Motor Trend Magazine mpg Transmission Analysis

Ann Crawford

October 2, 2017

Executive Summary

This paper uses regression models and exploratory data analysis on the mtcars dataset to examine transmision type (mannual or automatic) as a predictor for miles per gallon. Specifically address these questions:
1. Is an automatic or manual transmission better for MPG? In this case better means higher value for miles per gallon. 2. Quantify the MPG difference between automatic and manual transmissions.

Steps to Analysis

1. Examine the DataSet
2. Analyze a simple linear regression mpg to transmission: automatic vs manual
3. Review other perdictors using multivariate analysis to quantify transmission as a predictor.

Conclusion

The best fit model for predictor mpg uses an interation of weight and am. Transmission and weight are correlated. Including transmission and wt in regression analysis explains roughly 83% of the variance in mpg with a good residual plot. It does not make sense to discuss am as a predictor of mpg without considering the interaction of wt. ### Examine the DataSet The mtcars data frame contains 32 observations on 11 variables. vs,am, gear, carb and cyl can be treated as factors. For details consult help **?mtcars**

```
y <- mtcars$mpg
x <- mtcars$am
# check for nas
sapply(mtcars,function(x) sum(is.na(x)))

##  mpg  cyl disp   hp drat   wt qsec   vs   am gear carb
##    0    0    0    0    0    0    0    0    0    0    0

sapply(mtcars, function(x) length(unique(x)))

##  mpg  cyl disp   hp drat   wt qsec   vs   am gear carb
##   25    3   27   22   22   29   30    2    2    3    6
```

Analyze a simple linear regression mpg to transmission: automatic vs manual

Since question 1 specifcally asks to compare mpg to tranmission, the box plot shows that mannual transmissions have higher mpg in general with some variance. To further explore the variance and the value of am as a predictor, start with simple linear regression using transmisison as the single perdictor for mpg and review the residuals. fit1 <- lm(mpg ~ am , data= mtcars) e <- resid(fit1) The model fit returns a pvalue of 2.850207410^{-4} indicating statical significance at the 95% level. R squared = 0.3597989 indicating transmission accounts for roughly 36% of the variablity in mpg. ### Residual Plot shows errors with the model design The residual plot shows, the relationship is **not a simple linear regression**. The pattern in residual plot reveals errors with the model design. Residuals represent variation left unexplained by our model. Points should be radomly clustered around the 0. The fitted plot show that the standard error (shadded portion) tends to increase as am and mpg increase.

Multvariate Analysis

This section use multi-variate linear regression to explore how the other variables in the mtcars dataset may be related to mpg.

fit2 <- lm(mpg ~ ., data = mtcars) First fit a model with all the variables. But the results contradict the initial findings, where the p values show no statisical significance at the 95% level, all greater than 0.05. The pvalue for am is now `r summary(fit2)$coefficients[8,4]'. Wt has the lowest p value. This indicates and interation between the variables.

Checking the correlation of perdictors

The correction matrix, cor(mtcars[,1:11]), confirms correlation between the mtcars varaibles. It seems that wt is a strong predictor for the mpg response variable.

Ploting weight mpg and color coding the transmission shows a strong inverse relationship between wt and am.

