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 transmisions 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 

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^2, 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^2,
diagnostics; ANOVA, coeficients, 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 reveiw 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 high # cylinders, larger cubic inch displacement, higher horsepower, V-engine shape, automatic transmission, lwo # of gears, and high # of carburetors and the reverse seems to be correlated to 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 transmision 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

```r
library(ggplot2);library(dplyr)
data("mtcars")
head(mtcars,4)
```

```
##                 mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
```

```r
f2 <- lm(mpg ~ drat + am + cyl + vs + gear + carb, data = mtcars)
coef(summary(f2))
```

```
##               Estimate Std. Error    t value   Pr(>|t|)
## (Intercept) 19.4217139 10.1778170  1.9082396 0.06790743
## drat         1.4847375  1.6944829  0.8762186 0.38925304
## am           3.1389197  1.9498251  1.6098468 0.11998744
## cyl         -0.9891230  0.8005129 -1.2356116 0.22808843
## vs           0.8875264  2.0047454  0.4427128 0.66178044
## gear         1.0595862  1.4957354  0.7084049 0.48524733
## carb        -1.4654381  0.5433135 -2.6972238 0.01233725
```

```r
vif(f2)
```

```
##     drat       am      cyl       vs     gear     carb
## 3.197676 3.687659 7.962247 3.977237 4.744231 3.000045
```

```r
sqrt(vif(f2))
```

```
##     drat       am      cyl       vs     gear     carb
## 1.788205 1.920328 2.821745 1.994301 2.178126 1.732064
```

```r
f3 <- lm(mpg ~ I(drat + am) + carb, data = mtcars)
coef(summary(f3))
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)   8.513614  2.5122481  3.388843 2.038897e-03
## I(drat + am)  4.299534  0.5623180  7.646089 1.977436e-08
## carb         -2.002921  0.3330488 -6.013897 1.533216e-06
```

```r
confint(f3)
```

```
##                  2.5 %    97.5 %
## (Intercept)   3.375490 13.651739
## I(drat + am)  3.149464  5.449603
## carb         -2.684083 -1.321760
```

```r
vif(f3)
```

```
## I(drat + am)         carb
##      1.00043      1.00043
```

```r
sqrt(vif(f3))
```

