Fuel Efficiency of Automatic vs. Manual Transmission

Andrew Reynolds

September 25, 2015

Executive Summary

We use the `mtcars` dataset from the R datasets package to address the question, "Is an automatic or manual transmission better for MPG?" and we quantify the difference. We begin by reading the data, with no processing necessary. Next we do some exploratory analysis to give a preliminary answer to the posed question. Then we perform linear modeling and adjust for likely confounders. We find that vehicles with manual transmission get significantly better fuel efficiency than automatics. We show that this difference may simply be due to the facts that manual transmission vehicles tend to be lighter than automatics, and lighter vehicles get better fuel efficiency.

Loading the data and exploratory analysis

We load the `mtcars` dataset. No preprocessing is necessary.

```
data(mtcars)
```

To begin we compare the MPG with transmission. According to the "MPG by transmission" plot (**see Appendix p.3**), there seems to be a large difference in MPG between automatic and manual transmissions. The manual transmission vehicles get better fuel efficiency.

However, one or more of the other variables in the dataset could be confounding our results. Likely confounders are variables correlated with the transmission variable `am`, so we check which variables other than `mpg` (the first variable) are most closely correlated with `am` (the ninth variable).

```
cor(mtcars[-1])[9 - 1,]
```

```
##         cyl        disp          hp        drat          wt        qsec
## -0.52260705 -0.59122704 -0.24320426  0.71271113 -0.69249526 -0.22986086
##          vs          am        gear        carb
##  0.16834512  1.00000000  0.79405876  0.05753435
```

The variables most closely correlated with `am` are `disp`, `drat`, `wt`, and `gear`. These variables are all correlated with each other (we omit the verification of this fact), so we may pick a representative from among them and adjust for it in the analysis that follows. We pick `wt` because it's easy for a person to judge visually whether `wt` is high or low for a particular vehicle, which they may be considering purchasing for example. The other variables would work about as well, and would be useful if the hypothetical car buyer has access to information such as the displacement of the engine, etc.

Modeling

First we consider a line model with `mpg` as outcome and `am` as regressor.

```
simplefit <- lm(mpg~am, data = mtcars)
summary(simplefit)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

The `am` slope coefficient (7.245) is the expected change in `mpg` if one switches from an automatic to manual transmission. The p-value for both the intercept and the slope are very small, indicating that transmission is a significant indicator of MPG. Now we fit a linear model for `mpg` including `am` and `wt` as explanatory variables.

```r
fit <- lm(mpg ~ am + wt, data = mtcars)
summary(fit)$coef
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

Note the new `am` slope coefficient is small, and its p-value is very high, while the other p-values are very low. This shows that `am` is unnecessary in the linear model if `wt` is included. The data fail to give significant evidence that transmission makes a difference in MPG when considering vehicles of the same weight. This is also evident from the "MPG vs weight by transmission" plot (**see Appendix p.3**). We can clearly see the overall downward trend in the plot from left to right, which tells us that heavier cars get worse fuel efficiency. This trend exists within the two transmission types.

Another important feature of the "MPG vs weight by transmission" plot is that it shows that manual transmissions tend to be lighter than automatics. There aren't many instances of the two transmission types having similar weight to compare directly. And as shown with our linear model earlier, the difference in mpg for fixed weight is insignificant.

Diagnostics

Now we check for potential problems with our model. The hatvalues and the dfbetas are found in the panel plot on page 4 in the Appendix. There are three points with slightly elevated hatvalues, but it's not extreme enough to be worrisome. The dfbetas are all quite small, with one exception. It is not a problem, though, since the dfbetas for our variable of interest (`am`) barely exceed 0.4.

A few standard diagnostic plots are found in the panel plot on page 5 in the Appendix. There does seem to be a slight systematic trend in the "Residuals vs Fitted" plot, but much less so when standardized residuals are used. There little departure from normality, judging from the "Normal Q-Q" plot. The "Residuals vs Leverage" plot does show a few high-leverage points with moderate residuals. These are very few, and not too extreme. Overall it appears our fitted model is reliable.

Which transmission type gets better fuel efficiency?

