Regmods-016 Course Project

Duncan Turnbull

17 December 2014

Executive Summary

This project seeks to answer two key questions around fuel consumption based on car data from the 1974 Motor Trend US magazine: + "Is an automatic or manual transmission better for MPG". + "Quantify the MPG difference between automatic and manual transmissions" The data provided had 32 samples with 11 variables. Correlations amongst the variables suggests that weight and the number of cylinders were also variables of interest

Once consideration for weight and engine cylinders are included it does not appear that a significant differnce between manual and automatic transmission can be seen. The primary factors are the car weight and the engine size.

Look at the data

Looking particularly at how automatic and manual transmissions vary

1. Identify variables of interest

Potential factors in the data - am (Transmission 0 = auto, 1 = manual)

Look at multi variate model, rule out anything check VIF qsec (1/4 mile time) seems unlikely, maybe drat (rear axle ratio) too
Other variables probably affect fuel consumption similarly whether manual or auto i.e. weight, cylinders, displacement, horsepower, V/S number of carburetors The number of forward gears will probably influence the choice

2 .Final models

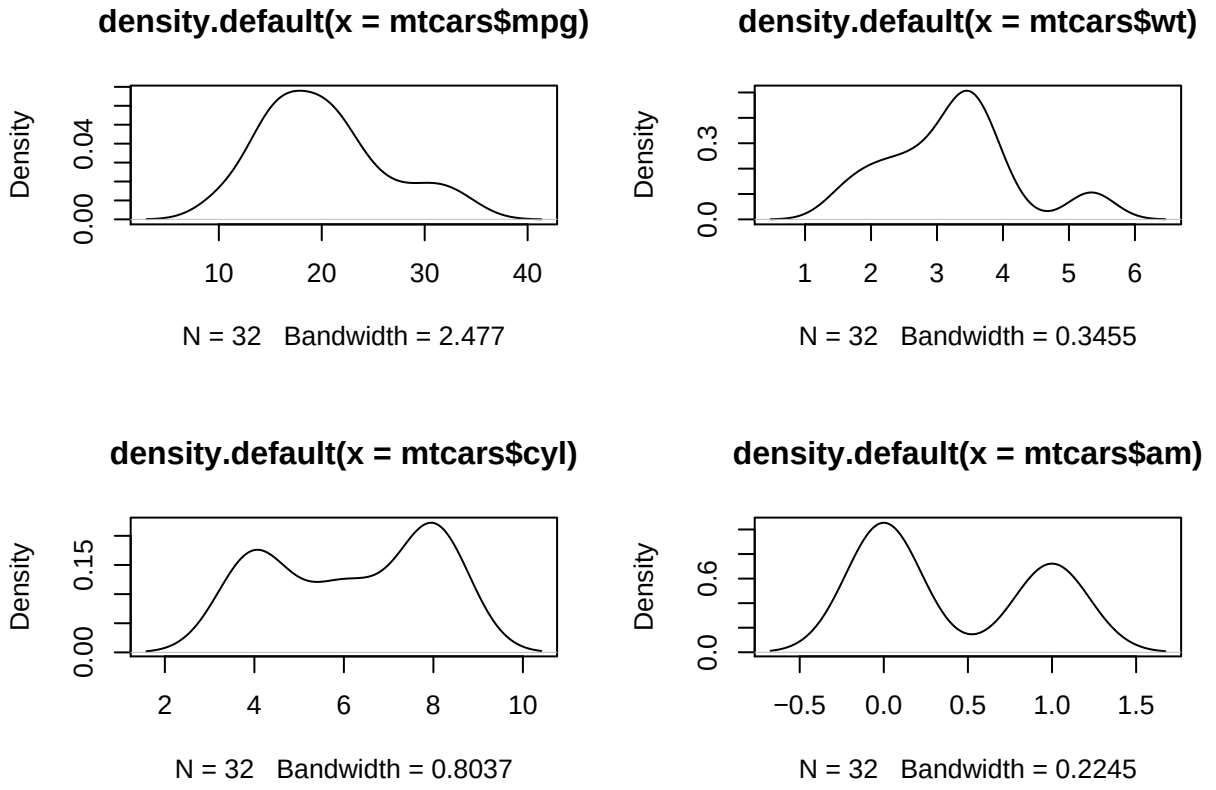
```
fit <- lm(mpg ~ factor(am) +wt ,mtcars)
```