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

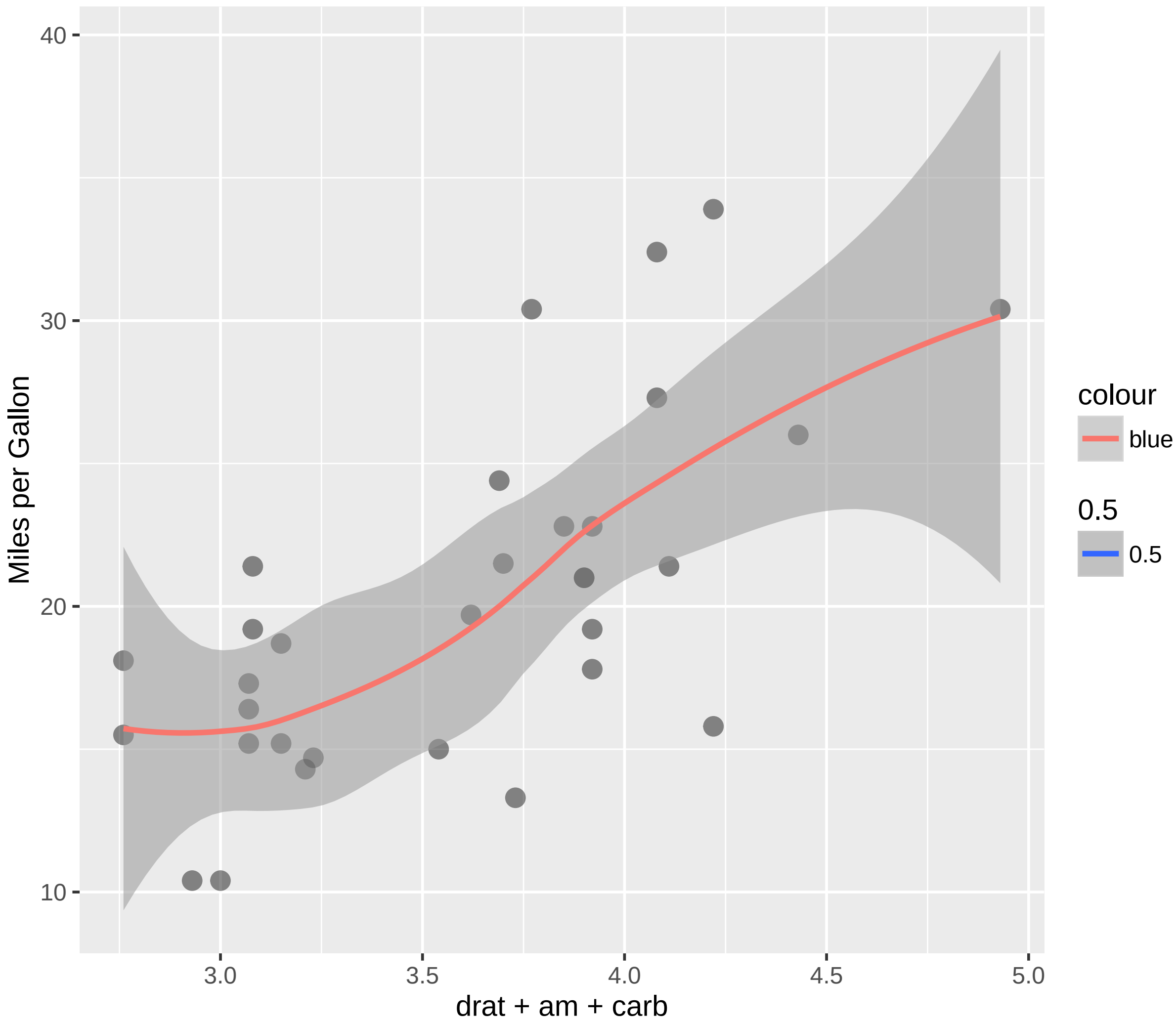
```r
x <- mtcars[,c(5,9,11)]
x
```

```
##                     drat am carb
## Mazda RX4           3.90  1    4
## Mazda RX4 Wag       3.90  1    4
## Datsun 710          3.85  1    1
## Hornet 4 Drive      3.08  0    1
## Hornet Sportabout   3.15  0    2
## Valiant             2.76  0    1
## Duster 360          3.21  0    4
## Merc 240D           3.69  0    2
## Merc 230            3.92  0    2
## Merc 280            3.92  0    4
## Merc 280C           3.92  0    4
## Merc 450SE          3.07  0    3
## Merc 450SL          3.07  0    3
## Merc 450SLC         3.07  0    3
## Cadillac Fleetwood  2.93  0    4
## Lincoln Continental 3.00  0    4
## Chrysler Imperial   3.23  0    4
## Fiat 128            4.08  1    1
## Honda Civic         4.93  1    2
## Toyota Corolla      4.22  1    1
## Toyota Corona       3.70  0    1
## Dodge Challenger    2.76  0    2
## AMC Javelin         3.15  0    2
## Camaro Z28          3.73  0    4
## Pontiac Firebird    3.08  0    2
## Fiat X1-9           4.08  1    1
## Porsche 914-2       4.43  1    2
## Lotus Europa        3.77  1    2
## Ford Pantera L      4.22  1    4
## Ferrari Dino        3.62  1    6
## Maserati Bora       3.54  1    8
## Volvo 142E          4.11  1    2
```

```r
y <- mtcars$mpg
n <- length(y)
par(mfrow = c(1,1), mar = c(4,4,2,2)) # set margin
g <- ggplot(mtcars, aes(x = drat, am, carb, y = y),)
```

```
## Warning: The plyr::rename operation has created duplicates for the
## following name(s): (``)
```

```r
g <- g + xlab("drat + am + carb")
g <- g + ylab("Miles per Gallon")
g <- g + geom_point(size = 3.0, col = "black", alpha = .45)
g <- g + geom_smooth(aes(method = "lm", col = "blue", alpha = .5))
g
```



Multivariate Linear Model Adjusted

Multivariate LM Nested & ANOVA table

With nested modeling method, models 3 & 6 each add a significate 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 varialbes of this model. The results show that "hp", "drat", and "wt" all have square rooted infaltion 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 H0 = The difference between Automatic and Manual transmission MPG = 0
        B Ha = The difference between Automatic and Manual transmission MPG != 0
        C Desired confidence interval = .95 (one sided) ??
```

Conclusions / Recommendations:

```
A
B
```

Are there other alternative analyses?

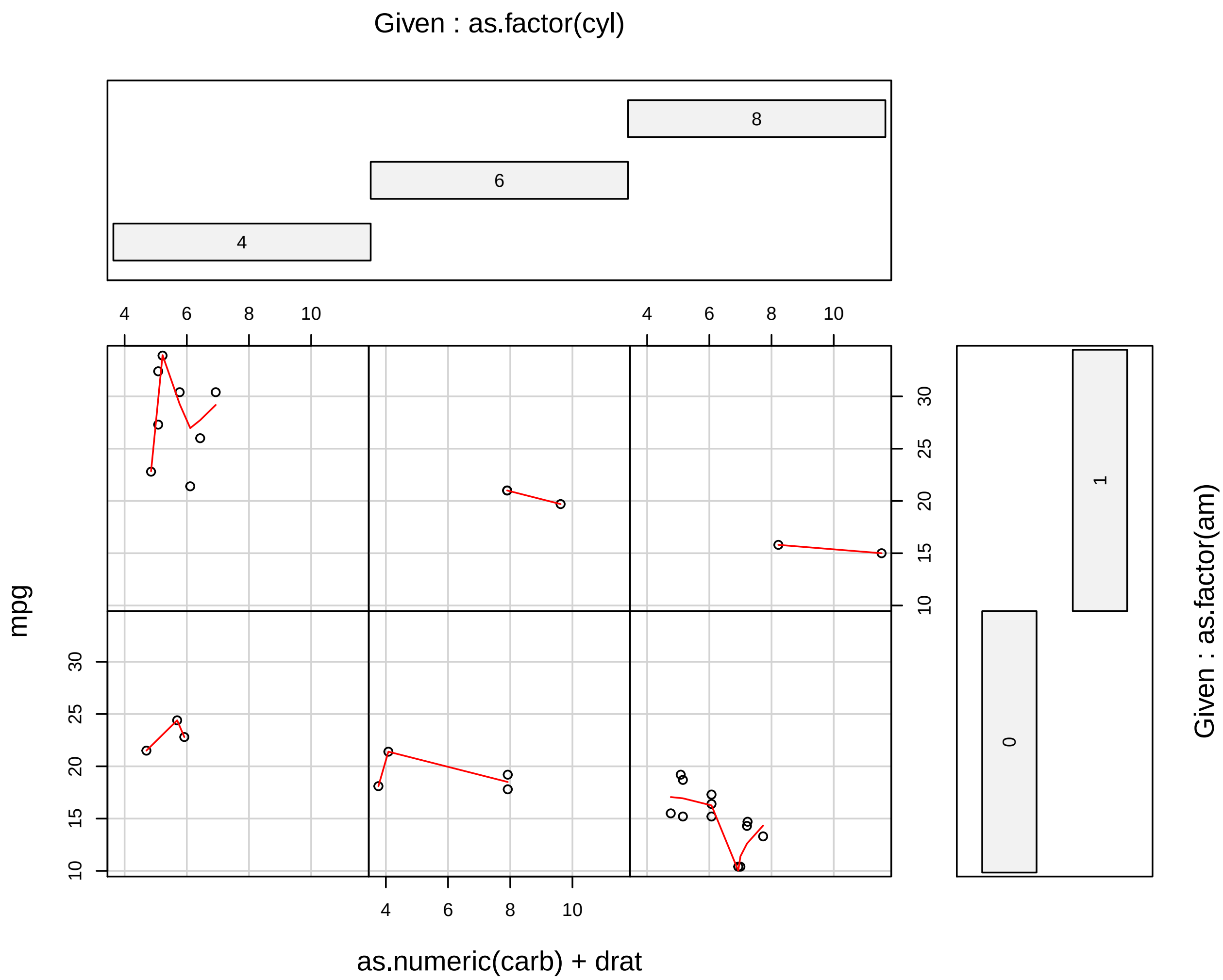
```
A VIF
B Challenge the results ?
C Measures of uncertainty 'e'
```

Appendix A: "Exploratory Graphical Analysis"

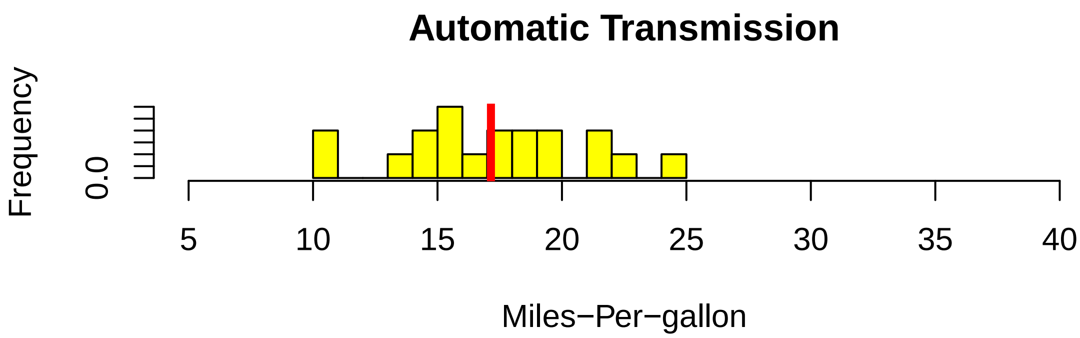
Pairs Plot:

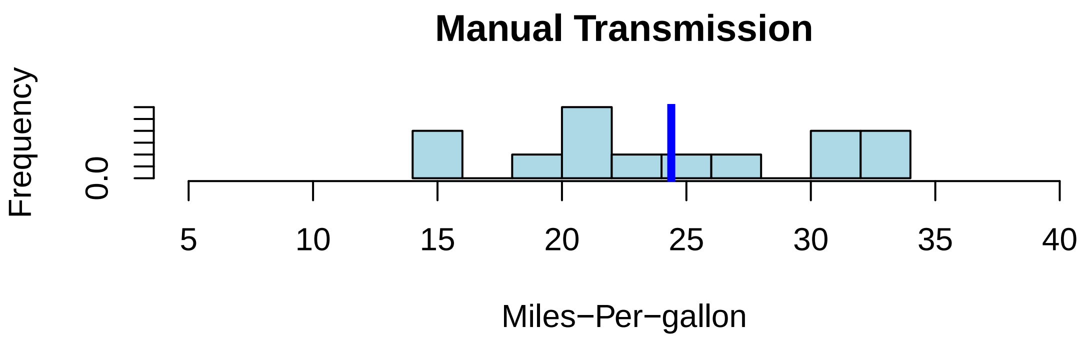
Interpretation: MPG decreases as the; # of cylinders increases, the engine displacement