The answer to our posed question must be that manual transmissions get better fuel efficiency, as we discovered initially. A 95% confidence interval for the difference in `mpg` is

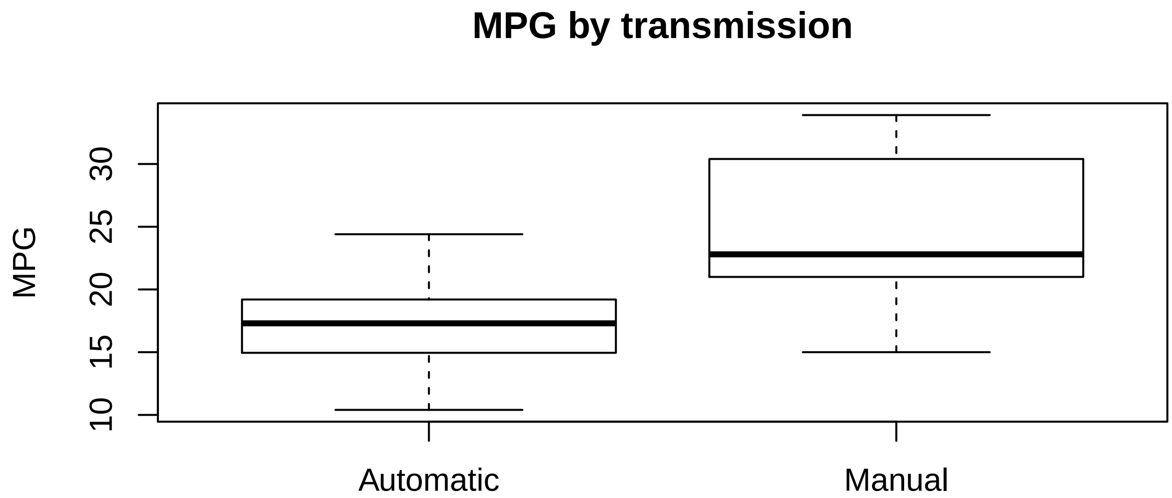
```r
confint(simplefit)[2,]
```

```
##    2.5 %   97.5 %
##  3.64151 10.84837
```

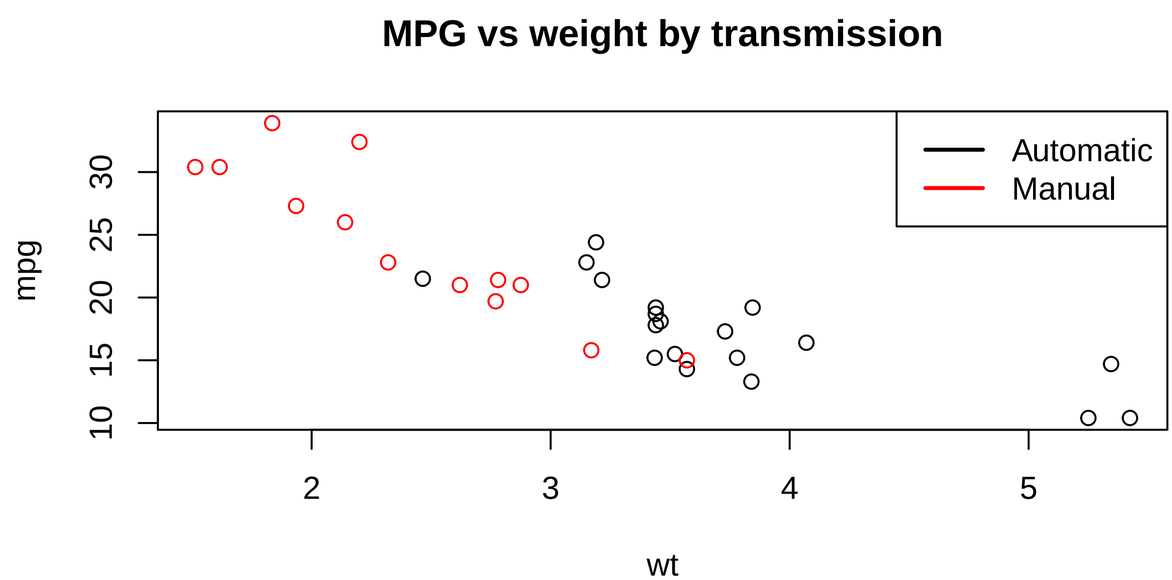
However, our analysis revealed that this difference may be due to the facts that lighter vehicles get better fuel efficiency and that manual transmission vehicles tend to be lighter than automatics. Adjusting for weight renders transmission type insignificant.

