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

james c walmsley

12/1/2016

Executive Summary:

```
## Add after completing analysis
```

Problem statement & questions to be answered:

```
## Assuming I work for Motor Trend, a magazine about the automobile industry. Looking at the data s
## They are interested in exploring the relationship between a set of variables and the miles per g
## They are particularly interested in the following two questions:
        ## 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? Was the report done in Rmd (knitr)?

Analysis considerations:

```
Descriptive
        any(is.na)
        head(data)
        str(data)
        summary(data)
Exploratory
        Pairs
        Histograms
        Boxplots
        QQ plots
OLS Ordinary least squares
        General least squares for linear equations
Regression to the mean - Simple linear regression
Statistical linear regression
        Basic - w additive Gaussian error
        Interpretation of regression coefficients (intercept, slope)
        Regression - prediction
Residuals
        Residual variation
        Influence
        Leverage
        Estimate residual variation
        R squared
Regression inference
        Parameters
        Confidence intervals
        Prediction
Multivariate regression analysis
        Linear models
        Two variable simple linear regression (additive) / (multiplicative)
        Summary coefficients
        Fitted values, residuals and residual variation
        Summary coefficients
        Model Adjustment
GLMs
        Linear
        Logistic
        Poisson
        Binary GLMs
                Odds
                Fitting
Poisson
        Count data
Predictive ~ NA
Causal ~ NA
Mechanistic ~ NA
```

Software environment:

```
System - session Info:
```

```r
sessionInfo()
```

```
## R version 3.3.1 (2016-06-21)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X 10.11.6 (El Capitan)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

```
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] magrittr_1.5    formatR_1.4     tools_3.3.1     htmltools_0.3.5
##  [5] yaml_2.1.13     Rcpp_0.12.7     stringi_1.1.1   rmarkdown_1.0
##  [9] knitr_1.14      stringr_1.1.0   digest_0.6.10   evaluate_0.9
```

Accessing data:

Getting the data:

```r
rm(list=ls()); library(UsingR); library(datasets); head(mtcars)
```

```
## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units

##
## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':
##
##     cancer

##                    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
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Raw data overview:

```
    Motor Trend 'mtcars' data set:
```

```
any(is.na(mtcars)); head(mtcars,5)
```

```
## [1] FALSE
```

```
##                    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
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
```

Processing the data:

```
    Transformations;
            1 factor variables 8:11;
            2 change variable labels in columns 8 & 9;
                    a Note; for column header 8 = vs; variable names = V-block, & S-block;
                    b Note; for column header 9 = am; variable names = A-type = A, & M-type;
```

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

```
##                    mpg cyl disp  hp drat    wt  qsec      vs     am gear
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46 V-block M-type    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02 V-block M-type    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61 S-block M-type    4
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44 S-block A-type    3
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02 V-block A-type    3
## Valiant           18.1   6  225 105 2.76 3.460 20.22 S-block A-type    3
##                   carb
## Mazda RX4            4
## Mazda RX4 Wag        4
## Datsun 710           1
## Hornet 4 Drive       1
## Hornet Sportabout    2
## Valiant              1
```

Exploratory analysis:

```
Histograms
Boxplots
Rug
Barplots
Scatterplots
Multiple plots
Graphing - base, lattice, ggpplot2
ABlines (h/v)
Confidence intervals
Standard error
Variance
Fitted lines
Heatmap
K-Means
Dimension Reduction
        PCA
        SVD
Figures: Exploratory see Appendix A
```

Statistical modeling, regression & model fit:

```
Assumptions:
        A
        B
        C
Simple Linear Regression
Multivariate Linear Regression
        lm - simple
        lm - multivariate
        lm - nested
        lm - remove the intercept (-1)
        lm - step function
Coefficients / Slope
Standard Error
T-Vales
pValues
Residuals
        Leverage
        Influence
Confidence Intervals
Residuals
Hatvalues
dfbetas
Influence Measures
Anova
        Chisq
Ancova
GLMs
```

Preliminary findings: quesions of interest: & interpretation of results:

A B C

Conclusions / recommendations:

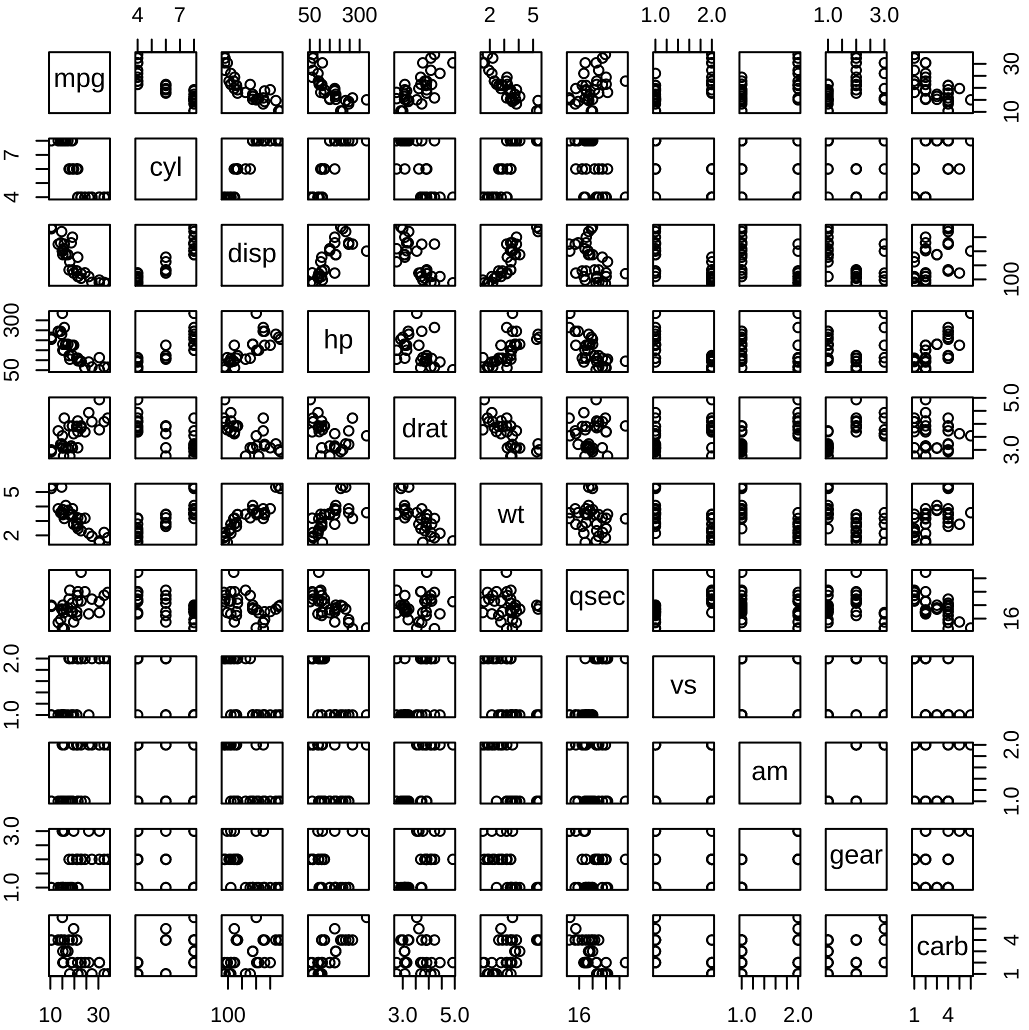
```
A
B
C
        1 Challenge the results ?
        2 Measures of uncertainty 'e'
```

What are some possible alternative analyses?

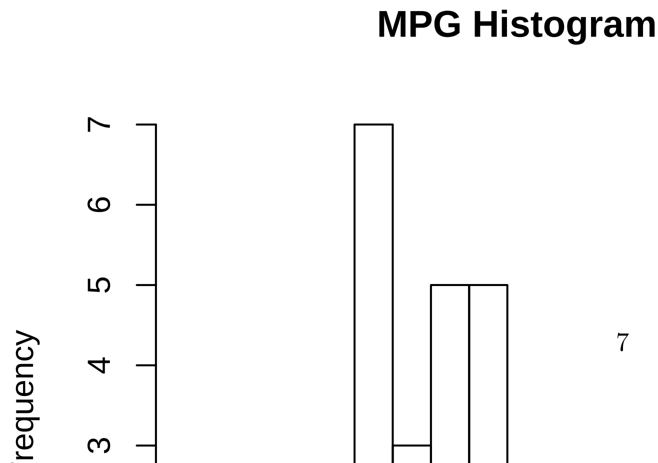
```
A
B
```

Appendix A

#Pairs



#Histograms



```
#Fitted
```

```
#To be inserted
```

```
#Residuals
```

```
#Residuals vs Fitted
```

=== END ===