

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

james c walmsley

12/1/2016

I. Executive Summary:

Add after completing analysis

II. Problem Statement & Questions to Answer:

Assuming I work for Motor Trend, a magazine about the automobile industry. Looking at the data set
They are interested in exploring the relationship between a set of variables and the miles per gallon
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
Exploratory
Regression to the mean - Simple linear regression
Multivariable regression analysis
Adjustments
Residuals, variation, diagnostics
Multiple variables & model selection
GLMs
Binary GLMs

IV. Software Environment:

System - session Info:

```
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
```

V. Accessing Data:

Getting the data:

VI. Raw Data Overview:

Motor Trend 'mtcars' data set:

```
any(is.na(mtcars)); colnames(mtcars)
```

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

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

VII. Processing 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 = Automatic = A, & Manual = M;
```

```
##           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-eng Manual   4    4  
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-eng Manual   4    4  
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 S-eng Manual   4    1
```

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  
Simple Linear Regression
```

lm - nested or step function

```
lm - remove the intercept (-1) ## ????
```

```
GLM - na
```

```
Binary GLM yes ## Need to revise this code
```

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

XII. Conclusions / Recommendations:

A
B
C

```
1 Challenge the results ?
```

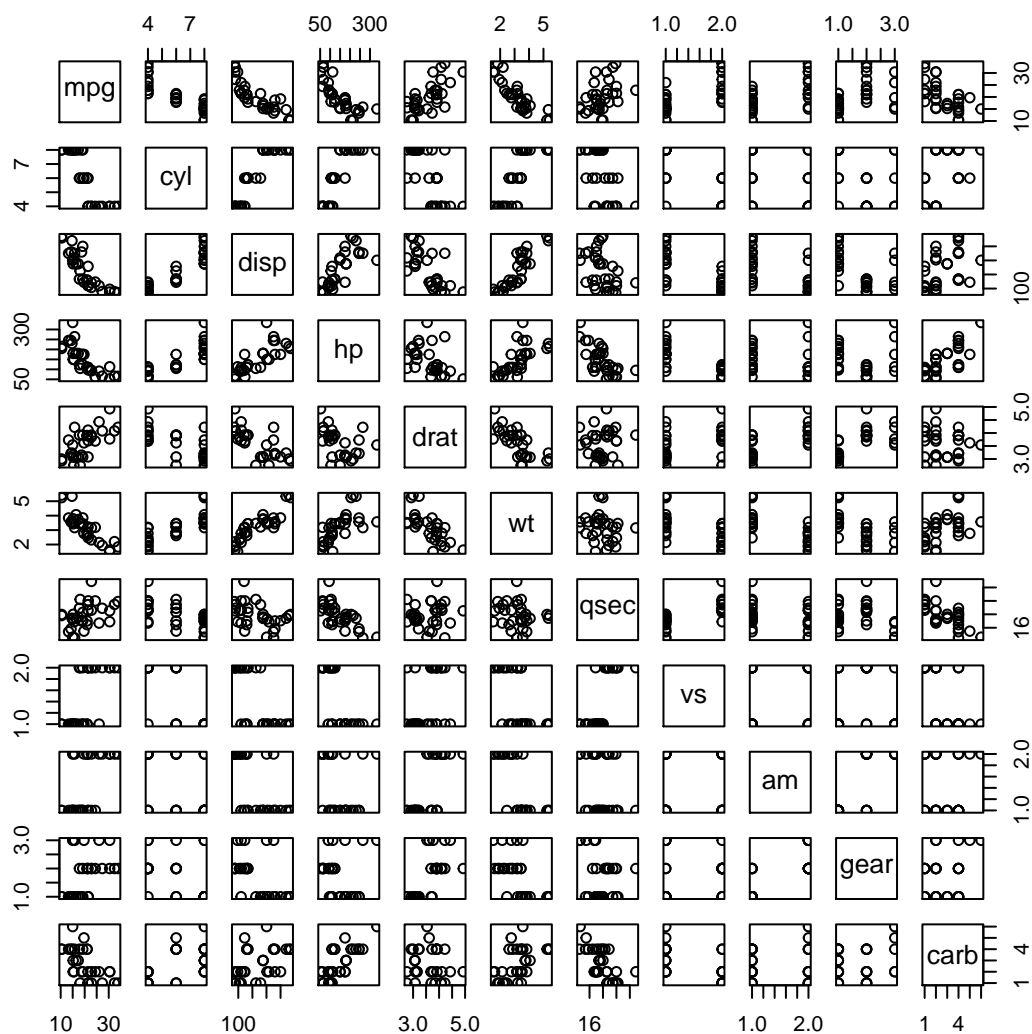
```
2 Measures of uncertainty 'e'
```

XIII. Are there other alternative analyses?

A
B

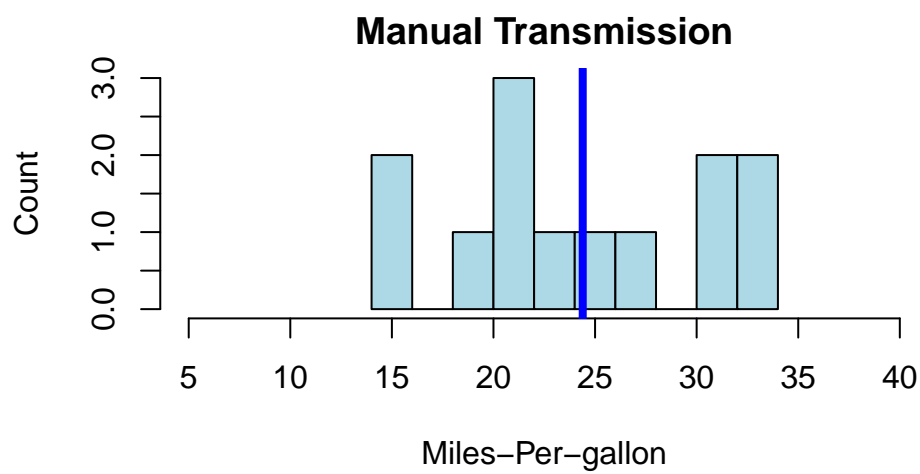
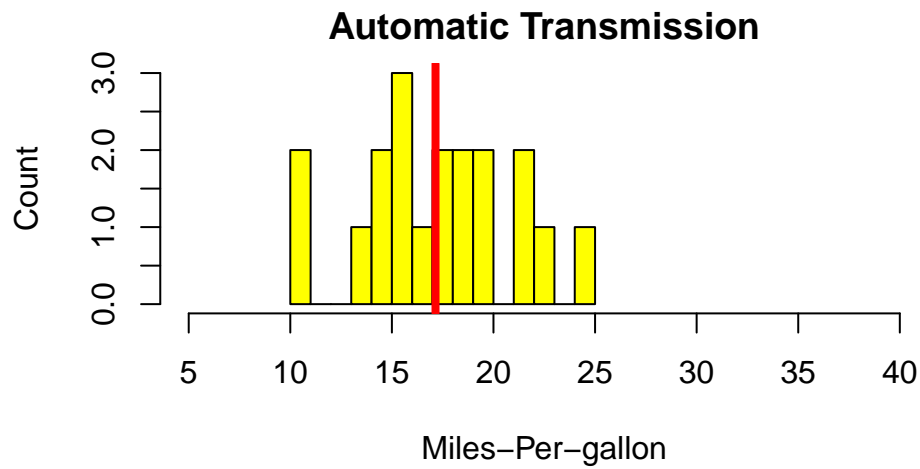
XIV. Appendix A, “Visual Analysis Plots”

