Impact of transmission type on gas mileage was inconclusive in the Motor Trend data set

Tim Wise

December 2015

### Executive Summary

In this report, we looked at Motor Trend cars data set (mtcars) to see if we could determine whether the type of transmission in a car, manual or automatic, had a significant effect on its gas mileage, and if so, to quantify that impact.

The mtcars data set includes descriptor variables like, weight, number of cylinders, displacement, transmission type, plus measured values like mileage and quarter mile time.

### Analysis Details

First, let's compare the mileage of cars with automatic transmissions to those with manual transmissions. **Figure 1** shows a boxplot of the mileage for each transmission type. It shows:  
- there is a sizeable difference in the median, with manual transmissions getting better gas mileage (22.8 MpG) than automatics (17.3 MpG).

A simple regression of mileage as a function of transmission type (mpg ~ am), see **Figure 2**, shows:  
- the difference in average mileage is significant (the amManual p-value of 2.8510^{-4} is less than the 95-th percentile threshold of 0.05)  
- manual transmissions get 3.64 to 10.8 MpG *better* than the average 17.1 MpG of automatic transmissions (the 95% confidence interval)  
- this fit accounts for just 33.8% percent of the variation in mileage (the adjusted R-square value)

BUT, looking at the residuals of this fit in **Figure 3**, we see that this fit is not good at all. The residuals should be random, yet we see a significant pattern that looks like parallel lines. This indicates the residuals contain information from other variables that should be included in our model.

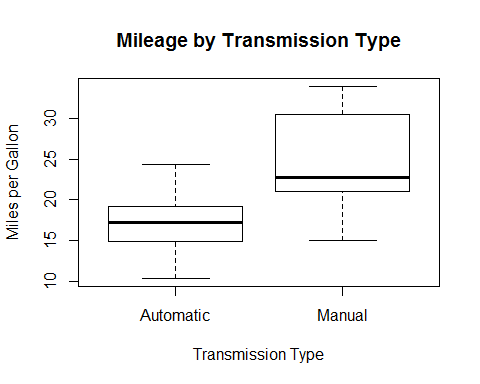
Let's investigate further, looking at what variables affect mileage the most. **Figure 4** shows the correlation between variables in the data set. In the first row, we see that mileage (mpg) is affected the most by weight (wt) with a correlation coefficient of -0.87. Let's explore the relationship between mileage and weight.

**Figure 5** plot the mileage verses weight for both transmissions types.

### Appendix

This section contains the figures referenced in the Analysis Details.

boxplot(mpg ~ am, data=D,   
 main="Mileage by Transmission Type",  
 xlab="Transmission Type",  
 ylab="Miles per Gallon")



**Figure 1**: Automobiles with manual transmissions *appear* to get better mileage than those with automatic transmissions

fit.mpg.vs.am <- lm(mpg ~ am, data=D)

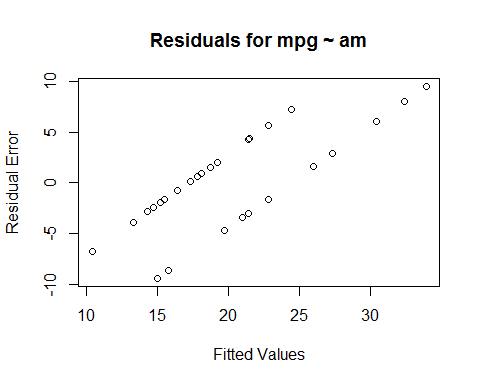
summary(fit.mpg.vs.am); round(confint(fit.mpg.vs.am), 2)

##   
## Call:  
## lm(formula = mpg ~ am, data = D)  
##   
## 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

## 2.5 % 97.5 %  
## (Intercept) 14.85 19.44  
## amManual 3.64 10.85

**Figure 2**: A regression fit of mpg ~ am regression indicates the greater mileage of manual transmissions (amManual) is statistically significant and ranges from 3.64 to 10.8 MpG better than automatics, but ...

plot(x=D$mpg, y=resid(fit.mpg.vs.am),   
 main="Residuals for mpg ~ am",   
 xlab="Fitted Values",  
 ylab="Residual Error")



**Figure 3**: The residuals of mpg ~ am are not random, show a significant pattern, and tells us this is a bad fit

corr.matrix <- round(cor(mtcars), 2)  
corr.matrix[lower.tri(corr.matrix,)] <- ""

as.data.frame(corr.matrix)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## mpg 1 -0.85 -0.85 -0.78 0.68 -0.87 0.42 0.66 0.6 0.48 -0.55  
## cyl 1 0.9 0.83 -0.7 0.78 -0.59 -0.81 -0.52 -0.49 0.53  
## disp 1 0.79 -0.71 0.89 -0.43 -0.71 -0.59 -0.56 0.39  
## hp 1 -0.45 0.66 -0.71 -0.72 -0.24 -0.13 0.75  
## drat 1 -0.71 0.09 0.44 0.71 0.7 -0.09  
## wt 1 -0.17 -0.55 -0.69 -0.58 0.43  
## qsec 1 0.74 -0.23 -0.21 -0.66  
## vs 1 0.17 0.21 -0.57  
## am 1 0.79 0.06  
## gear 1 0.27  
## carb 1

**Figure 4**: Mileage, first row, is highly correlated (|| > 0.80) to weight, number of cylinders, displacment, and horsepower. And those variables are highly correlated amongst themselves.

### Errata