Understanding mpg in terms of transmission type

C. Davis Buenger

August 21, 2016

Executive Summary

The following analyzes the factor of transmission type (automatic vs. manual) as a predictor of mpg on the mtcars data set. Initial analysis indicated that cars with manual transmission would have an 7.245 mpg greater than cars with automatic transmission and this result was significant up to a p value of 0.000285. However, the naive analysis biased our results by omitting several significant variables. When the number of cylinders was added as a predictor, the effect of transitioning from automatic to manual only increased the mpg by 2.56 with a statistical significance of 0.0584 (Not statistically significant). Furthermore, when weight was included as a predictor the effect of tranmition type disappeared completely. In this case, transitioning from a automatic to a manual actually decreased the mpg by 0.0236 with a statistical significance of 0.988. Thus transmission type does not produce significant effect on the mpg of a car.

Initial Analysis

Let us take a quick look shape of the mpg values of the cars with an automatic transmission and of the cars with a manual transmission. First let us calculate the mean mpg of each factor.

```
mtcars$am <- factor(mtcars$am,levels=c(0,1),labels=c("Automatic","Manual"))
mean(subset(mtcars,am=="Automatic")$mpg)
```

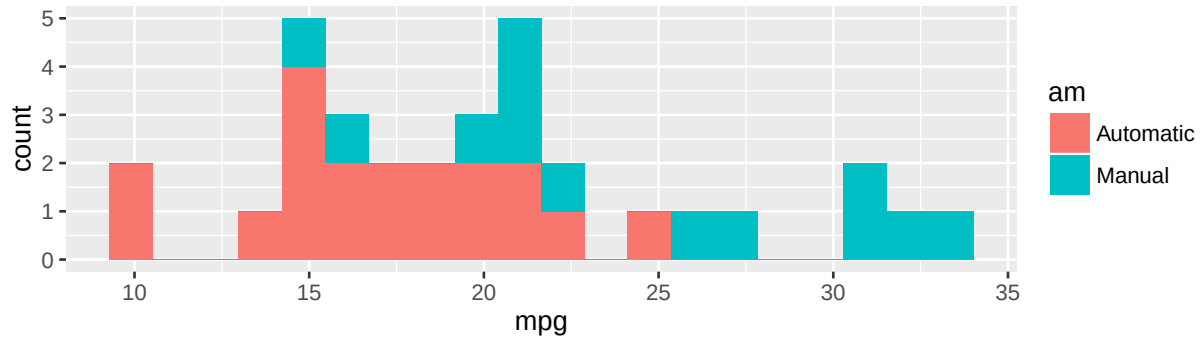
```
## [1] 17.14737
```

```
mean(subset(mtcars,am=="Manual")$mpg)
```

```
## [1] 24.39231
```

As you can see the mean of the cars with an automatic transmission perform significantly better than those with manual transmission. Next let us look at some histograms.

```
library(ggplot2)
qplot(mpg,geom="histogram",data=mtcars, bins=20, fill=am)
```



As evident by the graph above, the distribution of the orange automatic cars has lower mpg than the distribution of the turquoise manual cars.

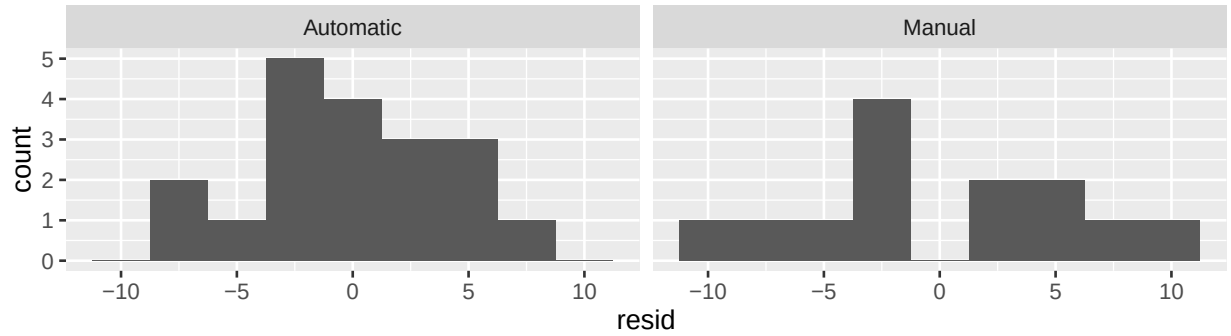
Let's create a linear model of mpg verses transmission type to assess the significance of the relationship. Note that since transmission type is a factor variable with two categories, by above we know that the linear model will have intercept 17.14737 with slope 7.24494. We perform a linear model to test the statistical significances of this relation and to test for outliers.

```
fit1<-lm(mpg~am, data=mtcars)
summary(fit1)
```

```
##
## 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 ***
## amManual       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
```