Pairs



Histograms

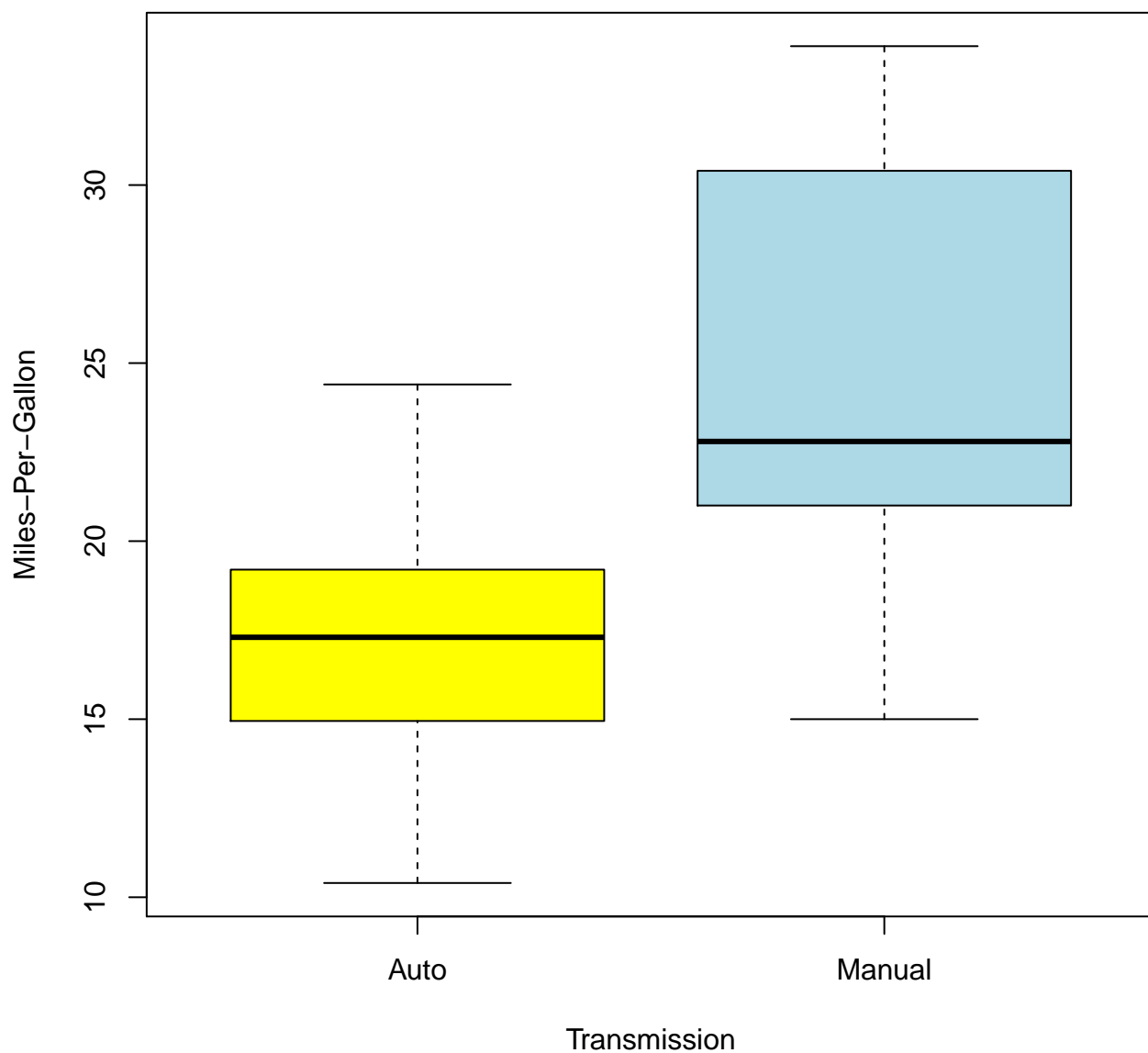
##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	V-block	Manual	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	V-block	Manual	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	S-block	Manual	4	1



