

# Regression Models Project - Motor Trend Data 'mtcars'

## Miles Per Gallon Analysis

*james c walmsley*

*12/1/2016*

### Executive Summary:

The mean "MPG" rating of vehicles with automatic transmissions is 17.15 mpg, and those with manual t

Upon further analysis of the relationship between vehicle design components and "MPG" ratings in th

Additionally, vehicle weight is highly correlated (-86%) with mpg ratings and transmission type is l

To answer the question of whether automatic or manual transmissions are better for gas mileage we w

### Problem Statement & Questions to Answer:

```
## Q1 "Is an automatic or manual transmission better for 'mpg'"
## Q2 "Quantify the MPG difference between automatic and manual transmissions"
```

---

### Grading - Criteria (remove on completion)!!!

```
####YES!!!! Did the report include an executive summary?
####???? Did the student answer the questions of interest or detail why the question(s) is (are) no
####???? Did the student quantify the uncertainty in their conclusions and/or perform an inference.
####???? Was the report brief (about 2 pages long) for the main body of the report and no longer th
```

```
####YES!!!! Did the student interpret the coefficients correctly?
####YES!!!! Did the student do some exploratory data analyses?
```

```
####YES!!!! Did the student fit multiple models and detail their strategy for model selection?
####Needs some more work
```

```
####YES!!!! Did the student do a residual plot and some diagnostics?
####YES!!!! Was the report done in Rmd (knitr) with pdf output?
```

---

## Analysis Considerations:

Descriptive - (dim, mean, sd, sigma<sup>2</sup>, str & summary) statistics

Exploratory - pairs, histograms, QQ, fitted, residualplots, boxplots  
& (multiple plots); T-Test

Analysis - OLS, simple linear regression, statistical linear regression,  
multivariate regression & model selection, logistic regression, pValues,  
adjustments, residuals, residual fit, predict fit, hatvalues, variance, & dfbetas, R<sup>2</sup>,  
diagnostics; ANOVA, coefficients, confint, correlation, covariance, variance inflation

## Software Environment: & System - session Info:

Set the Working Directory then get System & Session Info

## Accessing Data & Raw Data Overview: Motor Trend 'mtcars' data set:

Clean up the work space & get the data:

A data frame with 32 observations on 11 variables.

```
[, 1] mpg Miles/(US) gallon  
[, 2] cyl Number of cylinders (4,6,8)  
[, 3] disp Displacement (cu.in.)  
[, 4] hp Gross horsepower  
[, 5] drat Rear axle ratio  
[, 6] wt Weight (1000 lbs)  
[, 7] qsec 1/4 mile time  
[, 8] vs V/S (V = vee-block, S = straight-block)  
[, 9] am Transmission (0 = automatic, 1 = manual)  
[,10] gear Number of forward gears (3,4,5)  
[,11] carb Number of carburetors (1,2,3,4,6,8)
```

## Process Data: factor columns 2 & 8:11 (cyl,vs,am,gear,carb)

Order the data set from least mpg on the top row to highest mpg on the bottom row

## Descriptive Statistics

Visual review of the list of vehicles in order of mpg seems to indicate that there may be a strong negative correlation between mpg and the number of cylinders, larger cubic inch displacement, higher horsepower, V-engine shape, automatic transmission, low # of gears, and high # of carburetors and the reverse seems to be correlated to higher mpg ratings. This seems to indicate a high degree of multicollinearity.

## Exploratory Analysis:

See figures 1:4 in Appendix A  
Add narrative here!!

---

## Statistical Modeling, Regression & Model Fit:

Assumptions:

- A A correlation to mpg ratings may exist among multiple variables
- B

### Simple Linear Regression Plot

In this plot we examine the relationship between mpg ratings and vehicle weight and find that as mpg rating increases vehicle weight decreases.