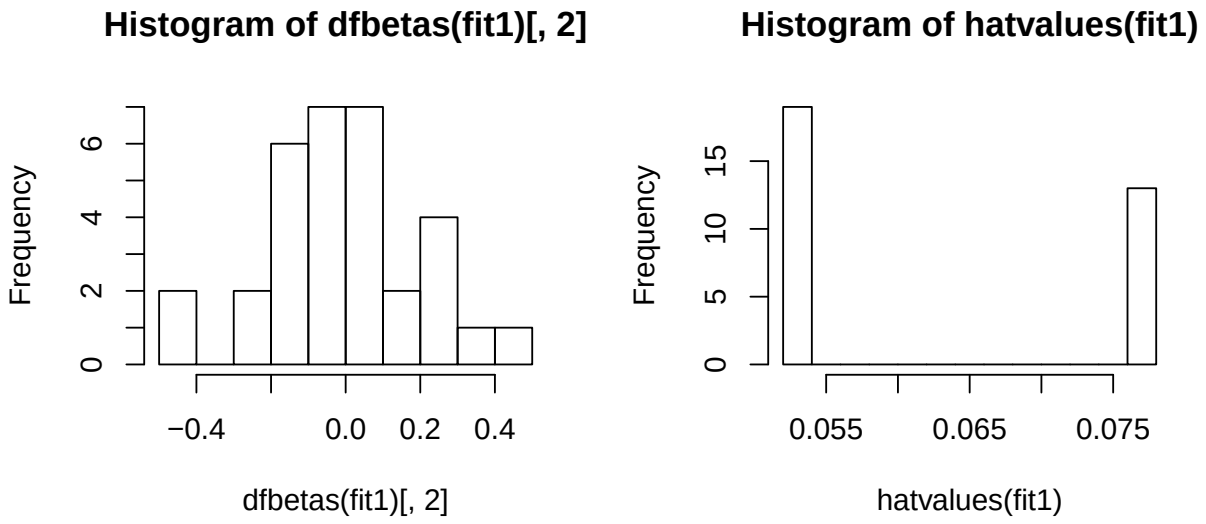
As seen above both the intercept and slope are statically significant at the 0.001 level. Now let us investigate the residuals of the fit and check their normality.

```
residualsFit<-data.frame(residuals(fit1),mtcars$am)
names(residualsFit)=c("resid","am")
qplot(resid,data=residualsFit, geom="histogram", facets= .~am,binwidth=2.5)
```



Observe that the residuals for each factor follow a "relatively" normal distribution. Additionally, the dfbetas values and hat values reveal no obvious outliers and no significantly large values.

```
par(mfrow=c(1,2))
hist(dfbetas(fit1)[,2],10)
hist(hatvalues(fit1),10)
```



These results might lead one to believe that changing a car from a manual transmission to an automatic transmission will increase the mpg by 7.245 mpg with a p value of 0.000285. However, we have not yet accounted for the potential bias due to not including sufficiently many predicting variables. We check for this next.

Investigating Confounding Variables

The mtcars data set includes 9 additional variables to transmission type and mpg. They are number of cylinders, displacement (cu.in.), gross horsepower, rear axle ratio, weight (1000 lbs), 1/4 mile time, V/S, number of forward gears, and number of carburetors. Of these, we will compare the effect of weight and cylinder number on mpg to the effect of transmission type on mpg.

First observe the following table.

