How Does Transmission Type Effect Milage?

Philip Bulsink

2017-02-19

Executive Summary

This report will outline the effect of transmission type (automatic vs. manual) on gas milage. The influence of other variables will be investigated as well as confounding factors. It is shown that, with all other factors taken into consideration, no milage improvement is seen for manual cars.

Introduction

The data to be studied is available in `R` in the `mtcars` frame. This data contains information on 32 automobiles from 1973-1974, listing a total of 10 variables on each car. This is easily loaded into `R` with one command.

```
data("mtcars")
```

Data Preparation

A few of the variables we will be investigating in this report can be considered 'factors', that is, of a non-numeric or a counted type. This includes transmission type, and may include number of gears, number of carburetors, number of cylinders, or if the engine is straight or V shaped.

```
mtcars$am<-factor(mtcars$am, labels = c("Automatic", "Manual"))
mtcars$vs<-factor(mtcars$vs, labels = c("V", "Straight"))
mtcars$cyl<-as.factor(mtcars$cyl); mtcars$carb<-as.factor(mtcars$carb); mtcars$gear<-as.factor(mtcars$ge
```

Data Exploration

A first glance at the data suggests that there is an effect on milage (mpg) by transimission type (am), whereby manual transmissions have better (higher) gas milage. See a plot of this data in Appendix 1.

We can quantify this by regressing the data over this factor:

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

From this, we can see that the average milage of cars with a manual transmission is approximately 7.24 better than that of an automatic transmission. The p value for this estimate is very low, at 2.85×10^{-4}, allowing us to reject the null hypothesis that there is no difference.

Detailed Data Analysis

What if the enhancement of milage in manual transmissions is due to some other factor? For example, in Appendix 2, we show the milage as a function of transmission type, but broken up by cylinder.

```
fit2 <- lm(mpg~am + cyl, data=mtcars)
```

If we include cylinders in our model, we then find that cars with a manual transmission is approximately 2.56 better than that of an automatic transmission. The benefit of a manual transmission is much reduced when considering cylinder count.

Each of the possible models from this data is built, followed by an ANOVA analysis to identify which factors are significant.

```r
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 + wt
## Model  4: mpg ~ am + cyl + wt + gear
## Model  5: mpg ~ am + cyl + wt + gear + hp
## Model  6: mpg ~ am + cyl + wt + gear + hp + carb
## Model  7: mpg ~ am + cyl + wt + gear + hp + carb + disp
## Model  8: mpg ~ am + cyl + wt + gear + hp + carb + disp + drat
## Model  9: mpg ~ am + cyl + wt + gear + hp + carb + disp + drat + vs
## Model 10: mpg ~ am + cyl + wt + gear + hp + carb + disp + drat + vs + qsec
##    Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1      30 721
## 2      28 264  2       456 28.43 7.9e-06 ***
## 3      27 183  1        82 10.16  0.0061 **
## 4      25 175  2         8  0.51  0.6106
## 5      24 150  1        25  3.13  0.0972 .
## 6      19 140  5        10  0.25  0.9329
## 7      18 128  1        12  1.46  0.2456
## 8      17 126  1         2  0.26  0.6165
## 9      16 122  1         4  0.52  0.4839
## 10     15 120  1         1  0.15  0.6997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that only the numbers of cylinders, and the weight affect the model in a statistically significant way, beyond the automatic/manual correllation. The rest of the variables, while reducing model variance, are not significant predictors.

Final Model

With this knowledge, we build the final model:

```r
fitfinal<-lm(mpg~am+cyl+wt, data=mtcars)
```

Thus, we can see that, with appropriate factors taken into consideration, the average manual transmission has a mpg of only 0.15 higher than that of an automatic transmission (at 33.75). However, the standard error of this estimated improvement is 1.3, which would include 0. Thus, a 0 increase in milage is within the error of the analysis.

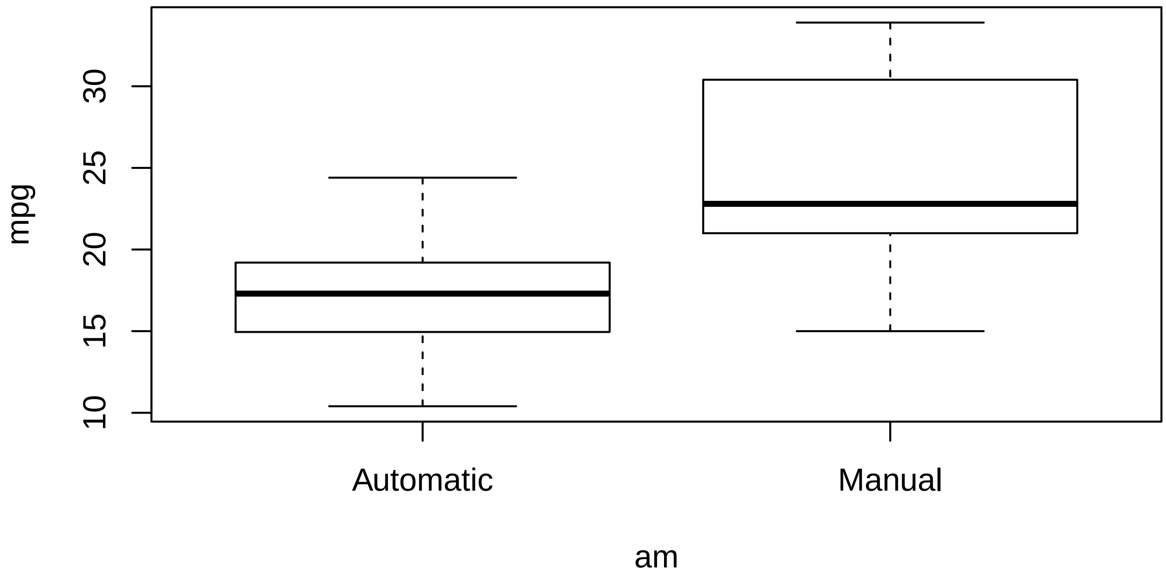
Analysis of the residuals (see Appendix 3) shows that there is no obvious trend, and the QQ plot appears approprate. The Shapiro-Wilk normality test returns p=0.26, which proves our residual normality, validating our ANOVA analysis.

Conclusion:

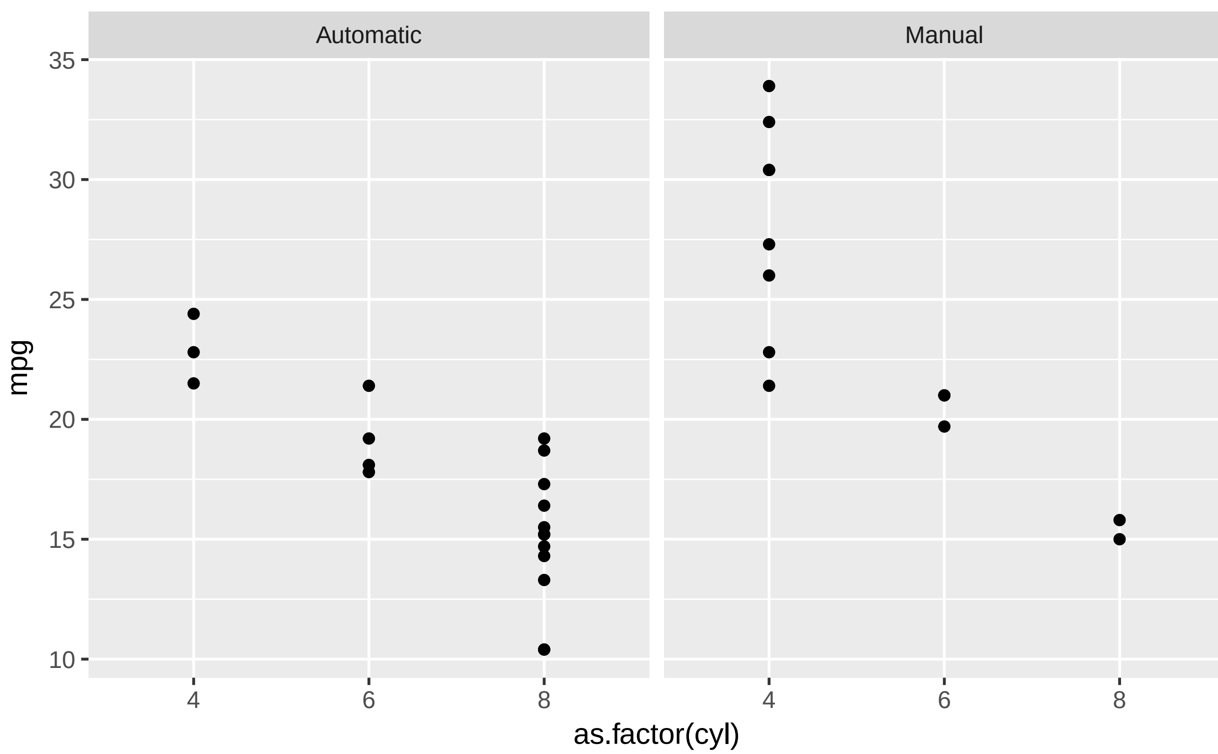
While it appeared at first that manual transmission cars had better gas milage, once additional relavant variables were considered, the null hypothesis cannot be rejected, that is, there's no statistical evidence for this argument. Manual transmissions are not responsible for increased gas milage.

Appendices

Appendix 1: Plot of milage vs. transmission type:



Appendix 2: Plot of milage vs. cylinder count and transmission type:



Appendix 3: Residuals Diagnostic Plots