### Bivariate Linear Model: # Note: manual trans & eng shape / vs are significant

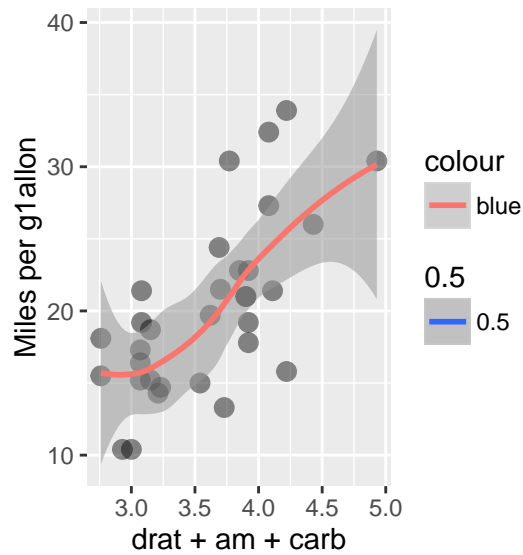
This plot examines adding to the weight component the transmission type in a bivariate model that highlights some vehicles with noticable leverage on the model results in particular the followign vehicles: for the low end of the mpg ratings ~ Chrysler Imperial, and for the high mpg ratings the ~ Fiat 128, Toyota Corolla,

### Multivariate Linear Model (all vars)

Considering the results of the VIF (variance inflation test) vif(fit) results indicate variance inflation of these regressors: cyl, disp, hp, drat, wt, qsec, vs, am, gear, & carb. Considering the sqrt(vi(fit) though we can choose those regressors that have values less than three which are: drat, vs, gear, carb and keep the one of interest which is am

**variables with low sqrt of variance inflation = cyl, drat, vs, am, gear & carb; The first multivariate model includes these regressors**

```
## I(drat + am)          carb
##      1.000215      1.000215
```



## Multivariate Linear Model Adjusted

### Multivariate LM Nested & ANOVA table

With nested modeling method, models 3 & 6 each add a significant variable to the fitted model and in this case model 3 is disp & model 6 is wt. Now use the vif (variable inflation) test on model six to check for any variance inflation among the variables of this model. The results show that “hp”, “drat”, and “wt” all have square rooted inflation values less than 3.0 so these can be accepted into the model of best fit (fbf1) along with the main variable of the study “am”

### Best Fit Modeling

Based on the nested modeling process followed by the anova table check then followed by the vif and sqrt(vif) test we decide to go with the following model labeled (fbf1) and notice that all of the sqrt(vif) values are less than 2.0 indicating a good model fit.

```
===== # DIAGNOSTICS A
```

## DIAGNOSTICS B

however the confint for drat, hp and wt each include zero

NEXT PROCESS: Prediction !!!!

## Preliminary Findings:

Questions of Interest:

A What other regressors if any correlated with mpg rating and transmission type?

B

Interpretation of Results:

A Using ANOVA table with Nested Multivariate Regression fit it is clear that the variable w  
 B Based on the  
 C

## Inference:

Hypothesis':

A  $H_0$  = The difference between Automatic and Manual transmission MPG = 0  
 B  $H_a$  = The difference between Automatic and Manual transmission MPG  $\neq$  0  
 C Desired confidence interval = .95 (one sided) ??

## Conclusions / Recommendations:

A  
 B

## Are there other alternative analyses?

A VIF  
 B Challenge the results ?  
 C Measures of uncertainty 'e'

