Regression Models Project - Motor Trend Data 'mtcars' Miles Per Gallon Analysis

james c walmsley

12/1/2016

I. Executive Summary:

```
## Add after completing analysis
NOTE: include some info on cor, confint, ChisSq?, VIF
```

II. 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)!!!

Did the student interpret the coefficients correctly?

Did the student do some exploratory data analyses?

Did the student fit multiple models and detail their 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?

YES Was the report done in Rmd (knitr) with pdf output?

III. Analysis Considerations:

```
Descriptive - (dim, str & summary) statistics
Exploratory - pairs, histograms, boxplots & (multiple plots)
Analysis - mean, sd, sigma^2, T-Test, OLS, simple linear regression,
statistical linear regression, multivariate regression & model selection,
pValues, adjustments, residuals (predict fit, residual fit (-1)), hatvalues,
variance, & dfbetas, R^2, diagnostics; ANOVA, coeficients, correlation,
covariance, confint, VIF
```

IV. Software Environment:

Set working directory:

System - session Info:

V. Accessing Data:

Clean up the work space & get the data:

```
rm(list=ls()); data("mtcars")
```

VI. Raw Data Overview: Motor Trend 'mtcars' data set:

A data frame with 32 observations on 11 variables.

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 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
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

VII. Process Data:

Transformationns = Factor variables 8:11; modify variable names for col 8 & 9

```
data(mtcars)
mtcars$vs <- factor(mtcars$vs, labels = c("V-eng", "S-eng")); mtcars$am <- factor(mtcars$am, labels = c
head(mtcars,3)
```

VIII. Exploratory Analysis:

Add narrative here!!

```
??mtcars
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
```

```
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : Factor w/ 2 levels "V-eng","S-eng": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

```
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

    See Appendix A, Figures 1:4

IX. Statistical Modeling, Regression & Model Fit:

    Assumptions:
            A A correlation to mpg ratings may exist among multiple variables
            B
            C

```
?mtcars
t.test(mpg ~ am, data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

X. 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

XI. 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) ??
```

