Regression Models Course Project

Andrea

Thursday, June 18, 2015

Executive Summary

The objective of this exercise is to analyze a set of variables within the *mtcars* dataset in relationship to miles per gallon (MPG). There are two questions to answer:

1. Is an automatic or manual transmission better for MPG
2. Quantify the MPG difference between automatic and manual transmissions

However, in the grading section of the project, there are questions relating to evaluating other models. The analysis is to consider multiple models for greater impacts as well as check for residuals and diagnostics. An outline of strategy for what models should be selected is also included. ## Dataset Definition The dataset mtcars contains 32 readings across 11 variables. The 11 variables reported in the mtcars dataset are: mpg - Miles/(US) gallon; cyl - Number of cylinders; disp - Displacement (cu.in.);hp - Gross horsepower; drat - Rear axle ratio; wt - Weight (lb/1000); qsec - 1/4 mile time; vs - engine type is either V or straight; am - Transmission (0 = automatic, 1 = manual); gear - Number of forward gears; carb - Number of carburetors.

Of these values, horsepower, 1/4 mile time, as well as mpg are results of design, not design criteria themselves. Therefore, since we are only evaluating for mpg, horsepower and 1/4 mile time are being thrown out of the evaluations so that their bias can be removed. ##Data Analyses The first question is an evaluation of which transmission type, either automatic(0) or manual(1) is better for mpg. The boxplots show that the manual transmissions (am = 1) have the highest median MPG rating. Therefore, manual transmissions on average have the better MPG.

However, in looking at the simple linear model for the dataset, it is apparent that transmission type does not tell the entire story. While the 95 percent confidence interval does not include zero, so we can reject the null hypothesis that there is no association between mpg and transmission type, the Adjusted R^2 = 34% (.3385) so 66 percent of the variation in mpg remains unexplained. There are clearly factors other than transmission type which should help us understand the effect on mpg.

When executing the correlation on mtcars, transmission type (am) has much smaller relationship to mpg than other factors such as weight, displacement, or cylinders. The correlations show that the lower amount of weight, cylinders, and displacement we should expect an increase in the mpg. These other items will be looked at for additional modeling.

```
##        mpg        cyl       disp       drat         wt         vs
##  1.0000000 -0.8521620 -0.8475514  0.6811719 -0.8676594  0.6640389
##         am       gear       carb
##  0.5998324  0.4802848 -0.5509251
```

Model Selection / Analysis

Using ANOVA to check which of the models has the most statistical impact: Analysis of Variance (Table1):

Variables	ResDf	RSS	Df	SumOfSq	F	Pr(>F)	Num Star
am	30	720.9					None

Variables	ResDf	RSS	Df	SumOfSq	F	Pr(>F)	Num Star
am+wt	29	278.32	1	442.58	46.115	1.867e-07	3
am+wt+cyl	28	191.05	1	87.273	12.791	0.001292	2
am+wt+cyl+disp	27	188.43	1	2.6212	0.3756	0.5451	None

The ANOVA results show that by adding weight and cylinders we can see improvements in the model, however taking into account displacement wouldn't add much as the Pr(>F) factor is high. The summary of the linear model for each of those shows the difference of the adjusted R^2 value to change from .8079 (am + wt + cyl) to .8122 (am + wt + cyl + disp) with the inclusion of the displacement. Both are 81%, an improvement over transmission alone so we only have 19% of our points unexplained. So the most parsimonious model - with transmission included, irregardless that its impact is less than the others - is Weight + Cylinders.

However, the best baseline to model would have been by weight, not by transmission. The summary of that linear model has an Adjusted R^2 value of 75%, significantly more impressive than the 34% of the transmission type.

Figures 1-3 show the charted residuals and fitted values of Transmission only, versus Weight only versus Transmission + Weight + Cylinders. The improvement is seen below in Table 2 with the smaller variations from zero.

Model	Min	1Q	Median	3Q	Max
AM	-9.39	-3.09	-0.29	3.24	9.5
WT	-4.54	-2.36	-0.13	1.41	6.87
AM+WT+CYL	-4.17	-1.53	-0.53	1.58	6.08

Analysis of Automatic versus Manual Transmissions

However, because the questions had to do with analyzing the transmissions, we can see from the model looking at mpg versus am that the automatic (Intercept point) is average 17.147 mpg and Manual = 24.392 mpg which is = Intercept + Standard Error of the factor(am). Therefore the difference between Automatic (am = 0) and manual transmissions (am =1) is the Standard Error of 7.245 mpg with Manual transmission having better (higher) mileage per gallong (See Fig.4). The coeffecients for the am only model are:

Intercept	am (1)
17.147368	7.24439

This is confirmed looking at the mtcars data separated by am: AMT<-mtcars[which(mtcars$am==0),] & MAT<-mtcars[which(mtcars$am==1),] (Fig4)

Mean mpg for Automatic	Mean mpg for Manual
17.14737	24.39231

Appendix

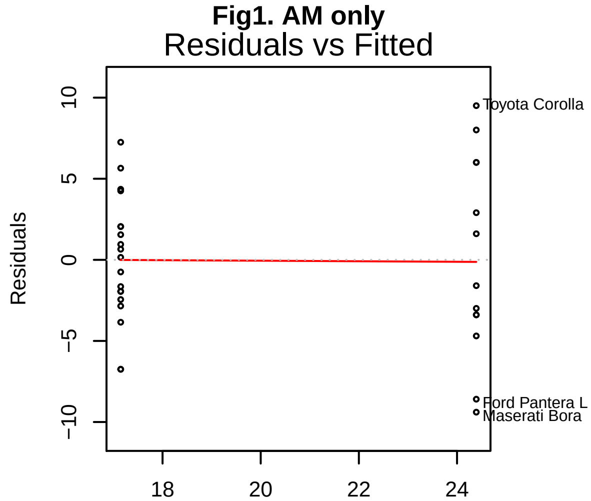

Fig1. AM only
Residuals vs Fitted
Residuals
Toyota Corolla
Ford Pantera L
Maserati Bora
10
5
0
−5
−10
18
20
22
24