increases, the horsepower increases, the weight increases, the rear axle ratio decreases, qsec

time decreases, the engine is a V-block, # of gears decreases, and the # of carburators 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 carburators decreases

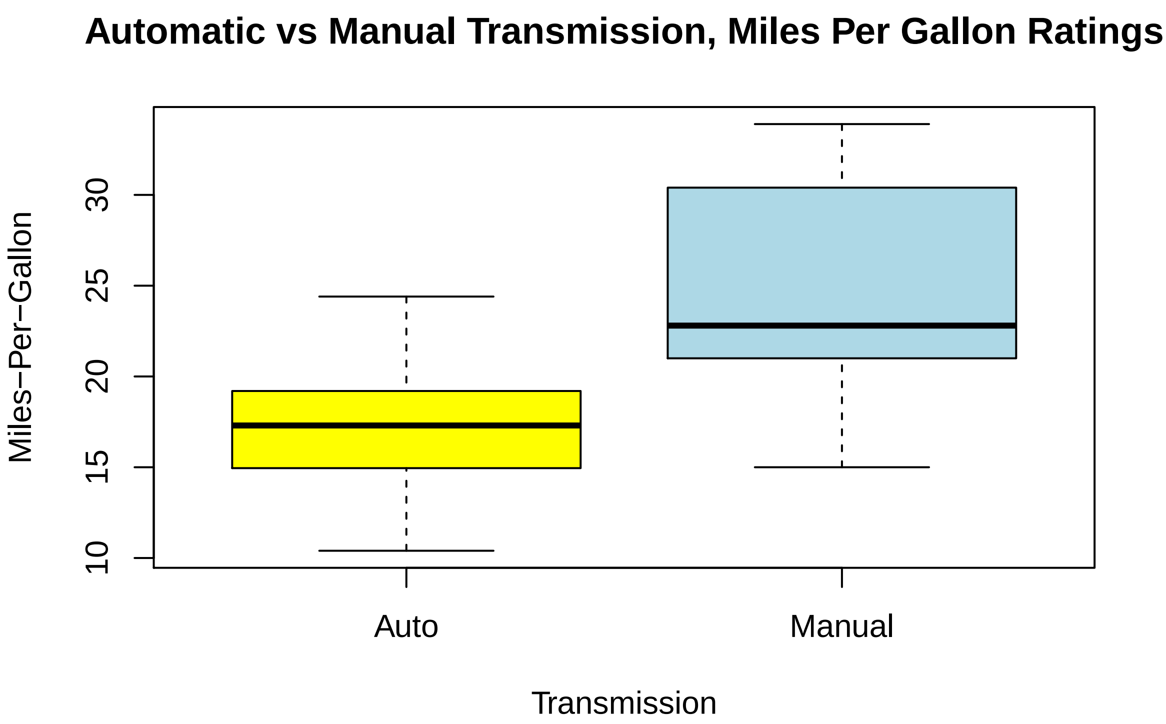


Histograms Plot





Box Plot



NOTE: use the cut function by 3 on MPG

=== END ===