Exploratory Data analyses Fit multiple models and detail strategy for model selection Did the student answer the questions of interest or detail why the question(s) is (are) not answerable? Did the student do a residual plot and some diagnostics? Did the student quantify the uncertainty in their conclusions and/or perform an inference correctly? Was the report brief (about 2 pages long) for the main body of the report and no longer than 5 with supporting appendix of figures? Did the report include an executive summary? Was the report done in Rmd (knitr)?

Appendix

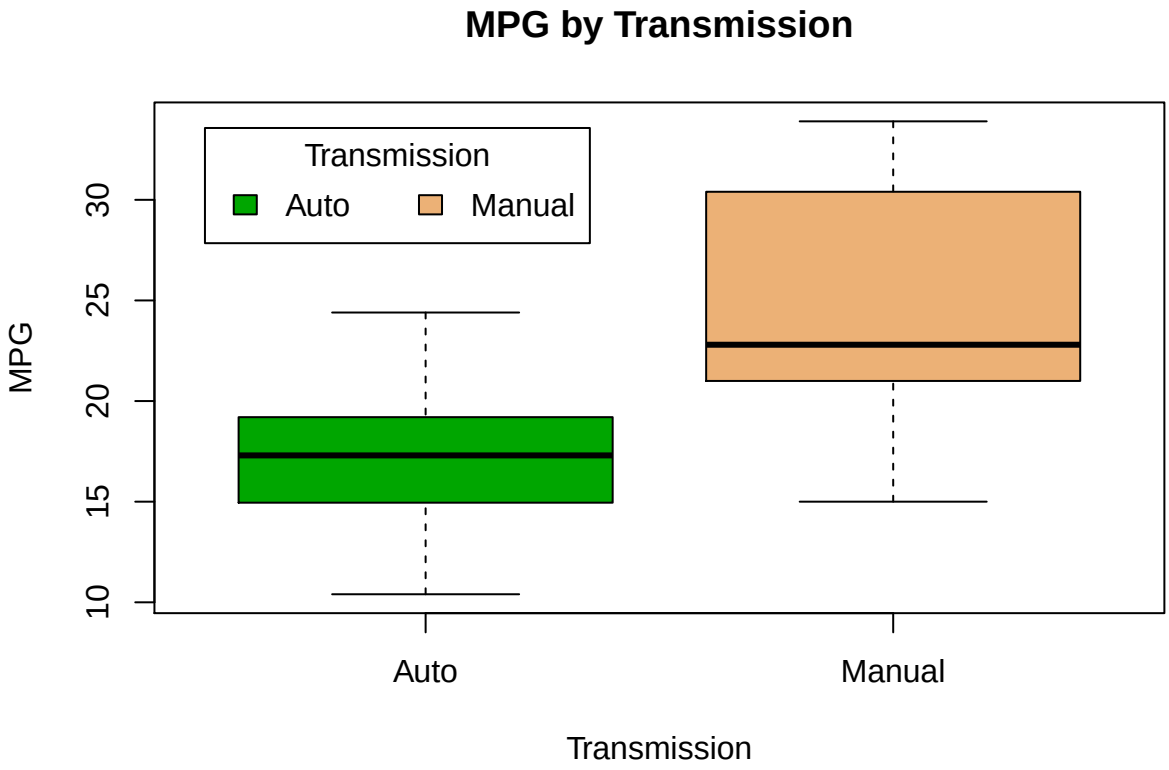
Density / distribution of interesting variable

```
par(mfrow=c(2,2)) ; plot(density(mtcars$mpg));plot(density(mtcars$wt));
plot(density(mtcars$cyl));plot(density(mtcars$am))
```



MPG by transmission

From the boxplot below it appears as if there are clear distinctions between the MPG for different transmission types. However there could be some confounding variables.