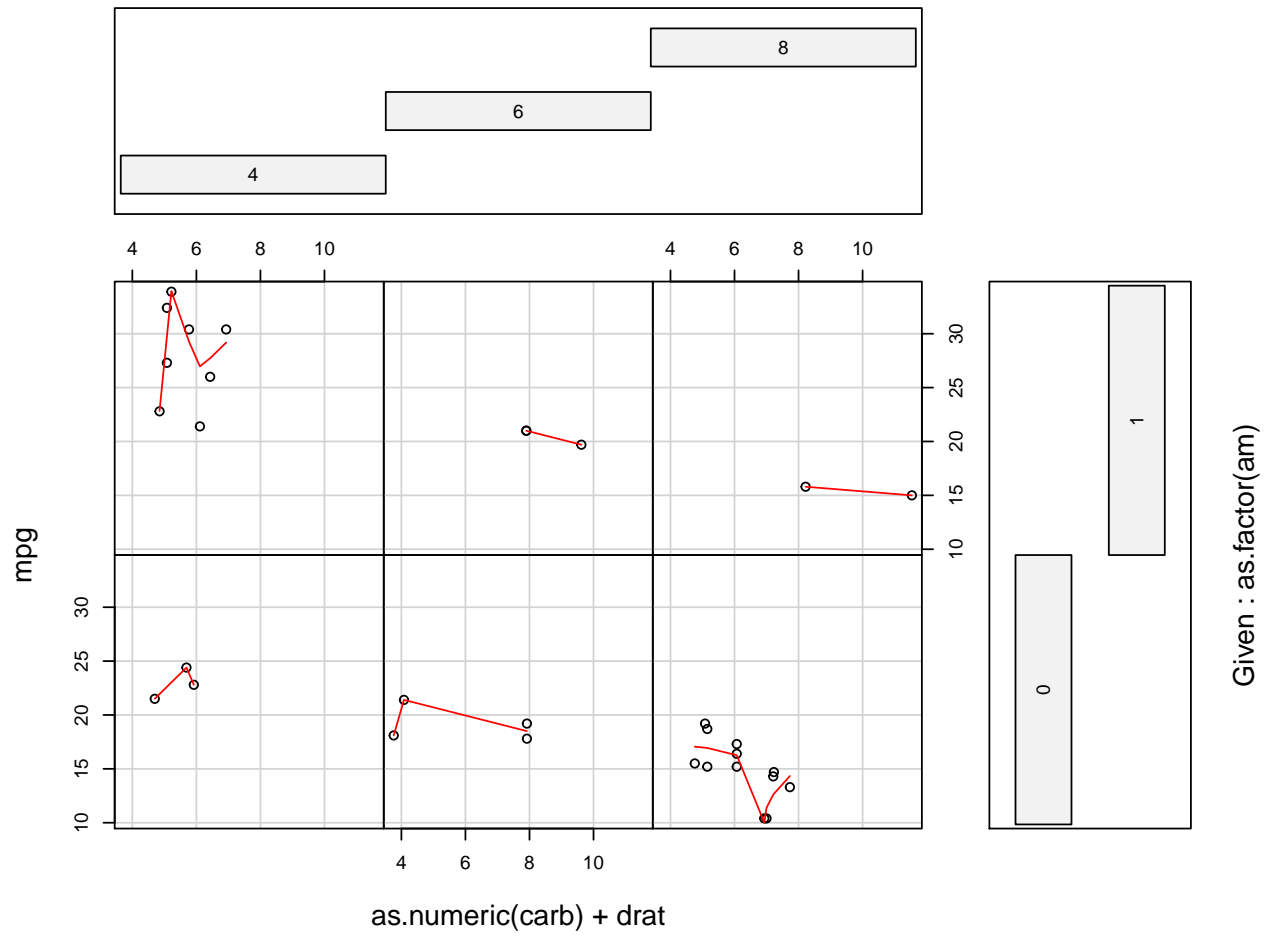
---

## Appendix A: “Exploratory Graphical Analysis”

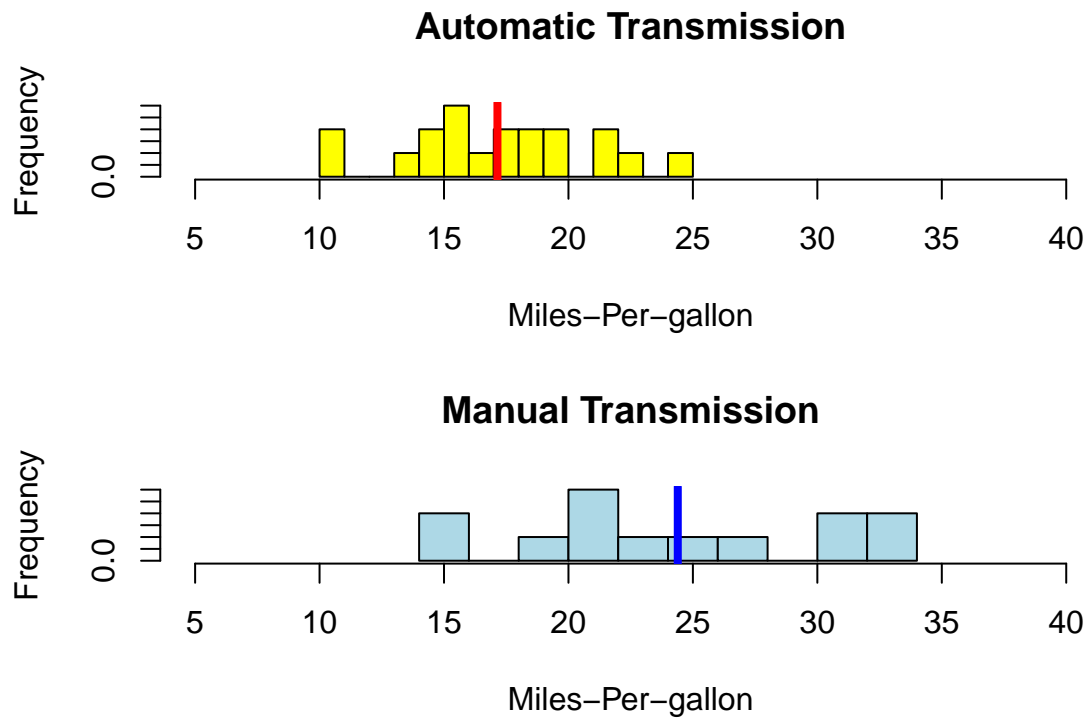
### Pairs Plot:

Interpretation: MPG decreases as the; # of cylinders increases, the engine displacement increases, the horsepower increases, the weight increases, the rear axle ratio decreases, qsec time decreases, the engine is a V-block, # of gears decreases, and the # of carburetors increases ~ conversely the MPG increases as the: rear axle ratio increases, qsec time increases, the engine is a Straight-block, transmission is manual, # of gears increases and the # of carburetors decreases

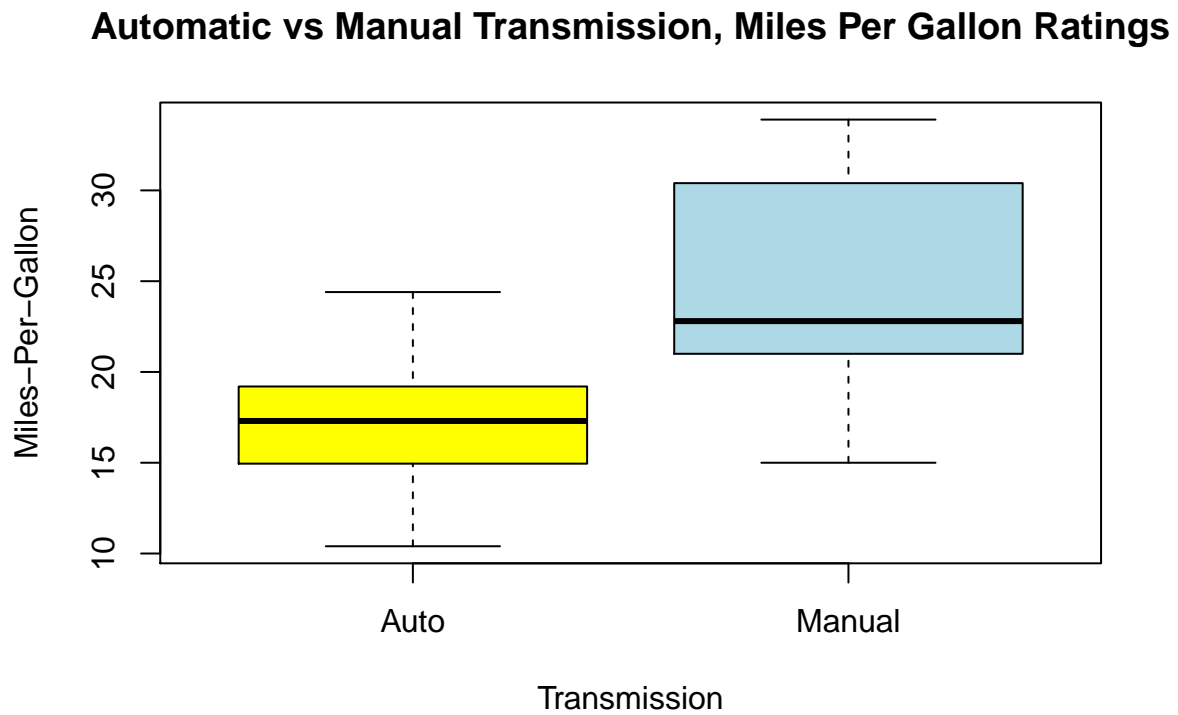
Given : as.factor(cyl)



## Histograms Plot



## Box Plot



**NOTE: use the cut function by 3 on MPG**

=== END ===