Appendix - Figures

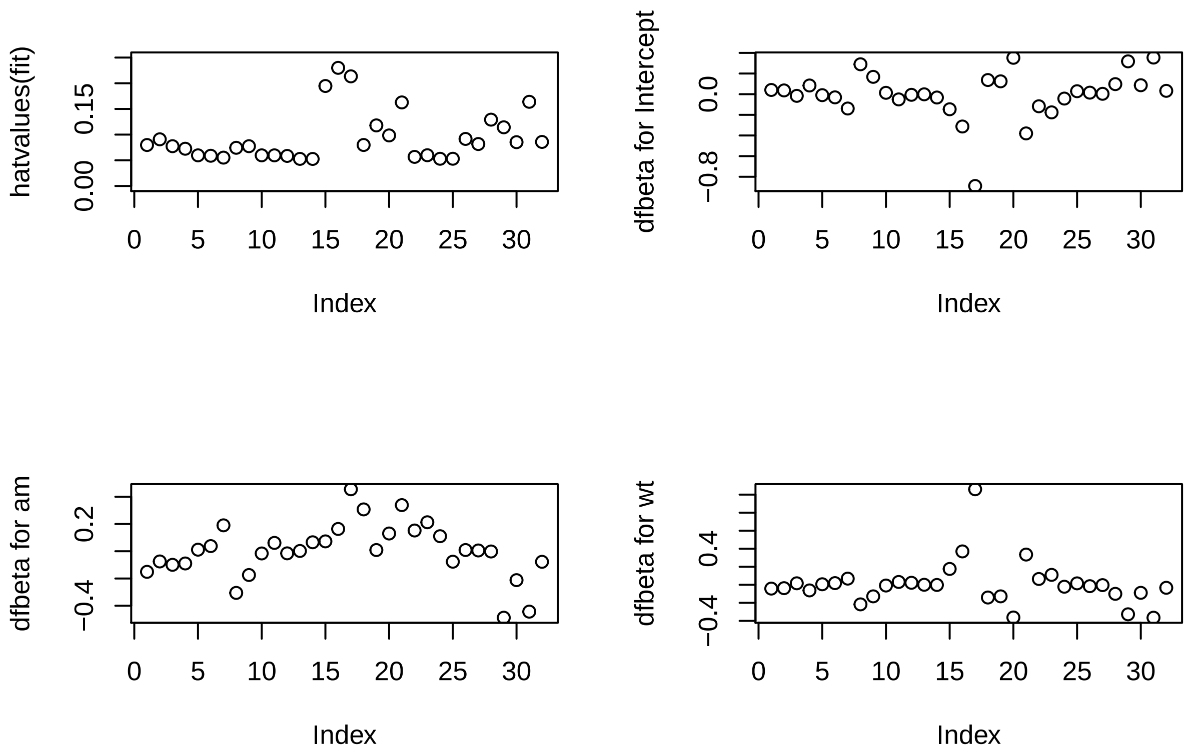
```
boxplot(mtcars$mpg ~ mtcars$am, main = "MPG by transmission", ylab = "MPG", xaxt = "n")
axis(side = 1, at = c(1,2), labels = c("Automatic", "Manual"))
```




```
plot(mpg ~ wt, col = as.factor(am), data = mtcars,
       main = "MPG vs weight by transmission")
legend("topright", c("Automatic", "Manual"), lty=c(1,1),
       lwd = 2, col = c("black", "red"))
```



```r
par(mfrow = c(2, 2))
plot(hatvalues(fit), ylim = c(0, .25))
plot(dfbetas(fit)[,1], ylab = "dfbeta for Intercept")
plot(dfbetas(fit)[,2], ylab = "dfbeta for am")
plot(dfbetas(fit)[,3], ylab = "dfbeta for wt")
```



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