Effect of type of transmission on Mileage

Koushik Kumaraswamy

Tuesday, October 21, 2014

Executive Summary

This study attempts to quantify the effects of type of transmission on mileage(mpg). The study concludes that manual transmission is better for mpg compared to automatic transmission. Change in mpg by moving to a manual transmission from automatic is estimated to be around 2.936 holding other parameters constant. The 95% confidence interval for this increase is [0.045 , 5.826], holding other parameters constant. In addition to the transmission type, weight and qsec parameters had significant impact on automobile mileage.

Exploratory data analysis

Goal of analysis is to answer key questions on relationship between mileage (mpg) and type of transmission(am) as found in the **mtcars** dataset in R. Refer to the help (?mtcars) for details on data set.

Appendix 1 pairs plot reveals that **mpg** seems to have a negative relationship with **cyl**,**disp**,**hp**, **wt**, and **carb** fields. A postive relationship is indicated between **mpg** and **drat** , **qsec**, and **gear**

Both **am** and **vs** are indicator fields where the classes corresponding to **1** have better mileage relative to **0** class.

Model Selection

We first fit **mpg** to **am** and initially ignore effects of other 9 predictors.

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## am             7.245      1.764   4.106 2.850e-04
```

p value of **beta_1** is significant and **beta_1** = 7.24 is the change in mpg between automatic (am=0) and manual (am=1) holding all other regressors constant. We still need to review the effects of the other predictors. So let's fit a maximal model using all predictors.

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788  0.6573  0.51812
## cyl         -0.11144    1.04502 -0.1066  0.91609
## disp         0.01334    0.01786  0.7468  0.46349
## hp          -0.02148    0.02177 -0.9868  0.33496
## drat         0.78711    1.63537  0.4813  0.63528
## wt          -3.71530    1.89441 -1.9612  0.06325
## qsec         0.82104    0.73084  1.1234  0.27394
## vs           0.31776    2.10451  0.1510  0.88142
## am           2.52023    2.05665  1.2254  0.23399
## gear         0.65541    1.49326  0.4389  0.66521
## carb        -0.19942    0.82875 -0.2406  0.81218
```

```
## [1] "Adjusted R^2: 0.806642318990986"
```

None of the **p** values are <0.05: hence not a good model . We need to reduce this to a “minimally adeqate” / parsimonious model by performing stepwise regression in backward direction. Start with **fullfit** above and take out least significant parameters in iterations until we have a model with the least # of significant predictors.

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Coefficients:
## (Intercept)           wt         qsec           am
##        9.62        -3.92         1.23         2.94
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    9.618     6.9596   1.382 1.779e-01
## wt            -3.917     0.7112  -5.507 6.953e-06
## qsec           1.226     0.2887   4.247 2.162e-04
## am             2.936     1.4109   2.081 4.672e-02
```

```
## [1] "Adjusted R^2: 0.833556080257604"
```

This has higher adjusted R^2 vs fullfit model inspite of using only 3/10th of the predictors and is our answer to capturing effects of all significant predictors.

Residual & other diagnostics

Reviewing the 2x2 diagnostic plot in Appendix 2, we find that: (a)Residual & standarized residual plot do not show significant patters in variation, (b)Normal QQ plots indicate that the errors are normally distributed, (c)Residual vs leverage plot do not indicate significant outliers

Getting to the answers

Firstly, We can safely say that manual transmission is better for MPG since the **am** coefficient that indicates the difference in mpg between 0(auto) and 1(manual) is positive.

Secondly, lets get the 95% confidence intervals for the parameters:

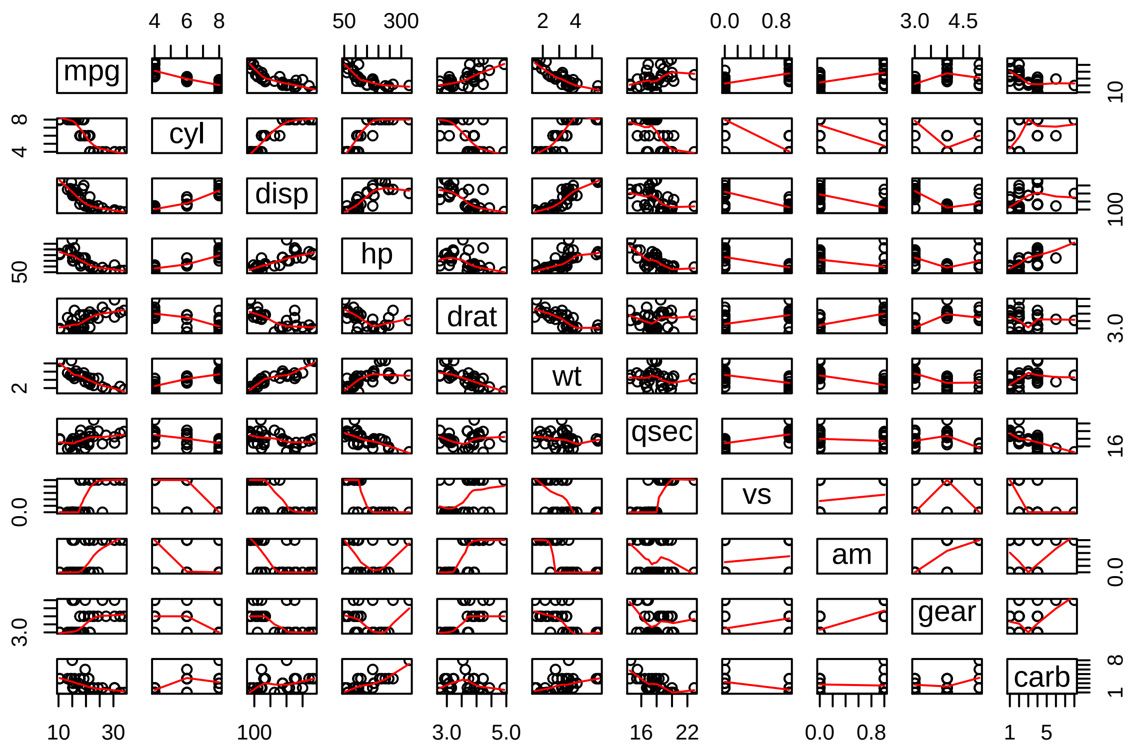
```
##                 2.5 % 97.5 %
## (Intercept) -4.63830 23.874
## wt          -5.37333 -2.460
## qsec         0.63457  1.817
## am           0.04573  5.826
```

Based on this we can say that 95% confidence interval for change in mpg by using manual transmission Vs Auto is between 0.045 and 5.826 holding other two regressors constant. The estimated value within this interval as noted from the final model coefficient is 2.936.

Appendix 1 : Pairs Plot

Initial plot of pairs of regressors:

```r
data(mtcars)
pairs(mtcars,panel=panel.smooth)
```



Appendix 2 : Diagnostic plots of final model

2x2 Diagnostic plot of final model:

```
par(mfrow=c(2,2))
plot(finalfit)
```