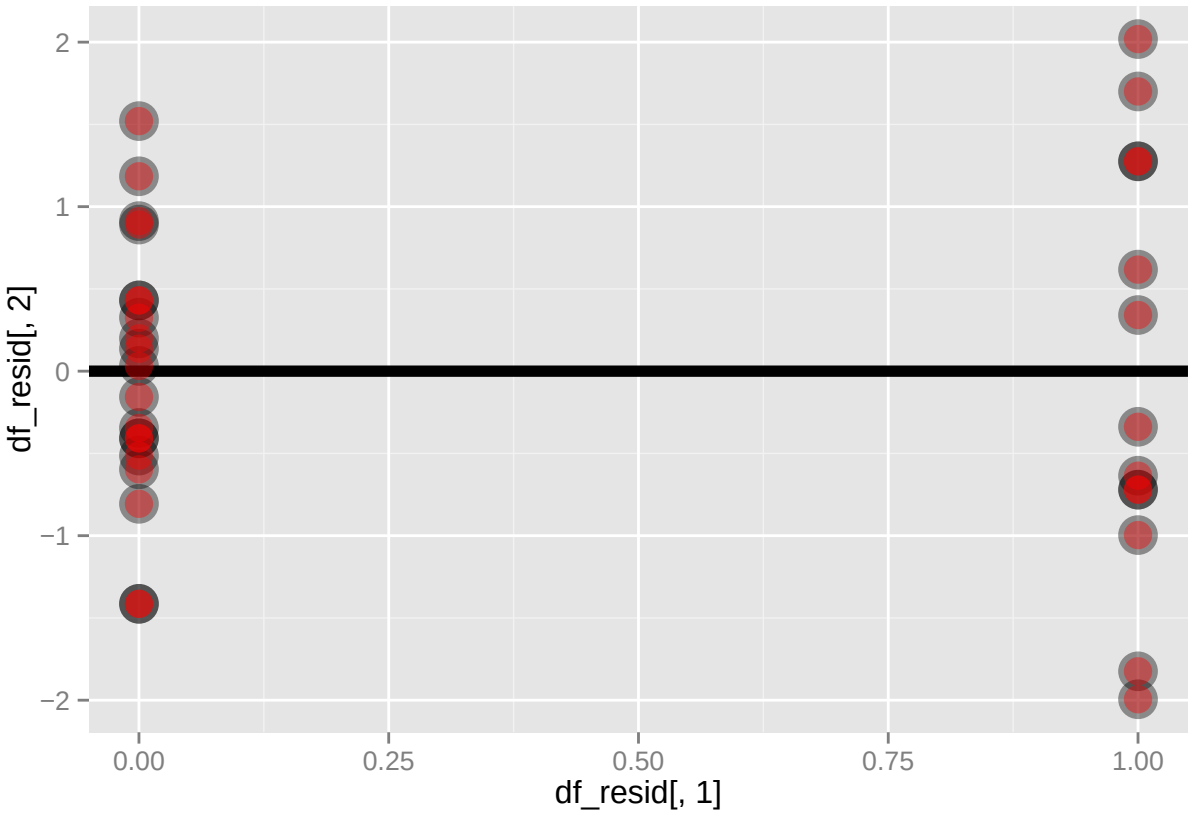
However, the model has a relatively poor R-square value of .3598. Given that there are several other factors that impact a car besides it's transmission it will probably be good to add another variable to increase confidence in the model.

Residual Analysis

```
library(broom)
df_resid<-cbind(augment(fit)$am, augment(fit)$.std.resid)


colnames(df_resid) <- c("am","residual")

rg<-ggplot(data.frame(x=df_resid[,1], y = df_resid[,2]), aes(x=df_resid[,1], y=df_resid[,2]))
 rg<- rg + geom_hline(yintercept=0, size=2)
 rg <- rg + geom_point(size = 7, colour = "black", alpha = 0.4)
 rg <- rg + geom_point(size = 5, colour = "red", alpha = 0.4)
 rg
```



Looking at the resideual above, there is a clear pattern: some residuals are associated with automatic transmissions, and others are not. It is worth noting the relationship between the p-value (test statistics) of the different transmisson types and the residuals. The residuals of the manual transmissions (the intercept) are much more dense compared to the automatic transmissions in the graph above. The p-value for the intercept is much smaller than the other variable, hence more variation is explained in the β_0 than in the β_1.