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:

```
## Assuming I work for Motor Trend, a magazine about the automobile industry. Looking at the data s
## Motor Trend is interested in exploring the relationship between a set of variables and the miles
## 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?

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

III. Analysis Considerations:

```
Descriptive - dim, str, summary
Exploratory - pairs, histograms, boxplots & (multiple plots)
Analysis - Mean, T-Test, Z-Test, covariance, OLS, regression to mean,
simple linear regression, multivariable regression & model selection,
Adjustments, Residuals (predict fit, residual fit (-1)), Variation &
Diagnostics; GLMs & Binary GLMs, counts, correlation, confint, ChiSq-Test, VIF
```

IV. Software Environment:

```
Set working directory:

System - session Info:
```

V. Accessing Data:

```
Getting the data:
```

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

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

```r
any(is.na(mtcars)); head(mtcars, 3)
```

VII. Processing Data:

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

```r
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!!
See Appendix A, Figures 1:4
```

IX. Statistical Modeling, Regression & Model Fit:

```
Assumptions:
        A Correlation exists among multiple variables
        B
        C
```

X.Preliminary Findings:

Questions of Interest: # & Interpretation of Results: A Based on the ANOVA table we can see model 4 is significant in relation to the variable for weight B 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
```

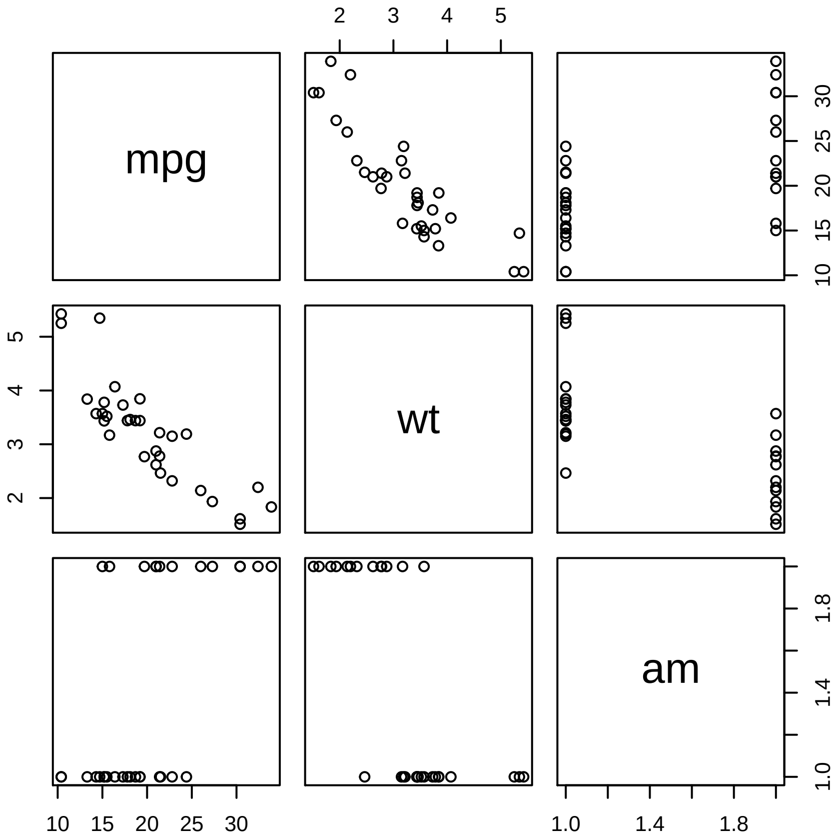
XIII. Are there other alternative analyses?