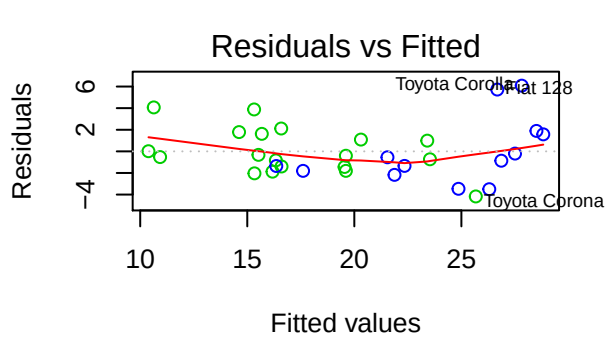
Correlations

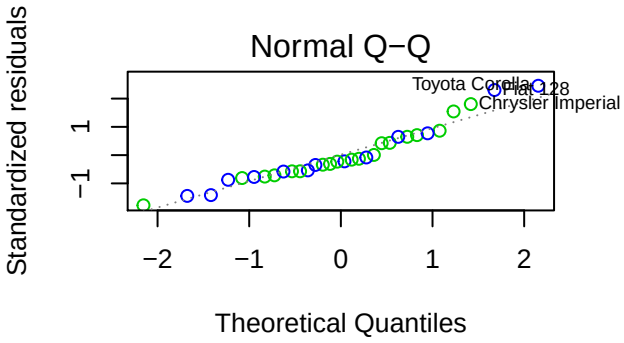
The correlation table shows strong correlation to weight (wt), number of cylinders (cyl), size of engines (disp), gross horsepower (hp), however hp, disp and cyl are all strongly correlated with each other suggesting only one factor is required. For our purposes we choose cyl, the number of cylinders.

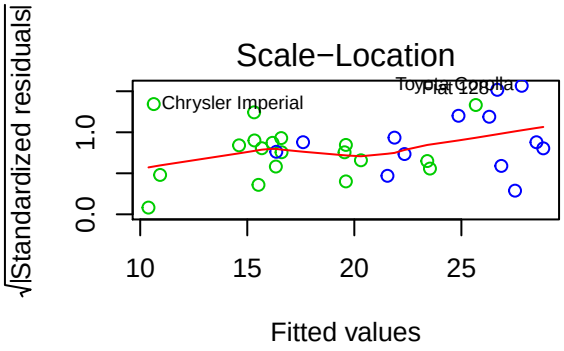
```
##            mpg        cyl       disp         hp       drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.6999381  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.7102139  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.4487591  0.6587479
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.7124406  1.0000000
##            qsec         vs         am       gear       carb
## mpg   0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
## cyl  -0.5912421 -0.8108118 -0.5226070 -0.4926866  0.5269883
## disp -0.4336979 -0.7104159 -0.5912270 -0.5555692  0.3949769
## hp   -0.7082234 -0.7230967 -0.2432043 -0.1257043  0.7498125
## wt   -0.1747159 -0.5549157 -0.6924953 -0.5832870  0.4276059
```

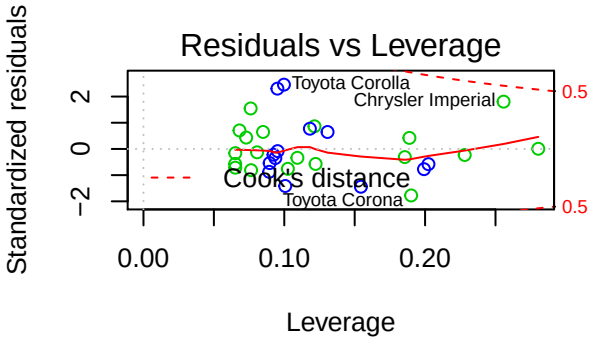
Residuals

The residual plots appear relatively normally distributed and show any observations with undue influence.


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona


Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Chrysler Imperial


Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Toyota Corolla
Chrysler Imperial
Toyota Corona
0.5
0.5