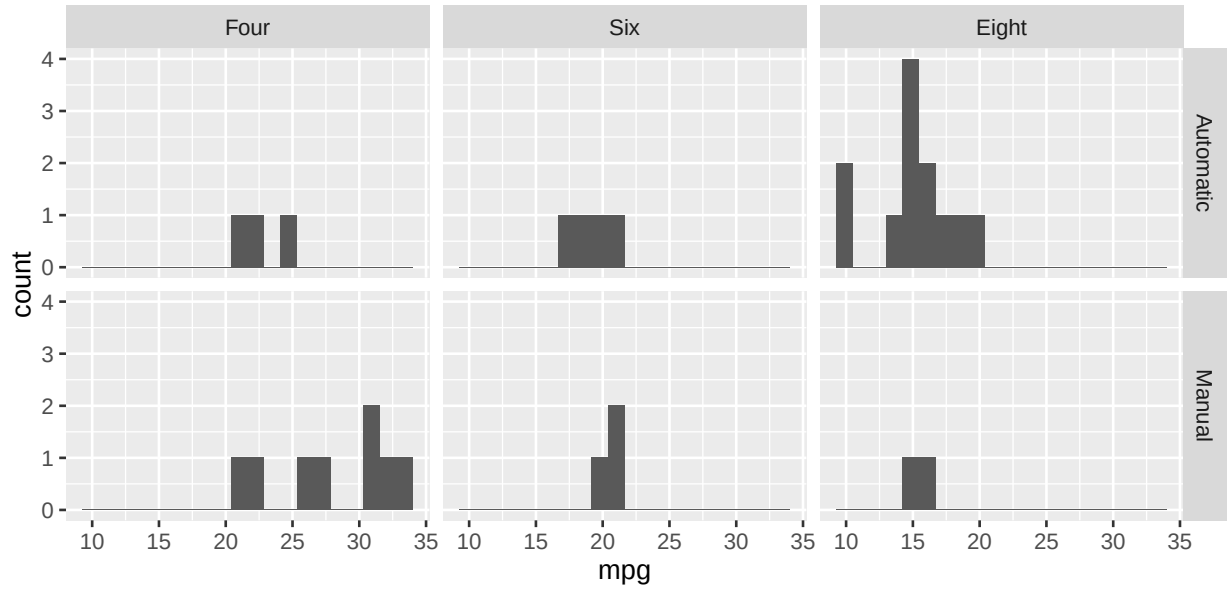
```
 mtcars$cyl <- factor(mtcars$cyl,levels=c(4,6,8),labels=c("Four","Six","Eight"))
table(mtcars$cyl,mtcars$am)
```

```
##
##         Automatic Manual
##   Four          3      8
##   Six           4      3
##   Eight        12      2
```

As you can see, the majority manual cars have 4 cylinders and the majority of automatic cars are 8 cylinders. Since mpg falls as the number of cylinders rises, the results above may just be due to the fact that more of the manual cars have four cylinders.

Comparing the mpg of automatic cars and manual cars restricted to each cylinder category reveals that there may be some improvement in four cylinder cars, but there is no apparent improvement for cars with six or eight cylinders.

```
qplot(mpg,geom="histogram",data=mtcars, bins=20, facets =am~cyl)
```



After performing a linear regression of mpg against cylinders and transmission type we observe that keeping cylinders fixed, a transition from an automatic transition to a manual one would produce a 2.56 mpg improvement. However, the p-value of this analysis is only 0.058457 and thus the result is not statistically significant.

Now, let us examine the effect of weight and transmission type on mpg. Observe the average weight of the automatic cars is 3769 lbs and the mean of the manual cars is 2411 lbs.

```
mean(subset(mtcars,am=="Automatic")$wt)
```

```
## [1] 3.768895
```

```
mean(subset(mtcars,am=="Manual")$wt)
```

```
## [1] 2.411
```

Again, it is reasonable to believe that our initial results which showed that manual cars had better gas mileage occurred because omission of the weight variable biased our analysis.

Fitting a linear regression with mpg as a function of transmission type and weight yields the following.

```
fit3<-lm(mpg~am+wt,data=mtcars)
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.32155    3.05464  12.218 5.84e-13 ***
## amManual    -0.02362    1.54565  -0.015    0.988
## wt          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

This linear regression predicts that changing a car from an automatic transmission to a manual on will decrease mpg by 0.02362 with a p-value of 0.988. Thus converting from an automatic transmission to a manual one does not have a significant effect on mpg.

Conclusions

We engaged in this study to determine if automatic transmissions or manual transmissions were better for mpg. Initial results indicated that perhaps manual cars had significantly higher mpg. Further analysis, however, indicated that after including weight or cylinder number in the linear regression, the effect of transmission type had no significant effect on mpg.