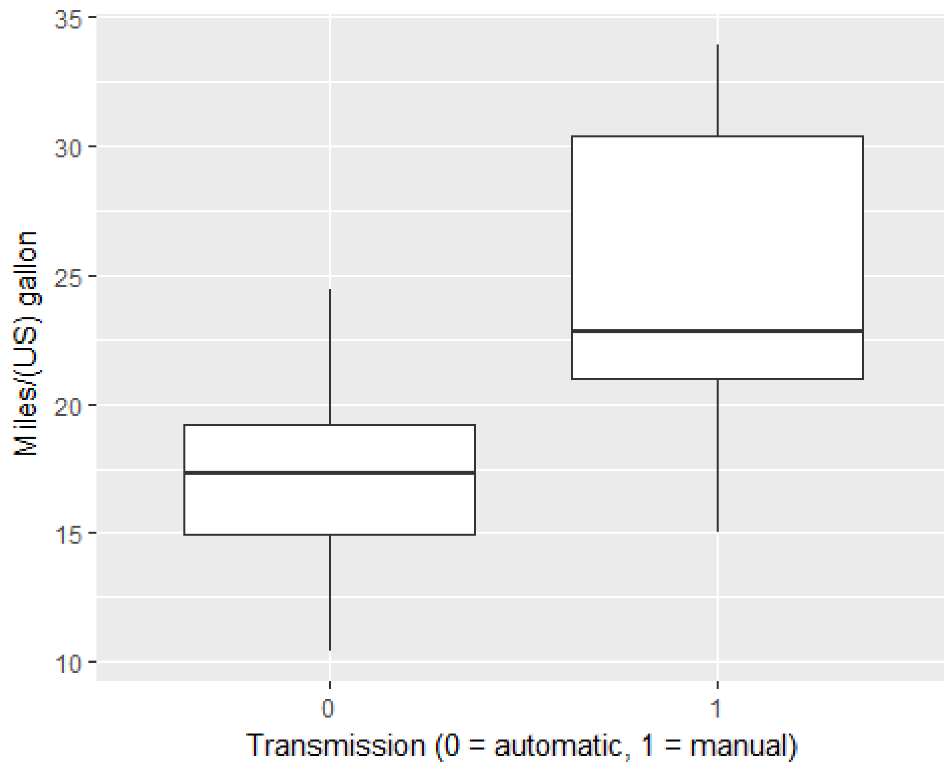
Multiple Linear Regression

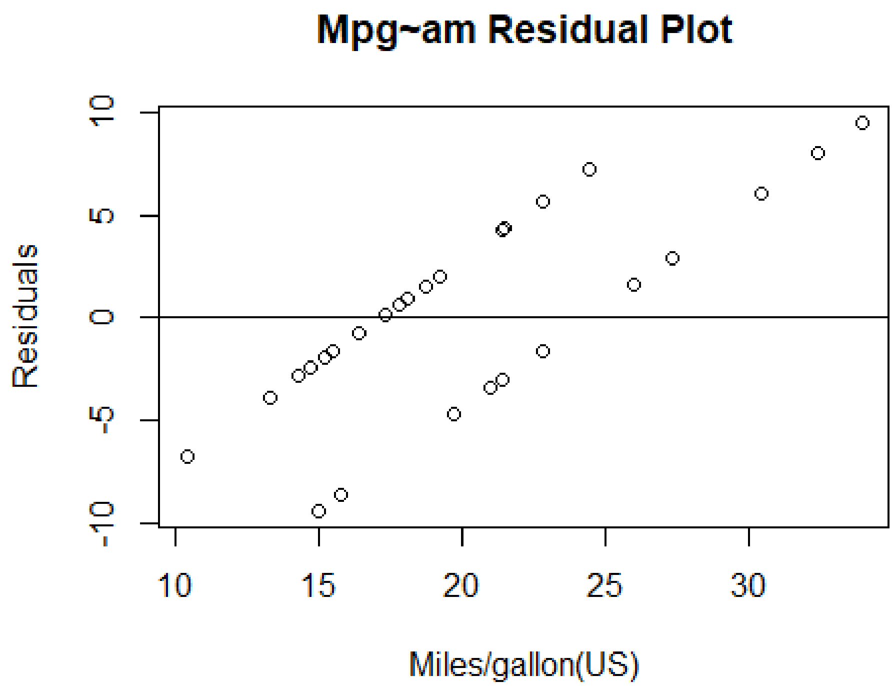
Weight as an interaction variable with transsmission is a better model for predicting mpg.

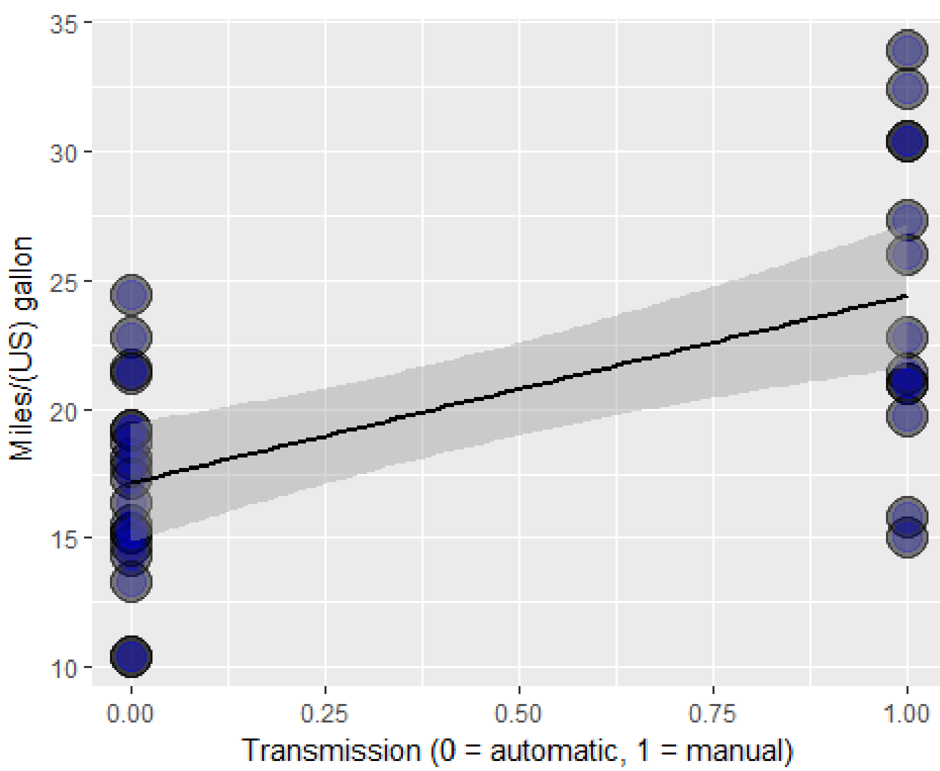
Weight is inversely related to mpg with a correclation between wt and mpg of -0.8676594.The correlation between am and wt is -0.6924953, fitting the mpg outcome with wt and am as interacting varables yields and good fit and a well behaved residual plot. returns a pvalue of 4.551182410^{-5} indicating statistical significance at the 95% level. R squared = 0.8330375 indicating transmission together with wt accounts for roughly 83% of the variablity in mpg.