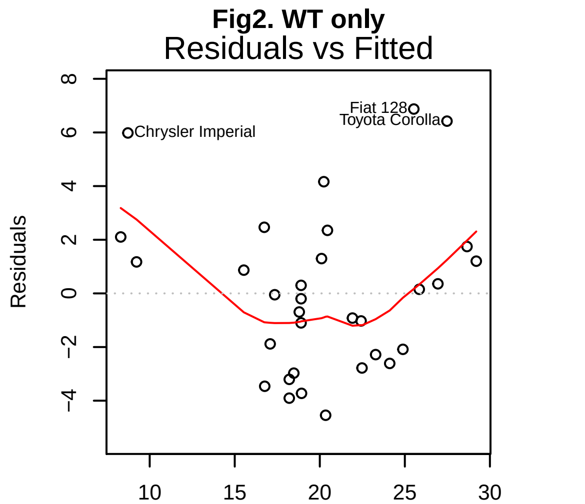

Fig2. WT only
Residuals vs Fitted
Residuals
Chrysler Imperial
Fiat 128
Toyota Corolla
8
6
4
2
0
−2
−4
10
15
20
25
30

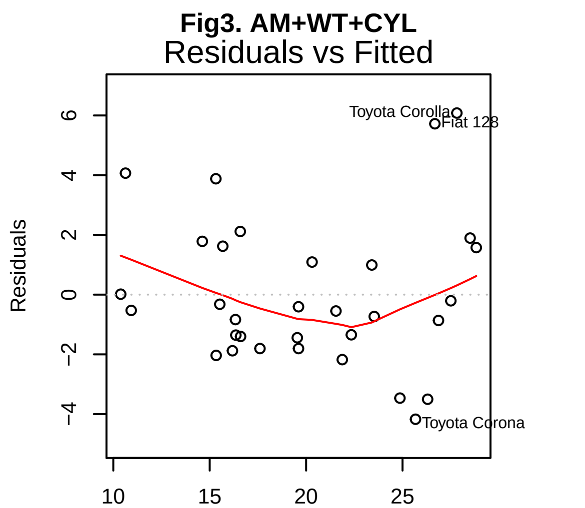

Fig3. AM+WT+CYL
Residuals vs Fitted
Residuals
Toyota Corolla
Fiat 128
Toyota Corona
6
4
2
0
−2
−4
10
15
20
25

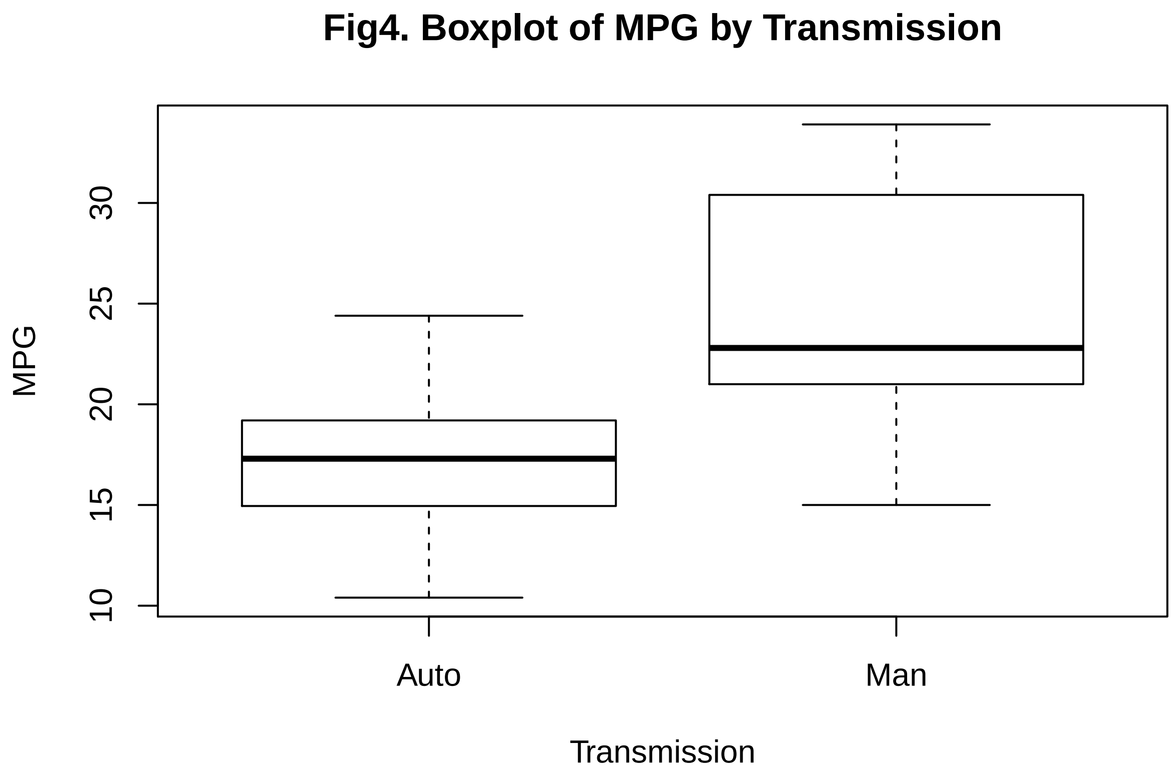
Fig4. Boxplot of MPG by Transmission
MPG
10
15
20
25
30
Auto
Man
Transmission

"'