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

XIV. Appendix A, "Graphical Analysis""

Pairs

```
pairs(mpg ~ ., data = mtcars[,c(1,6,9)])
```

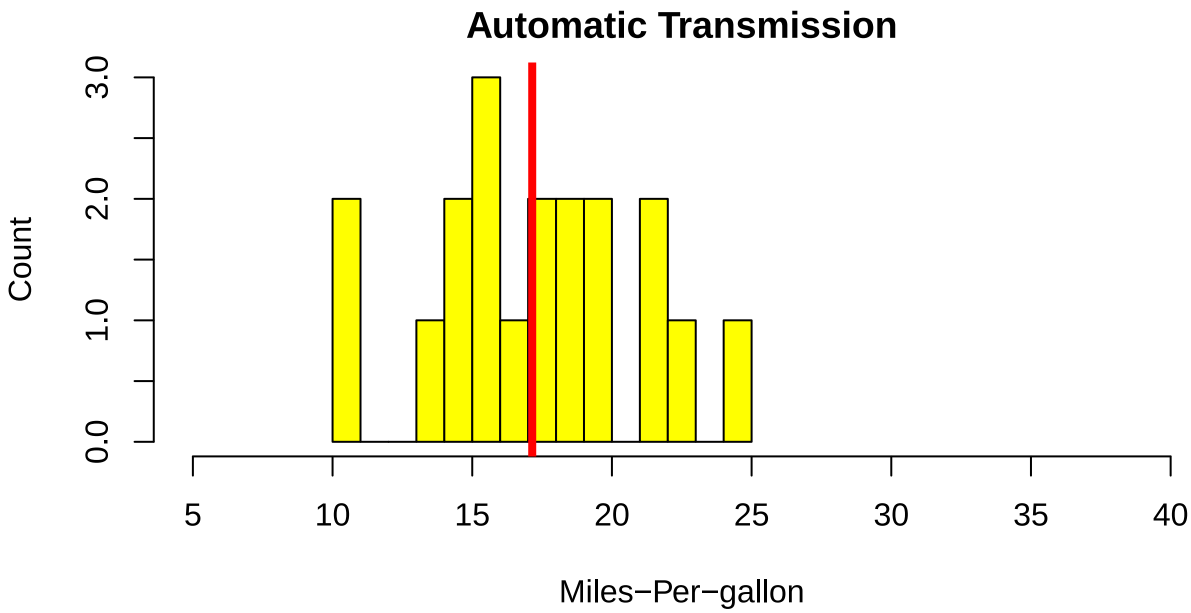


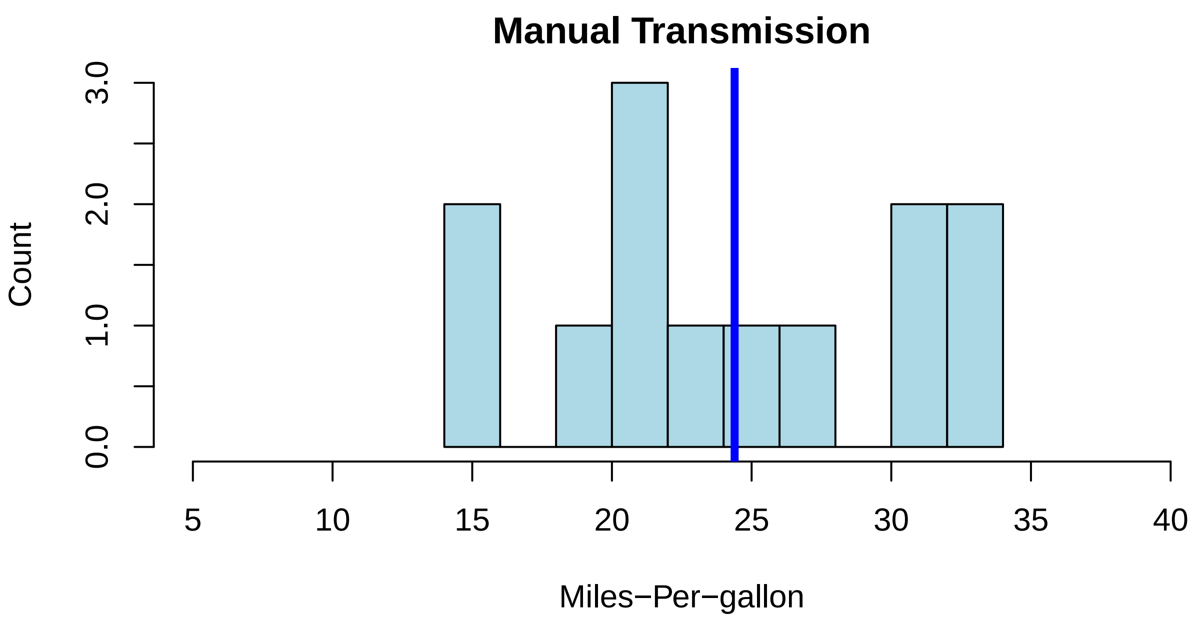
Histograms

```r
data("mtcars");par(mfrow = c(2,1), mar = c(4,4,2,1)) # set margin
mtcars$vs <- factor(mtcars$vs, labels = c("V-block", "S-block")); mtcars$am <- factor(mtcars$am, labels
head(mtcars,3)
```

```
##                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
```

```r
hist(mtcars$mpg[mtcars$am=="Auto"], breaks=10, xlab = "Miles-Per-gallon", ylab = "Count", main = "Automa
abline(v=mean(mtcars$mpg[mtcars$am=="Auto"]), col="red", lwd = 4)
hist(mtcars$mpg[mtcars$am=="Manual"], breaks=10, xlab = "Miles-Per-gallon", ylab = "Count", main = "Manu
abline(v=mean(mtcars$mpg[mtcars$am=="Manual"]), col="blue", lwd = 4)
```



