Fuel Economy and Transmission Type

kwu

Wednesday, June 17, 2015

In this report, I attempt to determine if there is a relationship between fuel economy (measured as mpg) and transmission type (manual/automatic) using the mtcars dataset. Further, if such a relationship exists, I will attempt to quantify it.

A scatter plot of all covariate pairs (Fig 1) reveals that there appears to be a higher mean mpg for manual transmission cars. However, many of the covariates appear to exhibit a relationship with mpg. Further, many of the covariates are themselves highly correlated with each other raising the risk of variance inflation and multicollinearity. The covariates cylinders (cyl), engine displacement (disp), and weight (wt) are highly correlated and our covariate of interest, transmission type (am), is strongly negatively correlated with wt. When we regress mpg against all of the covariates, none of them are significant at the 5% level, though wt stands out with the lowest p value which intuitively makes sense. Looking at the variance inflation of the model reveals cyl, disp, and wt appear to be highly related.

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07

##       cyl      disp        hp      drat        wt      qsec        vs
## 15.373833 21.620241  9.832037  3.374620 15.164887  7.527958  4.965873
##        am      gear      carb
##  4.648487  5.357452  7.908747
```

Therefore we attempt to reduce the number of covariates using Akaike information criterion (AIC) with the 'step' function in R. Working backwards from a model with all covariates, this retains wt, am, and 1/4 miel

time (qsec) as covariates (see appendix for source code). All 3 covariates are significant at the 5% level and the model has an adjusted R squared (0.8336) greater than the model with all the covariates (0.8066).

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Next we examine residuals of the model in Fig 2 to see if any bias has been introduced by removing some of the covariates. There does not appear to be evidence of heteroskedasticity or non-independence except for with respect to qsec.

My analysis suggests that transmission type does have a relationship with fuel economy even after adjusting for weight at the 5% significance level. The model predicts that a manual transmission increase fuel economy 2.9mpg compared to an automatic transmission adjusted for the weight of the car and its ability to accelerate. However, this prediction ignores the possibility of a non linear relationship or potential interaction effects between the covariates.

Appendix

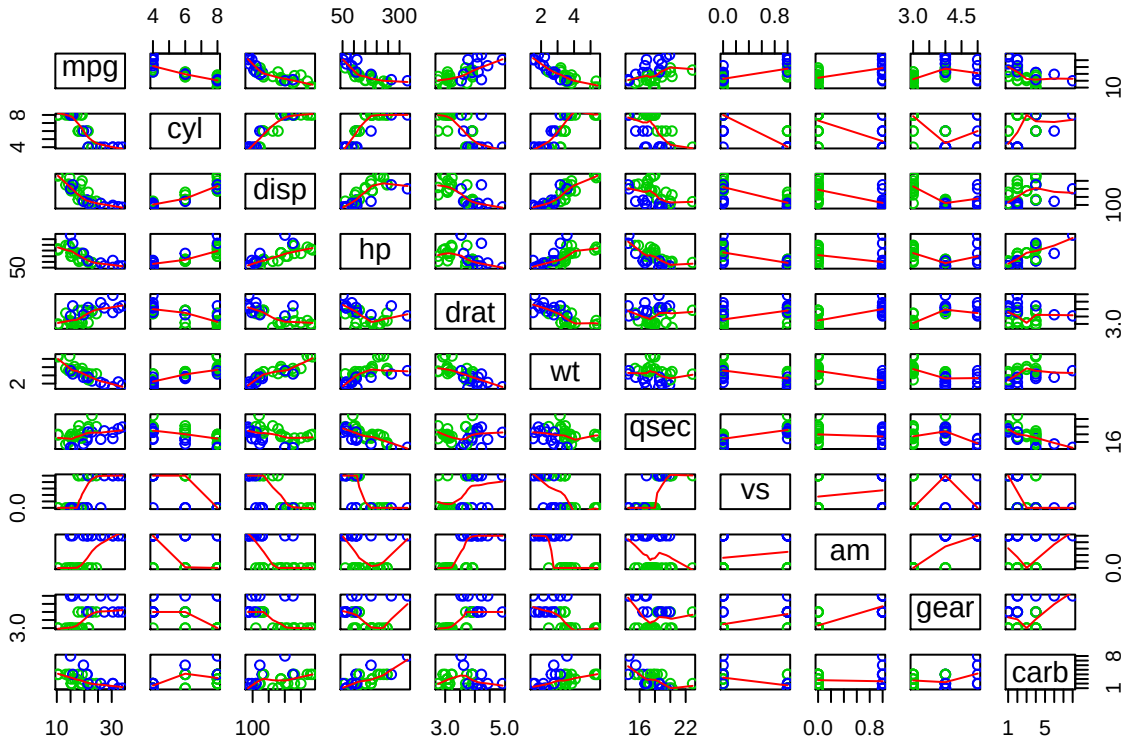

Figure 1

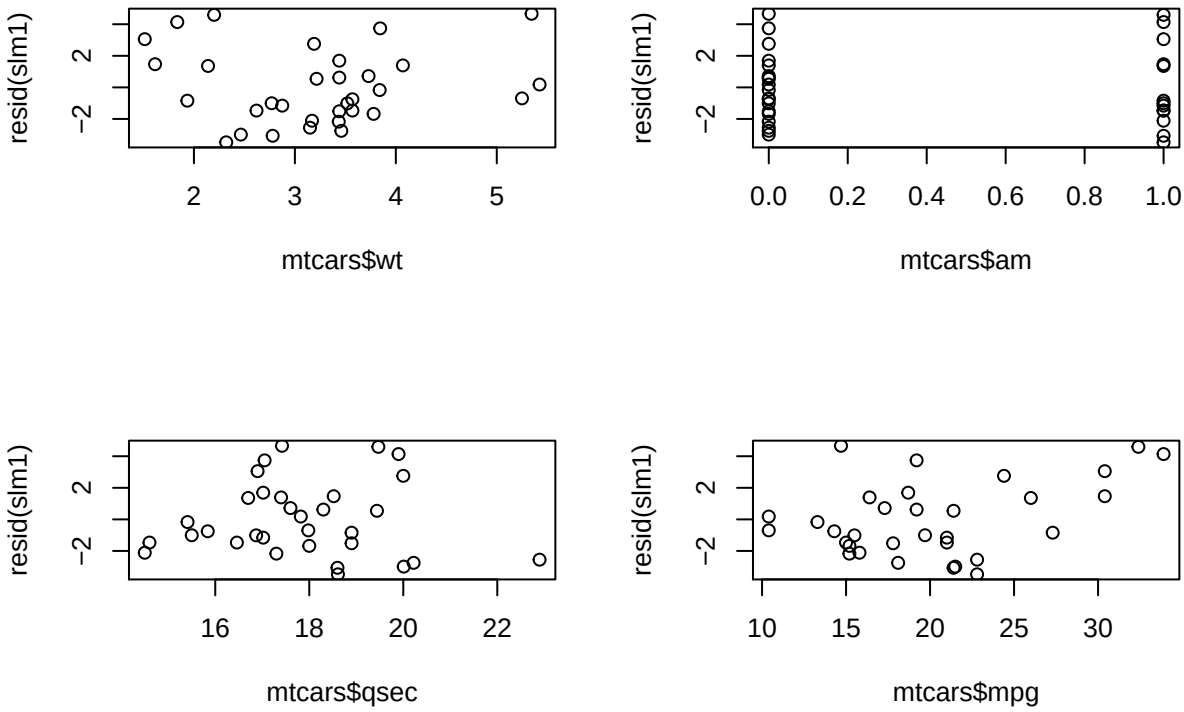

Figure 2

Source Code

```
lm1 <- lm(mpg ~ ., data = mtcars)
slm1 <- step(lm1, trace=0)
```