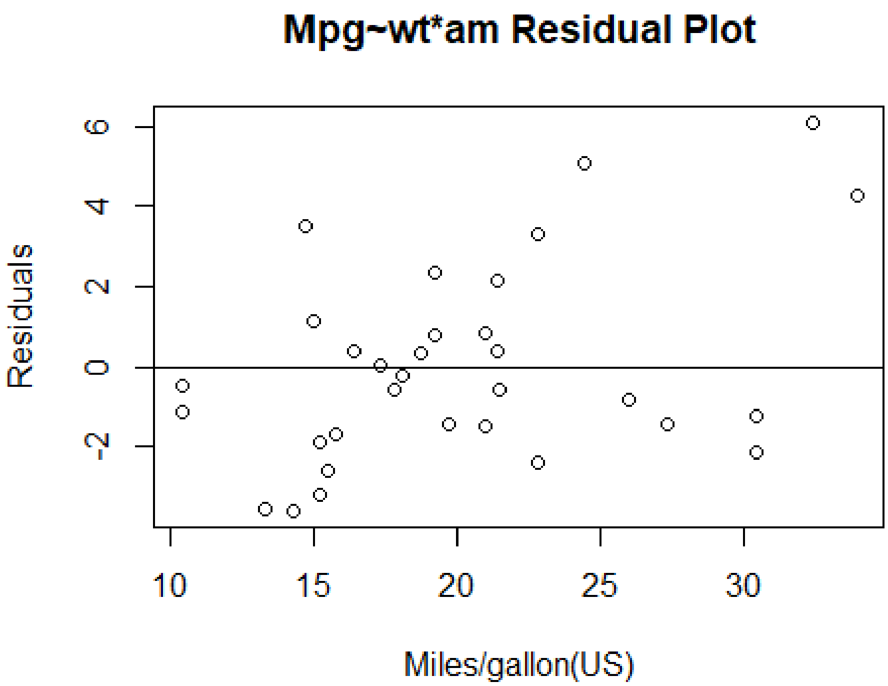
fitwtam <- lm(mpg~am*wt, mtcars)

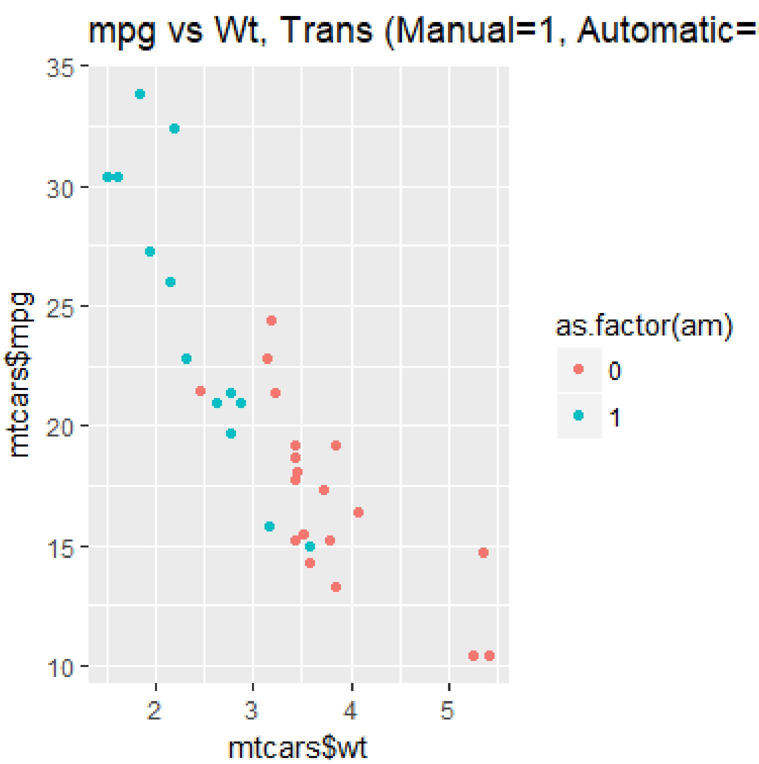
Appendix of Figures




Mpg~am Residual Plot
Residuals
Miles/gallon(US)
10
5
0
-5
-10
10
15
20
25
30


Miles/(US) gallon
Transmission (0 = automatic, 1 = manual)
35
30
25
20
15
10
0.00
0.25
0.50
0.75
1.00


Mpg~wt*am Residual Plot
Residuals
-2
0
2
4
6
10
15
20
25
30
Miles/gallon(US)


mpg vs Wt, Trans (Manual=1, Automatic=(
mtcars$mpg
10
15
20
25
30
35
2
3
4
5
mtcars$wt
as.factor(am)
0
1