XII. Conclusions / Recommendations:

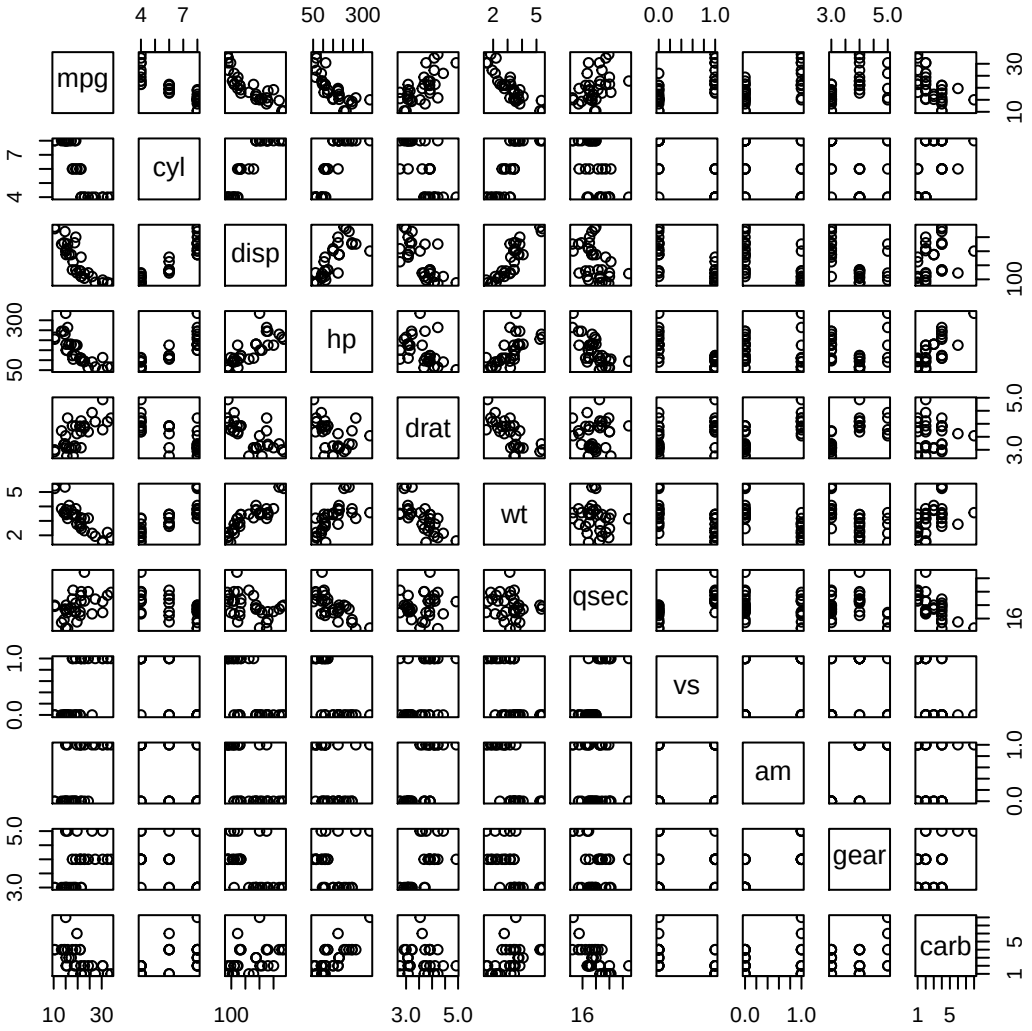
```
A
B
```

XIII. Are there other alternative analyses?

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

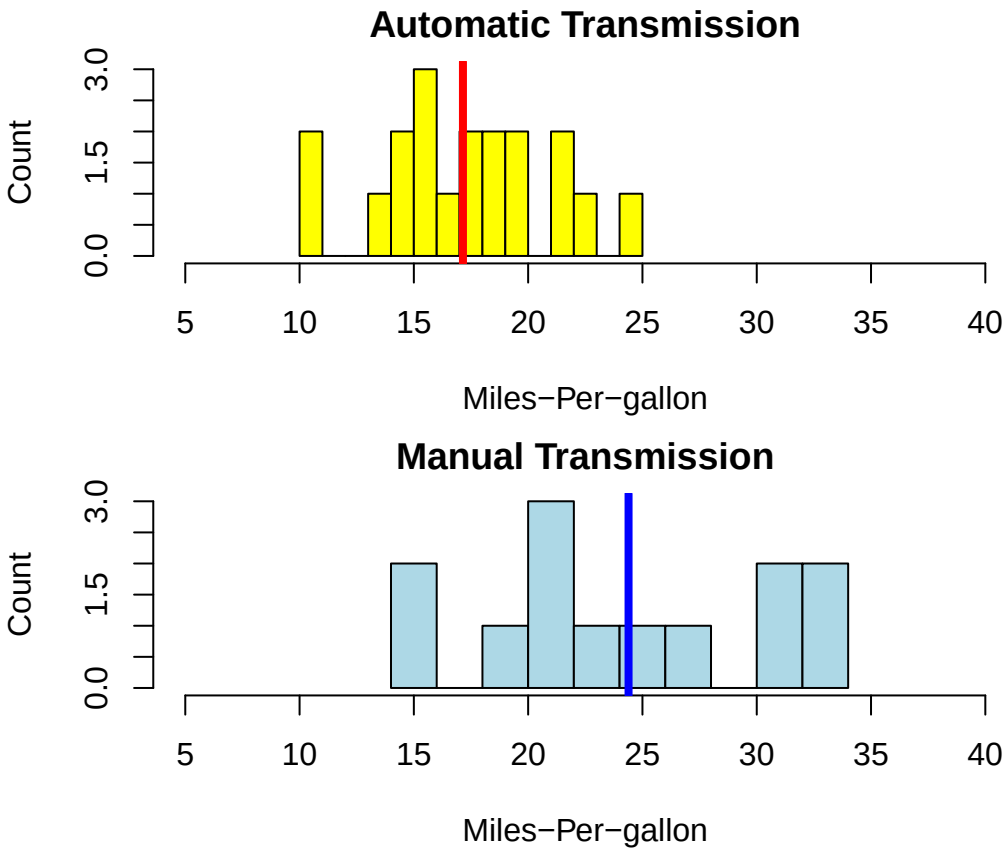
XIV. Appendix A, “Graphical Analysis”"

Pairs



Histograms

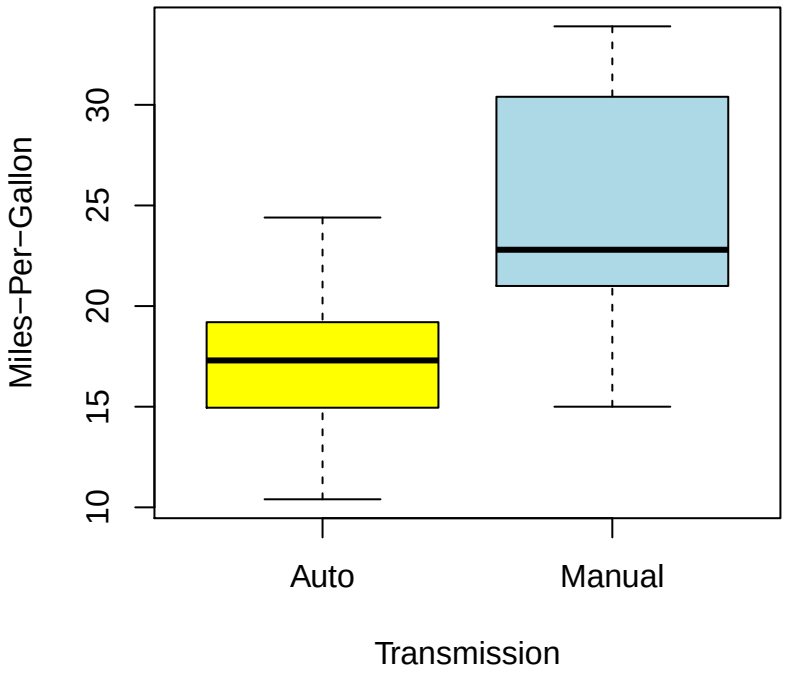
```
##                mpg cyl disp  hp drat    wt  qsec      vs     am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 S-block Manual    4    1
```



Box Plots

```
##                mpg cyl disp  hp drat    wt  qsec      vs     am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 S-block Manual    4    1
```

ɔmatic vs Manual Transmission, Miles Per Gallon



Simple Linear Regression Single Variable Plot

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
```

Bivariate Linear Model Regression plot

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## factor(am)1 -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

```
##                 Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)    31.416055  3.0201093 10.402291 4.001043e-11
## factor(am)1    14.878423  4.2640422  3.489277 1.621034e-03
## wt             -3.785908  0.7856478 -4.818836 4.551182e-05
## factor(am)1:wt -5.298360  1.4446993 -3.667449 1.017148e-03
```

Multivariate LM (all vars)-Residuals/Fitted/Residuals vs Fitted

Multivariate LM (all vars)-Residuals/Fitted/Residuals vs Fitted & Adjusted

Multivariate LM Nested Plot

=== END ===