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

Automatic vs Manual Transmission Miles Per Gallon



QQ Plots

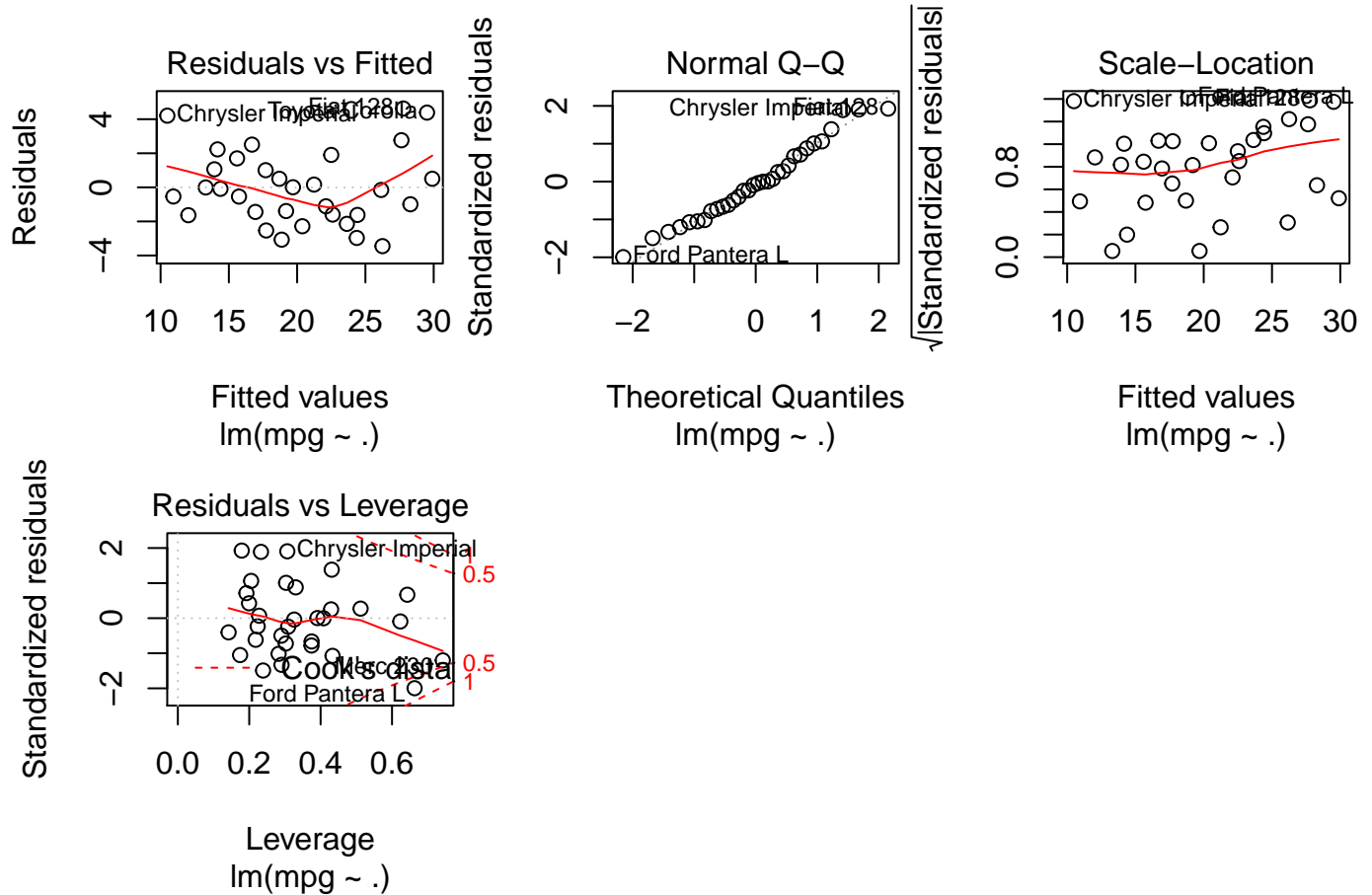
Bivariate Linear Model Regression plot

```
##           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 Linear Model (all vars)-Residuals/Fitted/Residuals vs Fitted

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	12.30337416	18.71788443	0.6573058	0.51812440
## cyl	-0.11144048	1.04502336	-0.1066392	0.91608738
## disp	0.01333524	0.01785750	0.7467585	0.46348865
## hp	-0.02148212	0.02176858	-0.9868407	0.33495531
## drat	0.78711097	1.63537307	0.4813036	0.63527790
## wt	-3.71530393	1.89441430	-1.9611887	0.06325215
## qsec	0.82104075	0.73084480	1.1234133	0.27394127
## vs	0.31776281	2.10450861	0.1509915	0.88142347
## am	2.52022689	2.05665055	1.2254035	0.23398971
## gear	0.65541302	1.49325996	0.4389142	0.66520643
## carb	-0.19941925	0.82875250	-0.2406258	0.81217871



#To be inserted

Residuals plot

Residuals vs Fitted

Generalized Linear Models

Binary Generalized Linear Models

=== END ===