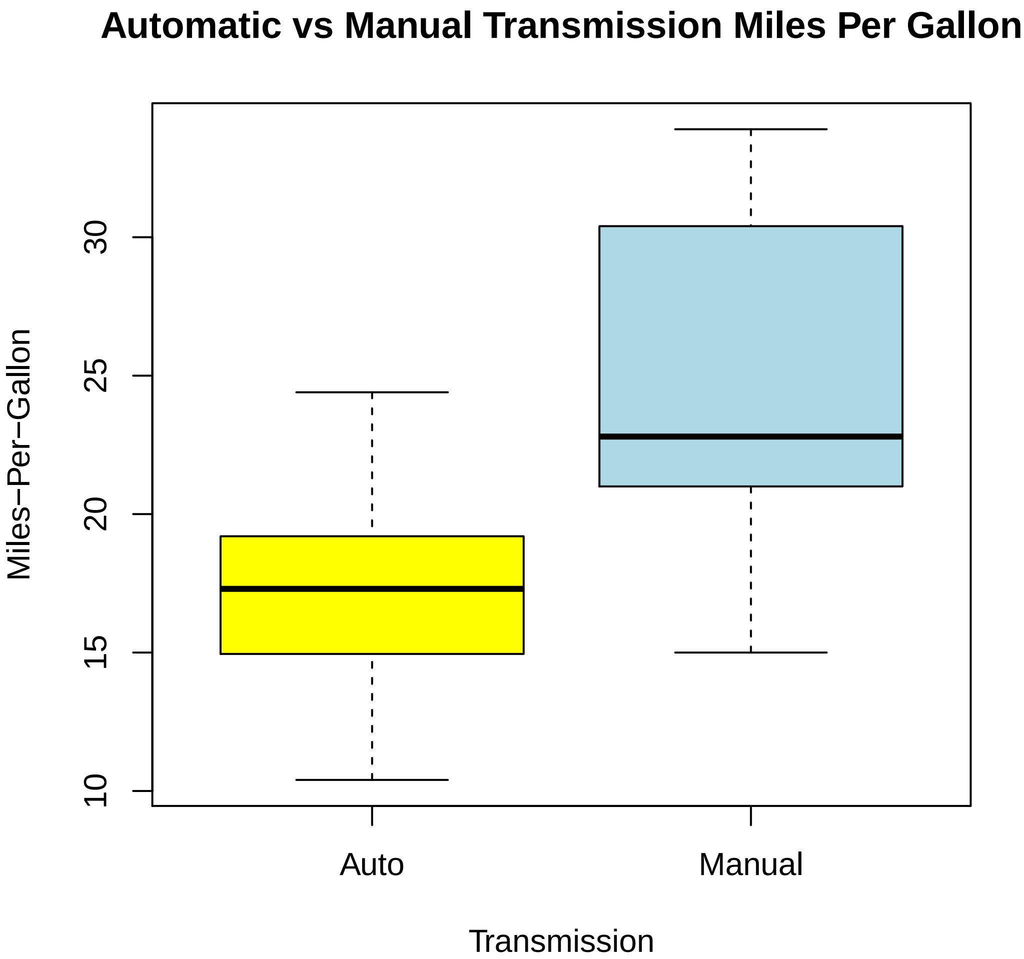


Box Plots

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

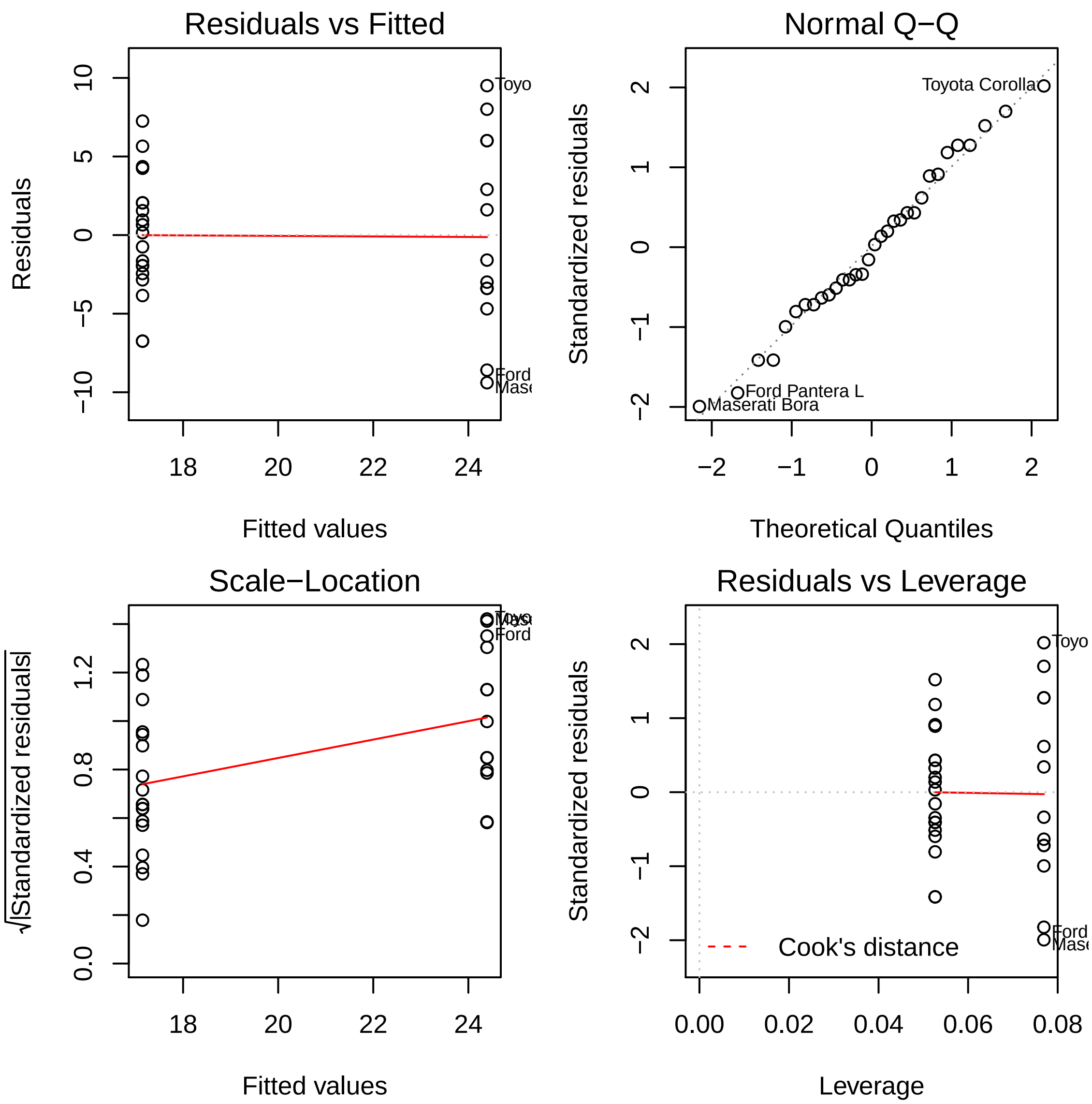
```
##                mpg cyl disp  hp drat    wt  qsec      vs     am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 S-block Manual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

```
boxplot(mpg ~ am, data = mtcars, col = (c("yellow","lightblue")), ylab = "Miles-Per-Gallon", xlab = "Tra
```



Simple Linear Regression Single Variable Plot

```
par(mfrow = c(2, 2), mar = c(4, 5, 2, 1));fslrm <- lm(mpg ~ am, data = mtcars);
coef(summary(fslrm)); plot(fslrm)
```



Bivariate Linear Model Regression plot

```
data("mtcars")
f2 <- lm(mpg ~ factor(am) + wt, data = mtcars);coef(summary(f2))
f3 <- lm(mpg ~ factor(am) * wt, data = mtcars);coef(summary(f3))
```

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

```
par(mfrow = c(2, 2), mar = c(4, 5, 2, 1))
mlr1 <- lm(mpg ~ ., data = mtcars);coef(summary(mlr1))
```

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

```r
par(mfrow = c(2, 2), mar = c(4, 5, 2, 1))
mlr1 <- lm(mpg ~ . -1, data = mtcars);coef(summary(mlr1))
```

Multivariate LM Nested Plot

Generalized Linear Models - na

Binary Generalized Linear Models (Example from course text need to apply to thi project)

=== END ===